Web-based Supplementary Materials for "Bayesian Approach for Addressing Differential Covariate Measurement Error in Propensity Score Methods"

Hwanhee Hong, Kara E. Rudolph, and Elizabeth A. Stuart

1 Full conditional posterior distributions for joint models

The likelihood of the observed data $\mathbf{O} = (Y, A, W, Z)$ is written as

$$L(Y, A, W \mid \boldsymbol{\alpha}, \gamma, \boldsymbol{\psi}, \boldsymbol{\xi}, X) \propto L(A \mid X, Z, \boldsymbol{\alpha}) L(W \mid A, X, \gamma, \delta, \sigma_{w \mid x,a}) L(Y \mid A, X, Z, \boldsymbol{\psi}, \sigma_{y \mid x,a,z}), \quad \text{(Web.1)}$$

where $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \alpha_2)$, $\boldsymbol{\psi} = (\psi_0, \psi_1, \psi_2, \psi_3)$, and $\boldsymbol{\xi} = (\sigma_{w \mid x,a}, \sigma_{y \mid x,a,z})$.

Suppose we assume $\pi(\boldsymbol{\theta}) \propto 1$, where $\boldsymbol{\theta}$ is the set of all model parameters except X and $\pi(\cdot)$ is the prior density. The relationship of X and Z in Equation (3) of the main manuscript can be used as a prior distribution for X, such as $\pi(X \mid Z) \sim N(\beta_0 + \beta_1 Z, \sigma^2_{x|z})$. Please refer to Section 3 of the main manuscript for the prior specification of other parameters. For illustration purposes, we estimate propensity scores using a probit regression rather than logistic because it allows us to derive posterior distributions in closed form by utilizing $L(A \mid X, Z, \boldsymbol{\alpha}) \propto \prod_{i=1}^{n} exp\{-0.5(A_i^* - \alpha_0 - \alpha_1 X_i - \alpha_2 Z_i)^2\}$, where A^* is sampled from normally distributed latent continuous data with a unit variance such that $A^* > 0$ when $A = 1$, and $A^* < 0$ otherwise (Albert and Chib, 1993). The full posterior distribution under all three measurement error scenarios can be written as

$$q(\boldsymbol{\theta}, X \mid \mathbf{O}) \propto L(Y, A, W \mid \boldsymbol{\theta}, X) \pi(X \mid Z, \boldsymbol{\theta}). \quad \text{(Web.2)}$$

We can sample the model parameters and X from (Web.2). Selected conditional posterior distributions under the mixed measurement error scenario can be specified as follows:

$$q(\gamma \mid \boldsymbol{\theta}, \mathbf{O}) \sim N\left(\frac{1}{\sum A_i}\sum_i A_i(W_i - X_i),\ \frac{\sigma^2_{w\mid x,a=1}}{\sum A_i}\right)$$

$$q(\sigma^2_{w\mid x,a} \mid \boldsymbol{\theta}, \mathbf{O}) \sim IG\left(\frac{\sum(A_i = a)}{2} - 1,\ 0.5\sum_{A_i=a}(W_i - X_i - \gamma A_i)^2\right)$$

$$q(X_i \mid \boldsymbol{\theta}, \mathbf{O}) \sim N\left(\frac{h(A_i^*, A_i, Z_i, W_i, Y_i)}{D_i},\ \frac{1}{D_i}\right),$$

where

$$D_i = \alpha_1^2 + \sigma^{-2}_{w\mid x,a_i} + \sigma_y^{-2}\psi_2^2 + \sigma_x^{-2},\ \text{and} \tag{Web.3}$$

$$\begin{aligned} h(A_i^*, A_i, Z_i, W_i, Y_i) = {} & (\sigma_x^{-2}\beta_0 - \alpha_1\alpha_0) + \alpha_1 A_i^* + (-\gamma\sigma^{-2}_{w\mid x,a_i} - \psi_1\psi_2\sigma_y^{-2})A_i \\ & + (\beta_1\sigma_x^{-2} - \alpha_1\alpha_2 - \psi_2\psi_3\sigma_y^{-2})Z_i + \sigma^{-2}_{w\mid x,a_i}W_i + \psi_2\sigma_y^{-2}Y_i. \end{aligned} \tag{Web.4}$$

