**Abstract**

In this manuscript, we use moderated nonlinear factor analysis (MNLFA), a flexible latent variable model, to obtain estimates of social integration and social status from school-based sociometric data. We use these estimates, known as factor scores, to predict smoking onset around the transition to high school, finding that lower integration and higher status predict greater smoking onset hazard. We employ a combination of structural equation modeling and survival models to find that deviance mediates the relationship between social status and smoking onset hazard at the transition to high school. We propose that MNLFA may be useful in generating scores to test hypotheses such as these, which are not able to be tested in most extant models for longitudinal network data.

**A latent variable approach to modeling social dynamics of smoking onset in adolescence**

Numerous developmental theories emphasize the role of the peer context as integral to the emergence of health risk behaviors over time (Prinstein & Giletta, 2016). Advances in social network analyses have allowed such relationships between peer context and adolescent health risk behaviors to be investigated. Although many variants may be found in the literature, there are generally two approaches to using social network analysis to understand adolescent health risk behavior. The first approach is to directly model the co-evolution of social networks and behaviors across multiple time points. This may be done using the stochastic actor-oriented modeling framework (SAOM; Snijders, 1996; Snijders, Van de Bunt, & Steglich, 2010), most often implemented using the SIENA program (Ripley et al., 2017), or through extensions of exponential random graph models (ERGM; Hanneke, Fu, & Xing, 2010; Krivitsky & Handcock, 2014; Robins, Pattison, Kalish, & Lusher, 2007). These models are well-suited to testing hypotheses about peer selection and influence because they directly take into account the time-specific set of connections between individuals in the network and allow these connections to be both predictors and outcomes of health risk behaviors. The second approach is an indirect one which involves first obtaining person-level indicators from sociometric data and then using these indicators in subsequent models to address questions about various relations between placement in the social network and behavior. Network indicators may be used to capture such constructs as social integration (i.e., embeddedness in dense friendship groups and connectedness to others in the network) and social status (i.e., relative social standing and influence in the network; Ennett et al., 2006). Such indicators, here termed person-level social network indices, may be extracted from network data and have the benefit of being easily transported to any other type of analysis.

Each of these two approaches has benefits and drawbacks, but fundamentally they differ in the types of hypotheses that they are best able to evaluate. Direct approaches, such as the SAOM framework, arguably represent the optimal way of understanding the relationship between connections within a given network and the spread of health risk behaviors over that network through selection and influence processes. However, to the extent that peer relationships and network factors serve to moderate or mediate other person-specific risk processes that drive health risk behaviors, these direct modeling approaches may be limited. In contrast, indirect approaches that use person-level indicators from social network analyses are more flexible and may be used to test, for example, developmental pathways spanning multiple domains in which social dynamics are simply one part of a person-level process taking place both within and outside of the peer network. For instance, Ennett et al. (2006) showed that parenting factors (strong bonds and appropriate demandingness) are more strongly associated with substance use for youth who are less integrated into peer networks at school (as measured by indices such as outdegree and reciprocity) or who have lower social standing relative to their peers (as measured by indices such as betweenness centrality). Such measures may similarly be used to test questions about mediation, such as whether greater youth depression leads to lower social integration and standing in a peer network, which then lead to greater substance use (Authors et al., under review). They may also be used to test complex intra-individual hypotheses, such whether indicators of low social status and integration contribute to a multivariate risk cascade that includes stress-related processes crossing the biological-psychological-social-institutional levels of analysis. Although direct approaches such as the SAOM may represent the state-of-the-art for testing questions about influence and selection, they are not able to model processes at the level of the individual such as these. Thus, there continues to be a need for indirect approaches which use person-level indicators of social network involvement, and which allow greater model flexibility to test some questions about the broader development context of health risk behaviors over time.

The current use of person-level social network measures to test such hypotheses entails multiple assumptions about measurement that may not always be tenable. Here we discuss these assumptions, introduce a latent variable model to test these assumptions (the moderated nonlinear factor analysis, MNLFA; Bauer and Hussong, 2009), and offer a potential solution in cases where these assumptions do not hold. The aims of this paper are twofold. First, we introduce the MNLFA and use it to obtain estimates of social integration and social status from person-level indices resulting from analyses of six longitudinal school networks. Second, we illustrate how to use of the resultant scores to test mediational hypotheses that could not be evaluated in the SAOM framework. Before introducing MNLFA, we motivate its use by considering in greater detail the problems associated with measurement of social dynamics using person-level social network indicators.

**Measurement and person-level social network indices in models for behavior**

In this paper, we focus on a commonly used indirect or two-step strategy in which some person-level index of social status or social integration is first obtained from a social network analysis, and then linked to a set of behavioral predictors or outcomes in subsequent analyses. In this literature, numerous operationalizations of such constructs of social integration and status are used, often with little account for their inter-relation. Table 1 shows a list of commonly-used person-level indices. Person-level indices associated with social integration include outdegree, out-of-school nominations, reciprocity, transitive triads, and intransitive triads. Person-level indices associated with social status are similarly plentiful and include indegree, three-step reach, Bonacich centrality, and betweenness centrality. Such indices of social status and integration have been important predictors of smoking onset and behavior in adolescence (Abel, Plumridge, & Graham, 2002; Alexander, Piazza, Mekos, & Valente, 2001; Ennett et al., 2006; Ennett et al., 2008; Lakon & Valente, 2012; Valente, Unger & Johnson, 2005), although modeling such measures without taking into account their inter-relation has resulted in a complicated pattern of findings. For example, social integration, operationalized using indices such as friendship reciprocity and neighborhood density (Ennett et al., 2006), negatively predicts smoking. However, the nature of the relationship between smoking and social status depends somewhat on how status is measured: whereas Alexander, Piazza, Mekos, and Valente (2001) find that students with more incoming nominations are more likely to smoke, Ennett and colleagues (2006) did not find any relationship between smoking and normed indegree, three-step reach centrality, betweenness centrality, or power centrality when these were all included in the model simultaneously.

The key advantages of using these measures is flexibility and reduced reporter bias: researchers may obtain a several indices of a subject’s social integration or social status that account for the structure of the network without relying on self- or teacher-reported measures with well-known biases. Like any variables, these indices may be used in complex models for behavior. However, this flexible practice makes several assumptions about what these indices mean and how they are measured. Notably, researchers assume that a person-level index represent some underlying feature of an individual’s experience in a social network, such as social status or integration. In the social sciences, quantities such as these are often referred to as latent variables. Though many different definitions of latent variables exist (Lord and Novick, 1968; Bollen, 2002; MacCallum and Austin, 2000), a latent variable can be generally defined as a construct that cannot be measured directly but which must be indexed by some observed indicator. Under the interpretation of person-level social network indices as arising from latent variables such as social integration and social status, the true network would be defined as the one whose state is exclusively a function of these latent variables.

A whole set of methods falling under the heading of factor analysis (Joreskog, 1969; Lord & Novick, 1968) have been designed to address the issues arising from measuring latent variables with observed indicators. However, the use of person-level indices from social network analyses typically makes two additional assumptions. First, the use of an index is treated as a perfect measure of the latent construct of interest, devoid of random or systematic error in measurement. Second, this practice often entails the assumption that the relationship between the construct and the index is not confounded by any extraneous variables. These issues of measurement error and covariate bias are discussed in turn.

**Measurement error and person-level social network indices.** Typically, measurement error in self-report measures arises from the interaction between the respondent and the question or between the respondent and the testing environment. This problem may be compounded for social network measures because person-level network indices depend not only on a given subject's responses but on those of the whole network; as a result, a wide variety of sources of measurement error are possible in person-level indices resulting from social network analysis.

