Inverse Propensity Score Weighting with a Latent Class Exposure: Estimating the Causal Effect of Reported Reasons for Alcohol Use on Problem Alcohol Use 16 Years Later

Bethany C. Bray[1,2] · John J. Dziak[1] · Megan E. Patrick[3] · Stephanie T. Lanza[1,4,5]

Abstract

Latent class analysis (LCA) has proven to be a useful tool for identifying qualitatively different population subgroups who may be at varying levels of risk for negative outcomes. Recent methodological work has improved techniques for linking latent class membership to distal outcomes; however, these techniques do not adjust for potential confounding variables that may provide alternative explanations for observed relations. Inverse propensity score weighting provides a way to account for many confounders simultaneously, thereby strengthening causal inference of the effects of predictors on outcomes. Although propensity score weighting has been adapted to LCA with covariates, there has been limited work adapting it to LCA with distal outcomes. The current study proposes a step-by-step approach for using inverse propensity score weighting together with the "Bolck, Croon, and Hagenaars" approach to LCA with distal outcomes (i.e., the BCH approach), in order to estimate the causal effects of reasons for alcohol use latent class membership during the year after high school (at age 19) on later problem alcohol use (at age 35) with data from the longitudinal sample in the Monitoring the Future study. A supplementary appendix provides evidence for the accuracy of the proposed approach via a small-scale simulation study, as well as sample programming code to conduct the step-by-step approach.

Keywords Latent class analysis · Causal inference · Propensity scores · Alcohol use · Motives · Reasons for drinking

Latent class analysis (LCA) has proven to be a useful tool in prevention research. It has been used in recent years to identify a wide array of within-individual patterns, such as exposure to

✉ Bethany C. Bray
  bcbray@psu.edu

1 The Methodology Center, Penn State, 404 Health & Human Development Bldg., University Park, PA 16802, USA

2 College of Health and Human Development, Penn State, 304 Health & Human Development Bldg., University Park, PA 16802, USA

3 Institute for Social Research, University of Michigan, 426 Thompson St., Ann Arbor, MI 48106, USA

4 Edna Bennett Pierce Prevention Research Center, Penn State, 314 Biobehavioral Health Bldg., University Park, PA 16802, USA

5 Department of Biobehavioral Health, Penn State, 219 Biobehavioral Health Bldg., University Park, PA 16802, USA

risk factors (Lanza and Rhoades 2013), substance use and other health behaviors (Cardoso et al. 2016; Gilreath et al. 2014; Héroux et al. 2012), and expectancies, attitudes, and norms related to decision-making (Stapleton et al. 2014). This work has contributed to our understanding of the etiology underlying the development of many negative behavioral and health outcomes targeted by prevention programs, as well as the effects of prevention programs (Jiang et al. 2012; Low et al. 2016; Spilt et al. 2013). As the use of LCA has expanded, increasingly complex research questions have been posed about the roles played by latent class variables in development and prevention. Recently, new methodologies for LCA with distal outcomes (Bakk and Vermunt 2016; Bray et al. 2015; Lanza et al. 2013d; Vermunt 2010) have made it easier to study the link between latent class membership and an outcome variable. However, these techniques do not explicitly address confounding by variables that may provide alternative explanations for observed relations. That is, although these new developments provide a way to regress an outcome variable on a latent class variable, and in some cases can include

control variables (Asparouhov and Muthén 2014), by themselves they do not capitalize on modern approaches to causal inference. This is a critical limitation for prevention research and the identification of causal mechanisms leading to negative outcomes.

There are numerous approaches to causal inference, many of which are based on the propensity score (e.g., matching, subclassification, and weighting). For an accessible introduction to propensity scores for causal inference readers are referred to Lanza et al. (2013b). In particular, inverse probability weighting based on the propensity score has become popular in prevention research, and has been used to examine causal effects of time-invariant (Coffman et al. 2012; Varvil-Weld et al. 2014) and time-varying (Bray et al. 2006) exposures and mediators (Coffman and Zhong 2012) and can accommodate other important aspects of developmental models, such as moderators of causal effects (Green and Stuart 2014). Importantly, the use of inverse propensity score weighting has already been demonstrated with LCA with covariates (Lanza et al. 2013c, 2016). In these demonstrations, one of the causal effects of interest was the average exposure effect (ATE; i.e., average treatment effect; causal effect of an exposure for the entire population) of a non-randomized, observed, and binary exposure on a latent class variable, which (by definition) was unobserved and multinomial. Such an approach is useful for understanding causal antecedents of latent class membership, but not consequences of membership. In the current study, we reverse the direction of the causal effect of interest and focus on the ATE of a non-randomized, latent, multinomial exposure (i.e., a latent class variable) on an observed, binary outcome. That is, we propose an approach for implementing inverse propensity score weighting with the currently recommended analytic approach to LCA with distal outcomes. The use of inverse propensity score weighting with a latent class exposure was considered previously by Schuler (2013), Schuler et al. (2014), and Yamaguchi (2015), all of whom compare alternative approaches. Yamaguchi (2015) considers theoretical and technical details in depth.

We demonstrate our proposed approach by examining the link between patterns of reasons for alcohol use during the transition out of high school and later problem alcohol use. Reasons for alcohol use have been associated with alcohol use and problem alcohol use (Patrick et al. 2018; Patrick and Schulenberg 2011; Patrick et al. 2011) and represent an important potential target for prevention programs. For example, drinking to cope with negative emotions may be a more important risk factor than drinking socially (e.g., see Merrill et al. 2014; Patrick et al. 2011). Based on the idea that multiple reasons may co-occur within individuals, previous research has used LCA to identify patterns of reasons for alcohol use (Coffman et al. 2007; Stapinski et al. 2016), marijuana use (Patrick et al. 2016), and e-cigarette use (Evans-Polce et al. 2017). These studies used LCA with covariates or LCA with distal outcomes either to predict patterns of reasons from covariates or to predict outcomes from patterns of reasons.

However, established associations have been correlational in nature. For example, Patrick et al. (2016) examined associations between patterns of reasons for marijuana use at ages 19/20 and problem marijuana use at age 35. Although they were able to control for several variables in the outcome analysis (e.g., gender, race/ethnicity, and parental education), the significant effects of latent class membership on the outcome do not represent ATEs because a rigorous approach to adjusting for confounding was unavailable. Critically, if adolescent patterns of reasons for use are associated with later outcomes, even after using causal inference methods to account for differences across people with different patterns, it provides evidence that patterns of reasons could be used to screen individuals into prevention programs and/or tailor the content of programs. Here, we examine associations between patterns of reasons for alcohol use at age 19 and problem alcohol use at age 35, and we use inverse propensity score weighting as a rigorous approach to adjust for many potential confounders at age 18 and to estimate ATEs. The approach proposed here can be used generally to estimate ATEs of latent class variables on distal outcomes: it accounts for both the latent class measurement model and the propensity for particular exposure group membership (i.e., latent class membership) in order to estimate an unbiased causal effect of latent class membership on a distal outcome.

The purpose of the current study is to provide a step-by-step empirical example of how to apply inverse propensity score weighting with a latent class exposure. This example is conducted with data from the US national Monitoring the Future (MTF) study and estimates the causal effect of reasons for alcohol use at age 19 on later problem alcohol use at age 35.

