MODELING EARLY ALCOHOL INITIATION: A COMPARISON OF LINEAR REGRESSION, LOGISTIC REGRESSION, AND DISCRETE TIME HAZARD MODELS

By

KETEVAN DANELIA

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
July, 2011

MODELING EARLY ALCOHOL INITIATION: A COMPARISON OF LINEAR REGRESSION, LOGISTIC REGRESSION, AND DISCRETE TIME HAZARD MODELS

Dissertation Approved:

Dr. Ron Cox

Dissertation Adviser

Dr. Robert Larzelere

Dr. Laura Hubbs-Tait

Dr. Melanie Page

Outside Committee Member

Dr. Mark E. Payton

Dean of the Graduate College

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION

Overview

Harmful use of alcohol has become one of the most important public health issues in the world according to the World Health Organization Report (WHO, 2005). Health problems associated with alcohol consumption include a wide range of diseases, health conditions, and high-risk behaviors, from mental disorders and road traffic injuries (especially among young people), to liver diseases and unsafe sexual behavior (WHO, 2005). Alcohol use and prevalence increase radically during early adolescence, from the ages of 12 through 15 years (WHO, 2005). The initiation of these behaviors early in adolescence leads to a greater risk of health-related diseases and disorders (Farrington, 2003; Moffitt, 1993; Sampson & Laub, 2003).

According to the "gateway drug theory," an adolescent who uses any one drug is more likely to use another drug. Alcohol and tobacco, followed by marijuana, are considered the first "gates" for most adolescents. Under this theory, alcohol or tobacco use precedes the use of marijuana, which precedes the use of other illicit drugs. Even though there are ongoing debates about this theory, many researchers tend to support it. The National Center on Addiction and Substance Abuse (CASA) provides the following information: Among 12-to 17-year-olds with no other problem behaviors, those who drank alcohol and smoked cigarettes at least once in the past month are 30 times more likely to smoke marijuana than those who didn't, and those who used all three gateway drugs (cigarettes, alcohol, marijuana) in the past month are almost 17 times likelier to use another illicit drug like cocaine, heroin, or LSD (Cengage, 2002). There is epidemiologic evidence relating early use of alcohol with initiation of illegal drugs in Mexican students (Herrera-Vazquez, Wagner, Velasco-Mondragon, Borges, & Lazcano-Ponce, 2004). A study conducted by Wagner, Velasco-Mondragon, Herrera-Vazquez, Borges, and Lazcano-Ponce (2005) shows that early onset use of alcohol is associated with excess risk of illegal drug use. These findings underscore the importance of targeting alcohol initiation for early intervention and prevention strategies.

Concerning the status of American youth and families, some researchers have concluded that the United States is a nation at risk with regard to alcohol and drug abuse (Weissberg, Walbergg, Obrien, & Kuster, 2003). Health-risking behaviors including alcohol use, tobacco use, and delinquent behavior have large costs to society (Mokdad, Marks, Stroup, & Gerberding, 2005; Owings, 2008; Woolf, 2006). A U.S. national report shows that American teenagers who use alcohol and tobacco usually initiate use between 12 to 16 years of age. Youthful substance abuse can be defined as the frequent use of alcohol or other drugs in a way which leads to a problem that extracts considerable costs on both personal and societal levels. Underage drinkers account for nearly 20% of the alcohol consumed in the US each year (Foster, Vaughan, Foster, & Califano, 2003). Adolescents between 12 to 16 years old who have ever used substances, such as alcohol and drugs are more likely at some point to have sold drugs, carried a handgun, or been in a gang than youth who have never used substances (Snyder & Sickmund, 1999). Earlier onset of initial alcohol use often signals future deficiencies in social functioning and physical and mental health (Friedman, Terras, & Zhu, 2004; Jones, et al., 2004; McGue & Iacono, 2005). For example, Hingson, Heeren, and Winter (2006) found that 45% of adults who began drinking by age 14 became dependent on alcohol at some point in their lives versus 9% who began drinking at age 21 or older. Studying prevalence and drug dependence among Americans from 15 to 54 years old, it was found that about 1 in 7 (14%) had a history of alcohol dependence and about 15% of drinkers had become alcohol dependent (Anthony, Warner, & Kessler, 1994). Behavioral signs of alcohol dependence may include: alcohol withdrawal symptoms (e.g., nervousness, shaking, irritability, and nausea); increased tolerance to alcohol; alcohol consumed in larger amounts or over a longer period than was intended; failure of attempts to stop drinking; considerable time devoted to activities associated with alcohol use or obtaining alcohol; neglected daily activities; and disregard for consequences of negative behaviors (Reyes, 1999). People who begin drinking before age 15 are four times more likely to develop alcohol dependence at some time in their lives compared with those who have their first drink at age 20 or older (Grant & Dawson, 1997). Early substance use by adolescents is important because it increases the likelihood of later substance abuse (Spoth, Guyll, & Day, 2002). Most studies conclude that earlier initiation of drugs, such as alcohol and marijuana, is associated with greater use of that drug, greater probability of involvement in more serious drugs, and greater involvement in deviant activities (Brunswick & Boyle, 1979; Margulies, Kessler, & Kandel, 1977; Kleinman, 1978)

Early Alcohol Initiation

Exploring a sequence of drug involvement leads not only to early drug initiation, which progresses to further drug involvement and abuse, but also to a very first stage of this sequence, which is an opportunity to try a drug. The first to draw attention to the first stage of drug involvement, referred to as drug exposure opportunity, was Robins (1977). A study conducted by Van Etten, Neumark, and Anthony (1997) shows that among persons who were given an opportunity to use marijuana, there are increases in the probability of progressing from first marijuana opportunity to first marijuana use. Furthermore, the transition from first opportunity to eventual marijuana use seems to depend on age at first opportunity. Wagner and Anthony (2002) showed that once the chance of marijuana use had occurred, tobacco smokers were more likely to engage in actual marijuana use. A study of youthful drug involvement in Chile found that the probability of marijuana use and the conditional probability of marijuana use (given an opportunity) are greater for users of alcohol only, tobacco only, and alcohol plus tobacco, as compared to non-users of alcohol and tobacco (Caris, Wagner, Rios-Bedoyae, & Anthony, 2009).

While an increasing number of studies have looked at age of first use as an independent variable, there are not many studies that model these first stages of alcohol involvement as an outcome, despite the obvious importance of exploring this topic. There are several reasons that only a few studies have looked at age of first use as an outcome variable, which I will briefly explore here. One reason for a lack of studies has been methodological concerns about the validity of self-report questionnaires measuring opportunity to use and first use of drugs (Van Etten & Anthony, 1999; Van Etten, Neumark, & Anthony, 1997). Most surveys looking at these variables use a retrospective design within a questionnaire or interview. There is a concern that users of alcohol, tobacco, and illicit drugs might report more or less completely and accurately than nonusers. However, since there is no perfectly valid estimation for age at first opportunity or for age at first use, self-report is still the most common means of collecting data for almost all studies (Wagner & Anthony, 2002).

Modeling Alcohol Initiation

A second, more difficult concern to resolve has been how to model age of first use. As occurs with any variable measuring a high-risk behavior in which a sizeable percentage of the population has never engaged, the distribution of age of first use tends to be bimodal, one mode for those who have engaged in a high-risk behavior and one for those who have never engaged. One solution (with attendant problems discussed below) is to eliminate the data for the portion of the population who have not engaged in the high risk behavior. In order to properly address the bimodal distribution of the outcome, many studies dichotomize the outcome and use logistic regression to predict the probability of initiation given the predictors included in the study (e.g., Tur, Puig, Pons, & Benito, 2003; Bekman, Cummins, & Brown, 2010; MacPherson, Magidson, Reynolds, Kahler, & Lejuez 2010; Carlini-Marlatt, Gazal-Carvalho, Gouveria, & Souza, 2003). With logistic regression an outcome variable has two possible values: either alcohol initiation or no initiation by the time of the interview.

Logistic regression. Because the dependent variable is not a continuous one, the goal of logistic regression is the classification of study participants in one of two categories of the dependent variable (e.g., alcohol user or no user) predicted by the independent variable. In other words, we are predicting the probability that a person will be classified into one as opposed to the other of the two categories. Because the probability of being classified into the first or lower valued category, $P(Y = 0)$, is equal to 1 minus the probability of being classified into the second or higher-valued category, $P(Y = 1)$, if we know one probability, we know the other (Menard, 2002). Interpretation of coefficients in logistic regression equation is different from those in linear regression equation. In linear regression the model coefficients have a straightforward interpretation where the coefficient of the predictor variable estimated the expected amount of change in the dependent variable for any one-unit increase in the independent variable (Pedhazur, 1997). Logistic regression reports *odds ratios* (OR) that are interpreted differently. First of all, it is important to understand the concept of odds. For our dichotomous outcome variable, the odds of membership in the alcohol user group are equal to the probability of membership in the alcohol user group divided by the probability of membership in the non-user group. For example, if the probability of membership in alcohol user group is .5, the odds are 1 (.5/.5); if the probability is .8 the odds are 4 (.8/.2). Obviously, if the odds ratio is 1 then both memberships are equally likely (e.g., as likely to be in the user group as in the non-user group). If the odds are more than 1, the probability of being an alcohol user is more likely than being non-user; and if less than 1, then alcohol using is less likely. Odds tell us how much more likely it is that participants are in the alcohol user group rather than a member of the non-user group (Wright, 2002).

Odds ratio estimates the multiplicative change in the odds of membership in the alcohol user group for a one-unit increase in the predictor and is computed by exponentiating the regression coefficient of the predictor variable (Wright, 2002). For example, if the regression coefficient is .75, the odds ratio is $e^{.75} = 2.12$. This means that the odds that study participants are in an alcohol user group (vs. not) is 2.12 times greater when the value of a predictor is increased one unit. An odds ratio of .5 indicate that the odds of being in alcohol user group (vs. not) decreases by half when predictor increases by one unit, i.e. there is a negative relationship between predictor and outcome. Under the null hypothesis of no effect, the odds ratio will be 1.00, meaning that the odds of being in the user group stays the same with increases in the corresponding predictor.

However using logistic regression to model the onset of a behavior can be problematic (Singer & Willet, 2003). Dichotomizing discards variation in age of alcohol initiation, which represents meaningful information because individuals initiate alcohol at different times of their lives. Individuals who started using alcohol at early age of their lives are different from those who initiated alcohol much later, but they become indistinguishable in a logistic regression analysis. Thus, using this technique does not consider nor provide any information about the early stages of alcohol initiation as compared to later initiation and therefore is not very useful for investigating age of first alcohol use.

Missing data. A second option that other researchers have chosen to accommodate the bimodal distribution of age of first use variables is to truncate the data and only study those who have initiated use. This option usually results in continuous, possibly normal distribution and allows for standard regression assumptions, but presents other problems. It is well-known that the scientific method involves making structured observations, drawing causal inferences based on observations, and generalizing study results beyond the study (Cozby, 2007; Dooley, 2001). Truncated data can be interpreted as systematically missing data, and can have consequences for all these activities associated with the scientific method. There are a wide range of consequences of having missing data as described in McKnight, McKnight, Sidani, and Figueredo (2007), but they all require data to be missing in a non-systematic way, at least after controlling for predictors in the analysis. Missing data can affect the reliability and validity of systematic observations. When drawing inferences from observations, missing data can affect the strength of the study design and the validity of conclusions about relationships between variables. When generalizing study results, missing data can limit the representativeness of the study sample, the strength of interventions, and other aspects of the study such as time or place about which we would like to generalize. The missing data have the potential to influence the validity of constructs of the study, i.e., how accurately the variables or constructs of interest are represented, or how well measures capture the variables or constructs. Beyond affecting the construct validity, missing data can affect both the reliability (stability and consistency) and validity (accuracy, generalizability) of research findings. These aspects are related to the internal validity of a study. If large portions of data are missing in a study, e.g., the data set used for analyses represents a smaller and potentially biased sample of participants that may lead to inaccurate and unstable parameter estimates. Consequently, the reliability and validity of study is jeopardized, which leads to weaker causal inferences regarding the relationships between variables and thus lower internal validity. Internal validity is often characterized as the extent to which a researcher can reasonably claim that a particular predictor is responsible for the observed outcome. The influence of other factors, i.e., confounds or alternative explanations for the outcome, weakens the inference that the predictor considered in the study caused the outcome. Those other factors are known as threats to internal validity (Shadish, Cook, & Campbell, 2002; Campbell & Stanley, 1966). Selection bias which refers to systematic differences on some characteristics between groups of individuals included in the study (Shadish et al., 2002) is one of the most recognized threats to internal validity. Such differences can influence study conclusions. Although the effects of missing data are potential problems for reliability and validity of study findings, the adverse effects on statistical procedures are almost always expected to be present. For example, statistical power (probability that a null hypothesis will be rejected given that it is false) is directly related to sample size. As the sample size decreases, statistical power decreases. Missing data also affects data analysis in how they distribute data and error. For example, commonly used analyses, like ANOVA and multiple regressions, require errors to be normally distributed. Failure to conform to these assumptions produces inaccuracies in the results and affects significance tests and parameter estimates. When analyses require multivariate normality, it is almost guaranteed that it will be adversely affected by missing data (McKnight et al., 2007). The most critical problem comes up when data is missing systematically. This always leads to a selection bias, which is the one of the most difficult problems in data analysis because it leads to wrong estimates and results. In summary, missing data that affect the strength, integrity, reliability, and validity of causal inference affect internal validity. Missing data can also influence the generalizability of findings. In particular, the observed effects can be attributable to the resulting sample that participated in the study. In conclusion, the missing data can affect the interpretation of findings in a single study, the synthesis of results across studies, and the knowledge and understanding in the field (McKnight et al., 2007).

The present study focuses on early stages of alcohol involvement and intends to increase understanding of initial opportunity to try alcohol and the transition to initial alcohol use. It is a secondary data analysis that draws from and expands on the Cox (2007) study that looked at factors associated with age of first use of various substances from a subsample of Venezuelan youth who had already initiated use. Since Cox used a multi-level regression analysis he could only study those participants who had initiated drug use due to a bimodal distribution of the outcome variable, which, as previously explained, is a violation of the normality assumption in regression analysis. The study focused on participants who were deemed at higher risk due to having initiated use and participants who reported no drug use were left out of the sample (about 16.1 % of the whole sample). Dropping those participants from the sample, as discussed previously, creates difficulties because it represents a selection bias problem. This non-trivial portion of the data must be considered systematically missing because it is not missing completely at random (MCAR) or even missing at random (MAR). MCAR applies when the probability that an observation (e.g., alcohol use) is missing is unrelated to the value of this observation or of any other observation. MAR applies when the probability that the observation is missing does not depend on the value of this observation after controlling for other variables in the statistical analysis. When data is not MCAR or MAR it is called nonignorable missing data (Allison, 2002). Ignorability basically means that there is no need to model the missing data to obtain unbiased statistical estimates. According to Allison (2002), for nonignorable missing data, a careful consideration of the appropriate model is necessary because results typically will be very sensitive to the choice of model, especially to how well it controls for systematic biases in the missing data.

Proposed analyses. A statistical technique that combines logistic and multiple regressions is *survival analysis*. Similar to logistic regression analysis, survival analysis detects participants with high-risk and no high-risk behaviors (alcohol initiation in this study). In addition, survival analysis evaluates early ages of high-risk behavior similar to traditional regression analysis but unlike regression analysis it overcomes the missing data problem by keeping participants with reported no high-risk behavior in the sample. A more thorough discussion of survival analysis will be provided later.

Another concern when modeling age of initiation of a high-risk behavior that needs to be addressed is the potential for non-independence of observations in the data. Much of the data studying adolescent initiation of substances were collected in schools or neighborhoods. As such, the data is nested and presents a challenge regarding how to model the dependent variable without violating the independence assumption required for regression analysis. When nesting occurs and the non-independence in not accounted for, standard errors are artificially small resulting in inflated parameter estimates (Pedhazur, 1997).

A recent methodological development has combined survival analysis with multilevel modeling. Multilevel survival analysis incorporates the best features of logistic regression, utilizes all of the available information in the data, and accounts for the nested structure of the data. However, there are only a handful of studies that has used survival analysis in a multilevel framework to both model onset and account for the nesting (Barber, Murphy, Axinn & Maples, 2000; Reardon, Brennan, & Buka, 2002; Steele, Goldstein & Browne, 2004). To better understand this method a brief description of the basic tenets of survival analysis is required here.

Survival analysis is usually used when the research question involves a test of "whether and when" as outlined by Singer and Willet (2003). The present study passes this test because it investigates whether or not alcohol initiation happened and when it happened, e.g., how many years have passed since the age of first opportunity of alcohol use. According to Singer and Willet (2003), besides having a research question leading to a survival analysis, it is also important to clearly examine methodological features that involve clearly defining a *target event*, such as an alcohol initiation occurrence investigated in this study; *beginning of time*, i.e., a starting point when nobody has yet experienced alcohol initiation, which is the age of first opportunity of using alcohol or any arbitrary age; and a *metric of clocking time* (meaningful scale in which event occurrence is recorded), i.e., years from the age of first opportunity of using alcohol or years from the arbitrary age. Event occurrence is a transition of an individual from one “state” (never initiating yet) to another “state” (having initiated).

Survival analysis is another term for *event history analysis* (the former is usually used in biomedical studies and latter in social studies) and both are used interchangeably in the research literature. The name comes from studying how long subjects of a study survive under different circumstances (Allison, 1984). In the language of survival analysis, the present study investigates how long participants survive until alcohol initiation after either having an opportunity of using alcohol or after some arbitrary age. In other words, what is the duration of time (survival) from a starting point (opportunity or an early age) until initiation?

The rationale for using survival analysis and not traditional regression-like analysis when investigating alcohol initiation lies in realizing the fact that not all participants have experienced alcohol use before the time of data collection. In the Cox (2007) data, 16.1% of the participants never initiated drugs by the time of the data collection. The problem when study participants have unknown event times is called *censoring,* and participants with unknown event times are called censored observations (Allison, 1984; Singer & Willet, 2003). The amount of censoring is usually related to the rate at which events occur and the length of data collection (Singer & Willet, 2003).

Alcohol initiation, or time of event occurrence, is measured in discrete time intervals because we only know the year in which alcohol use was initiated. A key concept in survival analysis is the *risk set,* which is the set of participants in this study who are at risk of alcohol initiation at each discrete time, e.g., year (Allison, 1984). A second key concept is the *hazard rate*. In the present study the hazard rate is the probability of alcohol initiation at a particular year to a particular participant, given that that participant did not initiate alcohol in any earlier time (Allison, 1984). The hazard rate is not an observed variable, but it is estimated from alcohol initiation and its timing (Allison, 1984). It represents the fundamental dependent variable of the event survival analysis in this study.

Proposed models. About 11.6% of participants of the sample said they never had an opportunity to try alcohol and therefore will be excluded from the analysis when we investigate alcohol initiation survival after participants have had an opportunity to use it. Excluding from the study those participants who have never been exposed to alcohol partially reduces the problem with missing data (compared to above mentioned 16.1% of participants). However, the excluded participants (11.6%) still represent a significant percentage of the sample. The problem can be resolved if the *beginning of time* is set to be an arbitrary age or the reported first use of alcohol. The beginning of time will be set at age 4 and all earlier reported ages will be discarded. Cox (2007) set the cutoff at age at 4 due to memory limitations of very young children (i.e., children in the 1-3 age range) and due to anecdotal evidence from focus groups with teachers and parents indicating that children as young as 4 years of age were used by others to traffic drugs into schools, and thus could have had a chance to use the substance (Cox, 2007). Additionally, self-report data from several youth (*n*=9) indicates that youth began some type of substance use at this age. This model includes all participants, alcohol initiators and non-initiators starting from age 4 and represents 99.8% of the whole sample. Four participants who reported age of first use at ages 2 and 3 are not included in the analysis.

The multilevel modeling framework allows taking into account the nested structure of the data, i.e., the non-independence of observations. For example, persons may be nested within schools, communities or within countries. Additionally, the use of a combination of survival analysis and multilevel models allows for the testing of the relationship between several risk factors and the time between age of first opportunity to use alcohol and age of first use of alcohol for one model and the time between age 4 and age of first use of alcohol for another model.

Due to their salience in the literature, the variables of family attention, adolescent externalizing behavior, socioeconomic status, and gender (variables to be defined later) will be the independent variables considered in this study. The present study will illustrate how those predictors affect the likelihood of alcohol initiation during the time from the first opportunity to use alcohol and from age 4 until its initiation. It needs to be noted that not all explanatory variables considered in the present study are exactly time invariant variables. Time-varying variables might create problems by leading to inaccurate results if the variables were to change over time. Explanatory variables usually are measured only once even in longitudinal studies and often are assumed to be time-invariant variables and are treated accordingly (Allison, 2002). In the present study, Family Attention and Externalizing behavior, while not traditionally considered time-invariant in the strictest sense, have been shown in the literature to be very resistant to change and are assumed to be time-invariant for the purposes of the present analysis (Loeber, 1982; Murphy, Wickramaratne, & Weissman, 2010). Loeber (1982) reviewed studies on the stability of antisocial behavior and showed that adolescents who exhibit high rates of antisocial behavior are more likely to persist in this behavior than children who initially display low rates of antisocial behavior. Studies showed that once high levels of antisocial behavior were established, adolescents tend to maintain rather than decrease levels of antisocial behavior. Murphy et al. (2010) explored parental bonding in 20-year follow up study and concluded that parental bonding maintained a non-significant mean level change over a 20-year period. Even though items that formed parental bonding are not exactly the same as items for Family Attention in this study, they are very similar.

The second proposed model will also allow for controlling for the age of first opportunity to use alcohol. In the second model where the *beginning time* is age 4 and not the age of first opportunity per se, controlling for age of first opportunity will eliminate those participants who reported no opportunity to try alcohol (due to listwise deletion). This, again, will lead to the missing data problem. One way of dealing with missing participants, i.e. with those who have never been exposed to alcohol is to recode the age of first opportunity to try alcohol, a continuous variable, so that participants with no opportunity to try alcohol are assigned the year they were interviewed plus one year (recognizing that they might have an opportunity to try alcohol later that year). Another situation that can be considered within this model is the one that eliminates the age of first opportunity to try alcohol as a control variable. The present study will evaluate this model using survival analysis considering both situations: 1) controlling for the age of first opportunity where the age of first opportunity is recoded as was described above; 2) without controlling for the age of first opportunity to try alcohol.

The main objective of this study is to compare and contrast three different approaches (i.e., survival analysis, multiple, and logistic regressions) to modeling age of first use as an outcome variable where predictors in all three of the models are the same but the statistical analysis employed for the evaluation of models are different. The first approach will be an extension of the Cox study and will estimate survival until alcohol initiation with and without having the opportunity of alcohol initiation. Here, a multilevel version of survival analysis will be employed to estimate and (1) early alcohol initiation when participants are observed from the age of first opportunity to use alcohol; and (2) early alcohol initiation when the beginning of time for observing them starts at age 4 controlling for a recoded age of first opportunity with no opportunity set at age of interview plus one year. The second approach will be a reanalysis of Cox's (2007) study which evaluated age of onset of use of all drugs with non-using subjects eliminated. In the current study, the outcome is a continuous variable - age of first use of alcohol. To examine this approach a standard linear regression analysis will be used and, like Cox (2007), cases not reporting alcohol use will be eliminated. Contrary to Cox's study, multilevel analysis will not be employed because multilevel survival analysis does not go beyond person-level variables in this study. The third approach will model age of first use as a dichotomous outcome variable (use vs. no use) with the logistic regression.

