An Empirical Evaluation of Alternative Approaches to Adjusting for Attrition When Analyzing Longitudinal Survey Data on Young Adults' Substance Use Trajectories

Running title: Longitudinal Attrition Adjustments

Yajuan Si, Brady T. West, Philip Veliz, Megan E. Patrick, John E. Schulenberg, Deborah D. Kloska, Yvonne M. Terry-McElrath, and Sean E. McCabe

Author affiliations: Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor (Yajuan Si, Brady T. West, Megan E. Patrick, John E. Schulenberg, Deborah D. Kloska, Yvonne M. Terry-McElrath); Department of Systems, Populations and Leadership, School of Nursing, University of Michigan, Ann Arbor (Philip Veliz); Center for the Study of Drugs, Alcohol, Smoking and Health, Department of Health Behavior and Biological Sciences, School of Nursing, University of Michigan, Ann Arbor (Sean E. McCabe)

Corresponding Author: Yajuan Si, PhD, Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor. ISR 4014, 426 Thompson St, Ann Arbor, MI 40104 (yajuan@umich.edu; 734-7646935).

Acknowledgement: The work was supported by research grants R01DA001411, R01DA016575, and R01DA031160 from the National Institute on Drug Abuse, National Institutes of Health.

Abstract

Objectives: Longitudinal survey data allow for the estimation of developmental trajectories of substance use from adolescence to young adulthood, but these estimates may be subject to attrition bias. Moreover, there is a lack of consensus regarding the most effective statistical methodology to adjust for sample selection and attrition bias when estimating these trajectories. Our objective is to develop specific recommendations regarding adjustment approaches for attrition in longitudinal surveys in practice.

Methods: Analyzing data from the national U.S. Monitoring the Future panel study following four cohorts of individuals from modal ages 18 to 29/30, we systematically compare alternative approaches to analyzing longitudinal data with a wide range of substance use outcomes, and examine the sensitivity of inferences regarding substance use prevalence and trajectories as a function of college attendance to the approach used.

Results: Our results show that analyzing all available observations in each wave, while simultaneously accounting for the correlations among repeated observations, sample selection, and attrition, is the most effective approach. The adjustment effects are pronounced in wave-specific descriptive estimates but generally modest in covariate-adjusted trajectory modeling.

Conclusions: The adjustments can refine the precision, and, to some extent, the implications of our findings regarding young adult substance use trajectories.

Keywords: Longitudinal trajectory modeling; substance use; selection bias; attrition; weighting.

Word counts: abstract (200), main text (4966).

INTRODUCTION

Substance use becomes more common in adolescence and typically peaks in young adulthood (Jager et al., 2013; McCabe et al., 2016, 2019; Schulenberg & Maggs, 2002; Schulenberg et al., 2005). Longitudinal surveys collect rich data about individual characteristics and enable the estimation of substance use prevalence and trajectory modeling, a key approach to understanding the developmental course and etiology of substance use. Examples include the Monitoring the Future (MTF) panel study (Schulenberg et al., 2021), the Population Assessment of Tobacco and Health Study (Hyland et al., 2017), and the National Longitudinal Study of Adolescent to Adult Health (Harris et al., 2019). However, the quality of prevalence and trajectory estimates can be attenuated by panel attrition. With declining response rates in surveys (Brick & Williams, 2013; de Leeuw, Hox, and Luiten, 2018), study respondents could be systematically different from attriters in terms of substance use outcomes, causing potential bias in estimates of developmental trajectories (Feldman & Rabe-Hesketh, 2012). Moreover, sample selection procedures for longitudinal studies are often complex in nature, including design features such as stratification, cluster sampling, and survey weights for probability samples and auxiliary variables that affect selection and response propensities for nonprobability samples. A failure to account for these selection features in estimation could affect the inferential validity and generalizability of descriptive summary measures and estimates of trajectory models.

Weighting approaches have been proposed in the survey statistics literature to simultaneously adjust for these complex sample design features and panel attrition (Heeringa, West, & Berglund, 2017). However, there is no clear consensus in the literature on whether or how to apply weighting adjustments for attrition in longitudinal trajectory modeling. The role of weighting adjustments in regression models has been a long-debated topic (Bollen et al., 2016). Longitudinal trajectory estimation introduces methodological challenges and may require wave-specific weighting adjustments. Alternative to weighting, multiple imputation (MI; Rubin, 1987) allows for the inclusion of variables that can be incomplete into the imputation model but is subject to computational burdens. The common practice is to use weighting adjustment for attrition and MI for item nonresponse when individuals only answer partial questions (Si et al., 2022a, 2022b).

Of the prior studies that have examined long-term trajectories of alcohol, marijuana, and prescription drug misuse during the transition from adolescence to young adulthood, analyses have generally been restricted to respondents providing complete information in all waves and ignored attrition (e.g., McCabe, Veliz, & Schulenberg, 2018). More recent studies focusing on long-term substance use trajectories have addressed attrition, typically using inverse propensity score weighting (McCabe et al., 2019; Patrick et al., 2016, 2018; Terry-McElrath et al., 2017). Using the MTF panel data, Keyes et al. (2020) estimated bias by imputing nonrespondents' outcomes. Nevertheless, appropriate statistical approaches to adjusting for attrition when estimating longitudinal trajectories with a model adjusting for various risk factors have received relatively little focus, especially for subgroups of particular interest to substance use researchers.

Various sociodemographic characteristics are known to be associated with substance use (e.g., Roghani, Nyarko, & Potter, 2021), and there are differences in substance use behaviors among young adults as a function of college attendance (e.g., Schulenberg et al., 2021). In general, binge drinking and non-medical misuse of prescription stimulants have higher prevalence among college students than non-college young adults, although levels have converged some in recent years. In contrast, non-college young adults tend to have higher prevalence of cigarette use, daily cannabis use, non-medical misuse of prescription sedatives/tranquilizers, and other illicit drugs, including heroin and methamphetamine, than those attending college.

Research has also focused on how key sociodemographic differences like educational attainment alter the developmental course of substance use with longitudinal data (e.g., Linden-Carmichael, et al., 2019), and some of these studies include various remedies to account for differential attrition. However, attrition still remains a strong concern with such panel studies, given that some groups with different substance use patterns (e.g., non-college attenders) are more likely to drop out of these long-term studies and may ultimately bias the results when assessing these key sociodemographic differences. This paper contributes to the literature by developing guidelines on the use of different attrition adjustment methods when analyzing longitudinal survey data.

The multi-cohort MTF panel study of teens and adults offers a unique data source that can be used to evaluate attrition effects on longitudinal trajectory modeling, when accounting for diverse socio-demographics and examining specific substance use outcomes of interest. Inverse propensity score weighting procedures can address attrition and have been used in many recent MTF substance use trajectory publications (Patrick et al., 2016, 2021; Terry-McElrath & Patrick, 2016a, 2016b; Terry-McElrath et al., 2019).

Using seven waves of longitudinal data from the MTF panel study as an example, we seek to 1) perform a systematic comparison of alternative approaches to analyzing longitudinal survey data with a wide range of substance use outcomes, and 2) examine the sensitivity of inferences about differences in estimated substance use prevalence and trajectories to the approaches used, focusing on trajectory differences as a function of college attendance. Our methodological examination focuses on three aspects:

1. Whether to use the complete cases (CCs) that respond to all waves (including those who do not answer all questions) or the available cases (ACs) that respond in any wave.
2. Comparing approaches to accounting for the correlation of repeated measures on the same individual.
3. Whether to apply weighting adjustments for attrition.

[Table 1 about here]

We investigate eight different methods based on these aspects, shown in Table 1. Appropriate decisions depend on the underlying missing data mechanism and the trajectory model specification (Little & Rubin, 2019). We evaluate the effects of the different adjustments on estimated prevalence and trajectories, and corresponding substantive inferences.

METHODS

Data

The MTF study began in 1975 and has been an ongoing epidemiological and etiological research project to study changes in the beliefs, attitudes, and behaviors of U.S. young people regarding substance use and other health risks (Miech et al., 2021; Schulenberg et al., 2021). Each year, about 15,000 12th grade students in approximately 133 public and private high schools nationwide participate in the MTF study. The data from students are collected with a multi-stage random sampling procedure, designed to secure new nationwide samples of 12th grade students each year. Beginning with the class cohort of 1976, a subsample from each 12th grade class (modal age 18) was followed up after high school on a continuing basis, with oversampling of students who report drug use.

The subsample selected for the panel study is a sample of U.S. individuals with modal age 18 who provided their sex and contact information. There were important features that affected the sample selection (e.g., drug use reporting, sex, and geographical strata of schools). One random half of each cohort began follow-up assessments one year after high school (modal age [hereafter referred to simply as "age"] 19) and the other random half two years after high school (age 20), with all being followed every two years through age 29/30; in this study, the two random halves were combined (e.g., follow-up 1 includes ages 19/20). The longitudinal follow-ups permitted examination of developmental changes within cohorts. We analyzed the MTF panel data to examine trajectories of substance use from late adolescence through young adulthood. We focused on four cohorts whose baseline data were collected from 2002 to 2005 (age 29/30 data collected in 2013-2017) to allow for six possible follow-up waves for any participant (i.e., first follow-up at 19/20 sixth follow-up at 29/30). We chose these specific cohorts because they were the first to include the latest updates to the wording of existing questions about specific substance use (i.e., non-medical misuse of prescription medications).

Measures

Our goal was to estimate the trajectories of substance use, adjusting for key risk factors for different subgroups defined by college attendance status. By virtue of the sampling design, at baseline respondents were all in high school. We began the trajectory modeling at the first follow-up of age 19/20. The outcomes were longitudinal measures of the following substance use behaviors: binge drinking (five or more drinks in a row during the past 2 weeks), cigarette smoking (in the past 30 days), marijuana use (in the past 12 months), non-medical prescription opioid misuse (in the past 12 months), a composite indicator of any non-medical prescription drug (NMPD) misuse (of four specific drug classes that included amphetamines, sedatives, tranquilizers, or opioids, in the past 12 months), and a composite indicator of any use of other selected illicit drugs, including LSD, other hallucinogens, cocaine, or heroin (in the past 12 months). The time-varying covariates included follow-up wave indicators, full-time four-year college attendance, and marital status. The time-invariant covariates were baseline characteristics including cohort (2002-2005), age in months, sex, race/ethnicity, high school grades, parental education, baseline measures of corresponding substance outcomes, and reported intent to attend a four-year college. The covariates were selected based on the substance use literature where the conditional interpretations adjusting other variables will be substantively meaningful. The candidate covariates in the attrition adjustment included all selection features and baseline characteristics related to the substance use outcomes. The full lists are available in eTables 1 and

2. Most of these measures are subject to item nonresponse, and our strategies for handling missing values are discussed in the eAppendix. We found that our findings are robust under different imputation methods, including MI, mainly due to the low rates of item nonresponse.