A Step-by-Step Example

The Causal Question

By estimating a set of ATEs, we are seeking to answer the following specific research question: "What differences in problem alcohol use at age 35 would be expected if all individuals in the population had a certain pattern of reasons for alcohol use at age 19, compared to if all individuals in the population had a different pattern of reasons?" In particular, we focus on the ATEs for differences in problem use vs. non-problem use or abstinence at age 35 if (a) all individuals in the population had a pattern of reasons for alcohol use at age 19 labeled Enhancement Reasons, compared to if all

individuals in the population had a pattern labeled Few Reasons and (b) all individuals had a pattern labeled Coping Reasons, compared to if all individuals had a pattern labeled Enhancement Reasons. Following standard notation (e.g., McCaffrey et al. 2013) for observed exposures, we define the ATE of membership in one particular latent class versus another as follows. Let $Y_i[t]$ be the value on the distal outcome Y which would have been observed if individual i had actually belonged to latent class t; this may or may not be the counterfactual for a given i and t, because class membership is unknown. Then, the ATE of membership in class t versus class t' is defined as the expectation $E(Y[t] - Y[t'])$ taken over all individuals (i.e., all values of i) in the population.

Participants

The example used longitudinal data from the MTF study. Annually, MTF collects national representative data from about 15,000 US high school seniors (i.e., 12th graders; Miech et al. 2017); approximately 2450 individuals from each cohort are selected for longitudinal follow-up (Schulenberg et al. 2017). The sample included senior year cohorts from 1976 to 1998 who provided data at age 35 from 1993 to 2015; the sample included all individuals who (1) were selected for longitudinal follow-up; (2) reported alcohol use in the past 12 months at age 19; and (3) provided data on (a) reasons for alcohol use at age 19, (b) all confounders at age 18, and (c) alcohol use and alcohol use problems at age 35. Only those individuals who reported using alcohol in the past 12 months were asked about their reasons for use. Note that MTF uses multiple randomly assigned forms to decrease respondent burden; reasons for alcohol use were included on one form (of five from 1976 to 1988, and of six from 1989 to 2015). All analyses used attrition weights to adjust for attrition from ages 18 to 35, calculated as the inverse of the estimated probability of participation at age 35, based on several demographic, educational, and substance use variables, as well as the original sampling weight correcting for over-sampling of substance users at age 18. The weighted analytic sample size was 6572 (52.9% female; 83.1% White, 7.3% Black, 4.8% Hispanic, 4.8% other).

Measures

Potential Confounders The example considered 15 potential confounders at age 18: cohort (1976 to 1998), gender (male vs. female), race (White vs. Black, Hispanic, and other), living in a two-parent family (yes vs. no), parents' highest level of education (some college or more vs. high school graduate or less), average grades in high school (A to D), college preparatory curriculum in high school (yes vs. no), definitely planned to graduate from a 2- or 4-year college (yes vs. no), number of days skipped school in the past month (0 to 11+),

importance of religion (not important to very important), average number of nights out with friends per week (< 1 to 6–7), number of cigarettes smoked in the past month (0 to 2+ packs per day), binge drinking (i.e., 5+ drinks in a row) in the past 2 weeks (0 to 10+ times), marijuana use in the past 12 months (yes vs. no), and other illegal substance use in the past 12 months (yes vs. no). Weighted descriptive statistics for all potential confounders are shown in Table 1.

Exposure: Latent Classes of Reasons for Alcohol Use At age 19, participants were asked, "On how many occasions (if any) have you used alcohol during the last 12 months?" (1 = 0 occasions, 2 = 1–2 occasions, 3 = 3–5 occasions, 4 = 6–9 occasions, 5 = 10–19 occasions, 6 = 20–39 occasions, 7 = 40 or more occasions). Non-users were excluded from analyses because they were not asked questions about reasons for use. At age 19, participants who indicated they had used alcohol at least once in the past 12 months were asked, "What have been the most important reasons for your using alcohol? (Mark all that apply.)" Responses to the 14 reasons were dichotomous (marked vs. unmarked). Reasons could be broadly categorized as social and recreational, coping with negative mood and experiences, utility and pleasure of effects, and compulsive use (see Patrick et al. 2011). These 14 binary variables represent the latent class indicators. Weighted descriptive statistics for all latent class indicators are shown in Table 1.

Outcome: Problem Alcohol Use at Age 35 At age 35, participants who indicated they had used alcohol in the past 5 years were asked, "Think back over the last five years. Did your use of alcohol cause you any of the following problems?" (0 = no, 1 = a little, 2 = some, 3 = a lot; coded as none vs. any). The 16 problems included, for example, "Caused you to behave in ways that you later regretted" and "Hurt your relationship with your spouse/partner or girlfriend/boyfriend." This measure does not provide a clinical diagnosis of alcohol use disorder, but items cover 8 of the 11 DSM-5 criteria for alcohol use disorder (Patrick et al. 2011; Schulenberg et al. 2015). For the outcome analysis, participants were categorized as problem users (i.e., had affirmative responses to two or more criteria; see Schulenberg et al. 2015) or abstinent/ non-problem users (i.e., had not used alcohol in the past 5 years or had used alcohol in the past 5 years but did not meet criteria for disorder).[1] Weighted descriptive statistics for the outcome are shown in Table 1.

---

[1] The earliest timeframe during which problem alcohol use could be assessed with the MTF data was the last 5 years prior to age 35. If there were individuals who developed problem alcohol use and fully recovered by age 30, we were unable to identify them as positive cases.

Table 1 Weighted descriptive statistics for the potential confounders, latent class indicators, and outcome

		Mean (std. dev.) or frequency	Range or valid %
Potential confounders			
Cohort		1986.5 (7.4)	1976–1998
Gender	Female	2273	52.9
	Male	2027	47.1
Race	White	3572	83.1
	Black	314	7.3
	Hispanic	207	4.8
	Other	206	4.8
Living in two parent family	No	931	21.7
	Yes	3369	78.4
Parents' highest level of education	High school or less	1545	35.9
	Some college or more	2755	64.1
Average grades in high school		6.1[B to B+] (2.4)	1[D]-9[A]
Took college prep curriculum	No	1762	41.0
	Yes	2537	59.0
Definitely plans to graduate college	No	1846	42.9
	Yes	2454	57.1
No. of days cut school in the past 30 days		1.7[0 to 1 day] (1.5)	1[0 days]–7[11+ days]
Importance of religion		2.6[little to pretty imp] (1.2)	1[not imp]–4[very imp]
Ave no. days per wk. going out		3.5[2 to 3 eves] (1.5)	1[< 1 eves]–6[6–7 eves]
Smoked cig in past 30 days		1.7[0 to < 1 cig/day] (1.5)	1[0 cigs]–7[2+ pks/day]
Drank 5+ drinks in the past 2 weeks		1.7[0 to 1 time] (1.4)	1[0 times]–6[10+ times]
Used marijuana in the past 12 months		1.7[0 to 1–2 times] (1.8)	1[0 occs]–7[40+ occs]
Used other substances in the past 12 months	No	3293	76.6
	Yes	1007	23.4
Latent class indicators[a]			
To have a good time with my friends		3428	79.7
Because it tastes good		2344	54.5
To relax or relieve tension		2224	51.7
To feel good or get high		2180	50.7
Because of boredom, nothing else to do		901	21.0
To get away from my problems or troubles		808	18.8
Because of anger or frustration		651	15.1
To fit in with a group I like		370	8.6
To get to sleep		324	7.5
To increase effects of other drug(s)		175	4.1
To seek deeper insights and understanding		167	3.9
To get through the day		45	1.0
To decrease/offset effects of other drug(s)		29	0.7
I am "hooked"—I feel I have to drink		25	0.6
Outcome			
Problem alcohol use	Abstinent/non-problem use	2011[b]	67.6[b]
	Problem use	966	32.5

