Web-based Supplementary Materials for "Accounting for Uncertainty in Confounder and Effect Modifier Selection when Estimating Average Causal Effects in Generalized Linear Models" by Chi Wang, Francesca Dominici, Giovanni Parmigiani, and Corwin Matthew Zigler

Web Appendix A: Prior and Posterior Distributions of $\boldsymbol{\beta}^{\alpha^Y}$ and $\boldsymbol{\theta}$

Assume that the priors of $\boldsymbol{\beta}^{\alpha^Y}$ and $\boldsymbol{\theta}$ are independent: $\pi(\boldsymbol{\beta}^{\alpha^Y}, \boldsymbol{\theta}) = \pi(\boldsymbol{\beta}^{\alpha^Y})\pi(\boldsymbol{\theta})$. We have

$$\begin{aligned}
p(\boldsymbol{\beta}^{\alpha^Y}, \boldsymbol{\theta}|D) \propto & p(\boldsymbol{Y}, \boldsymbol{X}, \boldsymbol{V}|\boldsymbol{\beta}^{\alpha^Y}, \boldsymbol{\theta})\pi(\boldsymbol{\beta}^{\alpha^Y}, \boldsymbol{\theta}) \\
= & p(\boldsymbol{Y}|\boldsymbol{\beta}^{\alpha^Y}, \boldsymbol{\theta}, \boldsymbol{X}, \boldsymbol{V})p(\boldsymbol{X}|\boldsymbol{\beta}^{\alpha^Y}, \boldsymbol{\theta}, \boldsymbol{V})p(\boldsymbol{V}|\boldsymbol{\beta}^{\alpha^Y}, \boldsymbol{\theta})\pi(\boldsymbol{\beta}^{\alpha^Y}, \boldsymbol{\theta}) \\
= & p(\boldsymbol{Y}|\boldsymbol{\beta}^{\alpha^Y}, \boldsymbol{X}, \boldsymbol{V})p(\boldsymbol{X}|\boldsymbol{V})p(\boldsymbol{V}|\boldsymbol{\theta})\pi(\boldsymbol{\beta}^{\alpha^Y}, \boldsymbol{\theta}) \\
\propto & [p(\boldsymbol{Y}|\boldsymbol{\beta}^{\alpha^Y}, \boldsymbol{X}, \boldsymbol{V})\pi(\boldsymbol{\beta}^{\alpha^Y})][p(\boldsymbol{V}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})] \\
\propto & p(\boldsymbol{\beta}^{\alpha^Y}|D)p(\boldsymbol{\theta}|D).
\end{aligned}$$

By taking integral over $\boldsymbol{\beta}^{\alpha^Y}$ and $\boldsymbol{\theta}$ on both sides of the formula, we can see that the proportional factor should be equal to 1. Therefore, the posteriors of $\boldsymbol{\beta}^{\alpha^Y}$ and $\boldsymbol{\theta}$ are independent

$$p(\boldsymbol{\beta}^{\alpha^Y}, \boldsymbol{\theta}|D) = p(\boldsymbol{\beta}^{\alpha^Y}|D)p(\boldsymbol{\theta}|D).$$

Therefore, these two parameters can be sampled separately in a Monte Carlo (MC) algorithm.

For the prior of $\boldsymbol{\theta}$, we assume $\pi(\boldsymbol{\theta}) \propto \Pi_{k=1}^{K}\theta_k^{-1}$ (Rubin, 1981). The posterior follows a Dirichlet distribution $D(n_1, \ldots, n_K)$

$$p(\boldsymbol{\theta}|\boldsymbol{V}) \propto \Pi_{k=1}^{K}\theta_k^{n_k-1}, \tag{A1}$$

where n_k counts the number of observed $\boldsymbol{V}$ values equal to the kth distinct value of $\boldsymbol{V}$. Note that if our goal is to estimate the average causal effect for the whole population, one should include all observations in the calculation of the posterior of $\boldsymbol{\theta}$. But if the goal is to estimate

the average causal effect for a certain subpopulation, only the observations belonging to the subpopulation should be used.

For $\boldsymbol{\beta}^{\alpha^Y}$, we assume by default a uniform prior, though in specific applications one can also specify informative priors. For example, one can consider a prior $\boldsymbol{\beta}^{\alpha^Y} \sim MVN(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a diagonal matrix with entries $[\{\log(100)/4\}^2, \{\log(15)/4\}^2 s_1^{-2}, \ldots, \{\log(15)/4\}^2 s_L^{-2}]$, with s_l^2 the sample variance of the lth predictor in outcome model $\boldsymbol{\alpha}^Y$ with L ($\leqslant 2M+1$) predictors. This prior centers coefficients at zero and concentrates prior mass away from extreme values that might represent, for example, a log odds ratio with magnitude greater than log(15) for a 1 standard deviation change in any predictor (in the case of logistic regression). Specifying an informative prior can be important when sample size is very small relative to the number of potential confounders, achieving an effect similar to regularizing the likelihood function to obtain more stable results (Makalic and Schmidt, 2011).

Web Appendix B: Sampling from the Posteriors

Sampling from the posterior of $(\boldsymbol{\alpha}^X, \boldsymbol{\alpha}^Y)$

We sample $(\boldsymbol{\alpha}^X, \boldsymbol{\alpha}^Y)$ from their joint posterior $p(\boldsymbol{\alpha}^X, \boldsymbol{\alpha}^Y|D)$ by iteratively sampling from $p(\boldsymbol{\alpha}^X|\boldsymbol{\alpha}^Y, D)$ and $p(\boldsymbol{\alpha}^Y|\boldsymbol{\alpha}^X, D)$. To sample from each of the two conditional posteriors, we use the MC^3 method (Madigan et al., 1995). Below, we illustrate the MC^3 method for drawing a sample from $p(\boldsymbol{\alpha}^Y|\boldsymbol{\alpha}^X, D)$. Sampling from $p(\boldsymbol{\alpha}^X|\boldsymbol{\alpha}^Y, D)$ is performed similarly. Define the neighborhood of $\boldsymbol{\alpha}^Y$ to be all the models in the outcome model space that has one more or one less main effect or interaction term than $\boldsymbol{\alpha}^Y$. Suppose the current Markov chain is in state $(\boldsymbol{\alpha}^X_{(0)}, \boldsymbol{\alpha}^Y_{(0)})$, we randomly draw a model $\boldsymbol{\alpha}^Y_{(1)}$ from the neighborhood of $\boldsymbol{\alpha}^Y_{(0)}$. The model is then accepted with probability

$$\min\left\{1, \frac{p(\boldsymbol{\alpha}^Y_{(1)}|\boldsymbol{\alpha}^X_{(0)}, D)}{p(\boldsymbol{\alpha}^Y_{(0)}|\boldsymbol{\alpha}^X_{(0)}, D)} = \frac{p(\boldsymbol{Y}|\boldsymbol{\alpha}^Y_{(1)}, \boldsymbol{X}, \boldsymbol{V})}{p(\boldsymbol{Y}|\boldsymbol{\alpha}^Y_{(0)}, \boldsymbol{X}, \boldsymbol{V})} \times \frac{p(\boldsymbol{\alpha}^Y_{(1)}|\boldsymbol{\alpha}^X_{(0)})}{p(\boldsymbol{\alpha}^Y_{(0)}|\boldsymbol{\alpha}^X_{(0)})}\right\}. \tag{A2}$$

Otherwise, the chain stays in state $\boldsymbol{\alpha}^Y_{(0)}$. In equation (A2), the prior odds $p(\boldsymbol{\alpha}^Y_{(1)}|\boldsymbol{\alpha}^X_{(0)})/p(\boldsymbol{\alpha}^Y_{(0)}|\boldsymbol{\alpha}^X_{(0)})$ can be calculated based on equations (7) and (8) and the Bayes factor can be approximated by a BIC approximation (Raftery, 1995; Lefebvre et al., 2014)

$$\frac{p(\boldsymbol{Y}|\boldsymbol{\alpha}^Y_{(1)}, \boldsymbol{X}, \boldsymbol{V})}{p(\boldsymbol{Y}|\boldsymbol{\alpha}^Y_{(0)}, \boldsymbol{X}, \boldsymbol{V})} \doteq \exp\left\{\frac{1}{2}(BIC_1 - BIC_0)\right\},$$

where BIC_0 and BIC_1 are the BICs from the two outcome models, respectively.