Note that we drop the conditional notation on the subscripts of σ_x and σ_y for simplicity in (Web.4). In the conditional posterior distribution of X_i, $\boldsymbol{\psi}$ and σ_y^2 are involved. Specifically, if ψ_2 is large in (Web.3), the posterior distribution variability of X_i would decrease, meaning that the posterior sample of X_i is more precise. For each draw of $\boldsymbol{\alpha}$ and X_i, we calculate posterior samples of propensity scores using $logit^{-1}(\alpha_0 + \alpha_1 X_i + \alpha_2 Z_i)$. Similarly, we obtain an estimate of the ATE using Equation (1) in the main manuscript at each iteration of the MCMC algorithm.

2 Additional results of simulation studies and data analysis

Web Table 1 shows bias, MSE, coverage probability, and the average width of 95% credible intervals of ATE in the simulation study when there is more imbalance of X and Z between treated and control groups (i.e., using $logit(Pr(A = 1 \mid X, Z)) = -2 + X + Z$). First, the overall trend is similar to that in Table 3 of the main manuscript for systematic and heteroscedastic measurement error scenarios. However, joint models tend to produce smaller *relative* biases (i.e., the ratio of biases between our models and the True model) than when the imbalance of X and Z between

two groups is small. Second, in the mixed measurement error scenario, the bias of the ATE under Joint_H is smaller than that under Joint_S, while the opposite is true in Table 3. That is, assigning a point mass prior to δ is more important than to γ when covariates differ a lot between treated and control groups. All these findings could be because propensity scores more effectively control for imbalances in X and Z between two groups, so the effect of measurement error on bias in treatment effect estimation is diminished. Third, it is obvious that two-step models perform poorly when X is a strong predictor of Y; TS_SH produces much larger bias and MSE with lower coverage probability than Joint_PM.

We also explore the impact of model misspecification under the mixed measurement error scenario; the results for outcome model misspecification and propensity score model misspecification are in Web Figures 1 and 2, respectively. The overall trend is similar to panels (e) and (f) of Figure 2 in the main manuscript. When outcome model is misspecified the bias and MSE are much larger in the Joint_inf and Joint_H models with a large X-Y association. This is expected because misspecification of the outcome model can also impair the imputation of X under the joint model. The two-step models are not influenced by the outcome model misspecification because the X-Y association is not incorporated when imputing X. On the other hand, propensity score model misspecification does not impact much on estimating the ATE as bias and MSE are just slightly larger than those estimated when propensity score model is correctly specified. These results agree with findings from Drake (1993).

Web Figure 3 compares prior and posterior distributions of parameters related to differential measurement error. These figures show whether those parameters really get much updating from the data or not. We generated a single dataset under mixed measurement error and fitted Joint_inf (using $\gamma \sim N(0,3)$ and $\sigma_{w\,|\,x,a} \sim Uniform(0.01,3)$), Joint_S ($\sigma_{w\,|\,x,a} \sim Uniform(0.01,3)$ with a point mass prior on γ), and Joint_H ($\gamma \sim N(0,3)$ with a point mass prior on $\sigma_{w\,|\,x,a}$). All parameters are updated pretty well in Joint_inf as the posteriors are not the same as their priors. In addition, we observe that using point mass prior on either parameter helps another parameter converge faster as dotted-line density plots have slightly less variable than dashed-line plots.

