Submitted to the Annals of Applied Statistics

SUPPLEMENT TO "FUSED COMPARATIVE INTERVENTION SCORING FOR HETEROGENEITY OF LONGITUDINAL INTERVENTION EFFECTS"

BY JARED D. HULING*, MENGGANG YU† AND MAUREEN SMITH†

*The Ohio State University** *and University of Wisconsin-Madison*†

1. Derivations for matching estimator. First, for a given value for the propensity a and $\pi_i = \pi(\boldsymbol{x}_i)$, denote

$$\hat{S}_{(1)} \equiv \frac{\sum_{t=1}^{K} \frac{1}{N_t} \sum_{i=1}^{N_t} w(A_i) I(A_i = 1) M(Y_{it}, \boldsymbol{\gamma}_t' \boldsymbol{x}_i \cdot A_i/2) I(\pi_i = a)}{\sum_{t=1}^{K} \frac{1}{N_t} \sum_{i=1}^{N_t} w(A_i) I(A_i = 1) I(\pi_i = a)} \tag{1}$$

and

$$\hat{S}_{(-1)} \equiv \frac{\sum_{t=1}^{K} \frac{1}{N_t} \sum_{i=1}^{N_t} w(A_i) I(A_i = -1) M(Y_{it}, \boldsymbol{\gamma}_t' \boldsymbol{x}_i \cdot A_i/2) I(\pi_i = a)}{\sum_{t=1}^{K} \frac{1}{N_t} \sum_{i=1}^{N_t} w(A_i) I(A_i = -1) I(\pi_i = a)}, \tag{2}$$

where the weight function $w(A)$ is such that both $E[w(A)I(A = 1)|\boldsymbol{X}]$ and $E[w(A)I(A = -1)|\boldsymbol{X}]$ are constants, denoted as c_1 and c_{-1}, respectively, conditional on $\pi(\boldsymbol{X}) = a$. For example, when $w(A) = I(A = 1)(C + 1) + I(A = -1)(C + 1)/C$, where C is the number of controls in the matched cluster corresponding to observations with $\pi(\boldsymbol{X}) = a$, $c_1 = a(C + 1)$ and $c_{-1} = (1 - a)(C + 1)/C$.

Intuitively, quantities $\hat{S}_{(1)}$ and $\hat{S}_{(-1)}$ should converge in probability to

$$S_{(1)} \equiv \frac{\sum_{t=1}^{K} E\left[w(A) I(A = 1) \cdot M(Y_t^{(1)}, \boldsymbol{\gamma}_t' \boldsymbol{X}/2) I(\pi(\boldsymbol{X}) = a)\right]}{\sum_{t=1}^{K} E\left[w(A) I(A = 1) I(\pi(\boldsymbol{X}) = a)\right]} \tag{3}$$

and

$$S_{(-1)} \equiv \frac{\sum_{t=1}^{K} E\left[w(A) I(A = -1) M(Y_t^{(-1)}, -\boldsymbol{\gamma}_t' \boldsymbol{X}/2) I(\pi(\boldsymbol{X}) = a)\right]}{\sum_{t=1}^{K} E\left[w(A) I(A = -1) I(\pi(\boldsymbol{X}) = a)\right]}, \tag{4}$$

respectively.

Keywords and phrases: Fused lasso, Precision medicine, Comparative effectiveness research, Electronic health records, Interaction modeling

By conditioning on $\boldsymbol{X}$ we have

$$\begin{aligned} S_{(1)} &= \frac{\sum_{t=1}^{K} E\left\{I(\pi(\boldsymbol{X})=a)E\left[w(A)I(A=1)M(Y_t^{(1)},\boldsymbol{\gamma}_t'\boldsymbol{X}/2)|\boldsymbol{X}\right]\right\}}{\sum_{t=1}^{K} E\left\{I(\pi(\boldsymbol{X})=a)E\left[w(A)I(A=1)|\boldsymbol{X}\right]\right\}} \\ &= \frac{\sum_{t=1}^{K} c_1 E\left\{I(\pi(\boldsymbol{X})=a)E\left[M(Y_t^{(1)},\boldsymbol{\gamma}_t'\boldsymbol{X}/2)|\boldsymbol{X}\right]\right\}}{\sum_{t=1}^{K} c_1 E\left[I(\pi(\boldsymbol{X})=a)\right]} \\ &= E\left[\frac{1}{K}\sum_{t=1}^{K} M(Y_t^{(1)},\boldsymbol{\gamma}_t'\boldsymbol{X}/2)|\pi(\boldsymbol{X})=a\right]. \end{aligned}$$