Defining each person in a given network as a node, and each tie between people (e.g., a friendship nomination) as edges, Wang et al. (2012) classify measurement error in networks as resulting from three potential sources: false positive or negative edges; false positive or negative nodes; and falsely aggregated or disaggregated nodes. These first two cases are particularly likely to result from school-based surveys. Edges may be spuriously added or omitted if a respondent intentionally names peers of perceived higher status, omits those of perceived lower status, or simply forgets to report names from all friends (Bernard et al., 1984; Marsden, 1990). Measurement error in social network measures may also arise from common limitations due to study design, such as limiting the criteria for prospective nodes' inclusion in the network (i.e., the boundary specification problem; Laumann, Marsden, and Prensky, 1989) or capping the number of nominations individuals can make. Additionally, absence from data collection may result in node-level missingness if absentees are omitted from the analysis entirely or edge-level missingness if absentees are nominated by others but, due to their absence, unable to make any nominations themselves.

A growing body of research suggests that spuriously added or omitted nodes or edges may lead to considerable bias in person-level measures of centrality and homophily (Borgatti et al., 2006; Costenbader & Valente, 2003; Kossinets, 2006; Smith and Moody, 2013). Additionally, this bias is more severe in smaller and less cohesive networks (Wang, 2012; Smith, Moody, and Morgan, 2017). Although there is a growing consensus that measurement error may lead to bias in person-level indices, this concern is generally ignored when researchers use them as predictors and outcomes in models. In the next section we will introduce a strategy that addresses this concern.

**Covariate-related bias and person-level indices.** The quality of measurement by person-level social network indices may vary across subjects who differ on gender, race, and age. For example, indegree may be a stronger indicator of social status for older subjects relative to younger ones, or girls relative to boys. When person-level social network indices are used to measure latent constructs, the interpretation of between-person differences must account for the possibility that the measurement process itself, rather than the construct being measured, differs between subjects. In other words, it is not always clear whether individuals of different genders, races or ages differ in the latent construct that the index is presumed to measure (say, social integration) or if apparent differences in the latent variable across groups actually reflect differences in how well the index measures the underlying construct across individuals of different genders, races, or ages. In the latent variable literature, these questions refer to the topic of measurement invariance (Byrne, Shavelson, & Muthen, 1989; Meredith, 1993; Millsap, 2012) or differential item functioning (DIF; Thissen, Steinberg, & Wainer, 1993; Osterlind & Everson, 2009). Under the assumption of measurement invariance, the relationship between an item (here, a person-level social network index) and the latent construct it is supposed to measure is identical across all individuals under study. In the presence of measurement invariance (i.e., in the absence of DIF), all variation in the observed variable can be attributed to the latent variable(s) it measures, rather than to background variables. If an index shows DIF on the basis of some covariate, however, that index does not necessarily mean the same thing across individuals.

Measurement invariance has not been studied formally in social network studies to our knowledge. However, the complex set of relationships between individuals who generate person-level social network measures has led many researchers to question how the meaning of these indices changes based on individual- and network- level characteristics. For instance, the effects of indices measuring popularity, such as proximity prestige (a weighted index of incoming nominations; Kornienko and Santos, 2014) and betweenness centrality (Faris & Felmlee, 2011; 2014), on mental health outcomes, such as depression and relational aggression, have been shown to be moderated by gender. Additionally, Faris and Felmlee (2011; 2014) found that female students had higher overall levels of betweenness centrality than male students in a sociometric study of peer relationships and aggression in schools. However, they point out that betweenness centrality may be differentially related to social status based on gender segregation in sociometric studies of schools. These authors note that "gender bridges" (i.e., students who have friends of different genders in highly gender-segregated schools) are likely to be of high status, but that this is not reflected in measures such as centrality (2011; pp. 54-55). As such, differences between male and female adolescents in centrality are not directly interpreted as female adolescents being of higher status than male adolescents; rather, these differences may be due to subtle discrepancies across genders in how status is operationalized.

**Moderated nonlinear factor analysis**

Above we review two potential threats to valid inferences about relations between social network constructs and behavior based on person-level social network indices: (1) that social network data is measured with error, and (2) that there may be systematic differences between individuals in the relationship between indices and the constructs they measure. One widely used, versatile set of techniques that deal with measurement error is known as factor analysis. Though a full review of factor analysis is not given here (see Hoyle, 2000; Ford, MacCallum, & Tait; 1986), the basic premise is that common variance among a collection of indicators is accounted for by a smaller set of latent variables that are not directly measured (Joreskog, 1969).

However, in most formulations of the common factor model, there is no explicit way to deal with the second problem of systematic bias in observed scores due to background characteristics: the observed variable is assumed to measure the latent variable equally well across all individuals in the sample. An extension of factor analysis, the moderated nonlinear factor analysis (MNLFA; Bauer & Hussong, 2009) handles this problem by incorporating covariates directly into the expressions for measurement parameters relating the observed and latent variables.

A schematic of an MNLFA for modeling social integration from person-level indices is shown in Figure 1; mathematical expressions for the model are given in the Appendix. Though Figure 1 depicts one of the models we apply in the current study (with the other being for social status), the following description generally applies to any instance of MNLFA. A latent variable model, the MNLFA hypothesizes that the indicators – in this case, transitive triads, intransitive triads, reciprocity, out-of-school friends – all measure the same construct, in this case social integration. These indicators are related to the latent variable through factor loadings, which are denoted by arrows between the latent variable and each indicator in Figure 1, represent the predicted change in indicators associated with a one-unit shift in the latent variable. Additionally, each indicator is characterized by an intercept, which represents the predicted value of the indicator when the value of the latent variable is zero.

The MNLFA allows covariates to explain observed differences in person-level indicators in two broad ways. First, the mean and variance of the latent variable may be affected by covariates, thus producing a corresponding change in the person-level index. Effects of covariates on the latent variable are referred to as impact. Suppose the measures in Figure 1 are used to measure social integration in an MNLFA. Suppose further that boys in a given network show lower overall outdegree than girls. This may be caused by mean impact, if boys are truly less socially integrated than girls in their peer networks; in this case the path from Male to the social integration latent variable (the MI path) would be negative.

Second, the parameters relating the indices to the latent variable may be affected by covariates. Effects of covariates on items, over and above the latent variable, are referred to as differential item functioning (DIF). Returning to the example of boys showing low outdegree than girls in the model in Figure 1, intercept DIF would occur if boys simply nominated fewer of their peers as friends than girls did, regardless of their true level of social integration; in this case, the path from Male to outdegree (ID) would be negative. Similarly, if outdegree simply isn't as strong an indicator of social integration for boys as it is for girls, this would manifest as loading DIF, and the path labeled LD would be negative.

Explicitly modeling impact and DIF allows substantively interesting effects of covariates on a construct of interest (i.e., impact) to be distinguished from effects of covariates on the items relating to some extraneous process outside of the latent variable (i.e., DIF). Numerous effects may occur at the person level (e.g., gender, race, SES) and, given the presence of data from multiple networks, at the network level (e.g., network size; network density; clustering), where the ability to explicitly model this distinction may be particularly useful. For instance, consider the example of network size. Because several person-level social network indices increase or decrease with network size, many measures such as betweenness centrality are normalized by dividing by network size. However, the relationships between some latent constructs and other person-level indices (e.g., outdegree and indegree, which may not increase deterministically with network size) may be affected by network size in ways that are more complex and difficult to model reliably.