Substance Use Trajectory Modeling

The developmental course of substance use can be affected or moderated by sociodemographic characteristics. Across six follow-up wave indicators defined by t_{ij} *(=1, 2, 3, 4, 5, 6),* we considered marginal (or population-averaged) logistic regression models for each of these binary outcome variables Y, where $Y_{ij} = 1$ if individual i indicated use of substances at measurement j, and otherwise, $Y_{ij} = 0$:

$$logit\left(Pr\left(Y_{ij} = 1\right)\right) = \beta_0 + \vec{\beta}_1 t_{ij} + \vec{\beta}_2 x_i + \beta_3 coll_{ij} + \beta_4 marry_{ij} + \vec{\beta}_5 t_{ij} * coll_{ij}. \quad (1)$$

Here the marginal model includes the primary coefficients of interest that identify possible differences in trajectories based on college status: the coefficients $\vec{\beta}_5$ for the interaction $t_{ij} * coll_{ij}$ between college attendance $(coll_{ij})$ and wave (t_{ij}). Treating the first follow-up wave (age 19/20) as the reference level, we introduced five dummy variables for the wave indicators, and the quantities of interest were the odds ratios, $(exp(\vec{\beta}_1), exp(\vec{\beta}_5))$, which were exponentials of the coefficients as five-dimensional vectors. The model adjusted for a vector of time-invariant individual-level measures x_i described above and a time-varying indicator of being married $(marry_{ij})$.

We compared different estimation approaches and evaluated their effects on the trajectory modeling of substance use outcomes. We first fit simple logistic regression models ignoring the correlation of the repeated measures within a sampled student. For both CCs and ACs, we considered unweighted analyses, unweighted analyses accounting for the correlation of the repeated observations on each individual, and weighted analyses accounting for attrition adjustments and individual clustering. Due to computational challenges in achieving estimation convergence using existing software, we do not explicitly consider multilevel modeling as an alternative subject-specific trajectory modeling approach (e.g., Heeringa, West, & Berglund, 2017) in this study.

Attrition Adjustment

We use the baseline sample as the benchmark and adjust for attrition. For both CC and AC weight construction, we treated the response indicators as binary outcomes and considered two approaches to predicting the probability of response: classification trees based on recursive partitioning (Breiman et al., 1984) and classical logistic regression models. The tree-based approach selects variables and their higher-order interactions through splitting rules that sequentially maximize predictive performance for the overall decision tree. The predictive performance of the classification tree depends on balancing the tree size and the true error rate based on a new test dataset, the goal of which is to avoid overfitting the training data used to construct the tree. We applied a conditional tree method that automatically stops splitting based on hypothesis tests and eliminates the pruning step (Hothorn, Hornik, & Zeileis, 2006).

The logistic regression model includes only main effects of the same set of covariates, enabling an assessment of whether any additional higher-order interactions in the tree-based approach improves our overall ability to predict response propensity. We performed forward selection techniques to select the best predictors in the logistic regression model. The candidate covariates in the attrition adjustment included all selection features and baseline characteristics related to the substance use outcomes, including the cohort indicator, the MTF sample stratification code, the oversampling indicator for 12th grade drug users for the panel, school type, age, sex, race/ethnicity, marital status, family structure, parental education, future plans after high school graduation (military or technical schools, etc.), frequencies of missing class due to different reasons, working status, weekly pay amounts from jobs and other sources, binge drinking in the past two weeks, use of various substances in the past 30 days/12 months/lifetime, and a host of variables concerning beliefs, high school activities and performance, and other problematic behaviors at baseline. The attrition adjustment will be effective if the selected covariates are strongly related to the aforementioned substance use outcomes.

For probability samples, the base weights adjust for the sample composition at baseline to match the target student population, and the attrition-adjusted weights are constructed by multiplying the base weights by the inverses of predicted probabilities of participating at a given wave (or for all follow-up waves). Here, we treated the MTF panel study as a quasi-probability sample by assigning base weights as 1 and using the inverse response propensity scores at each wave as the attrition-adjusted weights. In CC analyses, we created an indicator of whether the individual had participated in all six follow-up waves, and used the inverse of the predicted response propensities p_{cc} to construct the attrition-adjusted CC weight: $w_{cc} = \frac{1}{p_{cc}}$. In AC analyses, we created wave-specific response indicators of whether the individual has responded in one particular wave and constructed multiple attrition-adjusted AC weights based on the inverses of predicted AC response propensities p_{ac} as the attrition-adjusted AC weights: $w_{ac} = \frac{1}{p_{ac}}$, to fully utilize all available observations.

These adjustment approaches make different assumptions about the underlying response mechanism. The unweighted analysis assumes that the attrition results in data that are missing completely at random, and the attrition-adjusted weighting analysis assumes that the attrition results in data that are missing at random conditional on the baseline characteristics. The CC and AC analyses thus differ in terms of whether the response mechanisms and the variables that predict response propensities change across the follow-up waves.

Accounting for Clustering

To account for the correlation within each individual, we applied two methods: (1) a design-based approach, treating individuals as clusters in all of the weighted analyses incorporating the attrition adjustments and applying Taylor Series Linearization for variance estimation (Binder, 1983); and (2) generalized estimating equations (GEE) accounting for the clustering structure due to multiple responses per person with an exchangeable working correlation matrix specification. For the purpose of our comparison, we chose the exchangeable working correlation here to approximate the design-based analysis. We note that inferences related to the fixed coefficients in models fitted using GEE are generally robust against the possible misspecification of this working correlation structure, and model diagnostics and goodness of fit measures should be used to inform model selection in practice (Liang & Zeger, 1986).

We implemented tree-based methods using the contributed R package *party* (Hothorn et al., 2021) and fitted GEE models with the R package *geepack* (Halekoh, Højsgaard, & Yan, 2006). We accounted for the clustering via the individual numeric identifiers (MTF ID) and the (possibly adjusted) weights in the logistic regression models for the longitudinal substance use outcomes via the R package *survey* (Lumley, 2020), which utilizes weighted estimating equations assuming an exchangeable correlation structure within clusters and Taylor Series Linearization for variance estimation. We performed model diagnostics and evaluated prediction performance of the tree models based on the area under the Receiver Operating Characteristic curve (AUC), an aggregate measure of performance across all possible classification thresholds. We implemented MI via chained equations to handle item nonresponse with the R package *mice* (Van Buuren & Oudshoorn, 1999) with details given in the eAppendix. All analyses were performed in R 4.0.3 (R Core Team, 2020).

RESULTS

The MTF study collected baseline data from 9,802 sampled individuals in the 2002-2005 12th grade cohorts who were aged 29/30 in 2013-2017. The CC analyses included 2,257 individuals who responded in all six waves: 642 from the 2002 cohort, 549 from the 2003 cohort, 547 from the 2004 cohort, and 519 from the 2005 cohort. Considering all 9,802 sampled individuals who responded at baseline, the attrition rates across the six follow-up waves were 43.6%, 48.4%, 52.1%, 55.9%, 59.0%, and 61.8%, respectively, resulting in a total of 27,372 available observations for the AC analyses from 6,787 individuals who participated in at least one of the follow-up waves.

[Table 2 about here]

The eAppendix includes a detailed description of the constructed weights and the variables used in the tree methods. The AUC values and the descriptive summaries of the six AC weights and the one CC weight are shown in Table 2. The AUC values ranged between 0.65 and 0.68, indicating moderate prediction power of the tree models, the fit of which is determined by a tradeoff between prediction accuracy and tree sizes. The attrition-adjusted CC weights had the largest variability, and the distributions of the AC weights were similar, except for Wave 6 (age 29/30), where there was increased variability in the weights.

Wave-Specific Descriptive Statistics

[Table 3 about here]

We first considered prevalence estimates for each specific wave. The three groups of individuals (9,802 sampled individuals, 2,257 CC individuals, 6,787 AC individuals) had different sociodemographic characteristics, given in Table 3. Compared to all sampled baseline individuals (baseline), the majority of the CC respondents were female (66% in CC, 52% at baseline, and 57% in AC), white (81% in CC, 71% at baseline, and 75% in AC), and less likely to report any drug use (20% in CC, 28% at baseline, and 26% in AC). The AC individuals did not generally present different characteristics from all sampled individuals at baseline. However, the differences became apparent when we examined the respondents in each follow-up wave. We compared unweighted CCs and ACs, as well as the CCs and ACs with attrition-adjusted weights, in the descriptive summaries for college attendance status and substance use outcomes from the first (age 19/20) to the sixth (age 29/30) follow-up.

[Figure 1 about here]

Figure 1 depicts the estimated proportions of individuals attending college across time based on these four methods. The estimated proportions of individuals with full-time college attendance increased from age 19/20 to age 29/30. However, different approaches resulted in different values of the proportion estimates. The CC analysis yielded the highest estimated proportions of individuals attending college, around 60-72%, while the AC analysis yielded estimates around 50-67%. The attrition-adjusted, weighted analyses reduced both estimates, showing that the attritors were less likely to attend college across young adulthood.

[Figure 2 about here]

The prevalence estimates of substance use across time based on these four methods are presented in Figure 2. Across all analyses the prevalence of binge drinking increased and peaked at age 21/22, and then decreased in a monotone fashion, and the attrition weighting adjustments did not affect the trends. The AC and CC analyses presented similar decreasing trends in the proportions of 30-day cigarette smokers. The prevalence of annual marijuana use decreased from age 19/20 to age 25/26, remained similar until age 27/28, and decreased again through age 29/30. Weighting increased the prevalence most strongly in CC analyses. The attrition-adjusted CC weighting reduced the differences between CC and AC estimates of annual non-medical prescription opioid misuse prevalence, and the reduction effect was apparent between age 23/24 and 29/30.

We found similar results for the prevalence of composite annual NMPD misuse. As for composite annual illegal drug use, the prevalence substantially decreased between age 21/22 and age 25/26 and then remained similar, where the AC estimates were higher than those of the CCs and weighting tended to inflate the estimated rates.

Overall, across the six different substance use outcomes, the AC estimates were higher than the CC estimates with varying trajectories over time; attrition-adjusted weighting increased both estimates, and the effect was larger for CC than for AC. This indicates that the attriters who missed follow-up waves tended to be substance users, and the CC analysis underestimated prevalence as a result.

Trajectory Modeling

For each of the eight trajectory models, we collected the coefficients of the wave indicators and the interactions between wave and the college attendance indicators estimated based on the mean structure given in Equation (1). We present the odds ratios of the five categorical panel wave indicators for college attenders and non-attenders and their 95% confidence intervals in Figures

3-8. The y-axis was set at the same range for all eight plots inside each figure to facilitate comparison, and the plots with different scales are presented in the eAppendix. The trajectories varied across different methods and outcomes. Generally, AC estimates had lower variances than CC analyses due to larger sample sizes in the follow-up waves. The inclusion of weights increased variances in CC analyses, but the variance inflation was negligible in AC analyses.

[Figure 3 about here]

Figure 3 shows that the prevalence of binge drinking followed a non-linear trend, first increasing through age 23/24 in the AC analyses, and then decreasing after that through age 29/30. The attrition-adjusted AC analysis showed that college attendance reduced the prevalence of binge drinking at age 23/24 and age 27/28. All CC analyses showed that the peak was at age 21/22 and did not display any substantial differences between college attenders and non-attenders. The AC analyses yielded more efficient trajectory estimates of binge drinking with smaller variances than the CC analyses, and weighting slightly changed the estimates.

[Figure 4 about here]

Figure 4 shows that the estimated cigarette smoking rate for those attending college generally decreased over time under both CC and AC analyses. However, in the AC analyses, the trend was a non-linear, inverted U shape for college non-attenders, where the rates of cigarette smoking substantially increased from age 21/22 through 23/24. College attenders had a significantly lower probability of smoking than those who did not attend. Weighting did not affect these analyses. Like binge drinking, the AC estimates of cigarette smoking trajectories had lower variability than the CC estimates.

