Received: 12 March 2021 | Accepted: 26 August 2021

DOI: 10.1111/biom.13683

Biometrics A JOURNAL OF THE INTERNATIONAL BIOMETRIC SOCIETY WILEY

READER REACTION

Reader reaction to "Outcome-adaptive lasso: Variable selection for causal inference" by Shortreed and Ertefaie (2017)

Ismaila Baldé[1] | **Yi Archer Yang**[2] | **Geneviève Lefebvre**[1]

[1]Département de mathématiques, Université du Québec à Montréal, Montréal, Québec, Canada

[2]Department of Mathematics and Statistics, McGill University, Montréal, Québec, Canada

Correspondence
Geneviève Lefebvre, Département de mathématiques, Université du Québec à Montréal, C.P. 8888, Succursale Centre-ville, Montréal, QC H3C 3P8, Canada.
Email: lefebvre.gen@uqam.ca

Funding information
Institut des sciences mathématiques; Fonds de Recherche du Québec - Santé; Natural Sciences and Engineering Research Council of Canada

Abstract

Shortreed and Ertefaie introduced a clever propensity score variable selection approach for estimating average causal effects, namely, the outcome adaptive lasso (OAL). OAL aims to select desirable covariates, confounders, and predictors of outcome, to build an unbiased and statistically efficient propensity score estimator. Due to its design, a potential limitation of OAL is how it handles the collinearity problem, which is often encountered in high-dimensional data. As seen in Shortreed and Ertefaie, OAL's performance degraded with increased correlation between covariates. In this note, we propose the generalized OAL (GOAL) that combines the strengths of the adaptively weighted L_1 penalty and the elastic net to better handle the selection of correlated covariates. Two different versions of GOAL, which differ in their procedure (algorithm), are proposed. We compared OAL and GOAL in simulation scenarios that mimic those examined by Shortreed and Ertefaie. Although all approaches performed equivalently with independent covariates, we found that both GOAL versions were more performant than OAL in low and high dimensions with correlated covariates.

KEYWORDS

adaptive elastic net, causal inference, high-dimensional data, propensity score, variable selection

1 | INTRODUCTION

In a very interesting paper, Shortreed and Ertefaie (2017) introduced the outcome adaptive lasso (OAL) approach for variable selection in the causal inference framework. OAL was designed to target confounders and predictors of outcome, while excluding spurious covariates and covariates only associated with exposure. As explained therein, variables selected that way aim to yield an unbiased and efficient propensity score (PS) estimator. Shortreed and Ertefaie (2017) theoretically and empirically demonstrated that OAL is able to select all true confounders and predictors of outcome, and exclude the rest of covariates. The performance of the algorithm was examined in situations wherein the number of predictors was small or large relative to the number of observations. Indeed, the authors suggested that OAL is also adequate to be used in high-dimensional problems (p increasing with n), which are common in causal inference.

In high dimension, an ideal variable selection approach should enjoy the oracle property and deal with the collinearity problem (Zou & Zhang, 2009) that typically plagues such settings. OAL is based on the adaptive lasso method (Zou, 2006) and features the oracle property, but its ability to properly treat correlated predictors is questionable. As seen in simulation scenarios presented in Shortreed and Ertefaie (2017), OAL was observed increasingly biased and variable as the correlation between predictors increased (see Web Appendices on: https://onlinelibrary.wiley.com/doi/10.1111/biom.12679).

Similar degraded performance is known for lasso, where unstable solution paths are obtained when predictors are highly correlated (Zou & Hastie, 2005).

It has previously been shown that elastic net can overcome the collinearity problem exhibited by lasso (Zou & Hastie, 2005). Moreover, it is possible to transform the elastic net problem into an equivalent lasso problem by augmenting the data (Zou & Hastie, 2005). This last idea was transposed to adaptive elastic net, which permits variable selection consistency and encourages grouping effect that is either selection or omission of correlated variables together (Ghosh, 2011). Building on these works, we propose an outcome adaptive elastic net approach to improve on OAL, which we name the generalized OAL (GOAL). Our idea is to start from an outcome adaptive elastic net problem that can be transformed into Shortreed and Ertefaie's OAL representation by augmenting the data. Two different versions of our approach, namely, Naive GOAL and GOAL with PIRLS, were explored.