Similarly, $S_{(-1)}$ can be shown to be equal to $E\left[\frac{1}{K}\sum_{t=1}^{K} M(Y_t^{(-1)},-\boldsymbol{\gamma}_t'\boldsymbol{X}/2)|\pi(\boldsymbol{X})=a\right]$ and hence the difference $\hat{S}_{(1)}-\hat{S}_{(-1)}$ from the matched statum with $\pi(\boldsymbol{X})=a$ should converge to

$$E\left[\frac{1}{K}\sum_{t=1}^{K} M(Y_t^{(1)},\boldsymbol{\gamma}_t'\boldsymbol{X}/2)-M(Y_t^{(-1)},-\boldsymbol{\gamma}_t'\boldsymbol{X}/2)|\pi(\boldsymbol{X})=a\right].$$

Thus, by aggregating over all strata, we have

$$\begin{aligned} &\int \left\{S_{(1)}-S_{(-1)}\right\} \mathrm{d}F_{\pi(\boldsymbol{X})|A=1} \\ =& \int E\left[\frac{1}{K}\sum_{t=1}^{K} M(Y_t^{(1)},\boldsymbol{\gamma}_t'\boldsymbol{X}/2)-M(Y_t^{(-1)},-\boldsymbol{\gamma}_t'\boldsymbol{X}/2)|\pi(\boldsymbol{X})=a\right] f_\pi(a|A=1)\mathrm{d}a \\ =& \int E\left[\frac{1}{K}\sum_{t=1}^{K} M(Y_t^{(1)},\boldsymbol{\gamma}_t'\boldsymbol{X}/2)-M(Y_t^{(-1)},-\boldsymbol{\gamma}_t'\boldsymbol{X}/2)|\pi(\boldsymbol{X})=a,A=1\right] f_\pi(a|A=1)\mathrm{d}a \\ =& E\left[\frac{1}{K}\sum_{t=1}^{K} M(Y_t^{(1)},\boldsymbol{\gamma}_t'\boldsymbol{X}/2)-M(Y_t^{(-1)},-\boldsymbol{\gamma}_t'\boldsymbol{X}/2)|A=1\right], \end{aligned}$$

where $f_\pi(a|A=1)$ is the conditional density of $\pi(\boldsymbol{X})$ given $A=1$. Note that for any $\boldsymbol{x}$ the first order conditions of the weighting estimator results in

$$E\left[\frac{1}{K}\sum_{t=1}^{K} M(Y_t^{(1)},\boldsymbol{\gamma}_t'\boldsymbol{X}/2)-M(Y_t^{(-1)},-\boldsymbol{\gamma}_t'\boldsymbol{X}/2)|\boldsymbol{X}\right]=0.$$

For any $\boldsymbol{x}$ the first order condition of the matching estimator, however, yields

$$E\left[\frac{1}{K}\sum_{t=1}^{K} M(Y_t^{(1)},\boldsymbol{\gamma}_t'\boldsymbol{X}/2)-M(Y_t^{(-1)},-\boldsymbol{\gamma}_t'\boldsymbol{X}/2)|\boldsymbol{X},A=1\right]=0.$$