[Figure 5 about here]

The AC and CC analyses in Figure 5 for marijuana use yielded different findings. The CC methods failed to detect the decreasing trends at age 21/22 for college non-attenders. However, the AC analyses showed that the age effects on the prevalence of marijuana use decreased from age 19/20 through age 29/30. Weighting with attrition adjustments slightly increased the uncertainty of the odds ratio estimates for this outcome.

[Figure 6 about here]

In the AC analyses, non-medical prescription opioid misuse started decreasing for college attenders at age 23/24 through age 29/30, and the decrease started for non-college attenders at age 27/28, as shown in Figure 6. The CC analyses of prescription opioid misuse indicated that the rate increased between age 21/22 and age 25/26 for those without college education. The attrition-adjusted weighting had negligible effects on the trend and estimation uncertainty.

[Figure 7 about here]

The trends of NMPD misuse shown in Figure 7 are similar to those of prescription opioid misuse, except that in the AC analyses the decreases started at age 25/26 for college attenders and at age 29/30 for non-college attenders. The CC analyses had larger variances and failed to show the age variation. The effect of the weighting adjustments was small in this case.

[Figure 8 about here]

The AC estimates of the composite illicit drug use trends were also more efficient than those under the CC approach. Figure 8 indicates that the AC analysis of the composite illicit drug use showed that the prevalence increased between age 19/20 and 21/22 and then decreased between age 23/24 and 29/30 for non-college attenders, and the decrease occurred between age 23/24 and 27/28 for college attenders. The CC analyses did not show any age differences until age 27/28 for college non-attenders. We did not observe substantial effects of weighting adjustments for the illicit drug use outcomes.

Overall, the AC analyses with the clustering features and attrition-adjusted weights resulted in the most efficient estimates. The attrition-adjusted weighted AC analyses utilized more data and, more importantly, corrected attrition bias in the estimates, even though the effect is generally minor on the trend estimates. The CC analyses yielded different estimates with large variances, with less statistical power to detect age effects or moderation effects of college attendance.

The AC analyses showed decreasing age effects on the use of cigarettes, marijuana, prescription opioids, NMPDs, and illicit drugs, and moderation effects of college attendance on rates of binge drinking, the use of cigarettes, marijuana use, prescription opioid misuse, NMPD misuse, and illicit drug use. The CC-GEE analysis provided evidence supporting that the decreasing trends were significant, potentially because of the correlation adjustment, where individuals had participated in all follow-up waves and provided complete trajectories. However, the variance estimates under CC-GEE failed to account for the attrition-adjusted weights that inflate the variances. The weights played a negligible role in AC analyses of the trajectory modeling, which would generally be expected if the models are well-specified (Korn & Graubard, 1999), but weights were essential in the descriptive summary estimates.

DISCUSSION

Using national U.S. data from the MTF study, we evaluated the effects of attrition adjustment on longitudinal trajectory estimates of substance use from adolescence through young adulthood, and compared different approaches to this type of analysis. The eight methods considered varied in terms of using CCs or ACs, using GEE or not, and using attrition-adjusted weights or not. Their performances depended on the relationship between attrition and the substance use outcomes, the trajectory model specification and covariates included in the model, the variability of the weights, and the sample size. Overall, the weighted AC analysis adjusting for clustering and attrition appeared to be the most effective approach. These findings have important methodological, clinical, and policy implications.

Methodological Implications

We recommend using all available cases and including variables strongly related to the outcome in longitudinal trajectory modeling. The attrition weight construction accounts for all sample selection features and baseline characteristics related to the substance use outcomes. The role of weights is prominent in the descriptive prevalence estimates, but modest in the trajectory modeling. When we added the features that affect sample selection (e.g., strata) as covariates in the trajectory modeling, using the weights in estimation did not change the estimates, and the interpretation of coefficients became conditional on these features (results are not presented here and available upon request). When one has concerns about the outcome model specification, accounting for weights in the model can offer protection against model misspecification (Korn & Graubard, 1999; Kott, 2007; Winship & Radbill, 1994; Pfeffermann, 2011), in the sense that estimates will be unbiased with respect to the sample design. It is important to simultaneously adjust for correlations of repeated observations and attrition bias with weights, especially if weights are informative about the outcomes. Hence, the AC analysis with the attrition-adjusted weights appears to be the most effective approach. Yet, validation of empirical findings requires additional evidence.

The empirical comparison suggests a few directions for future methodological research. First, ideally a full factorial design based on Table 1 with 16 different approaches should be examined. We include IDs as a clustering variable in the design-adjusted weighted analyses, essentially mimicking GEE with an exchangeable correlation structure. The point estimates will be similar to those of weighted GEEs (Robins, Rotnitzky, & Zhao, 1995). However, the programs enabling weighted GEEs may give misleading variance estimates (Natarajan et al., 2008). Multilevel models could serve as subject-specific (as opposed to marginal) alternatives if there is interest in the explicit estimation of between-individual variances in trajectories. However, the appropriate implementation of multilevel modeling in this context requires weights for multiple levels of the data hierarchy: baseline weights and time-varying weights (Heeringa, West, & Berglund, 2017; Pfeffermann, 1993; Rabe-Hesketh & Skrondal, 2006). The estimation of multilevel models in our application cannot achieve convergence, and the computation of generalized linear mixed effects models with weights needs further developments, such as using Bayesian paradigms. Second, all methods considered here assumed missingness at random. Other studies have evaluated adjustment techniques assuming missingness not at random (e.g., Feldman & Rabe-Hesketh, 2012; Terry-McElrath et al., 2017; Deng et al., 2013; etc.), which allows for sensitivity analyses. However, different assumptions have been introduced for model estimation. Rigorous evaluation of different methods through simulation studies is still needed, and this work is ongoing. Third, MI approaches can be generalized to simultaneously handle attrition and item nonresponse (Si, Reiter, & Hillygus, 2015, 2016; Si, Palta, & Smith, 2020; Si et al., 2021). Extensions of current MI approaches are required to accommodate the complex challenges in trajectory modeling and implementation in practice, such as accounting for complex survey design features, non-monotone attrition, and a large number of mixed types of variables.

Substantive Implications

Our investigation can be extended to other substance use outcomes (e.g., substance use disorder) and other longitudinal studies to enhance estimates for clinical and policy purposes. Our results show different trajectories and implications for different drug classes. Consistent with other research (Arterberry, et al., 2020), we observe decreasing trends with age in the use of cigarettes, marijuana use, prescription opioid misuse, NMPD misuse, and illicit drug use, and that college attenders have lower prevalences of cigarette smoking and use of some substances during particular age periods. The differences based on college attendance have real-world implications for making clinical and policy decisions, such as determining population-level estimates and community resources that should be allocated for substance-related screening and interventions. The moderating effects vary across substance use outcomes and time. The non-linear trend of binge drinking prevalence showed the importance of trajectory modeling across multiple time points, with a peak value at age 21/22 (Patrick et al., 2019). It will also be valuable to examine other national longitudinal probability-based studies (e.g., Hyland et al., 2017; Harris et al., 2019), with different survey designs, instruments, measures, response patterns, and response rates.

CONFLICT OF INTEREST

None declared.

DATA SHARING AND ACCSSIBILITY

The restricted-use panel data of the Monitoring the Future study are analyzed via the NAHDAP Virtual Data Enclave Management System maintained by the Inter-university Consortium for Political and Social Research at the University of Michigan. The codes are publicly accessible at https://github.com/yajuansi-sophie/MTFattrition.

HUMAN STUDIES AND SUBJECTS

The study is exempted from the Independent Review Board as the secondary data analysis does not involve human subjects.

REFERENCES

Arterberry, B.J., Boyd, C.J., West, B.T., Schepis, T.S., & McCabe, S.E. (2020). DSM-5 substance use disorders among college-age young adults in the United States: Prevalence, remission and treatment. J Am Coll Health, 68(6), 650-657.

Binder, D. A. (1983). On the variances of asymptotically normal estimators from complex surveys. International Statistical Review, 51(3), 279–292.

Bollen, K.A., Biemer, P.P., Karr, A.F., Tueller, S., & Berzofsky, M.E. (2016). Are survey weights needed? A review of diagnostic tests in regression analysis. Annual Review of Statistics and Its Application, 3, 375–92.

Breiman, L., Friedman, J.H., Olshen, R.A., & Stone, C.J. (1984). Classification and Regression Trees, Wadsworth, Belmont, CA, Chapman & Hall, New York.

Brick, J.M., & Williams D. (2013). Explaining rising nonresponse rates in cross-sectional surveys. The ANNALS of the American Academy of Political and Social Science, 645, 36-59.

de Leeuw, E., Hox, J., & Luiten, A. (2018), International nonresponse trends across countries and years: An analysis of 36 years of labour force survey data. Survey Insights: Methods from the Field. Retrieved from https://surveyinsights.org/?p=10452.

Deng, Y., Hillygus. D.S., Reiter, J.P., Si, Y., & Zheng, S. (2013). Handling attrition in longitudinal studies: The case for refreshment samples. Statistical Science, 22, 238–56.

Feldman, B.J., & Rabe-Hesketh, S. (2012). Modeling achievement trajectories when attrition is informative. Journal of Educational and Behavioral Statistics, 37(6), 703–36.

Halekoh, U. Højsgaard, S., & Yan, J. (2006). The R Package geepack for Generalized Estimating Equations. Journal of Statistical Software, 15(2), 1–11.

Harris, K.M., Halpern, C.T., Whitsel, E., et al. (2019) Cohort Profile: The National Longitudinal Study of Adolescent to Adult Health (Add Health). Intern J Epidemiol. 48(5): 1415-1415k. doi: 10.1093.ije/dyz115 PMID: 31257425 PMCID: PMC6857761.

Heeringa, S.G., West, B.T., & Berglund, P.A. (2017). Applied Survey Data Analysis. 2nd ed. CRC Press.

Hothorn, T., Hornik, K., Strobl, C., & Zeileis, A. (2021). Party: A laboratory for recursive partytioning. Version 1.3-6. http://cran.r-project.org/web/packages/party/.

Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive rartioning: A conditional inference framework. Journal of Computational and Graphical Statistics, 15, 651-674.

Hyland, A., Ambrose, B.K., Conway, K.P., Borek, N., Lambert, E., Carusi, C., et al. (2017) Design and methods of the Population Assessment of Tobacco and Health (PATH) Study. Tob Control. 26(4):371–8.

Jager, J., Schulenberg, J.E., O'Malley, P.M., & Bachman, J.G. (2013). Historical variation in drug use trajectories across the transition to adulthood: the trend toward lower intercepts and steeper, ascending slopes. Development and Psychopathology, 25, 527-543.

Keyes, K.M., Jager, J., Platt, J., Rutherford, C., Patrick, M.E., Kloska, D.D., & Schulenberg, J. (2020). When does attrition lead to biased estimates of alcohol consumption? Bias analysis for loss to follow-up in 30 longitudinal cohorts. Int J Methods Psychiatr Res., 29(4),1-9.

Korn, E., & Graubard, B. (1999). Analysis of Health Surveys. John Wiley & Sons.

Kott, P. S. (2007). Clarifying some issues in the regression analysis of survey data. Survey Research Methods, 1,11-18.