Sampling from the Posteriors of $\boldsymbol{\beta}^{\alpha^Y}$ and $\boldsymbol{\theta}$

We sample $(\boldsymbol{\beta}^{\alpha^Y}, \boldsymbol{\theta})$ based on their posteriors, then calculate $\psi(\boldsymbol{\beta}^{\alpha^Y}, \boldsymbol{\theta})$. Since the posteriors of $\boldsymbol{\beta}^{\alpha^Y}$ and $\boldsymbol{\theta}$ are independent, they can be sampled separately. We do not require any re-sampling on $\boldsymbol{V}$ when obtaining samples of $\boldsymbol{\beta}^{\alpha^Y}$, and thus do not impose extra computational burden by considering $\boldsymbol{V}$ as random. MC samples of $\boldsymbol{\theta}$ can be generated based on equation (A1). Except in the case where (2) is a normal linear model, the posterior of $\boldsymbol{\beta}^{\alpha^Y}$ does not follow a known distribution. Therefore, we generate samples of $\boldsymbol{\beta}^{\alpha^Y}$ based on the Metropolis-Hastings algorithm. The Metropolis proposal distribution is centered at the current value of $\boldsymbol{\beta}^{\alpha^Y}$ and has variance-covariance matrix equal to a matrix $\boldsymbol{V}$ times a tuning parameter. When a noninformative prior is assumed, $\boldsymbol{V}$ is the large sample variance-covariance matrix of the MLEs. When an informative prior is assumed, $\boldsymbol{V}$ is the variance-covariance matrix estimated from a ridge regression model with penalty specified according to the assumed informative prior. In our simulations and real data analysis, we used the R MCMCpack package (Martin et al., 2011) to generate an MCMC sample of $\boldsymbol{\beta}^{\alpha^Y}$ from a given model $\boldsymbol{\alpha}^Y$. The tuning parameter was set to have the acceptance rate in the Metropolis-Hastings algorithm between 20% and 50%.

Web Appendix C: Regression Coefficients for the Two Simulation Scenarios in the Main Text

[Web Table 1 about here.]

Web Appendix D: Issues of the Stratification by Propensity Score Method When Sample Size is Small

When sample size is small, the estimated propensity scores are sparse so that some propensity strata may need to be merged together. The following table listed the number of times out of the 500 simulation replicates that merging strata is needed for the stratification method for the two simulation scenarios in Section 3.

[Web Table 2 about here.]

Web Appendix E: Standardized Bias B Value

Based on true or estimated propensity scores, we calculated the standardized bias B value (Rubin, 2007) as a measure of the balance between exposure groups for the two simulation scenarios in the main text of the paper (Web Table 3). Examining B values from the estimated propensity scores illustrates the difficulty in balancing covariates when there are a lot of potential confounders and limited number of observations. Values of B calculated using the estimated propensity score are much larger than those calculated using the true propensity score. The results illustrate that the data are simulated so that covariates could be balanced, but the difficulty in estimating the propensity score leads to difficulty in balancing covariates.

[Web Table 3 about here.]

Web Appendix F: Propensity Score Stratum-Specific Average Causal Effect for the Real Data

[Web Table 4 about here.]

Web Appendix G: Methods Comparison Based on Random Sub-Sampling the Real Data

[Web Table 5 about here.]

Web Appendix H: Simulation Results When There are Both Continuous and Binary Potential Confounders and There are Correlations Among Potential Confounders

In this section, we use simulation studies to evaluate the performance of BAC when the potential confounders are a) correlated; and b) a mixture of binary and continuous variables. We simulated $W_1, \ldots, W_{50}$ from multivariate normal distribution $MVN(0, \Sigma)$, where $\Sigma = (\sigma_{ij})_{50\times 50}$ and $\sigma_{ij} = 0.7^{|i-j|}$. Let $V_m = W_m$, for $m = 1, 3, 5, \ldots, 49$, and $V_m = I\{W_m > 0\}$, for $m = 2, 4, 6, \ldots, 50$. The set of potential confounders includes V_1 to V_{50}. Binary exposure and binary outcome variables were generated from

$$\text{logit}\{E(X_i|\boldsymbol{V}_i)\} = \delta_1 V_{1i} + \delta_2 V_{2i} + \delta_3 V_{3i} + \delta_4 V_{4i} + \delta_5 V_{5i} + \delta_6 V_{6i} + \delta_7 V_{7i} + \delta_8 V_{8i},$$

$$\text{logit}\{E(Y_i|X_i, \boldsymbol{V}_i)\} = \beta_X X_i + \beta_1 V_{1i} + \beta_2 V_{2i} + \beta_3 V_{3i} + \beta_4 V_{4i} + \beta_5 V_{5i} + \beta_6 V_{6i} + \beta_7 V_{7i} + \beta_8 V_{8i},$$

where $\delta_0 = -1$, $\beta_0 = -1.25$, and $\delta_1 = \ldots = \delta_8 = \beta_X = \beta_1 = \ldots = \beta_8 = 0.5$. We generated 500 independent simulation replicates for each sample size $n = 100$, 150, 300, or 500. Results are summarized in Web Table 6. BAC performs well under this simulation scenario. The ACE estimate is virtually unbiased and the RMSE is smaller than that based on the full model. The posterior inclusion probabilities (PIPs) for the 50 potential confounders are shown in Web Figure 1. Although true confounders and noise variables are correlated, BAC is able to distinguish them and assign large PIPs only to true confounders. Interestingly, continuous confounders (V_1, V_3, V_5 and V_7) tend to have higher PIPs than binary confounders (V_2, V_4, V_6 and V_8). This may due to the difference in the range of variable values. In this simulation scenario, continuous confounders were generated from normal distributions so that their values have a much wider range than that for binary confounders. Therefore,

those continuous confounders tend to have a larger impact on the exposure and the outcome values.

[Web Table 6 about here.]

[Web Figure 1 about here.]

Web Appendix I: An Alternative BAC Prior on $(\boldsymbol{\alpha}^X, \boldsymbol{\alpha}^Y)$ When There are Interactions Between Confounders and the Exposure in Absence of Main Effects

In practice, a confounder may interact with the exposure to affect the outcome in absence of its main effect. Rather than using equation (8), one may consider the following prior on $(\boldsymbol{\alpha}^X, \boldsymbol{\alpha}^Y)$, which does not restrict the main effect term to be included whenever an interaction is included in the outcome model.

$$\begin{aligned}
&\frac{P(\max\{\alpha_m^Y, \alpha_{m+M}^Y\} = 1 | \alpha_m^X = 1)}{P(\max\{\alpha_m^Y, \alpha_{m+M}^Y\} = 0 | \alpha_m^X = 1)} = \omega, \quad \frac{P(\alpha_m^Y = 1 | \alpha_m^X = 0)}{P(\alpha_m^Y = 0 | \alpha_m^X = 0)} = \frac{P(\alpha_{m+M}^Y = 1 | \alpha_m^X = 0)}{P(\alpha_{m+M}^Y = 0 | \alpha_m^X = 0)} = 1, \\
&\frac{P(\alpha_m^X = 1 | \max\{\alpha_m^Y, \alpha_{m+M}^Y\} = 0)}{P(\alpha_m^X = 0 | \max\{\alpha_m^Y, \alpha_{m+M}^Y\} = 0)} = \frac{1}{\omega}, \quad \frac{P(\alpha_m^X = 1 | \max\{\alpha_m^Y, \alpha_{m+M}^Y\} = 1)}{P(\alpha_m^X = 0 | \max\{\alpha_m^Y, \alpha_{m+M}^Y\} = 1)} = 1,
\end{aligned} \tag{A3}$$

where $m = 1, \ldots, M$. This prior increases the chance for either the main effect or the interaction term for a potential confounder to be included in the outcome model if the variable is in the exposure model. It also increases the chance for a potential confounder to be excluded from the exposure model if neither its main effect nor its interaction term is in the outcome model.