CHAPTER II

REVIEW OF LITERATURE

The main purpose of this chapter is to discuss some aspects of survival analysis which will be used in this study in combination with multilevel modeling and some methodological issues associated with it. The review of research studies will also investigate methods that are used for studying variables of interest to the present study and how those variables affect alcohol use.

Survival Analysis

As was already mentioned in the previous chapter, survival analysis studies how long subjects of a study survive until some event occurrence (e.g., alcohol initiation) under different circumstances. Obviously, occurrence of an event assumes a preceding time interval, which is its nonoccurrence. More specifically, a certain time period or duration of nonoccurrence must exist in order for an occurrence to be considered as an event. Survival analysis is used to study duration data, which represents the nonoccurrence of a given event (Yamaguchi, 1991). An event, which is alcohol initiation in this study, is defined by specifying a group of end points for duration intervals. . In the case of age of alcohol initiation, an event, i.e. alcohol initiation, is defined by the end point of the duration interval for having never used alcohol.

Another concept of the duration of the nonoccurrence of a given event is the risk period. The time period that represents the nonoccurrence of alcohol initiation can be divided into the period at risk and the period not at risk for initiating alcohol. The distinction between the risk and nonrisk periods requires assumptions. For example, we can assume that alcohol initiation can occur only for those individuals who had an opportunity to use alcohol. This assumption is more implicit and in fact, can be backed up by several research studies that emphasize the importance of being exposed to drug opportunities and then the transition from exposure to actual drug use (Van Etten et al., 1997; Van Etten & Anthony, 1999; Wagner & Anthony, 2002; Caris et al., 2009; Wilcox, Wagner, & Anthony, 2002). Benjet et al. (2007) state that an opportunity to use drugs is the first step of drug involvement. Along with other findings, they indicate that drug use is only possible given exposure to drug use opportunities. In a study conducted by Wagner and Anthony (2002) there is a clear implication that preventive strategies of drug use should be aimed at reducing drug use opportunities. Furthermore, many factors found to be related to drug use (gender, parental attention, socioeconomic status, etc.) may actually only be related to drug use to the extent that they relate to exposure to drug opportunities (Chen, Storr, & Anthony, 2005; Van Etten & Anthony, 1999).

An alternative would be to assume that all participants enter the risk period at the same age, which might be the youngest age observed in the given sample. Even though the youngest age of alcohol initiation was reported at age 2 in this data set, it was decided to set the earliest age at 4 years based on previous research findings in the Cox (2007) study and the high likelihood of errors in retrospective memory for events prior to age 4. The particular assumption made in defining the risk period becomes a characteristic for the model. The integrity of assumptions is very important for subsequent analysis. Study participants who are at risk, given the definition of the risk period, are considered to be the risk set of that time. According to Yamaguchi (1991), taking into consideration the distinction between the risk and non-risk periods, survival analysis can be defined either as the analysis of the duration for the nonoccurrence of an alcohol initiation during the risk period or as the analysis of rate of the alcohol initiation during the risk period.

The rate usually varies with time and among groups, and when it is attached to a particular moment in time, it is referred to as a hazard rate or transition rate, which was defined above. The term hazard rate comes from biostatistics, where the typical event is harmful. The term transition rate is more often used in sociology, where events are transitions between distinct states (Yamaguchi, 1991). The hazard rate (or hazard function) h(t) can also be defined in mathematical terms as the ratio of the unconditional instantaneous probability of having the event f(t) divided by the survival probability S(t), which is the probability of not having the event prior to time t:

$$h(t) = \lim_{\Delta t \to 0} [(P\ (t + \Delta t > T \geq t \mid T \geq t))/\Delta t] = f(t))/(S(t)$$

where T is the total duration of the risk period until an event occurs, and

$P\ (t + \Delta t) > T \geq t | T \geq t)$ indicates that probability that the event occurs during the time (t, t+ Δt) given that the event did not occur prior to time t. The unconditional instantaneous probability of having the event at time t, f(t) is also called the probability density function of T.

It has to be explained how the hazard rate depends on explanatory variables. For example, if there are just two explanatory variables, x_1 and $x_2(t)$ in year t then the first approximation of P(t) can be written as a linear function of those variables: $h(t) = a + b_1x_1$

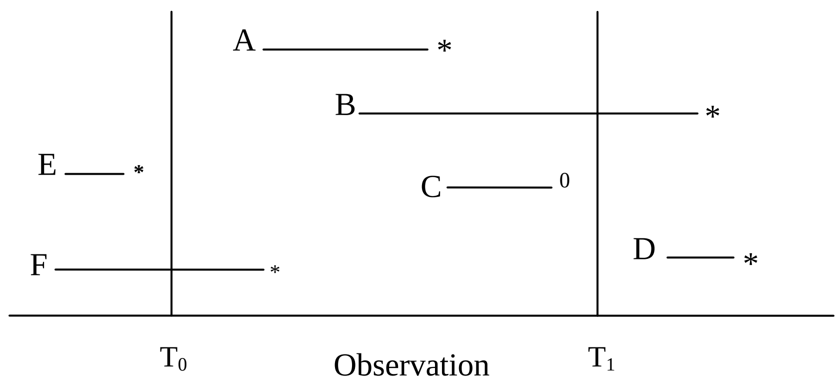
$+ b_2x_2(t)$, for t = 1, 2, …n. Because h(t) is a probability, it varies between 0 and 1, while the right-hand of the equation can be any real number. This kind of model can produce impossible predictions and consequently creates difficulties in computation and interpretation. The problem is avoided by taking the most commonly used logit transformation of h(t):

$\log(h(t) / (1- h(t))) = a + b_1x_1 + b_2x_2(t)$. As h($t$) varies between 0 and 1, the left-hand side of the equation varies between minus and plus infinity. The coefficients b1 and b2 show the change in the logit for each one-unit increase in x1 and x2, respectively (Allison, 1984).

There are two major methods for analyzing hazard rates: nonparametric methods which make few if any assumptions about the distribution of an event time and parametric methods which assume that the time until an event comes from a specific distribution, the most common being the exponential Weibull, and Gompertz distributions (Allison, 1984). Both methods can estimate the effects of covariates on hazard rates. Covariates that can be used in analysis may be time invariant, i.e., they do not vary throughout the duration of the time (gender, race, etc.) or time-variant (alcohol availability, perception of alcohol risk, etc.). Nonparametric methods do not specify the nature of the relation between time and hazard rates (Yamaguchi, 1991). Survival analysis can refer to an analysis based on either parametric or nonparametric hazard-rate models. Nonparametric models are used in this study.

Methods that assume that the time of event occurrence is measured exactly are known as “continuous-time” methods. In practice, time is almost always measured in discrete units (Singer & Willet, 2003). If these discrete units are very small, then time can be treated as a continuous measure. When the time units are large, e.g., years as in the present study, it is more appropriate to use discrete-time methods (Allison, 1984).

Censoring. One advantage of hazard-rate models for the analysis of duration data is its ability to deal with unknown event times, which is called censoring, as described in the previous chapter. Censoring exists when incomplete information is available about the duration of the risk period because of a limited observation period. Yamaguchi (1991) describes six distinct situations regarding censored observations (Fig.1)



Note: * = event occurrence; 0 = occurrence of event other than event of interest.
Source: Yamaguchi (1991).

Fig. 1. Left and right censored observations.

All participants are under observation from time T_0 (either age of first opportunity or age 4 in this study) and time T_1 (interview year in this study). Both times are assumed to be determined independently of subjects. The solid line indicated the risk period for each subject. The solid line with an asterisk (*) represents an occurrence of the event of interest (alcohol initiation), and the solid line with an open and end point (0) indicates that the risk period is terminated by an event other than alcohol initiation, e.g., the participant has been dropped from the sample. Yamaguchi (1991) explains differences among three distinct missing-data mechanisms using two variables, X and Y. Variable Y, which is the duration of the risk period up to the alcohol initiation, is subject to nonresponse (i.e. missing data) due to censoring. Variable X represents the timing of entry into the risk period, which is either AFO or age 4 in this study (as has been discussed above). Generally, three distinct missing-data mechanisms can be identified according to whether the probability of nonresponse to Y (1) depends on Y (and possibly X as well), (2) depends on X but not on Y, or (3) is independent of X and Y.

Allison (2002) refers to case (3) as a situation when data are missing completely at random (MCAR), which happens when the missing data are not systemically related to any variable in the model. A less stringent requirement is missing at random (MAR) when the nonresponse to Y (or missing on Y) is unrelated to the value of Y and to any predictors of Y after controlling for all X covariates in the model. Thus when there is case (2), the missing data are MAR, but the observed data are not observed at random (OAR). In this case, the observed missing data are random only within levels of X. The data are neither MAR nor OAR when case (1) is observed

If the missing data are not MAR, it is important to distinguish between situations where the data on Y are missing by a known mechanism and the data on Y are missing by an unknown mechanism. The most serious problem is when the missing data are not MAR and are missing by an unknown mechanism (Yamaguchi, 1991).

The typology of missing-data mechanisms can be applied to the different types of censoring depicted in Fig.1 and are explained in Yamaguchi (1991). For example, the entire risk period for Subject A falls within the period of observation, and thus this observation is not censored.

The risk period of Subject B starts during the period of observation, and this participant did not initiate alcohol use when the observation is terminated at T_1, i.e., at the year of the interview in this study. The subject's observation is right censored at $T_{1.}$ This type of censoring is typical for survey data. The value of Y is missing because the date of exit from the risk period for Subject B is not known, even though there is information about the duration of the risk period up to the censoring time.

The case of Subject B is also called a right-truncated observation. Truncation is a special case of censoring that is characterized by a partial observation of the duration data. Given that the timing of T_1 is determined independently of the hazard rate, survival analysis can handle this type of right censoring adequately. Among censored observations, right-truncated observations occur most frequently in social studies, and the ability of event history analysis to handle them is its major advantage over other analyses, such as linear or logit regression analyses. For right-truncated observations the missing data are not MAR but are missing by a known mechanism. The missing data are not MAR because the occurrence of censoring depends of the value of Y and the mechanism is known because we know when and how the observations are right truncated. For Subject C, the observation is right censored because an event other than the event of interest occurs during the observation period and takes the subject out of the risk set. This type of censoring is not under the control of the investigator. If the event that terminated the observation happened independently of the hazard rate of the event of interest then it is independent censoring. Virtually all survival analysis methods assume that the censoring times are independent of the time of event occurrence. It is possible to develop models which allow for dependence between censoring and times at which an event occurs but this is rarely done. The main reason for not developing such models is that it is impossible to test whether any dependence model is more appropriate than the independence model (Allison, 1984). When independent censoring applies, Subject C can be treated as an instance of right-censored observation technically in the same way as Subject B. Subject D represents a case in which the observation is fully censored on the right. Entry into the risk period occurs after the observation period and, the value of Y is missing for Subject D. In other words, the occurrence of full right censoring depends only on the particular variable X that represents the timing of entry into the risk period and does not depend on duration Y. The missing data of Y are MAR but the observed data of Y are not OAR.

The case of Subject E represents a case with full censoring on the left. Generally left censoring is much less manageable than right censoring, and the case of Subject E is the worst possible situation. The value of Y is missing for subject E. The data are neither MAR nor OAR. Besides, the missing-data mechanism is unknown because we do not know when and how the event occurred to make the value of Y missing. Unlike the case of full right censoring, the sample selection bias occurs as a function of the unknown values of the dependent variable Y. Full left censoring creates serious bias in parameter estimates unless the number of subjects with full left censoring is small. In this study, participants with age of first use younger than 4 are examples of full left censoring. Both subjects E and D are not in the risk set during the observation time for different reasons. Subject E has already experienced the event and D has not entered the risk set yet.

The case of Subject F represents a partially left-censored observation, which is also called left truncation. Here the data of Subject F cannot be used adequately. The beginning of the observation period is not equal to the beginning of the risk period for left-truncated duration data. This is the situation when the missing data are not MAR and are missing by an unknown mechanism.

Singer and Willet (2003) emphasize that left-censoring creates challenges that are not easily addressed even with the most sophisticated of survival methods. The most common advice is either redefining the beginning of time to coincide with a precipitating event (e.g., age of first opportunity) or eliminating left-censored data through design (e.g., starting as young as possible).

Even though censoring is a complicated issue, it is an advantage of survival analysis rather than a disadvantage because survival analysis can handle censored observations adequately in many situations. In the present study right censoring is used because the duration of time from having opportunity or from age 4 until initiation is not known because the event occurrence of alcohol initiation has not been observed. Right censoring is the most common situation. In this study left censoring is minimized by starting the risk set period at early age or when the opportunity of alcohol use first occurred.

Thus, in the present study time duration from the age of first opportunity (AFO) (in one model) and age 4 (in another model) until alcohol initiation is measured in discrete times, i.e. years, and time is censored for those who did not initiate alcohol by the end of the observation period. That is why we are using so called discrete-time survival model (Allison, 1982; Singer & Willet, 2003). For building this model constructing *a person-period data set* is necessary (Singer & Willet, 2003). Survival models usually require longitudinal data (when there are repeated interviews) but in this study we have retrospective data (a single interview) and person-period data set must be constructed from retrospective data to make the analysis possible. To obtain the person-period data set, each person has a period for each year he or she is in the risk set. The first period is the first year of risk (age of first opportunity or age 4). The last period is either the year of alcohol initiation or the age before the interview, whichever comes earlier. Any years in between constitute the other periods for that person in the person-period data set (This will be visualized with sample data in Chapter 3). All statistical aspects of analysis will be discussed in Chapter 3.

Review of Research Studies

First I will review studies of risk and protective factors in adolescent substance use. Reviewing risk and protective factors are relevant for this study since they are explanatory variables that potentially affect the outcome. Those factors are Externalizing behavior, which reflects participating in high-risk and delinquent behavior; Family Attention, which reflects adolescents bonding with parents and parental control; Socioeconomic Status, which reflects the type of residence and parental education; and Gender. Next I will review studies that address the issue of modeling the Age of Fist Use (AFU), which is the main response variable used for studying the hazard rate of alcohol initiation. I will also review several studies that address modeling of AFO of substance use because together with AFU they represent response variables in the model when we evaluate time duration from AFO to AFU of alcohol. And finally I will cover some of the recent studies that combine survival analysis with multilevel modeling methods to illustrate opportunities that this technique provides for studying different event occurrences in multilevel framework.

Overview of research on risk and protective factors. There are empirically identifiable patterns of behavior or contexts that serve as risk or protective factors in the development of substance use (Hawkins, Catalano, & Miller, 1992). According to Hawkins et al. (1992) certain characteristics of individuals and their personal environments are associated with a greater risk of adolescent drug abuse. Among these are physiological factors, family alcohol and drug behavior and attitudes, family conflict, low bonding to family, early and persistent problem behaviors, academic failure, low degree of commitment to school, peer rejection in elementary grades, antisocial behavior, association with drug-using peers, alienation and rebelliousness, attitudes favorable to drug use, and early onset of drug use. Many studies showed a strong association between externalizing behavior and substance abuse (Wells, Graham, Speechley, & Koval, 2004; Adalbjarnardottir & Hafsteinsson, 2001; Li & Feigelman, 1994; Jessor, 1993). Externalizing behavior is usually associated with aggression toward people and animals, damage of property, theft, and serious rule violations. Antisocial behavior during childhood predicts substance abuse during adolescence (Clark, Vanyukov, & Comelius, 2002). Another longitudinal study also shows that childhood externalizing behavior increases the likelihood of substance use in later adolescence (Adalbjarnardottir & Ranfnsson, 2001). On the other side, a child's antisocial behavior brings out aversive reactions by the parents, which then raise the child's aggressive behavior (Patterson, 1997). Parents may respond to adolescent antisocial behavior by raising their tolerance level for deviant behavior (Bell & Chapman, 1986) that may result in decreased attempts for dealing with problems. As a consequence, parents become less supportive and controlling when their adolescent's behavior becomes more aggressive and hostile.

Adolescents, regardless of their antisocial behavior, who characterized their parents as being attentive, were more protected against substance use than adolescents who perceived their parents as neglectful, both concurrently and longitudinally (Adalbjarnardottir & Hafsteinsson, 2001). It has been found that parental influences are the strongest and most direct early in the life of children when experimentation with substances takes place (Griffin, Botvin, Scheier, Diaz & Miller, 2000; Spooner, 1999). The longitudinal research conducted by Stice and Barrera (1992) showed that parental social support and control were generally negatively related to adolescent alcohol use and illicit substance use.

Other studies investigated the role of gender in the development of adolescent alcohol use. Females were found to have a reduced tendency to develop drinking problems across all ethnicities in comparison to males (Griffin et al., 2000). Recent research indicates that rates of alcohol use and alcohol dependence or abuse are higher among males than females (The NSDUH report, 2006). Wagner et al. (2005) found that early onset of alcohol/tobacco use is associated with excess risk of drug use among students of Morelos, Mexico, and that the risk is higher for males. A multinational collaborative epidemiological research study was conducted to estimate the occurrence and school-level clustering of drug involvement among school-attending adolescent youths in each of seven countries in Latin America. It was found that in comparison to females, males were more likely to use alcohol, tobacco, inhalants, marijuana, and illegal drugs; the odds ratio estimates showed high statistical significance and were 1.3, 2.1, 1.6, 4.1, and 3.2, respectively (Dormitzer et al., 2004).

The relationship between adolescents' SES and substance use has been studied by many researchers, but how they are associated remains controversial. A few studies have found that adolescents with low SES have more tendencies towards substance use. For example, Goodman and Huang (2002) showed that low SES was associated with greater alcohol use and greater cigarette and cocaine use among teenagers. When Reinherz, Giaconia, Hauf, Wasserman, and Paradis (2000) studied participants from a 3-year longitudinal study, they found that low SES and larger family size were associated with increased probability of substance abuse disorders in early adulthood. Hamilton, Noah, and Adlaf (2009) analyzed the Ontario Student Drug Use Survey and found that adolescents of ages between 12 and 19 years old who had parents with college degree were less likely to engage in drinking or illicit drug use. On the other hand, in a study of British adolescents Bellis et al. (2007) found that adolescents with more spending money were more likely to drink frequently, and drink in public. Similar results were obtained in a study of US college students where it was found that students with less spending money were less likely to drink and get drunk (Martin et al., 2009). Humensky (2010) analyzed data from the National Longitudinal Survey of Adolescent Health, a nationally-representative survey of secondary school students in the US. Results of the study indicated that higher parental income was associated with higher rates of drinking and marijuana use in early adulthood. Cox et al. (2010) found that youth in higher SES schools had an earlier age of onset for substance use than did youth in lower SES schools.

Overview of research on early onset of substance use. Studying early ages of substance use is very important because early initiation leads to several different negative outcomes in youth and adults. Grant and Dawson (1997) found that age at first drug use was a powerful predictor of lifetime drug abuse and drug dependence. The likelihood of drug abuse and dependence was determined as a function of ages of onset of drug used in a large representative sample of the US population. Numerous studies found that the earlier a child experiences alcohol or other drugs the greater is the risk of becoming involved in various problematic behaviors, which includes school failure, aggression, delinquency and later substance use and abuse (Jackson, Henriksen, Dickinson, & Levine, 1997; Kandel, 1982; Robins & Przybeck, 1985). The timing of adolescents' substance initiation and use is also of concern because the earlier people initiate substance use the greater and more harmful is the later use (Flory, Lynam, Milich, Leukefeld, & Clayton, 2004; DeWit, Adlaf, Offord, & Ogborne, 2000). Early initiation of substance appears to be a powerful precursor of later substance abuse, which is why exploring factors that contribute to early substance use initiation is the focus of many research studies. Kaplow, Curran, and Dodge (2002) examined predictors of early substance use in a longitudinal study of 295 children from kindergarten age to grade 6. Because of the low rate of substance use in each grade, measures from grades 4, 5, and 6 were combined to form an overall dichotomized measure of substance use (0 = no use, 1 = use) at any age. A series of hierarchical linear logistic regressions was used to test relations between several predictors and early childhood substance use. Results of the study indicated that the most significant predictors of early substance use are parenting and child functioning factors as opposed to more distal factors such as the neighborhood environment or socioeconomic status. Maternal parenting techniques ranging from reasoning to physical punishment were coded by frequencies with which mothers mentioned a certain technique. The mean score of verbal reasoning was calculated with higher scores indicating more frequent use of verbal reasoning. Increased verbal reasoning was significantly related to decreased likelihood of substance use initiation by the sixth grade. Findings of the study also suggest that children with parents who are involved in their school activities may be less likely to engage in early substance use. Parental abuse of substances when children were in kindergarten was strongly associated with an increased likelihood of child substance use by the sixth grade. Other studies found that adolescents who are strongly oriented toward their families show lower alcohol and illegal drug use than those adolescents who have weaker familial links (Andrews et al., 1997; Kuther, 2002).

As was previously emphasized in a review of research studies, early age of first use of substance among adolescents is much more alarming than later substance initiation since consequences of early substance involvement lead to greater problems. Consequently, studying the age of first use of substances to discover factors that affect age of first use is important for the prevention of various types of negative adolescent outcome. In order to study event occurrence and its timing (i.e. alcohol initiation in the present study), event history analysis needs to be employed (as was discussed above).

Exploring the age of first opportunity of substance use and understanding it as the early stage of substance involvement was emphasizes in a number of research studies (Van Etten & Anthony, 1997; Van Etten & Anthony, 1999; Benjet et al., 2007; Chen et al., 2005), as was mentioned earlier. When Van Etten and Anthony (2001) studied male-female differences in transitions from first drug use opportunity to first use in their longitudinal study of children 12 years or older, they studied the estimated probability of using a drug given that an opportunity to use the drug has been experienced. The reason for exploring AFO was based on results of their previous work where they found that male-female differences in the prevalence of drug use were due to different probabilities of having initial opportunity to try a drug rather than different probabilities of becoming drug user once the opportunity has occurred (Van Etten & Anthony, 1999). The estimates were based on the age of first opportunity to try a drug vs. age of first use of a drug. When the age of first opportunity to try a drug was equal to age of first use of a drug, it was defined as a rapid transition from the initial opportunity to first use. Then the estimated prevalence proportions and ratios were used to define the magnitude of male-female variation. They noted in the study that availability of time-to-event data to analyze time duration from first opportunity to first drug use would have been valuable for making a more precise definition of rapid transition. In other words, survival analysis was suggested for further research.