The first version solves the GOAL problem naively using the R function *lqa*. Developed by Ulbricht (2010) to fit penalized generalized linear models, *lqa* was used by Shortreed and Ertefaie (2017) to solve the OAL problem. To implement the proposed Naive GOAL estimator, only straightforward data manipulations are required before calling the R code provided by Shortreed and Ertefaie (2017) for OAL. The second version solves the GOAL problem via a penalized iteratively reweighted least squares (PIRLS) procedure. GOAL with PIRLS is based on a modified *lqa* function (referred herein as *mlqa*), which modifies the working response and weights of the Newton–Raphson update in the *lqa* function. To implement GOAL with the PIRLS estimator, we substitute the *lqa* function by *mlqa* in the R function.

In our work, which picks up simulation scenarios presented in Shortreed and Ertefaie (2017), both versions of GOAL were observed to perform similarly to OAL when predictors were uncorrelated. However, GOAL was seen to offer better performance than OAL when correlation between predictors was present.

Our note is structured as follows. Section 2 contains an overview of the methods. We present the GOAL approach in Section 2.1 and provide two simple algorithms (versions) for its implementation in Section 2.2. We describe the simulation study in Section 3 and corresponding results are presented in Section 4. We conclude with a discussion in Section 5.

2 | METHODS

We introduce the GOAL approach along with the data augmentation step that underlies both versions of our GOAL estimator. In our presentation, we adopt Shortreed and Ertefaie (2017)'s notation to describe variables and models. More precisely, we let $(\mathbf{X}, A, Y)$ denote the triplet of design matrix, treatment, and response, respectively, where $\mathbf{X} = (X_1, X_2, \ldots, X_p)$.

2.1 | Generalized outcome adaptive lasso

We briefly recall Shortreed and Ertefaie (2017) method (OAL) before introducing the proposed approach (GOAL). We assume the following PS model parametrized by α:

$$logit\{P(A=1|\mathbf{X})\} = \sum_{j=1}^{p} \alpha_j X_j.$$

Let $\ell_n(\alpha; A, \mathbf{X}) = \sum_{i=1}^{n}\{-a_i(x_i^T\alpha) + \log(1+e^{x_i^T\alpha})\}$ be the negative log-likelihood. OAL is an adaptive lasso penalty for logistic regression; it is defined as

$$\hat{\alpha}(OAL) = \arg\min_{\alpha} \left[\ell_n(\alpha; A, \mathbf{X}) + \lambda_n \sum_{j=1}^{p} \hat{\omega}_j |\alpha_j|\right], \quad (1)$$

where $\hat{\omega}_j = |\hat{\beta}_j^{ols}|^{-\gamma}$ such that $\gamma > 1$ and $(\hat{\beta}_A^{ols}, \hat{\beta}^{ols}) = \arg\min_{(\beta_A, \beta)} \|Y - \beta_A A - \mathbf{X}\beta\|_2^2$.

Shortreed and Ertefaie (2017) used the R function *lqa* to solve problem (1). Moreover, they proposed to minimize a weighted absolute mean difference (wAMD) between the treated and untreated groups to select the tuning parameter λ_n in the set

$$S_{\lambda_n} = \{n^{-10}, n^{-5}, n^{-2}, n^{-1}, n^{-0.75}, n^{-0.5}, n^{-0.25}, n^{0.25}, n^{0.49}\},$$

that is $\hat{\lambda}_n = \arg\min_{\lambda_n \in S_{\lambda_n}} wAMD(\lambda_n; \mathbf{X}, A)$, where

$$wAMD(\lambda_n; \mathbf{X}, A) = \sum_{j=1}^{p} \left|\hat{\beta}_j^{ols}\right| \left| \frac{\sum_{i=1}^{n} \hat{\tau}_i^{\lambda_n} X_{ij} A_i}{\sum_{i=1}^{n} \hat{\tau}_i^{\lambda_n} A_i} - \frac{\sum_{i=1}^{n} \hat{\tau}_i^{\lambda_n} X_{ij}(1-A_i)}{\sum_{i=1}^{n} \hat{\tau}_i^{\lambda_n}(1-A_i)} \right| \quad (2)$$

and $\hat{\tau}_i^{\lambda_n}$ is the inverse probability of treatment weight for individual i constructed using the PS model fitted from Equation (1).