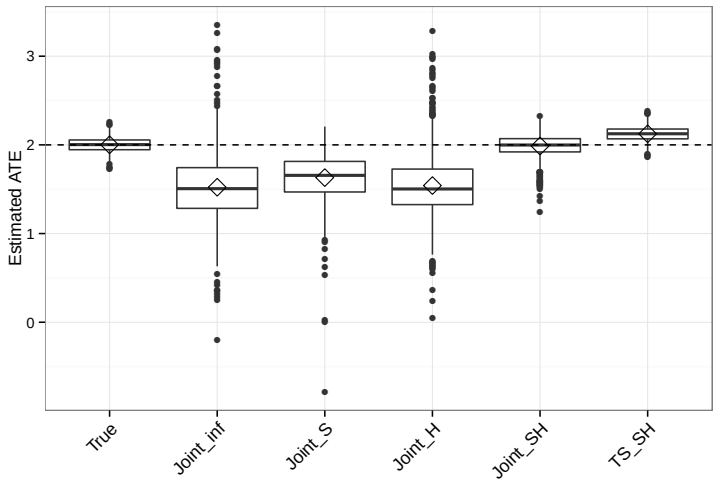
Web Figure 4 shows the ATE estimates in the illustrative example for each method using W_{sys} in Panel (a) and W_{het} in Panel (b). The pattern of results are similar as seen for W_{mix}, which are shown in Figure 5 in the main manuscript. Using mother-reported age, X, as the truth, the mean posterior ATE is 0.014, 95% CI: 0.005, 0.027, suggesting that living in a disadvantaged neighborhood is associated with a slightly higher probability of having a prevalent drug use or dependence disorder.

Using W_{sys} that represents systematic measurement error in the naïve approach results in a point estimate and 95% CI that are close to the true estimate: ATE=0.015, 95% CI: 0.004, 0.029. However, in this particular case, methods to correct for the measurement error do less well. Point estimates are similar to the truth for Joint_S1, Joint_S2, Joint_S3 (which is the same as Joint_S), and TS_S, but the confidence intervals are wider, especially for TS_S, and include 0 for all except Joint_S1. Performance is worse when a non-point mass prior is used (Joint_inf) as compared with a point-mass prior.

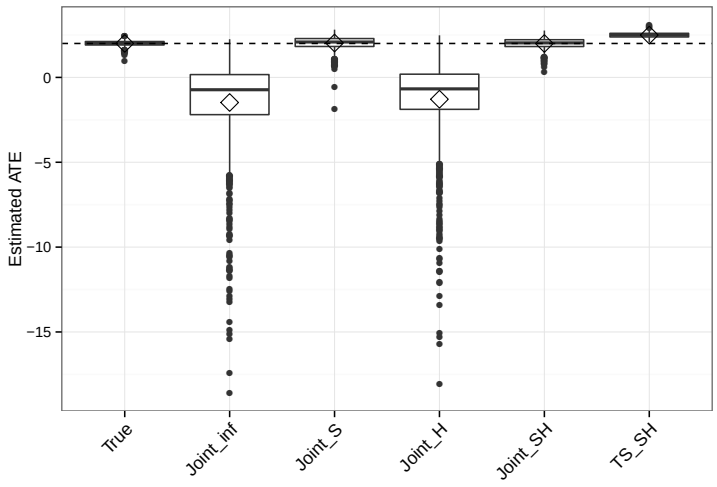
Using W_{het} that represents heteroscedastic measurement error in the naïve approach also results in a point estimate and 95% CI that are similar to the true estimate: ATE=0.018, 95% CI: 0.007, 0.031. In contrast to the systematic measurement error case, some methods to correct for measurement error offer slight improvements over the naïve approach. In particular, Joint_H3 (which is the same as Joint_H) and Joint_H4 move the point estimate closer to the true value and give the same inference. Although the confidence interval widens when a non-point mass prior is used (Joint_inf), the point estimate is closer to the true value.

Web Table 1: Bias, MSE, coverage probabilities, and average width of 95% credible intervals (95% CI width) of ATE in the simulation study when $logit(Pr(A = 1 \mid X, Z)) = -2 + X + Z$.