[a] Each alcohol use reason was reported as "marked" (i.e., endorsed as a reason for use) or "unmarked" (i.e., not endorsed). Reported frequencies and valid %s are for endorsement

[b] Non-users accounted for 148 participants (5.0%)

Valid % refers to the percent of non-missing cases

Step-by-Step Analytic Strategy

Step 1: Conduct Latent Class Model Identification and Selection

LCA was used to identify unique patterns of reasons to use alcohol. LCA is a type of finite mixture model that uses manifest items with categorical responses to divide a population into a set of mutually exclusive and exhaustive latent classes (Collins and Lanza 2010). In a standard LCA, two sets of parameters are of most interest. Latent class membership probabilities describe the distribution of the classes in the population. Item-response probabilities describe class-specific probabilities of providing particular responses to the latent class indicators. Classes are interpreted and named based on the patterns of item-response probabilities.

Models with one to six classes were compared using penalized fit criteria (e.g., AIC, BIC), solution stability, and theoretical interpretability; multiple sets of random starting values were used to assess model identification. Model fit and selection information are shown in Table 2; we selected the 4-class model for interpretation and further analysis, and posterior probabilities from this model were retained for use in step 2. For information about recommended approaches to LCA model identification and selection, interested readers are referred to Collins and Lanza (2010) and Lanza et al. (2013a).

Parameter estimates for the 4-class model are shown in Table 3. Class 1 (.53 prevalence) was characterized by comparatively low probabilities of all of the reasons for alcohol use and was labeled Few Reasons. Class 2 (.27) was characterized by comparatively high probabilities of "to have a good time with my friends" and "to feel good or get high" and comparatively low probabilities of the other reasons and was labeled Enhancement Reasons. Similarly, class 3 (.17) was characterized by comparatively high probabilities of "to relax or relieve tension," "to get away from my problems or troubles," and "because of anger or frustration" and was labeled Coping Reasons. Class 4 (.04) was labeled Diverse Reasons because it was characterized by comparatively high probabilities of all of the reasons for alcohol use.

Latent class exposures with more than two classes open the door for numerous causal questions, as is the case with observed multinomial exposures (Imbens 1999). A researcher may be interested in, for example, estimating the causal effects of (a) membership in one particular latent class versus another particular latent class, (b) membership in one particular latent class versus all other latent classes, or (c) all pairwise comparisons between latent classes. Here, to narrow the scope of the example and focus on the step-by-step approach, we consider the ATEs of membership in the Enhancement Reasons versus Few Reasons and the Coping Reasons versus Enhancement Reasons latent classes. Collectively, these ATEs include approximately 96% of the population and represent comparisons of classes that are expected to be at increasing risk for problem

alcohol use based on their configurations of alcohol use reasons.

Step 2: Calculate Propensity Scores and Inverse Propensity Score Weights

The estimated propensity score for individual i, $\hat{\pi}_i$, is based on a statistical model of selection into an exposure group, here a particular latent class of reasons for alcohol use. Propensity scores can be used to balance exposure groups on many potential confounders of the association between the exposure and outcome simultaneously (Lunceford and Davidian 2004; Robins et al. 2000; Rosenbaum and Rubin 1983), here the association between latent class membership of reasons for alcohol use and later problem alcohol use.

When an exposure, T, is observed and binary, propensity score estimates are often obtained using logistic regression: membership in the exposed ($T=1$) versus not exposed ($T=0$) group is predicted by the vector of potential confounders, $\mathbf{x}_i$. An individual's propensity score is his/her predicted probability of membership in the group to which he/she belongs, conditional on all potential confounders. Namely, $\hat{\pi}_i = P(T = 1|\mathbf{x}_i)$ for those in the exposed group and $\hat{\pi}_i = P(T = 0|\mathbf{x}_i)$ for those in the not exposed group. When the ATE is the estimand of interest, the inverse propensity score weights can be defined as $1/\hat{\pi}_i$. The probabilities $P(T = 1|\mathbf{x}_i)$ are derived from a logistic regression model with all potential confounders $(1, …, p)$ as predictors:

$$P(T = 1|\boldsymbol{x}_i) = \frac{\exp\big(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}\big)}{1 + \exp\big(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}\big)}$$

and $P(T = 0|\mathbf{x}_i) = 1 - P(T = 1|\mathbf{x}_i)$.

Now, suppose that the exposure is multinomial instead of binary (but still observed). If the exposure is multinomial with groups $1, …, n_T$, estimation of the propensity score needs to be generalized. This can be done by using multinomial logistic regression instead of binary logistic regression. An individual's estimated propensity score is still his/her probability of membership in the group to which he/she belongs, conditional on all potential confounders:

$$\hat{\pi}_i = P(T = T_i|\boldsymbol{x}_i) = \sum_{t=1}^{n_T} P(T = t|\boldsymbol{x}_i)I\{T_i = t\} \qquad (1)$$

where $I\{T_i = t\}$ is defined as 1 if $T_i = t$ and 0 otherwise. Now, however, the probabilities $P(T = t|\mathbf{x}_i)$ are estimated from a multinomial logistic regression model,

$$P(T = t|\boldsymbol{x}_i) = \frac{\exp\big(\beta_{0t} + \beta_{1t} x_{i1} + \beta_{2t} x_{i2} + \cdots + \beta_{pt} x_{ip}\big)}{\sum_{t'=1}^{n_T}\exp\big(\beta_{0t'} + \beta_{1t'} x_{i1} + \beta_{2t'} x_{i2} + \cdots + \beta_{pt'} x_{ip}\big)}$$

$$(2)$$

with appropriate identifiability constraints (e.g., the β coefficients for the reference class set to 0 to reflect the constraint that all probabilities must sum to 1). Note that each individual

Table 2 Model fit and selection criteria for reasons for alcohol use latent class analyses

No. of classes	LL	p	AIC	BIC	aBIC	Entropy
1	−14,969.80	14	29,967.60	30,051.58	30,007.09	1.00
2	−14,281.93	29	28,621.86	28,795.82	28,703.68	0.74
3	−14,131.44	44	28,350.88	28,614.82	28,475.02	0.70
4	**−14,045.55**	**59**	**28,209.11**	**28,563.03**	**28,375.56**	**0.72**
5	−13,993.39	74	28,134.78	28,578.68	28,343.55	0.66
6	−13,938.12	89	28,054.237	28,588.12	28,305.33	0.61

Models with more than six classes were not considered due to difficulty replicating the log-likelihood with multiple sets of random starting values. Bold font indicates the selected model.

