MULTIVARIATE BEHAVIORAL RESEARCH
2025, VOL. 60, NO. 1, 30–60

Causal Latent Class Analysis with Distal Outcomes: A Modified Three-Step Method Using Inverse Propensity Weighting

Trà T. Lê*, Felix J. Clouth*, and Jeroen K. Vermunt

Department of Methodology and Statistics, Tilburg University, Tilburg, The Netherlands

ABSTRACT
Bias-adjusted three-step latent class (LC) analysis is a popular technique for estimating the relationship between LC membership and distal outcomes. Since it is impossible to randomize LC membership, causal inference techniques are needed to estimate causal effects leveraging observational data. This paper proposes two novel strategies that make use of propensity scores to estimate the causal effect of LC membership on a distal outcome variable. Both strategies modify the bias-adjusted three-step approach by using propensity scores in the last step to control for confounding. The first strategy utilizes inverse propensity weighting (IPW), whereas the second strategy includes the propensity scores as control variables. Classification errors are accounted for using the BCH or ML corrections. We evaluate the performance of these methods in a simulation study by comparing it with three existing approaches that also use propensity scores in a stepwise LC analysis. Both of our newly proposed methods return essentially unbiased parameter estimates outperforming previously proposed methods. However, for smaller sample sizes our IPW based approach shows large variability in the estimates and can be prone to non-convergence. Furthermore, the use of these newly proposed methods is illustrated using data from the LISS panel.

KEYWORDS
Bias-adjusted three-step latent class analysis; distal outcomes; propensity score; causal inference

Introduction

Latent class (LC) analysis is a statistical technique used to classify individuals into unknown groups based on their responses to a set of observed indicators (Goodman, 1974; Lazarsfeld & Henry, 1968; Vermunt & Magidson, 2004). It is widely adopted in the social and behavioral sciences, and recently, becoming more and more popular in medical research, as many constructs of interest in these fields cannot be observed directly. Furthermore, researchers are often interested in investigating how class membership is related to external variables. i.e., covariates can be used to predict class membership or, in turn, class membership can be used to predict distal outcomes. Different methods have been developed to facilitate these types of research questions: a one-step approach (Bandeen-Roche et al., 1997; Dayton & Macready, 1988; Kamakura et al., 1994; Yamaguchi, 2000), a two-step approach (Bakk & Kuha, 2018), and a three-step approach (Bakk et al., 2013; Bolck et al., 2004; Vermunt, 2010). Here, we will focus on the three-step approach. Bias-adjusted three-step LC analysis consists of 1) estimating a standard LC model without covariates, 2) assigning subjects to classes using proportional or modal assignment, and 3) estimating the structural relations between latent classes, covariates, and distal outcomes while accounting for classification errors using the BCH or ML correction (Bakk et al., 2013).

LC modeling is commonly used for clustering individuals and investigating the relationship between the latent classes and distal outcomes. For instance, LC analysis has been used to investigate the effect of reasons for alcohol use classes on later problem alcohol use (Bray et al., 2019), identify profiles of substance use disorders and investigate their effect on criminal recidivism (Schmitter et al., 2021), or investigate the effect of comorbid pattern classes on breast cancer risk (Dalmartello et al., 2022). However, identifying causal relationships between latent classes and distal outcomes is challenging. Since the classes are by definition unobserved, it is impossible to randomize individuals into the latent classes and observational data needs to be leveraged to identify causal effects. As a result, individuals in different classes might differ systematically from each other on baseline characteristics resulting in individuals from these classes to be non-exchangeable and the latent class—distal outcome relationship to be confounded (for a discussion on exchangeability in observational data, see Hernán and Robins (2006)). Several causal inference techniques have been proposed to adjust for confounders and estimate causal effects with observational data. Particularly the use of inverse propensity weighting (IPW) (Austin, 2011; Imbens, 2004; Robins et al., 1994) has been proposed for estimating causal effects of latent classes on distal outcomes (Bray et al., 2019; Schuler et al., 2014; Yamaguchi, 2015).

IPW solves the issue of confounding in two steps. First, the propensity scores are estimated and weights based on the inverse of the propensity scores are constructed. Second, the effect of exposure on the outcome is estimated using the weights from the first step in the estimation procedure (Austin, 2011). Propensity scores reflect each individual's probability of exposure conditional on that individual's values on the confounding variables. As such, the propensity score reduces each individual's set of covariates to a single score (Robins et al., 2000; Rosenbaum & Rubin, 1983). The propensity score can be estimated using, e.g., logistic regression. Next, each individual can be weighted with an individual weight based on the inverse of the propensity score (Austin, 2011; Imbens, 2004). As such, an individual with a low probability of exposure that actually was exposed (hence, a combination that is rather uncommon in the data) will be up-weighted while an individual with a high probability of exposure that actually was exposed (hence, a common combination) will be down-weighted. The average treatment effect (ATE) on the distal outcome can then be estimated in a weighted analysis. Here, we consider the ATE as the average difference in the distal outcome when the entire population is moved from receiving the treatment to not receiving it (Austin, 2011; Imbens, 2004). However, different definitions for the causal effect such as the average treatment effect on the treated (ATT) are possible resulting in different formalizations of the weights (Austin, 2011; Imbens, 2004). Regardless of the definition of the causal effect, IPW has a conceptual advantage over other adjustment methods because it separates the confounding adjustment from the causal effect estimation. For instance, this allows for assessing the correct specification of the propensity score model independently from the causal effect estimation (Austin, 2011).

Using IPW to estimate the ATE of latent class membership on a distal outcome has been attempted before (Bray et al., 2019; Schuler et al., 2014; Yamaguchi, 2015). All of these approaches use stepwise LC analysis to include the IPW as fixed weights when estimating the ATE. However, they differ in the estimation methods for the IPW and the ATE. Schuler et al. (2014) first proposed a framework, in which the propensity scores are obtained from a standard three-step LC analysis with the confounders as covariates. Then, the IPW are included as fixed weights in the final step to estimate the ATE. In both steps however, the authors did not consider that classification errors need to be accounted for when using stepwise LC analysis. Essentially, estimated class membership was used here as true class membership, without accounting for the probability of classification errors which results in biased estimates for the ATE. Yamaguchi (2015) obtained the propensity scores from a one-step LC analysis instead and used the three-step method for estimating the ATE. However, also here the classification errors are not accounted for in the last step. In both approaches, the propensity scores are class-specific, that is, each individual receives a propensity score for each class. This is a relevant difference to cases where the exposure is also multi-categorical but observed as then only one weight corresponding to the realized exposure category is assigned (Imbens, 2004; McCaffrey et al., 2013). In contrast, latent class membership does not reflect a hard partitioning of exposure categories. Recently, Bray et al. (2019) proposed an approach that takes the classification errors into account. In both of the steps to estimate the propensity scores and the ATE, classification errors were accounted for using the BCH correction (Bolck et al., 2004; Vermunt, 2010). However, each individual was assigned only one propensity score (and thus, one weight), which is the weighted average of the propensity scores for all classes.

In this paper, we propose two novel approaches to include propensity scores in an LC analysis with distal outcomes by modifying the last step of the bias-adjusted three-step approach (Bakk et al., 2013; Vermunt, 2010). That is, 1) estimating the LC model of interest without the outcome or control variables, 2) assigning subjects to classes using proportional or modal assignment, and 3) estimating the ATE with either class-specific IPW as weights or propensity scores as control variables while accounting for classification errors using the BCH correction. The third step involves obtaining the propensity scores from an additional bias-adjusted three-step LC analysis with the confounders used as covariates. While our IPW strategy build upon previous work using a similar strategy, our proposal for using propensity scores as controls in a bias-adjusted three-step LC analysis is completely new. The paper is structured as follows: first, we present a review of the existing methods; second, our alternative approaches are described step by step in detail; third, a simulation study is conducted to evaluate their performance in comparison with the other methods; fourth, we illustrate our newly proposed methods on data from the LISS (Longitudinal Internet studies for the Social Sciences) panel; last, the paper is ended with a discussion and recommendations section.

A General model to include IPW in LC analysis with a distal outcome

This section reviews the different approaches to include IPW in an LC analysis with distal outcomes and proposes our alternative strategy using class-specific IPW with the BCH correction method. The three existing methods reviewed in this section can be summarized into a general three-step model as follows:

1. Estimate an LC model based on the indicator variables $\boldsymbol{Y}_i$. This first step might be done including the confounders $\boldsymbol{C}_i$ as covariates as in Yamaguchi (2015):

$$P(\boldsymbol{Y}_i|\boldsymbol{C}_i) = \sum_{t=1}^{T} P(X=t|\boldsymbol{C}_i)P(\boldsymbol{Y}_i|X=t),$$

   or excluding the confounders as in Schuler et al. (2014) and Bray et al. (2019):

$$P(\boldsymbol{Y}_i) = \sum_{t=1}^{T} P(X=t)P(\boldsymbol{Y}_i|X=t),$$

   with $t = 1, \ldots, T$ being the realization of the latent classes X.
2. Assign to each subject i class-assignment weights w_{is}, with $s = 1, \ldots, T$ being the realization of class assignment W. Using modal assignment, w_{is} is 1 if $P(X = s|\boldsymbol{Y}_i)$ is largest and 0 otherwise. Using proportional assignment, w_{is} is equivalent to $P(X = s|\boldsymbol{Y}_i)$.
3. Estimate the effect of class memberships on the distal outcome with IPW as fixed weights to control for confounding.
   a. When confounders were not included in Step 1, estimate the propensity scores as the class membership probabilities conditional on the set of confounders: $\hat{\pi}_{it} = P(X=t|\boldsymbol{C}_i)$ making use of the class assignments W.
   b. Compute the weights ipw_{it} as the inverse of the propensity scores; that is, $ipw_{it} = 1/P(X=t|\boldsymbol{C}_i)$.
   c. Estimate the ATE of class membership on the distal outcome Z_i with ipw_{it} as fixed weights.

Correcting for classification errors in Step 3 (a), involves estimating this model using the class assignments W:

$$P(W=s|\boldsymbol{C}_i) = \sum_{t=1}^{T} P(X=t|\boldsymbol{C}_i)P(W=s|X=t). \quad (1)$$

Regardless of the ML correction or the BCH correction being used, Step 3 (a) and (c) rely on obtaining the classification error probabilities as follows (Bakk et al., 2013):

$$P(W=s|X=t) = \frac{\frac{1}{N}\sum_{i=1}^{N} P(X=t|\boldsymbol{Y}_i)w_{is}}{P(X=t)}. \quad (1)$$

The BCH correction uses the elements of the inverse of the matrix with elements $P(W = s|X = t)$ (also referred to as the D^{-1} matrix) as weights, which we denote by d_{st}. Representing the class-specific density of the outcome variable by $f(Z_i|X=t)$, in Step 3 (c), the ATE can be estimated by maximizing the following pseudo-log-likelihood function:

$$\begin{aligned}\log L_{IPW} &= \sum_{i=1}^{N}\sum_{t=1}^{T} ipw_{it}\left(\sum_{s=1}^{T} d_{st}\cdot w_{is}\right) \log f(Z_i|X=t) \\ &= \sum_{i=1}^{N}\sum_{t=1}^{T} ipw_{it}\cdot w_{it}^{*} \log f(Z_i|X=t) \\ &= \sum_{i=1}^{N}\sum_{t=1}^{T} w_{it}^{**} \log f(Z_i|X=t) \quad (3)\end{aligned}$$

where w_{it}^{*} is the class-specify weight used in the standard BCH estimation. When multiplying this weight by ipw_{it} we get a new weight w_{it}^{**}. So, in fact, using IPW in a 3-step LC analysis simply involves using a modified set of weights. The ATE is obtained by comparing the estimated expected value of Z_i across latent classes.

As summarized in Table 1, the IPW based methods proposed so far differ in (1) the class assignment rule, (2) the correction for classification errors, and 3) the way propensity scores are obtained and used to construct IPW; that is, in the definition of w_{is}, d_{st}, and ipw_{it}. More specifically, Schuler et al. (2014) estimate the LC model including only the response indicators and assign individuals to classes using modal

Table 1. Summary of all methods considered in this study.

	Step 1: Estimation LC model	Step 2: Classification	Step 3a: Estimation PS	Step 3b: Construction IPWs	Step 3c: Estimation ATE
New 3-Step with IPW	LCA without covariates: $\log L = \log \sum_{i=1}^{N} \sum_{t=1}^{T} P(X=t)P(Y_i \mid X=t)$	Either modal: $w_{is} = 1$ if $P(X=s \mid Y_i)$ is largest and 0 otherwise, or proportional: $w_{is} = P(X=s \mid Y_i)$	$\hat{\pi}_{it} = P(X=t \mid C_i)$ Step-3 analysis with ML bias adjustment: $\log L_{ML} = \sum_{i=1}^{N} \sum_{s=1}^{T} w_{is} \log \sum_{t=1}^{T} P(X=t \mid C_i) P(W=s \mid X=t).$	Class specific IPWs for true classes: $ipw_{it} = 1/\hat{\pi}_{it}$	Step-3 analysis with BCH bias adjustment: $\log L_{IPW} = \sum_{i=1}^{N} \sum_{t=1}^{T} ipw_{it} \left(\sum_{s=1}^{T} d_{st} \cdot w_{is} \right) \log f(Z_i \mid X=t)$
New 3-Step with PS as Covariate	LCA without covariates: $\log L = \log \sum_{i=1}^{N} \sum_{t=1}^{T} P(X=t)P(Y_i \mid X=t)$	Either modal: $w_{is} = 1$ if $P(X=s \mid Y_i)$ is largest and 0 otherwise, or proportional: $w_{is} = P(X=s \mid Y_i)$	$\hat{\pi}_{it} = P(X=t \mid C_i)$ Step-3 analysis with ML bias adjustment: $\log L_{ML} = \sum_{i=1}^{N} \sum_{s=1}^{T} w_{is} \log \sum_{t=1}^{T} P(X=t \mid C_i) P(W=s \mid X=t)$	No IPWs needed	Step-3 analysis with ML or BCH bias adjustment: $\log L_{ML} = \sum_{i=1}^{N} \sum_{s=1}^{T} w_{is} \log \sum_{t=1}^{T} P(X=t) f(Z_i \mid X=t, \hat{\pi}_{it}) P(W=s \mid X=t)$ $\log L_{BCH} = \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\sum_{s=1}^{T} d_{st} \cdot w_{is} \right) \log f(Z_i \mid X=t, \hat{\pi}_{it})$
Bray et al. (2019)	LCA without covariates: $\log L = \log \sum_{i=1}^{N} \sum_{t=1}^{T} P(X=t)P(Y_i \mid X=t)$	Modal: $w_{is} = 1$ if $P(X=s \mid Y_i)$ is largest and 0 otherwise	$\hat{\pi}_i = \sum_{t=1}^{T} P(X=t \mid C_i) P(X=t \mid Y_i)$ Step-3 analysis with BCH bias adjustment: $\log L_{BCH} = \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\sum_{s=1}^{T} d_{st} \cdot w_{is} \right) \log P(X=t \mid C_i)$	Single IPW based on weighted sum of class-specific PSs: $ipw_i = 1/\hat{\pi}_i$	Step-3 analysis with BCH bias adjustment: $\log L_{IPW} = \sum_{i=1}^{N} \sum_{t=1}^{T} ipw_i \left(\sum_{s=1}^{T} d_{st} \cdot w_{is} \right) \log f(Z_i \mid X=t)$
Yamaguchi (2015)	LCA including covariates: $\log L = \log \sum_{i=1}^{N} \sum_{t=1}^{T} P(X=t \mid C_i) P(Y_i \mid X=t)$	Proportional: $w_{is} = P(X=s \mid Y_i, C_i)$	$\hat{\pi}_{it} = P(X=t \mid C_i)$ Available from Step 1	Class-specific IPWs for assigned classes: $ipw_{is} = 1/\hat{\pi}_{is}$	Step-3 analysis without bias adjustment: $\log L_{IPW} = \sum_{i=1}^{N} \sum_{s=1}^{W} ipw_{is} \cdot w_{is} \log f(Z_i \mid W=s)$
Schuler et al. (2014)	LCA without covariates: $\log L = \log \sum_{i=1}^{N} \sum_{t=1}^{T} P(X=t)P(Y_i \mid X=t)$	Modal: $w_{is} = 1$ if $P(X=s \mid Y_i)$ is largest and 0 otherwise	$\hat{\pi}_{is} = P(W=s \mid C_i)$ Step-3 analysis without bias adjustment: $\log L = \sum_{i=1}^{N} \sum_{s=1}^{W} w_{is} P(W=s \mid C_i)$	Class-specific IPWs for assigned classes: $ipw_{is} = 1/\hat{\pi}_{is}$	Step-3 analysis without bias adjustment: $\log L_{IPW} = \sum_{i=1}^{N} \sum_{s=1}^{W} ipw_{is} \cdot w_{is} \log f(Z_i \mid W=s)$

assignment. Hence, w_{is} is 1 for the class with the highest posterior membership probability and 0 for the other classes. Their class-specific propensity scores $\hat{\pi}_{is}$ are obtained in Step 3 (a) using a naive 3-step LC analysis with covariates, in which classification errors are not accounted for. Hence, the propensity scores are in fact estimated using the assigned class membership W instead of true class membership X (Figure 1.1) resulting in $ipw_{is} = 1/\hat{\pi}_{is}$. In Step 3 (c), the ATE is estimated with ipw_{is} as fixed weights, without taking into account the classification errors. Thus, d_{st} is an element of the identity matrix.