To evaluate the performance of this prior as compared to the prior in equation (8), we considered a simulation scenario similar to the second simulation scenario in the main text of the paper, except that V_1 and V_3 have no main effect ($\beta_1 = \beta_3 = 0$).We first examine PIPs for the 50 potential confounders and the 10 potential interaction terms. The PIPs based on the prior in equation (A3) are shown in Web Figure 2. For the two variables (V_1 and V_3) that

have interactions with X but do not have main effects on Y, their main effect terms have very low PIPs while their interaction terms have very high PIPs. In contrast, for the variable (V_5) that has both main effect and interaction terms, both terms have very high PIPs. The results suggest that the prior proposed in equation (A3) is able to correctly identify interactions in absence of main effects. As a comparison, the PIPs based on the prior in equation (8) are shown in Web Figure 3. For all the three variables (V_1, V_3 and V_5), both the main effect and the interaction terms have very high PIPs. Therefore, using the prior in equation (8) leads to high PIPs of main effect terms for variables which only have interactions with the exposure but do not have main effects. Next, we compare the estimation of ACE. Based on both priors, the coverage probabilities of 95% credible intervals are close to the desired value. But the sample standard error (SEE) based on the prior in equation (8) is slightly larger than that based on the prior in equation (A3). This is due to the fact that the prior in equation (8) makes BAC much more likely to include the main effect terms of V_1 and V_3 in the model and therefore adds more noise to the inference.

[Web Table 7 about here.]

[Web Figure 2 about here.]

[Web Figure 3 about here.]

Web Appendix J: Simulation Results When the Outcome Model is Misspecified

BAC depends on the assumption that the outcome is linearly associated with confounders through a generalized linear model. In this section, we use simulations to evaluate the performance of BAC when the outcome model is misspecified. We considered a scenario which has exponential, cubic, reciprocal, and log terms of confounders in the outcome model.

More specifically, the data were generated from the following models

$$\text{logit}(P(X_i = 1|V_i)) = \delta_0 + \delta_1 V_{1i} + \delta_2 V_{2i} + \delta_3 V_{3i} + \delta_4 V_{4i}$$

$$\text{logit}(P(Y_i = 1|X_i, V_i)) = \beta_0 + \beta_X X_i + \beta_1 \exp\{V_{1i}\} + \beta_2 V_{2i}^3 + \frac{\beta_3}{V_{3i}} + \beta_4 \log V_{4i},$$

where $\delta_0 = -4$, $\beta_0 = -3$, $\delta_1 =, \ldots = \delta_4 = \beta_X = 0.5$, $\beta_1 = \beta_2 = 0.1$, $\beta_3 = \beta_4 = 1$, and $V_{1i}, \ldots, V_{4i}$ were independently generated from Uniform$(0.1, 4)$. The set of potential confounders contains V_1 to V_4 as well as 46 other variables independently generated from Uniform$(0.1, 4)$. Results from 500 replications for sample size 100, 150, 300, or 500 are shown in Web Table 8.

[Web Table 8 about here.]

The bias in the ACE estimation based on BAC appears to be very small. The RMSE based on BAC is smaller than those based on propensity score methods. To further explore the impact of outcome model misspecification on ACE estimation, which is based on predicted values from the model, we plotted the predicted outcome values from the true outcome model against two misspecified models for each of the four true confounders (Web Figure 4): one misspecified model includes that true confounder but assumes linear associations (left panel); and the other misspecified model excludes that true confounder (right panel). The figure shows that excluding a true confounder has a larger impact on predicted values than misspecifiying the functional form of a true confounder. As shown in Web Figure 5, BAC is able to identify the true confounders and assign high probabilites for including them in the outcome model. Therefore, the bias appeares to be small although the inference is based on assuming linear associations in the outcome model. Furthermore, the ability of BAC to focus inference on a reduced set of factors that are true confounders offer efficiency gains relative to propensity score methods. BAC is designed primarily to identify the true confounders, and in doing so sacrifices some ability to avoid model misspecification by virtue of its reliance on a parametric model. The ability to identify the real confounders from a high dimensional set

of potential confounders is arguably more essential than guarding against misspecification of the functional form.

[Web Figure 4 about here.]

[Web Figure 5 about here.]

Web Appendix K: Simulation Results Comparing BAC with the Method Proposed By Schneeweiss et al. (2009)

In this section, we compare BAC to the method proposed by Schneeweiss et al. (2009) based on simulation studies. Because the method by Schneeweiss et al. (2009) only deals with the situation where the exposure, the outcome, and all potential confounders are binary, we considered the following simulation scenario that is under such situation. We considered 50 potential confounders (V_1 to V_{50}) independently generated from Bernoulli(0.5). The exposure and outcome variables were generated from

$$\text{logit}\{E(X_i|\boldsymbol{V}_i)\} = \delta_0 + \delta_1 V_{1i} + \delta_2 V_{2i} + \delta_3 V_{3i} + \delta_4 V_{4i} + \delta_5 V_{5i} + \delta_6 V_{6i} + \delta_7 V_{7i} + \delta_8 V_{8i} + \delta_9 V_{9i},$$

$$\begin{aligned}\text{logit}\{E(Y_i|X_i, \boldsymbol{V}_i)\} =& \beta_0 + \beta_X X_i + \beta_1 V_{1i} + \beta_2 V_{2i} + \beta_3 V_{3i} + \beta_4 V_{4i} + \beta_5 V_{5i} + \beta_6 V_{6i} + \beta_7 V_{7i} + \beta_8 V_{8i} \\ &+ \beta_9 V_{9i} + \beta_{10} V_{10i} + \beta_{11} V_{11i} + \beta_{12} V_{12i},\end{aligned}$$

where $\delta_0 = -2.4$, $\beta_0 = -3.45$, $\delta_1 = \delta_2 = \delta_3 = \beta_1 = \beta_4 = \beta_7 = \beta_{10} = 1$, $\delta_4 = \delta_5 = \delta_6 = \beta_X = \beta_2 = \beta_5 = \beta_8 = \beta_{11} = 0.5$, and $\delta_7 = \delta_8 = \delta_9 = \beta_3 = \beta_6 = \beta_9 = \beta_{12} = 0.1$.

We considered two variations of the method proposed by Schneeweiss et al. (2009): one takes the top 10 variables ($k = 10$) ranked by the method and includes them in the outcome model; and the other includes the top 20 variables ($k = 20$) in the model. Simulation results are presented in Web Table 9 and Web Figure 6. BAC provides unbiased estimate of the ACE with standard error smaller than that from the full model. In contrast, the method proposed by Schneeweiss et al. (2009) underestimates the standard error when sample size is small, and thus yields lower coverage probability than the desired value 0.95. Although

the coverage probability improves with increased sample size, the standard error of their estimator is larger than that from BAC and is close to that from the full model.

[Web Table 9 about here.]

[Web Figure 6 about here.]

Web Appendix L: Methods Compared Under Each Simulation Scenario

Web Table 10 lists the methods we compared under each simulation scenario.

[Web Table 10 about here.]

References

Lefebvre, G., Delaney, J. A., and McClelland, R. L. (2014). Extending the bayesian adjustment for confounding algorithm to binary treatment covariates to estimate the effect of smoking on carotid intima-media thickness: the multi-ethnic study of atherosclerosis. *Statistics in Medicine* **33,** 2797–2813.