A subsequent study was conducted by Wagner and Anthony (2002), where retrospective data were reorganized in person-period records prepared for survival analysis regression. To estimate the relative risk of having an opportunity to try marijuana in relation to prior use of alcohol or tobacco, a discrete-time survival regression model was used with covariate adjustments. Onset of tobacco smoking or alcohol use and opportunity to try marijuana were coded as time-varying characteristics (0 until the event occurrence and 1 after the event occurrence), whereas sex, age at interview and ethnicity were time invariant variables. Results showed that once marijuana exposure had occurred, the probability of initiating marijuana use depended on prior history of using alcohol or tobacco. In another study, Wagner et al. (2005) explored if patterns of the transition from early onset use of alcohol/tobacco to excess risk of drug use among students in the Mexican State of Morelos were similar to those observed in other countries. Early onset of alcohol or tobacco use was defined as first use by age 14 years. Then the cross-sectional data were re-organized into person–year records for survival analysis. To estimate the risk of drug use associated with early alcohol and tobacco use initiation, a Cox model for discrete-time survival analyses was used. Students were stratified by school to control for differences, such as different alcohol/tobacco use school policies. Results showed that male users of alcohol/tobacco were much more likely to use drugs compared with males who did not show early use of alcohol/tobacco. Females who did not show early alcohol/tobacco use were more likely to remain non-drug users.

Understanding the earliest stages of substance involvement is important primarily for prevention strategies that intend to prevent and control substance use and most importantly, substance dependence. Thus the first drug opportunity and the transition to the first drug use are relevant for prevention strategies. Studies using survival analysis in modeling AFO and AFU take into account the reality of early adolescent substance use. The advantages of survival analysis over most commonly used regression analyses are obvious (as discussed earlier) and important when studying timing (or duration of time) of event occurrence. Very few survival analysis studies have taken into account the multi-level structure of the data, which creates the problem of biased estimates of event occurrence when the data have a nested structure.

Overview of research on combination of survival analysis and hierarchical linear modeling. To predict earlier ages of substance use and effects of individual and contextual characteristics requires a combination of survival analysis and multi-level modeling methods if data have a nested structure. Survival analysis is traditionally used in medical and epidemiological studies where hazard rate is the dependent variable of the event history model and sometimes is called a hazard model. Two data analysis tools (even history analysis and multi-level modeling) were first combined by Barber, Murphy, Axinn, and Maples (2000). They developed a discrete-time multilevel hazard model and provide details regarding the assumptions that allow the regression coefficients to be estimated in a multilevel hazard analysis framework. There are only a few studies that investigate contextual characteristics in predicting early use of substances that employ multi-level discrete time hazard models. The Reardon et al. (2002) study was one of the first studies that estimated a multi-level discrete time hazard model. The study demonstrated a methodological approach to estimating contextual effects on substance use initiation using retrospective data. Their study investigated the effects of social context, such as neighborhood, on the timing of cigarette initiation. To accomplish this, the data were reorganized into a retrospective person-period data set from cross sectional data. Each individual record was converted into a number of person-year observations, one observation for every year from age 7 (beginning of the observation time) till censoring or the initiation of cigarette use. All cases were right-censored on the day of the participants' last birthday prior to the interview (More details about the censoring will be discussed in the next chapter). Effects of individual and neighborhood level variables on the likelihood of cigarette initiation at each age were examined. First a person-level discrete time hazard model was estimated and then a two-level discrete time hazard models were estimated. Contextual variables were treated as Level-2 variable and personal and age variables were combined together at Level 1. The study illustrated how these models could be estimated by multi-level software packages that are widely available (e.g., HLM).

There is a growing body of research that uses survival analyses in the social sciences. Browning, Levanthal, Brooks-Gunn (2005) used recently developed multilevel discrete-time event history techniques (Barber et al., 2000; Reardon et al., 2002) to model the onset of sexual behavior, where the dependent measure was the respondent's age at first sexual intercourse. The two-level discrete-time logit model was used to assess the hazard of sexual onset for every person in a certain neighborhood at a certain age. The first set of analyses was focused on individual, family, and peer influences on adolescent sexual behavior. In subsequent models family structural background (SES, composition, and size), family support and supervision, peer influences and developmental risk factors (positive peer attachment, peer deviance, pubertal development, prior problem behavior, sociability, and reading ability) were added to assess the extent to which these person-level factors account for racial and ethnic differences in the timing of first intercourse. In a study conducted by Bradshaw, Buckley, and Ialongo (2008) discrete-time survival analysis was used to model two service use variables (mental health, special education) in Grades 1–9 as a function of early symptom class membership. For this study, the event of interest was defined as the first receipt of services (mental health, special education) for each student. The timing of the event was recorded in discrete-time intervals (grade of first service receipt) so that discrete-time analysis could be used for modeling this event. The hazard probability of first service use in a given grade (i.e., the probability of a student experiencing a service use in that grade provided that a student had not used a service in an earlier grade) was related through a logistic function to early problem class membership. The nine variables that captured each student's grade in which he or she first used mental health services (between first and ninth grade) were coded either 1 if the service occurred or 0 if it did not. Once service use occurred, the remaining binary indicators were coded as missing, because the focus was on the first service use (because of its implication for later problems and relevance for school-based preventive interventions). The same procedure was used to create the special education service use survival variables. Child-level covariates of race, sex, and free or reduced-cost meal status (in first grade) were also controlled. Kim and Gray (2008) used three-wave panel data from the Domestic Violence Experience in Omaha, Nebraska. This study employed a discrete-time hazard model to examine a woman’s decision to leave the situation in which violence occurred based on four factors: financial independence, witness of parental violence, psychological factors, and the police response to the domestic violence call. Before implementing the discrete-time hazard model procedure, the three-wave panel data set was rearranged to create a person-period data set for the analyses. Each respondent had a separate observation for every wave until the event of interest (i.e., leaving). Then a series of discrete-time hazard models were calculated to address the hypotheses of that study.

Hawkins et al. (2008) explored effects of the Community Care intervention on initiation of delinquent behavior and substance use. Multilevel discrete-time survival analysis was used to assess the effects of the intervention on preventing the initiation of delinquent behavior and substance use between grades 5 and 7. Onset of substance use was measured by items consisting of the first reported lifetime use of any of four types drugs: alcohol, marijuana, cigarettes or other illicit drugs. Onset on delinquent behavior was measured as first occurrence of any one of nine types of delinquent behavior. Students who did not initiate delinquent behavior or substance use, respectively, during sixth or seventh grades were treated as right-censored observations (i.e., never experienced the event). To assess effects on students who had not yet initiated these behaviors, students who had already initiated delinquent behavior (22.2%) or substance use (27.5%) prior to the intervention were not included in the analyses. The analysis was implemented using the logit function for the dichotomous outcomes. Student- and community-level variables were included in the model as covariates to control for possible community differences; the intervention condition was included in the model as a community-level variable; and random effects were included to account for variation among students within communities, communities within matched pairs of communities, intervention effects across matched pairs of communities, and residual error. The effect of the intervention was estimated as the adjusted within-matched pair difference in community-level hazard of onset between the intervention and control communities, assuming proportional hazards over time, and was tested against the average variation in hazard of onset among the matched pairs of the intervention and control communities. In the present study time variable and person-level variables are separated at two different levels.

Other studies used survival analysis to model early substance use but not in a multi-level framework. Some studies that have not used survival analysis have identified parental involvement, monitoring, and support as protective factors that influence alcohol use. However, no studies have combined survival analysis with risk and protective factors with nested data with a multi-level analysis to provide unbiased estimates of substance initiation the way it is done in this study. Furthermore, to my knowledge no study accounted for drug exposure, the importance of which was discussed in several research studies, within this approach that combines discrete-time event history analysis with multilevel structure. This methodological approach combines the best features of linear multiple regression and logistic regression and takes advantage of multilevel modeling characteristics that allows for more accurate error estimations and consequently more accurate predictions. In addition this method handles unknown events, i.e., missing data, the most accurately.

Different statistical analyses lead to different estimates and results. The present study will investigate three different approaches in investigating substance use initiation where the major model will be the recently developed multi-level version of survival analysis, which combines the best features of logistic and linear regressions within a multilevel framework. The main objective of this study is to explore and compare this model with the more traditional statistical techniques of linear and logistic regressions.

In the approach where the combination of survival analysis and hierarchical linear modeling is used there are five models that will also be compared with each other. It can be hypothesized that in the model where the beginning of the risk period starts at the age of first opportunity (AFO) the hazard rate of alcohol initiation will be higher compared to the model when the beginning of the risk period is set at age 4 (Based on the research studies reviewed about significance of AFO). The present study was more exploratory and we did not put much emphasis on formulating precise hypotheses.

CHAPTER III

METHODS

This chapter presents the information about study participants, questionnaires that were used, description of how study variables were operationalized and data analysis strategies. The data for this study were collected in Venezuela for Dr. Ronald Cox's dissertation in 2007. Thus all the information about study participants, schools, and questionnaire used has already been reported in his dissertation (Cox, 2007).

Conceptual Definition of Variables

Initial opportunity to try alcohol and initial alcohol use are assessed by individual level variables, Age of first opportunity of using alcohol (AFO) and Age of first use of alcohol (AFU), correspondingly.

Age of first opportunity of using alcohol (AFO) **–** is conceptualized retrospectively by posing the question about the age when a participant remembered having a first opportunity to use alcohol.

Age of first use of alcohol (AFU) **–** is conceptualized retrospectively by asking questions about the age of first use of alcohol.

Family Attention (FA) **–** is conceptualized as the extent to which parents or guardians monitor their youth's behavior and demonstrate positive communication with their offspring.

Externalizing Behavior (EXT) – is conceptualized as the extent to which an adolescent participates in delinquent behavior and engages in risky behaviors.

Socioeconomic Status (SES) – is conceptualized as the positioning of the adolescent in a social structure by evaluating a type of neighborhood of residence (housing project, apartment, or house), level of parental/guardian education, number of people and bedrooms in their residence, and number of vehicles owned by the family.

Participants

A total of 1,831 students of ages 11 to 19 from 14 schools were surveyed in two school districts in Caracas, Venezuela. To control for false responses, questions on the first use of a fake drug (*Cadrina*) were included in the questionnaire. Among the 1,831 respondents, only 8 (0.4%) reported use of *Cadrina*. Under the assumption that mis-statements about a fake drug may signal for falsely positive reports about other drug experiences or general response errors in the completed questionnaires, they were excluded from the study. Three students that had more than 50% missing data were excluded (Cox, 2007). Thirty participants reported having age of first use of alcohol before they had an opportunity to use it and were excluded from the study. Additionally, five participants who reported using every substance once a day or more were considered outliers and were excluded. This left a total sample of 1,785 respondents.

Of the 1,785 students included in the analysis 945 respondents (52.9%) were female and 18 (1%) did not report gender. The question about age had students place themselves into one of five age cohorts. The first age cohort was from ages 11 to 12 (5.9%, *n*=105). The second age cohort was 13 to 14 (33.0%, *n*=589). The third age cohort was from 15 to 16 (40.1%, *n*=715). The fourth age cohort was 17 to 18 (19.9%, *n*=355).

The fifth cohort was age 19 or above (1.0%, *n*=20). Only one person (.1%) did not respond to the item regarding age.

A total of 828 students (46.4%) were from private schools. High schools in Venezuela include five grades, 7th-11th. The sample was approximately equally distributed among the five grades with *n*=336 in 7th, *n*=334 in 8th, *n*=373 in 9th, *n*=367 in 10th, and *n*=373 in 11th. The modal response to other demographic questions indicated that the most typical participants lived in the poorest housing area (*n*=1001, 55.4%), their families did not own a vehicle (*n*=706, 39.8%), lived in a home with 2-3 bedrooms (*n*=1096, 61.7%), and had 4-6 people living in their home (*n*=1090, 61.3%). Many respondents reported educational levels of the father and mother as having finished a post high school degree (*n*=540, 30.3% and *n*=536, 30.2% respectively) and 35.5% of fathers and 35.3% of mothers were reported as not having finished high school.

Measures

The PACARDO-V questionnaire. Data for this study were collected using a modified version of the PACARDO questionnaire (which stands for **PA**nama, **C**entral **A**merica, and **R**epublica **Do**mincana) questionnaire. The PACARDO was developed for use in a NIDA-funded grant "Cross-National Research in Clusters of Drug Use" (Dormitzer et al., 2004), and is a standardized self-administered questionnaire. It was previously administered to nationally representative samples of students in Central America, Panama, and the Dominican Republic and has 224 items (see Dormitzer, et al., 2004; and Dormitzer, 2004, for more information including psychometric properties). The primary instrument employed in the present study, the PACARDO-V (with the addition of the V for Venezuela) was adapted from the original PACARDO. The final version, PACARDO-V contains 112 items. Additionally, items from the PACARDO were modified in the PACARDO-V to reflect idiosyncrasies of the Venezuelan culture and language use and then pilot tested on this population (Cox, 2007).

Dependent variable in the study. Age of first alcohol use (AFU) was conceptualized as having tried alcohol the first time as measured in response to the standardized item, “How old were you first drank alcohol?” AFU is a continuous variable that ranged from 1 to 18 (0 = never used). Reported Age of AFU is continuous variable that ranged from 2 to 18. Frequencies of reported ages are presented in Table 3.1. and graphically in Fig. 3.1.

Table 3.1.

Age of First Use of Alcohol

AFU of Alcohol		
Age	Frequency	Percent (%)
0[a]	333	18.70
2	1	.10
3	3	.20
4	4	.20
5	7	.40
6	7	.40
7	13	.70
8	34	1.09
9	47	2.60
10	143	8.00
11	124	6.90
12	270	15.10
13	280	15.70
14	247	13.80
15	186	10.40
16	46	2.60
17	29	1.60
18	2	.10
Subtotal	1776	99.50
Missing	9	.50
Total	1785	100.00

Note: [a] Age of 0 indicates never used alcohol according to the respondent

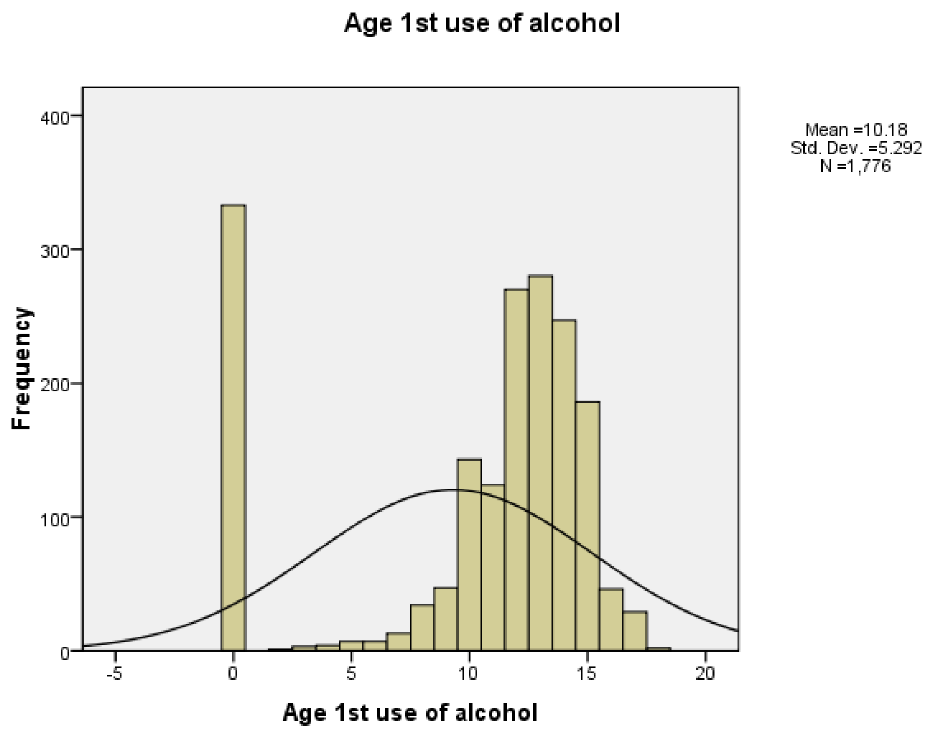


Fig. 3.1. Histogram of Age of First Use of Alcohol

Independent variables. Independent variables used in the study are: Age of First Opportunity (AFO) to use alcohol, Family Attention (FAM), Externalizing Behavior (EXT), Socioeconomic Status (SES), and Gender. These variables are chosen based on a literature review showing their significance as predictors or control variables of adolescence substance use.

Age of first opportunity of alcohol use (AFO) was conceptualized as having an opportunity to try alcohol as measured in response to the standardized item, “How old were you when you first had an opportunity to drink alcohol?” Reported AFO is a continuous variable that ranged from 2 to 16. Frequencies of reported ages are presented in Table 3.2 and graphically in Fig. 3.2.

Table 3.2.

Age of First Opportunity of Alcohol Use

Age of First Opportunity of Alcohol Use		
Age	Frequency	Percent (%)
0[a]	207	11.60
2	1	.10
3	3	.20
4	9	.50
5	23	1.30
6	21	1.20
7	33	1.80
8	67	3.80
9	92	5.20
10	214	12.00
11	154	8.60
12	309	17.0
13	261	14.60
14	196	11.00
15	144	8.10
16	32	1.80
17	11	.60
18	1	.10
Subtotal	1778	99.60
Missing	7	.40
Total	1785	100.00

Note: [a] Age of 0 indicates no opportunity of using alcohol according to the respondent

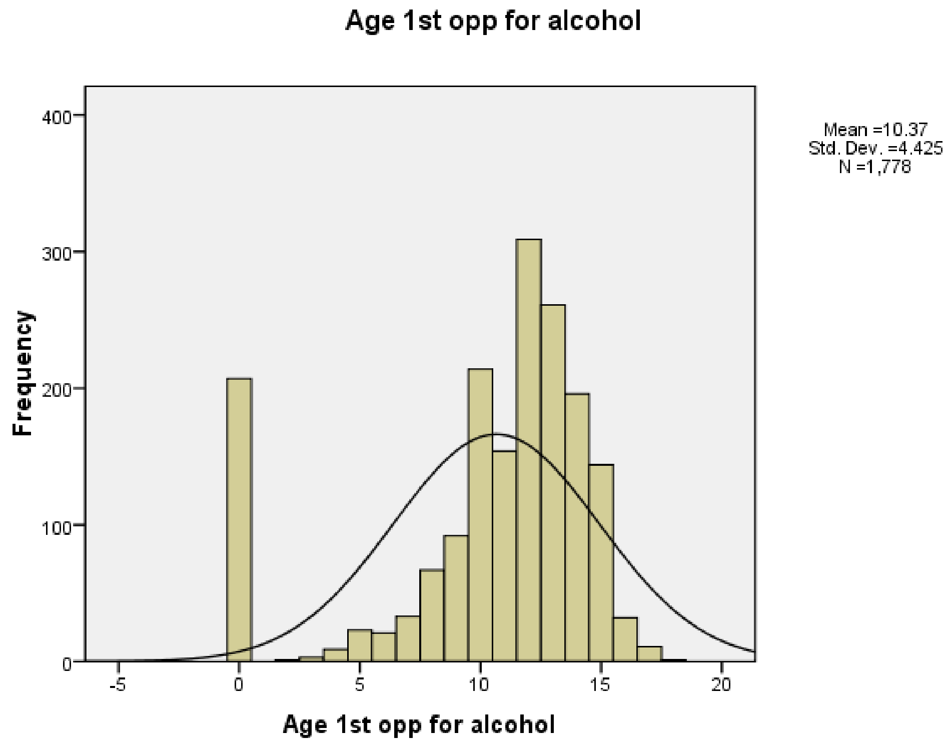


Fig. 3.2. Histogram of Age of First Opportunity of using Alcohol

Family Attention, Externalizing Behavior, and Socioeconomic Status are operationalized as empirically derived composite scores, or indexes. Items that are selected for indexes are purposely selected to correlate to some external criteria and not necessarily to each other, in contrast to scales (Cox, 2007; Streiner, 2003). Thus some reliability estimates, such as Chronbach's α coefficient, which measures internal consistency, or how items correlate with each other, might be negatively biased, or below a recommended reliability estimates, i.e. below .70. (Feldt & Charter, 2003).

Family Attention. The Family attention scale was adapted from Capaldi and Dishion (1988). Family Attention is conceptualized as a combination of items reflecting parental or guardians monitoring, positive affect and communication. Family Attention is measured by the following eight items from the PACARDO-V questionnaire:

V14. Are your parents or guardians aware of what you think or feel about things that are important to you?

V15. Are your parents or guardians aware of your likes and /dislikes?

V16. I always ask my parents for permission when I go out and have fun.

V17. Do you feel that your parents or guardians care about you?

V18. Are your parents or guardians often aware of where you are and what you are doing?

V25. Do you frequently have discussions with your parents/guardian that end in fights?

V26. My parents or guardians are always talking to me about how dangerous drugs are.

V27. My parents or guardians are always talking to me about how dangerous alcohol and cigarettes are.

Each item on the scale is scored as yes/no response (no = 1; yes = 2) such that higher scores indicate more family attention (V25 was recoded to be in the same direction as the rest of indicators). For each observation, scores on these eight items are summed to create an index and then standardized for making it easily interpretable (mean = 0, *SD* = 1.0). Descriptive statistics and reliability coefficient, Chronbach's α, are presented in Table 3.3 and graphically in Fig.3.3.

Table 3.3

Descriptive statistics of Family Attention

	Family Attention
Frequency	1519
Missing	296
Mean	.0000000
Std. Deviation	1.00000000
Chronbach's α (8)	.63
Original reliability α (8)	0.70
Skewness	-1.274
SE (skew)	.063

Note: (8) indicates 8 items
Original reliability is based on PACARDO (Dormitzer, 2004)

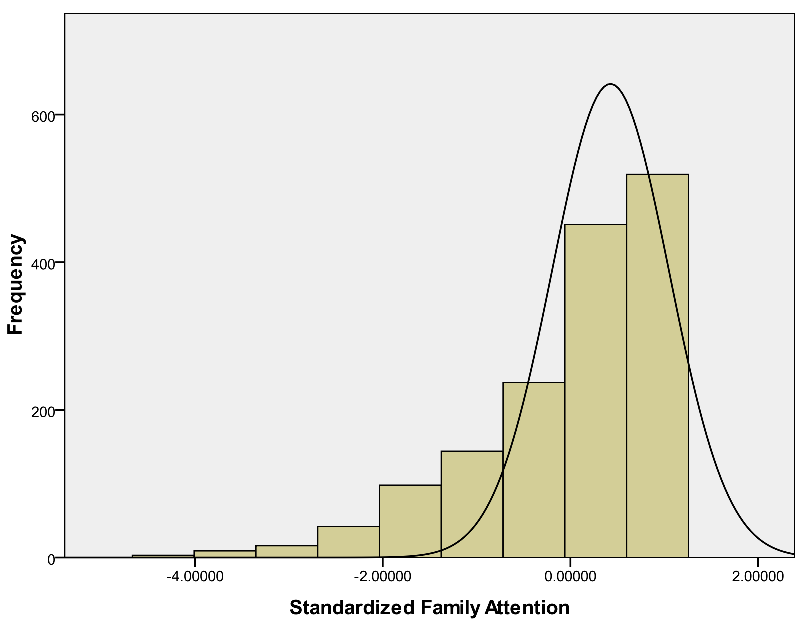


Fig.3.3. Histogram of Family Attention

In addition confirmatory factor analysis (CFA) was performed to insure an adequate relationship between the items and the Family Attention construct defined by them. Confirmatory factor analysis was implemented using Mplus v5.1. statistical software. Model fit was determined by four well-known fit indices that assess the magnitude of the discrepancy between the sample and fitted covariance matrices: Chi-Square, Comparative Fix Index (CFI), Tucker-Lewis Index (TLI), and Root Mean Square Error of Approximation (RMSEA). With large samples adequate fit is indicated by a normed Chi square (χ^2 model/*df*) ≤ 5 (Bollen, 1989), CFI > .95, TLI > .90 (Hu & Bentler, 1999), and/or a RMSEA $\leq .05$ (Kline, 2005) All fit indexes, indicated an adequate model fit (RMSEA 0.044, CFI 0.989, TLI 0.980). The Table 3.4 records the factor loading of the times on this latent construct. All loadings are above the conventional cut off point of .30.