Building upon the adaptive elastic net estimator (see Web Appendix A for a review), we define the GOAL estimator through the following optimization problem:

$$\hat{\alpha}(GOAL) = \arg\min_{\alpha} \left[\ell_n(\alpha; A, \mathbf{X}) + \lambda_1 \sum_{j=1}^{p} \hat{\omega}_j |\alpha_j| + \lambda_2 \sum_{j=1}^{p} \alpha_j^2\right], \quad (3)$$

where $\hat{\omega}_j$ is defined as in (1).

In the sequel, we propose two ways to solve problem (3) and obtain GOAL estimates.

Naive GOAL

Lemma 1 in Zou and Hastie (2005) was initially proposed for linear models to reexpress the elastic net problem into a lasso penalty. Algamal and Lee (2015a) directly applied this lemma for logistic regression to transform elastic net into a lasso problem on augmented data. Building on these works as well as on Ghosh (2011), Naive GOAL adopts an augmented adaptive lasso representation for logistic regression, as described below. Given the original design matrix and treatment data $(\mathbf{X}, A)$ and fixed (λ_1, λ_2), we create an augmented data set $(\mathbf{X}^*, A^*)$: $\mathbf{X}^* = \begin{pmatrix} \mathbf{X} \\ \sqrt{\lambda_2}\mathbf{I}_p \end{pmatrix}$, $A^* = \begin{pmatrix} A \\ 0_p \end{pmatrix}$, where $\mathbf{X}^* = (X_1^*, X_2^*, \ldots, X_p^*)$. Following Zou and Hastie (2005) and Algamal and Lee (2015a), we reexpress the GOAL estimator (3) as an OAL problem on the data $(\mathbf{X}^*, A^*)$:

$$\hat{\alpha}_N(GOAL) = \arg\min_{\alpha} \left[\ell_{n^*}(\alpha; A^*, \mathbf{X}^*) + \lambda_1 \sum_{j=1}^{p} \hat{w}_j |\alpha_j| \right], \tag{4}$$

where $n^* = n + p$.

The solution to (4) can be obtained using the *lqa* function without any further modification.

GOAL with PIRLS

Tibshirani (1996) solved the lasso problem for logistic regression by applying the original lasso algorithm for linear regression at each step of the PIRLS method. That is, the L_1-penalized logistic regression is viewed as a lasso-weighted least squares (lasso-WLS) problem at each iteration of the PIRLS algorithm. To solve the OAL problem, Shortreed and Ertefaie (2017) relied upon the *lqa* (Ulbricht, 2010) function to fit the penalized logistic likelihood using the PIRLS technique. Let $\tilde{\alpha}$ be the current estimate of α in the PIRLS and ℓ_Q be the quadratic approximation of ℓ_n. The Newton–Raphson update solution of OAL is obtained as

$$\hat{\alpha}_{PIRLS}(OAL) = \arg\min_{\alpha} \left[\ell_Q(\alpha; A, \mathbf{X}, Z, \mathbf{T}) + \lambda_1 \sum_{j=1}^{p} \hat{w}_j |\alpha_j| \right],$$

where $\ell_Q(\alpha; A, \mathbf{X}, Z, \mathbf{T}) = \frac{1}{2}\sum_{i=1}^{n} t_i (z_i - x_i^T \alpha)^2$, $\tilde{p}(x_i) = \frac{1}{1+\exp(-x_i^T \tilde{\alpha})}$,
$t_i = \tilde{p}(x_i)[1 - \tilde{p}(x_i)]$, $z_i = x_i^T \tilde{\alpha} + \frac{a_i - \tilde{p}(x_i)}{\tilde{p}(x_i)(1-\tilde{p}(x_i))}$, $Z = (z_1, \ldots, z_n)^T$, $\mathbf{T} = \mathrm{diag}(t_1, \ldots, t_n)$.