Liang, K.Y., & Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. Biometrika, 7, 13–22.

Linden-Carmichael, A.N., Kloska, D.D., Evans-Polce, R., Lanza, S.T., & Patrick, M.E. (2019). College degree attainment by age of first marijuana use and parental education. Substance Abuse, 40(1), 66-70.

Little, R. J. A., & Rubin, D.B. (2019). Statistical Analysis with Missing Data: Third Edition. New York: John Wiley & Sons.

Lumley, T. (2020) survey: analysis of complex survey samples. R package version 4.0.

McCabe, S.E., Kloska, D.D., Veliz, P., Jager, J., & Schulenberg, J.A. (2016). Developmental course of non-medical use of prescription drugs from adolescence to adulthood in the United States: national longitudinal data. Addiction, 111, 2166-2176.

McCabe, S.E., Veliz, P.T., Dickinson, K., Schepis, T.S., & Schulenberg, J.E. (2019). Trajectories of prescription drug misuse during the transition from late adolescence into adulthood: A national longitudinal multi-cohort study. Lancet Psychiatry, 6(10), 840-850.

McCabe, S.E., Veliz, P., & Schulenberg, J.E. (2018). How collegiate fraternity and sorority involvement relates to substance use during young adulthood and substance use disorders in early midlife: A national longitudinal study. J Adolesc Health, 62(3S), S35-S43.

Miech, R. A., Johnston, L. D., O'Malley, P. M., Bachman, J. G., Schulenberg, J. E., & Patrick, M. E. (2021). Monitoring the Future national survey results on drug use, 1975-2020: Volume I, Secondary school students. Ann Arbor: Institute for Social Research.

Natarajan, S., Lipsitz, S., Fitzmaurice, G., Moore, C., & Gonin, R. (2008). Variance Estimation in Complex Survey Sampling for Generalized Linear Models. Journal of the Royal Statistical Society. Series C (Applied Statistics), 57(1), 75-87.

Patrick, M.E., Kloska, D.D., Mehus, C., Terry-McElrath, Y.M., O'Malley, P.M., & Schulenberg, J.E. (2021). Key subgroup differences in age-related change from 18 to 55 in alcohol and marijuana use: U.S. national data. Journal of Studies on Alcohol and Drugs, 82(1), 93-102.

Patrick, M.E., Terry-McElrath,Y.M., Kloska, D.D., & Schulenberg, J.E. (2016). High-Intensity drinking among young adults in the United States: Prevalence, frequency, and developmental change. Alcoholism: Clinical and Experimental Research 40(9):1905-1912.

Patrick, M.E., Terry-McElrath, Y.M., Lanza, S. T., Jager, J., Schulenberg, J.E., O'Malley, P.M. (2019). Shifting age of peak binge drinking prevalence: Historical changes in normative trajectories among young adults aged 18 to 30. Alcohol Clin. Exp. Res., 43(2), 287-298.

Patrick, M.E., Veliz, P., Linden-Carmichael, A., & Terry-McElrath, Y.M. (2018). Alcohol mixed with energy drink use during young adulthood. Addict Behav, 84, 224-230.

Pfeffermann, D. (1993). The role of sampling weights when modeling survey data. International Statistical Review, 61(2), 317–37.

Pfeffermann, D. (2011). Modelling of complex survey data: Why model? Why is it a problem? How can we approach it? Survey Methodology, 37(2), 115–36.

Rabe-Hesketh, S., & Skrondal, A. (2006). Multilevel modelling of complex survey data. Journal of the Royal Statistical Society Series A, 169(4), 805–27.

R Core Team. (2020). R: A language and environment for statistical computing, https://www.R-project.org/

Robins, J.M., Rotnitzky, A., & Zhao, L.P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. Journal of the American Statistical Association, 90(429), pp.106-121.

Roghani, A., Nyarko, S.H., Potter, L. (2021). Smoking cigarettes, marijuana, and the transition to marriage among cohabiters in the USA. Glob Soc Welf.

Rubin, D.B. (1987). Multiple Imputation for Nonresponse in Surveys. New York: John Wiley & Sons.

Schulenberg, J.E. & Maggs, J.L. (2002). A developmental perspective on alcohol use and heavy drinking during adolescence and the transition to young adulthood. Journal of Studies on Alcohol, 14, 54-70.

Schulenberg, J.E., Merline, A.C., Johnston, L.D., O'Malley, P.M., Bachman, J.G., & Laetz, V.B. (2005). Trajectories of marijuana use during the transition to adulthood: the big picture based on national panel data. Journal of Drug Issues, 35, 255-279.

Schulenberg, J. E., Patrick, M. E., Johnston, L. D., O'Malley, P. M., Bachman, J. G., & Miech, R. A. (2021). Monitoring the Future national survey results on drug use, 1975-2020: Volume II, college students and adults ages 19-60. Ann Arbor: Institute for Social Research.

Schulenberg, J. E., Patrick, M. E., Kloska, D. D., Maslowsky, J., Maggs, J. L., & O'Malley, P. M. (2015). Substance use disorder in early midlife: A national prospective study on health and well-being correlates and long-term predictors. Substance Abuse: Research and Treatment.

Si, Y., Heeringa, S., Johnson D., Little, R.J.A., Liu, W., Pfeffer, F., & Raghunathan T. (2021), Multiple Imputation with Massive Data: An Application to the Panel Study of Income Dynamics. Journal of Survey Statistics and Methodology. https://doi.org/10.1093/jssam/smab038.

Si, Y., Little, R., Mo, Y., & Sedransk, N. (2022a). Nonresponse Bias Analysis in Longitudinal Educational Assessment Studies, under review, http://arxiv.org/abs/2204.07105.

Si, Y., Little, R., Mo, Y., & Sedransk, N. (2022b). A Case Study of Nonresponse Bias Analysis in Educational Assessment Surveys, under review. https://arxiv.org/abs/2104.04432.

Si, Y., Palta, M., & Smith, M. (2020). Bayesian profiling multiple imputation for missing hemoglobin values in electronic health records. Annals of Applied Statistics, 14(4), 1903–1924.

Si, Y., Reiter, J. P., & Hillygus, D. S. (2015). Semi-Parametric selection models for potentially non-ignorable attrition in panel study with refreshment samples. Political Analysis, 23, 92–112.

Si, Y., Reiter, J. P., & Hillygus, D. S. (2016). Bayesian latent pattern mixture models for handling attrition in panel studies with refreshment samples. The Annals of Applied Statistics, 10, 118–43.

Terry-McElrath, Y.M., O'Malley, P.M., Johnston, L.D., Bray, B.C., Patrick, M.E., & Schulenberg, J.E. (2017). Longitudinal patterns of marijuana use across ages 18-50 in a US national sample: a descriptive examination of predictors and health correlates of repeated measures latent class membership. Drug and Alcohol Dependence 171:70-83.

Terry-McElrath, Y.M., & Patrick, M.E. (2016a). Intoxication and binge and high-intensity drinking among U.S. young adults in their mid-twenties. Substance Abuse, 37(4), 597-605.

Terry-McElrath, Y.M., & Patrick, M.E. (2016b). Trends and timing of cigarette smoking uptake among U.S. young adults: survival analysis using annual national cohorts from 1976 to 2005. Addiction, 110(7), 1171-1181.

Terry-McElrath, Y.M., Patrick, M.E., Johnston, J.D., & Schulenberg, J.E. (2019). Young adult longitudinal patterns of marijuana use among U.S. National samples of 12th grade frequent marijuana users: a repeated-measures latent class analysi. Addiction, 114(6), 1035-1048.

Winship, C., & Radbill, L. (1994). Sampling weights and regression analysis. Sociological Methods & Research, 23, 230-257.

Table 1. Comparison and names of eight approaches investigated.

Method	CC	AC	Selection features	Correlation	Attrition-adjusted weight
CC	√		√		
AC		√	√		
CC-gee	√		√	√	
AC-gee		√	√	√	
CC-ID cluster	√		√	√	
AC-ID cluster		√	√	√	
CC-attr-w	√		√	√	√
AC-attr-w		√	√	√	√

- Note: CC: complete case analysis; AC: available case analysis; CC-gee: GEE with complete cases; AC-gee: GEE with available cases; CC-ID cluster: complete case analysis accounting for the cluster structure by individuals; AC-ID cluster: available case analysis accounting for the cluster structure by individuals; CC-attr-w: attrition-adjusted weighted analysis accounting for the cluster structure by individuals with complete cases; and AC-attr-w: attrition-adjusted weighted analysis accounting for the cluster structure by individuals with available cases.

Table 2. Area Under the Curve (AUC) values of the classification tree models for the attrition adjustment and descriptive summaries of different weights.

	Sample size	AUC	Weight				
			Min	25th Pctl	Median	75th Pctl	Max
Baseline-18	9802						
CC	2257	0.65	1.2	2.6	3.1	5.2	7.8
AC-19/20	5529	0.66	1.3	1.3	1.8	2.0	2.8
AC-21/22	5058	0.67	1.3	1.5	2.0	2.3	3.7
AC-23/24	4700	0.66	1.2	1.5	2.1	2.4	3.5
AC-25/26	4324	0.67	1.3	1.6	2.2	2.8	3.8
AC-27/28	4019	0.67	1.4	1.7	2.4	3.0	4.6
AC-29/30	3742	0.68	1.5	1.8	2.6	3.2	6.3

- Note: CC: complete case analysis; AC: available case analysis; Pctl: percentile.

Table 3. Descriptive summaries of selected baseline sociodemographics.

	Baseline	CC	AC	AC waves					
				19/20	21/22	23/24	25/26	27/28	29/30
Sample size	9802	2257	6787	5529	5058	4700	4324	4019	3742
Age in months (SD)	217(6)	216(5)	217(6)	217(5)	217(5)	216(5)	216(5)	216(5)	216(5)
Drug use reporting									
Yes	28%	20%	26%	25%	25%	25%	25%	24%	24%
No	72%	80%	74%	75%	75%	75%	75%	76%	76%
Sex									
Male	48%	34%	43%	41%	41%	40%	40%	39%	39%
Female	52%	66%	57%	59%	59%	60%	60%	61%	61%
Race/ethnicity									
Black	10%	5%	8%	8%	7%	7%	6%	6%	6%
White	71%	81%	75%	75%	77%	77%	78%	79%	79%
Asian	4%	4%	4%	4%	4%	4%	4%	4%	4%
Hispanic	11%	7%	9%	9%	9%	8%	8%	8%	8%
Other	4%	3%	4%	4%	4%	4%	4%	3%	3%

- Note: Baseline: sampled baseline individuals; CC: complete case analysis; AC: available case analysis; AC waves are six follow-up waves marked by the modal ages; SD: standard deviation.

Figure 1. Estimated proportion of participants with full-time four-year college attendance across six follow-up waves marked by the modal ages.

Figure 2. Prevalence estimates of substance use at baseline and six follow-up waves marked by the modal ages. The substance use measures include: binge drinking (five or more drinks in a row during the past 2 weeks), cigarette smoking (in the past 30 days), marijuana use (in the past 12 months), non-medical prescription opioid misuse (in the past 12 months), a composite indicator of any non-medical prescription drug (NMPD) misuse (of four specific drug classes that included amphetamines, sedatives, tranquilizers, or opioids, in the past 12 months), and a composite indicator of any use of other selected illicit drugs, including LSD, other hallucinogens, cocaine, or heroin (in the past 12 months).