In Yamaguchi (2015), the propensity scores are obtained directly from Step 1 since the LC model is estimated using both the response variables and the confounders (Figure 1.2). Thus, $\hat{\pi}_{it}$ is estimated from a one-step instead of a three-step approach. In Step 2, proportional assignment is used instead of modal assignment, so w_{is} is the posterior membership probabilities retained from Step 1. As in Schuler et al. (2014), classification errors are not accounted for implying that the d_{st} are the elements of the identity matrix.

Bray et al. (2019) address the issue of not accounting for classification errors in Schuler et al. (2014) and Yamaguchi (2015). Step 1 and 2 are similar to Schuler et al. (2014), that is, the LC model is estimated using only the response indicators and modal assignment is used, so w_{is} is either 1 or 0. The difference in their method lies in Step 3. In Step 3 (a), the propensity score is obtained from a bias-adjusted three-step LC analysis with confounders serving as covariates using the BCH method to account for the classification errors (Figure 1.3). A generalized formula to estimate the propensity score was proposed as follows[1]:

$$\hat{\pi}_i = \sum_{t=1}^{T} P(X = t|\mathbf{C}_i)P(X = t|\mathbf{Y}_i). \tag{4}$$

Here, the propensity score is the weighted average of $P(X = t|\mathbf{C}_i)$ over all classes, with the posterior membership probability retained from Step 1 as weights. Each subject is then only assigned one propensity score $\hat{\pi}_i$, and thus, one weight $ipw_i = 1/\hat{\pi}_i$, which contrary to Schuler et al. (2014) and Yamaguchi (2015) is no longer class-specific. In Step 3 (c), ipw_i is included as weights to estimate the ATE, taking into account the classification errors using the BCH method. Hence, d_{st} is an element of the D^{-1} matrix.

A New bias-adjusted three-step LC analysis using IPW

Upon reviewing the three existing methods, this paper proposes a new strategy. Here, we modify the final step of the bias-adjusted three-step LC analysis with a distal outcome proposed by Bakk et al. (2013) by including the ipw_{it} as fixed weights. It can also be seen as a combination of the strengths of the other methods: we keep the ipw_{it} class-specific as in Schuler et al. (2014) and Yamaguchi (2015) (though with different model specifications for estimating the propensity score), while accounting for classification errors in both Step 3 (a) (obtaining propensity score) and Step 3 (c) (estimating the ATE) using the BCH correction method as in Bray et al. (2019). Visually, our new method can be depicted identically to Figure 1.3. This section describes our strategy in detailed steps.

In Step 1, an LC model is estimated based on the observed indicators without confounders. In Step 2, subjects are assigned to classes using either proportional or modal assignment. In Step 3 (a), the propensity score is obtained from a bias-adjusted three-step LC analysis with covariates, using the ML correction method. Here, unlike in Bray et al. (2019), the ML method is used since it has been shown to be more efficient than the BCH method when the third step involves covariates (Vermunt, 2010). Thus, the propensity scores can be estimated by maximizing the following log-likelihood:

$$\log L_{ML} = \sum_{i=1}^{N}\sum_{s=1}^{T} w_{is}\log\sum_{t=1}^{T} P(X = t|C_i)P(W = s|X = t). \tag{5}$$

Finally, these class-specific ipw_{it} are used as fixed weights when estimating the ATE by maximizing the pseudo log-likelihood function in Equation 3. Similar to Bray et al. (2019), the BCH correction is used to account for the classification errors, hence, d_{st} is the element of the D^{-1} matrix. Note that robust standard errors should be used here for the BCH method (or when using proportional assignment regardless of the correction methods).

There is a particular reason for the choice of the BCH correction method in our final step. The BCH method transforms the data to represent the true classes by using weights (the inverse of the

[1]Note that Bray and colleagues report Equation (4) as $\hat{\pi}_i = \sum_{t=1}^{T} P(X = t|\mathbf{C}_i)P(X = t|\mathbf{Y}_i, \mathbf{C}_i)$. Thus, they propose that the posterior class membership probabilities that are used to weight the propensity scores need to be updated in light of the covariates. However, in the provided computer code the authors use $P(X = t|\mathbf{Y}_i)$ instead of $P(X = t|\mathbf{Y}_i, \mathbf{C}_i)$ which seems to be a more reasonable approach.

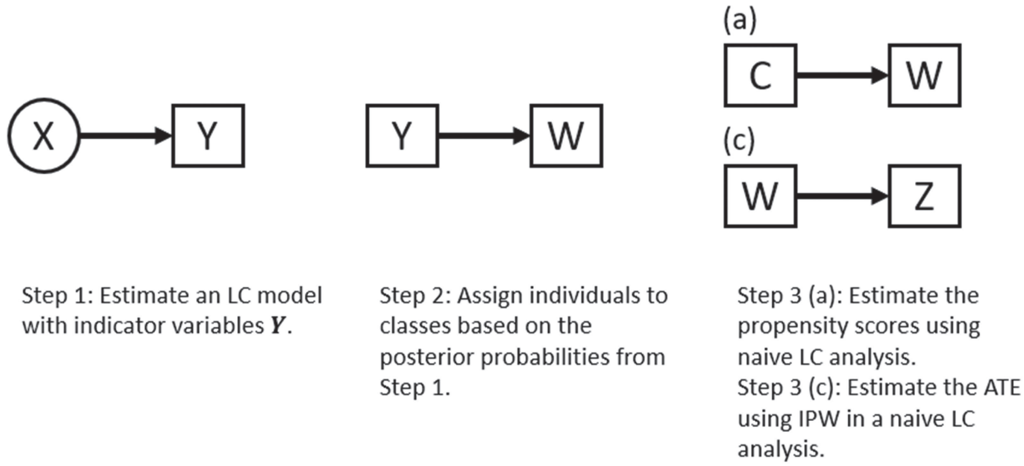

Figure 1.1. LC analysis using IPW as proposed by Schuler et al. (2014). Note that Step 3 (b): Constructing class-specific weights for each individual based on the inverse of the propensity scores is not shown.

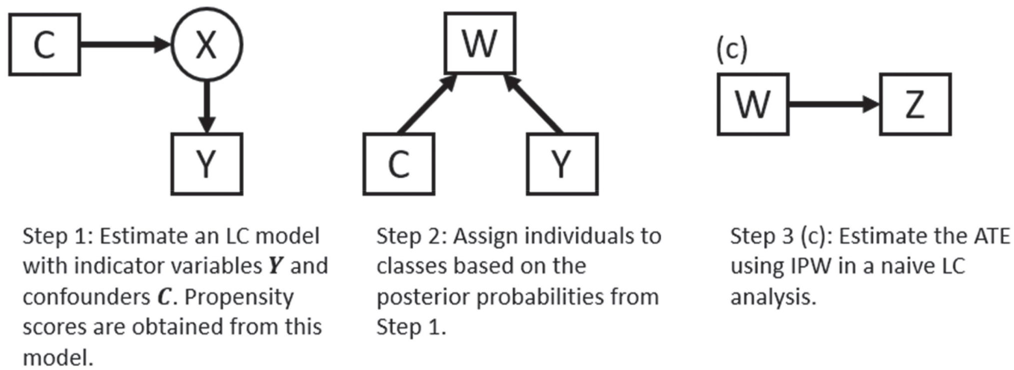

Figure 1.2. LC analysis using IPW as proposed by Yamaguchi (2015). Note that Step 3 (b): Constructing class-specific weights for each individual based on the inverse of the propensity scores is not shown.

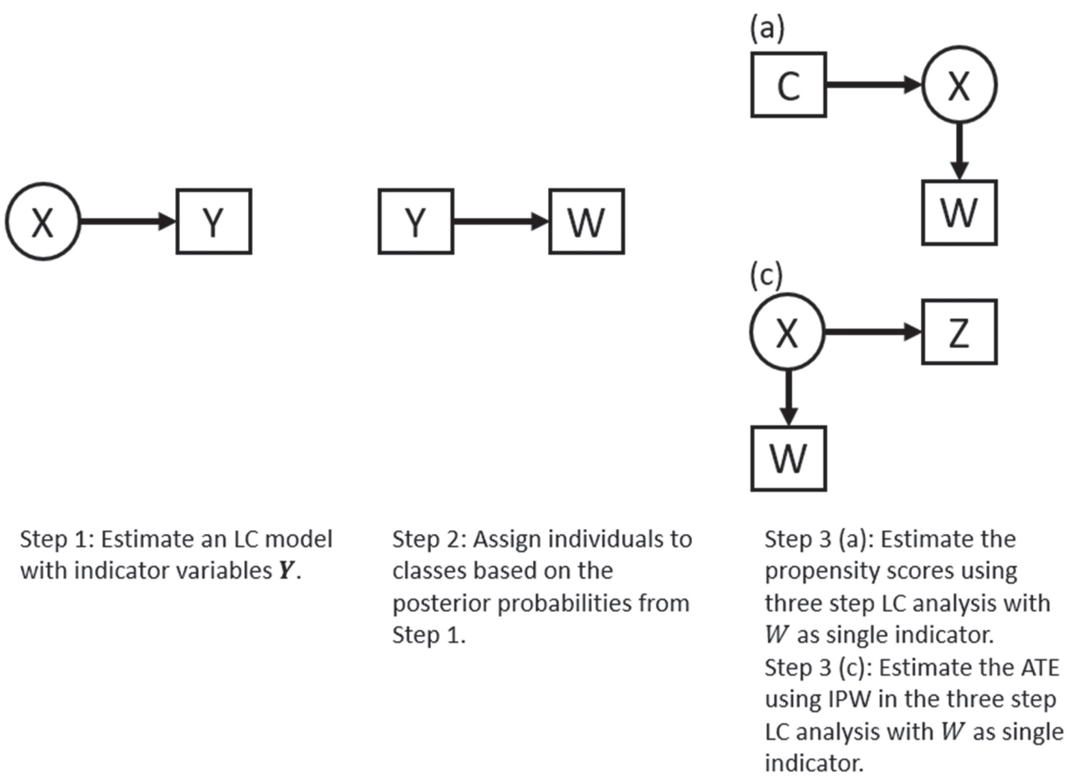

Figure 1.3. LC analysis using IPW as proposed by Bray et al. (2019). Note that Step 3 (b): Constructing single weights for each individual based on the inverse of the propensity scores is not shown.

classification errors), resulting in an expanded data set with T records per subject with classes t and weights $w^*_{it} = \sum_{s=1}^{T} d_{st} \cdot w_{is}$ (for a detailed explanation, see Bakk et al. (2013)). This expanded data set allows to link the class-specific ipw_{it} to class membership in the correct manner. This is not possible with the ML method where the LC model is estimated with the assigned class membership W as the single indicator and known error probabilities $P(W = s|X = t)$. Furthermore, while the previous steps could be estimated using the two-step method by Bakk and Kuha (2018), this last step of including class specific ipw_{it} is not possible in the second step of this approach.

A New bias-adjusted three-step LC analysis using the propensity score as covariate

As an alternative strategy to IPW, the propensity scores can also be included directly as control variables in the regression model for the outcome variable. After obtaining the class-specific propensity scores $\hat{\pi}_{it}$ in Step 3 (a), the final step of estimating the ATE can be done by maximizing the following log-likelihood function:

$$\log L_{BCH} = \sum_{i=1}^{N}\sum_{t=1}^{T} w^*_{it} \ \log f(Z_i|X = t, \hat{\pi}_{it}) \tag{6}$$

using the BCH correction (recommended for continuous outcome) or:

$$\log L_{ML} = \sum_{i=1}^{N}\sum_{s=1}^{T} w_{is} \ \log \sum_{t=1}^{T} P(X = t) f(Z_i|X = t, \hat{\pi}_{it}) P(W = s|X = t) \tag{7}$$

using the ML correction (recommended for categorical outcome). It is important to note that this method requires a correct specification of the relationship between the distal outcome and the propensity scores. We recommend using a flexible regression model with quadratic and interaction terms or splines for the propensity scores. Furthermore, as specified in Equations 6 and 7, we assume homogeneity of the ATE across strata of the propensity scores. This assumption could be relaxed by including interaction terms between exposure classes and the propensity scores. It is important to realize that the propensity scores $\hat{\pi}_{it}$ for each individual sum up to one over all classes. Therefore, propensity scores for only $t-1$ classes need to be included in the model. The ATE is obtained by comparing the estimated marginal expected value of Z_i across latent classes.

Simulation study

Design

A simulation study was conducted to evaluate the performance of our proposed strategy in comparison with the existing ones on three types of performance measures: bias of the ATE, bias of the standard errors (SE) of the ATE, and variation of the ATE.

The population model used is a latent class model with three classes for six dichotomous response indicators, an outcome variable Z, and two categorical confounders C_1 $(-0.5; 0.5)$ and C_2 $(-2; 1; 0; 1; 2)$. We investigated the methods' performances for two types of the outcome variable: Z is binary and Z is continuous with normal distribution. Adopting the same setup from Vermunt (2010), Class 1 is most likely to score high on all six indicators, Class 2 scores high on the first three and low on the last three indicators, and Class 3 scores low on all indicators.

When the outcome variable is binary, the BCH method can run into a problem when there are negative cell frequencies in the X-Z frequency table. This prevents the results from converging. It was first reported in the simulation study by Bakk et al. (2013), where the authors decided to delete the replications with the negative cell frequencies in the subsequent analyses. In this study, we chose to treat the binary outcome as continuous in the final step of estimating the ATE. This is possible because the ATE will be the same (the difference in means of Z or proportions of scoring "1" between two given classes). Furthermore, the SEs will be identical since we used robust SEs. The advantage of this option is that we will have results for all of the replications without convergence problems. The downside is that some of the replications will the have out-of-range estimated values for one of class-specific means of Z (below 0 or above 1), which we will report in the *Result* section.

We varied four factors in the simulation study: class separation level, sample size, confounding effect size, and treatment effect size. Class separation level and sample size have been shown to affect the performance of the bias-adjusted three-step LC analysis with an external variable (Bakk et al., 2013; Vermunt, 2010). Class separation was manipulated via the probabilities for the most likely response. We chose two levels of 0.80 and 0.90, corresponding to a moderate (entropy R-square of 0.65) and a good (entropy R-square of 0.90) separation condition, respectively. A good class separation is always ideal for the three-step approaches. The moderate separation condition can be perceived as the most common situation but also poses some challenges for the three-step methods. There is no added value to investigate the low class separation (e.g., response probability of 0.70 yielding entropy R-square of 0.36) since the three-step methods perform poorly in this condition (Bakk et al., 2013; Vermunt, 2010). The sample size is varied at three levels: 500, 1000, and 2500. The set up to manipulate the effect size and confounding strength was taken from Clouth et al. (2022). The effect of the confounders on the classes was modeled using the following logistic regression:

$$logit(X|C_1, C_2) = .5 + 1 * C_1 + \gamma * C_2 \quad (8)$$

with Class 1 as the reference group. The confounding effect was manipulated by varying the size of $\gamma = [1, 2, 3]$ for Class 3. For Class 2, γ was kept constant at 1. Note that one could also manipulate the strength of confounding through the relation between the confounders and the distal outcome Z.

For the binary outcome variable Z, the effect of the classes and the confounders on the outcome variable was modeled using logistic regression as follows

$$logit(Z|X, C_1, C_2) = 0 + 1 * C_1 + 1 * C_2 + 1 * X_2 + \beta * X_3 \quad (9)$$

with Class 1 as the reference group. X_2 and X_3 are two dummy variables denoting class membership in Class 2 and 3, respectively. The causal effect was manipulated by varying the size of $\beta = [1, 2, 3]$. For the continuous outcome variable Z, the effect of the classes and the confounders on the outcome variable was modeled as follows

$$E(Z|X, C_1, C_2) = 0 + 1 * C_1 + 1 * C_2 + 1 * X_2 + \beta * X_3. \quad (10)$$

The residual variance of Z was fixed at 10. The ATEs are defined as the average difference in means (continuous variable) or proportion of scoring "1" (binary variable) of Z in Class 2 and Class 3 compared to Class 1 averaged over all possible values of C_1 and C_2. In total, 2 (continuous and binary Z) x 2 (class separation levels) x 3 (effect sizes) x 3 (confounding sizes) x 3 (sample sizes) = 108 combinations of conditions were used to simulate data, with 500 replications per condition. The ATEs resulting from this setup are presented in Table 2.

A minor change was made to the Bray et al. (2019) method: in Step 3 (a) (the propensity score model), we used the ML correction method with proportional assignment instead of BCH method with modal

Table 2. ATEs of the population model for varying effect sizes β.

	Binary Z		Continuous Z	
	Class 2	Class 3	Class 2	Class 3
$\beta = 1$	0.164	0.164	1	1
$\beta = 2$	0.164	0.304	1	2
$\beta = 3$	0.164	0.402	1	3

assignment.[2] As explained above, the ML method is preferred to the BCH method for covariates.[3]

The simulation setup as described above results in equal sample sizes for all three classes. However, models with unequal class sizes are encountered frequently in practice as well. Therefore, additional scenarios with class sizes of (1) 45%, 45%, and 10% and (2) 80%, 10%, and 10% were investigated. Results for these scenarios are reported in the appendix.