2. Additional simulation results. In this section we present simulation results under the same exact settings as in the main text of this paper, but with the main effects $\phi_t(\boldsymbol{X}) = \frac{1}{4}(\boldsymbol{\beta}_{0t}'\boldsymbol{X})^2$ to inspect the validity of our approach with non-linear main effects. The signal-to-noise ratio of the interaction effects, defined as $\sqrt{Var(\boldsymbol{\gamma}_0(t)'\boldsymbol{X} \cdot A/2)}/\sqrt{Var(\epsilon_t + \phi_t(\boldsymbol{X}))}$, is roughly the same as the model with a linear form: $\phi_t(\boldsymbol{X}) = \boldsymbol{\beta}_{0t}'\boldsymbol{X}$. The results over all 500 simulation replications are presented in Figures 1, 2, and 3 and are similar to the scenarios where $\phi_t(\boldsymbol{X})$ is linear, which is perhaps unsurprising given the simulation results in Chen et al. (2017). In fact, the improvement of our proposed approach over the naive approach is slightly more apparent with nonlinear $\phi_t(\boldsymbol{X})$ than with linear main effects. Although under non-linear main effects scenarios the interaction effect signal-to-noise ratios are similar to those in scenarios with linear main effects, the range of response is different and larger for the non-linear main effects scenarios. In particular, there are more extreme values in the responses. This results in the results having more variability than for linear main effects scenarios.

References.

Chen, S., Tian, L., Cai, T., and Yu, M. (2017), "A general statistical framework for subgroup identification and comparative treatment scoring," *Biometrics*, pp. n/a–n/a.

Jared D. Huling
Department of Statistics
The Ohio State University
Columbus, OH 43210
E-mail: huling.7@osu.edu

Menggang Yu
Department of Biostatistics & Medical Informatics
University of Wisconsin-Madison
Madison, WI 53706
E-mail: meyu@biostat.wisc.edu

Maureen Smith
Department of Population Health Sciences,
Department of Family Medicine & Community Health, and
Health Innovation Program
University of Wisconsin-Madison
Madison, WI 53706
E-mail: maureensmith@wisc.edu

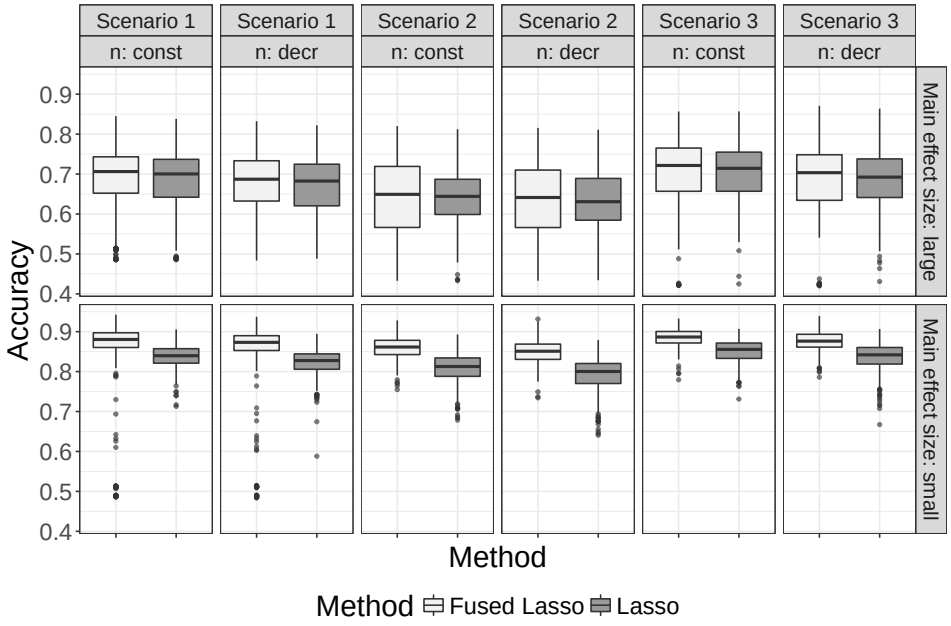

FIG 1. *Accuracy results from the simulation for each scenario and method. The accuracy of the estimated subgroups is evaluated on an independent test set of size 100000 for each simulation and each scenario. The columns of plots labeled "n: const" have sample sizes which are fixed for all time-points, i.e. have no dropout. The columns of plots labeled "n: decr" have sample sizes which decrease over time, i.e. a number of samples drop out after each time point.*