Madigan, D., York, J., and Allard, D. (1995). Bayesian graphical models for discrete data. *International Statistical Review/Revue Internationale de Statistique* **63,** 215–232.

Makalic, E. and Schmidt, D. F. (2011). Review of modern logistic regression methods with application to small and medium sample size problems. In *AI 2010: Advances in Artificial Intelligence*, pages 213–222. Springer.

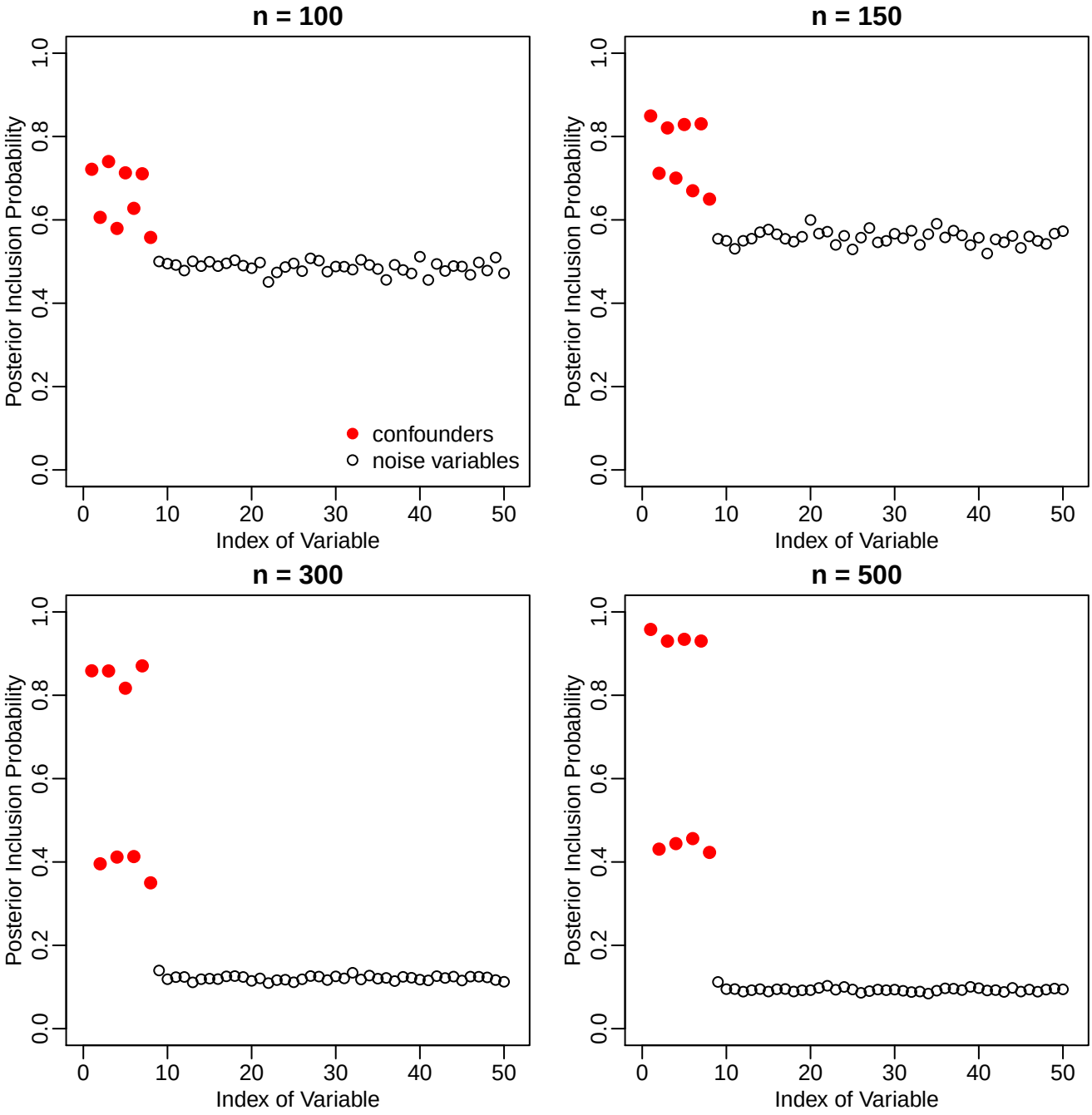
Martin, A. D., Quinn, K. M., and Park, J. H. (2011). Mcmcpack: Markov chain monte carlo in r. *Journal of Statistical Software* **42,** 1–21.

Raftery, A. E. (1995). Bayesian model selection in social research (with discussion). *Sociological Methodology (Peter V. Marsden, ed.)* **25,** 111–164.

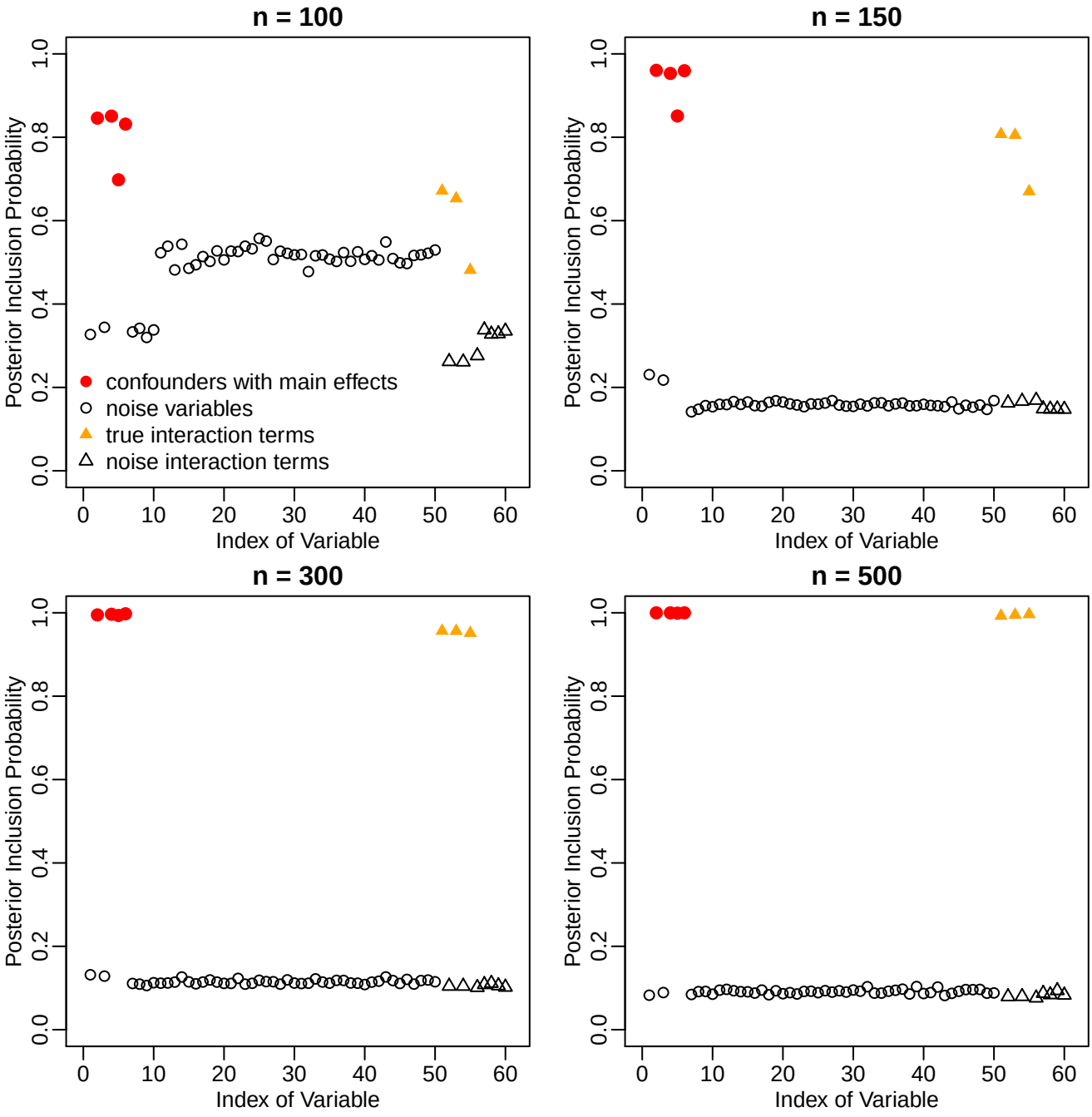
Rubin, D. B. (1981). The bayesian bootstrap. *The Annals of Statistics* **9,** 130–134.