All of the steps were conducted in LatentGOLD 6.1 (Vermunt & Magidson, 2021) and results were imported to R (RCoreTeam, 2022) to produce tables and figures. Three new options were implemented in LatentGOLD 6.1 to facilitate the procedure of the bias-adjusted three-step LC analysis with IPW. First, the Bray et al. (2019) method can now be easily carried out by specifying the option 'bray' in 'step 3'. Second, the 'propensity=(… .)' option was added to include the inverse of the estimated propensity scores as fixed weights in the last step. Third, an option was added to help speed up the simulation process in batch mode. Detailed code is available on GitHub (https://github.com/trale97/LCAdistaloutcomeIPW).

Results

Results of all simulation scenarios can be explored interactively using the R shiny app https://trale.shinyapps.io/lcasim/. As results are very similar for Class 2 and Class 3, we will only be reporting results for Class 3. Furthermore, we will only be reporting results for moderate class separation as this corresponds to a more realistic scenario. Results for good class separation are reported in Appendix 1. Figures 2 and 3 display the bias of the ATE estimates for the binary and continuous outcome variables averaged across 500 replications. Larger confounding and larger effect size seem to slightly increase bias. For sample size, there is

[2]According to their provided computer codes, it appeared that the ML method was used instead of the BCH method (option 'AUXILIARY = (R3STEP)' in Mplus).

[3]Furthermore, we tested the Bray et al. (2019) method with both the proportional and modal assignment in the final step. However, there is almost no difference in all three performance measures and modal assignment will be the focus for reporting results.

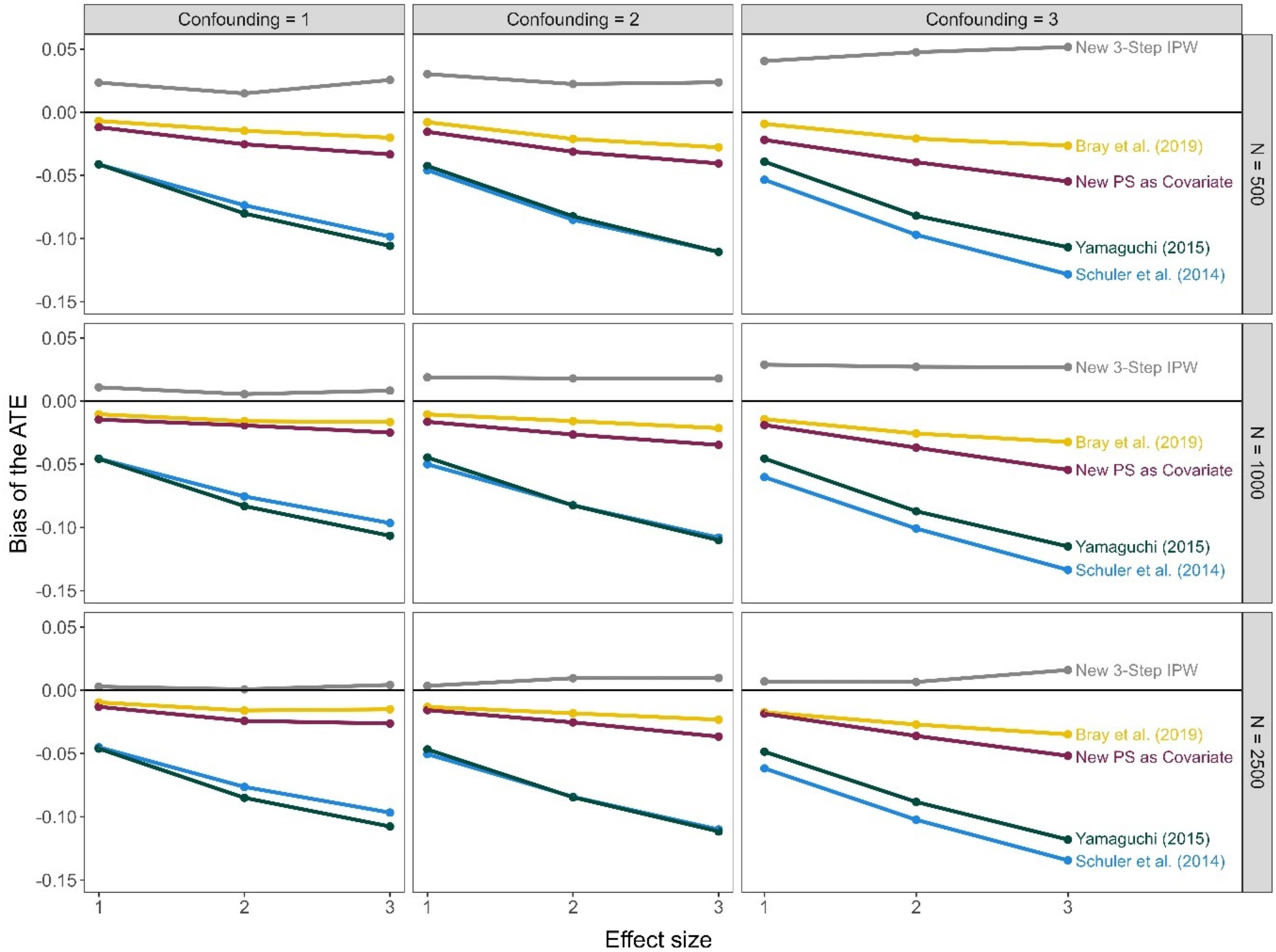

Figure 2. Results for the bias of the ATE for the binary distal outcome variable and moderate class separation. The bias is presented for varying levels of effect size, strength of the confounding effect, and sample size. Results are averaged over 500 replications.

no general pattern observable. For the binary outcome variable, both of our newly proposed methods as well as the method by Bray et al. (2019) perform well with the new IPW method slightly overestimating the ATE and the new propensity scores as covariates method and Bray et al. (2019) method slightly underestimating the ATE. Unsurprisingly, the methods by Schuler et al. (2014) and Yamaguchi (2015) are consistently more biased. For the continuous outcome variable, we observe essentially similar results. However, for a large sample size of $N = 2500$, both of our newly proposed methods show almost no bias clearly outperforming the other methods.

Figures 4 and 5 present the *SD* of the estimated ATE for the binary and continuous outcome variables, respectively. Increasing confounding seems to slightly increase the *SD* but effect size does not seem to affect the *SD*. For both, binary and continuous outcome variables, our newly proposed 3-Step IPW method is consistently outperformed.

Figures A1 and A2 (see Appendix) present the bias of *SE* for the binary and continuous outcome variable, respectively. None of the methods clearly outperforms the other methods in terms of bias of *SE*. For both, binary and continuous outcomes, the bias of *SE* decreases for larger sample sizes.

For some conditions with moderate class separation, our new 3-Step IPW method did not converge in all replications. Replications for which this was the case were not considered for summarizing the results. Tables A1 and A2 in the Appendix summarize the number of replications per condition for which the 3-Step IPW method did not converge. As can be seen, the condition with a strong confounding effect in combination with a small sample size is most problematic.

Real-Life example using data from the LISS panel

In this section, we illustrate our two newly proposed methods to investigate the relationship between mental

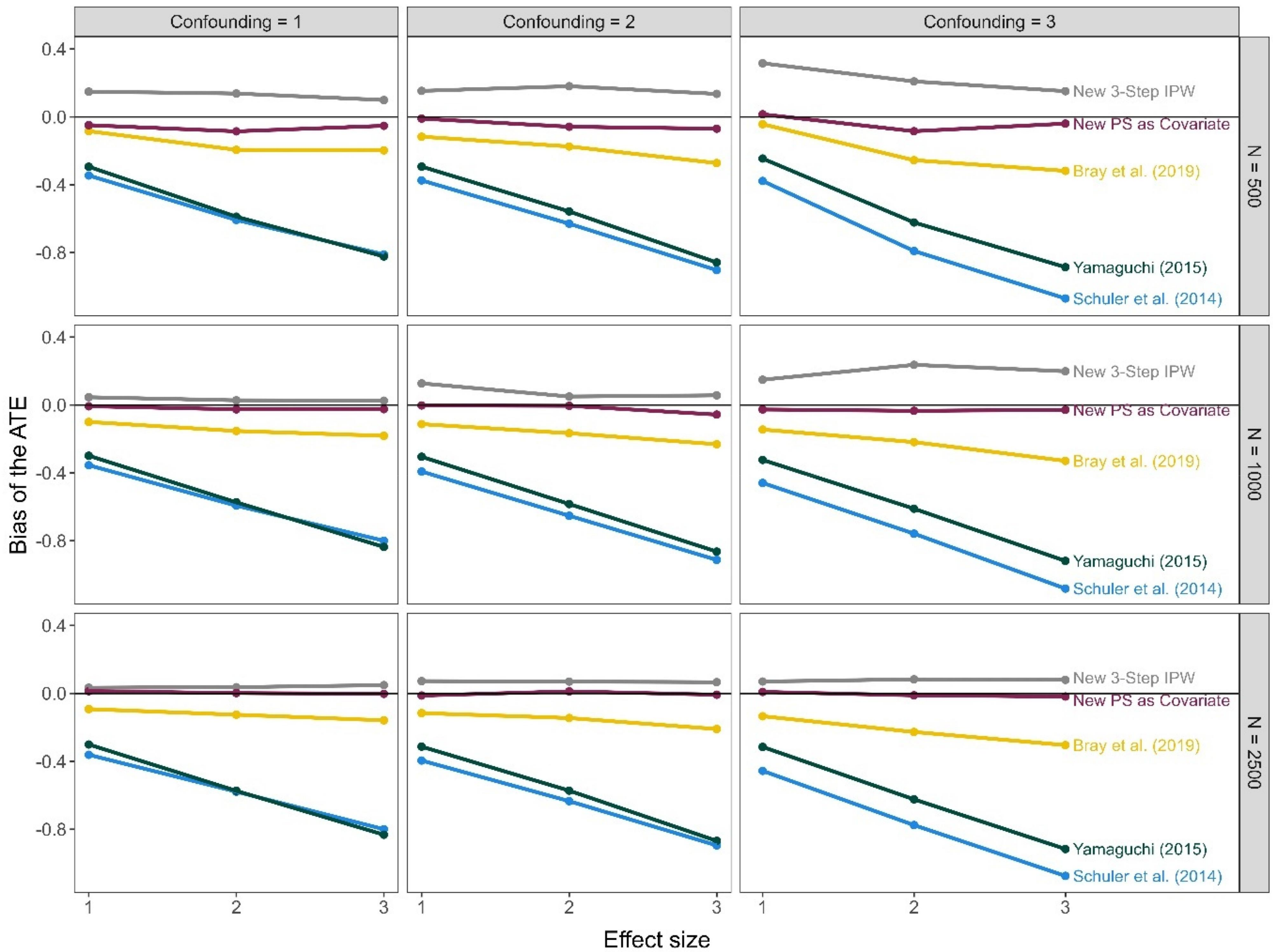

Figure 3. Results for the bias of the ATE for the continuous distal outcome variable and moderate class separation. The bias is presented for varying levels of effect size, strength of the confounding effect, and sample size. Results are averaged over 500 replications.

health and unemployment. Poor mental health and depression are leading contributors to the global burden of disease (Lépine & Briley, 2011) and key factors of economical productivity (Layard, 2013). Furthermore, poor mental health has been linked to lower levels of economic activity, lower earnings, more difficulties in both finding and retaining employment, and reduced financial security (Bubonya et al., 2019).

Data

In this paper we make use of data of the LISS (Longitudinal Internet studies for the Social Sciences) panel administered by Centerdata (Tilburg University, The Netherlands). The LISS panel is a representative sample of Dutch individuals who participate in monthly internet surveys. The panel is based on a true probability sample of households drawn from the population register. Households that could not otherwise participate are provided with a computer and internet connection. A longitudinal survey is fielded in the panel every year, covering a large variety of domains including health, work, education, income, housing, time use, political views, values and personality. More information about the LISS panel can be found at: www.lissdata.nl.

We used data from the 2021 wave of the LISS core study and excluded participants outside the labor force, that is, attending an educational program or being retired, resulting in a sample of 3567 participants. For some of the participants, the data contained missing values. While missing values on the indicators are addressed in LCA by using full-information maximum likelihood estimation, missing values on the confounders need to be imputed, preferably by multiple imputation. As this example data analysis serves an illustrative purpose, we decided to perform a single set of imputation using MICE (Van Buuren & Groothuis-Oudshoorn, 2011) and treating the data as fully observed for our analysis.

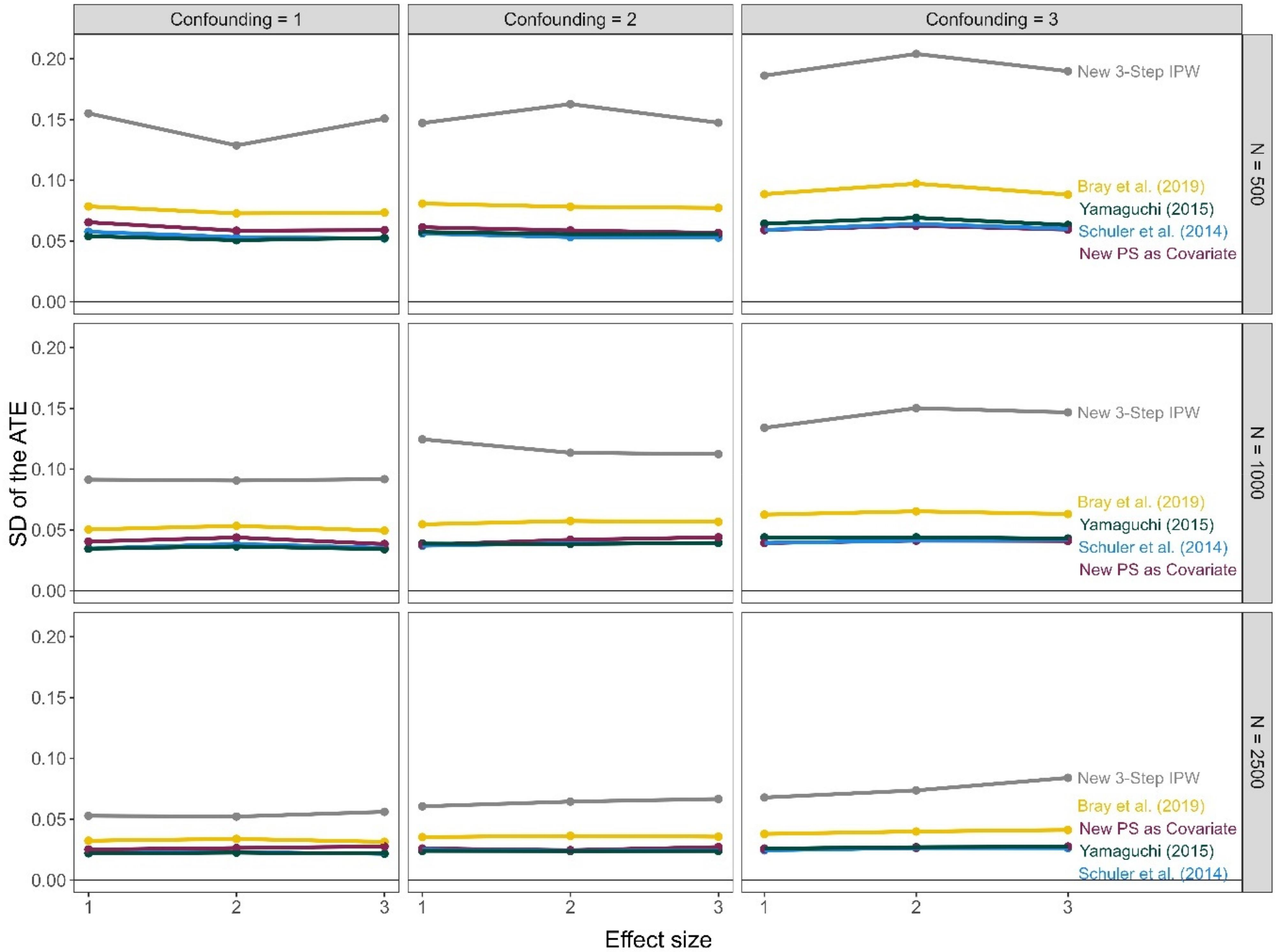

Figure 4. Results for the standard deviation (*SD*) of the ATE for the binary distal outcome variable and moderate class separation. The bias is presented for varying levels of effect size, strength of the confounding effect, and sample size. Results are averaged over 500 replications.

Measures

Outcome

The outcome in our study is employment status. "Being employed" is defined as conducting work for pay, either as an employee, self-employed professional, or assisting in a family business. "Being unemployed" is further defined as not conducting work for pay. This includes participants who performed unpaid (voluntary) work or care. Participants who were too young to perform any work for pay or who attended an educational program and individuals who reached retirement were excluded from the sample.

Exposure

Here, we consider "mental health" as exposure. Specifically, mental health is defined as the latent classes that are identified in the first step of the LC analysis based on the observed indicators "I felt very anxious", "I felt so down that nothing could cheer me up", "I felt calm and peaceful", "I felt depressed and gloomy", and "I felt happy". All items were scored on a six point Likert scale ranging from "never" to "continuously".

