Supplementary Material

The reference list includes background papers on the first order Generalized Estimating Equations (GEE1) as well as the Alternating Logistic Regressions (ALR) and other second-order multi-level GEE2 models, previously used to estimate predictive and explanatory relationships linking individual-level and group-level covariates in epidemiological research on villages and neighborhoods (e.g., see the cited works by Zeger et al., 1988; Liang & Zeger, 1990; Katz et al., 1993; Bobashev & Anthony, 2000; Petronis & Anthony, 2000; 2003). Prior GEE/ALR work on drug use clusters within schools based on nationally representative school survey data also has been published, some of it within *Drug and Alcohol Dependence* (Delva et al., 2000).

Within the GEE2 ALR modeling framework used in this research project, the individual student is the lowest level of analysis, and the age, sex, and race-ethnicity of the individual students become covariates included for control or statistical adjustment purposes while estimating the ALR slope parameter that links school-level cannabis risk perception at 't-1' as a covariate with a response variable framed in relation to the odds of being observed as a newly incident cannabis user at 't'. Within the GEE2 ALR framework, the pairwise odds ratio (PWOR) is derived from a specific parameter in the equation for the conditional expectation of cannabis smoking for a 12th-grader, conditioning on occurrences of newly incident cannabis smoking in other 12th-graders drawn from the same school or from different schools. Because only one level of clustering was of interest in this study, the equation's α parameter only has one value (Bobashev and Anthony, 2000). The predictive association between pairs of 12th-graders was modeled as follows:

$$\log\left(PWOR_{jk}\right) = \alpha_0 Z_{0jk}$$

The logarithm of the PWOR is a function of an indicator variable that expresses whether a pair of 12th-graders, *j* and *k*, are in the same school. Z_{jk} is a binary variable that takes value 1 when the pair belongs to the same school and takes value 0 otherwise (Bobashev and Anthony, 1998).

Concurrent with ALR estimation of the PWOR clustering parameter, there is estimation of the degree to which the binary response variable for newly incident cannabis use might depend upon either the individual-level or group-level covariates. In this approach, a logistic regression is iteratively used to control for potential school and student level variables:

$$\log(ODDS_j) = \beta_0 + \sum \beta_i X_i$$

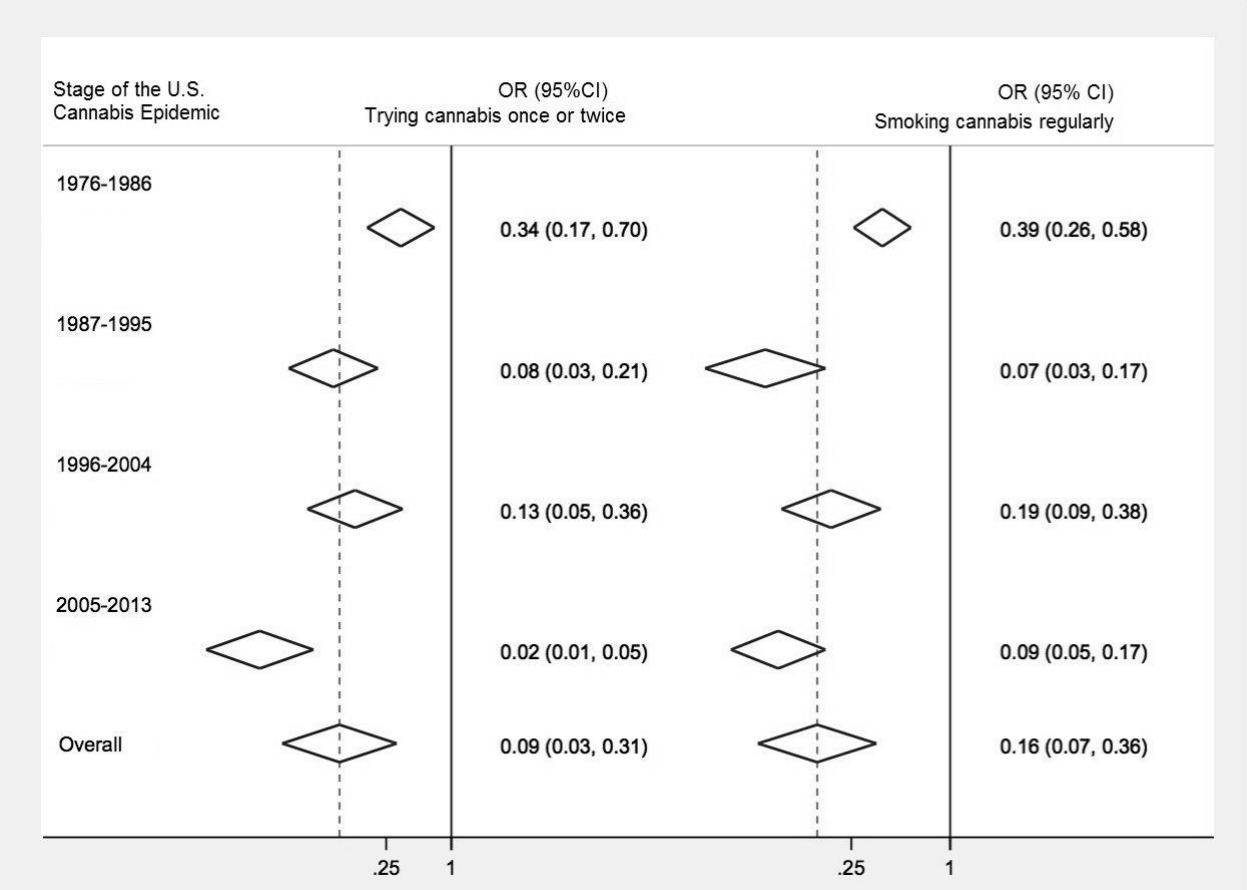
where X_is are the covariates. The βs yield odds ratio estimates linking covariate values with incidence odds of cannabis onsets. Estimation was via SAS version 9.4 'PROC GENMOD' with the LOGOR option on the REPEATED statement used to estimate PWORs, regression slope coefficients, and robust standard errors for each year (SAS Institute Inc., 1999).

In running the ALR, first an intercept only model was fitted to estimate associations of newly incident cannabis smoking within schools for each of the 38 years, 1976-2013. Because ALR produces PWOR estimates and also accommodates covariate adjustments, it was suspected that perceived risk might account for the odds of becoming a newly incident cannabis smoker and/or might account for the magnitude of clustering. Therefore, this suspected causal determinant of central interest was included next. By comparing the unadjusted and adjusted estimates of the PWOR after adding covariates, the risk effect could be estimated (Liang and Zeger, 1993). Covariate terms for school-level perceived risk of trying once or twice and then perceived risk of regular cannabis smoking were introduced to the ALR model. With the sample restricted to schools that were surveyed two years in a row, the proportion of students who

perceived great risk of cannabis use (either trying or regular smoking), as measured in the school at ‘t-1’ provided a school-level covariate to estimate odds of becoming a newly incident cannabis user at ‘t’.

A Prospective Study of Newly Incident Cannabis Use and Cannabis Risk Perceptions: Results from the United States Monitoring the Future Study, 1976-2013

Supplementary Figure 1. Estimated stage-specific meta-analysis summary odds ratios (OR) and 95% confidence intervals (CI) for the predictive association from a school-level 'great risk' perception about trying cannabis once or twice to the school-level odds of becoming a newly incident cannabis user (left column) and for the corresponding prediction based on 'great risk' of smoking cannabis regularly (right column), based on unadjusted Alternative Logistic Regressions. Data from "Monitoring the Future: Secondary School Students," United States, 1976-2013.



A Prospective Study of Newly Incident Cannabis Use and Cannabis Risk Perceptions: Results from the United States Monitoring the Future Study, 1976-2013

Supplementary Figure 2. Estimated stage-specific meta-analysis summary pairwise odds ratios (PWOR) and 95% confidence intervals (CI) for school-level clustering of newly incident cannabis use among 12th-graders, based on ALR models for 'great risk' perceptions about trying cannabis once or twice (left) and smoking cannabis regularly (right), without other covariates. Data from "Monitoring the Future: Secondary School Students," United States, 1976-2013.