Rubin, D. B. (2007). The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. *Statistics in Medicine* **26,** 20–36.

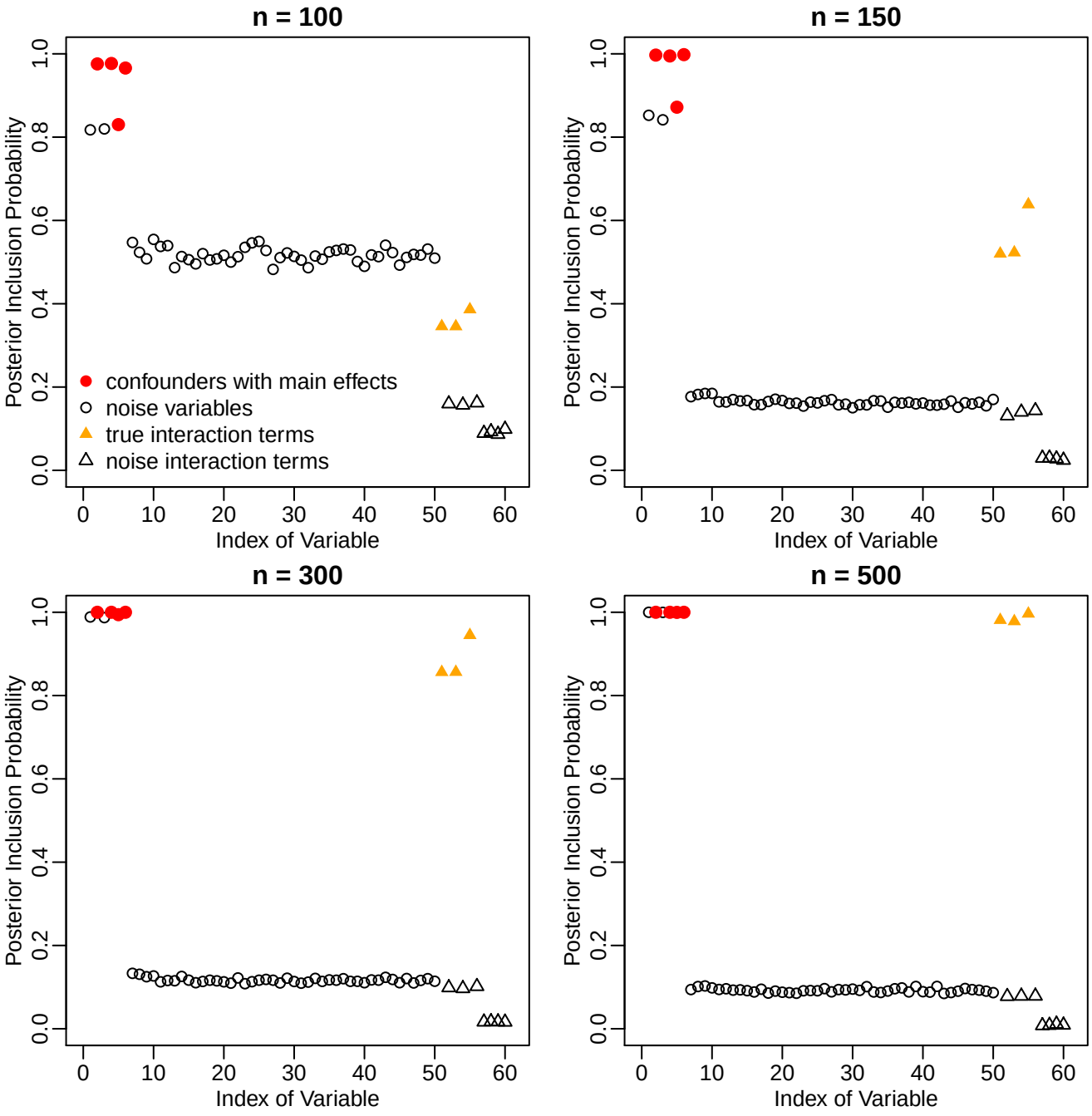
Schneeweiss, S., Rassen, J. A., Glynn, R. J., Avorn, J., Mogun, H., and Brookhart, M. A. (2009). High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. *Epidemiology (Cambridge, Mass.)* **20,** 512.



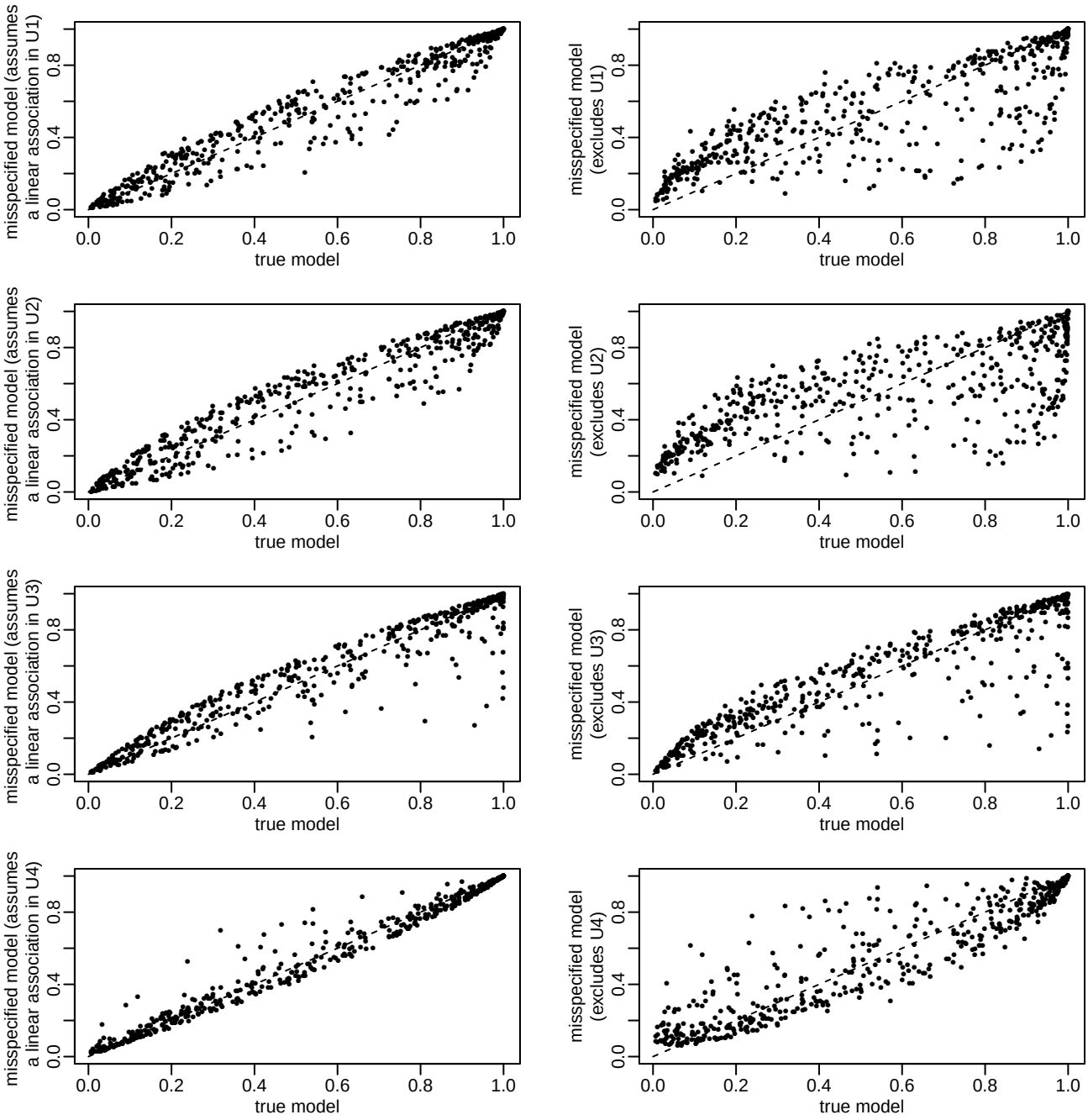
Web Figure 1. Marginal posterior inclusion probabilities of the 50 potential confounders in the simulation scenario presented in Web Appendix H, where there are both continuous and binary confounders and they are correlated.