LL log-likelihood, *p* number of free parameters, *AIC* Akaike information criterion, *BIC* Bayesian information criterion, *aBIC* sample size adjusted BIC

receives only one $\hat{\pi}_i$ because each individual belongs only to one exposure group.

When the exposure is latent, however, each individual's true exposure group, T_i, is unknown. When T_i is a multinomial latent variable (i.e., latent class variable), a reasonable substitute for the zero-one indicator variable $I\{T_i = t\}$ is $P(T_i = t | \mathbf{x}_i, \mathbf{u}_i)$, the posterior probability of membership in class t given the covariates $\mathbf{x}_i$ and the latent class indicator variables $\mathbf{u}_i$. Thus, we propose a generalized propensity score formula for latent class exposures:

$$\hat{\pi}_i = \sum_{t=1}^{n_T} P(T = t | \mathbf{x}_i) P(T = t | \mathbf{x}_i, \mathbf{u}_i). \tag{3}$$

This propensity score accounts for both the effects of the potential confounders and the unknown nature of class membership. Note that each individual still receives only one $\hat{\pi}_i$,

which is a weighted sum: the sum of the probabilities of membership in the exposure groups weighted by the posterior probabilities of membership in the exposure groups. Intuitively, we are broadening the concept of "the group to which he/she belongs" to incorporate imperfectly known group membership. The posterior probabilities $P(T = t | \mathbf{x}_i, \mathbf{u}_i)$ for each individual i and class t can be provided by any LCA software package.

When the exposure is latent, the β coefficients in Eq. 2 express the associations between covariates (i.e., potential confounders) and latent class membership; several methods exist to estimate these coefficients. Methods proposed to estimate associations between external variables and class membership include, among others, the unadjusted classify-analyze approach (also called a 3-step approach) and an adjusted classify-analyze approach (also called the

Table 3 Parameter estimates for the 4-class model of reasons for alcohol use

Latent class membership probabilities		1 Few Reasons 0.53	2 Enhancement Reasons 0.27	3 Coping Reasons 0.17	4 Diverse Reasons 0.04
	Overall	Item-response probabilities			
To have a good time with my friends	*0.80*	*0.68*	*0.97*	*0.86*	*0.98*
Because it tastes good	*0.55*	*0.53*	*0.58*	*0.51*	*0.73*
To relax or relieve tension	*0.52*	0.38	*0.51*	*0.87*	*0.91*
To feel good or get high	*0.51*	0.19	*0.98*	*0.66*	*0.96*
Because of boredom, nothing else to do	0.21	0.12	0.28	0.24	*0.87*
To get away from my problems or troubles	0.19	0.03	0.06	*0.76*	*0.74*
Because of anger or frustration	0.15	0.04	0.03	*0.57*	*0.70*
To fit in with a group I like	0.09	0.06	0.08	0.13	0.31
To get to sleep	0.08	0.05	0.05	0.13	0.36
To increase effects of other drug(s)	0.04	0.00	0.09	0.01	0.39
To seek deeper insights and understanding	0.04	0.02	0.03	0.06	0.25
To get through the day	0.01	0.00	0.00	0.01	0.18
To decrease/offset effects of other drug(s)	0.01	0.00	0.01	0.01	0.10
I am "hooked"—I feel I have to drink	0.01	0.00	0.00	0.00	0.12

Italic font indicates item-response probabilities > 0.50 to facilitate interpretation

BCH approach; Bolck et al. 2004; Vermunt 2010). The unadjusted classify-analyze approach is the traditional approach based on modal assignment where each individual is assigned to the latent class for which he/she has the highest posterior probability. Class membership is then treated as known (e.g., using dummy-coding) in a subsequent analysis model. This approach results in substantially attenuated β coefficients; subsequent approaches have been proposed as alternatives to reduce attenuation. A growing body of literature (e.g., Asparouhov and Muthén 2014; Bakk and Vermunt 2016; Bolck et al. 2004; Dziak et al. 2016) recommends the BCH approach for a wide array of applications. Similar to the unadjusted classify-analyze approach, the BCH approach is based on modal assignment and treatment of class membership as known in a subsequent analysis model. However, this analysis model is weighted to account for uncertainty in the modal assignments. Using the BCH approach prevents external variables from changing the meaning of the latent classes, accounts for uncertainty in modally assigned latent class membership and prevents attenuated β coefficients, and has been shown to perform accurately and robustly in simulations (see Bakk and Vermunt 2016 and Dziak et al. 2016 for accessible overviews of the BCH approach and comparisons to other approaches). Therefore, we propose the following approach to calculating propensity scores for latent class exposures that makes use of the BCH approach to estimate the β coefficients in Eq. 2. A simulation study examining the performance of this proposed approach is available in the online supplementary appendix and at methodology.psu.edu/downloads/appendices/lc-causal-exposures.

First, use the BCH approach to estimate the β coefficients for predicting latent class membership from the potential confounders. Second, substitute these estimated parameters and individuals' data into Eq. 2 and calculate the fitted probabilities for each individual, $P(T=t|\mathbf{x}_i)$, which are conditional on the potential confounders (i.e., covariates) but not the indicator variables. Third, calculate the estimated propensity score for each individual by combining these fitted probabilities with the posterior probabilities $P(T=t|\mathbf{x}_i, \mathbf{u}_i)$ (retained from step 1 above) using Eq. 3. After a propensity score $\hat{\pi}_i$ has been estimated for individual i, the inverse propensity score weight for individual i is calculated as $1/\hat{\pi}_i$. Because the MTF study design additionally involves attrition weights (Johnston et al. 2016), here denoted s_i, we calculated the final analysis weight for individual i as the product of the attrition weight and inverse propensity score weight, namely $s_i/\hat{\pi}_i$. Ridgeway et al. (2015) recommend the combined use of sampling weights and propensity weights in this way when both are required.

Step 3. Assess Overlap and Balance As with any other propensity score-weighted analysis, the quality of the weights needs to be assessed by examining overlap and balance. First, the degree of overlap in the estimated propensity scores between the exposure groups should be assessed. Poor propensity score overlap, where the range of propensity scores for individuals in one exposure group does not correspond to the range of propensity scores for individuals in another group, indicates the groups are too dissimilar to warrant causal inferences. Figure 1 uses boxplots to illustrate the overlap in inverse propensity score weights among the Few Reasons, Enhancement Reasons, and Coping Reasons latent classes that serve as the exposure groups of interest. Although there was not perfect overlap here, it was deemed adequate to warrant causal inferences, particularly for the ATE for Coping Reasons versus Enhancement Reasons.

Second, the improvement in balance between exposure groups after inverse propensity score weighting should be assessed. The goal of inverse propensity score weighting is to achieve balance; that is, to equate the exposure groups with respect to the distributions of the potential confounders. Balance is desirable because if the distributions of the potential confounders are equal, on average, across exposure groups, the groups may be compared directly. This is similar to the balance ideally achieved on potential confounders in a randomized controlled trial. Because these weights can only balance exposure groups with respect to included variables, it is important to include a comprehensive set of potential confounders, and sometimes interactions between them, in the propensity score model. Balance between exposure groups after weighting can be assessed by the standardized mean difference (SMD): the standardized difference in the means or proportions for a given potential confounder between two exposure groups (i.e., latent classes). SMD values close to 0 reflect that means or proportions between the groups are similar on average; SMD values less than 0.20 are typically considered indicative of good balance (Cohen 1988). SMDs on each potential confounder are compared between unweighted and inverse propensity score weighted samples.