Similarly to OAL, the Newton–Raphson update solution of GOAL is obtained as

$$\hat{\alpha}_{PIRLS}(GOAL) = \arg\min_{\alpha} \left[\ell_Q(\alpha; A, \mathbf{X}, Z, \mathbf{T}) + \lambda_1 \sum_{j=1}^{p} \hat{w}_j |\alpha_j| + \lambda_2 \sum_{j=1}^{p} \alpha_j^2 \right] \tag{5}$$

$$= \arg\min_{\alpha} \left[\ell_{Q^*}(\alpha; A^*, \mathbf{X}^*, Z^*, \mathbf{T}^*) + \lambda_1 \sum_{j=1}^{p} \hat{w}_j |\alpha_j| \right], \tag{6}$$

where $\mathbf{X}^* = \begin{pmatrix} \mathbf{X} \\ \sqrt{\lambda_2}\mathbf{I}_p \end{pmatrix}$, $A^* = \begin{pmatrix} A \\ 0_p \end{pmatrix}$, $Z^* = \begin{pmatrix} Z \\ 0_p \end{pmatrix}$, and $\mathbf{T}^* = \begin{pmatrix} \mathbf{T} & 0_{n\times p} \\ 0_{n\times p}^T & \mathbf{I_p} \end{pmatrix}$.

We prove the equality between Equations (5) and (6) in Web Appendix B.

2.2 | Implementation of GOAL

Selection of tuning parameters is a fundamental aspect of penalized model fitting. As we defined the GOAL problem by using two tuning parameters (λ_1, λ_2), our proposal is to balance the exposure groups on a two-dimensional surface. Following Zou and Hastie (2005), we first consider relatively small values of λ_2:

$$S_{\lambda_2} = \{0, 10^{-2}, 10^{-1.5}, 10^{-1}, 10^{-0.75}, 10^{-0.5}, 10^{-0.25}, 10^{0}, 10^{0.25}, 10^{0.5}, 10^{1}\}.$$

Then for each fixed $\lambda_2 \in S_{\lambda_2}$, the GOAL algorithms solve the OAL problem with augmented data defined as a function of the λ_2 value. In this step, we recall that λ_1 is selected on the basis of the wAMD defined in Equation (2). The chosen (λ_1, λ_2) is the one maximizing the balance between exposure groups (i.e., corresponding to the smallest wAMD). In the sequel, we summarize the steps to implement both versions of GOAL in Algorithms 1 and 2, respectively.

As seen above, although our first GOAL version (Algorithm 1) applies naively to a logistic model a data augmentation step that was originally developed for a linear model, our second GOAL version (Algorithm 2) performs the data augmentation within each iteration of the PIRLS. Also note that in step 4 in Algorithm 1 and step 7 in Algorithm 2, we compute the adaptive elastic net as discussed in (Ghosh, 2011; Zou & Hastie, 2005).

ALGORITHM 1 Naive GOAL

1: **Input**: Design matrix and treatment data $(\mathbf{X}, A)$

2: For each fixed λ_2 define: $\mathbf{X}^* = \begin{pmatrix} \mathbf{X} \\ \sqrt{\lambda_2}\mathbf{I}_p \end{pmatrix}$ and $A^* = \begin{pmatrix} A \\ 0_p \end{pmatrix}$

3: Call OAL algorithm with augmented data $(\mathbf{X}^*, A^*)$ to solve $\hat{\alpha}_N^*$(naive adaptive elastic net) = $\arg\min_{\alpha} \left[\ell_{n^*}(\alpha; \mathbf{X}^*, A^*) + \lambda_1 \sum_{j=1}^{p} \hat{w}_j |\alpha_j| \right]$

4: Compute $\hat{\alpha}_N$(adaptive elastic net) = $(1+\lambda_2)\hat{\alpha}_N^*$(naive adaptive elastic net)

5: **Output**: $\hat{\alpha}_N$(adaptive elastic net)

ALGORITHM 2 GOAL with PIRLS

1: **Input**: Design matrix and treatment data $(\mathbf{X}, A)$

2: For each fixed λ_2 define: $\mathbf{X}^* = \begin{pmatrix} \mathbf{X} \\ \sqrt{\lambda_2}\mathbf{I}_p \end{pmatrix}$ and $A^* = \begin{pmatrix} A \\ 0_p \end{pmatrix}$