Confounders

Furthermore, we identified the variables age, gender, subjective general physical health, household status, gross household income, education, and origin as confounders of the mental health—employment relationship.

Results

Step 1: Measurement model

Based on the Bayesian Information Criterium (BIC), Akaike Information Criterium (AIC), and maximum bivariate residual (BVR; Table 3), we selected a LC model with three classes. Note that with large sample sizes, these information criteria tend to decrease even for large numbers of classes. Here, we deemed the reduction in BIC, AIC, and BVR to not be substantial

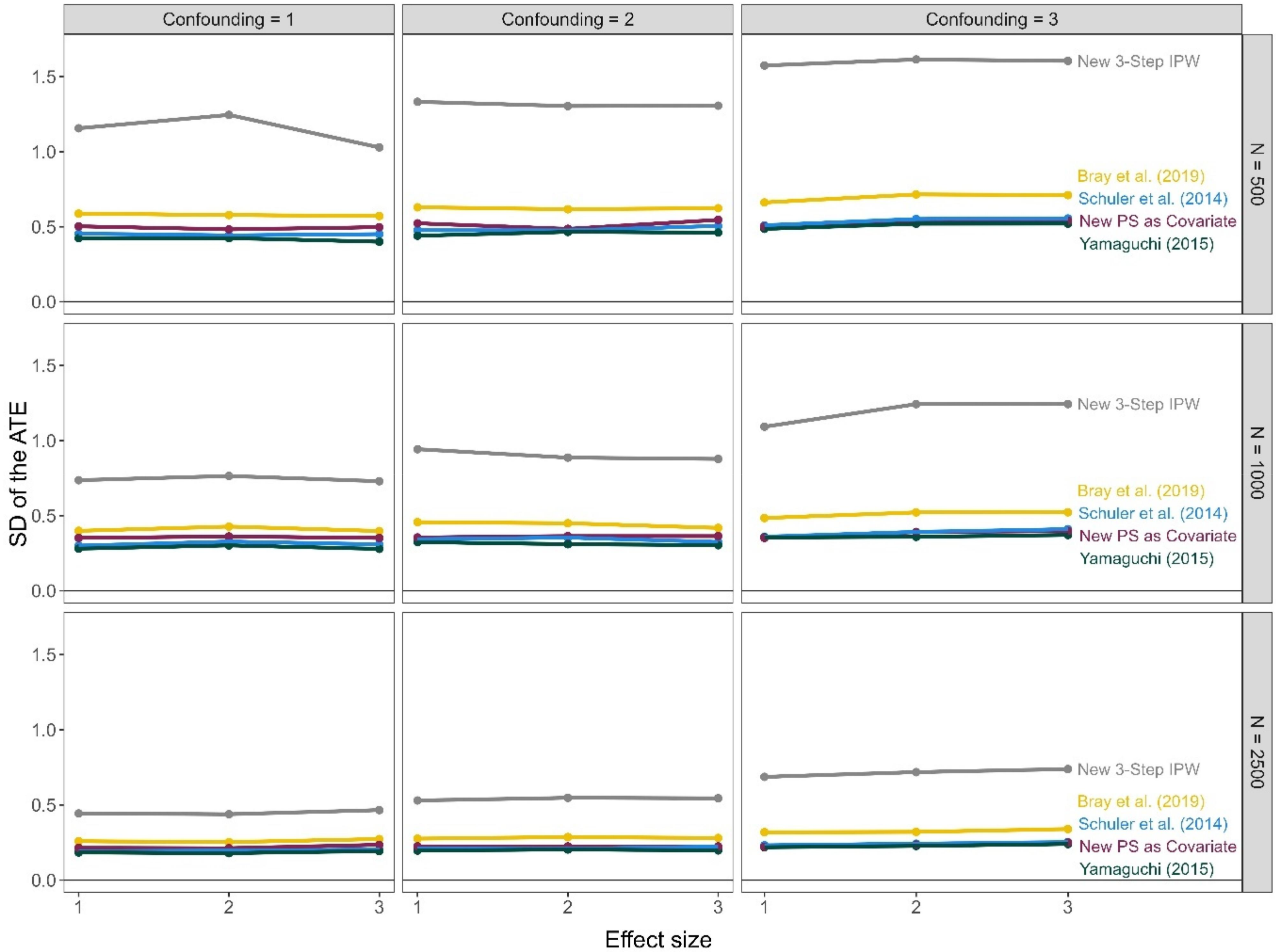

Figure 5. Results for the standard deviation (*SD*) of the ATE for the continuous distal outcome variable and moderate class separation. The bias is presented for varying levels of effect size, strength of the confounding effect, and sample size. Results are averaged over 500 replications.

Table 3. Goodness of Fit statistics. Log-likelihood values, Bayesian Information Criterium (BIC), Akaike Information Criterium (AIC), and maximum bivariate residual (BVR) values are presented for models with 1 – 5 classes. The selected solution is highlighted.

	Log-likelihood	BIC	AIC	Max. BVR
1 Class	−23860.0	47944.4	47789.9	3656.9
2 Classes	−20842.6	41938.8	41747.2	574.5
3 Classes	**−19984.3**	**40271.3**	**40042.6**	**214.5**
4 Classes	−19699.8	39751.4	39485.7	149.0
5 Classes	−19555.3	39511.4	39208.6	102.9

enough for justifying four classes. Class 1 (44.4%) is characterized by slightly elevated levels on all indicators and is labeled "slightly elevated mental health problems" (Figure 6). Class 2 (36.9%) is characterized by good to very good values on all indicators and is labeled "good mental health". Note that for the indicators "feel calm" and "feel happy", the scales are inverted so that low scores correspond to good values. Class 3 (18.7%) is characterized by high values on all indicators and is labeled as "poor mental health". In all three states, scores on the items "feel calm" and "feel happy" were slightly elevated.

Step 3 (a): IPW diagnostics

One important assumption for identifying the ATE when using IPW is positivity. Positivity refers to individuals in the different exposure groups having non-zero probabilities of membership in the other exposure groups (Hernán & Robins, 2006). When using propensity score methods, this assumption can be assessed by inspecting the overlap of propensity scores of the different exposure groups (Austin, 2011). Here, the exposure groups refer to the latent classes identified in *Step 1*. As can be seen in Figure 7, there is sufficient overlap of propensity scores in the LISS data. Furthermore, to identify the ATE, exchangeability is required. While there is no possibility of assessing exchangeability, we can check if our propensity score model correctly adjusts for the measured

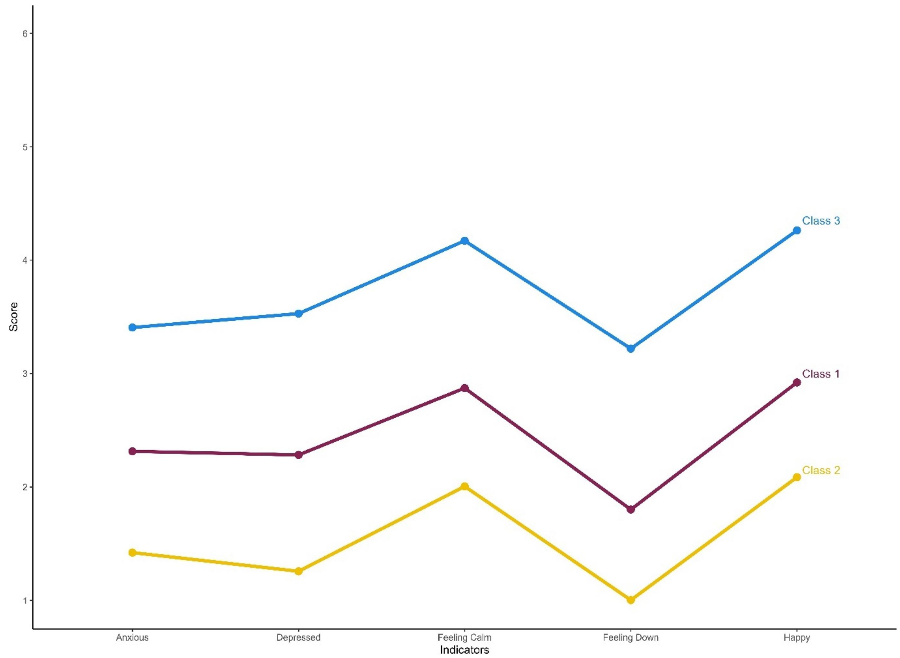

Figure 6. Composition of the latent classes as estimated in the Step 1 of the bias-adjusted three step LC analysis. Class specific average scores are presented for all five observed indicators.

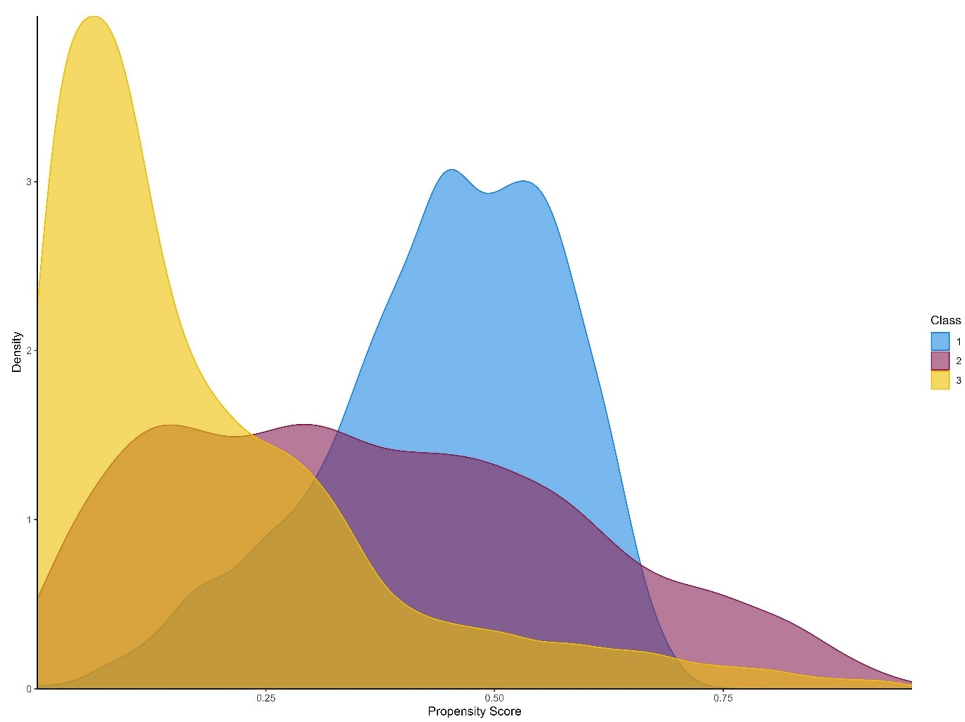

Figure 7. Overlap of the propensity scores for the three latent classes (exposure groups).

confounders. This can be investigated by assessing balance on the confounders between the exposure groups after weighting (Austin, 2011). As presented in Table 4, there are no significant differences in the confounders between the three latent classes indicating sufficient balance.

Step 3 (c): ATE

Table 5 shows the estimated probabilities of employment status for the three latent classes. As can be seen, both the IPW and the propensity score as covariate method estimate only minor differences in employment status between the classes. In detail,

Table 4. Descriptive statistics for the confounders used to estimate the propensity scores after weighting. Unless otherwise specified, marginal differences in probabilities/means [CI] are presented.

	Class 2	Class 3
Job		
Unemployed	0.033 [−0.057; 0.122]	−0.016 [−0.090; 0.057]
Employed	−0.033 [−0.122; 0.057]	0.016 [−0.057; 0.090]
Gender		
Female	0.010 [−0.064; 0.084]	0.013 [−0.081; 0.106]
Male	−0.010 [−0.084; 0.064]	−0.013 [−0.106; 0.081]
Age	−2.117 [−4.485; 0.250]	−1.392 [−4.073; 1.289]
Household Status		
Never been married	0.043 [−0.037; 0.124]	−0.048 [−0.125; 0.028]
Married	−0.050 [−0.124; 0.025]	0.020 [−0.074; 0.113]
Separated	0.009 [−0.010; 0.028]	0.004 [−0.004; 0.013]
Divorced	0.003 [−0.056; 0.062]	0.004 [−0.047; 0.056]
Widowed	−0.006 [−0.022; 0.010]	0.020 [−0.031; 0.071]
Household Income	−6907.4 [−16384.2; 2569.4]	−6926.5 [−15590.2; 1737.2]
Education		
Primary school	0.014 [−0.025; 0.054]	−0.008 [−0.030; 0.014]
Vmbo	0.012 [−0.058; 0.082]	−0.002 [−0.070; 0.065]
Havo/Vwo	−0.017 [−0.044; 0.010]	−0.014 [−0.069; 0.042]
Mbo	−0.009 [−0.072; 0.055]	0.040 [−0.047; 0.127]
Hbo	−0.004 [−0.073; 0.066]	−0.010 [−0.099; 0.080]
Wo	0.003 [−0.047; 0.052]	−0.006 [−0.071; 0.058]
Origin		
Dutch	−0.026 [−0.115; 0.063]	−0.009 [−0.080; 0.062]
First Gen. non-West	0.019 [−0.052; 0.090]	0.002 [−0.048; 0.051]
First Gen. West	−0.005 [−0.039; 0.028]	−0.001 [−0.032; 0.030]
Second Gen. non-West	0.001 [−0.041; 0.043]	−0.006 [−0.028; 0.016]
Second Gen. West	0.011 [−0.047; 0.070]	0.015 [−0.025; 0.054]
Health	−0.106 [−0.321; 0.110]	0.142 [−0.083; 0.366]

Table 5. Probabilities of employment status conditional on class membership. Estimates for the naive approach without adjustment, our newly proposed IPW method, and the propensity score as covariate method are presented.

	Naive approach		IPW		PS as covariate	
	Unemployed	Employed	Unemployed	Employed	Unemployed	Employed
Class 1	0.222	0.778	0.263	0.737	0.249	0.751
Class 2	0.221	0.779	0.313	0.687	0.250	0.750
Class 3	0.377	0.623	0.254	0.746	0.256	0.744

according to the IPW method, individuals in class 2 have a 5.0% higher probability (CI = [−4.0%; 14.0%]) and individuals in class 3 have a 0.9% lower probability (CI = [−8.3%; 6.5%]) of being unemployed compared to individuals in class 1. According to the propensity score as covariate method, individuals in class 2 have a 0.03% higher probability (CI = [−3.5%; 3.5%]) and individuals in class 3 have a 0.7% higher probability (CI = [−3.8%; 5.1%]) of being unemployed compared to individuals in class 1. In contrast, the naive approach that does not account for confounding shows a significant effect of class membership in class 3 on employment status. Individuals in class 2 have a 0.05% lower probability (CI = [−3.5%; 3.4%]) and individuals in class 3 have a 15.5% higher probability (CI = [10.8%; 20.3%]) of being unemployed compared to individuals in class 1.

Discussion

In this paper, we presented two novel approaches utilizing propensity scores for three-step bias adjusted LC analysis with distal outcomes. In the first approach, class-specific weights based on the inverse of the propensity scores are included in the estimation of the third step of three-step bias adjusted LC analysis. In the second approach, the propensity scores are used directly as covariates in the third step. Specifically the first approach solves some of the shortcomings of previously proposed methods by Schuler et al. (2014), Yamaguchi (2015), and Bray et al. (2019) by allowing class-specific weights while accounting for classification errors. In a simulation study, we showed that both Schuler et al. (2014) and Yamaguchi (2015) estimate the ATE with substantial bias. This is not surprising since for both methods, classification errors were not accounted for in the third step. In contrast, Bray et al. (2019) and our two newly proposed methods estimate the ATE without substantial bias. However, our IPW based approach has substantially larger standard errors. While IPW is known for producing larger standard errors in some occasions, it seems like implementing IPW in the third step of three-step bias adjusted LC analysis intensifies the problem. The propensity score as covariate methods does not have this problem, however, it is considerably less flexible than the IPW method. That is, while in the IPW method, the ATE can be estimated by simply including the latent classes and the weights in the third step, the propensity score as covariate method requires the correct specification of this third step. I.e., one might need to include quadratic terms of the propensity scores or splines to estimate the ATE without bias. Furthermore, the propensity score as covariate method assumes homogeneity of the treatment effect across strata of the propensity scores. This assumption might be unreasonable in practice and can be relaxed by including interaction terms between the treatment variable and the class-specific propensity scores in the model. Whether the flexibility of the IPW approach in comparison to the propensity score as covariate approach outweighs its relative inefficiency can not be determined generally but depends on the individual use case, for instance, available sample size.

