Published in final edited form as:

Multivariate Behav Res. 2015 ; 50(2): 184–196. doi:10.1080/00273171.2014.977433.

Modeling Cyclical Patterns in Daily College Drinking Data with Many Zeroes

David Huh, Debra L. Kaysen, and David C. Atkins

University of Washington

Abstract

Daily college drinking data often have highly skewed distributions with many zeroes and a rising and falling pattern of use across the week. Alcohol researchers have typically relied on statistical models with dummy variables for either the weekend or all days of the week to handle weekly patterns of use. However, weekend vs. weekday categorizations may be too simplistic and saturated dummy variable models too unwieldy, particularly when covariates of weekly patterns are included. In the present study we evaluate the feasibility of cyclical (sine and cosine) covariates in a multilevel hurdle count model for evaluating daily college alcohol use data. Results showed that the cyclical parameterization provided a more parsimonious approach than multiple dummy variables. The number of drinks when drinking had a smoothly rising and falling pattern that was reasonably approximated by cyclical terms, but a saturated set of dummy variables was a better model for the probability of any drinking. Combining cyclical terms and multilevel hurdle models is a useful addition to the data analyst toolkit when modeling longitudinal drinking with high zero counts. However, drinking patterns were not perfectly sinusoidal in the current application, highlighting the need to consider multiple models and carefully evaluate model fit.

Why do people drink to excess? Given the tremendous costs to human lives and society (Hingson, Heeren, Winter, & Wechsler, 2005; Perkins, 2002), research has increasingly focused on reasons and contexts that drive problem drinking. For example, among young college-aged adults, social motives may be a particular impetus to drink (Kuntsche, Knibbe, Gmel, & Engels, 2006; Mohr et al., 2005; Read, Wood, Kahler, Maddock, & Palfai, 2003). The focus on elucidating factors that may explain problematic drinking has driven substance use researchers to pursue intensive longitudinal designs in which participants report their alcohol consumption one or more times a day over a set time frame (e.g., 30 days; Kaysen et al., 2013). Not surprisingly, daily drinking among college students shows a regular pattern over days of the week, with greater drinking on weekends as opposed to weekdays. This systematic pattern has implications for the statistical analysis of daily use data.

The most common strategy for modeling weekly patterns in college drinking studies is to include a dummy variable for weekends, typically defined as Thursday through Saturday, versus weekdays (e.g., Neighbors et al., 2011). Although dummy variable approaches are easy to implement, there are a number of disadvantages. First, single dummy variable

indicators imply an abrupt change across days of the week, whereas drinking data tend to show a smoother transition over days of the week. Alternatively, multiple dummy variables (e.g., Simons, Dvorak, Batien, & Wray, 2010; Simpson, Stappenbeck, Varra, Moore, & Kaysen, 2012) can precisely capture shifts over time and may be useful when differences between specific days are of interest, but are cumbersome when covariates are involved. For example, when six dummy variables are used to represent the days of the week, assessing a covariate effect across time introduces six additional interaction terms.

The current paper introduces an alternative framework for modeling such data, by including cyclical regression covariates (i.e., sine and cosine parameters). Models with cyclical terms capture rising and falling trends over time, which may address the shortcomings of dummy variables through their ability to directly represent periodic patterns while still being more parsimonious relative to saturated dummy variable models. In addition, because alcohol use is a type of count outcome often containing many zeroes (i.e., non-drinking), we illustrate the use of cyclical terms in a type of count regression called a hurdle model (Atkins, Baldwin, Zheng, Gallop, & Neighbors, 2013; Hilbe, 2011), described below.

To date, cyclical models (also referred to as “cosinor models”) have been primarily used in the biomedical literature (e.g., Marler, Gehrman, Martin, & Ancoli-Israel, 2006; Qin & Guo, 2006) and ecology (e.g., Flury & Levri, 1999). Limited applications of cyclical models in the social sciences have included the examination of weekly patterns in sexual behavior (Bodenmann, Atkins, Schär, & Poffet, 2010), mood (e.g., Chow, Grimm, Fujita, & Ram, 2007), seasonal patterns in alcohol use (Uitenbroek, 1996), and primary care office visits following smoking cessation intervention (Land et al., 2012). However, the features of cyclical models may make them well-suited for longitudinal behavioral data, such as alcohol consumption among college students. As illustrated in Figure 1, a rising and falling pattern over time can be represented as a sinusoidal function with regression parameters that define the location (phase) and height (amplitude) of the peak.

This provides a more unified picture of longitudinal patterns than dummy variables, which divide complex patterns into pairwise contrasts at the potential cost of obscuring overall trends. Second, cyclical models can represent rising and falling trends with only a pair of parameters, and given this, they are less demanding to estimate and with fewer parameters to interpret relative to separate dummy variables for individual days. Thus, behavioral outcomes – such as daily college drinking data – that exhibit a regular rising and falling pattern with the day of the week may be reasonably approximated using a sinusoidal function.

As mentioned, a key characteristic of daily drinking data is its distribution. Figure 2a presents daily (total) drinks from Project DASH (Kaysen et al., 2013), a study of 176 heavy drinking female college students who were followed as part of a larger study on PTSD and drinking. Count outcomes, such as daily drinking data, are most appropriately modeled with discrete data distributions such as the Poisson or negative binomial. However, as is common with count outcomes, the distribution of daily drinks in this college population is strongly skewed with a high percentage of zeroes. Although count regression methods are seeing increasing use by substance use researchers (see, e.g., Neal & Simons, 2007), Poisson and

negative binomial models may not provide optimal fits to data such as those in Figure 2a, in which a large number of zeroes is quite distinct from the non-zero portion of the outcome (Atkins et al., 2013; Hilbe, 2011). In such cases the outcome is considered “zero-altered” relative to an ordinary count distribution.¹

On a substantive level, the zeroes may represent a key feature of the data rather than an ignorable nuisance. For example, the transition into a behavior, such as the decision to have the first drink, may be a distinct process from the degree to which one engages in that behavior once it starts (i.e., how much one drinks once drinking has commenced). Consequently, regression models that underrepresent the actual frequency of zeroes in such instances may be flawed both substantively and statistically, since the zeroes may represent an important aspect of the phenomena of interest. One type of two-part model, a hurdle model, assumes that a threshold must be crossed from zero into positive counts. As illustrated with the DASH data in Figure 2b, the drinking outcome is effectively divided into two outcomes, each modeled in its own regression equation. One outcome is a dichotomous variable representing zero drinks vs. any drinking and includes the entire sample. The second outcome, highlighted in grey, represents the amount of drinks when drinking. Thus, a hurdle model contains two sub-models: 1) a logistic regression for zero vs. not zero and 2) a zero-truncated count model (either Poisson or negative binomial) for non-zero counts (Atkins et al., 2013; Hilbe, 2011). In recent years, hurdle models have been increasingly applied in the psychological sciences to behavioral outcomes including substance use (e.g., Kaysen et al., 2013), gambling (e.g., Humphreys, Lee, & Soebbing, 2010), and health care utilization (e.g., Bethell, Rhodes, Bondy, Lou, & Guttman, 2010).