In this example, because we had a latent multinomial exposure, to examine balance, we used the BCH approach to estimate the within-class means/proportions for each potential confounder (one at a time) and calculated SMD = (difference in mean or proportion between two classes)/(overall standard deviation among all individuals). Table 4 shows the means/proportions on each potential confounder for each latent class, as well as the SMD for the two ATEs of interest (i.e., Enhancement versus Few Reasons and Coping versus Enhancement Reasons). Binge drinking (i.e., drank 5+ drinks in a row in the past 2 weeks, 1 = none, …, 6 = 10+ times) showed the poorest balance (SMD = .42 for Enhancement versus Few Reasons) prior to inverse propensity score weighting. As shown in Table 4, balance on the potential confounders across all latent classes (i.e., exposures) was substantially improved after applying the inverse propensity score weights; adequate balance was achieved on all potential confounders for the ATEs of interest.
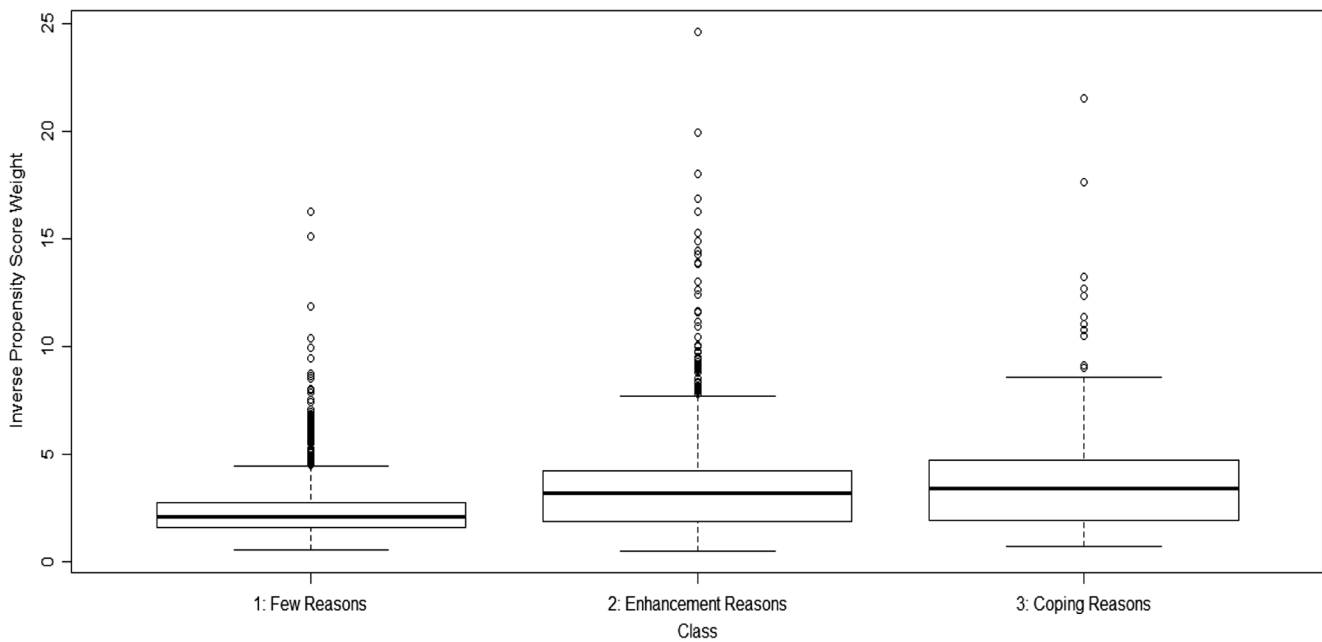
Fig. 1 Boxplots illustrating the overlap of the inverse propensity score weight distributions for the average exposure effect (ATE) of membership in the Enhancement Reasons versus Few Reasons and Coping Reasons versus Enhancement Reasons latent classes at age 19 on problem alcohol use vs. non-problem use/abstinence at age 35

Step 4. Conduct Outcome Analysis After confirming that overlap is adequate and balance is achieved with the inverse propensity score weights, the weighted outcome analysis to obtain the estimated ATEs is conducted. Typically, the outcome analysis to address the causal research question of interest is straightforward and involves a weighted regression model for the outcome on dummy-coded exposure group membership. Robust standard errors of effect estimates are recommended to account for the fact that the inverse propensity score weights are estimated. In this example, one appropriate outcome analysis is a weighted logistic regression model for problem alcohol use (versus non-problem use or abstinence) on dummy-coded latent class membership with the Few Reasons latent class coded as the reference.

We again used the BCH approach to obtain the necessary parameter estimates for predicting problem alcohol use at age 35 from latent class membership at age 19. In this example, we estimated both an unweighted ATE and a weighted ATE. The unweighted ATE included the attrition weights to adjust for the MTF study design only; the weighted ATE included the final analysis weight calculated in step 2 that combined the inverse propensity score weight with the MTF attrition weight. In both analyses, the weights were treated as survey sample weights and robust standard errors were requested.

Table 5 shows the unweighted and weighted effects of reasons for alcohol use latent class membership at age 19 on problem alcohol use at age 35, adjusting for potential confounders assessed at age 18. Even after adjusting for confounding, there was a significant causal effect of membership in the Enhancement Reasons latent class compared to membership in the Few Reasons latent class at age 19 on the chances of having problem alcohol use vs. non-problem use/abstinence at age 35. If all individuals in the population had Enhancement Reasons for alcohol use at age 19, an estimated 38% of them would have problem alcohol use at age 35, whereas if all individuals had Few Reasons at age 19, an estimated 22% of them would have problem alcohol use at age 35. That is, having Enhancement Reasons corresponds to being 1.17 times (odds ratio = $e^{0.16}$ = 1.17) more likely to have problem alcohol use compared to having Few Reasons. In comparison, those with Enhancement Reasons at age 19 were about equally likely to have problem alcohol use at age 35 as those with Coping Reasons (ATE = .07, $p > .05$).

Software Note We used Mplus (Muthén and Muthén 2015) to estimate the necessary models, and we used R (R Core Team 2015) to perform necessary calculations, such as calculating $P(T = t| \mathbf{x}_i)$ from Eq. 2 and $\hat{\pi}_i$ from Eq. 3. However, other software packages may be used. For example, SAS PROC LCA (Lanza et al. 2015) or Latent GOLD (Vermunt and Magidson 2015) could have been used instead of Mplus, and MathWorks' Matlab or Microsoft Excel could have been used instead of R. Annotated generic sample codes for Mplus and R are available in the online supplementary appendix and at methodology.psu.edu/downloads/appendices/lc-causal-exposures.