X-Y association	Model	Bias	MSE	Coverage probability	95% CI width
Systematic measurement error					
Low	True	0.019	0.027	0.771	0.372
	Naive	-0.105	0.103	0.643	0.475
	Joint_inf	0.031	0.047	0.985	1.319
	Joint_PM	-0.153	0.065	0.928	1.057
	TS_PM	0.146	0.039	0.935	0.850
High	True	0.071	0.268	0.896	1.491
	Naive	-0.421	1.210	0.772	1.891
	Joint_inf	0.138	0.469	0.968	2.778
	Joint_PM	0.056	0.274	0.944	1.859
	TS_PM	0.585	0.515	0.894	2.696
Heteroscedastic measurement error					
Low	True	0.019	0.027	0.771	0.372
	Naive	0.187	0.050	0.368	0.303
	Joint_inf	-0.219	0.087	0.911	1.176
	Joint_PM	-0.059	0.034	0.947	0.799
	TS_PM	0.175	0.048	0.886	0.851
High	True	0.071	0.268	0.896	1.491
	Naive	0.751	0.683	0.325	1.204
	Joint_inf	0.061	0.338	0.948	1.919
	Joint_PM	-0.002	0.314	0.951	1.813
	TS_PM	0.705	0.640	0.805	2.671
Mixed measurement error					
Low	True	0.019	0.027	0.771	0.372
	Naive	-0.029	0.070	0.665	0.428
	Joint_inf	-0.037	0.057	0.989	1.486
	Joint_S	-0.219	0.084	0.931	1.184
	Joint_H	-0.030	0.050	0.988	1.410
	Joint_SH	-0.055	0.034	0.944	0.790
	TS_SH	0.176	0.048	0.887	0.839
High	True	0.071	0.268	0.896	1.491
	Naive	-0.115	0.748	0.806	1.704
	Joint_inf	0.019	0.464	0.974	3.016
	Joint_S	0.064	0.286	0.942	1.920
	Joint_H	0.038	0.384	0.982	3.032
	Joint_SH	0.004	0.273	0.945	1.810
	TS_SH	0.702	0.648	0.800	2.665

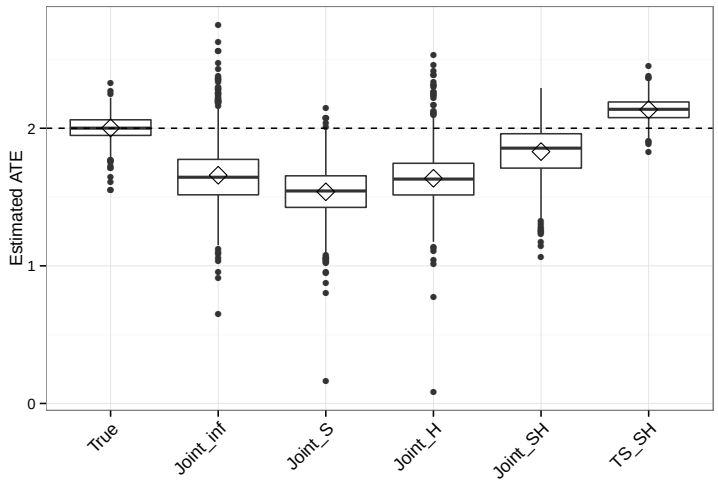


(a) Low X-Y association

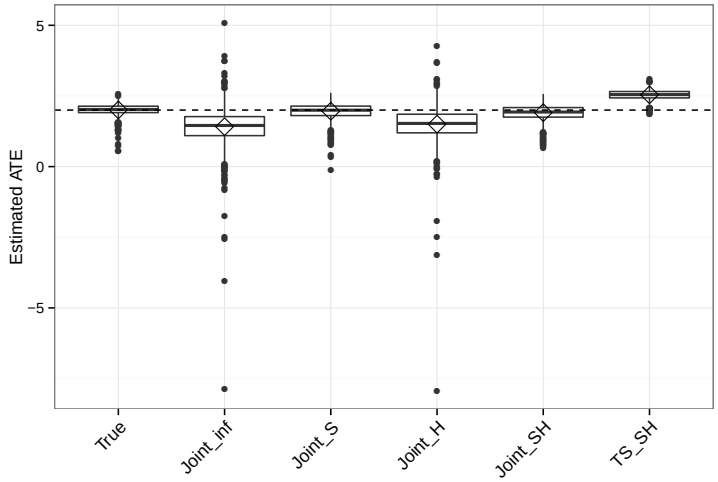


(b) High X-Y association

Web Figure 1: ATE estimates from the simulation study when the outcome model is misspecified under mixed measurement error.