The advantage of hurdle models is that they can capture potential differences in the processes that generate zero versus positive counts. For example, the point at which the probability of drinking is highest may fall on a different day than the when the quantity of drinking is at its highest. Furthermore, the variables that predict whether someone drinks at all may be different than those that predict how many drinks are consumed once drinking has started. Although we do not focus on it in the present article, a common alternative to the hurdle model that is a zero-inflated model. A zero-inflated model is a mixture model and assumes that there are two types of zeroes: zero counts and “excess zeroes” above and beyond what would be predicted by a count distribution. Hence, the logistic regression in a zero-inflated model predicts these excess zeroes, and consequently, the count model includes zeroes (and is not zero-truncated). The advantage of hurdle models is that they are easier to interpret because all of the zeroes are handled in the logistic regression, resulting in a clean distinction between zeroes and non-zero counts. Additionally, because the two component regressions can be fit separately, hurdle models are easier to estimate (Hilbe, 2011).

In principle, the combination of a multilevel hurdle model with cyclical covariates would address two of the key features inherent in daily college drinking data. However, key

¹Typically, “zero-altered” means “zero-inflated” in that there are excess zeroes beyond what a count distribution predicts; however, the hurdle model in particular can fit instances of zero-deflation in which there are fewer zeroes than predicted by a count distribution. Thus, we use the more general term “zero-altered” to refer to instances in which the count of zeroes is different from what a count distribution would predict.

methodological questions exist as to how to combine hurdle models and cyclical regression terms. Because the logistic and zero-truncated count portions of a hurdle model are effectively separate regressions, a combination of cyclical and/or dummy variables can be used across the two parts of the model. Is it appropriate to model trends in both portions of a hurdle model with cyclical parameters? In the context of the DASH study, descriptive plots indicate that the average proportion of drinking and the number of drinks when drinking (Figure 3) each exhibit a rising and falling pattern. This suggests that cyclical parameters are a reasonable candidate for representing both the presence and degree of alcohol consumption among heavy-drinking college women.

The objective of the present study is two-fold. The first is to introduce cyclical models and their use to psychological researchers. Cyclical models may be well-suited for many types of longitudinal data that appear in the psychological sciences, but are under-utilized because of the default emphasis on linear models. The second is to illustrate their feasibility in conjunction with hurdle models to evaluate a longitudinal alcohol outcome and to test a moderator of use over time: social drinking motives. Drinking motives in general have been highly studied and there is a large literature on their association with drinking severity (e.g., Kuntsche et al., 2006; Mohr et al., 2005; Read et al., 2003). Moreover, previous research has supported a prospective association between social motives, one type of drinking motive, and elevated weekend drinking behavior (Mezquita et al., 2011). Consequently, we chose social drinking motives to use in the present illustration to compare the cyclical, weekend dummy variable, and saturated dummy variable approaches. Although our primary aim is to show the utility of cyclical models for alcohol use data, we do note that no one has previously examined the association in daily alcohol data using a count regression approach.

Methods

Participants and Procedures

The present study is a secondary analysis of data from 176 female undergraduates (136 with a past history of sexual victimization and 40 with no past trauma history) over the age of 18 who were followed as part of a larger study on the association of PTSD symptoms and drinking (Kaysen et al., 2013). Study inclusion criteria were 1) consuming 4 or more drinks on one occasion at least twice in the past month, and 2) reporting either no history of trauma exposure *or* reporting sexual victimization (at least one incidence of childhood sexual abuse *or* one incidence of adult sexual assault that was not within the past three months), and 3) for those with a history of sexual victimization at least one intrusive and one hyperarousal PTSD symptom in the past month. The study included a baseline assessment followed by a 30-day monitoring period. PDA assessments took four minutes to complete on average and included questions on context, affect, PTSD symptoms, and alcohol use and related problems. For this secondary analysis, we focused on the 172 participants who provided data on (1) baseline self-reported social enhancement motives and (2) daily drinking during the 30 days of monitoring.

Measures

Social drinking motives—Five items from the Drinking Motives Questionnaire-Revised (Grant, Stewart, O'Connor, Blackwell, & Conrod, 2007) were used to assess social drinking motives (e.g., “Because it is what most of my friends do when we get together.”), and participants were asked to indicate the degree to which they drank for each of those reasons. Response options ranged from 1 = *never/almost never* to 5 = *almost always/always*. Items were averaged to form a composite of social drinking motives with higher scores reflecting greater endorsement of social motives for drinking. These composite scores were then standardized by centering the scores at the mean and dividing by the standard deviation. Cronbach’s alpha was .81 for the social motives subscale in the present study.

Daily alcohol consumption—Questions regarding alcohol focused on the participants’ alcohol use for the prior 24 hours. Participants were asked, “How many standard drinks have you had in the past 24 hours?” along with a definition of a standard drink described during the baseline assessment. Standard drinks could be entered into the PDA. If the participant did not consume alcohol, they could either type in 0 or select the response “I did not drink.”

Statistical Analyses

Multilevel hurdle negative binomial modeling was utilized for all statistical analyses. The hurdle models were estimated in two parts: (1) a logistic regression to model zeros versus non-zeros (i.e., no drinking vs. any drinking) and (2) a zero-truncated negative binomial regression² to model the subset of the data that are positive counts (i.e., amount of drinking when there is any drinking). Because the hurdle model clearly classifies the data as zero or non-zero, the overall likelihood can be factored into two pieces and fit separately, corresponding to the two sub-models (Hilbe, 2011).³ This is the strategy taken in the present analyses and the reason why hurdle models are easier to fit relative to zero-inflated models, in which the joint likelihood must be estimated together. The analyses were performed using maximum likelihood estimation in R v3.0.3 (R Core Team, 2014) with v0.7.7 of the glmmADMB package (Skaug, Fournier, Nielsen, Magnusson, & Bolker, 2013). Supplemental computer code for fitting the models in R and SAS, along with example data, is available online (<https://github.com/davidhuh/cyclicalmodels>).