Table 4 Balance table showing means/proportions and standardized mean differences on potential confounders

Potential confounder	Unweighted						Weighted					
	Few Reasons	Enhancement Reasons	Coping Reasons	Diverse Reasons	SMD for Enhancement vs. Few ATE	SMD for Coping vs. Enhancement ATE	Few Reasons	Enhancement Reasons	Coping Reasons	Diverse Reasons	SMD for Enhancement vs. Few ATE	SMD for Coping vs. Enhancement ATE
Cohort	1986.33	1986.45	1987.18	1986.59	0.02	0.12	1986.59	1986.76	1986.94	1984.91	0.03	0.03
Gender	0.45	0.53	0.39	0.73	0.16	0.28	0.48	0.48	0.50	0.47	0.01	0.04
Race: Black	0.10	0.04	0.04	0.10	0.34	0.02	0.07	0.07	0.07	0.05	0.01	0.02
Hispanic	0.06	0.04	0.04	0.04	0.11	0.04	0.04	0.04	0.06	0.02	0.01	0.10
Other	0.06	0.03	0.05	0.07	0.16	0.13	0.05	0.05	0.04	0.06	0.02	0.02
Living in two-parent family	0.78	0.80	0.76	0.79	0.05	0.12	0.78	0.78	0.79	0.79	0.01	0.03
Parents' highest level of education	0.59	0.75	0.61	0.72	0.34	0.29	0.64	0.66	0.62	0.73	0.04	0.08
Average grades in high school	6.13	6.17	5.96	5.24	0.02	0.11	5.96	6.05	6.06	5.99	0.04	0.01
Took college prep curriculum	0.57	0.67	0.55	0.43	0.20	0.26	0.56	0.59	0.62	0.52	0.06	0.06
Definitely plans to graduate college	0.57	0.61	0.54	0.42	0.09	0.15	0.56	0.57	0.58	0.51	0.01	0.03
No. of days cut school in the past 30 days	1.51	1.88	1.84	2.27	0.29	0.03	1.79	1.78	1.77	1.73	0.01	0.00
Importance of religion	2.74	2.44	2.60	2.52	0.30	0.16	2.61	2.60	2.67	2.39	0.00	0.05
Average no. days per week going out	3.39	3.80	3.55	3.96	0.33	0.10	3.60	3.60	3.52	3.70	0.00	0.02
Smoked cigarette in the past 30 days	1.55	1.74	1.95	2.29	0.15	0.16	1.82	1.78	1.71	1.90	0.03	0.06
Drank 5+ drinks in past 2 weeks	1.47	2.00	1.87	2.53	0.42	0.08	1.81	1.85	1.82	1.85	0.03	0.09
Used marijuana in past 12 months	1.42	2.06	1.76	2.52	0.35	0.16	1.83	1.84	1.73	1.83	0.00	0.07
Used other substances in the past 12 months	0.16	0.31	0.28	0.51	0.32	0.20	0.26	0.28	0.24	0.34	0.03	0.06

SMD standardized mean difference (absolute value) in potential confounder between two latent classes; SMDs > 0.20 are italicized to facilitate interpretation of large differences. *ATE* average exposure effect; the ATEs of interest were membership in the Enhancement Reasons versus Few Reasons and in the Coping Reasons versus Enhancement Reasons latent classes at age 19 on problem alcohol use vs. non-problem use/abstinence at age 35. Attrition weights adjusting for the Monitoring the Future study design and BCH weights adjusting for classification error were used in both the unweighted and weighted analyses

Table 5 Average exposure effects (ATEs) of reasons for alcohol use latent class membership at age 19 on problem alcohol use at age 35

	Mean (SE)		ATE	SE	Odds ratio
	Few Reasons	Enhancement Reasons			
Unweighted	0.20 (0.01)	0.41 (0.02)	*0.21*	0.03	*1.24*
Weighted	0.22 (0.02)	0.38 (0.02)	*0.16*	0.03	*1.17*
	Enhancement Reasons	Coping Reasons			
Unweighted	0.41 (0.02)	0.47 (0.03)	0.06	0.04	1.06
Weighted	0.38 (0.02)	0.45 (0.04)	0.07	0.05	1.08

Bold font indicates a significant effect at $\alpha = 0.05$. The ATEs of interest were membership in the Enhancement Reasons versus Few Reasons and in the Coping Reasons versus Enhancement Reasons latent classes at age 19 on problem alcohol use vs. non-problem use/abstinence at age 35. Attrition weights adjusting for the Monitoring the Future study design and BCH weights adjusting for classification error were used in both the unweighted and weighted analyses

ATE average exposure effect; *SE* standard error

Discussion

As prevention scientists continue to use LCA to identify population subgroups, many interesting research questions about the causal effects of latent class exposures may arise. The step-by-step empirical example presented here illustrates one approach to estimating these effects. In particular, we used data from the US national MTF study to show that odds of problem alcohol use at age 35 would be about 17% higher if all individuals in the population had Enhancement Reasons for alcohol use at age 19 compared to if all individuals in the population had Few Reasons, but no significant differences were found between having Enhancement Reasons and Coping Reasons. That is, even after adjusting for a variety of characteristics, certain patterns of alcohol use reasons were significantly associated with elevated risk for problem alcohol use relative to other patterns. This provides evidence that alcohol use reasons are important risk factors for subsequent alcohol use problems and supports the idea that reasons for alcohol use represent important screening criteria and/or potential targets for prevention and intervention.

In our example, the unweighted and weighted results were quite similar. Although this was somewhat unexpected, effect estimates may increase, decrease, or stay the same after inverse propensity score weighting and this should not be seen as a reason to or not to use modern causal inference methods. Additionally, because of the similarities in results here, more straightforward approaches, such as traditional covariate adjustment, may seem like attractive alternatives. A large body of literature has discussed the limitations of approaches where some number of potential confounders are added directly to the outcome analysis in order to control for them (e.g., Dehejia and Wahba 2002; Rubin 2001). Notably, propensity score approaches allowed for the efficient, simultaneous adjustment of many more potential confounders are typically possible with traditional covariate adjustment. In addition, an important advantage of using propensity score approaches over traditional covariate adjustment is their usefulness in recognizing and diagnosing issues with the assumptions of overlap and balance, critical to avoiding inappropriate extrapolation (Austin 2011; Rubin 1997; Schafer and Kang 2008). An accessible discussion of the relative advantages and disadvantages of propensity score approaches and traditional covariate adjustment is available from Zanutto (2006).

The example presented here was intended primarily to demonstrate the approach and so the scope was somewhat narrow. For example, here we compared the effects of only three of the four identified latent classes. Also, additional potential confounders or additional effects of confounders (e.g., interactions between confounders) could have been added to the propensity score model to increase the strength of our causal inferences. In addition, our analyses focused on ATEs of membership in one latent class compared to another latent class. In the context of a latent multinomial exposure, an ATE focuses on the effect if everyone in the population belonged to one latent class (e.g., Enhancement Reasons) compared to a scenario in which everyone in the population belonged to another latent class (e.g., Few Reasons). However, there are a number of other causal effects that may be of interest, such as the average exposure effect among the exposed (ATT; i.e., average treatment effect among the treated; Austin 2011). The form of the weights depends on the causal effect of interest (e.g., McCaffrey et al. 2013).

Some aspects of the proposed approach require further research. We used casewise deletion for missing data on the potential confounders for simplicity and because our missing data were minimal, but other approaches should be considered. For example, Mplus provides a way to include missing data on covariates using full information maximum likelihood, assuming they are continuous-normal (Muthén and Muthén 2015). More generally, a straightforward alternative to listwise deletion is mean imputation (White and Thompson 2005), which is commonly used for propensity score estimation. More sophisticated approaches such as multiple

imputation (Schafer and Graham 2002; van Buuren 2007) could be considered, but would need to account for the necessary BCH weights. Additionally, how to use doubly robust methods (Bang and Robins 2005; Kang and Schafer 2007; Tan 2010; Zhang et al. 2016) is an open question.