3: Initialize $\tilde{\alpha}$ to 0

4: Compute $\tilde{p}(x_i) = \frac{1}{1+\exp(-x_i^T\tilde{\alpha})}$, $t_i = \tilde{p}(x_i)[1-\tilde{p}(x_i)]$, $z_i = x_i^T\tilde{\alpha} + \frac{a_i - \tilde{p}(x_i)}{\tilde{p}(x_i)(1-\tilde{p}(x_i))}$, $i = 1, 2, \ldots, n$

5: Set $Z^* = \begin{pmatrix} Z \\ 0_p \end{pmatrix}$ and $\mathbf{T}^* = \begin{pmatrix} \mathbf{T} & 0_{n\times p} \\ 0_{n\times p}^T & \mathbf{I_p} \end{pmatrix}$, where $Z = (z_1, \ldots, z_n)^T, \mathbf{T} = \mathrm{diag}(t_1, \ldots, t_n)$

6: Call OAL algorithm with augmented data $(\mathbf{X}^*, A^*)$ to solve $\hat{\alpha}_I^*$(naive adaptive elastic net) = $\arg\min_{\alpha} \left[\ell_{Q^*}(\alpha; \mathbf{X}^*, A^*, Z^*, \mathbf{T}^*) + \lambda_1 \sum_{j=1}^{p} \hat{w}_j |\alpha_j| \right]$

7: Compute $\hat{\alpha}_I$(adaptive elastic net) = $(1+\lambda_2)\hat{\alpha}_I^*$(naive adaptive elastic net)

8: Update $\tilde{\alpha} = \hat{\alpha}_I$(adaptive elastic net)

9: Repeat $4-8$ until convergence of $\tilde{\alpha}$

10: Set $\tilde{\alpha}_I$(adaptive elastic net) $= \tilde{\alpha}$

11: **Output**: $\tilde{\alpha}_I$(adaptive elastic net)

3 | SIMULATION STUDY

The simulation study was designed to investigate the performance of GOAL, as compared to OAL, in higher and lower dimensional settings. We followed Shortreed and Ertefaie (2017) simulation setup to generate the data $(\mathbf{X}, A, Y)$. They simulated $X_i = (X_{i1}, X_{i2}, \ldots, X_{ip})_{1\le i\le n}$ from a multivariate standard Gaussian distribution with pairwise correlation ρ; binary treatment A from a Bernoulli distribution with $logit\{P(A_i = 1)\} = \sum_{j=1}^{p} \alpha_j X_{ij}$ and continuous outcome as $Y_i = \beta_A A_i + \sum_{j=1}^{p} \beta_j X_{ij} + \epsilon_i$ where $\epsilon_i \sim N(0,1)$ and $\beta_A = 0$ or 2. We varied both the sample size (n) and the number of covariates (p). To evaluate GOAL's performance, we examined all (n, p) combinations of the original paper of Shortreed and Ertefaie (2017), that is: $n = 200$ with $p = 100$ and $n = 500$ with $p = 200$ for the high-dimensional settings, and $n = 200, 500, 1000$ with fixed $p = 20$ for the low-dimensional settings.

In our simulations, we considered the same four scenarios as in Shortreed and Ertefaie (2017). These scenarios are defined as follows, where $\beta \in \mathbb{R}^p$ are the regression coefficients in the outcome model and $\alpha \in \mathbb{R}^p$ are the regression coefficients in the treatment model:

(a) Scenario 1 sets $\beta = (0.6, 0.6, 0.6, 0.6, 0, 0, 0, \ldots, 0)$ and $\alpha = (1, 1, 0, 0, 1, 1, 0, \ldots, 0)$;
(b) Scenario 2 sets $\beta = (0.6, 0.6, 0.6, 0.6, 0, 0, 0, \ldots, 0)$ and $\alpha = (0.4, 0.4, 0, 0, 1, 1, 0, \ldots, 0)$;
(c) Scenario 3 sets $\beta = (0.2, 0.2, 0.6, 0.6, 0, 0, 0, \ldots, 0)$ and $\alpha = (1, 1, 0, 0, 1, 1, 0, \ldots, 0)$;
(d) Scenario 4 sets $\beta = (0.6, 0.6, 0.6, 0.6, 0, 0, 0, \ldots, 0)$ and $\alpha = (1, 1, 0, 0, 1.8, 1.8, 0, \ldots, 0)$.