To model whether or not a person drank on a particular day, we made use of multilevel logistic regression. Let $\Pr[\text{DRINKS}_{ti} > 0]$ be the probability of individual i drinking one or more drinks on day t , and $\Pr[\text{DRINKS}_{ti} = 0]$ be the probability of individual i not drinking on day t . Thus, the following equation depicts a basic multilevel logistic model without moderators of time:

$$\log \left(\frac{\Pr[\text{DRINKS}_{ti} > 0]}{\Pr[\text{DRINKS}_{ti} = 0]} \right) = \beta_{0(B)} + \beta_{1(B)} \text{TIME}_{ti} + r_{0i(B)} + r_{1i(B)} \text{TIME}_{ti} \quad (1)$$

²There are two common formulations of negative binomial models, which define either a linear (NB1) or quadratic (NB2) relationship between the expected variance and mean outcome (Cameron & Trivedi, 1998; Hilbe, 2011). For the present analyses we utilized the NB1 formulation for the truncated negative binomial model because of better fit to the data.

³In theory, hurdle models with random effects could allow correlations between the random effects variance terms of the two sub-models. Thus far, we are unaware of examples of this formulation of a hurdle mixed model in the applied literature.

where i indexes individuals, t indexes time in days, and (B) identifies regression coefficients from the logistic model. The outcome is modeled as the natural logarithm of the odds (i.e., logit function) of an individual i drinking versus not drinking on day t , which constrains predictions to between 0 and 1. The predictor “TIME _{ti} ” is a placeholder for the slope of drinking over time, which is replaced by the three different parameterizations of time described in the next section. The average trajectory for the sample is defined by the intercept and slope of the fixed effects (i.e., $\beta_{0(B)} + \beta_{1(B)} \text{TIME}_{ti}$). The random intercept ($r_{0i(B)}$) represents deviations of individuals from the group average on the outcome whereas the random slope ($r_{1i(B)}$) represents deviations of individuals from the group average slope of time.

For days in which a person did drink, the number of drinks consumed was modeled using multilevel zero-truncated negative binomial regression. Let $E[\text{DRINKS}_{ti} \mid \text{DRINKS}_{ti} > 0]$ be the expected number of drinks of individual i on a particular drinking day t . Thus, the following equation depicts a basic multilevel zero-truncated negative binomial model without moderators of time:

$$\log(E[\text{DRINKS}_{ti} \mid \text{DRINKS}_{ti} > 0]) = \beta_{0(C)} + \beta_{1(C)} \text{TIME}_{ti} + r_{0i(C)} + r_{1i(C)} \text{TIME}_{ti} \quad (2)$$

where (C) identifies the regression coefficients from the count model. The outcome is modeled as the natural logarithm of the expected number of drinks for individual i on a particular drinking day t , which constrains predictions to positive counts greater than or equal to 1. The parameters for the zero-truncated negative binomial regression mirror those of the logistic regression.

Both the logistic and zero-truncated negative binomial models can be extended to include additional covariates and their interactions with time, to assess variables that predict patterns of drinking over time (i.e., moderators), which we describe later. We evaluated three parameterizations of time in the logistic and zero-truncated negative binomial regression models: (1) a single dummy code for weekend vs. weekday (2) separate dummy codes for each day of the week, and (3) sine (phase or peak location) and cosine (amplitude or peak magnitude) terms.

Parameterizations of time—To derive the WEEKEND dummy variable, we coded the day of the week as 0 = *weekday* (Sunday to Wednesday) and 1 = *weekend* (Thursday to Saturday). In the present college sample, drinking on Thursdays was more similar to that on Fridays and Saturdays than the other days of the week, which informed the decision to treat it as part of the weekend. Equation 3 is the zero-truncated negative binomial portion of the hurdle model with a weekend dummy variable predictor for time.

$$\log(E[\text{DRINKS}_{ti} \mid \text{DRINKS}_{ti} > 0]) = \beta_{0(C)} + \beta_{1(C)} \text{WEEKEND}_{ti} + r_{0i(C)} + r_{1i(C)} \text{WEEKEND}_{ti} \quad (3)$$

For the saturated dummy parameterization, we derived a set of six categorical contrasts of day of the week (Tuesday through Sunday). The excluded category (Monday) is the reference day whose mean is represented by the intercept ($\beta_{0(B)}$, $\beta_{0(C)}$) of the regression equation. Figure 1 indicated that both the probability of drinking and quantity of drinking

when drinking where lowest on Monday. Consequently, using Monday as the reference category permitted a contrast of the lowest and highest drinking days (i.e., Saturday). The choice of reference category does not alter the overall model fit when a saturated set of dummy codes is used. Equation 4 is the zero-truncated negative binomial portion of the hurdle with a saturated set of dummy variables for time.

$$\log(E[\text{DRINKS}_{ti} | \text{DRINKS}_{ti} > 0]) = \beta_{0(C)} + \beta_{1(C)} \text{TUE}_{ti} + \beta_{2(C)} \text{WED}_{ti} + \beta_{3(C)} \text{THU}_{ti} + \beta_{4(C)} \text{FRI}_{ti} + \beta_{5(C)} \text{SAT}_{ti} + \beta_{6(C)} \text{SUN}_{ti} + r_{0i(C)} + r_{1i(C)} \text{TUE}_{ti} + r_{2i(C)} \text{WED}_{ti} + r_{3i(C)} \text{THU}_{ti} + r_{4i(C)} \text{FRI}_{ti} + r_{5i(C)} \text{SAT}_{ti} + r_{6i(C)} \text{SUN}_{ti} \quad (4)$$

For the cyclical model, a sinusoidal function of time was modeled using the approach described by Fluri and Levri (1999). Equation 5 is the zero-truncated negative binomial portion of the hurdle model with cosine and sine transformed versions of the time variable. DAY_{ti} is a linear predictor for time coded 0 = *Monday* to 6 = *Sunday* for each day of the week.