Figure 3. Odds ratio estimates of panel effects on binge drinking (five or more drinks in a row during the past 2 weeks) prevalence for college attenders and non-attenders.

Figure 4. Odds ratio estimates of panel effects on cigarette smoking (in the past 30 days) prevalence for college attenders and non-attenders.

Figure 5. Odds ratio estimates of panel effects on the use of marijuana (in the past 12 months) for college attenders and non-attenders.

Figure 6. Odds ratio estimates of panel effects on the non-medical misuse of prescription opioids (in the past 12 months) for college attenders and non-attenders.

Figure 7. Odds ratio estimates of panel effects on the non-medical misuse of prescription drugs (NMPD) for four specific drug classes that included amphetamines, sedatives, tranquilizers, or opioids, in the past 12 months) for college attenders and non-attenders.

Figure 8. Odds ratio estimates of panel effects on the use of illicit drugs (a composite indicator of any use of other selected illicit drugs, including LSD, other hallucinogens, cocaine, or heroin, in the past 12 months) for college attenders and non-attenders.

Legend: CC: complete case analysis; AC: available case analysis; CC-gee: GEE with complete cases; AC-gee: GEE with available cases; CC-ID cluster: complete case analysis accounting for the cluster structure by individuals; AC-ID cluster: available case analysis accounting for the cluster structure by individuals; CC-attr-w: attrition-adjusted weighted analysis accounting for the cluster structure by individuals with complete cases; and AC-attr-w: attrition-adjusted weighted analysis accounting for the cluster structure by individuals with available cases.

APPENDIX

eTable 1. Baseline variables and their item missingness percentages for the Monitoring the Future study used in the attrition analysis.

Variables at baseline	Values	Missing
School strata indicator	1-27	0.0%
Age in months	187-279	0.4%
Cohort indicator	2002, 2003, 2004, 2005	0.0%
Oversampling indicator for 12th grade drug use	0.33, 1	0.0%
School type	Public, Catholic, non-Catholic private	0.0%
Family structure	None, 1 parent, 2 parents	0.7%
#smoked cigarettes (cig) in past 30 days	None, <1 cig/Day, 1-5/Day, 1 / 2 pack/Day, 1 pack/Day, 1.5 pack/Day, 2+ packs/Day	1.2%
#times of drinking in lifetime	0, 1-2, 3-5, 6-9, 10-19, 20-39, 40+ Occasions	3.5%
#times of 5+ drinking in a row in last 2 weeks	0, 1, 2, 3-5, 6-9, 10+ Occasions	5.0%
#times of marijuana use in lifetime	0, 1-2, 3-5, 6-9, 10-19, 20-39, 40+ Occasions	1.9%
#times of LSD use in lifetime	0, 1-2, 3-5, 6-9, 10-19, 20-39, 40+ Occasions	1.4%
#times of PSYD use in lifetime	0, 1-2, 3-5, 6-9, 10-19, 20-39, 40+ Occasions	1.4%
#times of cocaine use in lifetime	0, 1-2, 3-5, 6-9, 10-19, 20-39, 40+ Occasions	1.6%
#times of AMPH use in lifetime	0, 1-2, 3-5, 6-9, 10-19, 20-39, 40+ Occasions	1.3%
#times of SED/BARB use in lifetime	0, 1-2, 3-5, 6-9, 10-19, 20-39, 40+ Occasions	1.2%
#times of TRQL use in lifetime	0, 1-2, 3-5, 6-9, 10-19, 20-39, 40+ Occasions	1.0%
#times of heroin use in lifetime	0, 1-2, 3-5, 6-9, 10-19, 20-39, 40+ Occasions	1.1%
#times of NARC use in lifetime	0, 1-2, 3-5, 6-9, 10-19, 20-39, 40+ Occasions	1.3%
Sex	Male, Female	0.0%
Race/ethnicity	Black, White, Asian, Hispanic, Other	2.0%
Parental marital Status	Married, engaged, separate/divorced, single	1.1%
Parental education	Having college education (college graduate, graduate school) or not (grade school, some high school, high school graduate, some college)	7.1%
Mother paid job	No, sometimes, most time, all time	1.4%

Political preference	Strong Republican, mild Republican, mild Democratic, strong Democratic, Independent, No preference, other	4.6%
Political belief	Very conservative, conservative, moderate, liberal, very liberal, radical, none above	2.1%
Religious preference	Baptist, Churches of Christ, Disciples of Christ, Episcopal, Lutheran, Methodist, Presbyterian, United Church of Christ, Other Christian, Unitarian Universalist, Catholic, Eastern Orthodox, Jewish, Latter Day Saints, Muslim, Buddhist, other religion, none	2.9%
Attend religious service	Never, Rarely, 1-2 times every month, 1 each week or more	1.5%
Religion importance	Not, little, pretty, very important	1.6%
High school program	college preparation, general, vocational-technical, other	2.4%
Miss school due to illness (4 weeks)	None, 1 day, 2 days, 3 days, 4-5 days, 6-10 days, 11+ days	4.2%
Skip school (4 weeks)	None, 1 day, 2 days, 3 days, 4-5 days, 6-10 days, 11+ days	5.7%
Miss school due to others (4 weeks)	None, 1 day, 2 days, 3 days, 4-5 days, 6-10 days, 11+ days	4.9%
High school grades	D, C-, C, C+, B-, B, B+, A-, A	2.8%
Will do vocational-technical school	Definitely won't, probably won't, probably will, definitely will	6.1%
Will do Military	Definitely won't, probably won't, probably will, definitely will	6.7%
Will do 2yr college	Definitely won't, probably won't, probably will, definitely will	6.2%
Will do 4yr college	Definitely won't, probably won't, probably will, definitely will	4.7%
Will do Graduate/Professional school	Definitely won't, probably won't, probably will, definitely will	5.9%
Weekly work	None, 5 or fewer, 6-10, 11-15, 16-20, 21-25, 26-30, 30+ hours	3.8%
Pay from jobs during an average week	None, $1-5, $6-10, $11-20, $21-35, $36-50, $51-75, $76-125, $126+	5.8%
Pay from other sources during an average week	None, $1-5, $6-10, $11-20, $21-35, $36-50, $51-75, $76-125, $126+	6.9%
#evenings go out per week	<1, 1, 2, 3, 4-5, 6-7	4.1%
Frequency going out with a date	Never, 1/month or fewer, 2-3/month, 1/week, 2-3/week, 3+/week	4.6%

Average # of miles driving a car, truck or motorcycle per week	None, 1-10, 11-50, 51-100, 101-200, >200 miles	4.2%
#times receiving tickets (12 months)	None, 1, 2, 3, 4+	5.4%
#times receiving tickets after drinking (12 months)	None, 1, 2, 3, 4+	5.8%
#times receiving tickets after marijuana use (12 months)	None, 1, 2, 3, 4+	5.8%
#times receiving tickets after other drug use (12 months)	None, 1, 2, 3, 4+	6.0%
#accidents during past 12 months (12 months)	None, 1, 2, 3, 4+	6.0%
#accidents after drinking	None, 1, 2, 3, 4+	6.2%
#accidents after marijuana use	None, 1, 2, 3, 4+	6.3%
#accidents after other drug use	None, 1, 2, 3, 4+	6.3%

eTable 2. Item missingness percentages for variables used in the trajectory modeling.

	Baseline-age	Follow-up waves (modal age)					
	18	19/20	21/22	23/24	25/26	27/28	29/30
College attendance*	4.7%	1.8%	2.1%	1.5%	1.4%	1.6%	1.7%
Marital status	1.1%	0.7%	0.5%	0.6%	0.5%	0.4%	0.6%
Binge drinking	5.0%	4.2%	3.6%	3.1%	2.5%	2.5%	2.5%
Cigarette smoking in the past 30 days	1.2%	1.9%	1.9%	1.7%	1.6%	1.5%	1.3%
Marijuana use in the past 12 months	2.2%	2.1%	2.1%	2.0%	1.6%	1.5%	1.3%
LSD use in the past 12 months	1.4%	1.7%	1.7%	1.6%	1.4%	1.1%	1.1%
Other hallucinogen use in the past 12 months	1.4%	1.6%	1.5%	1.5%	1.3%	0.9%	1.1%
Cocaine use in the past 12 months	1.6%	1.7%	1.5%	1.5%	1.4%	1.4%	1.4%
Amphetamines misuse in the past 12 months	1.4%	1.4%	1.2%	1.1%	0.9%	0.9%	1.1%
Sedative misuse in the past 12 months	1.2%	1.1%	1.0%	1.2%	0.9%	0.9%	0.9%
Tranquilizer misuse in the past 12 months	1.0%	1.1%	1.1%	1.1%	0.8%	1.1%	1.0%
Heroin use in the past 12 months	1.1%	1.0%	1.0%	1.0%	1.0%	1.0%	0.8%
Opioid misuse in the past 12 months	1.3%	1.4%	1.2%	1.2%	1.1%	0.9%	0.9%

*The baseline measure is on the intent to attend college.

Item Nonresponse

The proportions of cases with missing data for these MTF variables are generally low, as shown in eTables 1 and 2; we, therefore, do not expect substantial effects of different imputation methods on the results. We apply different strategies to handle the item nonresponse associated with different variables depending on their role in the analysis. Our findings are robust under different imputation methods, including multiple imputation (MI). MI approaches modeling the dependency structure and accounting for the imputation model uncertainty are generally preferred.

We started from the 54 baseline characteristics included in the attrition analysis. When we tried imputing them, many variables had multiple categories that needed multinomial-logit regression models, for example, the variable measuring religious preference had 18 levels, and only 3 out of 54 baseline variables were continuous. Hence, the sequential imputation models in the R packages *mice* and *mi* do not work well here. To avoid complexity, their missing values in the baseline characteristics listed in eTable 1 were not imputed in the paper but treated as another category since the goal is to predict attrition using the tree-based models.

We implemented a two-stage MI process to impute missing items of the key covariates and substance use measures in the trajectory modeling.

We applied the first-stage of MI based on the subset of baseline characteristics that were used in the trajectory modeling: the cohort indicator (2002-2005), age in months, sex, race/ethnicity, high school grades, parental education, the baseline measures of the corresponding substance use outcomes (binge drinking, cigarette smoking in the past 30 days, marijuana use in the past 12 months, LSD use in the past 12 months, other hallucinogen use in the past 12 months, cocaine use in the past 12 months, amphetamines misuse in the past 12 months, sedative misuse in the past 12 months, tranquilizer misuse in the past 12 months, heroin use in the past 12 months, opioid misuse in the past 12 months), and reported intent to attend a four-year college. The imputation models and methods were tailored to the variable types and included: predictive mean matching for age, multinomial-logit models for race/ethnicity and grades, and logistic regressions for binary outcomes. We saved one dataset of completed baseline measures after MI and used it as the starting point for the following MI implementation of the time-varying measures.

The second-stage of the MI imputed the missing items in the time-varying measures using follow-up wave indicators, full-time four-year college attendance, marital status, the follow-wave measures of substance use outcomes (binge drinking, cigarette smoking in the past 30 days, marijuana use in the past 12 months, LSD use in the past 12 months, other hallucinogen use in the past 12 months, cocaine use in the past 12 months, amphetamines misuse in the past 12 months, sedative misuse in the past 12 months, tranquilizer misuse in the past 12 months, heroin use in the past 12 months, opioid misuse in the past 12 months). We reshaped the data in the long format where each row was one measurement and every participant had multiple rows of repeated measures. We had to ignore the clustering structure in the imputation models as two-level imputation models failed to converge. This could lead to improper imputation results as the imputation model was not congenial with the analysis models that accounted for clustering effects and weights.