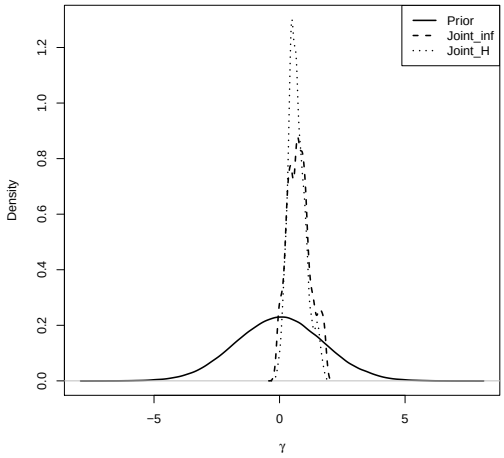


(a) Low X-Y association

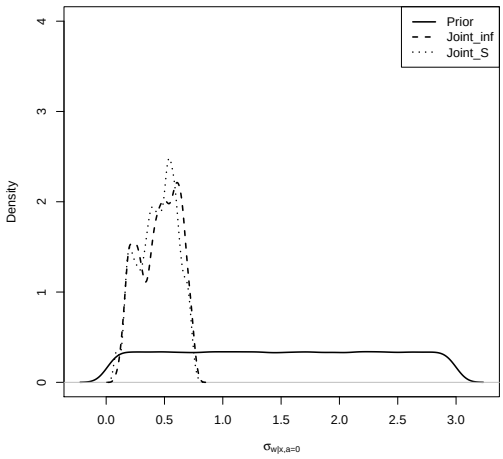


(b) High X-Y association

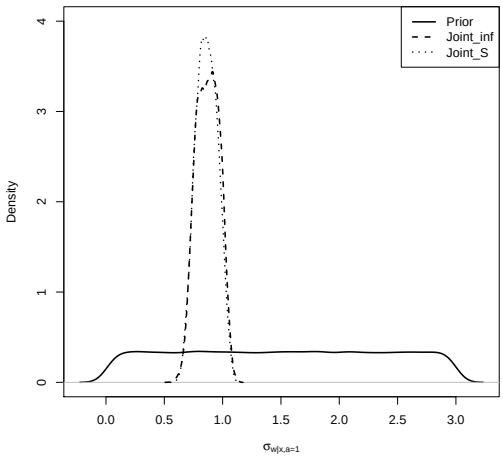
Web Figure 2: ATE estimates from the simulation study when the propensity score model is misspecified under mixed measurement error.