$$\begin{aligned} \log(E[\text{DRINKS}_{ti} | \text{DRINKS}_{ti} > 0]) &= \beta_{0(C)} + \beta_{1(C)} \cos\left(\frac{2\pi}{7} \text{DAY}_{ti}\right) \\ &+ \beta_{2(C)} \sin\left(\frac{2\pi}{7} \text{DAY}_{ti}\right) \\ &+ r_{0i(C)} + r_{1i(C)} \cos\left(\frac{2\pi}{7} \text{DAY}_{ti}\right) \\ &+ r_{2i(C)} \sin\left(\frac{2\pi}{7} \text{DAY}_{ti}\right) \end{aligned} \quad (5)$$

Note that the DAY_{ti} variable is multiplied by a constant ($2\pi / 7$), where 7 is the period of the cycle and π is the mathematical constant π equal to approximately 3.14159. Separate cosine and sine functions are then applied to generate two predictors of time. The regression coefficients estimated from these two predictors have the following interpretation: β_1 is the amplitude or the height of the wave from midpoint to peak, and β_2 is the phase or the location of the peak of the wave from the origin (i.e., Monday).

Baseline models—Multilevel logistic and zero-truncated negative binomial regressions were fit separately with a weekend dummy variable, cyclical terms, and saturated set of dummy variables. These baseline regressions were first estimated with random intercepts, allowing participants to vary in their average probability of any drinking (logistic sub-model) and quantity of drinks when drinking (zero-truncated negative binomial sub-model). The necessity of random effect slopes was assessed via χ^2 likelihood ratio (deviance) tests.⁴ If the random slope terms were significant, they were retained in the final baseline model.

⁴A χ^2 test was used to evaluate whether there was non-zero variation of the random slopes in the logistic and zero-truncated negative binomial models. Stoel (2007) notes that traditional likelihood ratio tests tend to be underpowered when it involves fixing the variance of a random effect (e.g., a random slope) to zero, which is the case when testing against a model with random intercepts only. However, the more sensitive test would not likely change the conclusions since the traditional approach was sufficient to identify variation in the random slopes.

Moderation models—The main effect of social drinking motives and its interaction with time was added to baseline models to evaluate moderation effects of time under each time parameterization. The random effects structure used in the baseline models was retained. Equation 5 illustrates a cyclical zero-truncated negative binomial regression portion of a hurdle model with a single baseline covariate (i.e., moderator) of time. SDMOTIVES_{0i} represents the social drinking motives at baseline ($t = 0$) for participant i .

$$\begin{aligned} \log(E[\text{DRINKS}_{ti}] | \text{DRINKS}_{ti} > 0) = & \beta_{00(C)} + \beta_{01(C)} \text{SDMOTIVES}_{0i} + \\ & \beta_{10(C)} \cos\left(\frac{2\pi}{7} \text{DAY}_{ti}\right) + \beta_{11(C)} \left(\cos\left(\frac{2\pi}{7} \text{DAY}_{ti}\right) \times \text{SDMOTIVES}_{0i}\right) + \\ & \beta_{20(C)} \sin\left(\frac{2\pi}{7} \text{DAY}_{ti}\right) + \beta_{21(C)} \left(\sin\left(\frac{2\pi}{7} \text{DAY}_{ti}\right) \times \text{SDMOTIVES}_{0i}\right) + \\ & r_{0i(C)} + r_{1i(C)} \cos\left(\frac{2\pi}{7} \text{DAY}_{ti}\right) + r_{2i(C)} \sin\left(\frac{2\pi}{7} \text{DAY}_{ti}\right) \end{aligned} \quad (5)$$

For the cyclical and saturated dummy variable parameterizations where time was represented by multiple terms, interactions of social motives with all slope terms of time were included.

Model comparison—The weekend dummy variable, cyclical, and saturated dummy variable approaches were evaluated across several dimensions, including fit to the data, parsimony, ease of use, and interpretation. Each of the approaches was evaluated in the logistic and zero-truncated negative binomial regressions to identify the best-fitting parameterization in each.

Since the cyclical and dummy variable specifications are not subsets of each other, the Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) were used for model comparison, which are appropriate for comparisons of non-nested models.⁵ For each model, comparisons were made within sub-model (i.e., logistic and zero-truncated negative binomial) as well as overall model. Theoretically, the BIC statistic represents the probability that a given model out of the possible models is the true one (Raftery, 1995). In contrast, the AIC statistic assesses the ability of the model to predict new data (Kuha, 2004). The practical difference between the two rests in the degree to which each penalizes extra parameters; the BIC tends to favor more parsimonious models relative to the AIC.⁶ We follow the general recommendation of evaluating model fit using both the BIC and AIC (O'Connell & McCoach, 2008). Per Raftery (1995), a difference of 2 to 5 reflects “weak” evidence for the lower BIC model, 6 to 10 reflects “strong” evidence, and greater than 10 reflects “very strong” evidence.⁷ Burnham and Anderson (2004) proposed similar cutoffs

⁵The formulas for the BIC and AIC were as follows:

$$\text{BIC} = \text{Deviance} + \text{Number of parameters} \times \ln(\text{Number of observations})$$

$$\text{AIC} = \text{Deviance} + 2 \times \text{Number of parameters}$$

⁶The calculation of BIC and AIC in multilevel models is less straightforward than in cross-sectional analyses since the determination of the number of parameters and sample size (i.e., degrees of freedom) is less clear due to correlated observations (Gelman & Hill, 2006). For consistency, we computed the number of parameters in each regression equation as the total count of fixed and random effects. For the calculation of BIC, there is no accepted standard regarding basing the sample size on the number of individuals or the number of observations (O'Connell & McCoach, 2008). Consequently, we used the number of observations, since this is the default in R as well as most commercial statistical packages.

for comparing AIC values with differences of 3 to 7 reflecting “considerably less support” for the higher AIC model and differences of 10 or higher reflecting “essentially no support” For both BIC and AIC, lower values correspond with better fit.