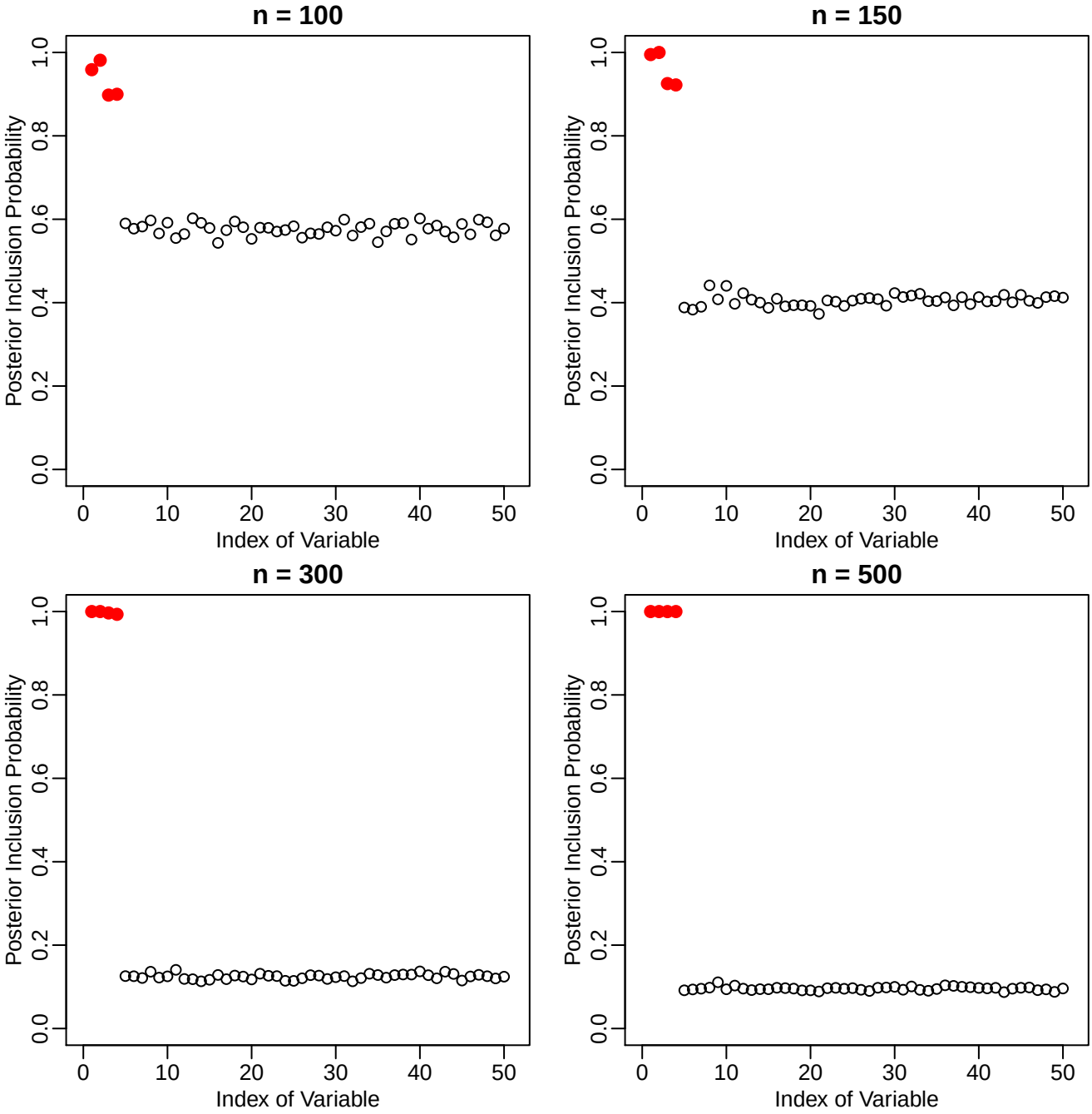
Web Figure 2. Marginal posterior inclusion probabilities of the 50 potential confounders and 10 potential interaction terms in the simulation scenario presented in Web Appendix I, where there are interactions between confounders and the exposure in absense of main effects. Results are based on BAC with the prior in equation (A3).



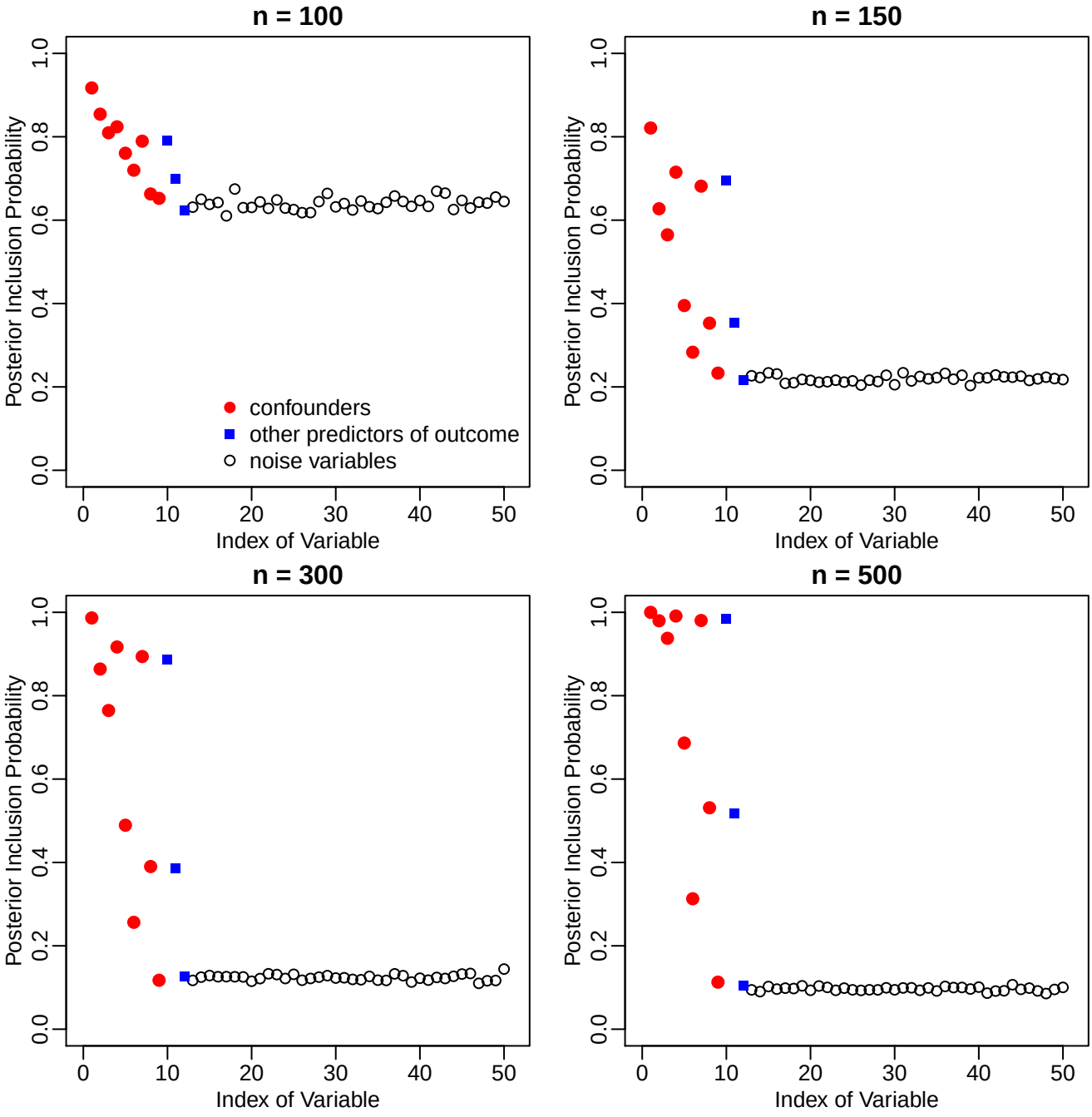
Web Figure 3. Marginal posterior inclusion probabilities of the 50 potential confounders and 10 potential interaction terms in the simulation scenario presented in Web Appendix I, where there are interactions between confounders and the exposure in absense of main effects. Results are based on BAC with the prior in equation (8).