An additional problem with the IPW approach can arise when there are extreme weights. Extreme weights are a well known problem of IPW based methods. However, our newly proposed three-step method using the BCH correction amplifies this problem as, for some individuals, extreme IPW weights are multiplied with extreme BCH weights. In some cases, these new weights can become so extreme that they cause convergence problems. Especially with increasing numbers of classes, propensity scores might become extremely small as class membership in some of the more extreme classes is very unlikely for some individuals. Intentionally or not, Bray et al. (2019) solve this issue by averaging the IPW weights over all classes for each individual. However, this is not a common approach and it has not been investigated how this approach performs, e.g., regarding its balancing properties. More traditionally, this problem is addressed by truncating the IPW weights (or propensity scores) (Cole & Hernán, 2008). For instance, individuals with weights larger than the 99% percentile might be assigned the weight of the 99% percentile. An alternative approach to truncating would be the use of Bayesian shrinkage priors on the multinomial logistic regression coefficients in the third step. Such priors force positivity on the class definitions by preventing the posterior class membership probabilities to take on values close to zero. While both, truncating and utilizing shrinkage priors might solve the convergence problems, they come at the cost of worsening balance. To this point, it remains an open question how feasible it is to achieve sufficient balance with larger numbers of classes, as in this case, weights might need to be truncated to a larger degree.

As has been shown in this study, the use of IPW weights for exposures that are latent requires a stepwise analysis approach. All methods presented, Schuler et al. (2014), Yamaguchi (2015), Bray et al. (2019), as well as our two newly proposed methods utilize a three-step approach of (1) estimating a measurement model, (2) classifying individuals, and (3) estimating structural models for propensity scores and ATEs. While these models vary in their specifications of the propensity scores and the correction methods used in the third step, they all follow this general framework. However, there is an alternative stepwise approach for relating auxiliary variables to the latent classes, the two-step method (Bakk & Kuha, 2018). Similar to the three-step method, in the first step, a measurement model including only the indicator variables is estimated. However, the structural model is then estimated in a second step where a full model is specified including indicator variables and covariates or distal outcomes. Crucially, in this structural model, parameters for the item response probabilities are fixed to the values that were estimated in the first step. As such, the two-step method does not require a classification step and, as a consequence, a correction for misclassifications. Generally, the two-step method can be used whenever the three-step method is appropriate. However, for the IPW method proposed in this study, we exploit the fact that the BCH correction extends the dataset with one record per class per observation facilitating the use of multiple weights per observation. This procedure is only possible with the BCH correction and the two-step approach can therefore not be utilized. For our second method of using the propensity score directly as covariates in the structural model, the two-step approach could be used.

The identification of the ATE relies on the assumption of positivity. Using IPW based methods, this is usually checked by assessing overlap of the propensity scores between the exposure groups. Here, this would mean assessing overlap of the propensity scores between the assigned classes. However, this seems to be unreasonable considering that there might not be any hard partitioning and every individual is receiving multiple propensity scores. In this study, we presented the overlap of all propensity scores for all individuals to assess balance. However, one might argue that positivity is already violated if a single individual has a close to zero probability of class membership in any class. Again, this becomes increasingly likely with increasing number of classes. To this point, it remains an open question how to best assess the assumption of positivity when using bias-adjusted three-step LC analysis. Furthermore, the identification of the ATE relies on the assumption of exchangeability or no unmeasured confounding. While it is impossible to show that no unmeasured confounding is present, it is possible to investigate if confounding due to measured confounders is accounted for correctly. I.e., the correct specification of the propensity score model can be assessed by checking balance between the latent classes after weighting. However, using bias-adjusted three-step LC analysis, this is not a straightforward task. As balance needs to be assessed for true class membership X rather than assigned class membership W, this cannot be done as usual, for instance, using the *survey* or *tableone* packages in R. Rather, an additional third step of bias-adjusted three-step LC analysis needs to be estimated with all measured confounders as distal outcomes of the latent classes using the BCH correction. Then, balance is achieved if the parameter estimates of this model are non-significant.

More generally, considering latent variables as exposures in causal inference is not without criticism. For instance, VanderWeele (2022) questions the causal efficacy of latent variables. He argues that the potential outcomes under the different exposures are not well-defined as latent variables almost never resemble a unidimensional exposure. That is, class membership in one of the mental health classes as identified in this study might reflect different sets of scores on the underlying indicator variables. Class membership therefore does not reflect exactly the same version of this exposure resulting in a less well defined estimand for the causal effect. To account for this, VanderWeele (2022) proposes a new model of measurement based on the theory for causal inference under multiple versions of treatment. It remains a topic for future research how bias-adjusted three-step LC analysis can be implemented in this framework. However, this issue is not exclusive for latent variables but for most exposures in the social sciences. Careful consideration about loosening the definition of the same version of exposure and accepting less precisely defined causal effects is warranted (Kaufman, 2019).

Conclusion

Bias-adjusted three step LC analysis is a popular tool for estimating the effect of class membership on distal outcomes. Here, we proposed two extensions based on the propensity score to adjust for confounding and estimate the ATE, i.e., bias-adjusted three step LC analysis using IPW and bias-adjusted three step LC analysis using the propensity scores as covariates. While both methods perform well in terms of bias, the IPW approach is less efficient and can run into convergence issues when the distal outcome is discrete. We recommend the use of bias-adjusted three step LC analysis with IPW for continuous distal outcomes with larger sample sizes and the use of bias-adjusted three step LC analysis with the propensity scores as covariates for discrete distal outcomes. Both options are implemented in the statistical software program LatentGOLD 6.1.

Article Information

Conflict of interest disclosures: Each author signed a form for disclosure of potential conflicts of interest. No authors reported any financial or other conflicts of interest in relation to the work described.

Ethical principles: The authors affirm having followed professional ethical guidelines in preparing this work. These guidelines include obtaining informed consent from human participants, maintaining ethical treatment and respect for the rights of human or animal participants, and ensuring the privacy of participants and their data, such as ensuring that individual participants cannot be identified in reported results or from publicly available original or archival data.

Funding: This work was supported by Grant 628.001.030 from The Netherlands Organisation for Scientific Research (NWO).

Role of the funders/sponsors: None of the funders or sponsors of this research had any role in the design and conduct of the study; collection, management, analysis, and interpretation of data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

References

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, *46*(3), 399–424. https://doi.org/10.1080/00273171.2011.568786

Bakk, Z., & Kuha, J. (2018). Two-step estimation of models between latent classes and external variables. *Psychometrika*, *83*(4), 871–892. https://doi.org/10.1007/s11336-017-9592-7

Bakk, Z., Tekle, F. B., & Vermunt, J. K. (2013). Estimating the association between latent class membership and external variables using bias-adjusted three-step approaches. *Sociological Methodology*, *43*(1), 272–311. https://doi.org/10.1177/0081175012470644

Bandeen-Roche, K., Miglioretti, D. L., Zeger, S. L., & Rathouz, P. J. (1997). Latent variable regression for multiple discrete outcomes. *Journal of the American Statistical Association*, *92*(440), 1375–1386. https://doi.org/10.1080/01621459.1997.10473658

Bolck, A., Croon, M., & Hagenaars, J. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis*, *12*(1), 3–27. https://doi.org/10.1093/pan/mph001

Bray, B. C., Dziak, J. J., Patrick, M. E., & Lanza, S. T. (2019). 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. *Prevention Science: The Official Journal of the Society for Prevention Research*, *20*(3), 394–406. https://doi.org/10.1007/s11121-018-0883-8

Bubonya, M., Cobb-Clark, D. A., & Ribar, D. C. (2019). The reciprocal relationship between depressive symptoms and employment status. *Economics and Human Biology*, *35*, 96–106. https://doi.org/10.1016/j.ehb.2019.05.002

Clouth, F. J., Pauws, S., Mols, F., & Vermunt, J. K. (2022). A new three-step method for using inverse propensity weighting with latent class analysis. *Advances in Data Analysis and Classification*, *16*(2), 351–371. https://doi.org/10.1007/s11634-021-00456-5

Cole, S. R., & Hernán, M. A. (2008). Constructing inverse probability weights for marginal structural models. *American Journal of Epidemiology*, *168*(6), 656–664. https://doi.org/10.1093/aje/kwn164

Dalmartello, M., Vermunt, J., Parazzini, F., Serraino, D., Giacosa, A., Crispo, A., Negri, E., Levi, F., Pelucchi, C., & La Vecchia, C. (2022). Comorbidity patterns, family history and breast cancer risk: A latent class analysis. *Journal of Epidemiology and Community Health*, *76*(10), 867–872. https://doi.org/10.1136/jech-2022-219279

Dayton, C. M., & Macready, G. B. (1988). Concomitant-variable latent-class models. *Journal of the American Statistical Association*, *83*(401), 173–178. https://doi.org/10.1080/01621459.1988.10478584

Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, *61*(2), 215–231. https://doi.org/10.1093/biomet/61.2.215

Hernán, M. A., & Robins, J. M. (2006). Estimating causal effects from epidemiological data. *Journal of Epidemiology and Community Health*, *60*(7), 578–586. https://doi.org/10.1136/jech.2004.029496

Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics*, *86*(1), 4–29. https://doi.org/10.1162/003465304323023651

Kamakura, W. A., Wedel, M., & Agrawal, J. (1994). Concomitant variable latent class models for the external analysis of choice data. *International Journal of Marketing Research*, *11*(5), 451–464. https://doi.org/10.1016/0167-8116(94)00004-2

Kaufman, J. S. (2019). Commentary: Causal inference for social exposures. *Annual Review of Public Health*, *40*(1), 7–21. https://doi.org/10.1146/annurev-publhealth-040218-043735

Layard, R. (2013). Mental health: The new frontier for labour economics. *IZA Journal of Labor Policy*, *2*(1), 1–16. https://doi.org/10.1186/2193-9004-2-2

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Houghton Mill.

Lépine, J.-P., & Briley, M. (2011). The increasing burden of depression. *Neuropsychiatric Disease and Treatment*, *7*(Suppl 1), 3–7. https://doi.org/10.2147/NDT.S19617

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*(19), 3388–3414. https://doi.org/10.1002/sim.5753

RCoreTeam (2022). *R: A language and environment for statistical computing*. RFoundation for Statistical Computing. https://www.r-project.org/.

Robins, J. M., Hernán, M., & Brumback, B. (2000). Marginal structural models and causal Inference in epidemiology. *Epidemiology* , *11*(5), 550–560. https://doi.org/10.1097/00001648-200009000-00011

Robins, J. M., Rotnitzky, A., & Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, *89*(427), 846–866. https://doi.org/10.1080/01621459.1994.10476818

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55. https://doi.org/10.1093/biomet/70.1.41

Schmitter, M., Vermunt, J., Blaauw, E., & Bogaerts, S. (2021). Risk classes of patients diagnosed with substance use disorders in Dutch forensic psychiatric centers. *The Journal of Forensic Practice*, *23*(1), 39–52. https://doi.org/10.1108/JFP-08-2020-0036

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 Research Methodology*, *14*(4), 232–254. https://doi.org/10.1007/s10742-014-0122-0

Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Mice: multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*(3), 1–67. http://www.jstatsoft.org/ https://doi.org/10.18637/jss.v045.i03

VanderWeele, T. J. (2022). Constructed measures and causal inference: Towards a new model of measurement for psychosocial constructs. *Epidemiology*, *33*(1), 141–151. https://doi.org/10.1097/EDE.0000000000001434

Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, *18*(4), 450–469. https://doi.org/10.1093/pan/mpq025

Vermunt, J. K., & Magidson, J. (2004). Latent class analysis. In D. Kaplan (Ed.), *The Sage Handbook of Quantitative Methodology for the Social Sciences*. (pp. 175–198) Sage Publications.

Vermunt, J. K., Magidson, J. (2021). *LG-Syntax User's Guide: Manual for Latent GOLD Syntax Module Version 6.0*. http://www.statisticalinnovations.com

Yamaguchi, K. (2000). Multinomial logit latent-class regression models: An analysis of the predictors of gender-role attitudes among Japanese women. *American Journal of Sociology*, *105*(6), 1702–1740. https://doi.org/10.1086/210470

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*. (RIETI Discussion Paper Series 15–E090 Issue.

Appendix

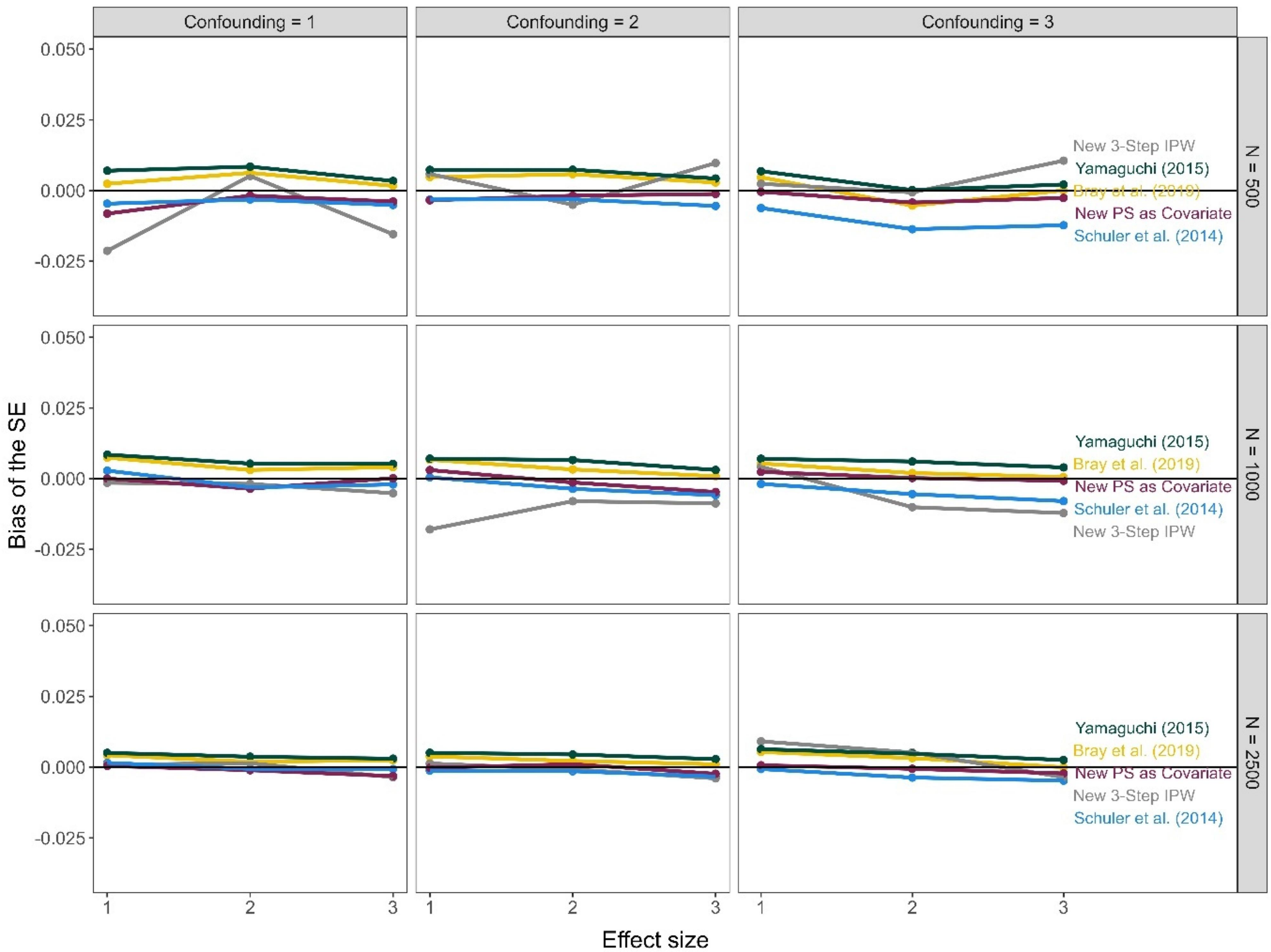

Figure A1. Results for the bias of the standard error (SE) of the ATE for the binary distal outcome variable and moderate class separation. The bias is presented for varying levels of effect size, strength of the confounding effect, and sample size. Results are averaged over 500 replications.