Results

Assessing Random Slopes for Time

Prior to comparing the different parameterizations of time, we examined the necessity of including random effects for time in each of the hurdle sub-models. The saturated dummy variable specification could not be evaluated under a full set of random effects in either the logistic or zero-truncated negative binomial sub-models as a model of that complexity could not be estimated using maximum likelihood. Given this limitation, we evaluated a simple linear random slope of time in models with a saturated set of dummy variables. The zero-truncated negative binomial regressions with random slopes evidenced the best fit in all three specifications of time based on χ^2 likelihood ratio tests, so random slopes were included in the count portion of the hurdle models. The logistic regression with random slopes for time did not evidence improvement over a model with only random intercepts for any of the parameterizations of time, so only random intercepts were included in the logistic portion of the hurdle models.

Model Comparisons

Table 1 summarizes the BIC and AIC fit statistics for the logistic and zero-truncated negative binomial portions of the hurdle model for each of the three specifications of time. Fit statistics are reported separately for each of the component regressions with the best-fitting specification in each portion of the hurdle model noted. An estimate of overall fit is provided when the same specification of time is applied in both portions of the hurdle model. An additional estimate of overall fit is provided for the hurdle model combining the best-fitting specification of time in the logistic regression with the best-fitting specification of time in the zero-truncated negative binomial regression.

Baseline models—By sub-model (i.e., logistic vs. zero-truncated negative binomial), there was “very strong” evidence that the logistic regression with full dummy variables was a better model of any drinking vs. no drinking than both the weekend (BIC = 41) and cyclical (BIC = 50) logistic models. There was strong evidence the weekend logistic model was a better model than the cyclical logistic model (BIC = 9). In contrast, there was very strong evidence that the zero-truncated negative binomial regression with cyclical terms was a better model of the amount of drinking when drinking than the weekend (BIC = 49) and saturated dummy variable (BIC = 59) specifications. There was strong evidence that the weekend zero-truncated negative binomial model was a better model than the saturated dummy variable specification (BIC = 10). The overall best-fitting model combined a logistic regression with a saturated set of dummy variables and a zero-truncated negative binomial regression with cyclical terms. The best-fitting models according to the AIC statistics were the same as for the BIC.

⁷Under Raftery’s (1995) guidelines, a 6-point difference on BIC is analogous to a P-value of 0.05.

Moderation models—The pattern of BIC and AIC statistics in the moderation models was comparable to the baseline models. As with the baseline models, both the BIC and AIC favored the cyclical terms in the zero-truncated negative binomial regression, but preferred a saturated set of dummy variables in the logistic regression.

Model Predictions

Figures 4 and 5 include regression tables and plots of the point estimates and confidence intervals summarizing the coefficient estimates from the baseline and moderation models, respectively.

Baseline models—The logistic and zero-truncated negative binomial sub-models with weekend, cyclical, and saturated dummy variable specifications each identified statistically significant change across time in the probability of any drinking and the expected number of drinks when drinking, respectively. The portrayal of the time effect varied for each of the specifications. With a weekend dummy variable, the time effect was characterized as a Thursday to Saturday elevation in both the probability of drinking (logistic sub-model) and the number of drinks when drinking (zero-truncated negative binomial sub-model). With the cyclical terms, the amplitude (cosine) and phase parameters identified a significant peak in both the probability (logistic sub-model) and quantity of drinking (zero-truncated negative binomial sub-model). With the saturated dummy variable specification, the trend was detected as elevations in the probability of drinking on 4 days (Tuesday and Thursday-Saturday) compared with Monday and in the quantity of drinks when drinking on 6 days (Tuesday-Sunday) compared with Monday.

Moderation models—The addition of the social drinking motives moderator of time increased the number of fixed effects in the weekend dummy, cyclical, and saturated dummy models by 2, 4, and 12, respectively. The logistic and zero-truncated negative binomial sub-models with weekend, cyclical, and saturated dummy variable predictors all detected statistically significant moderation of the probability of any drinking and the expected number of drinks when drinking, respectively, by baseline social drinking motives. As a way to interpret each of the models but also visually contrast the predictions from each model, Figure 6 plots the predictions made by each of the time specifications in each portion of the hurdle model. The portrayal of the how the time effect was moderated by social motives varied according to how time was parameterized.

Moderation in the logistic sub-model with a weekend dummy variable was detected as a slightly greater rise from weekday to weekend in the probability of any drinking among individuals with higher social drinking motives. In the logistic sub-model with cyclical terms, the moderation finding was detected as a flatter peak in drinking at higher baseline social motives, indicating a more similar probability of any drinking across days of the week among those with higher motives. In the logistic sub-model with a saturated set of dummy variables, the moderation finding was detected as greater pairwise elevations in the probability of drinking on 2 days (Thursday and Friday) compared with Monday among individuals with higher social drinking motives. The poorer performance of the logistic sub-model with cyclical terms based on the BIC and AIC statistics coincided with an apparent

divergence in late week drinking predictions in the cyclical vs. the saturated dummy variable models seen in the two upper left panels of Figure 6. Specifically, the cyclical terms under-predicted the probability of any drinking on Fridays and Saturdays and over-predicted drinking probability on Sundays.

Moderation in the zero-truncated negative binomial sub-model with a weekend dummy variable was detected as a slightly smaller increase in the amount of drinking from weekday to weekend among individuals with higher social drinking motives. Relatedly, drinkers with high social motives drank more on average than those with low social motives. In the zero-truncated negative binomial regression with cyclical terms, the moderation finding was reflected in a more flattened pattern of drinking across days of the week indicating greater consistency in the amount of drinking during the week among those with higher motives. In the zero-truncated negative binomial regression with a saturated set of dummy variables, the moderation finding was detected as greater pairwise elevations in the quantity of drinks when drinking on 2 days (Tuesday and Sunday) compared with Monday among individuals with higher social drinking motives. There was an apparent divergence in early-week drinking predictions under the cyclical vs. the saturated dummy variable specifications of time as seen in the two lower left panels of Figure 6. The cyclical parameterization over-predicted the number of drinks when drinking on Mondays among individuals with higher social drinking motives.

Generally, the predictions from the saturated dummy variable models were less precise, as evidenced by larger confidence intervals than the cyclical and weekend dummy variable models.