We ran the two-stage MI for 30 iterations and kept 10 multiply imputed datasets for inferences. We had 10 replicates of the available cases in the MTF study, and MI did not handle attrition here. We do acknowledge that the single, univariate imputation approaches fail to account for the correlations among variables and the imputation uncertainty.

eTable 3. Coefficient estimates in the logistic regression model for binge drinking based on available cases in the comparison of single imputation and multiple imputation of the missing item values.

Coefficients:	Single imputation		Multiple imputation		
	Estimate	Std. Error	Estimate	Std. Error	Fraction of missing information
(Intercept)	0.19	0.58	0.17	0.59	0.03
Reference level: none					
College intent at age 18	0.47	0.06	0.43	0.06	0.06
Reference level: Cohort 2002					
Cohort 2003	-0.06	0.04	-0.06	0.04	0.02
Cohort 2004	-0.07	0.04	-0.07	0.04	0.02
Cohort 2005	-0.06	0.04	-0.06	0.04	0.02
Age	0.00	0.00	0.00	0.00	0.02
Reference level: Male					
Female	-0.55	0.03	-0.55	0.03	0.03
Reference level: White					
Asian	-0.62	0.07	-0.64	0.08	0.07
Black	-0.94	0.06	-0.97	0.07	0.09
Hispanic	-0.29	0.05	-0.29	0.05	0.04
Other	-0.36	0.07	-0.35	0.07	0.02
Reference level: without college degrees					
Parental education	0.09	0.03	0.08	0.03	0.04
Reference level: other high school grades					
C-	-0.05	0.17	-0.05	0.17	0.03
C	-0.24	0.15	-0.20	0.16	0.06
C+	-0.34	0.15	-0.32	0.15	0.07
B-	-0.33	0.14	-0.30	0.15	0.05
B	-0.43	0.14	-0.41	0.15	0.06
B+	-0.37	0.14	-0.36	0.15	0.05
A-	-0.43	0.14	-0.41	0.15	0.06
A	-0.59	0.14	-0.57	0.15	0.05
Reference level: not married					
Marital status	-0.83	0.04	-0.83	0.04	0.02
Reference level: no					
Binge drinking at age 18	1.23	0.03	1.23	0.03	0.03
Reference level: Non-college attendance at age 19/20					
Non-college attendance at age 21/22	0.38	0.06	0.36	0.07	0.05
Non-college attendance at age 23/24	0.42	0.07	0.42	0.07	0.06
Non-college attendance at age 25/26	0.29	0.07	0.26	0.07	0.04

Non-college attendance at age 27/28	0.28	0.08	0.27	0.08	0.04
Non-college attendance at age 29/20	0.06	0.08	0.04	0.08	0.05
College attendance at age 19/20	0.12	0.03	0.12	0.03	0.03
College attendance at age 21/22	-0.07	0.09	-0.05	0.09	0.04
College attendance at age 23/24	-0.20	0.09	-0.18	0.09	0.07
College attendance at age 25/26	-0.14	0.09	-0.09	0.09	0.05
College attendance at age 27/28	-0.22	0.10	-0.20	0.10	0.03
College attendance at age 29/30	-0.20	0.10	-0.16	0.10	0.05

We used binge drinking as an example and fit a logistic regression model with covariates including: cohort indicator (2002-2005), age in months, sex, race/ethnicity, high school grades , parental education, the baseline binge drinking measures, follow-up wave indicators, full-time four-year college attendance, and marital status.

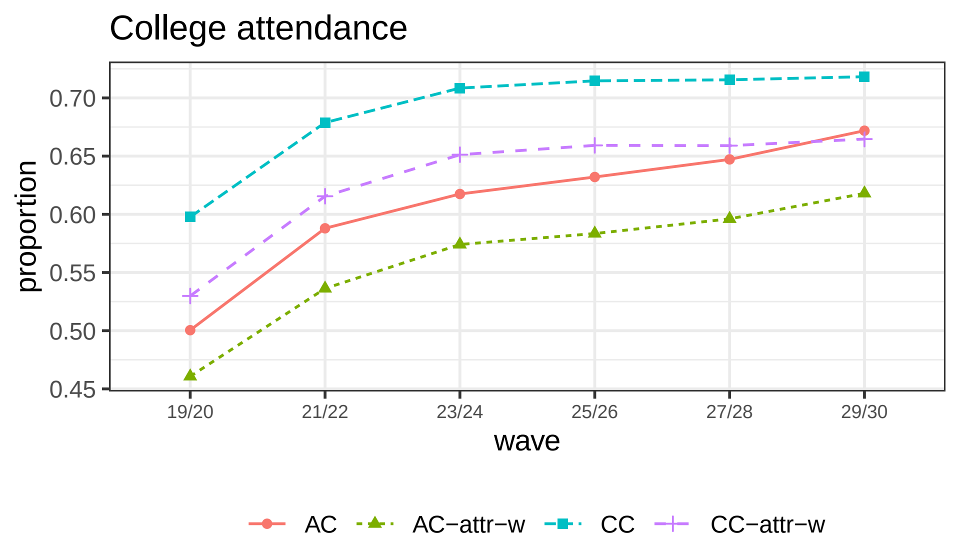
eTable 3 gave the coefficient estimates in the logistic regression models for binge drinking based on available values in the comparison of single imputation and multiple imputation of the missing item values. Overall, we found that single imputation yielded similar estimates and variances comparing to MI. The fraction of missing information (Rubin, 1987) was small, indicating that the uncertainty due to the imputation model was small. Hence, we can use a single imputed dataset of the missing items as our available cases and move forward with the weighted analyses.

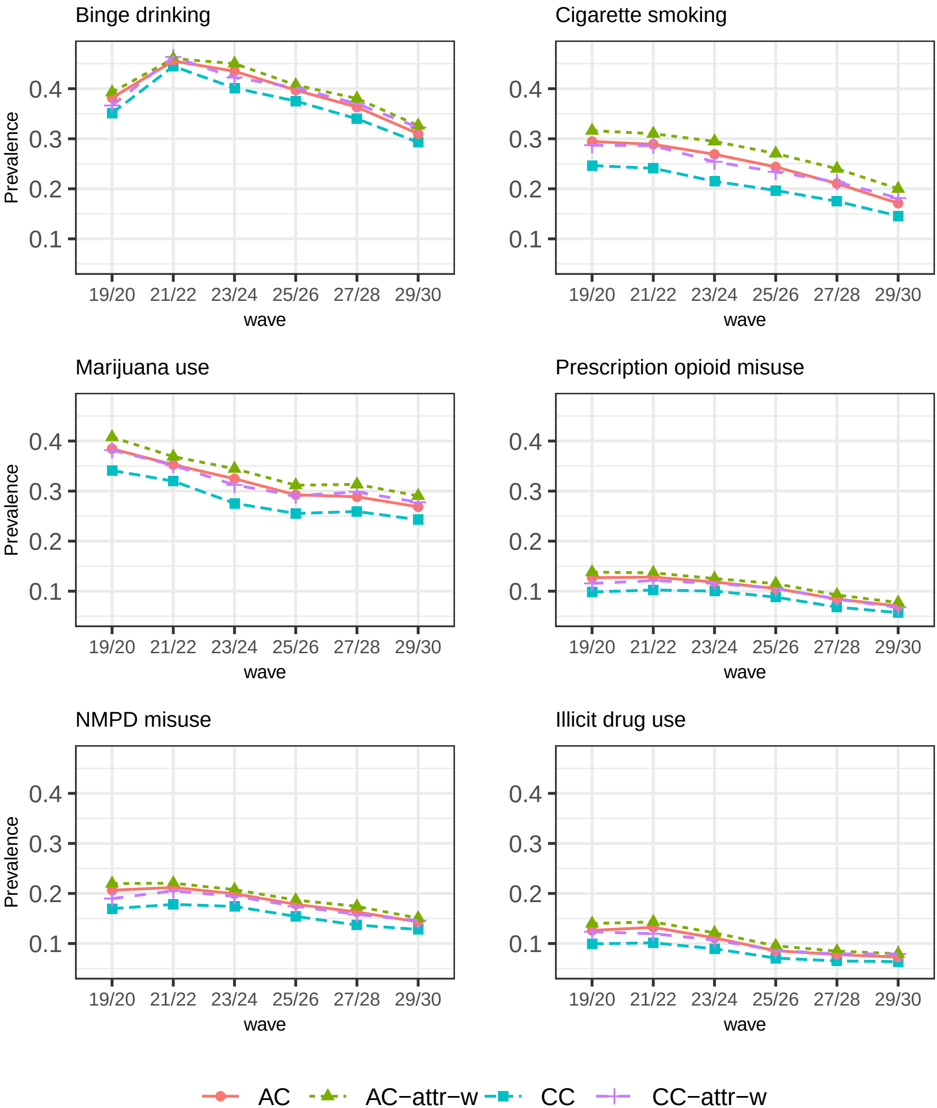
Attrition-Adjusted CC Weights

The logistic regression model for the CC response propensity after forward variable selection showed that the baseline participants who had low average high school grades, were male, were not from a two-parent family structure, got weekly pay from jobs, often missed classes, planned to enter the military, engaged in binge drinking in the past two weeks and cigarette smoking in the past 30 days, and were lifetime users of opioids tended to skip follow-up waves. Relative to the final logistic regression model, the tree retained interactions between high school grades and the following covariates: sex, race/ethnicity, family structure, indicator of oversampling drug users, working status, religious preference, frequencies of missing school classes due to other non-health related reasons, and lifetime drinking frequencies at baseline. The AUC values of the tree and logistic regression model were 0.70 and 0.74, respectively. The tree approach thus had slightly lower predictive performance, but the resulting weights were less variable, which can improve efficiency of the estimates. The resulting CC weights ranged from 1.2 to 46.0. To exclude the extreme values, we trimmed all weights larger than the 95th percentile of the computed weights down to a value of 11.3. The resulting trimmed weights had a mean value of 4.1 with a standard deviation of 2.4.