Table 3.4

CFA for Family Attention

Construct	Observed V	Estimates
Family Attention	V14	.654
	V15	.589
	V16	.530
	V17	.395
	V18	.490
	V25[a]	.424
	V26	.546
	V27	.560

[a] Reversed scored

Externalizing Behavior. Externalizing behavior (EXT) on the original PACARDO was adapted from the Drug Use Screening Inventory (Tarter & Hegedus, 1991) for use in research on non-clinical samples (as cited in Cox, 2007). EXT is operationalized as a composite score on the following items from PACARDO-V questionnaire:

V38. Have you ever belonged to a gang?

V40. Have you intentionally damaged another person's belongings during the last school year?

V41. Have you stolen anything during the last school year?

V42. Have you done anything risky or dangerous during the last school year?

V43. Is it true that the majority of the time you don't do your homework?

V47. Have you skipped school two or more days in a single month during the last school year?

V48. Have you ever been suspended from school?

V49. Have your grades gotten worse during this past year?

V51. I have seriously thought about dropping out of school.

Each item on the scale was scored as a yes/no response (no = 1; yes = 2) such that higher scores indicate more externalizing behavior problems. For each participant, scores on these nine items are summed up to create an index and then standardized for easier interpretation (mean = 0, *SD* = 1.0). Descriptive statistics and Chronbach's α, are presented in Table 3.5 and graphically in Fig.3.4.

Table 3.5

Descriptive statistics for Externalizing Behavior

Externalizing Behavior	
Frequency	1671
Missing	144
Mean	.0000000
Std. Deviation	1.00000000
Chronbach's α (9)	.64
Original reliability α(19)	.83
Skewness	.816
SE (skew)	.060

Note: (9) indicates nine items;
Original reliability is based on PACARDO (Dormitzer, 2004)

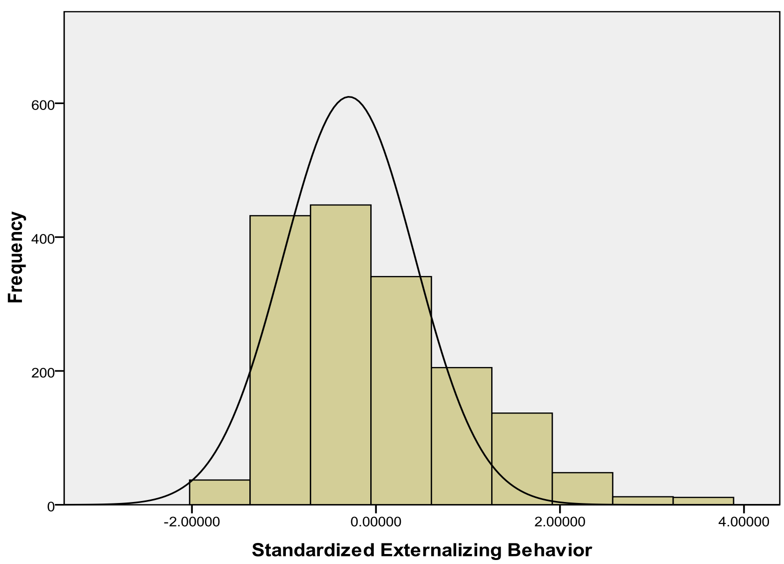


Fig. 3.4. Histogram for Externalizing Behavior

A confirmatory factor analysis was employed to test for the fit of the items to the Externalizing Behavior construct. CFA revealed an adequate fitting model (χ^2 /df < 5). Additional fit indexes also suggested a good fit RMSEA (.045), CFI (.981), and TLI (.961). CFA showed good factor loadings for the Externalizing Behavior latent construct (all above .40), presented in Table 3.6.

Table 3.6

CFA for Externalizing Behavior

Construct	Observed V	Estimates
Externalizing Behavior	V38	.653
	V40	.586
	V41	.530
	V42	.398
	V43	.489
	V47	.428
	V48	.541
	V49	.565
	V51	.418

Socioeconomic Status. Socioeconomic Status (SES) was measured as a composite score of the following five items from PACARDO-V questionnaire:

V6. What type of neighborhood do you live (ordinal scale scored 1-3).

V7. How many vehicles does your family have (ordinal scale scored 1-5)?

V9. How many bedrooms does your house have (ordinal scale scored 1-5)?

V12. What academic grade did your father (or the person who is like your father) achieve (ordinal scale scored 1-5)?

V13. What academic grade did your mother (or the person who is like your father) achieve (ordinal scale scored 1-5)?

For each observation, scores on these five items are summed and then the sum standardized for interpretability (mean = 0, *SD* = 1.0). Larger number on scale indicated better neighborhood, more vehicles, bedrooms, and higher level of education. Descriptive statistics of SES are presented in Table 3.7 and graphically in Fig. 3.5.

Table 3.7

Descriptive Statistics for Socioeconomic Status

	SES
Frequency	1774
Missing	41
Mean	.0000000
Std. Deviation	1.00000000
Chronbach's α	.64
Skewness	.116
SE (skew)	.058

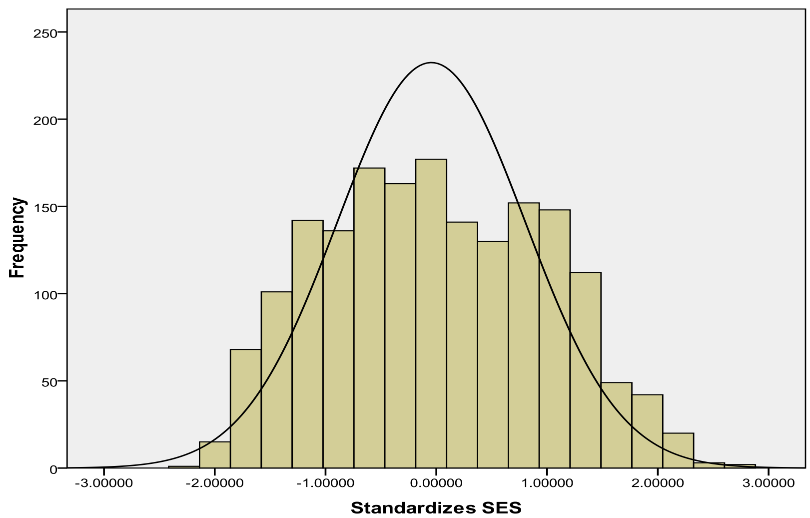


Fig. 3.5. Histogram for SES

A confirmatory factor analysis for SES showed an adequate fitting model (χ^2 /df < 5). Additional fit indexes also suggested a good fit RMSEA (.035), CFI (.978), and TLI (.981). CFA showed adequate factor loadings for Socioeconomic Status latent construct, as presented in Table 3.8.

Table 3.8

CFA for SES

Construct	Observed V	Estimates
	V6	0.653
	V7	0.589
SES BY	V9	0.52
	V12	0.398
	V13	0.49

In this study I considered three different approaches to model the outcome of our interest (alcohol initiation), which are: 1) combination of Survival Analysis with Hierarchical Linear Modeling; 2) Multiple Regression; and 3) Logistic Regression. Before analyzing different statistical models I utilized descriptive analysis to characterize the sample.

Descriptive Statistics

Determining age of participants. Exact ages of the participants when they were interviewed were not readily available from the data, because they indicated which of five age cohorts they belonged to in 2-year age groupings. Which of the two ages in any age cohort was estimated by cross tabulating age by their reported school grades. Table 3.11 shows the most common ages in each grade, ranging from age 13 for 7th graders to 17 for 11th graders. Children in each grade were assigned the modal age for that grade or the nearest age to that modal age that fell within their reported 2-year age cohort. Table 3.11 shows the estimated interview age for each combination of grade and age cohort.

Table 3.9

Crosstabulation of Age by School Grade

Age		School grade 7th	8th	9th	10th	11th	Total
11-12	Count	103	4	0	0	0	107
	%within age	96.3%	3.7%	0.0%	0.0%	0.0%	100.0%
	Estimated interview age	12	12				
13-14	Count	**207**	**271**	114	5	2	599
	%within age	34.6%	45.3%	19.0%	0.8%	0.3%	100.0%
	Estimated interview age	13	14	14	14	14	
15-16	Count	28	60	**242**	**280**	115	725
	%within age	3.9%	8.3%	33.4%	38.7%	15.8%	100.0%
	Estimated interview age	15	15	15	16	16	
17-18	Count	1	9	22	86	**243**	361
	%within age	0.3%	2.5%	6.1%	23.7%	67.5%	100.0%
	Estimated interview age	17	17	17	17	17	
≥19	Count	0	0	0	3	17	20
	%within age	0.0%	0.0%	0.0%	15.0%	85.0%	100.0%
	Estimated interview age				19	19	
Total	Count	339	344	378	374	377	1812
	%within age	18.7%	19.0%	20.8%	20.6%	20.8%	100.0%

Note. Boldface used to indicate the modal age group for each grade.

Descriptive analyses were performed to describe the sample by demographic characteristics in different ways: to assess the opportunity to use alcohol and alcohol initiation by gender, age cohorts, socioeconomic status, and school type (public vs.

private), to calculate the percentage of the sample that reported alcohol initiation before the interview year and during the interview year, which shows what part of the sample will be dropped (the reasons for discarding some information about the sample is discussed later).

Combination of Survival Analysis and Hierarchical Linear Modeling

Three models were constructed when we used this recently developed methodology of combining survival analysis and hierarchical linear modeling. The models that were developed utilizing this analysis are known in literature as Hazard Models because they estimate the hazard rate of event occurrence. The description of how models were developed is presented below.

Model 1 - Time from first opportunity to first use. A combination of hierarchical linear modeling with discrete-time survival analysis will be used to predict the time between the age of first opportunity of using alcohol to the first use of alcohol. This uses a two-level hierarchical linear model with censored data.

First the base model was tested to investigate the hazard rate of alcohol initiation by years from the age of first opportunity, which are considered as occasions nested within students. Predictors considered as second level variables are person-related variables: family attention, externalizing behavior, socioeconomic status, age of first opportunity of using alcohol, and gender. Effects of all these independent variables on the hazard rate of alcohol initiation were tested.

Combination of survival analysis and HLM. The discrete time event history analysis requires a person-period data set (Singer & Willet, 2003). The construction of the data set is adapted from Reardon et al. (2002). For each individual, the data contains a set of dummy variables starting with the AFO year and including each year after the AFO, until either the first use of alcohol (AFU) or the year before the interview year, whichever comes earlier. This way, each year from AFO until either AFU or their age at the interview is given a dummy variable to indicate whether alcohol initiation occurred during that year. Dummy variables indicate whether an individual ever initiated alcohol and the age at which that occurred, specifically in how many years after AFO.

Participants who did not initiate alcohol during the observation period are censored. According to Singer and Willett (2003), the validity of hazard analysis is based on the assumption that censoring is noninformative and right-censored. Censoring is noninformative if it is independent of event occurrence. In the present study all participants who remained after the censoring date are assumed to be representative of everyone who would have remained without censoring because the observation period ended. In the present study right censoring is used because the duration of time until initiation is not known because the event occurrence of alcohol initiation has not been observed. Censoring occurs at the same point in time for all individuals, which is also called fixed censoring and making any further assumption about the nature of censoring is unnecessary (Allison, 1984). Thus the partial age year they were interviewed, all years after their interview year, and all years after the first year of alcohol use (AFU), i.e., alcohol initiation, will be dropped. Following Reardon et al. (2002), partial age year was dropped so that it would not influence (downward) the hazard rate at that age because those individuals who did not report initiating alcohol at the interview year should have also remained in the sample and would have been treated as non-initiators when in fact they could have initiated alcohol later that year had they been observed for the entire year. Even though some data from the sample would be discarded, it reduces the bias of the estimated hazard rate, which leads to a more accurate estimation of the model.

A sample of the data is presented in Table 3.12 to visualize censoring and timing of alcohol initiation, also called event occurrence.

Table 3.10

Sample Person-Period Data Set

ID	AFO	AFU	Age at Interview	Years from AFO (dummy variables) Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Alcohol Initiation
1	12	13	14	1	0	0	0	0	0	0
1	12	13	14	0	1	0	0	0	0	1
2	11	15	15	1	0	0	0	0	0	0
2	11	15	15	0	1	0	0	0	0	0
2	11	15	15	0	0	1	0	0	0	0
2	11	15	15	0	0	0	1	0	0	0
3	8	13	16	1	0	0	0	0	0	0
3	8	13	16	0	1	0	0	0	0	0
3	8	13	16	0	0	1	0	0	0	0
3	8	13	16	0	0	0	1	0	0	0
3	8	13	16	0	0	0	0	1	0	0
3	8	13	16	0	0	0	0	0	1	1
4	13	0[a]	19	1	0	0	0	0	0	0
4	13	0	19	0	1	0	0	0	0	0
4	13	0	19	0	0	1	0	0	0	0
4	13	0	19	0	0	0	1	0	0	0
4	13	0	19	0	0	0	0	1	0	0
4	13	0	19	0	0	0	0	0	1	0

[a]Right-censored (missing) AFU.

For each student, the first row represent the year of their AFO. The last row represents either the year they initiated alcohol use (AFU) or the year prior to their age at the time of the interview (due to right censoring of the data as discussed above). For example, for participant 2 the information about his or her alcohol initiation is discarded because AFU, or alcohol initiation, happened at the age of the interview and therefore is censored. For participant 3, the last year from AFO shows that six years have passed since AFO before he or she initiated alcohol. This type of censoring gives unbiased estimate of hazard rates as shown by Malacane, Murphy, and Collins (1997).

The research objective of the study is to examine the effect of years from the AFO of alcohol use and also the effects of person-level variables, such as family attention, externalizing behavior, socioeconomic status, and gender (independent variables described above) on the likelihood of initiation of alcohol at each age. To accomplish this task, event history analysis needs to be combined with multi-level modeling, which is also called hierarchical linear modeling (HLM) according to Raudenbush and Bryk (2002). Employing hierarchical linear modeling with censored data is considered to be the most appropriate method for investigating predictors of the hazard rate of event occurrence (Reardon et al., 2002). Again, the rationale for taking a hierarchical linear approach is to account for the nested structure of the data. Dummy coded years from AFO are nested within persons and all above mentioned independent variables and are incorporated as Level-2 variables (distinct from Reardon et al., 2002, who put time- and person-level variables at Level-1). By arranging the variables this way we take the full advantage of multilevel modeling, which incorporates different error terms for different levels resulting in Type 1error rates than nonhierarchical methods (Singer & Willet, 2003). In this study two-level hierarchical linear modeling with censored data will be used and estimated by the HLM version 6 statistical program.

First, a discrete-time hazard model would be estimated to observe the hazard rate of alcohol initiation based on years from AFO for every year until initiation or censoring. The most convenient way to estimate this Level-1 model is the logit link function because this model is binary (Raudenbush & Bryk, 2002). Definitions of the logit and link function can be found in Raudenbush and Bryk (2002). By definition, the logit function in logistic regression is a special case of the link function for a binary outcome variable. The link function itself is a transformation and is used to model responses when a dependent variable is supposed to be nonlinearly related to the predictors. The Logit link function is represented by the equation:

$\eta_{ij} = \log (h_{ij} / (1- h_{ij}))$, where h_{ij} is the hazard of alcohol initiation for person i at year *j* from AFO.

In fact, logit is the log-odds of h_{ij}, where odds are the ratio of two probabilities for any mutually exclusive binary states (as was defined before).

Predictors considered as second level variables will be person-related independent variables: AFO of alcohol use, family attention, externalizing behavior, socioeconomic status, and gender. Their effect on the hazard rate of alcohol initiation will be analyzed.

Model 2(a) - Time from age 4 to first use. The differences between this model and Model-1 are: 1) the beginning time in Model 2 is set at age 4 and not at AFO of alcohol (which is the beginning time in Model-1) and 2) controlling for AFO at level 2 when AFO is recoded so that participants who reported not having an opportunity of having alcohol are estimated to have their first opportunity at the year after the year they were interviewed. This way of recoding for AFO seems to be reasonable because it takes into consideration the distribution of AFO. We can see from Fig 3.1 that the mode of the distribution is 12 after which the frequency of AFO decreases. Recoding AFO for those students reporting no opportunity to the age of the interview plus one year roughly mimics the shape of the distribution of AFO and the possibility of having the opportunity to try alcohol the year after the interview. Person related variables, family attention, externalizing, behavior, gender and socioeconomic status will be covariates at level-2. Here also a combination of hierarchical linear modeling with discrete-time survival analysis will be used to predict the time from the age of 4 to the first use of alcohol. The effect of the same covariates on the hazard rate of alcohol initiation will be also analyzed.

Model 2(b) – Without controlling for AFO. This model is analogous to *Model-2(a)* with the exception that it does not control for the effect of AFO of alcohol (at level-2) on the hazard rate of alcohol initiation. Here also a combination of hierarchical linear modeling with discrete-time survival analysis will be used to predict the time from the age of 4 to the first use of alcohol and possible effects of family attention, externalizing behavior, gender, and socioeconomic status on the hazard rate of alcohol initiation.

Multiple Regression Analysis Model

The main response variable in this second approach where we used multiple regression analysis is the year of first use of alcohol, which was described above. According to Cox (2007) youth reporting first drug use from ages 2-3 would most likely need to rely on a third person report due to memory limitations of very young children (i.e., 1-3). There were four cases that reported such a young age of alcohol consumption and they will be discarded from the analysis. Due to the high frequency of reporting "never used alcohol" (coded as 0), the distribution of AFU of alcohol is bimodal. Therefore, regression analysis cannot be used since bimodal distribution violates the assumption of homogeneity of variance and normality. Even more importantly, a valid score on a relevant age is missing and thus cannot be included in multiple regression analysis of the age of first use. To adjust for this assumption, cases reporting AFU as 0 will be excluded from the analysis. Therefore multiple regression analysis will accommodate the modeling of AFU and make it possible to evaluation the effects of independent variables on the age of first use of alcohol. Similar to Model 2, multiple regression was done with and without incorporating the age of first opportunity as a predictor. I did not take into consideration the fact that students were nested within schools as did Cox (2007) in his study, to make it easier to compare multiple regression models with previously introduced discrete time hazard models.

Logistic Regression Analysis Model

The main response variable, AFU of alcohol was dichotomized when we constructed a model using logistic regression analysis. Participants who reported no use of alcohol were coded as 0 and all other responses were coded as 1. Logistic regression analysis was utilized to evaluate the influence of the age of first opportunity to use alcohol, family attention, externalizing behavior, gender, and socioeconomic status on alcohol initiation. Again, this approach only allows determining alcohol initiating vs. no-initiating by the time of the interview. Logistic model is represented by the following equation:

$$\mathrm{logit}(P(Y = 1)) = \ln[(P(Y = 1)/P(Y = 0)] = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_n X_n,$$

where P is the probability of event occurrence (alcohol initiation); Y is the indication of alcohol initiation (yes=1, no=0); X_k are independent variables; and b_k are logistic regression coefficients, i.e., parameter estimates.

In the present approach covariates are the same as in previous models, i.e., the age of first opportunity to use alcohol, family attention, externalizing behavior, gender, and socioeconomic status. The same logistic regression model without incorporating the age of first opportunity to use alcohol was also estimated.

CHAPTER IV

RESULTS

Descriptive Statistics

The descriptive analyses were conducted to provide more detailed information about variability of alcohol use and opportunity to use alcohol by some demographic characteristics of the sample. Among 942 females included in the study, 183 (19.4%) reported never having tried alcohol, and out of 816 males, 146 (17.9%) reported never having tried alcohol, as shown in Fig. 4.1.

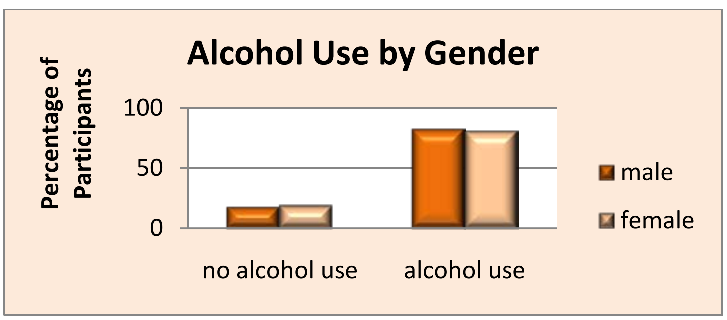


Fig. 4.1. Estimated Percentages of Participants Using Alcohol by Gender

Out of 942 females who participated in this study 12.5% (118) of them never had an opportunity to try alcohol, whereas out of 818 male participants only 10.6% (87) males reported having no opportunity to try alcohol as shown in Fig. 4.2.

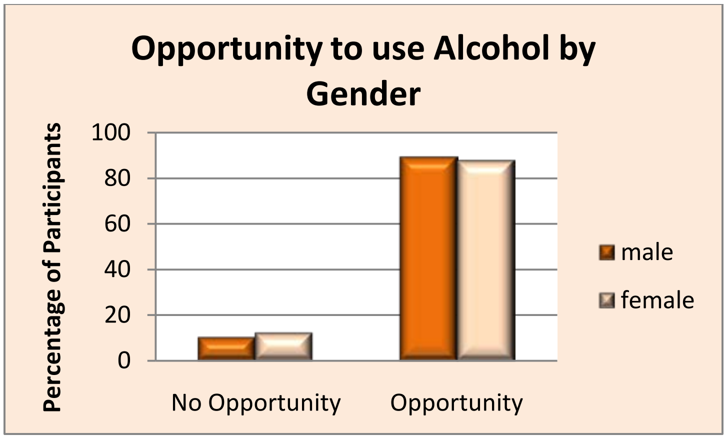


Fig 4.2. Estimated Percentages of Participants Who Had the Opportunity to Use Alcohol

The mean age of opportunity to try Alcohol was about 11.7 years old. Table 4.1 depicts what percentage of the sample had an opportunity to try alcohol in relation to gender and other demographic variables such as age cohorts and school type (public vs. private). A larger percentage of males and youth in private schools were exposed to opportunities to use alcohol. However, chi-square tests of independence did not reveal any significant statistical differences between genders and types of school on the opportunity to use alcohol (χ^2 [1] = 1.5, p = 0.21 and χ^2 [1] = 1.3, p = 0.24 respectively).

Table 4.1.