(a) γ

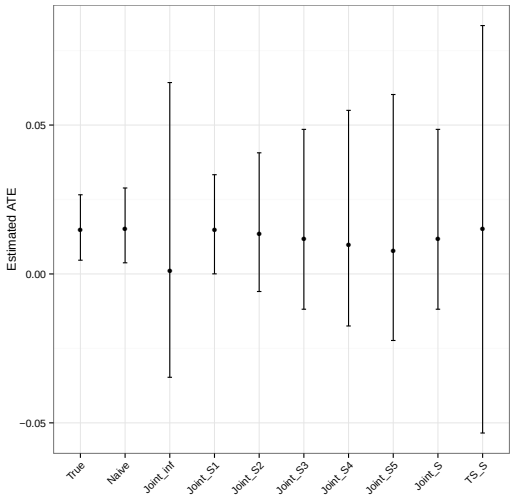


(b) $\sigma_{w \mid x,a=0}$

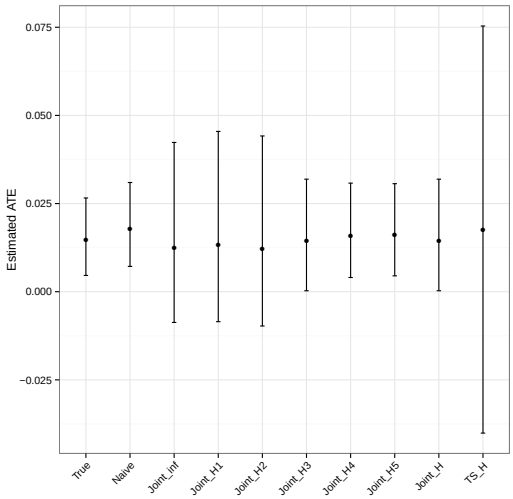


(c) $\sigma_{w \mid x,a=1}$

Web Figure 3: Comparison of prior and posterior distributions of γ, $\sigma_{w \mid x,a=0}$, and $\sigma_{w \mid x,a=1}$ in Joint_inf, Joint_S, and Joint_H under mixed measurement error



(a) Systematic measurement error



(b) Heteroscedastic measurement error

Web Figure 4: Estimated ATE and 95% CIs by method in the illustrative example using (a) W_{sys} and (b) W_{het}. The ATE is the average effect of living in a disadvantaged neighborhood on probability of past-year drug abuse or dependence. N=1,000.

3 JAGS code

```
## JAGS code to fit the Joint model
## when there is mixed differential measurement error in a covariate
## To fit Two-Step model, comment out "outcome models"
## z: correctly measured continuous covariate, observed
## x: correctly measured continuous covariate, not observed
## w: mismeasurement of x, observed
## t: treatment assignment, binary
## y: outcome, continous

model{

  for(i in 1:N) {
    t[i] ~ dbern(p[i])
    logit(p[i]) <- beta[1] + beta[2]*x[i] + beta[3]*z[i]

    w[i] ~ dnorm(mu_w[i], prec_w[t[i]+1])
    mu_w[i] <- x[i] + gamma*t[i]

    x[i] ~ dnorm(mu_x[i], prec_x)
    mu_x[i] <- alpha[1] + alpha[2]*z[i]

    y[i] ~ dnorm(mu_y[i], prec_y)
    mu_y[i] <- psi[1] + psi[2]*t[i] + psi[3]*x[i] + psi[4]*z[i]
  }

  for (j in 1:3) { beta[j] ~ dnorm(0, 1/3) }
  # Assume weakly-informative prior on gamma
  gamma ~ dnorm(0, 1/3)
  # When assuming point-mass prior on gamma
  #gamma <- 1 # assign a proper number here
  for (j in 1:2) { alpha[j] ~ dnorm(0, 1/3) }
  for (j in 1:4) { psi[j] ~ dnorm(0, 1/3) }

  for (j in 1:2) {
    prec_w[j] <- 1/pow(sig_w[j],2)
    sig_w[j] ~ dunif(0.01, 3)
  }
  prec_x <- 1/pow(sig_x, 2)
  sig_x ~ dunif(0.01, 3)
  prec_y <- 1/pow(sig_y, 2)
  sig_y ~ dunif(0.01, 3)

  delta <- (sig_w[2]/sig_w[1])-1
```

```
  for (i in 1:N) {
    ate_weight[i] <- ifelse(t[i]==1, 1/p[i], 1/(1-p[i]))
    ate_num0[i] <- y[i]*(1-t[i])*ate_weight[i]
    ate_den0[i] <- (1-t[i])*ate_weight[i]
    ate_num1[i] <- y[i]*t[i]*ate_weight[i]
    ate_den1[i] <- t[i]*ate_weight[i]
  }

  ATE <- (sum(ate_num1[])/sum(ate_den1[])) - (sum(ate_num0[])/sum(ate_den0[]))
}
```

References

Albert, J.H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American statistical Association* **88**, 669–679.

Drake, C. (1993). Effects of misspecification of the propensity score on estimators of treatment effect. *Biometrics* **49**, 1231–1236.