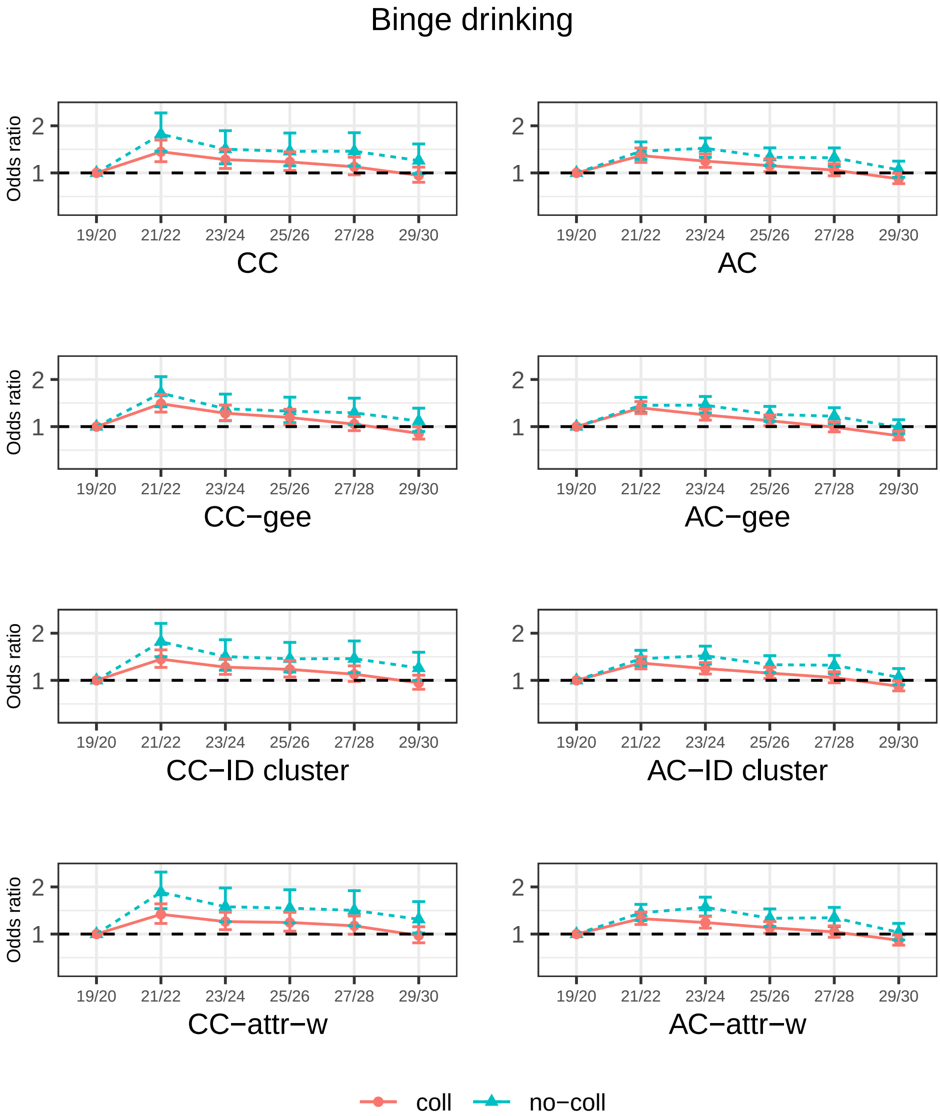
Attrition-Adjusted AC Weights

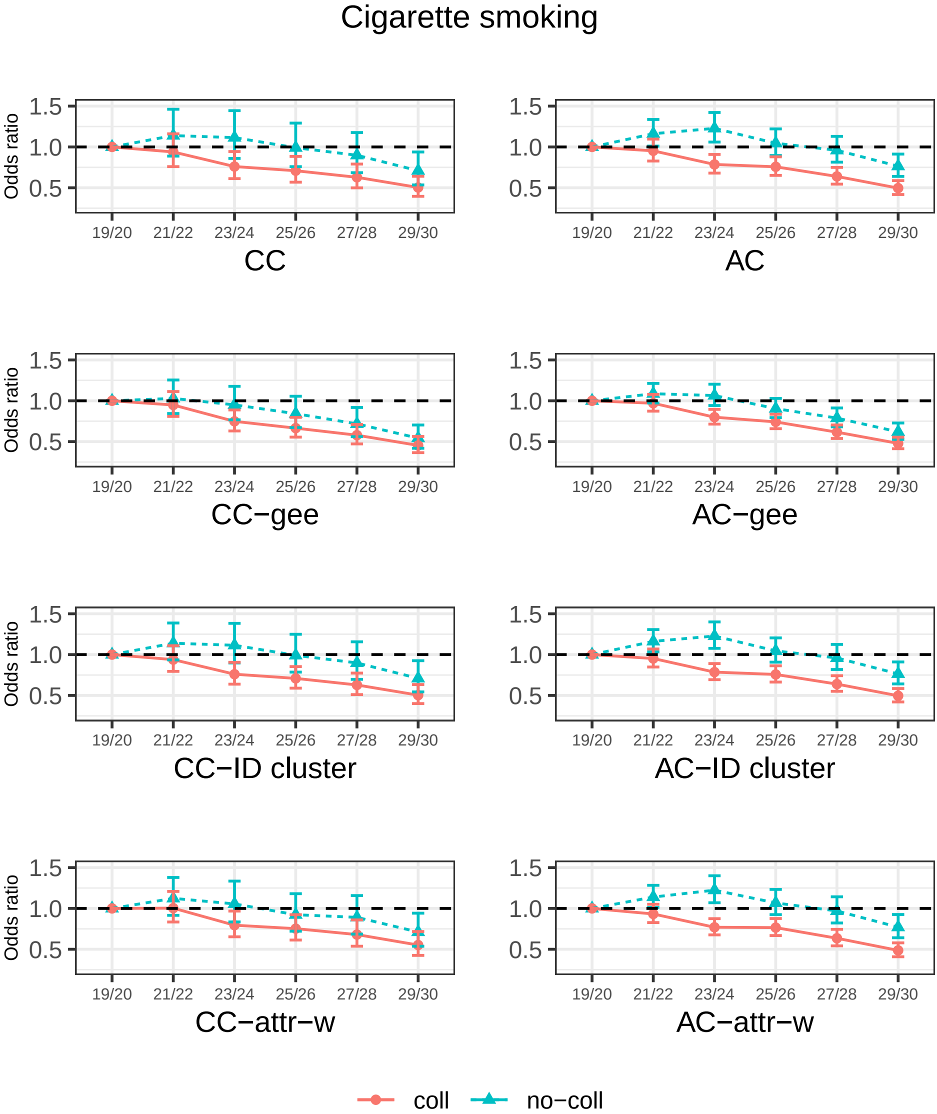
We performed wave-specific AC weighting adjustments. The logistic regression models selected different predictors across the six waves. For example, for the age 19/20 wave, respondents tended to have higher grades in high school, be female, live with both parents, plan to attend graduate/professional school, do not work, often drive, and not use drugs at baseline. The tree complexity and selected predictors or higher-order interactions also differed across waves, emphasizing the importance of performing wave-specific adjustments. At the age 19/20 wave, the constructed tree identified the interactions between high school grades, sex, family structure, substance use outcomes and college plans after high school as important predictors. Generally, across the six waves, the trees identified the interactions between high school grades, sex, race/ethnicity, family types, and substance use outcomes, with AUC values ranging from 0.67 to 0.68. We trimmed the six sets of AC weights at their wave-specific 95th percentiles respectively to reduce the variability of the weights.

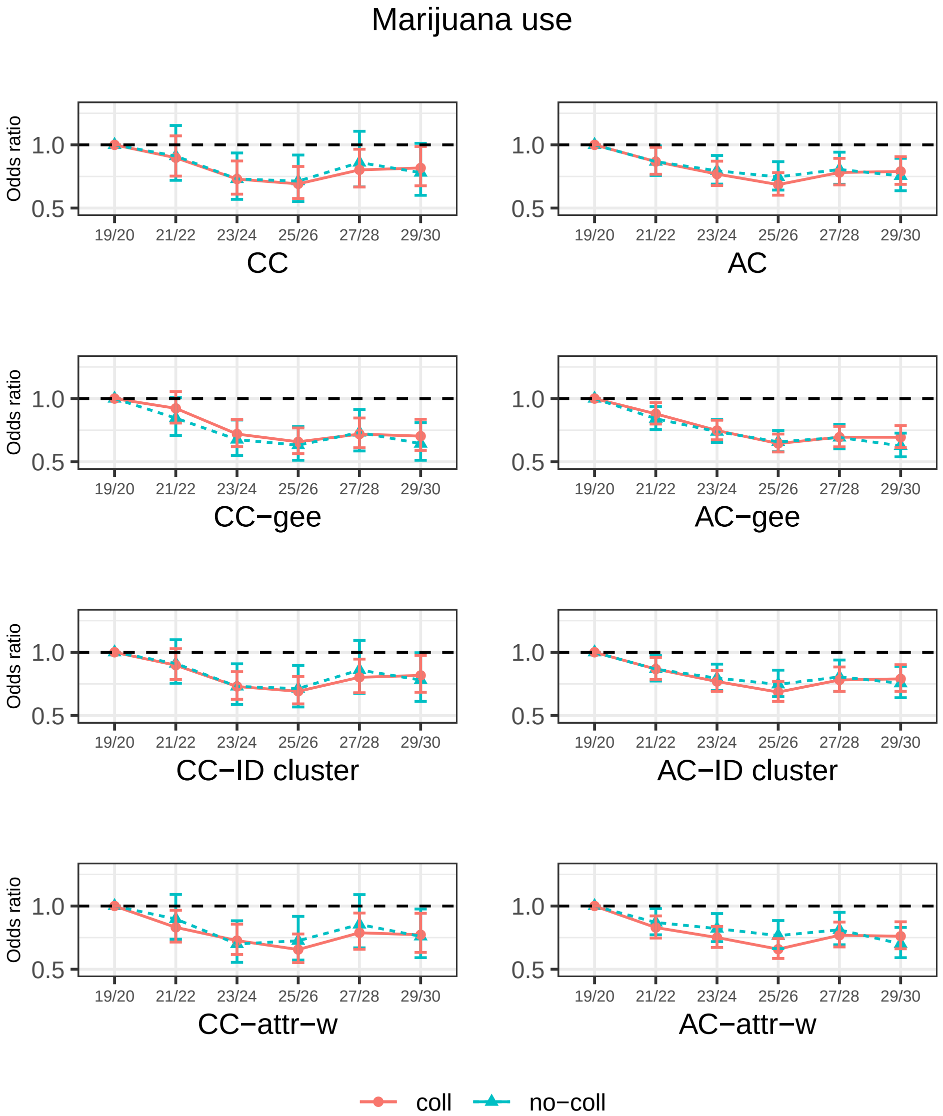

College attendance
proportion
0.70
0.65
0.60
0.55
0.50
0.45
19/20
21/22
23/24
25/26
27/28
29/30
wave
AC
AC−attr−w
CC
CC−attr−w