**Illustration of MNLFA in creating latent factors for social integration and status**

We now demonstrate the use of MNLFA to assess the measurement of social status and integration by person-level social network indices. After assessing the measurement of these constructs using MNLFA, we generate scores which are used to examine the effects of social status and integration on smoking onset. Finally, we test the hypothesis that deviance mediates the effect of social status on smoking onset using a structural equation model.

**Sample and Measures**

**Participants.** Data come from the Context Study, which was designed to support investigation of individual and contextual factors (i.e., family, peer social network, school, and neighborhood contexts) that influence the development of substance use and other problem behavior from early to late adolescence. A full description of participants and study design is available elsewhere (Ennett et al., 2006; Authors, submitted), but these are described briefly here. The study used a cohort-sequential design in which three cohorts of adolescents in the 6th, 7th, and 8th grades from three complete school districts in three primarily rural North Carolina counties were enrolled in the study and surveyed in school every six months for five data collection waves. Adolescents in two of the three school districts were surveyed in two additional waves, six and 12 months later. At wave 1, adolescents were enrolled in all 13 schools with middle grades (grades 6, 7, 8) in the three study school districts; three of these schools were alternative schools that included middle and high school grades. Beginning with wave 2, when the first adolescents transitioned to high schools, the school sample added all six high schools in the districts. The school sample size fluctuates across waves depending on the inclusion of middle and high schools and due to a single school system not participating at waves 6 and 7. The school-based design allowed measurement of peer social networks bounded by school enrollment and, during middle grades, by grade. Networks were defined on the basis of high school; as such, “school” here refers to the high school a student ultimately attends, even if they are in middle school at a given time point. Thus, there were a total of six schools, denoted School A-School F. Sample sizes for each school were *N* = 1677 for School A, *N* = 996 for School B, *N* = 493 for School C, *N* = 1642 for School D, *N* = 1015 for School E, and *N* = 1175 for School F.

At each of the seven waves, adolescents completed a self-report battery which assessed mental health, peer and family relationships, and alcohol, tobacco and other substance use. Additionally, at each wave students completed a sociometric survey in which they were asked to nominate up to five of their closest friends, starting with their best friend. Nominations were made using a student directory, which contained an alphabetical listing of students and a unique four-digit identification code for each student. Out-of-school nominations could also be made using the identification number “0000.”

**Measures. *Demographic measures.*** Demographic measures included adolescent-reported gender (effect-coded for analysis), race/ethnicity, child-reported parental education level, and grade (ranging over waves from Spring of Grade 6 to Grade 12). Though originally in ordinal scale, parental education was trichotomized to represent low (high school or less), medium (more than high school but less than a 4-year degree), and high levels of education (4-year degree or more), with medium used as the reference category. Additionally, due to low sample size of Hispanic/Latino students across schools, comparisons across race were limited to White and Black students, with White used as the reference category.

***Smoking and deviance measure****s.* Smoking onset was measured at each wave with the item, “How much have you ever smoked in your life?” There were seven response options: “none at all, not even a puff” (0); “1 or 2 puffs, but not a whole cigarette” (1); “3 or more puffs but not a whole cigarette” (2); “1 to 2 whole cigarettes” (3); “3 to 5 whole cigarettes” (4); “6 to 20 whole cigarettes” (5); and “more than 20 whole cigarettes at once” (6). Onset as a binary variable representing any amount of smoking, which took a value of 1 if the subject had ever smoked at all -- i.e., if the subject gave a response above 0 on the original item -- and 0 otherwise.

Deviance was a factor score computed from 15 items, in five-point ordinal scale, from the Problem Behavior Frequency Scale (Farrell, Kung, White, & Valois, 2000); the computation of deviance scores is described in greater detail by Authors (submitted).

***Social network measures****.*The above social network analysis resulted in a directed network, from which nine person-level social network indices were obtained using a combination of the UCINET program (Borgatti, Everett, and Freeman, 2002) and a set of SAS macros authored by James Moody. Person-level measures, listed in Table 1, were assumed to measure two separate but related latent constructs: social status and social integration. Measures of social status included normed betweenness, Bonacich centrality, three-step reach, and indegree. Measures of social integration included transitive triads, intransitive triads, reciprocity, out-of-school friends (reverse scored so that more out-of-school friends represents lower social integration), and outdegree.

A plurality of these indicators are count variables (i.e., indegree; outdegree; reciprocated ties; out-of-school friends). We evaluated the density plots of each and, determining that neither zero-inflation nor overdispersion were present, chose a Poisson distribution to model these person-level indices for all MNLFAs moving forward. The remaining indicators, including Bonacich centrality, three-step reach, normed betweenness, transitive triads, and intransitive triads, were continuous, but the distributions of these variables varied widely from one school to the next. Because it was untenable to use different link functions to model them across schools (e.g., a lognormal distribution and a normal distribution in the other), we chose to recode these indicators as categorical variables as shown in Table 1 by binning data at percentiles (i.e., a median split for Bonacich and three-step reach centrality; at 33% and 66% for transitive triads, intransitive triads, and normed betweenness). This choice ensured comparability of the indicators across schools and allows us to reduce model complexity.

**Generating social integration and social status scores using MNLFA**

MNLFA models were fit separately to six nonoverlapping samples corresponding to the six high schools to allow for relationships between the person-level indices, latent variables, and covariates to differ among the schools. The goal of the MNLFA fitting was to create factor scores indexing social status and social integration (see Table 1) that take into account impact and DIF due to the following covariates: gender, race, grade, network size, parental education, and cohort. We followed the sequence of steps described by Curran et al. (2014; 2017) for fitting an MNLFA and generating scores. Though a detailed description and rationale for these steps is provided elsewhere (Curran et al. 2014; Curran et al., 2017; Gottfredson et al., submitted), we give a general account of the procedure here[[1]](#footnote-1).

First we established that social status and integration were each unidimensional by fitting a confirmatory factor analysis (CFA; Joreskog, 1969) Because the response distributions of the outcomes were parameterized using nonlinear link functions, the usual set of fit statistics (e.g., RMSEA, CFI) were not available. However, loadings were generally large (standardized loadings > .5) and significant, providing indirect evidence that social integration and status were well-measured by the factors. Given the complexity of fitting multivariate MNLFAs and our confidence that each construct was well-measured by its constituent indicators, we proceeded with fitting two univariate MNLFA’s, one for social integration and one for social status. This step confirmed that outdegree, number of transitive triads, number of intransitive triads, reciprocity, and number of out-of-school friends were suitable measures of social integration; and that indegree, Bonacich centrality, three-step reach, and normed betweenness were suitable measures of social status. Additionally, we conducted graphical analyses in which the relationship between each indicator and each covariate was visually examined. Visual inspection revealed differences in each of the items according to one or more covariates, indicating the potential presence of at least some covariate effects -- i.e., potential impact or DIF.

Following these exploratory steps, we drew a calibration sample consisting of one randomly-sampled observation for each individual. This was done to account for nesting of observations (i.e., multiple time points within a given person), as is standard in applications of IRT and MNLFA (Swaminathan and Hambleton, 2013; Bauer and Hussong, 2009). Parameter estimates were obtained (as described below) using this calibration sample, after which scores were generated for the entire sample.

We then sequentially tested for impact and DIF on the basis of covariates. That is, for each covariate effect, a model containing that covariate's impact on the mean of the latent variable was fit, as were models testing each covariate's potential DIF effects on each item. In the current set of models, it was not hypothesized that either social integration or social status would show more variance according to any of the covariate; therefore, variance impact was omitted. A penultimate model containing all impact and DIF effects found to be significantly different from zero in these itemwise tests was then fit. Finally, nonsignificant effects were pruned using a Benjamini-Hochberg correction for multiple comparisons (Thissen, Steinberg, and Kuang, 2002).