Discussion

Intensive longitudinal designs have become increasingly common in the substance use literature and more generally in the psychological sciences. With respect to college drinking data, the most common approach to modeling non-linear patterns over time has been the inclusion of a dummy variable for weekend versus weekday. In the present article we demonstrate that this default approach is not always optimal since it fails to account for the continuous rise and fall in alcohol consumption across days of the week. We illustrated a practical alternative using cyclical regression covariates that directly models rising and falling patterns as a sinusoidal function. To date this approach has been confined primarily to analyzing circadian and seasonal patterns in biological processes. However, cyclical patterns also occur in behavioral outcomes such as alcohol use (e.g., Uitenbroek, 1996).

In the present study we compared cyclical regression covariates against two common approaches to modeling drinking data: a weekend vs. weekday dummy variable and multiple dummy variables for each day of the week. A key strength of this study is the use of real data from an intensive longitudinal study of drinking in a high-risk sample of college women. In the context of alcohol use outcomes, an added challenge is a high proportion of zeroes, which are frequently ignored in statistical analyses. Therefore, we implemented each of the three time specifications in a multilevel hurdle model that divided the zeroes and positive counts into separate outcomes. The best-fitting longitudinal model of drinking in the present study combined a full set of dummy variables for each day of the week to predict the

probability of drinking in concert with cyclical terms to predict the quantity of drinking when drinking.

Notably, the cyclical terms represented the rise and fall in drinking quantity well, while either of the dummy variable approaches better represented time trends in the probability of any drinking. Thus not all aspects of drinking followed a sinusoidal pattern. Specifically, the number of drinks when drinking had an intraweek pattern that was approximately sinusoidal, but the superiority of the weekend dummy variable with respect to predicting the probability of drinking suggests that there was some homogeneity in the probability of drinking during weekdays and weekends rather than a continuous rise and fall.

Next, we extended each of the main effect models of time to include social drinking motives as a moderator of daily drinking. All three time specifications identified statistically significant moderation by social drinking motives, both in the probability of any drinking and the number of drinks when drinking. At a substantive level, it is of interest that social motives were significant moderators of drinking and in the cyclical relationship of drinking across the week among college women, the majority of whom had experienced sexual victimization and were experiencing at least some PTSD symptoms. Relatively few studies have explored the role of social motives among trauma exposed samples, focusing instead on the role of coping motives, and to a lesser extent, enhancement motives (Dixon, Leen-Feldner, Ham, Feldner, & Lewis, 2009; Kaysen et al., 2007; Lehavot, Stappenbeck, Luterek, Kaysen, & Simpson, 2013; Stappenbeck, Bedard-Gilligan, Lee, & Kaysen, 2013). These findings suggest that social motives may be an important moderator of drinking broadly in college women.

Although representing each day of the week as a dummy variable provided accurate estimates of drinking on each day, there were disadvantages. One drawback was the inability to estimate a complete set of random slopes in the saturated model, which required modifying the model with a simple linear random slope. Another potential disadvantage with a saturated set of dummy variables is the division of time into a multiplicity of contrasts. Separate dummy variables may be appropriate when differences between specific days are of interest. However, when intraweek trends are approximately sinusoidal, the cyclical term approach has more statistical power since it simultaneously leverages information from all points in time for a more powerful longitudinal test. The use of a full set of dummy variables was particularly unwieldy when evaluating a moderator variable for time. The addition of a single covariate increased the number of fixed effects in the weekend, cyclical, and saturated dummy models by 2, 4, and 12, respectively.

It is important to consider the limitations of the present study, many of which inform directions for future research. First, the selection of the “best” model to the daily drinking data was based primarily on model fit statistics (i.e., BIC and AIC). Second, because it was not possible to estimate a full set of random effects with a full set of dummy variables, the comparison with the cyclical and weekend variable models was not directly parallel. Although outside the scope of the present paper, in principle, other differential specifications between fixed and random effects are possible, such as specifying dummy variables as fixed effects in concert with cyclical random slopes, and vice versa. Third, the present study

focused on intraweek patterns in daily drinking, however, the cyclical covariate approach can be applied to longer intervals (e.g., seasonal patterns) or shorter ones. Finally, because the data came from a sample of trauma-exposed college women, the findings may not generalize to a general college population. The hurdle model approach used in the present study would likely be appropriate with data from a general college population, where the proportion of non-drinking women would likely be even higher. However, there may be instances in heavier-drinking samples that a logistic regression may not be necessary (e.g., when most or all of the data is non-zero).

Conclusions

So are cyclical terms preferable over either a simple dummy variable for weekend vs. weekday or a saturated set of dummy variables for each day of the week? No single model will be preferable over any other on all data for all research questions. However, the cyclical terms represent a compromise of sorts between the simplicity of a weekend dummy variable that might “over-smooth” the weekly pattern of drinking and the specificity of a full set of dummy variables for days of the week, which lacks parsimony. As the pattern over time becomes more strictly sinusoidal, a cyclical model will be increasingly preferred given its parsimony over a full dummy variable model and a better fit to the data than a single dummy variable. The present analyses suggest that the combination of cyclical terms with hurdle regression is a reasonable option for analyzing longitudinal drinking with high zero counts. However, the current illustration also provides a realistic example in which careful attention is necessary to gauge the fit of the model to the data. A cyclical model provided a more parsimonious approach to the multiple dummy-variable model while directly capturing increasing and decreasing trends in college drinking data, but trends in both the probability and quantity were not perfectly sinusoidal, which highlights the importance of taking care to evaluate the fit of models using cyclical terms.

Even when a cyclical covariate approach does not turn out to be a good fit in a particular instance, valuable insights may be gained. In the context of alcohol data, we recommend that cyclical terms be considered as a routine option for modeling daily college drinking data alongside the more commonly used dummy variable approaches. Additionally, since high frequencies of zeroes are a common characteristic of substance use and other behavioral outcomes, even in higher-drinking samples, a zero-altered model, such as the hurdle model, is the recommended regression approach for such data.

Acknowledgments

This work was supported by the National Institute on Alcohol Abuse and Alcoholism (NIAAA) grants R01AA019511 and T32AA007455. The data utilized in the present study from Project DASH were collected through support from NIAAA grant R21AA016211. The authors would like to thank Scott Baldwin and Kevin Hallgren for their comments on prior versions of this manuscript.