Web Figure 4. Predicted outcome values based on the true model vs. misspecified models for the simulation scenario presented in Web Appendix J, where the outcome model contains non-linear terms of confounders. Two misspecified models were considered for each of the four true confounders: a model includes that true confounder but assumes linear association (Left Panel), and a model excludes that true confounder (Right Panel).



Web Figure 5. Marginal posterior inclusion probabilities of the 50 potential confounders in the simulation scenario presented in Web Appendix J, where the outcome model contains non-linear terms of confounders.



Web Figure 6. Marginal posterior inclusion probabilities of the 50 potential confounders in the simulation scenario presented in Web Appendix K.

Web Table 1
Regression coefficients for the two simulation scenarios in the main text

First simulation scenario				Second simulation scenario			
Exposure model		Outcome model		Exposure model		Outcome model	
δ_1	1	β_X	0.5	δ_0	-1.5	β_X	0.5
δ_2	1	β_1	1	δ_1	0.5	β_0	0.5
δ_3	1	β_2	0.5	δ_2	0.5	β_1	0.5
δ_4	0.5	β_3	0.1	δ_3	0.5	β_2	0.5
δ_5	0.5	β_4	1	δ_4	0.5	β_3	0.5
δ_6	0.5	β_5	0.5	δ_5	0.5	β_4	0.5
δ_7	0.1	β_6	0.1	δ_6	0.5	β_5	0.5
δ_8	0.1	β_7	1			β_6	0.5
δ_9	0.1	β_8	0.5			β_7^*	-0.5
		β_9	0.1			β_8^*	-0.5
		β_{10}	1			β_9^*	-0.5
		β_{11}	0.5				
		β_{12}	0.1				

*: coefficients for interaction terms

Web Table 2

Number of times out of the 500 simulation replicates that merging strata is needed for the stratification method for the two simulation scenarios in Section 3

		Needed to merged into less than 5 strata		Needed to merged into a single stratum	
		PS_F	PS_S	PS_F	PS_S
First Simulation	$n = 100$	497	485	440	460
Scenario	$n = 150$	255	129	8	9
Second Simulation	$n = 100$	478	427	330	379
Scenario	$n = 150$	142	60	20	5

Web Table 3

Standardized bias B values calculated based on estimated propensity socres (PS_F, PS_S, and twang) and true propensity scores (True PS) for the two simulation scenarios in the main text of the paper. Results are averaged over 500 replicates.

	Scenario One				Scenario Two			
	Estimated PS			True PS	Estimated PS			True PS
	PS_F	PS_S	twang		PS_F	PS_S	twang	
$n = 100$	1.72	1.67	4.68	0.90	1.59	1.53	4.75	0.56
$n = 150$	1.68	1.52	4.14	0.90	1.50	1.29	4.23	0.58
$n = 300$	1.32	1.18	3.41	0.90	1.08	0.92	3.50	0.57
$n = 500$	1.14	1.05	2.92	0.89	0.90	0.79	3.08	0.58

Web Table 4

Propensity score stratum-specific average causal effect

Quantile	Average age	Surgery ($n = 7037$)			Nonsurgery ($n = 8023$)			Stratum-specific average causal effect	Confidence Interval
		Total	Readmitted		Total	Readmitted			
		No.	No.	%	No.	No.	%		
$< 1^{st}$	83	821	169	20.6	2191	509	23.2	−0.026	(−0.059, 0.006)
$1^{st} - 2^{nd}$	78.3	1237	279	22.6	1775	478	26.9	−0.044	(−0.075, −0.013)
$2^{nd} - 3^{rd}$	74.7	1526	352	23.1	1500	440	29.3	−0.063	(−0.094, −0.031)
$3^{rd} - 4^{th}$	71.6	1669	320	19.2	1347	414	30.7	−0.116	(−0.147, −0.085)
$> 4^{th}$	68.2	1784	345	19.3	1210	373	30.8	−0.115	(−0.147, −0.083)

Numbers were calculated based on the PS_F method. The results from the PS_S method are similar.

Web Table 5
Estimation of ACE from samples generated by randomly selecting 0.5%, 1%, or 2% of the brain tumor Medicare data

		full model	PS_F	PS_S	PS_{RF}	PS_{RS}	$twang_N$	$twang_F$	$twang_S$	BAC_N	BAC_I
Sample 0.5% of the data ($n = 75$)	BIAS	0.002	—	—	—	—	0.008	0.025	0.008	—	0.007
	SEE	0.147	—	—	—	—	0.100	0.122	0.100	—	0.089
	SSE	0.120	—	—	—	—	0.089	0.114	0.089	—	0.078
	RMSE	0.120	—	—	—	—	0.089	0.117	0.089	—	0.078
	CP	0.99	—	—	—	—	0.95	0.99	0.95	—	0.97
Sample 1% of the data ($n = 150$)	BIAS	0.005	0.004	0.002	0.013	0.001	0.005	0.010	0.005	0.000	0.001
	SEE	0.080	0.075	0.077	0.147	0.091	0.074	0.082	0.074	0.069	0.067
	SSE	0.076	0.085	0.078	0.137	0.127	0.063	0.076	0.063	0.070	0.061
	RMSE	0.077	0.085	0.078	0.137	0.127	0.063	0.077	0.063	0.070	0.061
	CP	0.97	0.91	0.95	0.96	0.84	0.98	0.98	0.98	0.94	0.96
Sample 2% of the data ($n = 300$)	BIAS	0.003	0.003	0.002	0.003	0.006	0.005	0.002	0.005	0.002	0.005
	SEE	0.054	0.053	0.053	0.091	0.054	0.054	0.055	0.054	0.050	0.048
	SSE	0.050	0.052	0.051	0.065	0.066	0.045	0.052	0.045	0.048	0.045
	RMSE	0.050	0.052	0.051	0.065	0.066	0.045	0.052	0.045	0.048	0.045
	CP	0.96	0.95	0.96	0.99	0.89	0.98	0.98	0.98	0.96	0.96

For each method, the estimated ACEs from randomly selected samples were compared to the ACE estimate from the whole dataset, which serves as the "true" ACE value. "—" indicates results are unavailable.