The above yielded a total of twelve final models, one for social integration and social status within each of the six high schools. From each of these final models, modal a posteriori (MAP; Bock & Aitkin, 1981) scores for social integration and social status were obtained. Each MAP score represents an individual’s estimated level of social status or social integration, which is then used as a predictor or outcome in subsequent models.

**Results.** Descriptive statistics for factor scores within all schools are shown in Table 2. In all schools, as expected, there are strong positive correlations between each score and its constituent indicators; additionally, there are strong positive correlations between the two scores. Notably, the means of social integration and social status scores differ widely across the schools. This is likely due to the inclusion of different covariates in the models for each school's means. Thus, it was critical to include school as a fixed effect, as well as all covariates used in generating scores, in all subsequent analyses.

Table 3 summarizes all significant covariate effects found in the aMNLFAs across all six schools. Importantly, DIF effects for social network size were necessary to include for a few indices including indegree, transitive triads, and intransitive triads, which are all directly proportional to network size. Thus, loading and intercept effects of network size were included on indegree in the social status model and for transitive and intransitive triads in the social integration model, regardless of whether item-wise tests were significantly different from zero. As shown, schools differed substantially in which covariates were linked to either differences in the factor means, or DIF in individual items. Among mean impact parameters, the most frequently observed effect was a negative effect of grade on both social integration ( in Schools A, C, and F, respectively) and social status ( in Schools B, C, and E, respectively). Additionally, social integration was lower in subjects with lower levels of parental education in schools B, D, E, and F (where  respectively), and social status was higher for Black students in schools B and E (where  respectively) and trivially lower in school D (where ). This means, in general, that older students and students with lower parental levels of education were less socially integrated, and that younger students and Black students were of higher social status, within their networks.

DIF effects were less pervasive, but most frequently associated with gender and grade. In particular, in five of six schools, intercept DIF for reciprocity was found on the basis of gender, such that male subjects reported lower overall levels of reciprocity after controlling for social integration ( in Schools A, B, C, D, and F). This indicates that male students had fewer nominated friends reciprocate their nominations, even after controlling for social integration. DIF was also found for transitive and intransitive ties in the model for social integration, as well as Bonacich centrality, indegree, and three-step reach in the model for social status.

**Modeling smoking onset from social integration and social status**

After scores representing social integration and social status were generated, we used these scores as predictors in a series of models for smoking onset. In particular, we examined the temporal relationship between social integration, social status, deviance, and smoking onset.

**Discrete-time hazard models of smoking onset.** The model for smoking onset during high school was a discrete-time survival model (Allison, 1984), which models the probability of a binary event taking place during some fixed number of time periods. Discrete-time survival analysis is described in greater detail elsewhere (see Allison, 1984; Singer & Willett, 1993; Willett & Singer, 1991). However, the two main functions under discussion are the hazard function, which represents the probability that a subject’s experiences smoking onset during a given time interval, and the survival function, which represents the probability that a subject has not experienced smoking onset by the end of a given time interval. The hazard of smoking onset was modeled as a function of several predictors. All covariates used in the estimation of the MNLFA (race, gender, cohort, parental education, network size and school) s. Aside from network size, which is treated as time-varying, all predictors were included as time-invariant covariates. The predictors of interest here were social status, social integration, and deviance. Though they were available at all time points, the independent variables -- social integration, social status, and deviance -- were considered here as time-invariant predictors, owing to the difficulty of interpreting time-varying effects. Average values of social integration, social status, and deviance were computed across all time points and used as predictors here.

Finally, structural relationships between social status and deviance in the prediction of smoking hazard were examined using a discrete-time survival mediation model (DTSMM; Fairchild et al., 2015), a type of structural equation model which allows for a sequence of multiple temporally-ordered variables to predict a survival process. Whereas the regressions above may determine whether subjects’ average levels of social integration, social status, and deviance predict greater smoking hazard overall, the DTSMM helps to determine the mechanism by which these variables affect one another and in turn lead to greater smoking hazard. The DTSMM is part of a growing set of models which treat the hazard of event occurrence as a latent variable in a structural equation model (Raykov et al., 2017; Muthen and Masyn, 2005). In the absence of structural relationships (e.g., a mediated path predicting survival), the DTSMM is identical to a typical discrete-time survival model. Preliminary analyses indicated that no mediated relationship existed between social integration and smoking onset. Therefore, we focus exclusively on social status here.

A key assumption of inferences based on the DTSMM is temporal precedence – that is, that the predictor precedes the mediator, which precedes the outcome (MacKinnon, Fairchild, and Fritz, 2007). Therefore, to include sequentially ordered predictors, mediators, and outcomes, we focused exclusively on a subset of cases and time points for these analyses. Hazard functions began at fall of 7th grade; predictors (social integration and status) were measured at spring of 6th grade; and the mediator (deviance) was measured at fall of 7th grade. Due to the cohort-sequential design of the study, only cases in one cohort (Cohort 1) were measured in spring of 6th grade, and these cases were only measured at seven time points: spring of 6th grade, fall of 7th grade, spring of 7th grade, fall of 8th grade, spring of 8th grade, fall of 9th grade, and fall of 10th grade. Thus, the sample was limited to Cohort 1 (N = 1236), and included only these seven time points.

**Regression models.** Four discrete-time hazard regression models were fitted. First, a baseline model (Model 1) with no predictors determined the general pattern of smoking onset hazard across the study period. Second, a model with all demographic and methodological covariates (Model 2) tested whether smoking hazard was invariant across gender, race, parental education, cohort, school, and network size. Two models adding the effects of social integration, social status, and deviance were then fit (Models 3a and 3b). Model 3a included social integration, social status and deviance factor scores as predictors. For comparison, Model 3b included as predictors the individual person-level social network indices, in their original scales, used in the calculation of factor scores. Parameters from all discrete-time hazard models fit to the data are shown in Table 4.

Model 1 shows that the hazard of smoking onset starts at  in fall of sixth grade and accelerates relatively quickly in middle school, with more than 50% of the sample initiating smoking before spring of 7th grade. Model 2, which added predictor effects from all covariates, showed that smoking onset probability was slightly higher among Black participants. Strong differences were observed with respect to parental education, with low parental education linked to higher risk for smoking and high parental education linked to lower risk for smoking relative to the medium-education group.

Model 3a, which added factor scores for the mean levels of social integration, social status, and deviance, showed that, as expected, higher average levels of social integration were linked to lower hazard of smoking onset whereas higher levels of social status and deviance were linked to higher hazard of smoking onset. Model-predicted survival curves for subjects one standard deviation above and below the mean of social integration and status, holding all other predictors at sample averages, are shown in Figure 2. The probability of smoking onset increases rapidly during the middle school years for all subjects, but is lower overall for individuals of low social status or, to a lesser extent, high social integration.

In Model 3b, which included raw person-level indices as predictors, findings were more mixed. Three of the five indicators of social integration were linked with lower overall smoking onset probability; outdegree and the number of intransitive triads were not. Two of the four indicators of social status were predictive of smoking hazard, but these effects went in opposite directions: indegree was positively related to smoking onset hazard, whereas Bonacich centrality was negatively related to smoking onset hazard.