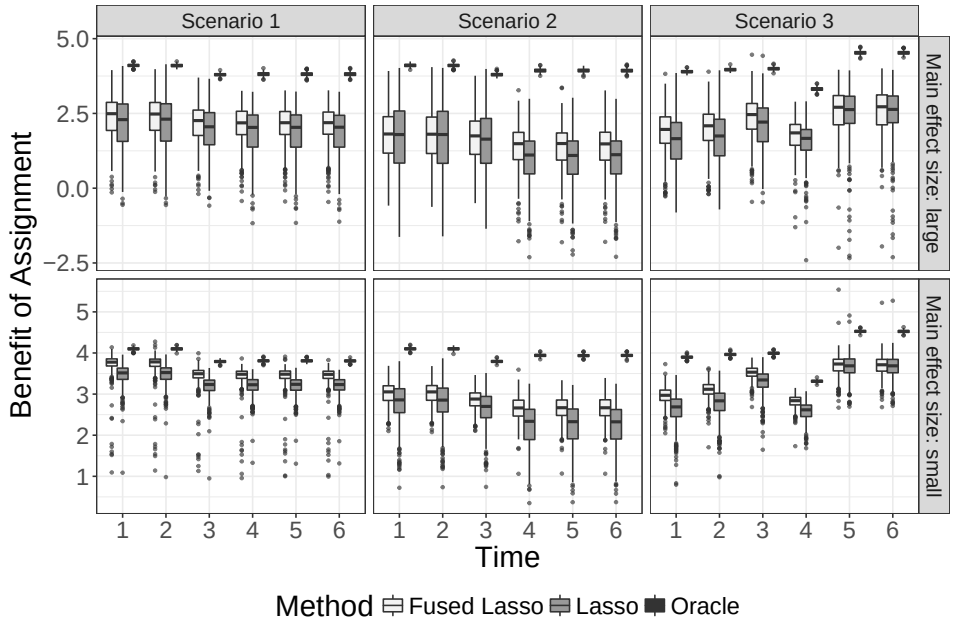

FIG 2. *Benefits of treatment assignment results from the simulation for each time point and method with sample sizes that decrease by 10 after each time point. The benefit of assignments is evaluated on an independent test set of size 100000 for each simulation and each time point.*

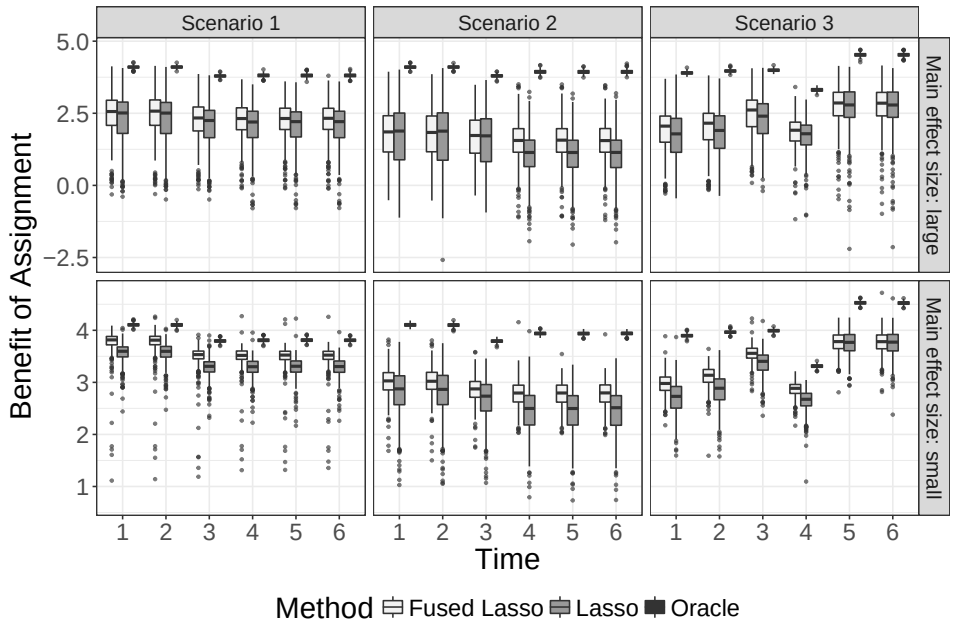

FIG 3. *Benefits of treatment assignment results from the simulation for each time point and method with constant sample sizes in time. The benefit of assignments is evaluated on an independent test set of size 100000 for each simulation and each time point.*