Another topic for future research is the investigation of alternative approaches to estimating the propensity scores. We used multinomial logistic regression because (a) it is a natural extension of the commonly used logistic regression for binary exposures, (b) it is straightforward to implement with a latent class variable when the BCH approach is used, and (c) it is readily available in LCA software. However, when exposures are observed, alternatives to logistic regression can be advantageous, such as generalized boosted models (see McCaffrey et al. 2013). Using these alternatives with a latent exposure may require modification, in order to take into account the fact that only the posterior probabilities of each individual's class memberships are available, not the class memberships themselves. A related topic is whether the generalized propensity scores proposed here should be trimmed in cases where corresponding inverse propensity score weights are very large. In our simulation study, trimming the weights did not appear to improve the results when the BCH approach was used, and so we did not trim our weights in the current study. However, more work needs to be done in this area. Additionally, the newly proposed "overlap weights" (Li et al. 2016) have favorable practical and theoretical properties, and should be investigated in this context.

In conclusion, the step-by-step approach presented in the current study can be used to estimate ATEs of latent class exposures in many areas of prevention science using LCA to understand population subgroups. Integrating inverse propensity score weighting with LCA with distal outcomes provides a rigorous way to adjust for potential confounders when investigating the effects of latent class membership on outcomes. As increasingly complex research questions are posed about the roles played by latent class variables in development and prevention, the integration of these two methods provides a new tool for the identification of causal mechanisms leading to negative outcomes.

Compliance with Ethical Standards

Conflict of Interest   The authors declare that they have no conflicts of interest.

Ethical Approval   All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent   Where appropriate, informed consent and assent were obtained from all individual participants included in this study.

Disclaimer   The content is solely the responsibility of the authors and does not necessarily represent the official views of NCRG, NIDA, or the National Institutes of Health.

References

Asparouhov, T., & Muthén, B. (2014). Auxiliary variables in mixture modeling: Three-step approaches using Mplus. *Structural Equation Modeling, 21*, 329–341.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research, 46*, 399–424.

Bakk, Z., & Vermunt, J. K. (2016). Robustness of stepwise latent class modeling with continuous distal outcomes. *Structural Equation Modeling, 23*, 20–31.

Bang, H., & Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics, 61*, 962–972.

Bolck, A., Croon, M., & Hagenaars, J. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis, 12*, 3–27.

Bray, B. C., Almirall, D., Zimmerman, R. S., Lynam, D., & Murphy, S. A. (2006). Assessing the total effect of time-varying predictors in prevention research. *Prevention Science, 7*, 1–17.

Bray, B. C., Lanza, S. T., & Tan, X. (2015). Eliminating bias in classify-analyze approaches for latent class analysis. *Structural Equation Modeling, 22*, 1–11.

Cardoso, J. B., Goldbach, J. T., Cervantes, R. C., & Swank, P. (2016). Stress and multiple substance use behaviors among Hispanic adolescents. *Prevention Science, 17*, 208–217.

Coffman, D. L., & Zhong, W. (2012). Assessing mediation using marginal structural models in the presence of confounding and moderation. *Psychological Methods, 17*, 642–664.

Coffman, D., Patrick, M. E., Palen, L., Rhoades, B. L., & Ventura, A. (2007). Why do high school seniors drink? Implications for a targeted approach. Prevention Science, 8, 241–248.

Coffman, D. L., Caldwell, L. L., & Smith, E. A. (2012). Introducing the at-risk average causal effect with application to HealthWise South Africa. *Prevention Science, 13*, 437–447.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale: Laurence Erlbaum.

Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. New York: Wiley.

Dehejia, R. H., & Wahba, S. (2002). Propensity score matching methods for non-experimental causal studies. *Review of Economics and Statistics, 84*, 151–161.

Dziak, J. J., Bray, B. C., Zhang, J., Zhang, M., & Lanza, S. T. (2016). Comparing the performance of improved classify-analyze

approaches for distal outcomes in latent profile analysis. *Methodology, 12*, 107–116.

Evans-Polce, R. J., Patrick, M. E., & Miech, R. (2017). Patterns of reasons for vaping in a national sample of adolescent vapers. Paper presented at the Society for Prevention Research 25th Annual Meeting: "Prevention and Public Systems of Care: Research, Policy and Practice," Washington.

Gilreath, T. D., Astor, R. A., Estrada Jr, J. N., Benbenishty, R., & Unger, J. B. (2014). School victimization and substance use among adolescents in California. Prevention Science, 15, 897–906.

Green, K. M., & Stuart, E. A. (2014). Examining moderation analyses in propensity score methods: Application to depression and substance use. *Journal of Consulting and Clinical Psychology, 82*, 773–783.

Héroux, M., Janssen, I., Lee, D. C., Sui, X., Hebert, J. R., & Blair, S. N. (2012). Clustering of unhealthy behaviors in the aerobics center longitudinal study. *Prevention Science, 13*, 183–195.

Imbens, G. (1999). The role of the propensity score in estimating dose-response functions (Tech. Work. Paper No. 237). Cambridge: National Bureau of Economic Research. Retrieved from https://www.nber.org/papers/t0237.pdf.

Jiang, L., Beals, J., Zhang, L., Mitchell, C. M., Manson, S. M., Acton, K. J., ... & Special Diabetes Program for Indians Demonstration Projects. (2012). Latent class analysis of stages of change for multiple health behaviors: Results from the Special Diabetes Program for Indians diabetes prevention program. Prevention Science, 13, 449–461.

Johnston, L. D., O'Malley, P. M., Bachman, J. G., Schulenberg, J. E., & Miech, R. A. (2016). *Monitoring the Future national survey results on drug use, 1975–2015: Volume 2, college students and adults ages 19–55*. Ann Arbor: Institute for Social Research, The University of Michigan.

Kang, J. D. Y., & Schafer, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data (with discussion and rejoinder). *Statistical Science, 22*, 523–539.

Lanza, S. T., & Rhoades, B. L. (2013). Latent class analysis: An alternative perspective on subgroup analysis in prevention and treatment. *Prevention Science, 14*, 157–168.

Lanza, S. T., Bray, B. C., & Collins, L. M. (2013a). An introduction to latent class and latent transition analysis. In J. A. Schinka, W. F. Velicer, & I. B. Weiner (Eds.), *Handbook of psychology* (Vol. 2, 2nd ed., pp. 691–716). Hoboken: Wiley.

Lanza, S. T., Moore, J. E., & Butera, N. M. (2013b). Drawing causal inferences using propensity scores: A practical guide for community psychologists. *American Journal of Community Psychology, 52*, 380–392.

Lanza, S. T., Coffman, D. L., & Xu, S. (2013c). Causal inference in latent class analysis. *Structural Equation Modeling, 20*, 361–383.

Lanza, S. T., Tan, X., & Bray, B. C. (2013d). Latent class analysis with distal outcomes: A flexible model-based approach. *Structural Equation Modeling, 20*, 1–26.