**Discrete-time survival mediation model.** Parameter estimates for the DTSMM are shown in Figure 3. Logit parameters of the baseline smoking onset hazard, shown in the rightmost portion of the figure, are comparable to their corresponding values in the discrete-time hazard models estimated for the larger sample. Among the covariates, only low parental education and attendance to School C significantly increased the hazard of smoking onset.

The primary goal of the DTSMM was to test whether social status in spring of 6th grade predicted deviance in fall of 7th grade, which in turn predicted greater smoking onset hazard – i.e., whether social status in spring of 6th grade exerted an indirect effect on smoking hazard, mediated by deviance in fall of 7th grade. However, to address the possibility of an alternate causal mechanism (i.e., deviance in spring of 6th grade predicting social status in fall of 7th grade, which predicts smoking onset), we included both social status and deviance at both time points. Deviance in the spring of 6th grade predicted higher deviance, but not higher social status, in the fall of 7th grade. In addition, social status in the spring of 6th grade predicted both higher social status and higher deviance in the fall of 7th grade. Additionally, deviance in the fall of 7th grade predicted smoking onset hazard, but social status in the fall of 7th grade did not.

Mediation analysis enables the magnitude of the total indirect path (the path through all mediators) and specific indirect paths (the path through any given mediator) from social status and deviance in fall of 6th grade to smoking onset hazard; corresponding standard errors are calculated using the delta method (Sobel, 1992). The hypothesized positive indirect effect of social status in spring of 6th grade on smoking onset hazard, transmitted through deviance in fall of 7th grade, was significantly different from zero (effect size =0.036, SE = 0.018, *p* = .047). Additionally, the indirect effect of deviance in spring of 6th grade on smoking onset hazard, transmitted through deviance in fall of 7th grade, was significantly different from zero (effect size =0.138, SE = 0.020, *p* < .001). Neither of the two possible indirect effects including social status in the fall of 7th grade was significantly different from zero. Thus, both deviance and social status in the spring of 6th grade were linked to higher levels of deviance in the fall of 7th grade, which increased the overall hazard of smoking onset.

**Discussion**

In the current analysis we applied moderated nonlinear factor analysis (MNLFA; Bauer and Hussong, 2009), a flexible latent variable technique, to person-level indices resulting from a social network analysis. This allowed for composite indices representing social status and social integration to be extracted from longitudinal network data in a way which accounted for error in measurement and potential confounding of covariates. We then linked these scores to the hazard of smoking onset during the transition from middle to high school. As hypothesized, adolescents with higher levels of social status and lower levels of social integration were more likely to smoke throughout the study period. Finally, subsequent mediation analyses indicated that deviance in the fall of 7th grade mediated the effect of social status in 6th grade on later smoking hazard during the transition between middle and high school.

Our methodological aim in introducing and demonstrating MNLFA was to give researchers a tool for summarizing complex social dynamics, and placing these social dynamics within models for human growth and development. This widens the potential use of sociometric data to the testing of hypotheses in which social dynamics are a portion of a complex within-person process. This flexibility is necessary if researchers are to integrate sociometric data into the study of broader etiological models of health risk behaviors, which necessarily integrates information across contexts (e.g., biological, cognitive, and social) and time scales (e.g., moments, weeks, and semesters) to distinguish between normal and abnormal development (Cicchetti & Rogosch, 2002). Future work may focus on using MNLFA to pool network data across multiple different studies covering different but overlapping age ranges in the service of this goal, as has been its primary use with self-report data (Curran et al., 2014; Hussong, Curran, and Bauer, 2013).

The findings linking higher social status and lower social integration to smoking are consistent with results from studies using single person-level social network indices to the development of smoking behavior in adolescence (Abel, Plumridge, & Graham, 2002; Alexander, Piazza, Mekos, and Valente, 2001; Ennett et al., 2006; Ennett et al., 2008; Lakon & Valente, 2012; Valente, 2005). When entered into the model separately, these indices were also linked to smoking hazard. The findings were partially consistent with those based on the factor scores. Most indices measuring social integration were linked to lower levels of smoking hazard. However, links from social status to smoking hazard were much less consistent, with both positive and negative effects from putative indices of social status. This may occur because the effect of each person-level index was a partial effect, controlling for the overlapping variance with the other. The use of MNLFA treats this overlapping variance as a strength, as opposed to a weakness: it represents a latent factor, in this case social status, measured by all relevant indices in the aggregate. This is a key advantage of latent variable approaches such as MNLFA.

The mediation finding in the current study places deviance temporally between social status and smoking onset hazard over the course of middle and high school. If there are indeed causal pathways between social status, deviance, and smoking onset, these relationships must be further probed with respect to other mediators and time scales. For instance, given the complex pattern of variables predicting smoking onset (Conrad, Flay, & Hill, 1992), it will be important to investigate whether smoking-specific constructs such as tobacco-related knowledge, refusal self-efficacy, and peer smoking norms also mediate the relationship between social status and smoking onset. Additionally, given that the predictors and sequelae of early-onset and late-onset smoking are different (e.g., Grant 1998; Saules et al., 2004), this mediated pathway may differ over smoking outcomes and stages of development. As demonstrated in the current report, MNLFA may be applied to sociometric data in the service of both of these goals.

Although we believe that MNLFA is a useful addition to the arsenal of tools used to understand the role of the peer network in health risk behaviors, methodological advancement to the MNLFA framework are still needed. As noted by Steglich, Snijders, and Pearson (2010), the problem of network dependence (i.e., the fact that the network structure of the cases renders the assumption of independent residuals untenable) is incompletely addressed by most methods outside of models for network dynamics such as the SAOM (Snijders, 1996; Snijders, Van de Bunt, & Steglich, 2010). The MNLFA partially addresses network dependence indirectly by controlling for covariates, as residuals in the person-level indices will be independent to the extent that covariates such as gender, race, and age account for some of the dependence between observations. However, the problem of network dependence when using MNLFA scores should be investigated further in future work. One exciting possibility is the use of MNLFA scores as predictors in a SAOM, in place of person-level indices -- e.g., allowing a factor score representing social status, rather than centrality, to affect the probability of tie formation. This would allow the structural and temporal dependence of network observations to be directly modeled, while accounting for measurement error in the network. Future work will focus on assessing the ability of MNLFA to rectify measurement issues in person-level social network indices through simulation studies. Our hope is that these assessments will further expand MNLFA's usefulness to the study of social dynamics in health and development.

**References**

Abel, G., Plumridge, L., & Graham, P. (2002). Peers, networks or relationships: strategies for understanding social dynamics as determinants of smoking behaviour. *Drugs: education, prevention and policy, 9*(4), 325-338.

Alexander, C., Piazza, M., Mekos, D., & Valente, T. (2001). Peers, schools, and adolescent cigarette smoking. *Journal of adolescent health, 29*(1), 22-30.

Allison, P. D. (2014). Event history and survival analysis: *Regression for longitudinal event data* (Vol. 46). SAGE publications.

Bauer, D. J., & Hussong, A. M. (2009). Psychometric approaches for developing commensurate measures across independent studies: traditional and new models. *Psychological methods, 14*(2), 101.

Bernard, H. R., Killworth, P., Kronenfeld, D., & Sailer, L. (1984). The problem of informant accuracy: The validity of retrospective data. *Annual review of anthropology, 13*(1), 495-517.

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual review of psychology, 53*(1), 605-634.

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). *Ucinet for Windows: Software for social network analysis.* Analytic Technologies, Harvard, MA. Available at: http://www.analytictech.com/ucinet.htm.