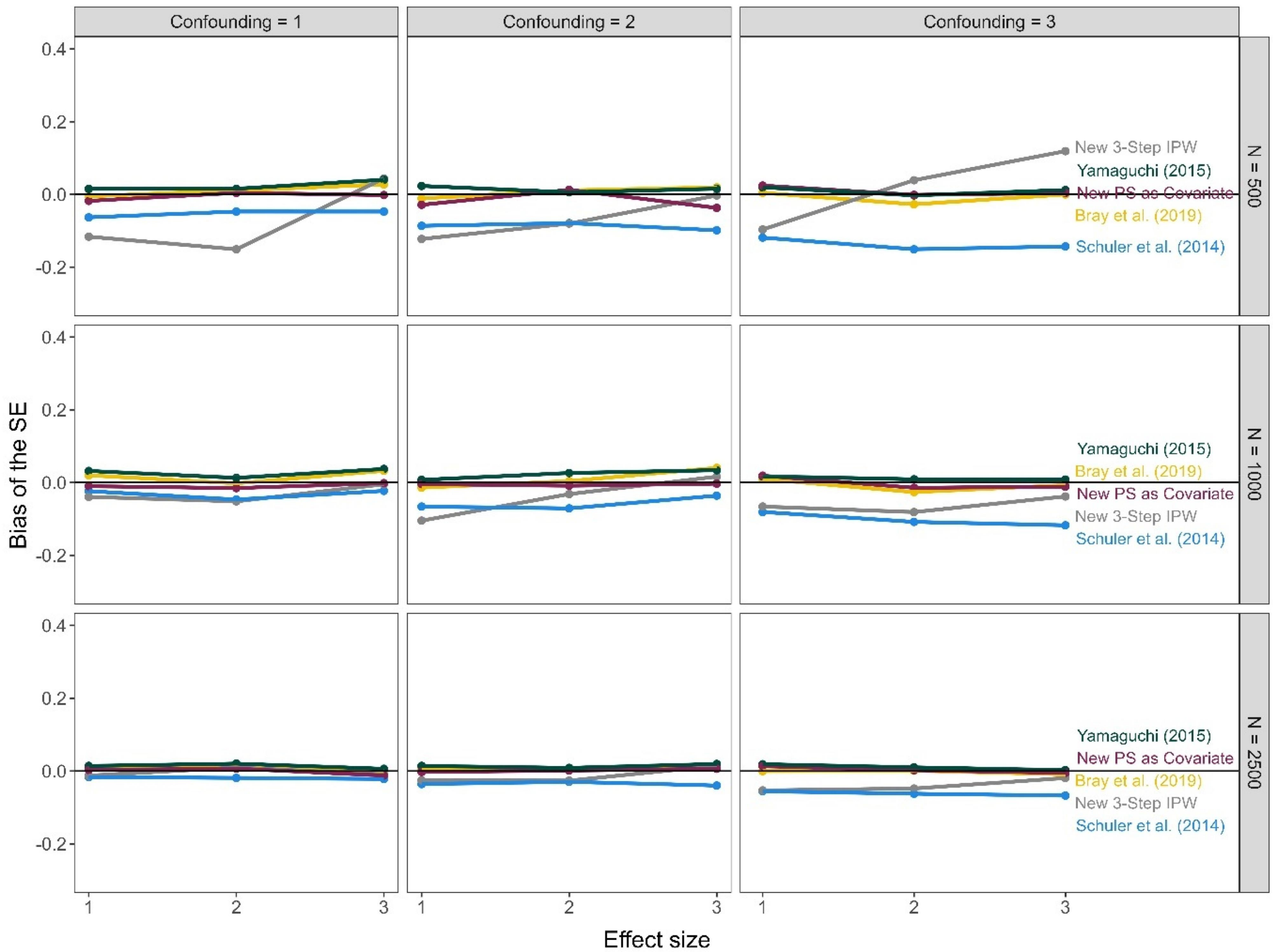

Figure A2. Results for the bias of the standard error (SE) of the ATE for the continuous distal outcome variable and moderate class separation. The bias is presented for varying levels of effect size, strength of the confounding effect, and sample size. Results are averaged over 500 replications.

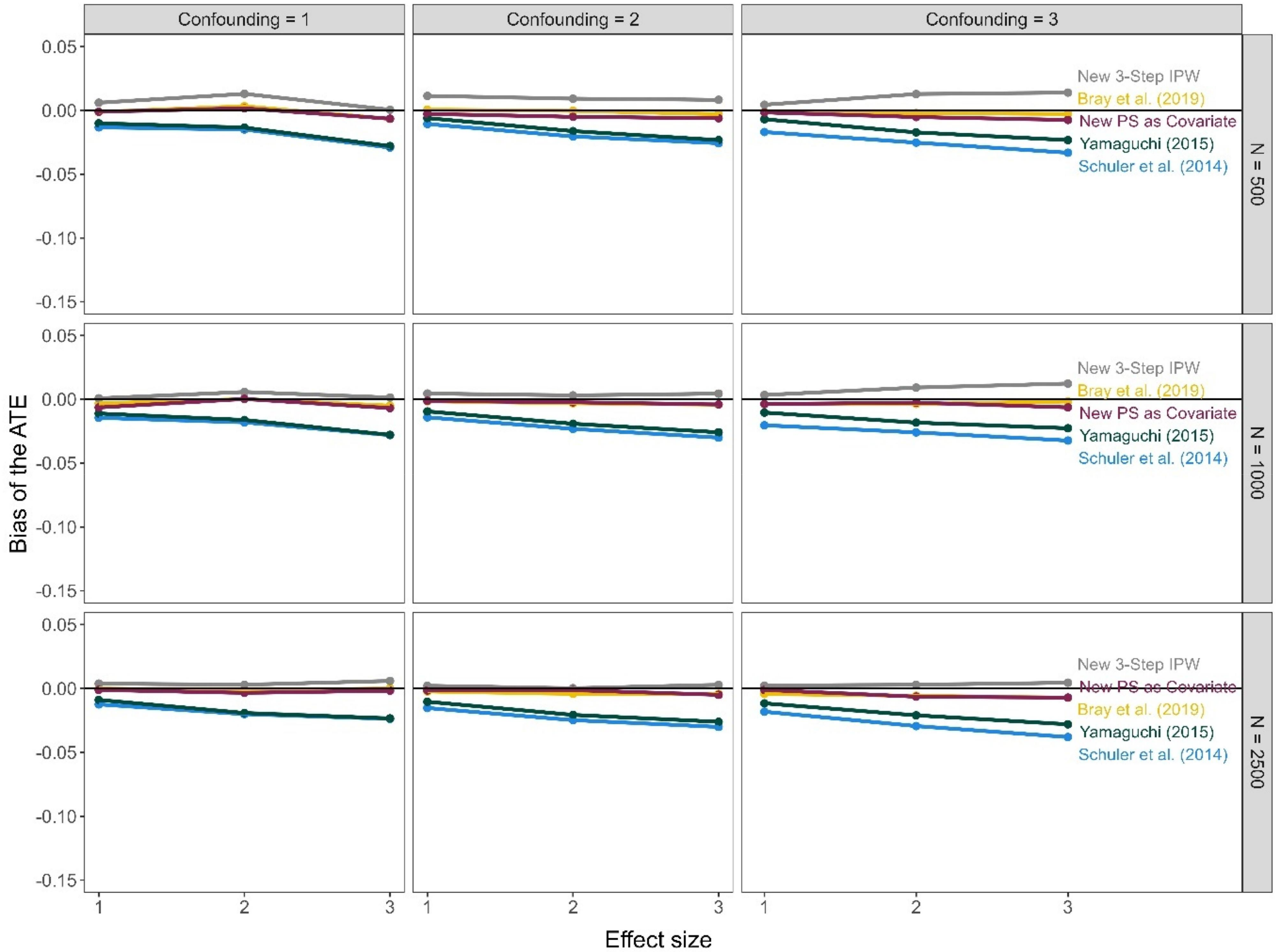

Figure A3. Results for the bias of the ATE for the binary distal outcome variable and good class separation. The bias is presented for varying levels of effect size, strength of the confounding effect, and sample size. Results are averaged over 500 replications.

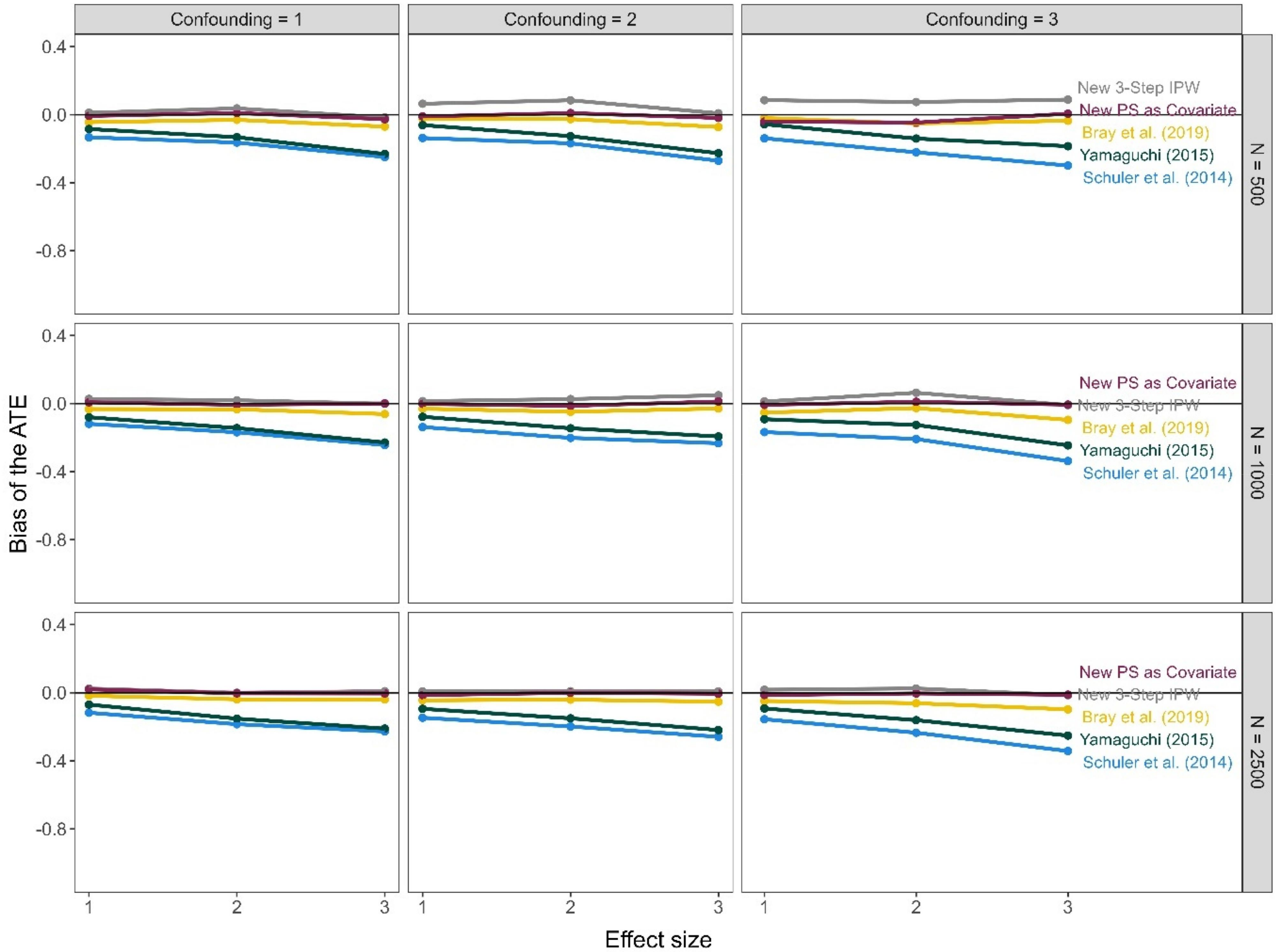

Figure A4. Results for the bias of the ATE for the continuous distal outcome variable and good class separation. The bias is presented for varying levels of effect size, strength of the confounding effect, and sample size. Results are averaged over 500 replications.

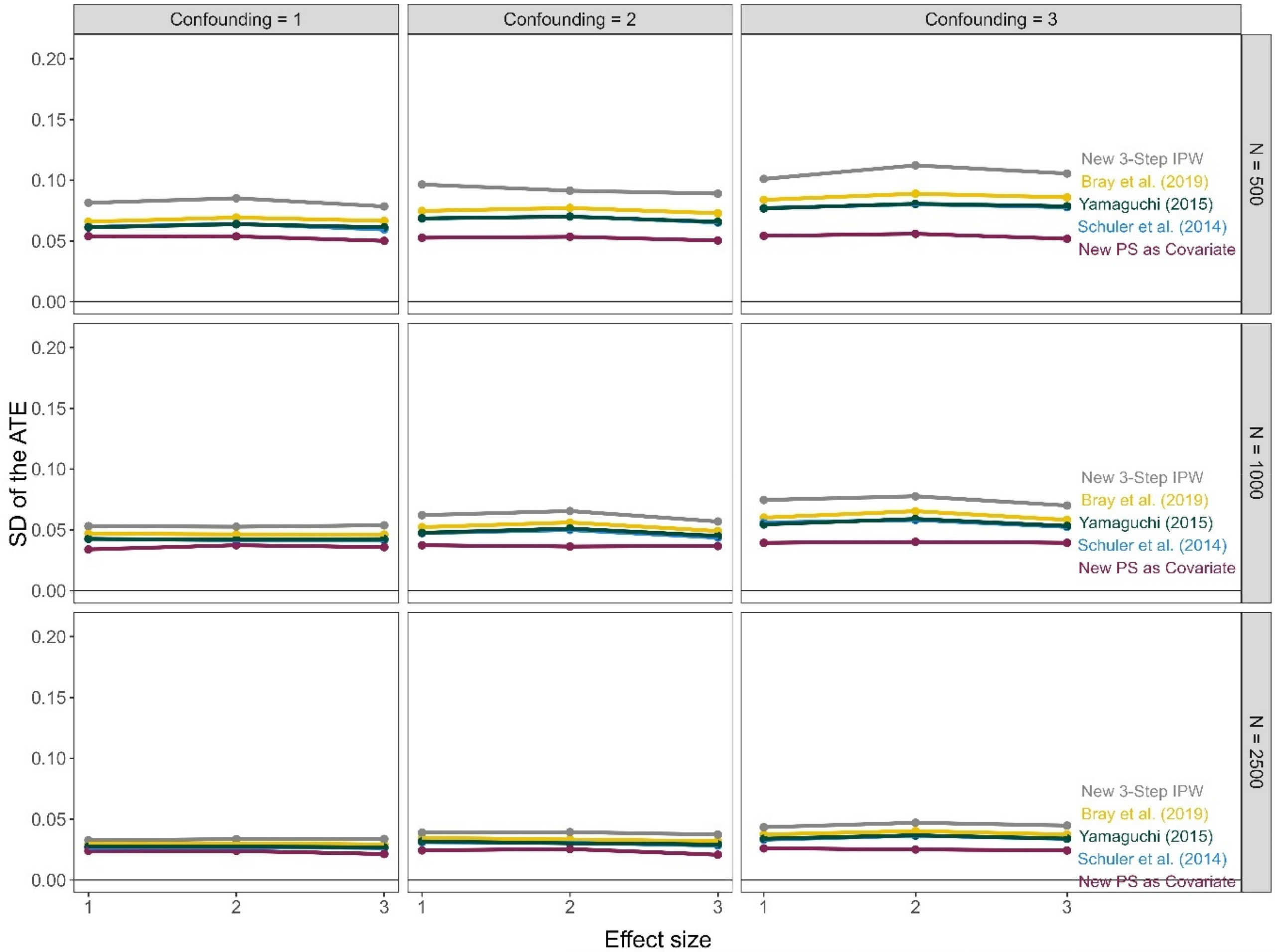

Figure A5. Results for the standard deviation (SD) of the ATE for the binary distal outcome variable and good class separation. The bias is presented for varying levels of effect size, strength of the confounding effect, and sample size. Results are averaged over 500 replications.

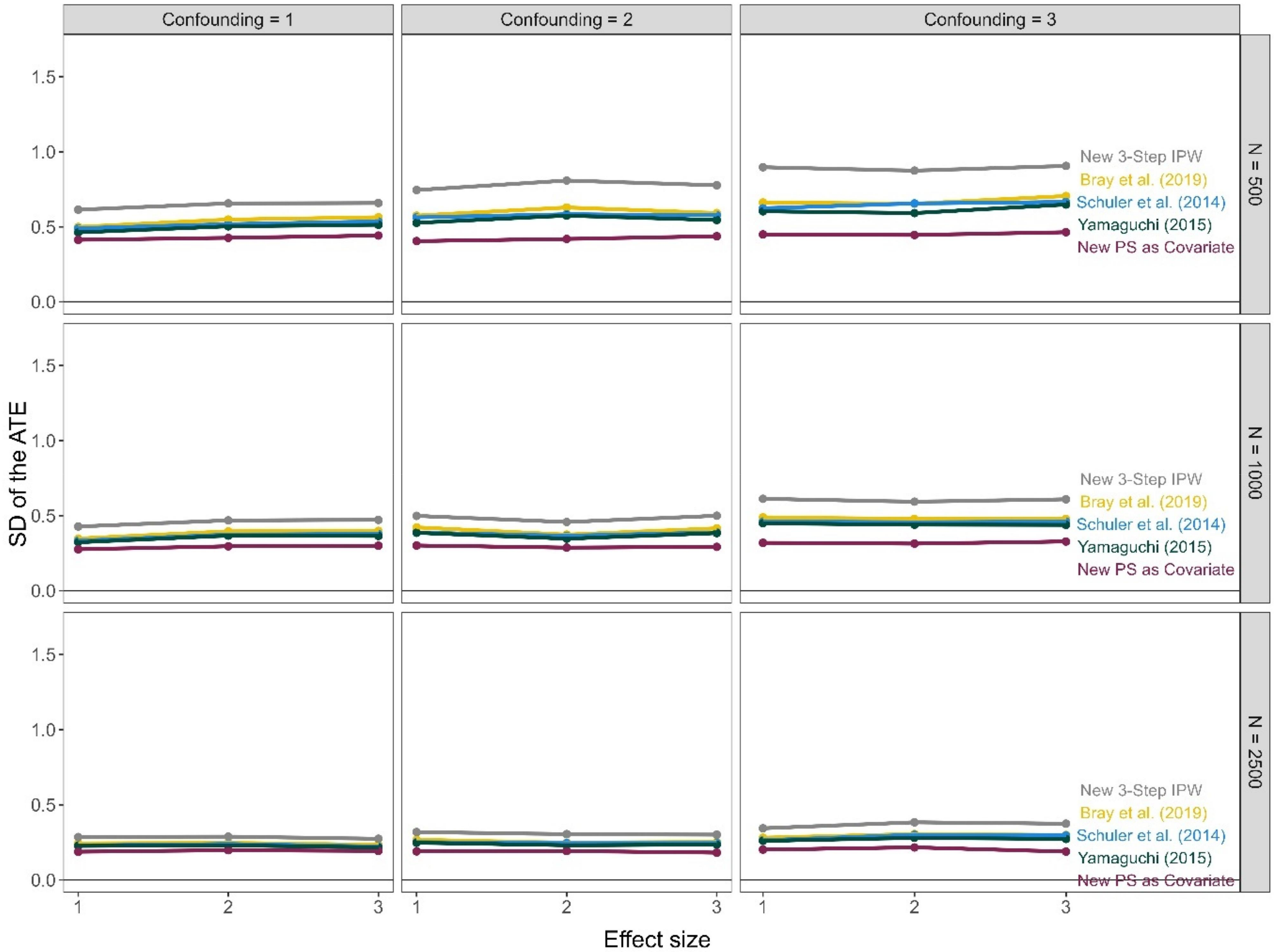

Figure A6. Results for the standard deviation (SD) of the ATE for the continuous distal outcome variable and good class separation. The bias is presented for varying levels of effect size, strength of the confounding effect, and sample size. Results are averaged over 500 replications.