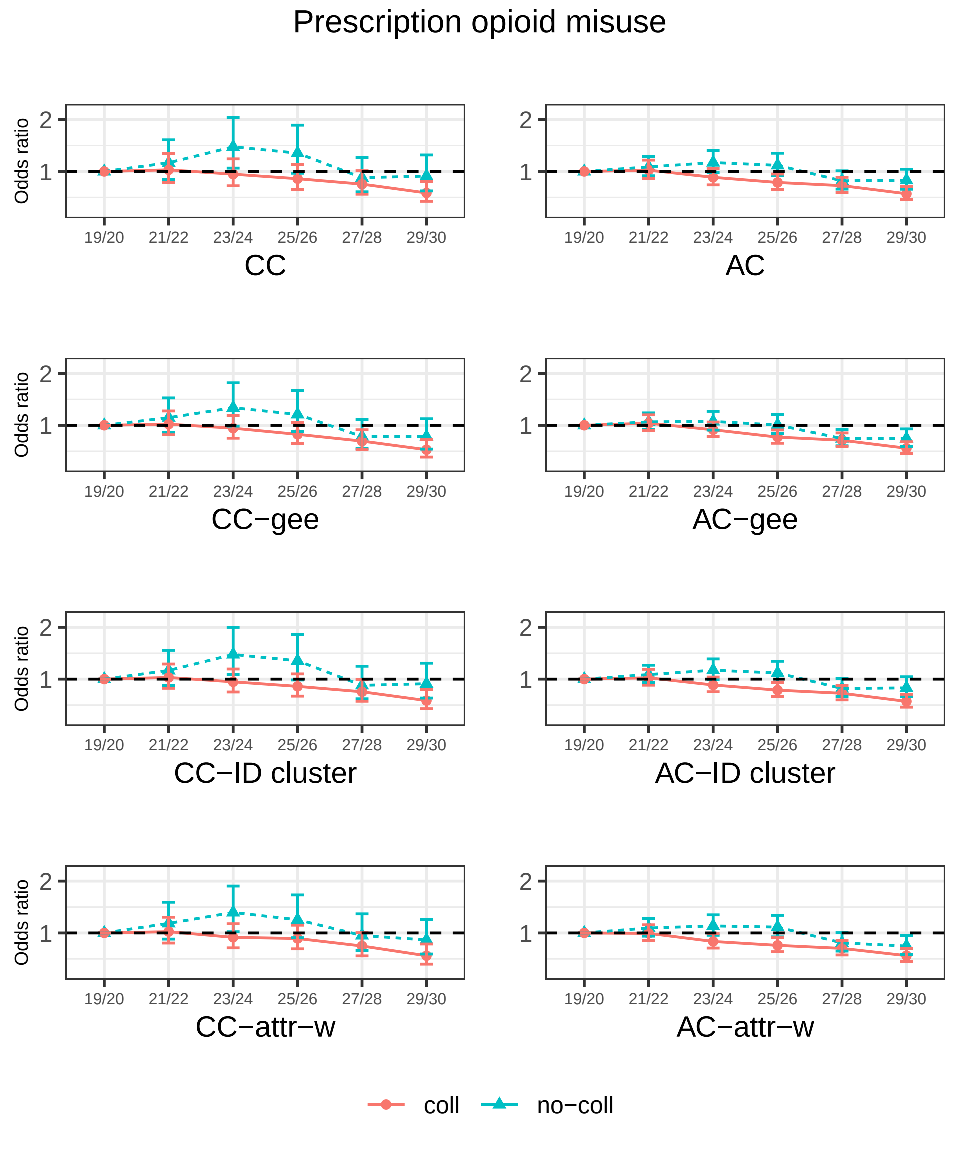


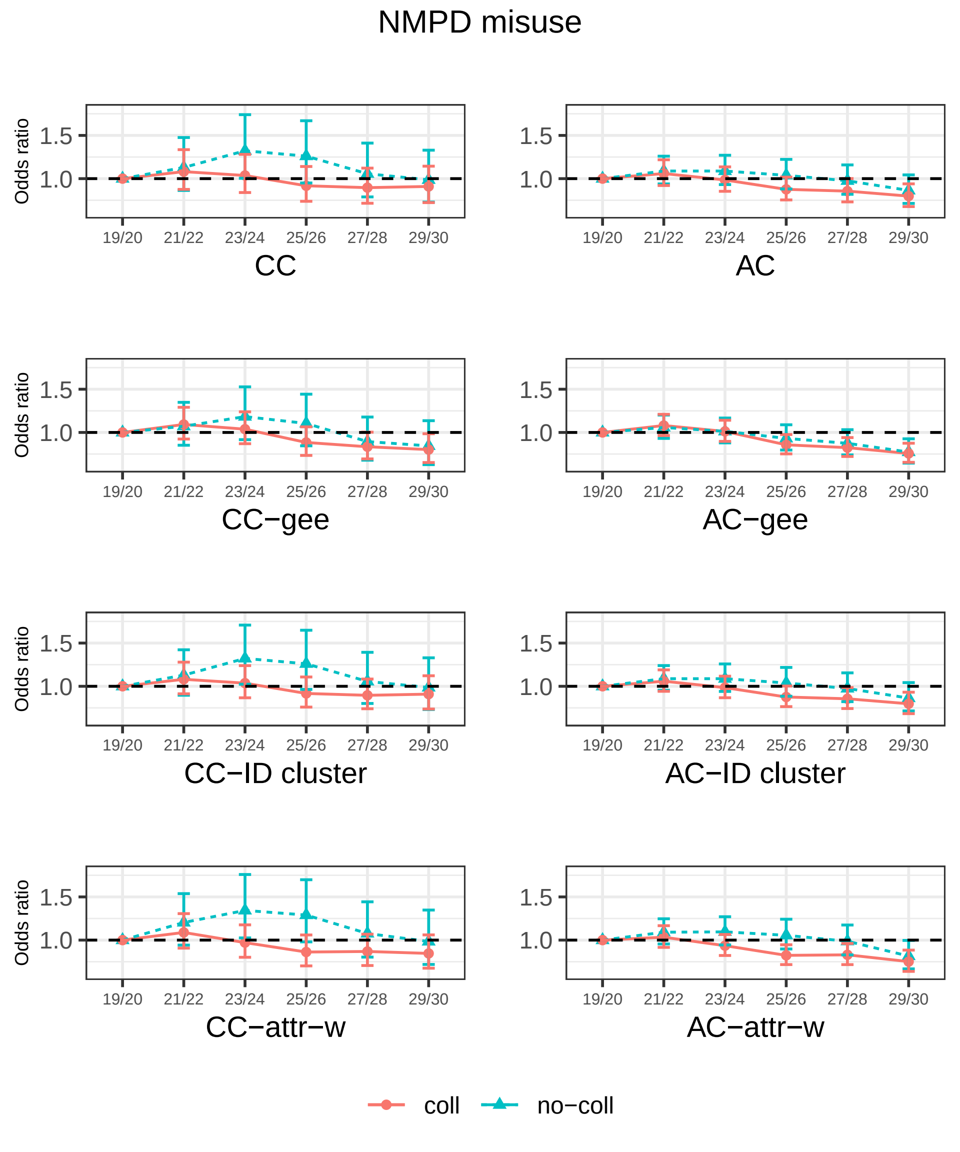
Binge drinking
Cigarette smoking
Marijuana use
Prescription opioid misuse
NMPD misuse
Illicit drug use
Prevalence
0.1
0.2
0.3
0.4
19/20
21/22
23/24
25/26
27/28
29/30
wave
AC
AC−attr−w
CC
CC−attr−w


Binge drinking
Odds ratio
2
1
19/20
21/22
23/24
25/26
27/28
29/30
CC
AC
CC−gee
AC−gee
CC−ID cluster
AC−ID cluster
CC−attr−w
AC−attr−w
coll
no−coll

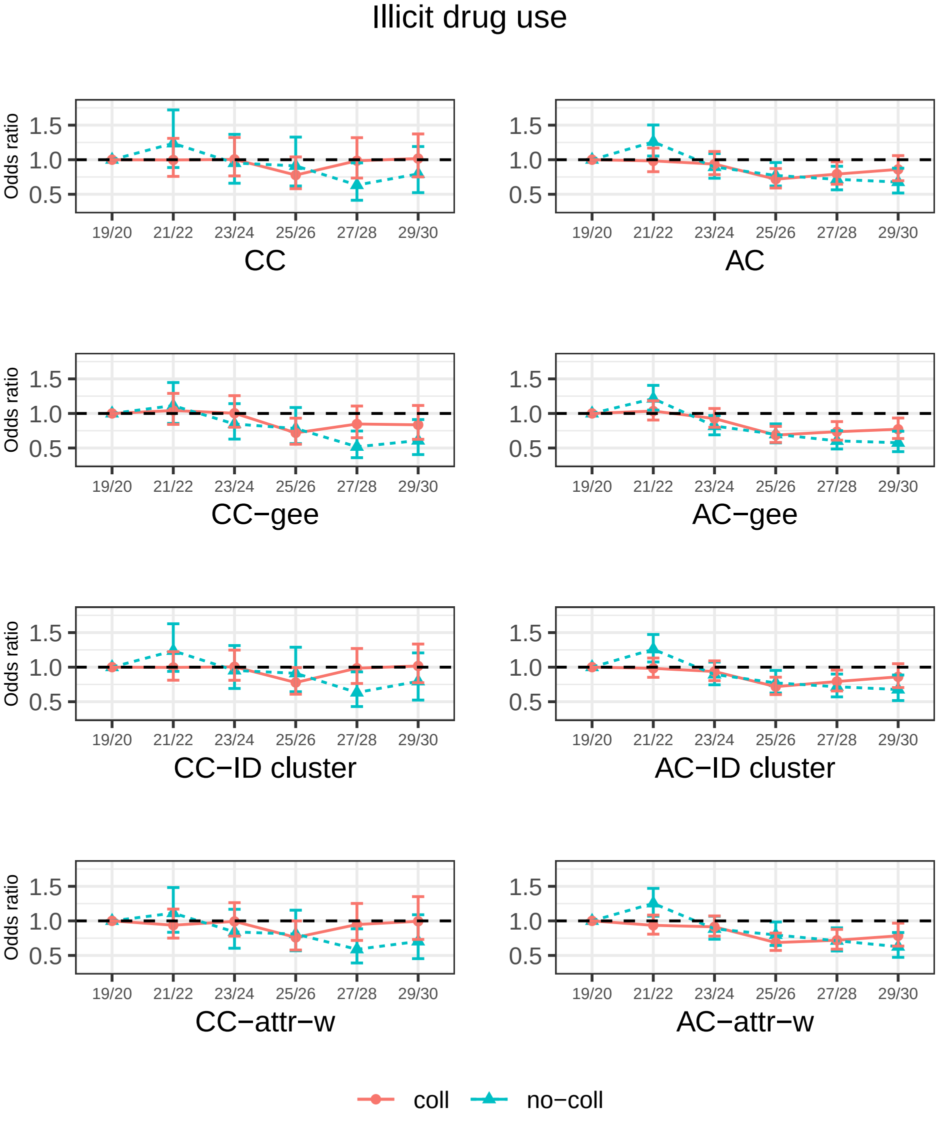

Cigarette smoking
Odds ratio
1.5
1.0
0.5
19/20
21/22
23/24
25/26
27/28
29/30
CC
AC
CC−gee
AC−gee
CC−ID cluster
AC−ID cluster
CC−attr−w
AC−attr−w
coll
no−coll


Marijuana use
Odds ratio
1.0
0.5
19/20
21/22
23/24
25/26
27/28
29/30
CC
AC
CC−gee
AC−gee
CC−ID cluster
AC−ID cluster
CC−attr−w
AC−attr−w
coll
no−coll


Prescription opioid misuse
Odds ratio
2
1
19/20
21/22
23/24
25/26
27/28
29/30
CC
AC
CC−gee
AC−gee
CC−ID cluster
AC−ID cluster
CC−attr−w
AC−attr−w
coll
no−coll



NMPD misuse
Odds ratio
1.5
1.0
19/20
21/22
23/24
25/26
27/28
29/30
CC
AC
CC−gee
AC−gee
CC−ID cluster
AC−ID cluster
CC−attr−w
AC−attr−w
coll
no−coll



Illicit drug use
Odds ratio
1.5
1.0
0.5
19/20
21/22
23/24
25/26
27/28
29/30
CC
AC
CC−gee
AC−gee
CC−ID cluster
AC−ID cluster
CC−attr−w
AC−attr−w
coll
no−coll