Demographics by Alcohol Use Opportunity

	Alcohol Use Opportunity				
	Total Youth	No opportunity		Opportunity	
	N (100%)	N	%	N	%
Age in years (mean (SD))				11.74[a]	(2.5)[b]
11-12	104	30.0	28.8	74.0	71.2
13-14	587	104.0	17.7	483.0	82.3
15-16	712	53.0	7.4	659.0	92.6
≥17	374	20.0	5.3	354.0	94.7
Gender					
Female	942	118.0	12.5	824.0	87.5
Male	818	87.0	10.6	731.0	89.4
SES					
Low	954	112.0	11.7	842.0	88.3
High	784	88.0	11.2	696.0	88.8
Type of School					
Public school	954	119.0	12.5	835.0	87.5
Private school	824	88.0	10.7	736.0	89.3

[a] Mean age.
[b] standard deviation.

Table 4.2 shows the percentage of the sample that initiated alcohol and never initiated alcohol (with and without given an opportunity to try it) by gender, age cohorts and school type. The average age of first alcohol use was about 12.5 years old. Alcohol use seemed to be more common among males than females, though the difference was not significant (χ^2 [1] = 0.7, p = 0.41). There were no significant differences in alcohol use by type of school (χ^2 [1] = 0.6, p = 0.44) or socioeconomic status (SES) (χ^2 [1] = 0.5, p = 0.48). Among those who reported an opportunity to use alcohol, there are no gender differences in initiation of alcohol use (92.1% females initiated alcohol compared to 91.9% of males who initiated alcohol).

Table 4.2

Demographic Differences in Lifetime Prevalence of Alcohol Use

	Lifetime Prevalence of Alcohol Use							
	Never Used		Used		Total 100%	Never Used (given the opportunity)		Total (given opportunity) 100%
	N	%	N	%	N	N	%	N
Age (years) (mean (SD))			12.5[a]	(2.2)[b]				
11-12	46	44.2	58	55.8	104	16	21.6	74
13-14	154	21.2	432	73.7	586	50	10.4	482
15-16	94	13.2	617	86.8	711	41	6.2	658
≥17	39	10.4	335	89.6	374	19	5.4	354
Gender								
Female	183	19.4	759	80.6	942	65	7.9	824
Male	146	17.9	670	82.1	816	59	8.1	729
SES								
Low	184	56.6	768	80.7	952	72	8.6	840
High	141	18.0	643	82.0	784	53	7.6	696
School of attendance								
Public school	185	19.4	768	80.6	953	66	7.9	834
Private school	148	18.0	675	82.0	823	60	8.2	735

[a] Mean age.
[b] standard deviation.

Model Comparisons

To model the age of first use of alcohol, three applications of discrete-time hazard models were conducted. Alcohol initiation was also analyzed using the more traditional statistical analyses of multiple regression and logistic regression for comparative purposes. All survival analyses represent modified versions of the Reardon et al. (2002) study and were implemented using the HLM software program. In contrast to Reardon et al. (2002), the present study implemented a two-level model with time-varying variables at Level-1 and person-level variables at Level 2. The advantage of implementing survival analysis in HLM is that it can handle nested data. In the present sample years are nested within students, which are nested within schools. The school level (level-3) is not included here to simplify the analyses and to maximize the comparability among all five models. Prior to performing a multi-level survival analyses, a person-period data set was constructed and data arranged accordingly, as was discussed in details in Chapter Three.

The three hazard models differ first on whether the analyses started following children from the time when they had their first opportunity to use alcohol or from age four. There are two versions of survival analysis that observe children by age starting at age four, because the results depend upon whether Age of First Opportunity was included as a predictor.

Two-Level Discrete-Time Hazard Model 1. The first survival analysis estimates the number of years that children "survive" until they used alcohol starting with their age of first opportunity. This was estimated from the year-by-year hazard rate (probability) of initiating alcohol use. After estimating a baseline model of how the hazard rates changed during these years, person-level variables were added to determine whether they predict increased or decreased hazard rates of using alcohol. Person-level variables included gender, socioeconomic status (SES), family attention (FAM), externalizing behavior (EXT), and age of first opportunity of using alcohol (AFO). As was already mentioned in a previous chapter, right censoring is used in this analysis. For every participant, the last possible year in the data is the year prior to the interview year following Reardon et al. (2002), because the interview year was only a partial year. In addition, right censoring eliminated all years after their interview year. All years after the first use of alcohol were also dropped from the analyses. Out of 1774 students, 1248 (70.3%) reported alcohol initiation prior to their interview year, 174 (9.8%) reported alcohol initiation during their interview year, and the rest of them did not report alcohol initiation. We found that for 19 (1.1%) students alcohol initiation supposedly occurred after the estimated age at which they were interviewed. This result is impossible, but is due to the approximate estimation of interview age.

We began by estimating a baseline two-level model in the Hierarchical Linear Modeling software package that does not include person-level predictors and includes only a random effect on the intercept. Here YR_{ij} is a dummy indicator of year j, starting from the AFO of alcohol use for person i, which defines the entire hazard curve of alcohol initiation. The baseline discrete-time hazard model is represented by the following equation at Level 1:

$$\eta_{ij} = \ln (h_{ij} / (1- h_{ij})) = \beta_{0i} + [\beta_1 YR1 + \beta_2 YR2 + \ldots + \beta_8 YR8] = \beta_{0i} + \sum \beta_{ji} (YR_{ij}) \qquad (1)$$

where η_{ij} is the log odds of alcohol initiation, h_{ij} is the hazard rate of probability of alcohol initiation, YR_{ij} is a dummy indicator of year j from AFO for student i, the coefficients $\beta_1, \beta_2, \ldots \beta_8$ are the intercept parameters indicating the conditional log odds that students whose covariate values are all zero will initiate alcohol use in each j^{th} year, given that they have not initiated alcohol in prior years. The β_{0i} parameter is the Level-1 intercept.

Person-level variations in the log odds will be added at Level 2 in HLM. For the baseline model, the Level-2 equations are:

$$\beta_{0i} = u_{0i} \qquad (2)$$

$$\beta_{ji} = \gamma_{i0}, \text{ for all } j \text{ years.}$$

where u_{0i} is the random effect of Level-2 person level variables. Note that β_{0i} is predicted only by the random effect, not by the usual Level-2 intercept. This sets the overall mean intercept to 0 for the entire sample, which makes the β_{ji} coefficient for each dummy code an estimate of the log odds of alcohol initiation for that year, which together can be used to estimate the baseline hazard curve across those years. The γ_{j0} coefficients set each β_{ji} coefficient equal to a Level-2 intercept.

Variations in the β_{ji} coefficients in magnitude and direction over the years observed define the shape of the logit hazard function and estimate how the risk of alcohol initiation increases, decreases or remains stable over time. Table 4.3 depicts the estimated parameters of the baseline model. The results show that the log odds of alcohol initiation are the highest at the age of first opportunity to try alcohol and then it declines till the 6th year from the age of first opportunity. At the 6th year from AFO the log odds of alcohol initiation increases. The statistical tests in Table 4.3 test whether the log odds differ significantly from zero, which is equivalent to testing whether the odds differs from 1 or whether the hazard differs from 0.5. The multivariate hypothesis test option in HLM can be used to test whether the log odds vary between two years. For instance, differences between the log odds of the hazard of alcohol initiation at the 5^{th} and 6^{th} years were tested with one test vector using 1 and -1 for the coefficients being compared and zeroes elsewhere. The composite null hypothesis test for $\beta_5 = \beta_6$ was rejected (χ^2 [1] = 14.2, $p < 0.01$), which indicated that the log odds of the hazard of alcohol initiation significantly differ from each other for those two years.

Table 4.3.

Two-Level Discrete Time Hazard Model 1

	Model 1 (baseline)		Model 1	
Predictor	Log odds (SE)	Odds ratio	Log odds (SE)	Odds ratio
Year 1	0.18** (0.05)	1.19	0.02 (0.14)	1.02
Year 2	-0.44** (0.09)	0.64	-0.45** (0.16)	0.63
Year 3	-0.39** (0.12)	0.68	-0.20 (0.17)	0.82
Year 4	-0.59** (0.16)	0.55	-0.27 (0.28)	0.76
Year 5	-1.12** (0.24)	0.32	-0.69* (0.27)	0.50
Year 6	0.05 (0.25)	1.05	0.71* (0.29)	2.03
Year 7	-0.54 (0.41)	0.58	0.28 (0.44)	1.31
Year 8	-1.09 (0.65)	0.34	-0.24 (0.67)	0.78
AFO			0.23**(0.02)	1.27
FAM			-0.04 (0.04)	0.96
EXT			0.04 (0.05)	1.04
SES			0.12**(0.05)	1.12
Gender			0.11 (0.09)	1.12
-2LL	3780.88		3797	

SE - Standard Error; LL - Log Likelihood
**p* < 0.05. ***p* < 0.01.

A graphical picture of the timing of alcohol initiation is provided by the hazard function (Fig. 4.3). The shape of the hazard curve is determined by the β_j coefficients and was obtained by converting the log odds of alcohol initiation for each year to hazard probabilities for those years by computing $\varphi_{ij} = 1/ (1 + \exp\{-\eta_{ij}\})$, where η_{ij} is the log of the odds of alcohol initiation.

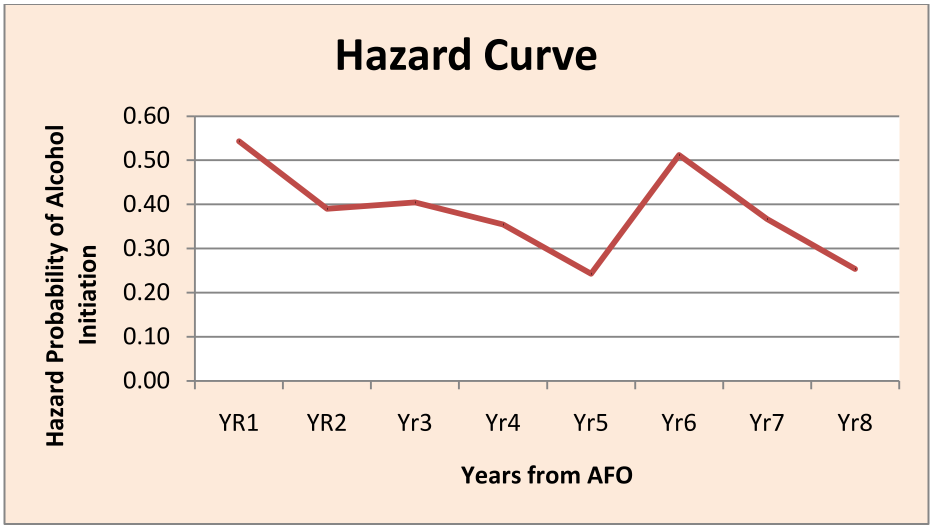


Fig. 4.3. Baseline Hazard Curve of Alcohol Initiation (Model 1)

This hazard curve depicts the overall shape of the hazard of alcohol initiation when there are no effects of covariates but nesting of occasions within persons is taken into consideration. It shows that more than 50% of the participants in the study were likely to initiate alcohol during the first year they had an opportunity to try alcohol. After the first year from AFO the hazard curve declines and then increases again at the 6^{th} year and then declines again.

The following figure depicts the survival probabilities until alcohol initiation for the baseline model (Fig. 4.4.). Each year shows the estimated proportion of the sample that has yet to initiate alcohol use during the observation period up to that year. Note that less than 10% of the sample remains in the risk set during the 6th year, which makes year-to-year changes in hazard rates less stable due to the smaller *n*.

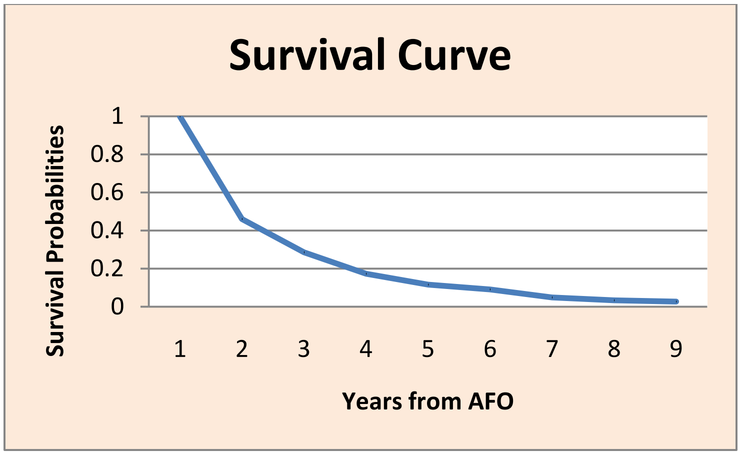


Fig. 4.4. Baseline Survival Curve of Alcohol initiation (Model 1).

The next interest in building this model was to investigate the effects of person-level covariates on the overall elevation of the hazard curve of the model. According to the proportional odds assumption, covariates can increase or decrease the hazard rate by the same proportion at all ages, while the overall shape of the hazard curve remains the same. The two-level discrete-time hazard model (Model 1) can be expressed in the following form, starting with the following equation at Level 1:

$$\eta_{ij} = \log (h_{ij} /(1- h_{ij})) = \beta_{0i} + \sum_{j=1}^{8} \beta_{ji} (YR_{ij}) \qquad (3)$$

The equations at Level 2 are:

$$\beta_{0i} = \sum_{n=1}^{5} \gamma_{0n} X_{in} + u_{0i} \qquad (4)$$

$$\beta_{ji} = \gamma_{i0}, \text{ for all } j \text{ years}$$

where X_{in} (n = 1, 2,…, 5) are five person-level covariates for person *i*, β_{ji} are intercept parameters, one per year, for eight years in the survival analysis, YR_{ij} is a dummy indicator of year *j*, and u_{0i} is a random effect of person-level variables. This model does not include a γ_{00} intercept on the Level-2 equation for β_{0i} because there is no omitted dummy variable. By omitting the γ_{00} intercept, the coefficient for the dummy indicator for each year is the log odds of alcohol initiation occurring in that year, conditional on all person-level variables equaling zero.

This model assumes that the effects of all Level-2 covariates on the hazard probability are the same for each time interval, i.e., that the log odds difference is constant over time. This assumption is referred as the proportional odds assumption (Singer & Willet, 2003). Each covariate can only predict changes in the intercept, not in distinct changes for specific years. This assumption can be tested for every covariate included in the model. The test is whether the effect of a covariate differs for different years in the data, i.e., and X_{in}*$Year_{ij}$ interaction. Following the example from Reardon et al. (2002), the level-two proportional odds assumption can be tested by first creating cross-product vectors by multiplying each covariate X_{in} by each dummy- coded year vector YR_{ij}. These interaction vectors X_{in}*YR_{ij} were added to the Level-1 model along with the main effect of X_n at Level 2 for one covariate at a time. As Reardon et al. (2002) indicate, the crucial test is that the coefficients for all of the cross-product terms are equal to each other, except for random variations. To test this assumption in the HLM multivariate test option, we created k-1 vectors (e.g., 8 - 1 = 7), each with -1 for the reference year, 1 for another year, and zeroes elsewhere (Reardon, personal communication, July 9, 2011). This is similar to the usual method for testing the main effect of a categorical factor in multiple regression with effect coding (Pedhazur, 1997). The chi-square test for this interaction was tested for every covariate and did not show any significant variations of the effects of a given covariate over the observed years from

AFO (χ^2 [8] $\leq$11.78, *p*s > 0.2). Thus the null hypotheses were not rejected, supporting the proportional odds assumption for each covariate.

Results for the final Two-level Discrete Time Hazard Model (Model 1) are given in the right-hand column of Table 4.3. The odds for the years now shows the hazard curve when all covariates are 0. Age of first opportunity was centered at 10 and the rest of the predictors were z-scored except for gender. Only two person-level variables, SES and AFO, had significant effects on the elevation of the overall hazard curve of alcohol initiation from the age of first opportunity. When the socioeconomic status of students increases by one standard deviation, or the AFO increases by one year, the log odds of alcohol initiation increases by 0.12 ($p < .01$) and 0.23 ($p < .01$), indicating 12% and 27% increases in the odds of alcohol initiation respectively.

Two-Level Discrete-Time Hazard Model 2. Whereas *Model 1* investigated the hazard rate of initiating alcohol use during years starting with the age of first opportunity, *Model 2* estimates the hazard rate for chronological ages, starting from age 4. *Model 2* includes those participants who did not report the age of first opportunity to use alcohol, who were dropped in the previous *Model 1* because there was no age of first opportunity for them to start the observation period. This is the main difference between *Model 1* and *Model 2* and all censoring issues that were discussed for *Model 1* apply for *Model 2*. The present model includes two versions: the first one, Model 2 (a), controls for age of first opportunity and the second one, Model 2 (b), does not control for the age of first opportunity.

The baseline discrete-time hazard model is represented by the following Level-1 equation:

$$\eta_{ij} = \log (h_{ij} / (1- h_{ij})) = \beta_{0\iota} + [\beta_4 AGE4 + \beta_5 YR5 + \ldots + \beta_{16} AGE16] = \beta_{0\iota} + \sum \beta_{ij} (AGE_{ij}) \quad (5)$$

where η_{ij} is the log odds of alcohol initiation, AGE_{ij} is a dummy indicator of year j from age 4 for student i, the coefficients $\beta_{4,}$ β_5, … β_{16} are the intercept parameters indicating the conditional log odds that students whose covariate values are all zero will initiate alcohol use in each j^{th} year, given that they have not initiated alcohol in prior years.

Person-level variations in the log odds will be added at Level 2 in HLM. For the baseline model, the Level-2 equations are:

$$\beta_{0i} = u_{0i} \quad (6)$$

$$\beta_{ij} = \gamma_{i0}, \text{ for all } j \text{ years.}$$

where u_{0i} is the random effect of Level-2 person level variables.

The first data column in Table 4.4 depicts estimated parameters of the baseline model. The results show that the log odds of alcohol initiation hazard are very low at early ages. The log odds of alcohol initiation are the highest when students reach the age of 15, from which it decreases at age 16. Differences between the log odds of alcohol initiation at age 15 and 16 were tested utilizing a multivariate hypothesis test for fixed effects. The composite null hypothesis test for $\beta_{15} = \beta_{16}$ was rejected (χ^2 [1] = 11.5, $p <$ 0.01), which indicated that the log odds of alcohol initiation at age 15 and 16 significantly differ from each other.

Table 4.4.

Two-Level Discrete Time Hazard Model 2

	Model 2 (baseline)		Model 2 (a)		Model 2 (b)	
Predictor	Log odds (SE)	Odds	Log odds (SE)	Odds ratio	Log odds (SE)	Odds ratio
Age 4	-6.35** (0.58)	0.001	-28.53**(7.23)	0.001	-6.17**(0.59)	0.002
Age 5	-5.65** (0.58)	0.004	-18.85**(4.20)	0.001	-5.47**(0.42)	0.004
Age 6	-5.49** (0.38)	0.004	-13.25**(2.32)	0.001	-5.31**(0.39)	0.005
Age 7	-4.86** (0.27)	0.008	-10.04**(1.17)	0.001	-4.68**(0.29)	0.009
Age 8	-3.94**(0.18)	0.019	-7.63**(0.59)	0.001	-3.75**(0.21)	0.023
Age 9	-3.55**(0.15)	0.029	-5.67**(0.37)	0.003	-3.35**(0.19)	0.035
Age 10	-2.37**(0.09)	0.093	-3.89**(0.24)	0.020	-2.16**(0.14)	0.116
Age 11	-2.39** (0.09)	0.091	-3.06**(0.21)	0.046	-2.15**(0.15)	0.116
Age 12	-1.39**(0.07)	0.250	-1.32**(0.18)	0.265	-1.11**(0.13)	0.330
Age 13	-1.04**(0.08)	0.356	-0.29 (0.18)	0.745	-0.72**(0.13)	0.486
Age 14	-0.59**(0.09)	0.554	0.63**(0.20)	1.886	-0.22 (0.14)	0.799
Age 15	0.09 (0.12)	1.091	1.82**(0.27)	6.196	0.52**(0.17)	1.691
Age 16	-0.81**(0.24)	0.444	1.33**(0.42)	3.798	-0.36 (0.27)	0.694
Fam. Attention			-0.07 (0.05)	0.929	-0.18**(0.04)	0.83
Ext. Behavior			0.04 (0.05)	1.044	0.15**(0.04)	1.15
SES			0.17**(0.05)	1.195	0.24**(0.04)	1.28
Gender			0.07 (0.10)	1.078	-0.17*(0.07)	0.84
AFO			- 0.19 (0.24)	0.83		
AFO* Age 4			-3.34**(1.00)	0.035		
AFO* Age 5			-2.30**(0.66)	0.100		
AFO* Age 6			-1.59**(0.44)	0.203		
AFO* Age 7			-1.25**(0.32)	0.284		
AFO* Age 8			-1.08**(0.27)	0.337		
AFO* Age 9			-0.82**(0.26)	0.440		
AFO* Age 10			-0.84**(0.25)	0.430		
AFO* Age 11			-0.71**(0.24)	0.491		
AFO* Age 12			-0.66**(0.25)	0.517		
AFO* Age 13			-0.56*(0.25)	0.572		
AFO* Age 14			-0.41 (0.25)	0.666		
AFO* Age 15			-0.34 (0.25)	0.715		
AFO* Age 16			-0.29 (0.26)	0.746		
-LL	22494.7		17118.6		22166.1	

Note. SE - Standard Error; LL - Log Likelihood.

*p < 0.05. **p < 0.01.

The shape of a baseline Hazard Curve is presented in Fig.4.5.

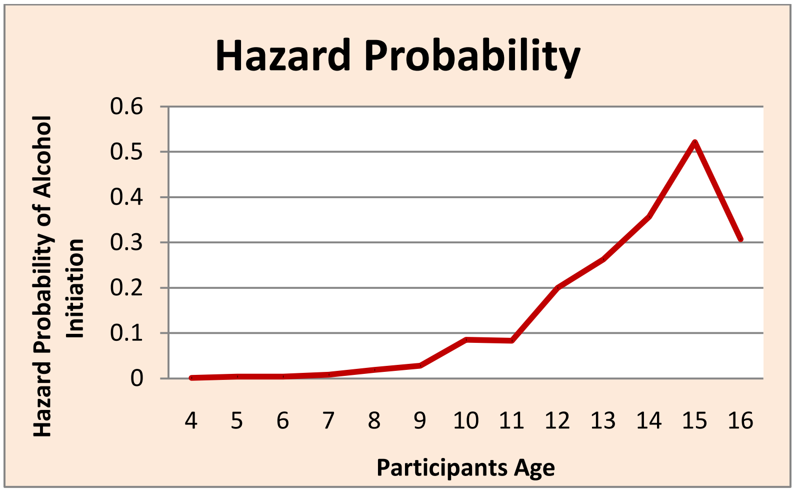


Fig. 4.5. Baseline Hazard Curve of Alcohol Initiation (Model 2)

The hazard probability graph shows the likelihood of alcohol initiation when there are no effects of any covariates on the hazard of alcohol initiation (Fig. 4.5). Note that the near-zero hazard rates at young ages correspond to a flat survival curve, whereas the high hazard rate at age 15 corresponds to a steep decline in the survival curve (Fig. 4.6). There is a minimal risk of alcohol initiation until students become 10 years old, the age when the risk started increasing. A multivariate hypothesis test for fixed effects did not indicate a significant difference between participants at age 10 and 11 in their hazard of alcohol initiation (χ^2 [1] = 0.02, p >0.5). Figure 4.5 shows that 52.2% of those who never used alcohol prior to age 15 do so during that year.