In each scenario, the first two covariates are confounders, the third and fourth covariates are outcome predictors (unrelated to treatment), the fifth and sixth covariates are exposure predictors (unrelated to outcome), and the rest are spurious covariates (i.e., $p - 6$ spurious covariates). Four different correlations ($\rho = 0, 0.2, 0.5, 0.75$) between covariates were investigated, where the first three values were considered by Shortreed and Ertefaie (2017). We refer the interested reader to the original paper (Shortreed & Ertefaie, 2017: Section 4.1, Section 6, and Web appendices) for more details on the simulation. For each scenario, we obtained estimates for the average treatment effect (ATE) using the IPTW estimator (Lunceford & Davidian, 2004) with the PS model fitted using either OAL or GOAL. We compared OAL and GOAL approaches based on the bias, standard error (SE), and mean squared error (MSE) of resulting IPTW estimators for the ATE. For variable selection, we used the proportion of times each predictor was selected for inclusion in the PS model (tolerance was 10^{-8}) under 1000 simulations.

4 | RESULTS

Figure 1 and Web Table C1 (first two sections) present results associated with Scenarios 1–2 in the high-dimensional settings ($p/n = 100/200, 200/500$), displaying the bias, SE and MSE of OAL and GOAL estimators for the ATE under a grid of increasing values for ρ (0, 0.2,