Borgatti, S. P., Carley, K. M., & Krackhardt, D. (2006). On the robustness of centrality measures under conditions of imperfect data. *Social networks, 28*(2), 124-136.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological bulletin, 105*(3), 456.

Cicchetti, D., & Rogosch, F. A. (2002). A developmental psychopathology perspective on adolescence. *Journal of consulting and clinical psychology, 70*(1), 6.

Conrad, K. M., Flay, B. R., & Hill, D. (1992). Why children start smoking cigarettes: predictors of onset. *Addiction, 87*(12), 1711-1724.

Costenbader, E., & Valente, T. W. (2003). The stability of centrality measures when networks are sampled. *Social networks, 25*(4), 283-307.

Curran, P. J., McGinley, J. S., Bauer, D. J., Hussong, A. M., Burns, A., Chassin, L., ... & Zucker, R. (2014). A moderated nonlinear factor model for the development of commensurate measures in integrative data analysis. *Multivariate behavioral research, 49*(3), 214-231.

Ennett, S. T., Bauman, K. E., Hussong, A., Faris, R., Foshee, V. A., Cai, L., & DuRant, R. H. (2006). The peer context of adolescent substance use: Findings from social network analysis. *Journal of research on adolescence, 16*(2), 159-186.

Ennett, S. T., Foshee, V. A., Bauman, K. E., Hussong, A., Cai, L., Reyes, H. L. M., ... & DuRant, R. (2008). The social ecology of adolescent alcohol misuse. *Child development, 79*(6), 1777-1791.

Fairchild, A. J., Abara, W. E., Gottschall, A. C., Tein, J. Y., & Prinz, R. J. (2015). Improving our ability to evaluate underlying mechanisms of behavioral onset and other event occurrence outcomes: A discrete-time survival mediation model. *Evaluation & the health professions, 38*(3), 315-342.

Faris, R., & Felmlee, D. (2011). Status struggles: Network centrality and gender segregation in same-and cross-gender aggression. *American Sociological Review, 76*(1), 48-73.

Faris, R., & Felmlee, D. (2014). Casualties of social combat: School networks of peer victimization and their consequences. *American Sociological Review, 79*(2), 228-257.

Farrell, A. D., Kung, E. M., White, K. S., & Valois, R. F. (2000). The structure of self-reported aggression, drug use, and delinquent behaviors during early adolescence. *Journal of clinical child psychology, 29*(2), 282-292.

Ford, J. K., MacCallum, R. C., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel psychology, 39*(2), 291-314.

Grant, B. F. (1998). Age at smoking onset and its association with alcohol consumption and DSM-IV alcohol abuse and dependence: results from the National Longitudinal Alcohol Epidemiologic Survey. *Journal of substance abuse, 10*(1), 59-73.

Hambleton, R. K., & Swaminathan, H. (2013). *Item response theory: Principles and applications*. Springer Science & Business Media.

Hanneke, S., Fu, W., & Xing, E. P. (2010). Discrete temporal models of social networks. *Electronic Journal of Statistics, 4*, 585-605.

Hoyle, R. H. (2000). Confirmatory factor analysis. *Handbook of applied multivariate statistics and mathematical modeling,* 465-497.

Hussong, A. M., Curran, P. J., & Bauer, D. J. (2013). Integrative data analysis in clinical psychology research. *Annual Review of Clinical Psychology, 9,* 61-89.

Knott, M., & Bartholomew, D. J. (1999). *Latent variable models and factor analysis* (No. 7). Edward Arnold.

Jessor, R., & Jessor, S. L. (1977). Problem behavior and psychosocial development: A longitudinal study of youth.

Joreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis*. Psychometrika, 34*(2), 183-202.

Kornienko, O., & Santos, C. E. (2014). The effects of friendship network popularity on depressive symptoms during early adolescence: Moderation by fear of negative evaluation and gender. *Journal of youth and adolescence, 43*(4), 541-553.

Kossinets, G. (2006). Effects of missing data in social networks. *Social networks, 28*(3), 247-268.

Krivitsky, P. N., & Handcock, M. S. (2014). A separable model for dynamic networks*. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 76*(1), 29-46.

Lakon, C. M., & Valente, T. W. (2012). Social integration in friendship networks: The synergy of network structure and peer influence in relation to cigarette smoking among high risk adolescents. *Social science & medicine, 74*(9), 1407-1417.

Laumann, E. O., Marsden, P. V., & Prensky, D. (1989). The boundary specification problem in network analysis. *Research methods in social network analysis, 61*, 87.

Lord, F. M. M. R. Novick (1968), *Statistical Theories of Mental Test Scores*. Reading, PA: Addison-Wesley.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual review of psychology, 51*(1), 201-226.

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annu. Rev. Psychol., 58*, 593-614.

Marsden, P. V. (1990). Network data and measurement. *Annual review of sociology, 16*(1), 435-463.

McCullagh, P., & Nelder, J. A. (1989). Generalized Linear Models, no. 37 in *Monograph on Statistics and Applied Probability*.

Mercken, L., Snijders, T. A., Steglich, C., Vartiainen, E., & De Vries, H. (2010). Dynamics of adolescent friendship networks and smoking behavior. *Social Networks, 32*(1), 72-81.

Mercken, L., Steglich, C., Sinclair, P., Holliday, J., & Moore, L. (2012). A longitudinal social network analysis of peer influence, peer selection, and smoking behavior among adolescents in British schools. *Health Psychology, 31*(4), 450.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525-543.

Millsap, R. E. (2012). *Statistical approaches to measurement invariance.* Routledge.

Muthén, B., & Masyn, K. (2005). Discrete-time survival mixture analysis. *Journal of Educational and Behavioral statistics, 30*(1), 27-58.

Osterlind, S. J., & Everson, H. T. (2009). *Differential item functioning* (Vol. 161). Sage Publications.

Prinstein M. J. & Giletta, M. (2016). Peer relations and developmental psychopathology. In D. Cicchetti (Ed.), *Developmental psychopathology* (3rd edition). Hoboken, NJ: Wiley.

Raykov, T., Gorelick, P. B., Zajacova, A., & Marcoulides, G. A. (2017). On the Potential of Discrete Time Survival Analysis Using Latent Variable Modeling: An Application to the Study of the Vascular Depression Hypothesis*. Structural Equation Modeling: A Multidisciplinary Journal*, 1-10.

Ripley, R. M., Snijders, T. A. B., Boda, Z., Vörös, A., & Preciado, P. (2017). *Manual for SIENA version 4.0* (version February 16, 2017). Department of Statistics, Nuffield College: Oxford: University of Oxford.

Robins, G., Pattison, P., Kalish, Y., & Lusher, D. (2007). An introduction to exponential random graph (p\*) models for social networks. *Social networks, 29*(2), 173-191.

Saules, K. K., Pomerleau, C. S., Snedecor, S. M., Mehringer, A. M., Shadle, M. B., Kurth, C., & Krahn, D. D. (2004). Relationship of onset of cigarette smoking during college to alcohol use, dieting concerns, and depressed mood: results from the Young Women's Health Survey. *Addictive behaviors, 29*(5), 893-899.

Singer, J. D., & Willett, J. B. (1993). It’s about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of educational statistics, 18*(2), 155-195.

Smith, J. A., & Moody, J. (2013). Structural effects of network sampling coverage I: Nodes missing at random. *Social networks, 35*(4), 652-668.

Smith, J. A., Moody, J., & Morgan, J. H. (2017). Network sampling coverage II: the effect of non-random missing data on network measurement. *Social networks, 48*, 78-99.