The following figure shows the survival probability curve until alcohol initiation (Fig. 4.6).

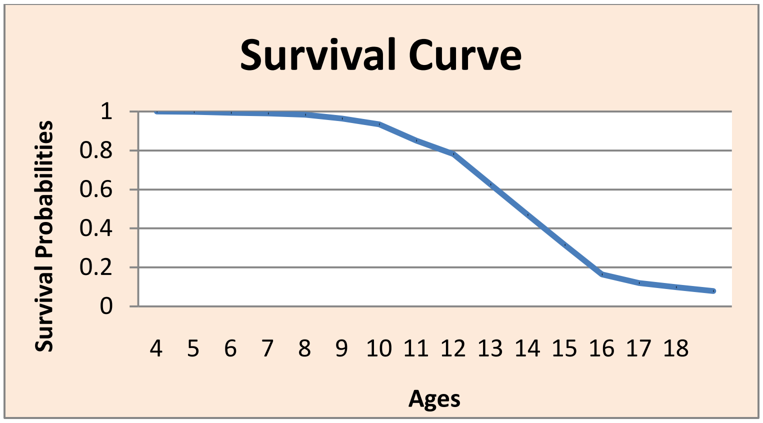


Fig. 4.6. Baseline Survival Curve of Alcohol initiation (Model 2).

Model 2 (a). The next step in building this model is to explore the fixed effects of covariates on the hazard curve of the model, i.e., how its overall elevation varies in relation to gender, age of first opportunity, externalizing behavior, family attention and socioeconomic status. For this model, missing values for age of first opportunity were recoded as one year older than at the age of interview to ensure keeping the most participants in the analysis (as discussed in Chapter 2). Note that the same estimate to replace missing values for AFO would not help in Model 1, because they would still have never entered the observation period before the interview age. This two-level discrete-time hazard model (Model 2(a)) can be expressed in the form of the following equation:

$$\eta_{ij} = \log (h_{ij} / (1- h_{ij})) = \beta_{0i} + \sum \beta_{ji} (AGE_{ij}),\ \ j = 4, 5, \ldots, 16 \qquad (7)$$

The equations at Level 2 are:

$$\beta_{0i} = \sum_{n=1}^{5} \gamma_{0n} X_{in} + u_{0i} \qquad (8)$$

$$\beta_{ij} = \gamma_{i0}, \text{ for all } j \text{ years}$$

where X_{in} (n = 1, 2,…, 5) are five person-level covariates for person i, γ_{j0} are intercept parameters for each year from age 4, AGE_{ij} is a dummy indicator of each year from age 4 to age 16, and u_{0i} is a random effect of person-level variables. This model does not include a γ_{00} intercept on Level-2 equation for $\beta_{\tilde{0}i}$, because there is no omitted dummy variable for any year. By omitting the γ_{00} intercept, the coefficient for the dummy indicator for each year is the log odds of alcohol initiation occurring in that year, conditional on all person-level variables equaling zero.

This model was also tested for the proportional odds assumption for Level-2 covariates. Chi-square tests for all covariates except for AFO ($\chi^2[12] = 87.636$, $p < .01$) did not show any significant differences between the effects of a given covariate across the years of observation (all other χ^2 $[13] \leq 16.80)$, ps > 0.20) and thus we failed to reject the null hypotheses for the interactions of all covariates with age except for AFO. We may conclude that the effect of AFO on the log odds of alcohol initiation was not the same for every year starting from age 4. For the final estimation of the model, all interaction vectors of AFO*AGE were entered at Level-1 to model the interaction of AFO * Age. The results for the final Model 2(a) are presented in the middle data columns of Table 4.4. Main effects for ages from age 4 to age 16 show hazard rates of alcohol initiation for males (due to coding them as zeros), when participants are at the grand-mean of AFO, Family Attention (FAM), Socioeconomic status (SES), and Externalizing behavior (EXT). The coefficients for the interaction vectors ($AFO*AGE_{ij}$) portray how the effect of AFO varies across ages. In these results, delaying a child's AFO by one year decreases the odds of alcohol initiation by 96.5% at age 4, but only by 25.4% at age 16. Only one person-level variable, SES, had a significant effect on the overall hazard of alcohol initiation. Increasing SES by one standard deviation predicts elevation of the log odds of alcohol initiation hazard by 0.17, indicating a 19% increase of the odds of alcohol initiation for any given year.

Because this model did not satisfy the proportional odds assumption and yielded extreme results for some combinations of age and age of first opportunity, the next model simplifies the survival analysis by dropping Age of First Opportunity as a predictor.

Model 2 (b). This model is the same as previous *Model 2(a)* except that it does not include age of first opportunity as a covariate at Level 2. When the effect of AFO on the hazard of alcohol initiation was dropped, all other person-level covariates in the model became significant (last columns of Table 4.4). If students are from higher socioeconomic level and show more externalizing behavior the logs odd of alcohol initiation hazard increases significantly at each age. More precisely, every one standard deviation increase in SES and externalizing behavior increases the log odds of the hazard by $\gamma = 0.24$ ($p < .01$) and $\gamma = 0.15$ ($p < .01$), indicating 28% and 15% increases in the odds of alcohol initiation respectively. If participants experience more family attention, e.g., an increase of one standard deviation, the log odds of alcohol initiation hazard decreases ($\gamma = -0.18$, $p < .01$), indicating about a 17% decrease in the odds of alcohol initiation. Gender also had a significant effect on the log odds of alcohol initiation, which are less for girls than for boys by $\gamma = -0.17$ ($p < .05$), meaning that the odds of alcohol initiation for females are 16% less than for males.

Multiple Regression Model. A less adequate analysis would be to use multiple regression analysis to predict the age of first use of alcohol (AFU) as a continuous dependent variable. Participants who reported not using alcohol were necessarily dropped from this analysis because they had no age of first use. The purpose of using this model is to estimate the effects of the same covariates used in the previous hazard models to predict age of first alcohol use, i.e., Gender, Family Attention (FAM), Socioeconomic Status (SES), Age of First Opportunity (AFO). This model is represented by the following equation:

$$AFU = b_0 + b_1 \text{ Gender} + b_2\text{EXT} + b_3\text{FAM} + b_4\text{AFO} + b_5\text{SES} \qquad (9)$$

where b_0 indicates the intercept of the equation, i.e., the predicted age of alcohol initiation when all other predictors are zero including AFO at its centered value of 10. Each of $b_{1,}$ $b_{2,}$ $b_{3,}$ b_4 and b_5 coefficients indicated the predicted effect of that variable, controlling for all other predictors in the equation.

Before implementing the regression analysis, a correlation analysis was conducted to check how the predictors are associated with each other and with the outcome. Results of the correlation analysis are presented in Table 4.5.

Table 4.5.
Intercorrelations of All Variables

Variables	1	2	3	4	5	6
1. AFU	1					
2. Gender	0.10**	1				
3. EXT	- 0.12**	- 0.19**	1			
4. SES	- 0.19*	- 0.11**	0.10**	1		
5. FAM	0.10**	- 0.10**	- 0.30**	0.04	1	
6. AFO	0.82**	0.14**	- 0.15**	- 0.18**	0.13**	1

Note. AFU – age of first use of alcohol; EXT- externalizing behavior; SES – socioeconomic status; FAM – family attention; AFO – age of first opportunity of using alcohol; $N = 1439$; *$p < .05$, **$p < .01$

The correlation results showed that every predictor is significantly associated with the age of first alcohol use. The largest correlation is from age of first opportunity to try alcohol (r = 0.82, p < .01). The age of first opportunity is more strongly associated with the other predictors in the model than is age of first use, except for essentially equal correlations with SES.

The results of the multiple regression analysis show that all predictors collectively account for a statistically significant proportion of the variance in predicting the age of first use of alcohol (F (5, 1390) = 562.75, p < .01). The model summary indicates that R^2 = 0.669 meaning that 66.9% of the variance in age of first use of alcohol can be explained by all the predictors included in the analysis. The standardized β coefficients show significance for AFO (β = 0.81, p < .01) and for SES (β = - 0.04, p = .01), which means that every standard deviation increase in AFO, will delay the age of first use of alcohol by 0.8 of a standard deviation when controlling for all other predictors, and every one standard deviation increase of socioeconomic status of participants will result in a decrease in the predicted age of first use of alcohol by 0.04 of a standard deviation after controlling for all other predictors in the model.

Next we ran the same multiple regression model without controlling for age of first opportunity to compare this model to the survival analysis models. The equation for the model is represented by the equation:

$$\text{AFU} = b_0 + b_1\,\text{Gender} + b_2\text{EXT} + b_3\text{FAM} + b_4\text{SES} \qquad (10)$$

The results of the multiple regression analysis show overall significance of predictors (F (4, 1391) = 20.37, p < .01) indicating a significant collective effect of the independent variables in predicting the age of first use of alcohol. The model summary indicated that only 5.5% of the variance in age of first use of alcohol can be explained by the combination of gender, participants' externalizing behavior, family attention, and socioeconomic status (i.e., $R^2 = 0.055$). All predictors have a statistically significant effect on predicting the age of first use of alcohol except for externalizing behavior ($p = 0.054$). More precisely, the standardized β coefficients indicate (when controlling for all other predictors in the model) that increasing family attention by one standard deviation will delay the age of first use of alcohol by 0.10 standard deviation; a one standard deviation increase in externalizing behavior and in socioeconomic status were associated with decreasing the first use of alcohol by 0.05 ($\beta = -0.054$, $p = .054$) and .17 ($\beta = -0.17$, $p = .01$) of a standard deviation of the age of first use of alcohol respectively. The age of first use of alcohol for girls is 0.074 of standard deviation older than the age of first use for boys ($\beta = 0.074$, $p < .01$).

Logistic Regression Model. The last model that is considered in the present study predicts alcohol use by participants using a logistic regression analysis. Logistic regression analysis is used to analyze a dichotomous outcome. Logistic regression accomplishes this by predicting the logit transformation of the dichotomous outcome variable. Basically the logistic regression predicts the logit of the outcome from a set of multiple predictors while controlling for all other predictors in the model, similar to multiple regression. As was discussed earlier, the logit is the natural logarithm of odds of the outcome, i.e., P/ (1-P), where P is the probability of the outcome happening. In this model alcohol use is the outcome variable, which is obtained from Age of First Use of alcohol (AFU) by dichotomizing participant responses as 0 when they did not report an age of first use of alcohol and as 1 when they indicated an age of first alcohol use.

Among 1776 participants of the study, 333 (18.8%) reported never using alcohol and 1443 (81.3%) reported using alcohol. We investigated if the likelihood of alcohol use is related to the same predictors that were used in previous models, i.e., Gender, Family Attention, Externalizing Behavior, and Age of First Opportunity to try alcohol (AFO). The equation for the relationship between the alcohol use (AlcUse) and predictors is represented by the following equation:

$$\text{Logit (AlcUse)} = b_0 + b_1 \text{ Gender} + b_2\text{EXT} + b_3\text{FAM} + b_4\text{AFO} + b_5\text{SES} \quad (11)$$

where b_0 is the intercept coefficient, i.e., the log odds of alcohol initiation when all other predictors are zero; and b_1, b_2, b_3, b_4 and b_5 indicate the expected change in the log odds of alcohol initiation for a one unit increase in the corresponding predictor when controlling for all other predictors in the equation.

The logistic regression analysis was carried out by the Logistic Regression procedure in SPSS v.18. By default the logistic regression analysis output first provides an estimate of an intercept-only model, which is also called the null or baseline model. It includes no predictors. An improvement over this baseline model is tested by examining two inferential omnibus statistical tests: the chi-squared and Score tests. Both tests produced the same conclusions for the present data (Table 4.6), namely that the logistic regression model with all predictors provides better estimates of who was most likely to use alcohol than the null model. For example, the score test indicates that the predictors as a group significantly improved the model.

Table 4.6

Logistic Regression Analysis for estimating likelihood of Alcohol Use

Predictor	b	*SE* b	Wald's χ^2	*df*	*p*	e^b (odds ratio)
Age of First Opportunity	- 0.01	0.04	0.02	1	0.878	0.99
Externalizing Behavior	0.26*	0.11	5.14	1	0.023	1.30
Socioeconomic Status	0.13	0.10	1.57	1	0.210	1.13
Family Attention	- 0.22$^+$	0.12	3.43	1	0.064	0.81
Female Gender	- 0.08	0.20	0.16	1	0.690	0.92
Constant	2.58**	0.16	253.23	1	0.001	13.16
Test			χ^2	*df*	*p*	
Overall model evaluation						
Score test			14.45	5	0.013	
Chi-square			15.65	5	0.008	

Note: *N* = 1776
*p< .05, **p< .01, $^+$p = 0.06

The statistical significance of each individual regression coefficient was tested using a Wald chi-square statistic (Table 4.6). According to Table 4.6 externalizing behavior is a significant predictor of alcohol use for participants from the study. The model suggests that youth who are higher on externalizing behavior are more likely to use alcohol (*p* = 0.02) In other words, an increase of one standard deviation in externalizing behavior increases the log odds of being in the alcohol use group by 0.26, which indicates a 30% ($e^{0.26}$ = 1.30) increase in the odds of alcohol use, holding all other variables constant. Family attention has only marginal significance in predicting having used alcohol. Namely, a one standard deviation increase in family attention decreases the odds of being in the alcohol user group by 19% ($e^{-0.22}$ = 0.81). No other variables have a significant effect on the odds of being in alcohol users' group in this model.

Next I ran the same logistic regression without including the age of first opportunity in the equation (to follow the same procedures I used before). The equation for the model is represented by the equation:

$$\text{Logit (AlcUse)} = b_0 + b_1 \text{ Gender} + b_2\text{EXT} + b_3\text{FAM} + b_4\text{SES} \quad (8)$$

The results of the logistic regression showed overall significance of the model above the baseline model (Score test = 69.23, $df = 4$, $p < 0.01$). The statistical significance of individual predictors indicated that only two predictors, Externalizing Behavior and Family Attention significantly predicted the likelihood of alcohol initiation. More precisely, a one standard deviation increase in externalizing behavior increased the log odds initiation alcohol use by 0.43, which corresponds to increasing the odds of alcohol initiation by about 54% ($e^{0.43} = 1.537$). A one-unit increase in family attention decreased the log odds of alcohol initiation by 0.36, which indicated decreasing the odds of initiating alcohol by about 30% ($e^{-0.36} = 0.697$). Effects of gender and socioeconomic status were not found to be significant in predicting odds of alcohol use.

Comparing Models

A summary of the main results obtained from the five models are presented in the following table (Table 4.6).

Table 4.7

Main Results of all Models

	Age of First Opportunity (AFO)	Family Attention	Externalizing Behavior	Socioeconomic Status	Gender
Discrete Time Hazard Model 1	γ = 0.23 **; OR = 1.27	γ = - 0.04 ; OR = 0.96	γ = 0.04 ; OR = 1.04	γ = 0.12 **; OR = 1.12	γ = 0.11; OR = 1.12
Discrete Time Hazard Model 2					
Model 2 (a) Control for AFO	γ = - 0.19$^{N/A}$; OR = 0.83$^{N/A}$	γ = - 0.07; OR = 0.93	γ = 0.04; OR = 1.04	γ = 0.17**; OR = 1.19	γ = 0.07; OR = 1.08
Model 2 (b) No control for AFO		γ = - 0.18**; OR = 0.83	γ = 0.12**; OR = 1.15	γ = 0.24**; OR = 1.28	γ = - 0.17*; OR = 0.84
Multiple Regression Model					
Control for AFO	β = 0.81**	β = - 0.01	β = 0.01	β = - 0.04*	β = - 0.02
No control for AFO		β = 0.10**	β = - 0.05*	β = - 0.17**	β = 0.07**
Logistic Regression Model					
Control for AFO	b = - 0.01; OR = 0.99	b = - 0.22^{+}; OR = 0.81	b = 0.26*; OR = 1.30	b= 0.13; OR = 1.13	b = - 0.08; OR = 0.92
No control for AFO		b = - 0.36**; OR = 0.70	b = 0.43**; OR = 1.54	b = 0.07; OR = 1.08	b = 0.03; OR = 1.03

Note: γ -effect on hazard rate of alcohol initiation; β - standardized multiple regression coefficient; b - log odds of alcohol use; $^{N/A}$ – not applicable; *p< .05, **p< .01, $^{+}$p = 0.06

Several points need to be considered regarding the estimated coefficients. Coefficients for discrete-time hazard models estimate the effects of covariates on the hazard rate of alcohol initiation equally over the years during which participants were observed. In the second hazard model (Model 2 (a)) the age of first opportunity of using alcohol did not have similar effects on years during the observation period and those estimated coefficients are not presented in the summary table. Due to a violation of proportional odds assumption for the age of first opportunity of using alcohol, the coefficient for estimating the hazard rate of alcohol use is not applicable, as depicted in the summary table. Standardized regression coefficients from multiple regression predict the impact of covariates on the age of alcohol initiation for those participants who initiate alcohol use whereas the logistic regression coefficients estimate the likelihood of being in either the alcohol user or non-user group.

Age of first opportunity seems to have the strongest predictive value for the first hazard model (Model 1) and for the multiple regression model. When the other predictors are significant, they naturally have opposite signs in multiple regression as in the corresponding hazard models. For example, higher Socioeconomic status (SES) predicts higher hazard rates of alcohol initiation (+γs and OR > 1.00), which results in a younger predicted age of alcohol initiation (-βs). SES appears to be a significant predictor in all models except for the logistic regression model. Externalizing behavior and family attention demonstrate similar patterns of statistically significant effects. In both cases, they significantly predict the timing of alcohol initiation only when age of first opportunity is not in the model. Similarly gender is a significant predictor of the age of first use only when the age of first opportunity is not incorporated in the models.

CHAPTER V

DISCUSSION

The main purpose of this study was to compare different statistical approaches in modeling alcohol initiation. We did this by modeling age of first use of alcohol and investigating how it was affected by variables that have been shown in the literature to have an impact on alcohol initiation. For developing these models we used a recently developed methodological approach, which is a combination of survival analysis with multilevel modeling (Reardon et al., 2002). Two discrete-time hazard models were developed within this methodological framework, the second of which has two versions (with and without the Age of First Opportunity (AFO) in the model). Two more models were developed to investigate alcohol initiation using more commonly used statistical methods: multiple and logistic regressions. Each of the regression models also includes two versions (with and without AFO in the model).

Several studies emphasized the importance of AFO in studying substance use and initiation (e.g., Van Etten & Anthony, 1999; Van Etten and Anthony, 2001; Caris et al., 2009), which was included in our models as one of the covariates and was found to be the most influential variable in the models leading to interesting results that are discussed below.

Next I explain how models differ from each other based on different ways of modeling the outcome, criteria for inclusion or exclusion of cases, and differences in results. Advantages of hazard models and implications for further research will be discussed along with some limitations of the present study.

Model Comparisons

Modeling outcome. The main differences among the models are based on how the age of first use is modeled as an outcome variable. The most relevant traditional analysis is multiple regression, which predicts age of first use as a continuous variable, but drops participants who have not used alcohol yet because their age of first use has not occurred yet and is thus unknown. Logistic regression predicts users vs. non-users, which puts non-users back in the analysis, but ignores distinctions about when alcohol was first used. Discrete-time hazard models predict the hazard of initiating alcohol use for every year of the observation period. Predicting the hazard rate of initiating use is similar to predicting the year of first use in multiple regression, but unlike multiple regression, hazard models retain non-users in the analysis for the years in which they delayed initiating alcohol use. In this study two hazard models were compared that differed on when the observation period began that deals with left-censoring issues. Survival analysis relies on the assumption that observation starts at the initial age of interest, which is the assumption of left-censoring at random, and that participants are observed until the alcohol initiation or until censored at random, which is the assumption of right-censoring (Allison, 1984; Yamaguchi, 1991). In the first hazard model (Model 1), when participants are observed from the time when they had an opportunity to try alcohol, year 1 is the first year in which it is possible for them to initiate alcohol use, which is their first year of opportunity. In the second, age-related hazard model (Model 2), year 1 is age 4, which was our estimate of the earliest age children could decide whether to initiate alcohol use or not.

Handling independent variables in models. In multiple regression, the independent variables are predicting the actual age of alcohol initiation. In logistic regression the independent variables are predicting the log odds of alcohol use vs. non-use. In survival analysis, the independent variables are predicting changes in the log odds of alcohol initiation. These changes are assumed to apply equally over the observation period, according to the usual proportional odds assumption, which was satisfied for all predictors except for AFO in Model 2 (a).

Distribution of age of first use. In the multiple regression model the mean age of first use of alcohol is predicted whereas in the logistic regression the actual age of alcohol use is ignored, because only alcohol use vs. no use is predicted. In hazard models baseline models show how the hazard of alcohol initiation varies during the observation period. The hazard curve reflects the hazard of alcohol initiation from the first age of opportunity in Model 1 and from age 4 in Model 2.

Exclusion criteria. The way of modeling the outcome determines who is excluded or retained in that particular model. The logistic regression model included everyone, but it cannot estimate the timing of the alcohol initiation whereas hazard models can. Participants who did not report age of first use (AFU) were excluded from the multiple regression model but retained in logistic regression model and in both hazard models if missing AFO data did not exclude them. Participants who did not report their AFO prior to their interview age were generally excluded from Model 1 and all models that controlled for AFO. That problem was overcome in Model 2 (a) by providing an estimate of a possible age of opportunity, which was the age at the time of interview plus 1 year.

Differences in Results

The results obtained in the present study varied across the models, depending especially on whether the age of first opportunity (AFO) was incorporated or not in that particular model. After summarizing the similarities and differences in the results from the four substantive predictors the section summarizes the more complicated case of the effect of AFO itself.

Substantive predictors. It should be noted that in the multiple regression and hazard models, coefficients of predictors have opposite signs (except for AFO), but they explain the variation in alcohol initiation conceptually in the same direction. For example, a higher hazard rate in a hazard model is equivalent to a younger age of alcohol initiation.

Gender. Gender predicted early use of alcohol in both analyses where the age of first opportunity (AFO) was not included in the model. Gender predicted a later age of using alcohol for girls compared to boys, but this is accounted for by the gender difference on AFO. Most studies regarding alcohol use have found gender to be one of the most significant predictors of alcohol initiation and use (Griffin et al., 2000l; Wagner et al., 2005; Dormitzer et al., 2004). The reason that gender was not found to have a significant impact in other models that included AFO might be because gender is more strongly associated with AFO than with age of first use (Fig. 5.1).

.