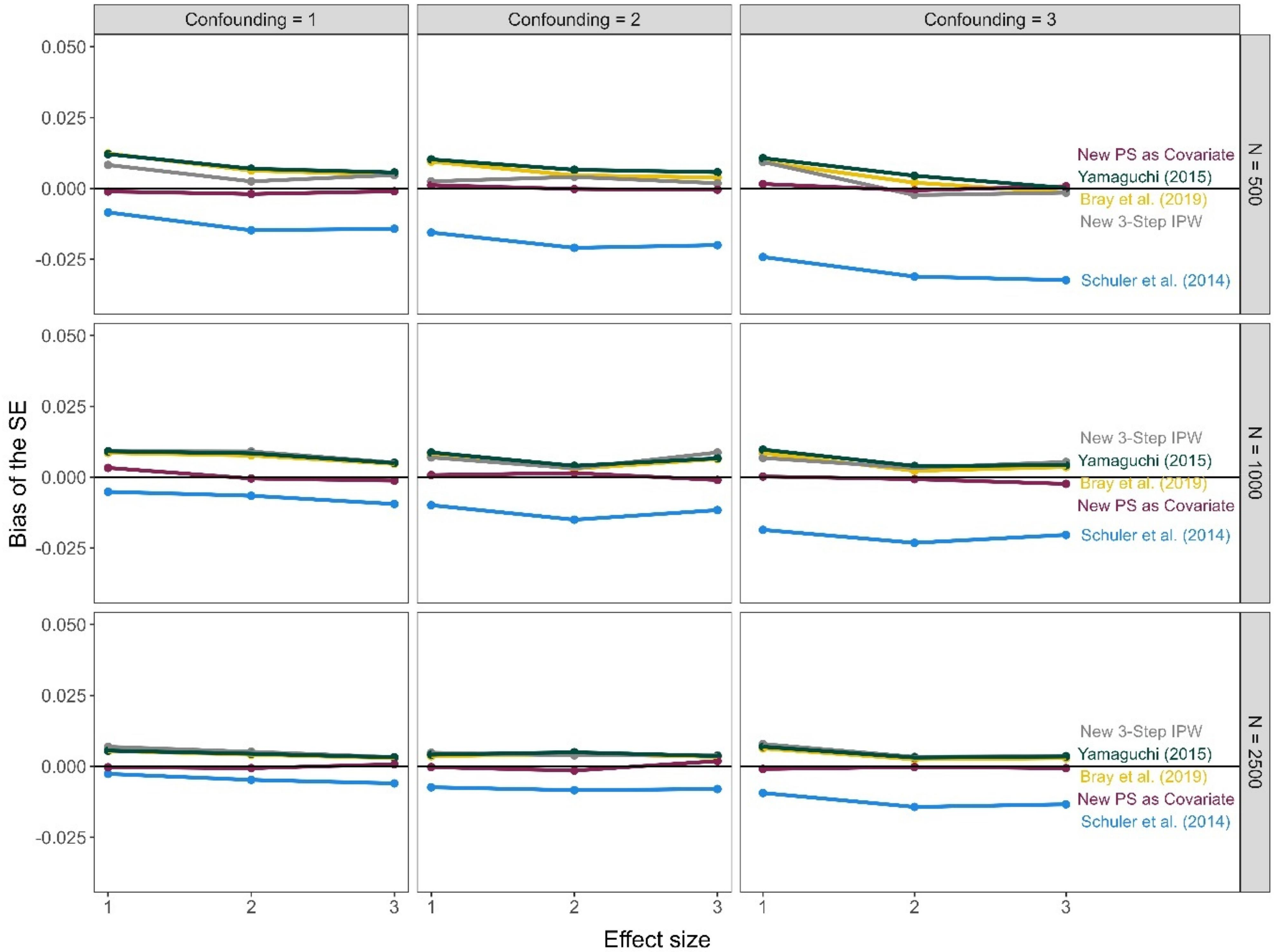

Figure A7. Results for the bias of the standard error (SE) of the ATE for the binary distal outcome variable and good class separation. The bias is presented for varying levels of effect size, strength of the confounding effect, and sample size. Results are averaged over 500 replications.

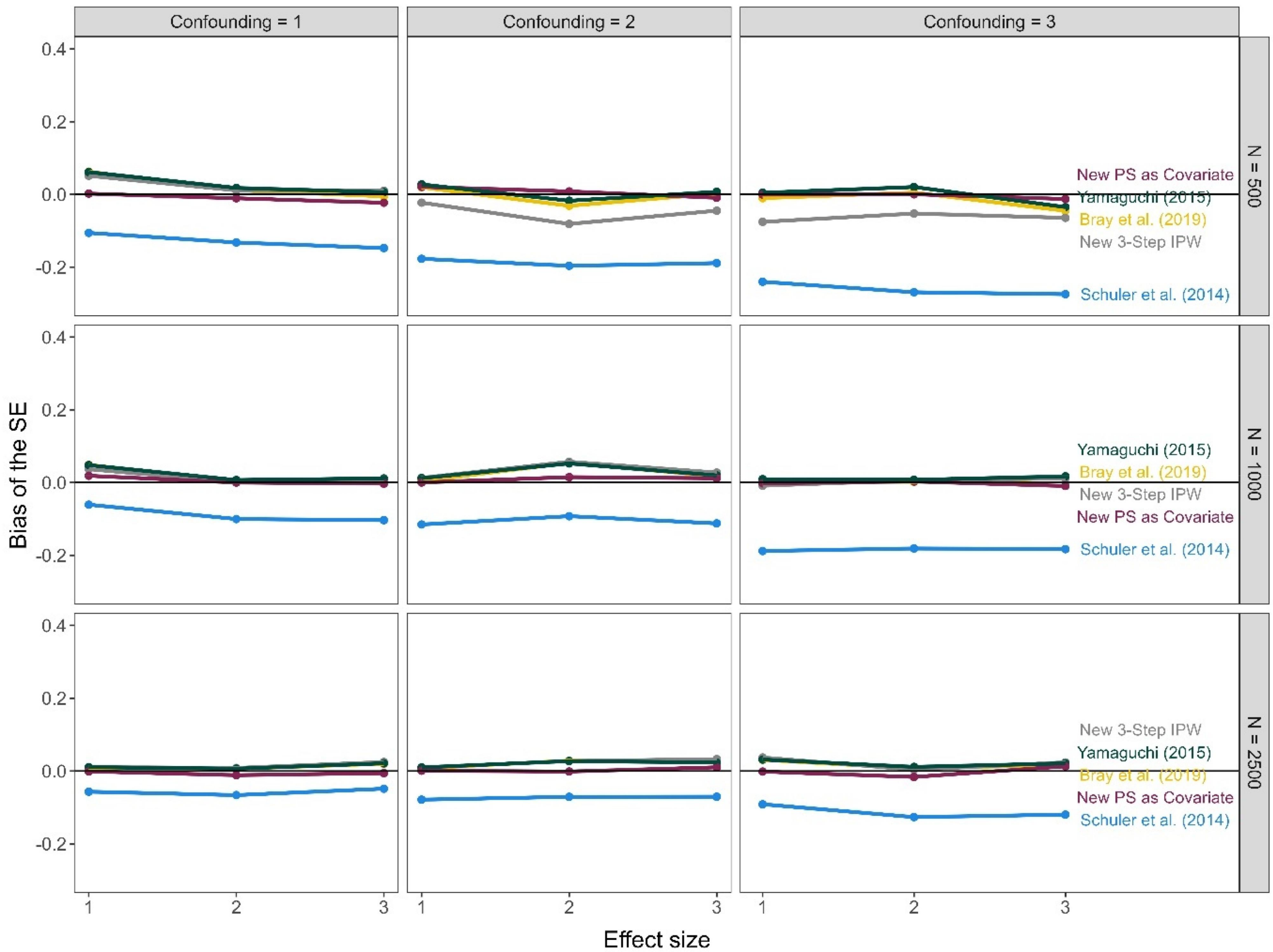

Figure A8. Results for the bias of the standard error (SE) of the ATE for the continuous distal outcome variable and good class separation. The bias is presented for varying levels of effect size, strength of the confounding effect, and sample size. Results are averaged over 500 replications.

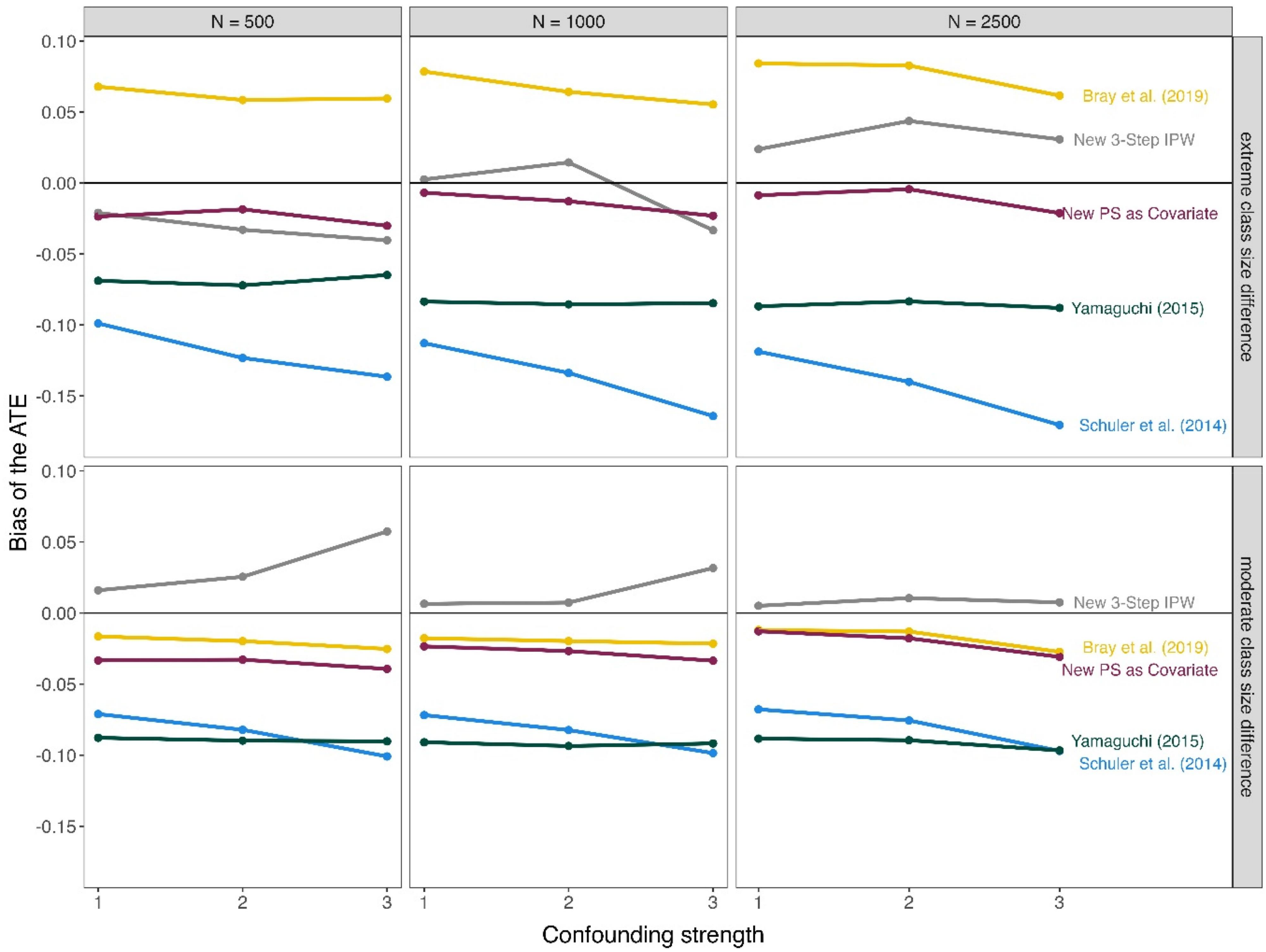

Figure A9. Results for the bias of the ATE for the binary distal outcome variable and good class separation. The bias is presented for varying levels of class size difference, strength of the confounding effect, and sample size. Results are averaged over 500 replications.

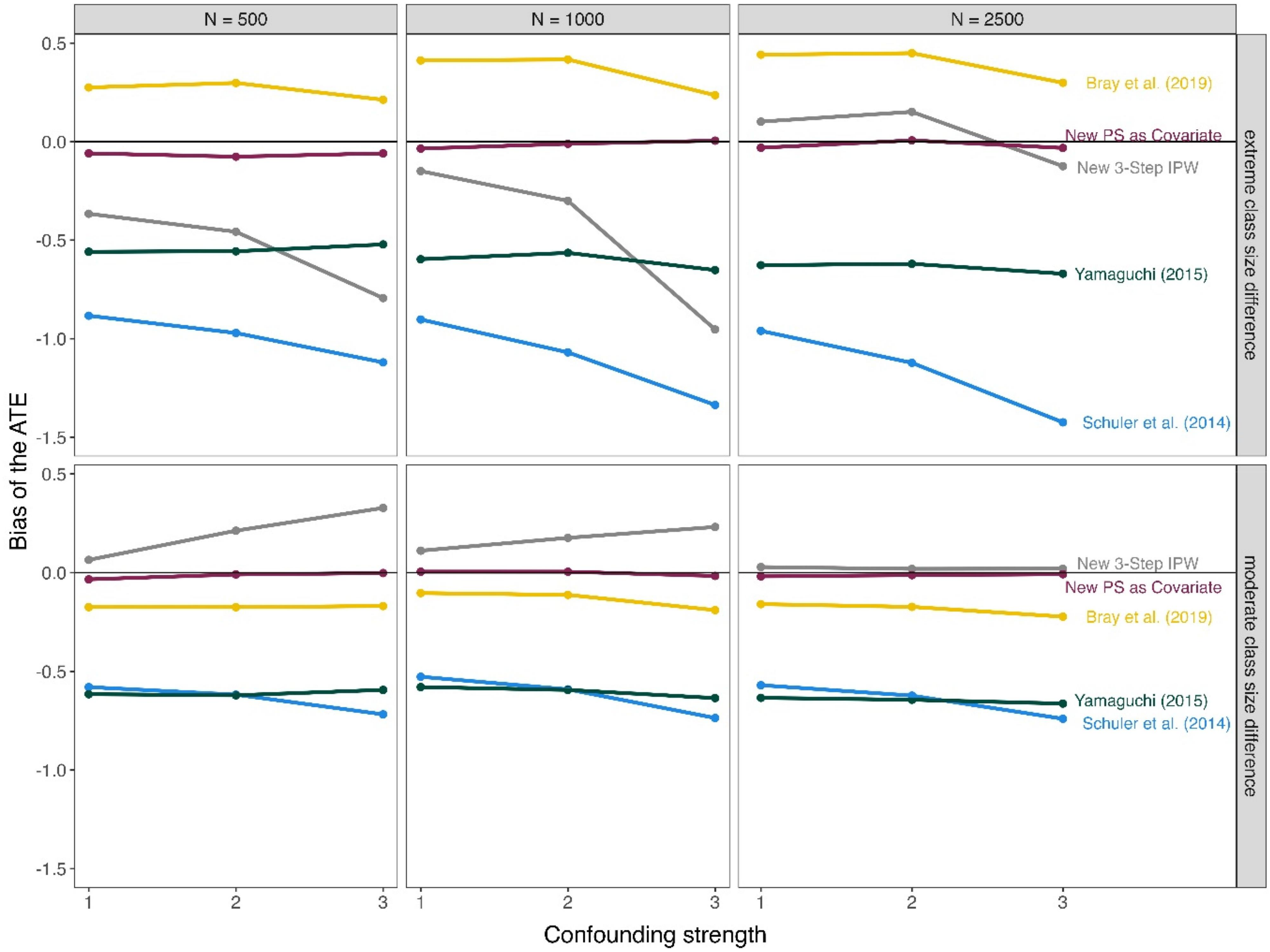

Figure A10. Results for the bias of the ATE for the continuous distal outcome variable and good class separation. The bias is presented for varying levels of class size difference, strength of the confounding effect, and sample size. Results are averaged over 500 replications.