Web Table 6

Estimation of ACE for the simulation scenario presented in Web Appendix H, where there are both continuous and binary confounders and they are correlated.

		true model	full model	BAC_N	BAC_I
$n = 100$	BIAS	0.005	0.064	—	0.033
	SEE	0.107	0.282	—	0.083
	SSE	0.105	0.221	—	0.079
	RMSE	0.105	0.230	—	0.086
	CP	0.94	0.99	—	0.94
$n = 150$	BIAS	0.006	0.071	—	0.024
	SEE	0.082	0.196	—	0.069
	SSE	0.084	0.174	—	0.074
	RMSE	0.084	0.188	—	0.078
	CP	0.94	0.99	—	0.92
$n = 300$	BIAS	0.002	0.006	0.006	0.009
	SEE	0.055	0.092	0.055	0.054
	SSE	0.053	0.061	0.054	0.049
	RMSE	0.053	0.061	0.054	0.050
	CP	0.95	0.98	0.95	0.96
$n = 500$	BIAS	0.002	0.004	0.005	0.007
	SEE	0.042	0.049	0.043	0.042
	SSE	0.041	0.045	0.042	0.040
	RMSE	0.041	0.045	0.042	0.041
	CP	0.96	0.95	0.95	0.95

"—" indicates results are unavailable.

Web Table 7

Estimation of ACE for the simulation scenario presented in Web Appendix I, where there are interactions between confounders and the exposure in absense of main effects.

		true model	full model	BAC_N prior in equation (8)	BAC_N prior in equation (A3)
$n = 100$	BIAS	0.072	—	0.196	0.004
	SEE	0.639	—	1.092	1.062
	SSE	0.547	—	0.844	0.826
	RMSE	0.551	—	0.866	0.825
	CP	0.96	—	0.97	0.97
$n = 150$	BIAS	0.057	0.179	0.088	0.061
	SEE	0.496	27.001	0.582	0.577
	SSE	0.510	0.793	0.570	0.564
	RMSE	0.513	0.812	0.577	0.567
	CP	0.94	0.99	0.95	0.94
$n = 300$	BIAS	0.008	0.013	0.024	0.034
	SEE	0.323	0.480	0.360	0.348
	SSE	0.323	0.384	0.341	0.337
	RMSE	0.322	0.384	0.341	0.338
	CP	0.93	0.97	0.95	0.95
$n = 500$	BIAS	0.007	0.022	0.041	0.040
	SEE	0.247	0.299	0.264	0.256
	SSE	0.246	0.271	0.253	0.249
	RMSE	0.246	0.272	0.256	0.252
	CP	0.95	0.95	0.96	0.95

"—" indicates results are unavailable.

Web Table 8

Estimation of ACE based on the simulation scenario presented in Web Appendix J, where the outcome model contains non-linear terms of confounders.

		true model	PS_F	PS_S	PS_{RF}	PS_{RS}	$twang_N$	$twang_F$	$twang_S$	BAC_N	BAC_I
$n = 100$	BIAS	0.007	0.174	0.175	0.048	0.057	0.109	0.053	0.089	—	0.038
	SEE	0.080	0.097	0.096	0.241	0.113	0.101	0.110	0.104	—	0.079
	SSE	0.083	0.110	0.108	0.211	0.183	0.076	0.098	0.116	—	0.077
	RMSE	0.083	0.206	0.205	0.216	0.192	0.133	0.111	0.146	—	0.086
	CP	0.94	0.53	0.53	0.99	0.68	0.87	1.00	0.83	—	0.93
$n = 150$	BIAS	0.002	0.048	0.038	0.021	0.043	0.104	0.037	0.063	—	0.020
	SEE	0.064	0.095	0.090	0.205	0.097	0.085	0.091	0.081	—	0.068
	SSE	0.065	0.137	0.124	0.197	0.142	0.059	0.077	0.091	—	0.069
	RMSE	0.065	0.145	0.130	0.198	0.149	0.120	0.085	0.110	—	0.072
	CP	0.93	0.78	0.79	0.96	0.79	0.85	1.00	0.84	—	0.93
$n = 300$	BIAS	0.002	0.020	0.017	0.009	0.037	0.096	0.016	0.025	0.006	0.009
	SEE	0.045	0.072	0.068	0.188	0.066	0.062	0.062	0.054	0.051	0.052
	SSE	0.046	0.075	0.066	0.121	0.100	0.043	0.064	0.062	0.051	0.049
	RMSE	0.046	0.077	0.068	0.121	0.106	0.105	0.066	0.067	0.052	0.050
	CP	0.94	0.92	0.93	1.00	0.76	0.72	0.98	0.89	0.94	0.96
$n = 500$	BIAS	0.002	0.013	0.011	0.025	0.031	0.086	0.004	0.006	0.007	0.004
	SEE	0.034	0.053	0.050	0.109	0.055	0.051	0.047	0.040	0.040	0.040
	SSE	0.034	0.046	0.043	0.082	0.079	0.035	0.041	0.042	0.038	0.037
	RMSE	0.034	0.048	0.045	0.086	0.085	0.093	0.041	0.042	0.039	0.037
	CP	0.94	0.96	0.98	0.98	0.77	0.65	0.97	0.94	0.95	0.97

"—" indicates results are unavailable.

Web Table 9
Estimation of ACE with comparison to Schneeweiss's method based on the simulation scenario presented in Web Appendix K.

		true model	full model	Schneeweiss's method $k = 10$	Schneeweiss's method $k = 20$	BAC_N	BAC_I
$n = 100$	BIAS	0.006	0.056	0.010	0.001	—	0.006
	SEE	0.110	0.260	0.086	0.090	—	0.082
	SSE	0.106	0.223	0.124	0.149	—	0.087
	RMSE	0.106	0.229	0.124	0.149	—	0.087
	CP	0.95	1.00	0.82	0.74	—	0.94
$n = 150$	BIAS	0.002	0.025	0.008	0.002	—	0.002
	SEE	0.085	0.197	0.078	0.079	—	0.075
	SSE	0.083	0.132	0.095	0.107	—	0.076
	RMSE	0.083	0.134	0.096	0.107	—	0.076
	CP	0.94	1.00	0.87	0.85	—	0.94
$n = 300$	BIAS	0.001	0.002	0.005	0.003	0.001	0.002
	SEE	0.058	0.071	0.061	0.062	0.056	0.056
	SSE	0.055	0.060	0.060	0.065	0.055	0.052
	RMSE	0.055	0.060	0.060	0.065	0.055	0.052
	CP	0.96	0.98	0.96	0.95	0.97	0.97
$n = 500$	BIAS	0.001	0.002	0.003	0.002	0.000	0.001
	SEE	0.044	0.048	0.048	0.050	0.043	0.043
	SSE	0.043	0.046	0.047	0.048	0.044	0.042
	RMSE	0.043	0.046	0.047	0.048	0.044	0.042
	CP	0.95	0.96	0.94	0.95	0.94	0.95

"—" indicates results are unavailable.

Web Table 10
Methods Compared Under Each Simulation Scenario.

	true model	full model	*Ad hoc*$_1$	*Ad hoc*$_5$	PS$_F$	PS$_S$	PS$_{RF}$	PS$_{RS}$	twang$_N$	twang$_F$	twang$_S$	Schneeweiss's method	BAC$_N$	BAC$_I$
Scenario One	x	x	x	x	x	x	x	x	x	x	x		x	x
Scenario Two	x	x			x	x	x	x	x	x	x		x	
Web Appendix G		x			x	x	x	x	x	x	x		x	x
Web Appendix H	x	x											x	x
Web Appendix I	x	x											x	
Web Appendix J	x				x	x	x	x	x	x	x		x	x
Web Appendix K	x	x										x	x	x