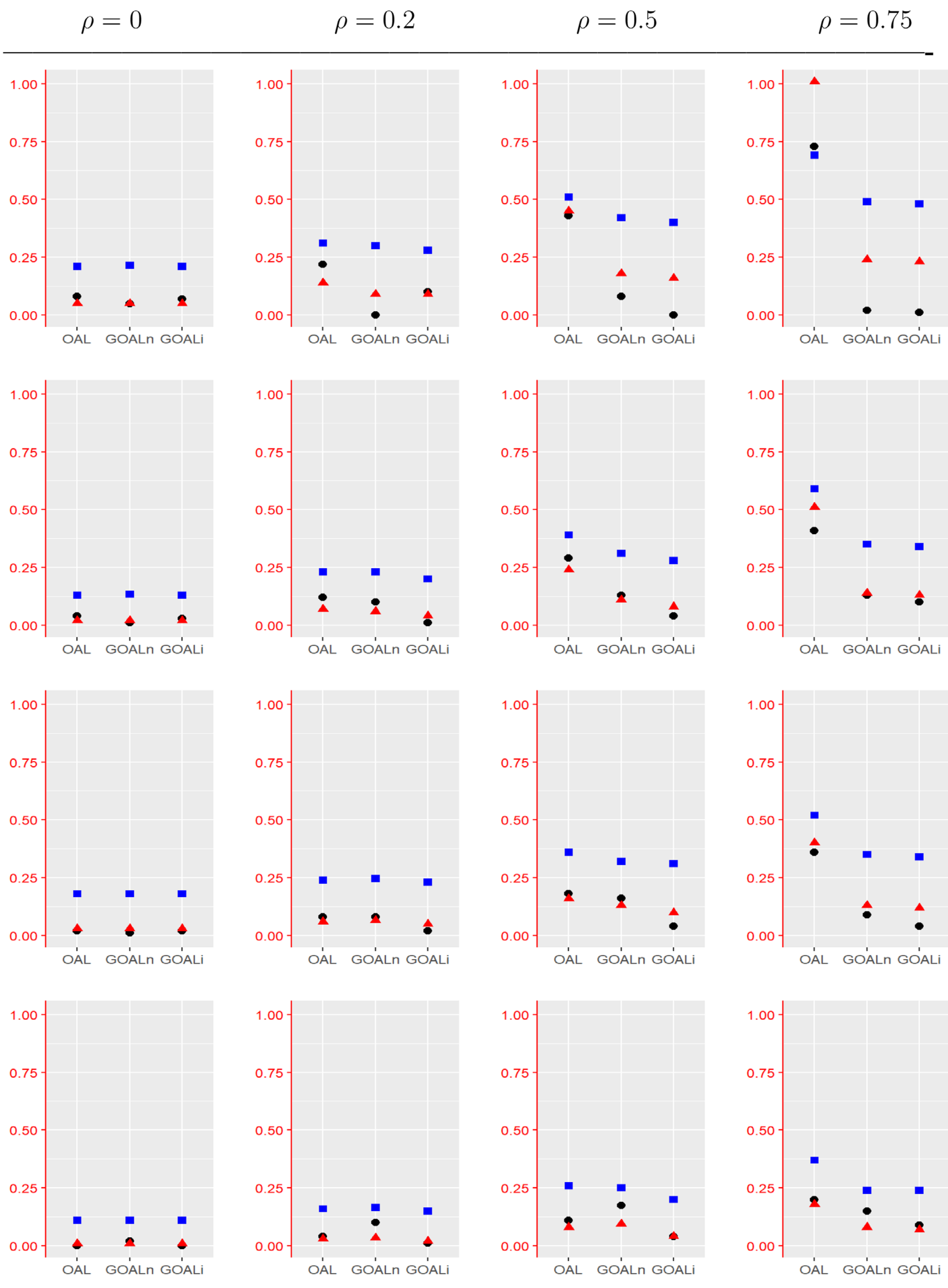

FIGURE 1 Absolute bias (circle), standard error (square), and mean squared error (triangle) of IPTW estimator for the average treatment effect (ATE) for OAL, naive GOAL (GOALn), and penalized iteratively re-weighted least squares GOAL (GOALi) under Scenarios 1 and 2 (based on 1000 IPTW estimates). The ratios p/n = 100/200, 200/500 are presented in rows 1 and 2, respectively, for Scenario 1 and in rows 3 and 4, respectively, for Scenario 2. This figure appears in color in the electronic version of this article

0.5, 0.75) when the true ATE is 0. Due to space constraints, the corresponding results for Scenarios 3–4 are presented in Web Figure C1 and Web Table C1 (last two sections). In addition, box plots of ATE estimates for OAL and GOAL are presented in Web Figures C2–C5 by scenario and p/n ratios, separately for each ρ value. All the results for the low-dimensional settings ($n = 200, 500, 1000$ and $p = 20$) when the true ATE is 0 are presented in Web Appendix C (Web Table C2 and Web Figures C6–C9). Selected complementary results for wAMD and variable selection in the high- and low-dimensional settings are also found in Web Appendix C. As similar results were obtained when the true ATE was 2, we omit their presentation for all scenarios. In the sequel, we refer to Naive GOAL as GOALn and GOAL with PIRLS as GOALi.

In the high-dimensional settings ($p/n = 100/200, 200/500$), all three estimators (OAL, GOALn, and GOALi) performed similarly when $\rho = 0$ (refer to Figure 1 (Scenarios 1–2), Web Figure C1 (Scenarios 3–4), and Web Table C1). When $\rho = 0.5$ or 0.75, GOAL was found systematically less variable than OAL, and either GOALi or both GOALi and GOALn exhibited less bias than OAL. GOALi's bias was small or relatively small for all ρ values. The difference between GOAL and OAL estimators was the largest when $\rho = 0.75$. Notably, the MSE of OAL was found at least twice the MSE of GOAL under this correlation value.

In the low-dimensional settings ($n = 200, 500, 1000$ with $p = 20$), we found that GOALi performed generally much like GOALn for both bias and variance in all scenarios (refer to Web Table C2 and Web Figures C6–C9). OAL performed similarly to GOAL estimators when $\rho = 0$. However, when $\rho > 0$, GOAL yielded ATE estimators that were both less biased and less variable than OAL, with the difference between GOAL and OAL becoming more marked as ρ increased. GOAL estimators exhibited small biases for all ρ values, whereas OAL's bias increased with correlation.

In Web Figures C10–C13, we present the wAMD between exposure groups for OAL and GOAL estimators over 1000 simulations for combinations ($n = 200$, $p = 100$) and ($n = 200$, $p = 20$) with $\rho = 0, 0.75$. In the high-dimensional setting with $\rho = 0$, OAL and GOALi performed similarly with respect to wAMD values, whereas GOALn yielded almost systematically larger wAMD values in comparison with GOALi (Web Figure C10). In the low-dimensional setting with $\rho = 0$, GOAL produced overall smaller or similar wAMD values than OAL across scenarios (Web Figure C12). When $\rho = 0.75$, both settings yielded wAMD values for OAL that were remarkably larger than for GOAL (Web Figure C11 and Web Figure C13). This greater imbalance between exposure groups for OAL is in line with the larger bias observed for OAL as compared to GOAL in the high- and low-dimensional settings with $\rho = 0.75$.