References

- Atkins DC, Baldwin SA, Zheng C, Gallop RJ, Neighbors C. A tutorial on count regression and zero-altered count models for longitudinal substance use data. *Psychology of Addictive Behaviors*. 2013; 27:166–177. [PubMed: 22905895]

- Bethell J, Rhodes AE, Bondy SJ, Lou WYW, Guttman A. Repeat self-harm: Application of hurdle models. *British Journal of Psychiatry*. 2010; 196:243–244. [PubMed: 20194549]
- Bodenmann G, Atkins DC, Schär M, Poffet V. The association between daily stress and sexual activity. *Journal of Family Psychology*. 2010; 24:271–279. [PubMed: 20545400]
- Burnham KP, Anderson DR. Multimodel inference understanding AIC and BIC in model selection. *Sociological Methods & Research*. 2004; 33:261–304.
- Chow, S-M.; Grimm, KJ.; Fujita, F.; Ram, N. Exploring cyclic change in emotion using item response models and frequency-domain analysis. In: Ong, AD.; van Dulmen, MHM., editors. *Oxford Handbook of Methods in Positive Psychology*. New York, NY: Oxford University; 2007. p. 362-379.
- Dixon LJ, Leen-Feldner EW, Ham LS, Feldner MT, Lewis SF. Alcohol use motives among traumatic event-exposed, treatment-seeking adolescents: Associations with posttraumatic stress. *Addictive Behaviors*. 2009; 34:1065–1068. [PubMed: 19632052]
- Flury BD, Levri EP. Periodic logistic regression. *Ecology*. 1999; 80:2254–2260.
- Gelman, A.; Hill, J. *Data analysis using regression and multilevel/hierarchical models*. New York, NY: Cambridge University; 2006.
- Grant VV, Stewart SH, O'Connor RM, Blackwell E, Conrod PJ. Psychometric evaluation of the five-factor Modified Drinking Motives Questionnaire — Revised in undergraduates. *Addictive Behaviors*. 2007; 32:2611–2632. [PubMed: 17716823]
- Hilbe, JM. *Negative binomial regression*. 2nd ed.. New York, NY: Cambridge University; 2011.
- Hingson R, Heeren T, Winter M, Wechsler H. Magnitude of alcohol-related mortality and morbidity among U.S. college students ages 18–24: Changes from 1998 to 2001. *Annual Review of Public Health*. 2005; 26:259–279.
- Humphreys BR, Lee YS, Soebbing BP. Consumer behaviour in lottery: The double hurdle approach and zeros in gambling survey data. *International Gambling Studies*. 2010; 10:165–176.
- Kaysen, D.; Atkins, DC.; Simpson, TL.; Stappenbeck, CA.; Blaynew, JA.; Lee, CM.; Larimer, ME. *Psychology of Addictive Behaviors*. Advance online publication; 2013. Proximal relationships between PTSD symptoms and drinking among female college students: Results from a daily monitoring study.
- Kaysen D, Dillworth TM, Simpson T, Waldrop A, Larimer ME, Resick PA. Domestic violence and alcohol use: Trauma-related symptoms and motives for drinking. *Addictive Behaviors*. 2007; 32:1272–1283. [PubMed: 17098370]
- Kuha J. AIC and BIC: Comparisons of assumptions and performance. *Sociological Methods & Research*. 2004; 33:188–229.
- Kuntsche E, Knibbe R, Gmel G, Engels R. Who drinks and why? A review of socio-demographic, personality, and contextual issues behind the drinking motives in young people. *Addictive Behaviors*. 2006; 31:1844–1857. [PubMed: 16460883]
- Land TG, Rigotti NA, Levy DE, Schilling T, Warner D, Li W. The effect of systematic clinical interventions with cigarette smokers on quit status and the rates of smoking-related primary care office visits. *PLoS ONE*. 2012; 7:e41649. [PubMed: 22911834]
- Lehavot K, Stappenbeck CA, Luterek JA, Kaysen D, Simpson TL. Gender differences in relationships among PTSD severity, drinking motives, and alcohol use in a comorbid alcohol dependence and PTSD sample. *Psychology of Addictive Behaviors*. 2013
- Marler MR, Gehrman P, Martin JL, Ancoli-Israel S. The sigmoidally transformed cosine curve: a mathematical model for circadian rhythms with symmetric non-sinusoidal shapes. *Statistics in Medicine*. 2006; 25:3893–3904. [PubMed: 16381069]
- Mezquita L, Stewart SH, Ibáñez MI, Ruipérez MA, Villa H, Moya J, Ortet G. Drinking motives in clinical and general populations. *European Addiction Research*. 2011; 17:250–261. [PubMed: 21691118]
- Mohr CD, Armeli S, Tennen H, Temple M, Todd M, Clark J, Carney MA. Moving beyond the keg party: A daily process study of college student drinking motivations. *Psychology of Addictive Behaviors*. 2005; 19:392–403. [PubMed: 16366811]

- Neal DJ, Simons JS. Inference in regression models of heavily skewed alcohol use data: A comparison of ordinary least squares, generalized linear models, and bootstrap resampling. *Psychology of Addictive Behaviors*. 2007; 21:441–452. [PubMed: 18072826]
- Neighbors C, Atkins DC, Lewis MA, Lee CM, Kaysen D, Mittmann A, Rodriguez LM. Event-specific drinking among college students. *Psychology of Addictive Behaviors*. 2011; 25:702–707. [PubMed: 21639597]
- O’Connell, AA.; McCoach, DB., editors. *Multilevel modeling of educational data*. Greenwich, CT: Information Age; 2008.
- Perkins HW. Surveying the damage: A review of research on consequences of alcohol misuse in college populations. *Journal of Studies on Alcohol and Drugs*. 2002; 63:91–100.
- Qin L, Guo W. Functional mixed-effects model for periodic data. *Biostatistics*. 2006; 7:225–234. [PubMed: 16207823]
- Raftery AE. Bayesian model selection in social research. *Sociological Methodology*. 1995; 25:111–164.
- R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2014. Retrieved from <http://www.R-project.org/>
- Read JP, Wood MD, Kahler CW, Maddock JE, Palfai TP. Examining the role of drinking motives in college student alcohol use and problems. *Psychology of Addictive Behaviors*. 2003; 17:13–23. [PubMed: 12665077]
- Simons JS, Dvorak RD, Batien BD, Wray TB. Event-level associations between affect, alcohol intoxication, and acute dependence symptoms: Effects of urgency, self-control, and drinking experience. *Addictive Behaviors*. 2010; 35:1045–1053. [PubMed: 20685044]
- Simpson TL, Stappenbeck CA, Varra AA, Moore SA, Kaysen D. Symptoms of posttraumatic stress predict craving among alcohol treatment seekers: Results of a daily monitoring study. *Psychology of Addictive Behaviors*. 2012; 26:724–733. [PubMed: 22369221]
- Skaug H, Fournier D, Nielsen A, Magnusson A, Bolker B. *glmmADMB: Generalized linear mixed models using AD model builder*. R Package Version 0.7.4. 2013
- Stappenbeck CA, Bedard-Gilligan M, Lee CM, Kaysen D. Drinking motives for self and others predict alcohol use and consequences among college women: The moderating effects of PTSD. *Addictive Behaviors*. 2013; 38:1831–1839. [PubMed: 23261496]
- Stoel RD, Garre Francisca G, Dolan Conor, van den Wittenboer Godfried. On the likelihood ratio test in structural equation modeling when parameters are subject to boundary constraints. *Psychological Methods*. 2007; 11:439–55. [PubMed: 17154756]
- Uitenbroek DG. Seasonal variation in alcohol use. *Journal of Studies on Alcohol*. 1996; 57:47–52. [PubMed: 8747501]