Snijders, T. A., Van de Bunt, G. G., & Steglich, C. E. (2010). Introduction to stochastic actor-based models for network dynamics. *Social networks, 32*(1), 44-60.

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological methodology, 13*, 290-312.

Snijders, T. A. (1996). Stochastic actor‐oriented models for network change. *Journal of mathematical sociology, 21*(1-2), 149-172.

Steglich, C., Snijders, T. A., & Pearson, M. (2010). Dynamic networks and behavior: Separating selection from influence. *Sociological methodology, 40*(1), 329-393.

Takane, Y., & De Leeuw, J. (1987). On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika, 52*(3), 393-408.

Thissen, D. (2001). Software for the computation of the statistics involved in item response theory likelihood-ratio tests for differential item functioning. Chapel Hill: University of North Carolina.

Thissen, D., Steinberg, L., & Kuang, D. (2002). Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *Journal of Educational and Behavioral Statistics, 27*(1), 77-83.

Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models.

Valente, T. W., Unger, J. B., & Johnson, C. A. (2005). Do popular students smoke? The association between popularity and smoking among middle school students. *Journal of Adolescent Health, 37*(4), 323-329.

Veenstra, R., Dijkstra, J. K., Steglich, C., & Van Zalk, M. H. (2013). Network–behavior dynamics. *Journal of Research on Adolescence, 23*(3), 399-412.

Wang, D. J., Shi, X., McFarland, D. A., & Leskovec, J. (2012). Measurement error in network data: A re-classification. *Social Networks, 34*(4), 396-409.

Willett, J. B., & Singer, J. D. (1993). Investigating onset, cessation, relapse, and recovery: why you should, and how you can, use discrete-time survival analysis to examine event occurrence. *Journal of consulting and clinical psychology, 61*(6), 95.

Table 1. Summary of all node-level indices used in the analyses.

|  |  |  |  |
| --- | --- | --- | --- |
| **Node-level index** | **Factor measured** | **Definition** | **Response distribution** |
| Indegree | Social status | Tendency to be popular: the number of in-school alters who nominate ego as friend (relative to the number of possible nominations). | Poisson |
| Bonacich centrality | Social status | Tendency to have popular friends: the extent to which an adolescent has many friends who themselves have many friends.  | Binary |
| Betweenness centrality | Social status | Tendency to connect adolescents who are not directly linked by a friendship tie: the proportion of all shortest social distance paths that include ego.  | Three-level ordinal |
| Three-step reach centrality | Social status | Tendency to be in close social distance to others: the proportion of the network that can reach ego in three ties or less.  | Binary |
| Outdegree | Social integration | Tendency to choose friends: the number of in-school friendship nominations made by ego (up to 5).  | Three-level ordinal |
| Reciprocity | Social integration | Tendency to have reciprocated friendships: the number of ego’s friendship nominations reciprocated by alter.  | Poisson |
| Transitive triads | Social integration | Tendency to be a friend of a friend’s friend: the number of all triads containing ego where an alter’s friend is a friend of ego.  | Three-level ordinal |
| Intransitive triads | Social integration | Conceptual complement to transitive triads: the number of all triads containing ego where an alter’s friend is not a friend of ego. | Three-level ordinal |
| Out of network friends | Social integration | Tendency to have friends not in the peer network: the number of nominations to friends not in the network. | Poisson |

Table 2. Descriptive statistics of factor scores across schools.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|   | **Social Status** |   | **Social Integration** |   | ***R*(Social Status, Social Integration)** |
|   | Mean | SD |   | Correlations with Indicators |   | Mean | SD |   | Correlations with Indicators |   |
|   |   | Bonacich Centrality | Three-Step Reach | Indegree | Normed Betweenness |   |   | Outdegree | Out-of-school friends  | Reciprocity | Transitive Triads | Intransitive Triads |   |
| **School A** | 0.066 | 0.824 |   | 0.563 | 0.710 | 0.781 | 0.644 |   | 0.024 | 0.734 |   | 0.483 | 0.415 | 0.601 | 0.489 | 0.804 |   | 0.650 |
| **School B** | 0.261 | 0.849 |   | 0.464 | 0.789 | 0.849 | 0.658 |   | 0.119 | 0.714 |   | 0.717 | 0.599 | 0.590 | 0.501 | 0.815 |   | 0.746 |
| **School C** | -0.318 | 0.861 |   | 0.375 | 0.715 | 0.826 | 0.609 |   | 0.538 | 0.799 |   | 0.573 | 0.439 | 0.548 | 0.567 | 0.772 |   | 0.702 |
| **School D** | 0.017 | 0.818 |   | 0.375 | 0.697 | 0.837 | 0.595 |   | 0.029 | 0.738 |   | 0.613 | 0.486 | 0.641 | 0.479 | 0.798 |   | 0.751 |
| **School E** | 0.252 | 0.928 |   | 0.540 | 0.778 | 0.798 | 0.643 |   | 0.106 | 0.776 |   | 0.731 | 0.554 | 0.596 | 0.547 | 0.776 |   | 0.686 |
| **School F** | -0.134 | 0.914 |   | 0.508 | 0.771 | 0.800 | 0.671 |   | 0.073 | 0.885 |   | 0.676 | 0.513 | 0.533 | 0.567 | 0.826 |   | 0.742 |

Table 3. Abridged impact and DIF results for all MNLFA models fitted.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|   |   |   | **Social Integration** |   | **Social Status** |
|   |   |   | Mean impact |   | Item-specific DIF effects |   | Mean impact |   | Item-specific DIF effects |
| **Covariates** |   |   | Outdegree | Out-of-school Friends | Reciprocity | Transitive Triads | Intransitive Triads |   |   | Bonacich Centrality | Three-step Reach | Indegree | Normed Betweenness |
| *Main effects* |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|   | Black |   | C |   |   |   |   |   |   |   | BDE |   |   |   |   |   |
|   | Cohort 1 |   | AD |   |   |   |   |   |   |   |   |   |   |   |   |   |
|   | Cohort 2 |   | CD |   |   |   |   |   |   |   | CD |   |   |   |   |   |
|   | Grade |   | ACF |   |   |   | E | A | E |   | BCE |   | B |   | A |   |
|   | High Parental Education |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|   | In High School |   |   |   |   |   |   |   |   |   | F |   |   |   |   |   |
|   | Male |   | A |   |   |   | ABCDF | AC |   |   | EF |   |   | F |   |   |
|   | Low Parental Education |   | BDEF |   |   |   |   |   |   |   | CF |   |   |   |   |   |
|   | Network Size |   |   |   |   |   |   | *ABCDEF* | *ABCDEF* |   | C |   |   | D | *ABCDEF* |   |
| *Interactions* |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|   | Black\*Grade |   |   |   |   |   |   |   |   |   | B |   |   |   |   |   |
|   | Cohort 1\*Grade |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|   | Cohort 2\*Grade |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|   | High Parental Education\*Grade |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|   | Male\*Grade |   |   |   |   |   |   | A |   |   | E |   |   |   |   |   |
|   | Low Parental Education\*Grade |   |   |   |   |   |   |   |   |   | BC |   |   |   |   |   |
|   | Network Size\*Grade |   | F |   |   |   |   | E |   |   |   |   |   |   | A |   |

*Note.* The letter codes used above are as follows. A = Effect found in School A; B = Effect found in School B; C = Effect found in School C; D = Effect found in School D; E = Effect found in School F; School E; F = Effect found in School F. Effects included as defaults (i.e., effects of network size on transitive ties, intransitive ties, and indegree) are shown in italics.