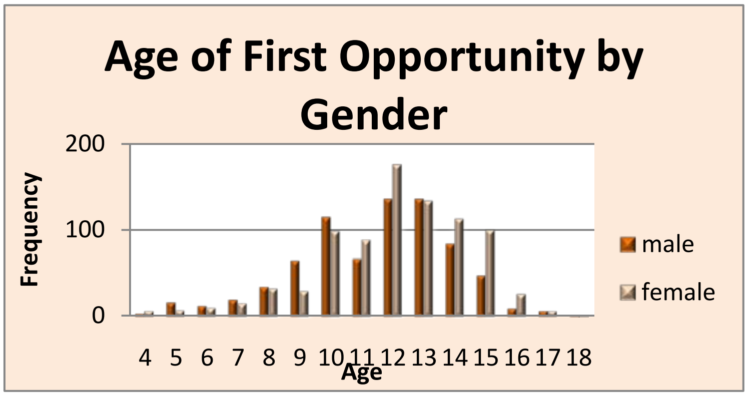


Fig. 5.1. Age of First Opportunity of Using Alcohol by Gender

At early ages boys seem to have more opportunity to try alcohol until they reach the age of ten after which girls started having more opportunity to use alcohol. After controlling for age of first opportunity, there is no gender difference in the hazard of initiating alcohol use in the first multiple regression model, in Hazard Model 1, or in Hazard Model 2(a). As was implied by findings from other studies, many factors that are associated with drug use may only be due to drug use opportunities. Namely, gender differences in drug use are a function of early opportunities to use drugs (Chen et al., 2005; Van Etten & Anthony, 1999).

Socioeconomic status. Socioeconomic status (SES) significantly predicts a higher hazard of early alcohol initiation in all models except for logistic regression, which predicts usage vs. non-usage. When predicting variation in the age of alcohol initiation, socioeconomic status seems to explain it to some extent, even after controlling for the age of first opportunity. Research that study the influence of socio economic status on alcohol use remain controversial. Some studies showed increase substance use for higher socioeconomic status (e.g., Humensky, 2010) and some studies have completely opposite results (e.g., Reinherz, et al., 2000). Cox (2007) found that SES had a significant inverse effect on age of use of alcohol when SES was incorporated as a contextual variable (i.e., at Level-2) and did not have a significant effect on age of first alcohol use when it was incorporated at Level-1. Thus the SES of the school but not the individual affected earlier versus later initiation of alcohol use. More research is needed to tease out the reasons behind the mixed results in the literature.

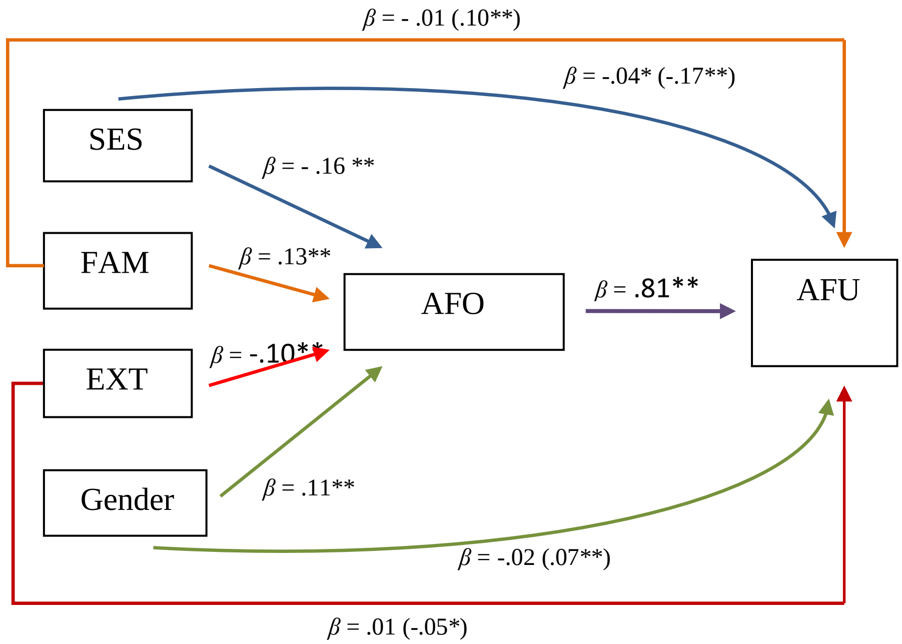
Externalizing behavior. Externalizing behavior is a significant predictor of age of alcohol initiation when the age of first opportunity to use alcohol is not incorporated in the model. When age of first opportunity is not controlled for, increased externalizing behavior predicts a younger age of alcohol initiation in multiple regression and higher hazard rate of alcohol initiation in Hazard Model 2 (b). These results are consistent with findings from the research literature regarding the positive relation between externalizing behavior and alcohol use that was reviewed in Chapter 2, but the results of the present study indicate that the effect of externalizing behavior on age of first alcohol use is entirely due to its effect on the opportunity for using alcohol. Externalizing behavior also has a significant impact on whether or not study participants are in the alcohol user or non-user group regardless of the age of first opportunity before their interview age. Youth with higher externalizing behavior most likely will initiate alcohol before the time of their interview. In all other models externalizing behavior does not significantly predict alcohol initiation beyond what is predicted by the age of first opportunity.

Family Attention. Family attention has the same impact on alcohol initiation and use as externalizing behavior, only in the opposite direction. More family attention delays the age of first use of alcohol and predicts a lower hazard rate of alcohol initiation when age of first opportunity is not incorporated in multiple regression and hazard models.

Similar to externalizing behavior, family attention does not predict variations in alcohol initiation beyond what is predicted by the age of first opportunity to use alcohol. Increases in family attention predicts significantly less chance to end up in the alcohol user group regardless of the age of first opportunity to use alcohol according to logistic regression analysis.

When models do not control for AFO all other independent variables of this study become significant predictors of alcohol initiation in all models except for the logistic regression model where controlling for AFO did not make statistically significant differences in predicting alcohol use. These findings support a meditational model, which accounts for the overall pattern of findings. As discussed in Chapter 2, others have also found that many of the factors associated with drug use may only be related to drug use to the extent that they predict exposure to drug use opportunities (Chen, et al., 2005; Van Etten & Anthony, 1999). In the present study we found that AFO apparently fully mediates the effect of family attention, externalizing behavior and gender in predicting early alcohol use. This conclusion is true under the assumption that family attention and externalizing behavior do not change over time. As was discussed earlier, these variables were found to be very resistant to change over time (Loeber, 1982; Murphy, et al., 2010). Mediation effect in the Multiple Regression Model is depicted in the following figure (Fig. 5.2). Coefficients in parenthesis denote standardized regression coefficients when AFO is not incorporated in the Model.

Mediation Model



*p< .05, **p< .01

Fig. 5.2. Mediation in the Multiple Regression Model

Effects of AFO. The apparent effect of AFO on the alcohol-use outcomes varies much more across the models for more complicated reasons. The simplest case is multiple regression, in which AFO is strongly correlated with AFU, accounting for over 66% of the variance in AFU by itself. That leaves only 34% of the variance to be uniquely predicted by the other predictors. Only SES accounted for variance beyond what was predicted by AFO. In logistic regression, however, AFO is missing for most of the non-user group, decreasing the size of that group from 18.8% to 8.9% of the sample. If the other 9.9% are omitted from the logistic regression model, AFO is not associated with being in the alcohol-user vs. non-user group. When we substituted an older age for the missing AFO, however, (as was done in Hazard Model 2), then a significant negative relation between AFO and alcohol use was observed, i.e., a younger age of opportunity was associated with alcohol use. More precisely, a one year increase in AFO decreases the odds of alcohol use by about 34% ($e^{-0.42} = 0.66$).

The age of first opportunity (AFO) is represented in Hazard Model 1 in two ways. First, it determines the initial year of the observation period. For example, Year 1 occurs at age 6 if AFO = 6, whereas Year 1 occurs at age 16 when AFO = 16 (see Table 5.1). Second, the positive coefficient for AFO indicates that the hazard rate of alcohol initiation in the first year of opportunity to use alcohol is higher if AFO is older. The association of AFO on the hazard rate of initiating alcohol in Model 1 can be explained for three different groups with early, middle, and older AFO (AFO of 4-8; 9 – 12; or 13 – 18 years old) in Table 5.1, created from a cross-tabulation. In every column, the hazard rate of initiating alcohol use is higher for those with older AFOs in the last row than those with younger AFOs in the first row. In the first column, for example, 16-year-olds have a 2/3 chance of initiating alcohol use if that is their first opportunity, whereas less than 1/3 of 6-year-olds start using alcohol if that is their first opportunity. This explains why AFO has a positive coefficient in predicting higher hazard rates when the observation period begins with their first year of opportunity. On the other hand, by comparing the same ages (e.g., the boldfaced proportions), a larger proportion of those with an earlier age of first opportunity have started using alcohol by any age selected. By comparing the cumulative use of alcohol for the same ages, the data replicate the usual finding in the literature that early opportunities to drink alcohol are associated with a higher cumulative use of alcohol at any given age.

Table 5.1

Cumulative Alcohol Use by Year and Age for Early, Middle, and Late Opportunities for Alcohol

	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7
Pr(Alcohol Use if AFO = 4-8)	0.31	0.43	0.56	0.64	0.70	0.82	**0.88**
Mean age	7	8	9	10	11	12	**13**
Pr(Alcohol Use if AFO= 9-12)	0.5	**0.69**	**0.82**	0.90	0.93	0.97	**0.98**
Mean age	11.5	**12.5**	**13.5**	14.5	15.5	16.5	**17.5**
Pr(Alcohol Use | AFO=13-18)	0.67	**0.86**	0.93	0.96			
Mean age	16.5	**17.5**	18.5	19.5			

Note. Pr = Probability of cumulative alcohol use up to that year or age.

The age-related Hazard Model (Model 2) adds another complication when AFO is a predictor. Although age determines entry into the observation period instead of AFO, including AFO as a predictor violates the proportional odds assumption. The effect of AFO is much larger at younger ages than at older ages (see Table 4.4). When modeling the age of first opportunity as an interaction term with years (the way to handle violations of the proportional odds assumption), alcohol initiation is impossible by definition for some combinations of age and AFO. That is, for any age prior to the age of first opportunity, the odds are zero and the log odds of zero are not defined (i.e., minus infinity). This may violate an implicit assumption of survival analysis that observed participants must be at risk of experiencing the event from the time they enter the observation period (Singer & Willett, 2003).

Advantages of Hazard Models

In summary, there are several advantages of hazard models for predicting the age of first alcohol use. Hazard models predict not only occurrence of alcohol initiation but also its timing. The baseline hazard curve gives the overall picture of the likelihood of alcohol initiation at every year over the observed period of time. Another advantage of hazard models is their ability to appropriately handle cases with unknown age of alcohol initiation. Those who have not initiated alcohol prior to the interview contribute exactly what is known about them, i.e., that they have not initiated alcohol yet. They were observed for a specified number of years prior to the interview (depending on when they entered the risk set), which is known as right-censoring. The ability to handle missing information is the most advantageous when studying rarely occurring events. In investigating alcohol initiation, only 18.6% reported no alcohol use, but when investigating illicit drugs, for example, many more participants do not report drug use prior to the interview. For modeling early ages of drug initiation or any other event occurrence, when a large percentage of participants have not yet experienced the event, hazard models may be the most appropriate statistical approach.

Contrasts from Multilevel Hazards Model by Reardon et al. (2002). Hazard models are even more advanced when it is possible to evaluate them in a multilevel framework. Often data on adolescent behaviors are collected from different schools, neighborhoods, counties, etc. These sampling techniques create a nested data structure, which violates the independence assumption of traditional regression analysis. Therefore, a methodological approach that combines survival analysis within a hierarchical linear modeling framework is needed to address both the missing data and the possible problems from violations of the assumption of independence of observations. The Reardon et al. (2002) study was one of the first to develop a methodological framework for employing the combination of survival analysis within a multilevel approach using HLM software from which the construction of hazard models of the present study were adapted. In the study conducted by Reardon et al. (2002), the time variable (represented by dummy coded observed years) and person-level variables are incorporated at the Level-1. In the present study I further developed Reardon and colleagues work by conceptualizing observed years as occasions nested within students. Occasions belong to Level-1, whereas person-level variables are incorporated at Level-2. I believe that this is an advantage of the present model because multilevel modeling incorporates different error terms for different levels of the data, which leads to more accurate Type-1 error rates (Raudenbush & Bryk, 2002). It is possible, however, that this is a trivial difference in this case of modeling binary outcomes, because they include no residual term at Level-1.In the present study participants were observed not only from age 4 (Model 2), similar to the example from Reardon et al. (2002), but also from age of their first opportunity to use alcohol (Model 1). After developing two-level discrete time hazard models, it is not difficult to take models to the next level by adding contextual variables on Level-3 of a multi-level model.

Limitations

Some limitations of the study need to be addressed. The first limitation is that the data were retrospective. The main limitation of retrospective data is their reliability. Participants might have memory limitations about when they first initiated or were exposed to alcohol. The second limitation of the study was the absence of the exact ages of participants. To overcome this problem approximate age was calculated based on age cohorts and school grades. After the age was estimated, calculations of percentage of the sample that initiated alcohol after the interview year showed that for 19 (1.1%) participants the age of alcohol initiation happened to be after their interview age (a discrepancy due to the imprecise age estimation). Third, in retrospective and cross-sectional data there is a lack of information about time-varying covariates. The covariates are measured only once, which does not create problems with stable characteristics (sex or ethnicity), but it becomes a problem in measuring any other changing factors. Lastly, more insights should be gained in studying time varying covariates and ways to incorporate them in analyses.

Further Research and Implications

In the present study two-level hazard models are developed that can easily be extended to the next level to investigate effects of contextual variables, e. g., school-level variables, on the timing of first substance use. In Cox's (2007) study, SES was not a significant within-school predictor of age of alcohol initiation, but school differences in mean SES did predict earlier alcohol initiation. The present study could not distinguish those types of effects from each other, because nesting of students within schools was not incorporated into this study. This illustrates the need to consider three-level models to differentiate within- and between-school effects on alcohol initiation.

Age of First Opportunity plays a significant role in alcohol initiation. Taking into account mediation models may provide important information for prevention efforts. Variables that predict AFU also seem to predict AFO as well. The same models developed for this study can be used in modeling AFO. Our results suggest that interventions targeted at delaying AFO could be very beneficial for prevention efforts.

Since approximately a third of the sample did not initiate use during the first year they had an opportunity to do so, Age of First Opportunity may also be a moderator for the effect of the predictors in this study on the age of first alcohol use. Age of first opportunity can be differentiated by a time lag between AFO and AFU. In other words, a rapid transition from AFO to AFU (e.g., if alcohol initiation occurs at the same year as AFO) can be considered on the one hand and a longer time lag between AFO and AFU (e.g., if alcohol initiation occurs after one (or more) years after AFO) on the other hand. This will provide insight into which predictors (family attention, externalizing behavior, socioeconomic status, and gender) have a significant effect on alcohol initiation dependent on the time duration from AFO to AFU.

Conclusion

In conclusion, all models considered in the present study have their own advantages. The main advantages of hazard models is in their ability to handle a particular kind of missing data called right censoring, such as youth who report delaying their initiation of alcohol use for all years covered in a given study. In investigating alcohol initiation, only about 18% reported no use of alcohol in this study, but when investigating illicit drugs, many more participants will be in a no-user group. For modeling early ages of drug initiation or any other event occurrence, when a vast majority of participants have not yet experienced it, hazard models should be considered. If the research interest is not in investigating when an event occurs or how the outcome varies by time, then multiple and logistic regressions might be more appropriate.

REFERENCES

Adalbjarnardottir , S., & Hafsteinsson, L. G. (2001). Adolescents' perceived parenting styles and their substance use: Concurrent and longitudinal Analyses *Journal of Research on Adolescence 11*, 401–423.

Adalbjarnardottir, S., & Rafnsson L. (2001). Perceived control in adolescent substance use: Concurrent and longitudinal analyses. *Psychology of Addictive Behaviors, 15,* 25-32. Allison, P. D. (1984). Event history analysis. Newbury Park, CA: Sage Publications.

Allison, P. D. (2002). Missing data. Thousand Oaks, CA: Sage Publications.

Andrews, J. A., Hops, H., Ary, D., Tildesley, E., & Harris, J. (1993). Parental influence on early adolescent substance use: Specific and nonspecific effects. *Journal of Early Adolescence, 13*, 285-310.

Anthony, J. C., Warner, L. A., & Kessler, R. C. (1994). Comparative epidemiology of dependence on tobacco, alcohol, controlled substances, and inhalants: Basic findings from the National Comorbidity Survey. *Experimental and Clinical Psychopharmacology, 2*(3), 244-268.

Barber, J. S., Murphy, S., Axinn, W. G., & Maples, J. (2000). Discrete-time multilevel hazard analysis. *Sociological Methodology, 30*, 201-235.

Bekman, N. M., Cummins, K., & Brown, S. A. (2010). Affective and personality risk and

.

cognitive mediators of initial adolescent alcohol use. *Journal of Studies on Alcohol and Drugs*, *71*(4), 570-580.

Bellis, M., Hughes, K., Morleo, M., Tocque, K., Hughes, S., Allen T., … Fe-Rodriguez, E, (2007). Predictors of Risky Alcohol Consumption in Schoolchildren and Their Implications for Preventing Alcohol-Related Harm. *Substance Abuse Treatment, Prevention and Policy*, 2-15.

Benjet, C., Borges, G., Medina-Mora, M. E., Blanco, J., Zambrano, J., Orozco, R., . . . Rojas, E. (2007). Drug use opportunities and the transition to drug use among adolescents from the Mexico City Metropolitan Area. *Drug and Alcohol Dependence, 90*(2-3), 128-134.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Bradshaw, C., Buckley, J., & Ialongo, N (2008). School-based service utilization among urban children with early onset educational and mental health problems: The squeaky wheel phenomenon. *School Psychology Quarterly, 23*(2), 169-186.

Browning, C. R., Leventhal, T., & Brooks-Gunn, J. (2005). Sexual initiation in early adolescence: The nexus of parental and community control. *American Sociological Review 70*, 758-778.

Brunswick, A. F., & Boyle, J. M. (1979). Patterns of drug involvement; Developmental and secular influences on age at initiation. *Youth & Society, 2*, 139-162.

Caris, L., Wagner, F. A., Rios-Bedoyae, C. F., & Anthony, J. C. (2009). Opportunities to use drugs and stages of drug involvement outside the United States: Evidence from the Republic of Chile. *Drug and Alcohol Dependence, 102*(1-3), 30-34.

Cozby, P. C. (2007). *Methods in Behavioral Research* (9th ed.). New York: McGraw-Hill.

Dooley, D. (2001). *Social Research Methods* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.

Dormitzer, C. M., Gonzalez, G. B., Penna, M., Bejarano, J., Obando, P., Sanchez, M., . . . Anthony, J. C. (2004). The PACARDO research project: youthful drug involvement in Central America and the Dominican Republic. *Revista Panamericana De Salud Publica-Pan American Journal of Public Health, 15*(6), 400-416.

Capaldi, D. M. & Dishion, T. J. (1988). *Psychometric Properties of Fourteen Latent Constructs from the Oregon Youth Study*. New York, Berlin, Heidelberg, London, Paris, Tokyo: Springer-Verlag.

Carlini-Marlatt, B., Gazal-Carvalho, C., Gouveria, N., & Souza, M. (2003). Drinking practices and other health-related behaviors among adolescents of Sao Paulo City, Brazil. *Substance Use and Misuse, 38,* 905-932.

Cengage, G. (2002). Gateway drug theory. In L. Breslow (Ed.), *Encyclopedia of Public Health*. Retrieved from: http://www.enotes.com/public-health-encyclopedia

Chen, C.Y., Storr, C. L., & Anthony, J. C. (2005). Influences of parenting practices on the risk of having a chance to try cannabis. *Pediatrics, 115*(6), 1631-1639.

Clark, D. B., Vanyukov M., & Comelius, J. (2002). Childhood antisocial behavior and adolescent alcohol use disorders. *NIAAA of the National Institute of Health, 26,* 109-115.

Cox, R. B. (2007). *Toward an eco-developmental theory of adolescent substance use in Venezuela* (Doctoral Dissertation). Available from Dissertations & Theses database. (UMI No. 3282080).

DeWit, D. J., Adlaf, E. M., Offord, D. R., & Ogborne, A. C. (2000). Age at first alcohol use: A risk factor for the development of alcohol disorders. *American Journal of Psychiatry, 157*, 745 - 750.

Farrington, D. P. (2003). Developmental and life-course criminology: Key theoretical and empirical issues - The 2002 Sutherland Award Address. *Criminology, 41*(2), 221-255.

Feldt, L. S., & Charter, R. A. (2003). Estimating the reliability of a test split into two parts of equal or unequal length. *Psychological Methods, 8*, 102-109.

Flory, K., Lynam, D., Milich, R., Leukefeld, C., & Clayton, R. (2004). Early adolescent through young adult alcohol and marijuana use trajectories: Early predictors, young adult outcomes, and predictive utility. *Developmental Psychopathology, 16*, 193-213.

Foster, S. E., Vaughan, R. D., Foster, W. H., & Califano J. A. Jr. (2003). Alcohol consumption and expenditures for underage drinking and adult excessive drinking. *Journal of the American Medical Association, 289*, 989-95.

Friedman, A. S., Terras, A., & Zhu, W. (2004). Early adolescent substance use/abuse as predictor to employment in adulthood: Gender differences. *Journal of Child & Adolescent Substance Abuse, 13*(4), 49-60.

Goodman, E., & Huang, B. (2002). Socioeconomic status, depressive symptoms and adolescent substance abuse. *Archives of Pediatric and Adolescent Medicine*, *156,* 448-453.

Grant, B. F., & Dawson, D. A. (1997). Age at onset of alcohol use and its association with DSM-IV alcohol abuse and dependence: Results from the National Longitudinal Alcohol Epidemiologic Survey. *Journal of Substance Abuse, 9*, 103-110.

Griffin, K. W., Botvin, G. J., Scheier, L. M., Diaz, T., & Miller, N. L. (2000). Parenting practices as predictors of substance use, delinquency, and aggression among urban minority youth: Moderating effects of family structure and gender. *Psychology of Addictive Behaviors*, *14* (2), 174-184.

Hamilton, H., Noah, S., & Adlaf, E. (2009). Perceived financial status, health and maladjustment in adolescence. *Social Science & Medicine*, *68*(8), 1527-1534.

Hawkins J. D., Catalano R. F., & Miller J. Y. (1992). Risk and protective factors for alcohol and other drug problems in adolescence and early adulthood. *American Psychological Association*, *112*, 64-105.

Hawkins, J. D., Brown, E. C., Oesterle, S., Arthur, M. W., Abbott, R. D., & Catalano, R. F. (2008). Early effects of Communities That Care on targeted risks and initiation of delinquent behavior and substance Use. *Journal of Adolescent Health, 43*(1), 15-22.

Herrera-Vazquez, M., Wagner, F. A., Velasco-Mondragon, E., Borges, G., & Lazcano-Ponce, E. (2004). Onset of alcohol and tobacco use and transition to other drug use among students from Morelos, Mexico. *Salud Publica De Mexico, 46*(2), 132-140.

Hingson, R.W., Heeren, T., Winter, M. R. (2006). Age at drinking onset and alcohol dependence: age at onset, duration, and severity. *Archives of Pediatrics & Adolescent Medicine, 160*(7), 739-746.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55.

Humensky, J. (2010). Are adolescents with high socioeconomic status more likely to engage in alcohol and illicit drug use in early adulthood? *Substance Abuse Treatment, Prevention, and Policy, 5*(1), 19.