Figure 1.
Correspondence between cyclical terms and components of the longitudinal trends they represent. P = Period (length of time it takes for the cyclical pattern to repeat).

Figure 2.

Histogram of daily drinking counts in Project DASH. The left panel illustrates the zero-altered count outcome and the right panel illustrates the division into dichotomous and zero-truncated count outcomes.

Figure 3.

Plots of the probability of any drinking and number of drinks when drinking by time

Logistic Regression

Weekend Dummy	B	SE	p
Intercept	-1.60	0.10	<.001
Weekend vs. Weekday	1.38	0.09	<.001

Cyclical	B	SE	p
Intercept	-1.01	0.09	<.001
Amplitude (cosine)	-0.78	0.06	<.001
Phase (sine)	-0.61	0.06	<.001

Saturated Dummy	B	SE	p
Intercept	-1.81	0.15	<.001
Tuesday vs. Monday	0.44	0.17	.01
Wednesday vs. Monday	0.15	0.17	.40
Thursday vs. Monday	0.86	0.17	<.001
Friday vs. Monday	2.04	0.16	<.001
Saturday vs. Monday	1.83	0.17	<.001
Sunday vs. Monday	0.18	0.18	.30

Zero-truncated Negative Binomial Regression

Weekend Dummy	B	SE	p
Intercept	1.03	0.05	<.001
Weekend vs. Weekday	0.36	0.04	<.001

Cyclical	B	SE	p
Intercept	1.15	0.04	<.001
Amplitude (cosine)	-0.17	0.04	<.001
Phase (sine)	-0.21	0.04	<.001

Saturated Dummy	B	SE	p
Intercept	0.82	0.09	<.001
Tuesday vs. Monday	0.27	0.09	<.01
Wednesday vs. Monday	0.23	0.10	.02
Thursday vs. Monday	0.38	0.09	<.001
Friday vs. Monday	0.57	0.09	<.001
Saturday vs. Monday	0.67	0.09	<.001
Sunday vs. Monday	0.30	0.11	.01

Figure 4.
Multilevel hurdle negative binomial model results with weekend dummy, cyclical, and saturated dummy variables (baseline models)

Figure 5. Multilevel hurdle negative binomial model results with weekend dummy, cyclical, and saturated dummy variables (moderation models)

Figure 6.

Predicted drinking by low (-1 SD), mean, and high ($+1$ SD) social drinking motives

Table 1

Comparison of Model Fit via Bayesian and Akaike Information Criterion

	Logistic Regression		Zero-truncated Negative Binomial Regression		Overall	
	BIC/AIC	BIC/ AIC	BIC/AIC	BIC/ AIC	BIC/AIC	BIC/ AIC
<i>Baseline model</i>						
- Weekend dummy	3756/ 3738	41 / 72	4267 / 4242	49 / 59	8023 / 6698	90 / 113
- Cyclical	3765 / 3741	50 / 74	4218 / 4183	(ref)	7983 / 6660	50 / 74
- Saturated dummy	3716 / 3666	(ref)	4277 / 4227	59 / 44	7992 / 6616	59 / 31
- Saturated/Cyclical	-	-	-	-	7933 / 6585	(ref)
<i>Moderation model</i>						
- Weekend dummy	3768 / 3738	4 / 65	4260/ 4225	44 / 59	8028 / 6686	47 / 106
- Cyclical	3785 / 3742	20 / 69	4216 / 4167	(ref)	8001 / 6649	20 / 69
- Saturated dummy	3765 / 3673	(ref)	4282 / 4198	66 / 31	8046 / 6602	66 / 22
- Saturated/Cyclical	-	-	-	-	7981 / 6580	(ref)

Note. In each cell the BIC is reported first, followed by the AIC. For the baseline and moderation models, the lowest BIC and AIC in the logistic sub-models, zero-truncated negative binomial sub-models, and overall are emphasized in bold. The difference in BIC/AIC (BIC/ AIC) is calculated with respect to the best fitting model. ref = referent model.

Table 2
Coding of the Weekend Dummy Variable, Saturated Dummy Variable, and Cyclical Terms

	DAY	WEEKEND	Saturated Dummy Variables					Cyclical Terms	
			TUE	WED	THU	FRI	SAT	PHASE	AMPLITUDE
Monday	0	0	0	0	0	0	0	1.0000000	0.0000000
Tuesday	1	0	1	0	0	0	0	0.6234898	0.7818315
Wednesday	2	0	0	1	0	0	0	-0.2225209	0.9749279
Thursday	3	1	0	0	1	0	0	-0.9009689	0.4338837
Friday	4	1	0	0	0	1	0	-0.9009689	-0.4338837
Saturday	5	1	0	0	0	0	1	-0.2225209	-0.9749279
Sunday	6	0	0	0	0	0	0	0.6234898	-0.7818315

Note. WEEKEND = Weekend vs. Weekday; TUE = Tuesday vs. Monday, WED = Wednesday vs. Monday, THU = Thursday vs. Monday, FRI = Friday vs. Monday, SAT = Saturday vs. Monday, Sunday vs. Monday