Table 4. Parameter estimates from all discrete-time survival models.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|   |   |   |   | Model 1 |   | Model 2 |   | Model 3a |   | Model 3b |
| **Model Fit** |  |  |  |  |  |  |  |  |  |  |  |
|  |  | Num.parameters | 11 |  | 23 |  | 26 |  | 33 |
|  |  | Loglikelihood | -10176.790 |  | -8164.790 |  | -7758.55 |  | -7747.598 |
|  |  | AIC | 20375.580 |  | 16375.580 |  | 15569.100 |  | 15561.196 |
|  |  | BIC | 20449.857 |  | 16530.882 |  | 15744.639 |  | 15783.989 |
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| **Model Parameters** |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | β | SE(β) |  | β | SE(β) |  | β | SE(β) |  | β | SE(β) |
| **Baseline Hazard Logit** |  |  |  |  |  |  |  |  |  |  |
|  |  | Spring 6th | -2.316\*\* | 0.053 |  | 0.121 | 0.095 |  | -0.735\*\* | 0.105 |  | 0.270 | 0.153 |
|  |  | Fall 7th | -2.527\*\* | 0.061 |  | 0.055 | 0.101 |  | -0.713\*\* | 0.108 |  | 0.292 | 0.156 |
|  |  | Spring 7th | -1.552\*\* | 0.041 |  | -0.046 | 0.076 |  | -0.798\*\* | 0.084 |  | 0.207 | 0.141 |
|  |  | Fall 8th | -1.930\*\* | 0.053 |  | -0.325\*\* | 0.084 |  | -1.049\*\* | 0.091 |  | -0.045 | 0.146 |
|  |  | Spring 8th | -1.012\*\* | 0.038 |  | -0.485\*\* | 0.054 |  | -1.184\*\* | 0.063 |  | -0.208 | 0.13 |
|  |  | Fall 9th | -1.202\*\* | 0.051 |  | -1.015\*\* | 0.079 |  | -1.653\*\* | 0.084 |  | -0.709\*\* | 0.142 |
|  |  | Spring 9th | -1.929\*\* | 0.077 |  | -1.771\*\* | 0.095 |  | -2.383\*\* | 0.099 |  | -1.433\*\* | 0.152 |
|  |  | Fall 9th | -1.493\*\* | 0.078 |  | -1.836\*\* | 0.11 |  | -2.405\*\* | 0.112 |  | -1.538\*\* | 0.161 |
|  |  | Spring 10th | -2.351\*\* | 0.145 |  | -2.603\*\* | 0.165 |  | -3.138\*\* | 0.167 |  | -2.243\*\* | 0.204 |
|  |  | Fall 11th | -1.400\*\* | 0.13 |  | -2.086\*\* | 0.163 |  | -2.665\*\* | 0.164 |  | -1.809\*\* | 0.203 |
|  |  | Fall 12th | -1.544\*\* | 0.235 |  | -2.644\*\* | 0.254 |  | -3.206\*\* | 0.255 |  | -2.369\*\* | 0.284 |
| **Predictor Effects** |  |  |  |  |  |  |  |  |  |  |  |
|  | **Methodological Control Variables** |  |  |  |  |  |  |  |  |  |
|  |  | Cohort 1  |  |  |  | -1.658\*\* | 0.076 |  | -1.221\*\* | 0.078 |  | -1.308\*\* | 0.082 |
|  |  | Cohort 2 |  |  |  | -0.916\*\* | 0.062 |  | -0.685\*\* | 0.064 |  | -0.752\*\* | 0.065 |
|  |  | School A |  |  |  | -0.065 | 0.038 |  | -0.113\*\* | 0.041 |  | -0.016 | 0.042 |
|  |  | School B |  |  |  | -0.121\*\* | 0.048 |  | -0.205\*\* | 0.052 |  | -0.065 | 0.052 |
|  |  | School C |  |  |  | 0.044 | 0.066 |  | 0.532\*\* | 0.077 |  | 0.038 | 0.094 |
|  |  | School E |  |  |  | 0.295\*\* | 0.048 |  | -0.075 | 0.053 |  | 0.183\*\* | 0.055 |
|  |  | School F |  |  |  | 0.219\*\* | 0.046 |  | 0.246\*\* | 0.048 |  | 0.257\*\* | 0.051 |
|  |  | Network Size |  |  |  | 1.030\*\* | 0.085 |  | 0.982\*\* | 0.077 |  | 1.022\*\* | 0.084 |
|  | **Demographic Control Variables** |  |  |  |  |  |  |  |  |  |
|  |  | Black |  |  |  | 0.043 | 0.022 |  | 0.007 | 0.023 |  | 0.021 | 0.023 |
|  |  | Male |  |  |  | 0.028 | 0.019 |  | 0.000 | 0.02 |  | -0.032 | 0.021 |
|  |  | High Parental Ed. |  |  |  | -0.079\*\* | 0.029 |  | -0.213\*\* | 0.031 |  | -0.225\*\* | 0.031 |
|  |  | Low Parental Ed. |  |  |  | 0.197\*\* | 0.021 |  | 0.24\*\* | 0.022 |  | 0.220\*\* | 0.022 |
|  | **Social Dynamics and Deviance** |  |  |  |  |  |  |  |  |  |  |
|  |  | Deviance |  |  |  |  |  |  | 0.508\*\* | 0.021 |  | 0.497\*\* | 0.021 |
|  |  | Social Status |  |  |  |  |  |  | 0.516\*\* | 0.049 |  |  |  |
|  |  |  | Indegree |  |  |  |  |  |  |  |  |  | 0.029 | 0.026 |
|  |  |  | Normed Betweenness |  |  |  |  |  |  |  |  | 0.044 | 0.028 |
|  |  |  | Bonacich Centrality |  |  |  |  |  |  |  |  | -0.571\*\* | 0.141 |
|  |  |  | Three-step In-Reach |  |  |  |  |  |  |  |  | 2.373\*\* | 0.65 |
|  |  | Social Integration |  |  |  |  |  |  | -0.240\*\* | 0.054 |  |  |  |
|  |  |  | Outdegree |  |  |  |  |  |  |  |  |  | 0.110\*\* | 0.051 |
|  |  |  | Reciprocity |  |  |  |  |  |  |  |  |  | -0.069\*\* | 0.029 |
|  |  |  | Out-of-school Friends (reversed) |  |  |  |  |  |  |  | -0.212\*\* | 0.039 |
|  |  |  | Transitive Triads |  |  |  |  |  |  |  |  |  | -0.041\*\* | 0.019 |
|   |   |   | Intransitive Triads |   |   |   |   |   |   |   |   |   | 0.001 | 0.005 |

Figure 1. Moderated nonlinear factor analysis with impact and DIF.



Figure 2. Predicted inverse survival curves for subjects of high and low social integration and status.



Figure 3. Results from the discrete-time survival mediation model.



*Note:* Parameter effects represented by bold paths were significant at the  level; they are shown with their standard errors in parentheses. Schools A, B, C, E, and F are compared to School D, which was the largest school. Smoking onset hazard is parameterized as a latent variable which is related to all smoking onset indicators by constraining their loadings to 1 (see Muthen and Masyn, 2005 for more details). The numbers to the right of each indicator are the log-odds of smoking onset in that time period.

1. This procedure is automated in a new R package, aMNLFA, which interfaces with Mplus to conduct all of these steps (Gottfredson, 2018). However, this package was not yet available at the time of the current analyses, and thus all analyses were conducted in SAS Version 9.3. [↑](#footnote-ref-1)