Presentation of variable selection results (see Web Figures C14–C15) is done in the Web Appendix.

5 | DISCUSSION

We presented a note on the OAL, a penalized variable selection approach for causal inference proposed by Shortreed and Ertefaie (2017). Although OAL was observed to have good performance in the low-dimensional settings with uncorrelated covariates examined by Shortreed and Ertefaie (2017), this approach was found to yield a biased IPTW estimator of the ATE in the high- and low-dimensional settings with correlated covariates. Our proposed approach GOAL was designed to improve on OAL by combining ideas from the adaptive lasso to achieve the oracle property and elastic net to address the collinearity problem. Our results showed that GOAL performed similarly or better than OAL in terms of balance between exposure groups and estimation accuracy in the same simulation scenarios studied by Shortreed and Ertefaie (2017). In particular, both versions of GOAL yielded IPTW estimators that were markedly less biased and variable than OAL in high and low dimensions with strongly correlated covariates. We found that Naive GOAL (GOALn) performed similarly to GOAL with PIRLS (GOALi) in the high-dimensional settings with uncorrelated covariates, but the latter was found less biased and slightly less variable than the former, in general, with correlated covariates. In the low-dimensional settings, GOALn and GOALi performed equivalently for every level of correlation between covariates investigated. Noting that the data augmentation step used for Naive GOALn is not readily justified for logistic models, GOALn's key advantage is that it can be estimated directly with the R code provided by Shortreed and Ertefaie (2017) after the data are modified. Although GOALi is also simple to implement, it is not as convenient as naive GOAL regarding the simplicity of the algorithm.

In this note, we performed all simulations with modest and large number of covariates (p) when $p < n$. A potential extension of this work would be to generalize GOAL for the case $p \geqslant n$. Indeed, owing to the properties of adaptive elastic net, GOAL appears well equipped for tackling this case. However, GOAL (as well as OAL) is based on the ordinary least squares (ols) weights $\hat{w}$ that require a full rank model for their estimation. GOAL's extension to ultra-high dimension will thus necessitate some modification to how the adaptive weights are defined. This will be investigated in a future study.

ACKNOWLEDGMENTS

Baldé gratefully acknowledges financial support from the Institut des sciences mathématiques (ISM). This work was funded by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Fonds de recherche du Québec–Santé (FRQ-S). G.L. is an FRQ-S Research Scholar.

DATA AVAILABILITY STATEMENT

The R code used for performing the simulation study is available in the Supporting Information section.

ORCID

Geneviève Lefebvre https://orcid.org/0000-0002-7155-8218

REFERENCES

Algamal, Z.Y. & Lee, M.H. (2015a) Applying penalized binary logistic regression with correlation based elastic net for variables selection. *Journal of Modern Applied Statistical Methods*, 14(1), 168–179.

Ghosh, S. (2011) On the grouped selection and model complexity of the adaptive elastic net. *Statistics and Computing*, 21, 451–462.

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.

Shortreed, S.M. & Ertefaie, A. (2017) Outcome-adaptive lasso: variable selection for causal inference. *Biometrics*, 73(4), 1111–1122. https://doi.org/10.1111/biom.12679.

Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 58, 267–288.

Zou, H. (2006) The adaptive lasso and its oracle properties. *Journal of the American Statistical Association: Series B*, 101, 1418–1429.

Zou, H. & Hastie, T. (2005) Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B*, 67, 301–320.

Zou, H. & Zhang, H.H. (2009). On the adaptive elastic-net with a diverging number of parameters. *The Annals of Statistics*, 37, 1733–1751.

SUPPORTING INFORMATION

Web appendices referenced in Section 2.1 and in Section 4, as well as examples of R code for implementing the generalized outcome adaptive lasso (GOAL), are available with this paper at the Biometrics website on Wiley Online Library.

How to cite this article: Baldé, I., Yang, Y.A., & Lefebvre, G. (2023) Reader reaction to "Outcome-adaptive lasso: Variable selection for causal inference" by Shortreed and Ertefaie (2017). *Biometrics*, 79, 514–520. https://doi.org/10.1111/biom.13683