Lanza, S. T., Dziak, J. J., Huang, L., Wagner, A., & Collins, L. M. (2015). PROC LCA & PROC LTA users' guide (Version 1.3.2). University Park: The Methodology Center, Penn State. Retrieved from http://methodology.psu.edu

Lanza, S. T., Schuler, M. S., & Bray, B. C. (2016). Latent class analysis with causal inference: The effect of adolescent depression on young adult substance use profile (Chp. 16, pp. 385–404). In W. Wiedermann & A. von Eye (Eds.), Causality and statistics. Hoboken: Wiley.

Li, F., Lock Morgan, K., & Zaslavsky, A. M. (2016). Balancing covariates via propensity score weighting. *Journal of the American Statistical Association. Advance online publication*. https://doi.org/10.1080/01621459.2016.1260466.

Low, S., Smolkowski, K., & Cook, C. (2016). What constitutes high-quality implementation of SEL programs? A latent class analysis of second step® implementation. *Prevention Science, 17*, 981–991.

Lunceford, J. K., & Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study. *Statistics in Medicine, 23*, 2937–2960.

McCaffrey, D. F., Griffin, B. A., Almirall, D., Slaughter, M. E., Ramchand, R., & Burgette, L. F. (2013). A tutorial on propensity score estimation for multiple treatments using generalized boosted models. *Statistics in Medicine, 32*, 3388–3414.

Merrill, J. E., Wardell, J. D., & Read, J. P. (2014). Drinking motives in the prospective prediction of unique alcohol-related consequences in college students. *Journal of Studies on Alcohol and Drugs, 75*, 93–102.

Miech, R. A., Johnston, L. D., O'Malley, P. M., Bachman, J. G., Schulenberg, J. E., & Patrick, M. E. (2017). *Monitoring the Future national survey results on drug use, 1975–2016: Volume I, secondary school students*. Ann Arbor, MI: Institute for Social Research, The University of Michigan.

Muthén, L.K. and Muthén, B.O. (2015). Mplus User's guide (7th ed.) Los Angeles, CA: Muthén & Muthén.

Patrick, M. E., & Schulenberg, J. E. (2011). How trajectories of reasons for alcohol use relate to trajectories of binge drinking: National panel data spanning late adolescence to early adulthood. *Developmental Psychology, 47*, 311–317.

Patrick, M. E., Schulenberg, J. E., O'Malley, P. M., Johnston, L., & Bachman, J. (2011). Adolescents' reported reasons for alcohol and marijuana use as predictors of substance use and problems in adulthood. *Journal of Studies on Alcohol and Drugs, 72*, 106–116.

Patrick, M. E., Bray, B. C., & Berglund, P. (2016). Reasons for marijuana use among young adults and long-term associations with marijuana use and problems. *Journal on Studies of Alcohol and Drugs, 77*, 881–888.

Patrick, M. E., Evans-Polce, R., Kloska, D. D., Maggs, J. L., & Lanza, S. T. (2018). Age-related changes in associations between reasons for alcohol use and high-intensity drinking across young adulthood. Journal of Studies on Alcohol and Drugs.

R Core Team (2015). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org.

Ridgeway, G., Kovalchik, S. A., Griffin, B. A., & Kabeto, M. U. (2015). Propensity score analysis with survey weighted data. *Journal of Causal Inference, 3*, 237–249.

Robins, J. M., Hérnan, M. A., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology, 11*, 550–560.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*, 41–55.

Rubin, D. B. (1997). Estimating causal effects from large data sets using propensity scores. *Annals of Internal Medicine, 127*, 757–763.

Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services & Outcomes Research Methodology, 2*, 169–188.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147–177.

Schafer, J. L., & Kang, J. (2008). Average causal effects from nonrandomized studies: A practical guide and simulated example. *Psychological Methods, 13*, 279–313.

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, 9*(Suppl 1), 41–57.

Schulenberg, J. E., Johnston, L. D., O'Malley, P. M., Bachman, J. G., Miech, R. A., & Patrick, M. E. (2017). *Monitoring the*

Future national survey results on drug use, 1975–2016: Volume II, college students and adults ages 19–55. Ann Arbor: Institute for Social Research, The University of Michigan.

Schuler, M. S. (2013). *Estimating the relative treatment effects of natural clusters of adolescent substance abuse treatment services: Combining latent class analysis and propensity score methods. Unpublished doctoral dissertation.* Baltimore: Johns Hopkins University Retrieved from https://jscholarship.library.jhu.edu/bitstream/handle/1774.2/36988/SCHULER-DISSERTATION-2014.pdf.

Schuler, M. S., Leoutsakos, J. S., & Stuart, E. A. (2014). Addressing confounding when estimating the effects of latent classes on a distal outcome. *Health Services Outcomes and Research Methodology, 14,* 232–254.

Spilt, J. L., Koot, J. M., & van Lier, P. A. (2013). For whom does it work? Subgroup differences in the effects of a school-based universal prevention program. *Prevention Science, 14,* 479–488.

Stapinski, L. A., Edwards, A. C., Hickman, M., Araya, R., Teesson, M., Newton, N. C., et al. (2016). Drinking to cope: A latent class analysis of coping motives for alcohol use in a large cohort of adolescents. *Prevention Science, 17,* 584–594.

Stapleton, J. L., Turrisi, R., Cleveland, M. J., Ray, A. E., & Lu, S. E. (2014). Pre-college matriculation risk profiles and alcohol consumption patterns during the first semesters of college. *Prevention Science, 15,* 705–715.

Tan, Z. (2010). Bounded, efficient and doubly robust estimation with inverse weighting. *Biometrika, 97,* 661–682.

Van Buuren, S. (2007). Multiple imputation of discrete and continuous data by fully conditional specification. *Statistical Methods in Medical Research, 16,* 219–242.

Varvil-Weld, L., Crowley, D. M., Turrisi, R., Greenberg, M. T., & Mallett, K. A. (2014). Hurting, helping, or neutral? The effects of parental permissiveness toward adolescent drinking on college student alcohol use and problems. *Prevention Science, 15,* 716–724.

Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis, 18,* 450–469.

Vermunt, J. K., & Magidson, J. (2015). *Upgrade manual for Latent GOLD 5.1.* Belmont, MA: Statistical Innovations.

White, I. R., & Thompson, S. G. (2005). Adjusting for partially missing baseline measurements in randomized trials. *Statistics in Medicine, 24,* 993–1007.

Yamaguchi, K. (2015). Extensions of Rubin's causal model for a latent-class treatment variable: An analysis of the effects of employers' work-life balance policies on women's income attainment in Japan. Research Institute of Economy, Trade and Industry Discussion Paper Series (No. 15-E-090). Tokyo, Japan: The Research Institute of Economy, Trade and Industry. Retrieved from http://www.rieti.go.jp/jp/publications/dp/15e090.pdf.

Zanutto, E. L. (2006). A comparison of propensity score and linear regression analysis of complex survey data. *Journal of Data Science, 4,* 67–91.

Zhang, Z., Liu, W., Zhang, B., Tang, L., & Zhang, J. (2016). Causal inference with missing exposure information: Methods and applications to an obstetric study. *Statistical Methods in Medical Research, 25,* 2053–2066.