Jackson, C., Henriksen, L., Dickinson, D., & Levine, D.W. (1997). The early use of alcohol and tobacco: Its relation to children's competence and parents' behavior. *American Journal of Public Health, 87,* 359–364.

Jessor, R. 1993. Successful adolescent development among youth in high-risk settings. *American Psychologist, 48*, 117-126.

Jones, D. R., Macias, C., Barreira, P. J., Fisher, W. H., Hargreaves, W. A., & Harding, C. M. (2004). Prevalence, severity, and co-occurrence of chronic physical health problems of persons with serious mental illness. *Psychiatric Services, 55*(11), 1250-1257.

Kandel, D. B. (1982). Epidemiological and psychosocial perspectives on adolescent drug use. *Journal of the American Academy of Clinical Psychiatry, 21*, 328–347.

Kaplow, J. B., Curran P. J., & Dodge K. A. (2002). Child, parent and peer predictors of early-onset substance use: A multisite longitudinal study. *Journal of Abnormal Child Psychology, 30*, 199-216.

Kim, J., & Gray, K. A. (2008). Leave or stay? Battered women's decision after intimate partner violence. *Journal of Interpersonal Violence, 23*, 1465-1482.

Kleinman, P. (1978), Onset of addiction; a first attempt at prediction. *International Journal of the Addictions*, *13*, 1217-1235.

Kline, R. B. (2005). Principles and practice of structural equation modeling. Second Edition. NY: Guilford.

Kuther, T. L. (2002). Rational decision perspectives on alcohol consumption by youth: Revising the theory of planned behavior. *Addictive Behaviors, 27*, 35-47.

Larsen, K., & Merlo J (2005). Appropriate assessment of neighborhood effects on individual health: Integrating random and fixed effects in multilevel logistic regression. *American Journal of Epidemiology 161*, 81–88.

Loeber, R. (1982). The stability of antisocial and delinquent child behavior: A review. *Child Development 53*, 1431-1446.

Li, X. M., & Feigelman, S. (1994). Recent and intended drug trafficking among male and female urban African-American early adolescents. *Pediatrics, 93*(6), 1044-1049.

MacPherson, L., Magidson, J. F., Reynolds, E. K., Kahler, C. W., & Lejuez, C. W. (2010). Changes in sensation seeking and risk-taking propensity predict increases in alcohol use among early adolescents. *Alcoholism-Clinical and Experimental Research, 34*(8), 1400-1408.

Malacane, L. A., Murphy, S. A., & Collins, L. M. (1997). *Determining the risk set for* *hazard functions when survival methods are applied to social science and prevention data.* University Park, PA: The Methodology Center, Pennsylvania State University.

Margulies, R, Kessler, R. C., & Kandel, D. (1977). A longitudinal study of onset of drinking among high school students. *Journal of Studies on Alcohol and Drugs, 38*, 897-912.

Martin, B. A., McCoy, T. P., Champion, H., Parries., M. T., DuRant, R. H., Mitra, A., & Rhodes, S. D (2009). The Role of Monthly Spending Money in College Student Drinking Behaviors and their Consequences. *Journal of American College Health*, *57*(6), 587-596.

McGue, M., & Iacono, W. G. (2005). The association of early adolescent problem behavior with adult psychopathology. *American Journal of Psychiatry, 162*(6), 1118-1124.

McKnight, P. E., McKnight, K. M., Sidani, S., & Figueredo, A. J. (2007). *Missing data.* New York, NY: The Guilford Press.

Menard, S. (2002). *Applied logistic regression analysis* (2nd ed.). Thousand Oaks, CA: Sage University Paper.

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

Mokdad, A. H., Marks, J. S., Stroup, D. F., & Gerberding, J. L. (2005). Actual Causes of Death in the United States, 2000. *Journal of the American Medical Association, 291*(10), 1238-1245.

Owings, L. (2008). Making youth violence prevention a national priority. *American Journal of Preventive Medicine*, *34*(3), S3-S4.

Patterson, G. R. (1997). Performance models for parenting: A social interactional perspective. In J. E. Grusec & L. Kuczynski (Eds.), *Parenting and children's internalization of values: A handbook of contemporary theory* (pp. 193-226). New York: Wiley.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Reinherz, H., Giaconia, R., Hauf, A., Wasserman, M., & Paradis, A. (2000). General and Specific Childhood Risk Factors for Depression and Drug Disorders by Early Adulthood. *Journal of the American Academy of Child and Adolescent Psychiatry, 39*(2), 223-231

Pedhazur, E. J. (1997). *Multiple regression in behavioral research* (3rd ed.). New York: Harcourt Brace College Publishers.

Reardon, S. F., Brennan, R. T., & Buka, S. L. (2002). Estimating multi-level discrete-time hazard models using cross-sectional data: Neighborhood effects on the onset of adolescent cigarette use. *Multivariate Behavioral Research, 37*(3), 297-330.

Reyes, A. (1999). Alcohol addiction – Alcohol dependence. *Medicine online*. Retrieved from: http://www.medicineonline.com/articles/a/2/Alcohol-Addiction/Alcohol-Dependence.html

Robins, L. N. (1977) Estimating addiction rates and locating target populations: how decomposition into stages helps. In Rittenhouse J. D. (ed). *The Epidemiology of*

Heroine and other Narcotics. Research monograph series 16. Washington DC, National Institute on Drug Abuse.

Robins, L. N., & Przybeck, T. R. (1985). Age of onset of drug use as a factor in drug and other disorders. In C. L. Jones & R. J. Battjes (Eds.), *Etiology of drug abuse: Implications for prevention* (pp. 178–192).Washington, DC: National Institute on Drug Abuse.

Sampson, R. J., & Laub, J. H. (2003). Life-course disasters? Trajectories of crime among delinquent boys followed to age 70. *Criminology*, *41*(3), 555-592.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Boston: Houghton-Mifflin.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis*. New York: Oxford University Press.

Snyder, H.N., & Sickmund, M. (1999). Juvenile Offenders and Victims: 1999 National Report.Washington, DC: National Center for Juvenile Justice, 1999.

Spoth R. L., Guyll, M., & Day, S. X. (2002). Universal family-focused interventions in alcohol-use disorder prevention: cost-effectiveness and cost-benefit analyses of two interventions. *Journal of Studies on Alcohol and Drugs, 63*(2), 219-28.

Steele F., Goldstein H. & Browne W. (2004). A general multilevel multistate competing risks model for event history data, with an application to a study of contraceptive use dynamics. *Statistical Modeling, 4,* 145-159.

Stice, E., & Barrera, Jr. (1995). A longitudinal examination of the reciprocal relations between perceived parenting and adolescents' substance use and externalizing behaviors. *Developmental Psychology, 31*(2), 322-334.

Streiner, D. L. (2003). Being inconsistent about consistency: When coefficient alpha does and doesn't matter. Journal of Personality Assessment, 80, 217-222.

Tarter, R., & Hegedus, A. (1991). The Drug Use Screening Inventory: Its application in the evaluation and treatment of alcohol and drug abuse. *Alcohol Health and Research World*, *15,* 65–75.

Tur, J. A., Puig, M. S., Pons, A., & Benito, E. (2003). Alcohol consumption among school adolescents in Palma De Mallorca. *Alcohol & Alcoholism, 38*(3), 243-248.

Van Etten, M. L., & Anthony, J. C. (2001). Male-female differences in transitions from first drug opportunity to first use: Searching for subgroup variation by age, race, region, and urban status. *Journal of Women's Health & Gender-Based Medicine, 10*(8), 797-804.

Van Etten, M. L., & Anthony, J. C. (1999). Comparative epidemiology of initial drug opportunities and transitions to first use: marijuana, cocaine, hallucinogens and heroin. *Drug and Alcohol Dependence, 54*(2), 117-125.

Van Etten, M. L., Neumark, Y. D., & Anthony, J. C. (1997). Initial opportunity to use marijuana and the transition to first use: United States, 1979-1994. *Drug and Alcohol Dependence*, *49*(1), 1-7.

Wagner, F. A., & Anthony, J. C. (2002). Into the World of Illegal Drug Use: Exposure Opportunity and Other Mechanisms Linking the Use of Alcohol, Tobacco, Marijuana, and Cocaine. *American Journal of Epidemiology, 155*(10), 918-925.

Wagner, F. A., Velasco-Mondragon, H. E., Herrera-Vazquez, M., Borges, G., & Lazcano-Ponce, E. (2005). Early alcohol or tobacco onset and transition to other drug use among students in the State of Morelos, Mexico. *Drug and Alcohol Dependence*, *77*(1), 93-96.

Weissberg, R.P., Walberg, H.J., O'Brien M.U., & Kuster, C.B. (2003). *Long-term trends in the well-being of children and youth.* Washington, DC: Child Welfare League of America Press.

Wells, S., Graham, K., Speechley, M., & Koval, J. (2005). Drinking patterns, drinking contexts and alcohol-related aggression among late adolescent and young adult drinkers. *Addiction, 100*(7), 933-944.

Wilcox, H. C., Wagner, F. A., & Anthony, J. C. (2002). Exposure opportunity as a mechanism linking youth marijuana use to hallucinogen use. *Drug and Alcohol Dependence*, *66*(2), 127-135.

Woolf, S. H. (2006). The big answer: rediscovering prevention at a time of crisis in health care. *Harvard Health Policy Review, 7*(2), 5-20.

World Health Organization. (2005). Global Status Report on Alcohol. Geneva, World Health Organization, Retrieved October 12, 2010 from World Health Organization Website: http://www.who.int/substance_abuse/publications/alcohol/en/

Wright, R. E. (1995). Logistic regression. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding multivariate statistics* (pp. 217-245). Washington, DC: American Psychological Association.

Yamaguchi, K. (1991). *Event history analysis*. London, United Kingdom: Sage Publications.

APPPENDICES

.

Appendix A

SYNTAX FOR CREATING PERSON-PERIOD DATA SET IN SPSS

1. Model 1 where beginning time is AFO

*Creating person-period data set for Level-1.

*The loop command adds cases for each of 18 age groups for the children in the study.

*Keep all variables that might be analyzed in the eventual multilevel modeling analyses.

*V104 (AFU) 0 needs to be assigned as missing data to keep those you never used alcohol (i.e. V104 = 0) in the sample.

compute age = 1.
format age (f8.0).
loop age = 1 to 18.
.xsave outfile =' FILE DIRECTORY: \FILE NAME *sav'*

end loop.
execute.

*Next, get the file that was created, which is now a person-period file.

get file = 'FILE DIRECTORY: \FILE NAME. *sav'*

*Creating a new variable indicating number of years (each raw) after AFO.
*Create dummy codes setting all dummy codes to 0.

compute yrdum0 = 0.
compute yrdum1 = 0.
compute yrdum2 = 0.
compute yrdum3 = 0.
compute yrdum4 = 0.
compute yrdum5 = 0.
compute yrdum6 = 0.
compute yrdum7 = 0.
compute yrdum8 = 0.
compute yrdum9 = 0.
compute yrdum10 = 0.

compute yrdum11 = 0.
compute yrdum12 = 0.
compute yrdum13 = 0.

compute yrdum14 = 0.
compute yrdum15 = 0.
compute yrdum16 = 0.
compute yrdum17 = 0.
compute yrdum18 = 0.
execute.

Select if (V103 GT 0).

compute yrsfmAFO = age - V103.
compute YFU = V104 - V103.
compute YRINT = ageintv - V103.
execute.

*Now finalize the dummy codes by changing the correct value to 1 to indicate the child's age in that row.

If (yrsfmAFO = 0) yrdum0 =1.
if (yrsfmAFO = 1) yrdum1 = 1.
if (yrsfmAFO = 2) yrdum2 = 1.
if (yrsfmAFO = 3) yrdum3 = 1.
if (yrsfmAFO = 4) yrdum4 =1.
if (yrsfmAFO = 5) yrdum5 =1.
if (yrsfmAFO = 6) yrdum6 = 1.
if (yrsfmAFO = 7) yrdum7 = 1.
if (yrsfmAFO = 8) yrdum8 = 1.
if (yrsfmAFO = 9) yrdum9 = 1.
if (yrsfmAFO = 10) yrdum10 =1.
if (yrsfmAFO = 11) yrdum11 =1.
if (yrsfmAFO = 12) yrdum12 =1.
if (yrsfmAFO = 13) yrdum13 = 1.
if (yrsfmAFO = 14) yrdum14 =1.
if (yrsfmAFO = 15) yrdum15 = 1.
if (yrsfmAFO = 16) yrdum16 = 1.
if (yrsfmAFO = 17) yrdum17 = 1.
if (yrsfmAFO = 18) yrdum18 = 1.
execute.

*Next create a drop variable, to indicate which ages will get dropped from the data.

compute drop=0.

*This is an example of right sencoring

*Right-censored ages need to be dropped for three reasons:

* (1) drop the partial age year when they were interviewed

* (2) drop all years after their interview year.

* (3) drop all years after the first year they used the substance.

*age1 variable is created to handle the fractional years in these data for AgeIntv. If the age of the interview is an integer representing the latest birthday, we do not need to do this, but use age and ageintv instead.

*compute age1 = age + 1.

if (yrsfmAFO lt 0) drop = 1.
if (yrsfmAFO GE YRINT) drop = 1.
if (yrsfmAFO GT YFU) drop = 1.

*Then permanently select only those cases (person-years) that were not dropped.

select if (drop = 0).

*Now a variable that indicates the year in which each child used the substance need to be created.

*This should always be the last year (row) in that child's data. However, those partial years are

right-censored, including cases in which the youth reported using the substance for the first time during the last partial year, the one during which they were interviewed.

compute Alcint = 0.
if age = V104 Alcint =1.
format Alcint (f8.0).
execute.

*Save data as FILE NAME.sav.

*Creating Level2 Data file.

```
SORT CASES BY ID1.
AGGREGATE
/OUTFILE='FILE DIRECTORY'
/PRESORTED
/BREAK=ID1
/V2_mean=MEAN(V2)
/Gender_mean=MEAN(Gender)
/V103_mean=MEAN(V103)
/AFOnew_mean=MEAN(AFOnew)
/SESZ_mean=MEAN(ZSESav)
/EXTBZ_mean=MEAN(ZEXTBav)
/FAMATENZ_mean=MEAN(ZFamAtenav)
/AFOc_mean=MEAN(AFOc)
/id2=mean(id2).
```

2. Model 2 where beginning time is Age 4.

*The loop command adds cases for each of 18 age groups for the children in the study.

*Here we start from age 4.

```
compute age=4.
format age (f8.0).
loop age=4 to 18.
.xsave outfile = 'FILE DIRECTORY: \FILE NAME.sav'

end loop.
EXECUTE.
```

*Next, we need to get the file we created, which is now a person-period file.

```
get file =  'FILE DIRECTORY: \FILE NAME. sav'
```

*create dummy codes for all ages starting from age 4, first by setting all dummy codes to 0.

```
compute agedum4=0.
compute agedum5=0.
compute agedum6=0.
compute agedum7=0.
compute agedum8=0.
compute agedum9=0.
compute agedum10=0.

compute agedum11 = 0.
compute agedum12 = 0.
compute agedum13=0.
```

compute agedum14=0.
compute agedum15=0.
compute agedum16=0.
compute agedum17=0.
compute agedum18=0.

*Finalizing the dummy codes by changing the correct value to 1 to indicate the child's age in that row.

if (age = 4) agedum4=1.
if (age = 5) agedum5 = 1.
if (age = 6) agedum6=1.
if (age = 7) agedum7 = 1.
if (age = 8) agedum8 = 1.
if (age = 9) agedum9 = 1.
if (age = 10)agedum10 = 1.

if (age = 11) agedum11=1.
if (age = 12) agedum12=1.
if (age = 13) agedum13=1.
if (age = 14) agedum14=1.
if (age =15) agedum15 = 1.
if (age = 16) agedum16=1.
if (age = 17) agedum17 = 1.
if (age = 18) agedum18 = 1.
execute.

**Next create a drop variable, to indicate which ages will get dropped from the data.*

compute drop=0.

*This is an example of right sencoring
*We drop right-censored ages for three reasons:

*   (1) drop the partial age year when they were interviewed
*   (2) drop all years after their interview year.
*   (3) drop all years after the first year they used the substance.

*age1 variable is created to handle the fractional years in these data for AgeIntv. If the

age of the

interview is an integer representing the latest birthday, we do not need to do this, but use

age and ageintv instead.

*compute age1 = age + 1.

if (age ge ageintv) drop=1.
if (age gt V104) drop=1.

*Then permanently select only those cases (person-years) that were not dropped.

select if (drop = 0).
*Now a variable that indicates the year in which each child used the substance need to be

created.

compute Alcint = 0.
if age = V104 Alcint =1.
format Alcint (f8.0).
execute.

*Creating Level2 Data file.

SORT CASES BY ID1.
AGGREGATE
/OUTFILE='FILE DIRECTORY: \FILE NAME. sav'
/PRESORTED
/BREAK=ID1
/V2_mean=MEAN(V2)
/Gender_mean=MEAN(Gender)
/V103_mean=MEAN(V103)
/AFOnew_mean=MEAN(AFOnew)
/SESZ_mean=MEAN(ZSESav)
/EXTBZ_mean=MEAN(ZEXTBav)
/FAMATENZ_mean=MEAN(ZFamAtenav)
/AFOc_mean=MEAN(AFOc)
/id2=mean(id2).

Note: All SPSS commands are italicized.

APPENDIX B
INSTITUTIONAL REVIEW BOARD APPROVAL

Oklahoma State University Institutional Review Board

Request for Determination of Non-Human Subject or Non-Research

4. Determination of "Research".
45 CFR 46.102(d): *Research* means a systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge. Activities which meet this definition constitute research for purposes of this policy whether or not they are conducted or supported under a program which is considered research for other purposes.

One of the following must be "no" to qualify as "non-research":

A. Will the data/specimen(s) be obtained in a systematic manner?
☒ No ☐ Yes

B. Will the intent of the data/specimen collection be for the purpose of contributing to generalizable knowledge (the results (or conclusions) of the activity are intended to be extended beyond a single individual or an internal program, e.g., publications or presentations)?
☐ No ☒ Yes

5. Determination of "Human Subject".
45 CFR 46.102(f): *Human subject* means a living individual about whom an investigator (whether professional or student) conducting research obtains: (1) data through intervention or interaction with the individual or (2) identifiable private information. Intervention includes both physical procedures by which data are gathered (for example venipuncture) and manipulations of the subject or the subject's environment that are performed for research purposes. Interaction includes communication or interpersonal contact between investigator and subject. Private information includes information about behavior that occurs in a context in which an individual can reasonably expect that no observation or recording is taking place, and information which has been provided for specific purposes by an individual and which the individual can reasonably expect will not be made public (for example, a medical record). Private information must be individually identifiable (i.e., the identity of the subject is or may be ascertained by the investigator or associated with the information) in order for obtaining the information to constitute research involving human subjects.

A. Does the research involve obtaining information about living individuals?
☒ No ☐ Yes
If no, then research does not involve human subjects, <u>no other information is required</u>.
If yes, proceed to the following questions.

All of the following must be "no" to qualify as "non-human subject":

B. Does the study involve intervention or interaction with a "human subject"?
☒ No ☐ Yes

C. Does the study involve access to identifiable private information?
☒ No ☐ Yes

D. Are data/specimens <u>received</u> by the Investigator with identifiable private information?
☒ No ☐ Yes

E. Are the data/specimen(s) coded such that a link exists that could allow the data/specimen(s) to be re-identified?
☒ No ☐ Yes
If "*Yes*," is there a written agreement that prohibits the PI and his/her staff access to the link?
☐ No ☐ Yes

Oklahoma State University Institutional Review Board

Request for Determination of Non-Human Subject or Non-Research

6. **Signatures**

Signature of PI Dan Li **Date** 7/12/11

Signature of Faculty Advisor ______ **Date** 7-12-11
(If PI is a student)

☒ Based on the information provided, the OSU-Stillwater IRB has determined that this project **does not** qualify as human subject research as defined in 45 CFR 46.102(d) and (f) and **is not subject to oversight by the OSU IRB.**

☐ Based on the information provided, the OSU-Stillwater IRB has determined that this research **does** qualify as human subject research and **submission of an application for review by the IRB is required.**

Shelia M. Kennison
Dr. Shelia Kennison, IRB Chair

7-12-11
Date

VITA

Ketevan Danelia

Candidate for the Degree of

Doctor of Philosophy

Thesis: MODELING EARLY ALCOHOL INITIATION: A COMPARISON OF LINEAR REGRESSION, LOGISTIC REGRESSION, AND DISCRETE TIME HAZARD

Major Field: Human Sciences

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Human Sciences at Oklahoma State University, Stillwater, Oklahoma in December, July, 2011.

Completed the requirements for the Master of Business Administration at Oklahoma State University, Stillwater, Oklahoma in 2004.

Completed the requirements for the Bachelor of Science in Mathematics at Tbilisi State University, Tbilisi, Georgia in 1984.

Name: Ketevan Danelia Date of Degree: July, 2011

Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: MODELING EARLY ALCOHOL INITIATION: A COMPARISON OF LINEAR REGRESSION, LOGISTIC REGRESSION, AND DISCRETE TIME HAZARD MODELS

Pages in Study: 127 Candidate for the Degree of Doctor of Philosophy
Major Field: Human Sciences

Scope and Method of Study: In social science research there is often a need to study the occurrence of a rare event whose distribution is not normal and whose data structure is nested. Common statistical methods for these questions require either the violation of important statistical assumptions or the mishandling of missing data. For data that involve whether an event occurs and when it occurs, the most appropriate statistical model are discrete-time hazard models. However, until recently a method that uses discrete-time hazard models and appropriately adjusts the standard errors to account for the nested structure of the data did not exist. The present study develops three models that combine discrete-time hazard models and hierarchical linear modeling, to model Age of First Use of alcohol, and compares and contrasts these models with more commonly used multiple regression and logistic regression models. To illustrate the advantages of this method, the study evaluates the effects of several common covariates of alcohol use, such as Age of First Opportunity (AFO) of using alcohol, Family Attention (FA), Externalizing Behavior (EXT), Socioeconomic Status (SES), and Gender in a sample of 1785 youth from Caracas, Venezuela.

Findings and Conclusions: Age of first opportunity of using alcohol appears to be the most influential variable in the models. The highest hazard rate of alcohol initiation was found at the first year of opportunity to use alcohol. The results obtained in this study varied across models depending on whether or not AFO was included in models as a covariate. When models did not control for AFO all other independent variables of this study become significant predictors of alcohol initiation in all models except for the logistic regression model where controlling for AFO did not make statistically significant differences in predicting alcohol use. Even though all models considered in the present study have their own advantages, hazard models are seen as the most appropriate in modeling age of first alcohol use. The main advantages of hazard models is in their ability to handle a particular kind of missing data called right censoring, such as youth who report delaying their initiation of alcohol use for all years covered in a given study. In investigating alcohol initiation, only about 18% reported no use of alcohol in this study, but when investigating illicit drugs, many more participants will be in a no-user group. For modeling early ages of drug initiation or any other event occurrence, when a vast majority of participants have not yet experienced it, hazard models should be used.

ADVISER'S APPROVAL: Ron Cox, PhD