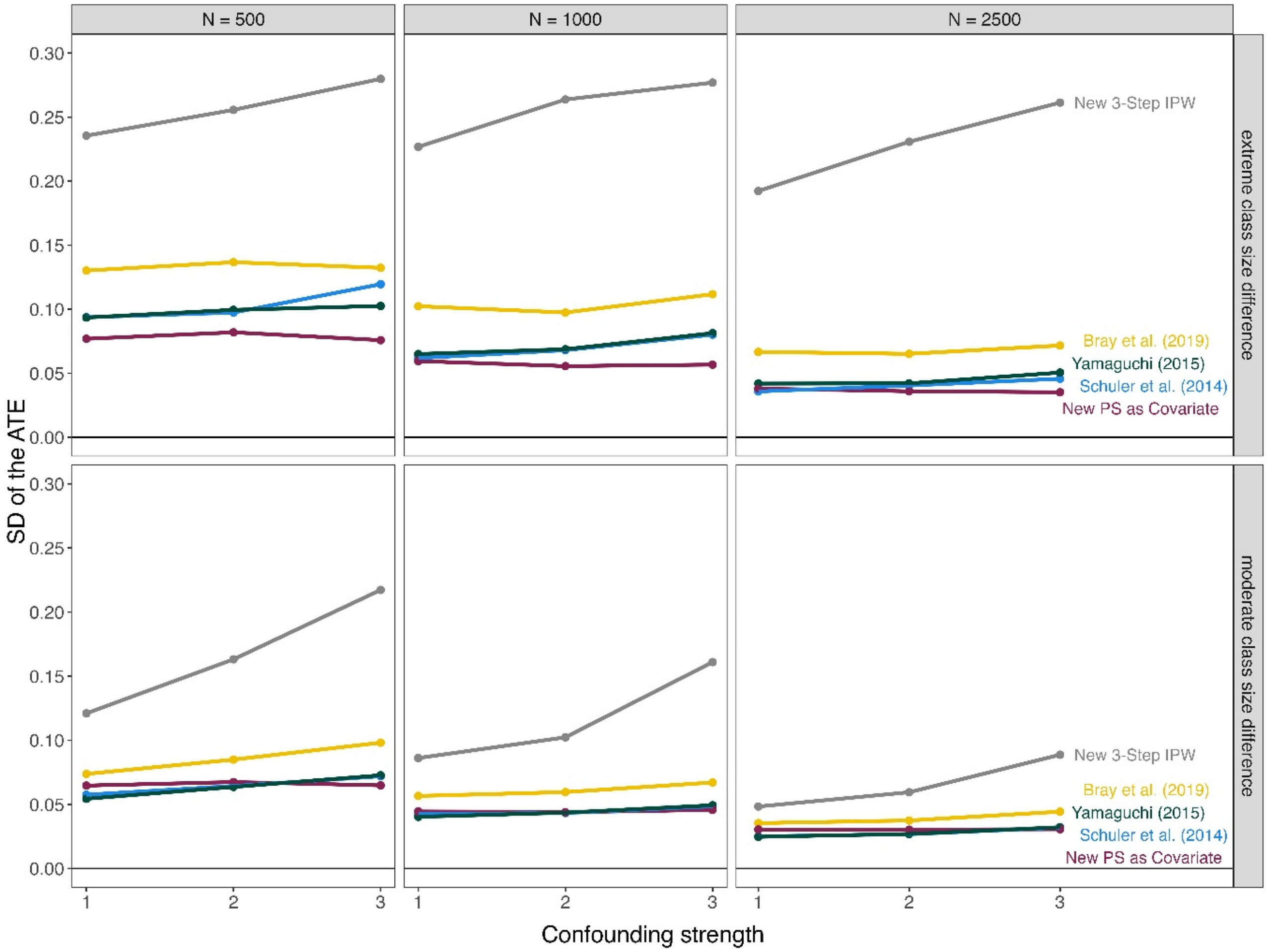

Figure A11. Results for the standard deviation (SD) of the ATE for the binary distal outcome variable and good class separation. The bias is presented for varying levels of class size difference, strength of the confounding effect, and sample size. Results are averaged over 500 replications.

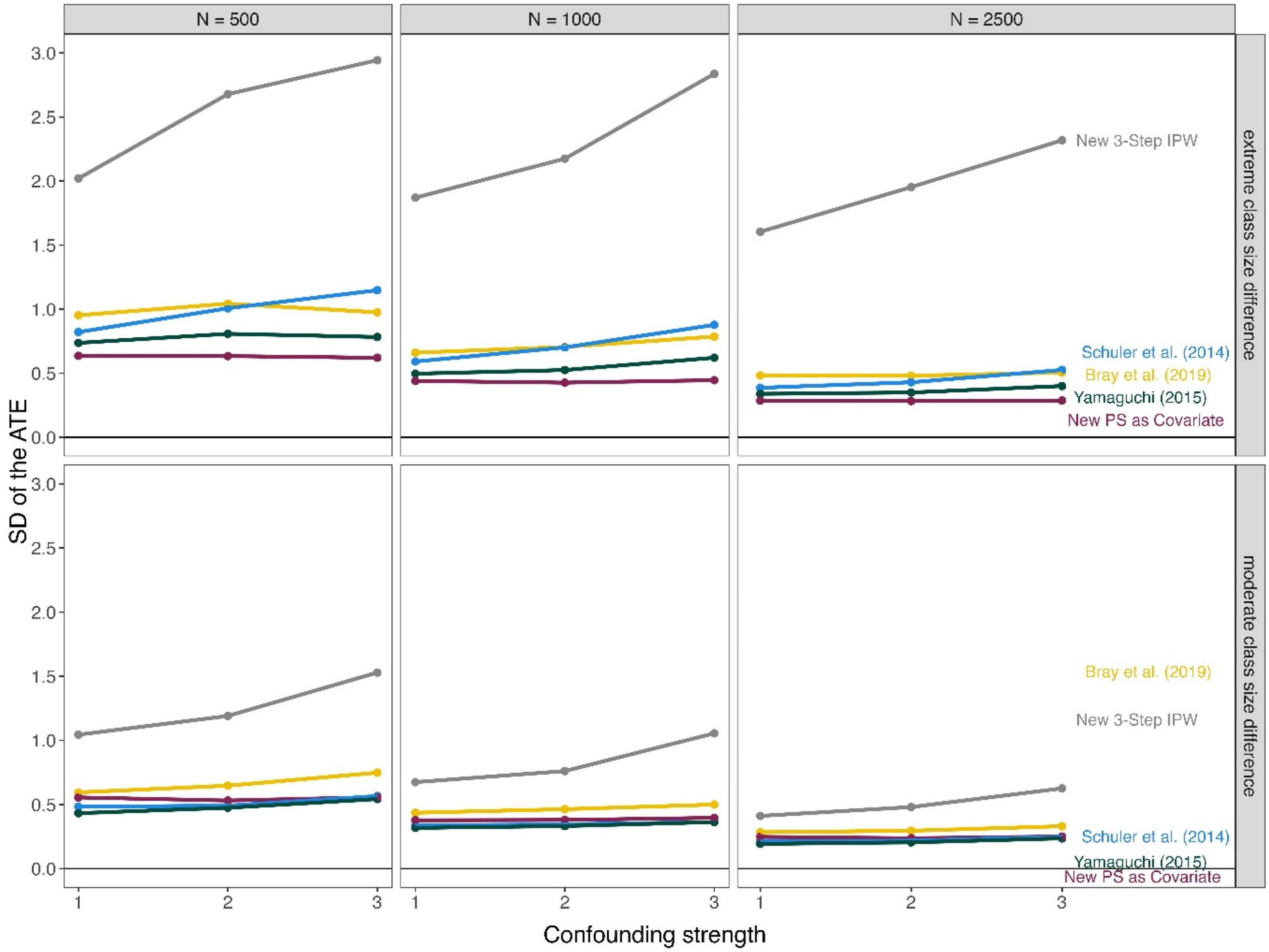

Figure A12. Results for the standard deviation (SD) of the ATE for the continuous distal outcome variable and good class separation. The bias is presented for varying levels of class size difference, strength of the confounding effect, and sample size. Results are averaged over 500 replications.

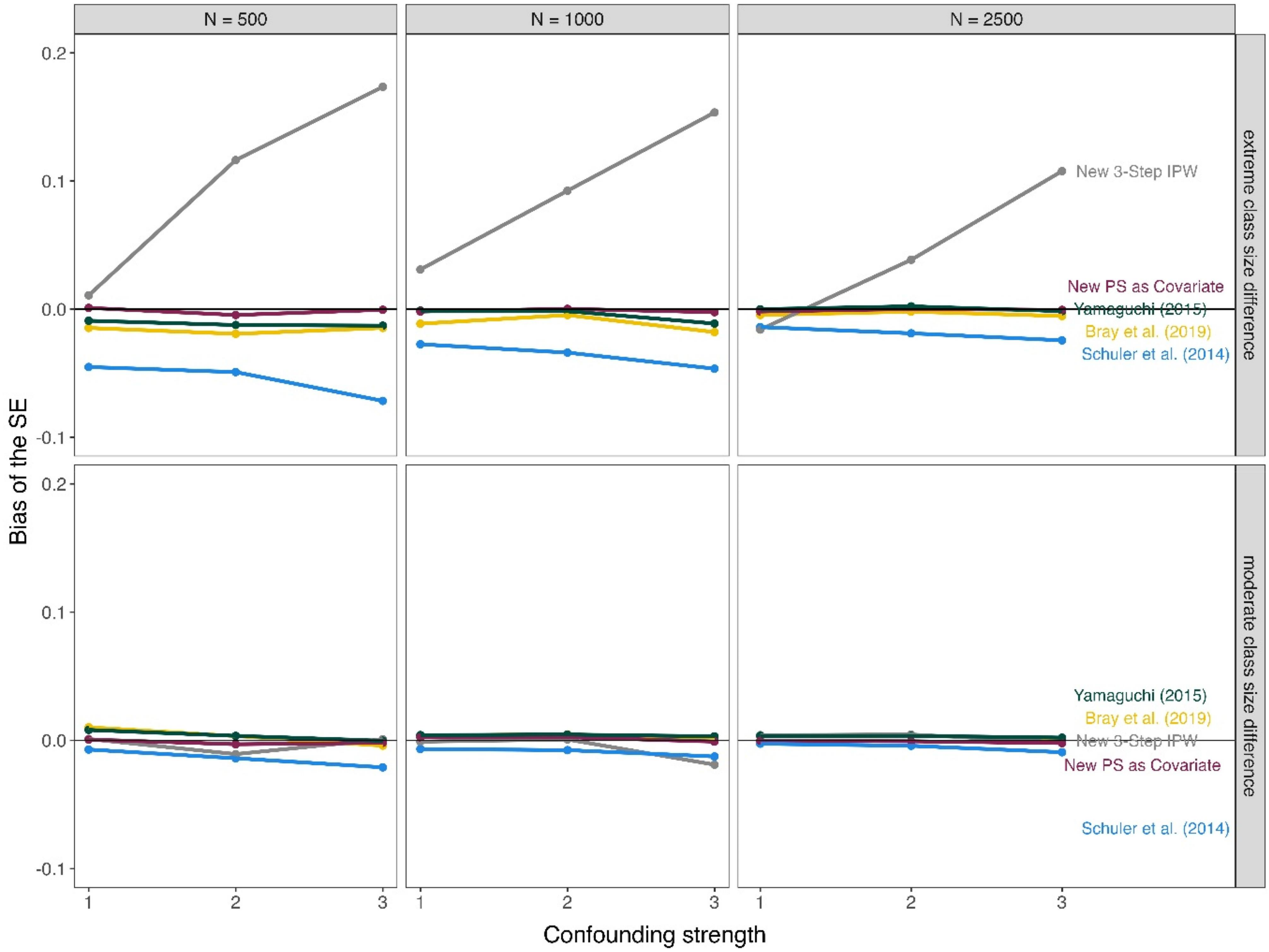

Figure A13. Results for the bias of the standard error (SE) of the ATE for the binary distal outcome variable and good class separation. The bias is presented for varying levels of class size difference, strength of the confounding effect, and sample size. Results are averaged over 500 replications.

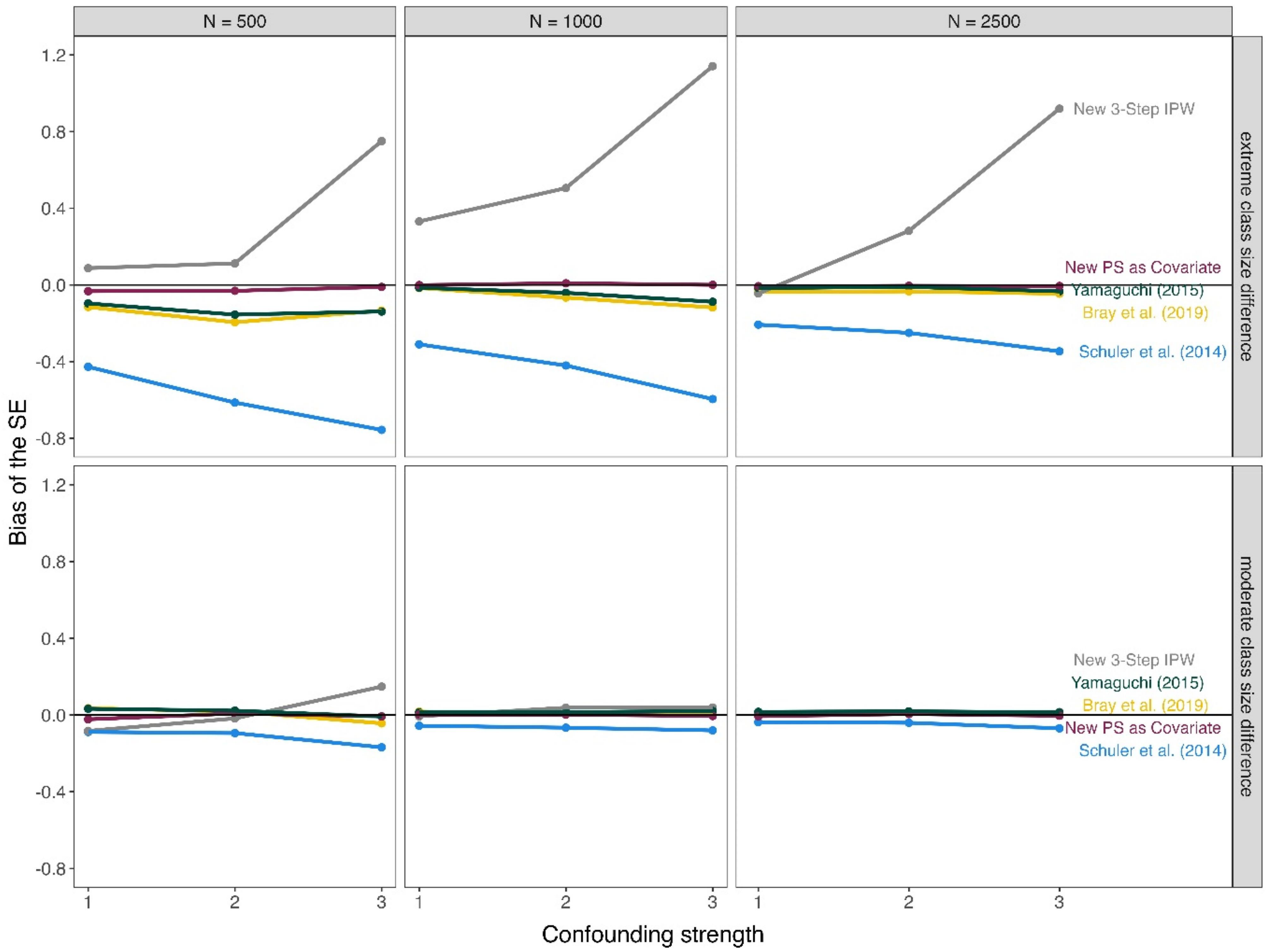

Figure A14. Results for the bias of the standard error (SE) of the ATE for the continuous distal outcome variable and good class separation. The bias is presented for varying levels of class size difference, strength of the confounding effect, and sample size. Results are averaged over 500 replications.

Table A1. Number of replications in the moderate class separation condition with a binary outcome for which the new 3-Step IPW method did not converge.

Sample Size	Confounding Strength	Effect Size	Number of Replications
500	1	1	4
500	1	2	8
500	1	3	2
500	2	1	7
500	2	2	6
500	2	3	8
500	3	1	13
500	3	2	15
500	3	3	16
1000	1	2	1
1000	1	3	1
1000	2	3	1
1000	3	2	2
1000	3	3	3

Table A2. Number of replications in the moderate class separation condition with a continuous outcome for which the new 3-Step IPW method did not converge.

Sample size	Confounding strength	Effect size	Number of replications
500	1	1	7
500	1	2	3
500	1	3	4
500	2	1	7
500	2	2	10
500	2	3	8
500	3	1	18
500	3	2	9
500	3	3	21
1000	1	3	1
1000	2	3	1
1000	3	1	3
1000	3	2	1
1000	3	3	3
2500	3	3	1