Web-based Supplementary Materials for "Comparing treatments via the propensity score: stratification or modeling?" by Jessica A. Myers and Thomas A. Louis

February 10, 2012

Web Supplement A

In this section, we present R code used to apply each analysis approach discussed in Section 2 of the main text. The function `mseHat` returns estimated mean squared error (MSE) for the stratified estimator given the data and a set of cut points for stratification. This function is used by the `optim` function to find the partition that minimizes estimated MSE. The function `strat.est` returns the treatment effect estimate and variance for the stratified approach.

```
## Estimate the GAM
library(mgcv)
gfit <- gam(y ~ z + s(ps, k=5))
delta.gam <- gfit$coefficients[2]
var.gam <- summary(gfit)$se[2]^2

## Save the estimated GAM for use in estimating MSE
g <<- function(x) predict(gfit, newdata = data.frame("z"=0, "ps"=x))
```

```
## Find EF t and use as starting values to find optimal t for K = 6
tef <- quantile(ps, 1:5/6)
topt <- optim(tef, mseHat, method = "L-BFGS-B",
              lower = 0, upper = 1)$par

## Calculate treatment effect estimates and variances for EF and
## optimal stratifications
se.ef <- strat.est(tef, y, z, ps)
delta.ef <- se.ef$delta
var.ef <- se.ef$var
se.opt <- strat.est(topt, y, z, ps)
delta.opt <- se.opt$delta
var.opt <- se.opt$var

## functions
mseHat <<- function(t){
  K <- length(t)+1
  t <- c(0, t, 1)
  f1hat <- approxfun(density(ps[z==1], from = 0, to = 1))
  f0hat <- approxfun(density(ps[z==0], from = 0, to = 1))
  strata <- cut(ps, t, include.lowest = TRUE)
  n <- table(strata)
  n1 <- table(strata[z==1])
  n0 <- table(strata[z==0])
  v1 <- tapply(y[z==1], strata[z==1], var)
```

```
  v0 <- tapply(y[z==0], strata[z==0], var)
  v <- v1/n1 + v0/n0
  b <- 1:K
  for(k in 1:K) {
    M1 <- integrate(f1hat, lower = t[k], upper = t[k+1])$value
    M0 <- integrate(f0hat, lower = t[k], upper = t[k+1])$value
                                        # estimated bias
    b[k] <- integrate(function(u) g(u)*(f1hat(u)/M1 - f0hat(u)/M0),
                      lower = t[k], upper = t[k+1],
                      subdivisions = 500)$value
  }
  kp <- !is.na(v) ## strata with infinite variance are removed
  w <- n[kp]/sum(n[kp])
  sum(b[kp]*w)^2 + sum(v[kp]*w^2) # bias^2 + variance
}

strat.est <- function(t, y, z, ps) {
  t <- c(0, t, 1)
  strata <- cut(ps, t, include.lowest = TRUE)
  n <- table(strata)
  n1 <- table(strata[z==1])
  n0 <- table(strata[z==0])
  v1 <- tapply(y[z==1], strata[z==1], var)
  v0 <- tapply(y[z==0], strata[z==0], var)
  v <- v1/n1 + v0/n0
  m1 <- tapply(y[z==1], strata[z==1], mean)
```

```
  m0 <- tapply(y[z==0], strata[z==0], mean)
  delta <- m1-m0
  kp <- !is.na(var)
  w <- n[kp]/sum(n[kp])
  list("delta" = sum(delta[kp]*w), "var" = sum(var[kp]*w^2),
       "K" = sum(kp))
}
```

Web Supplement B

In this section, we present the complete results of simulations, as described in Sections 3 and 4 of the main text. The results for data simulated under the additive model are given in Figure 1 in the main text and Web Figures 1-2. The results for data simulated under the log-additive model are given in Web Figures 3-5. The results for the simulation study of methods using a poorly-estimated propensity score in Section 4 are shown in Web Figures 6-7.

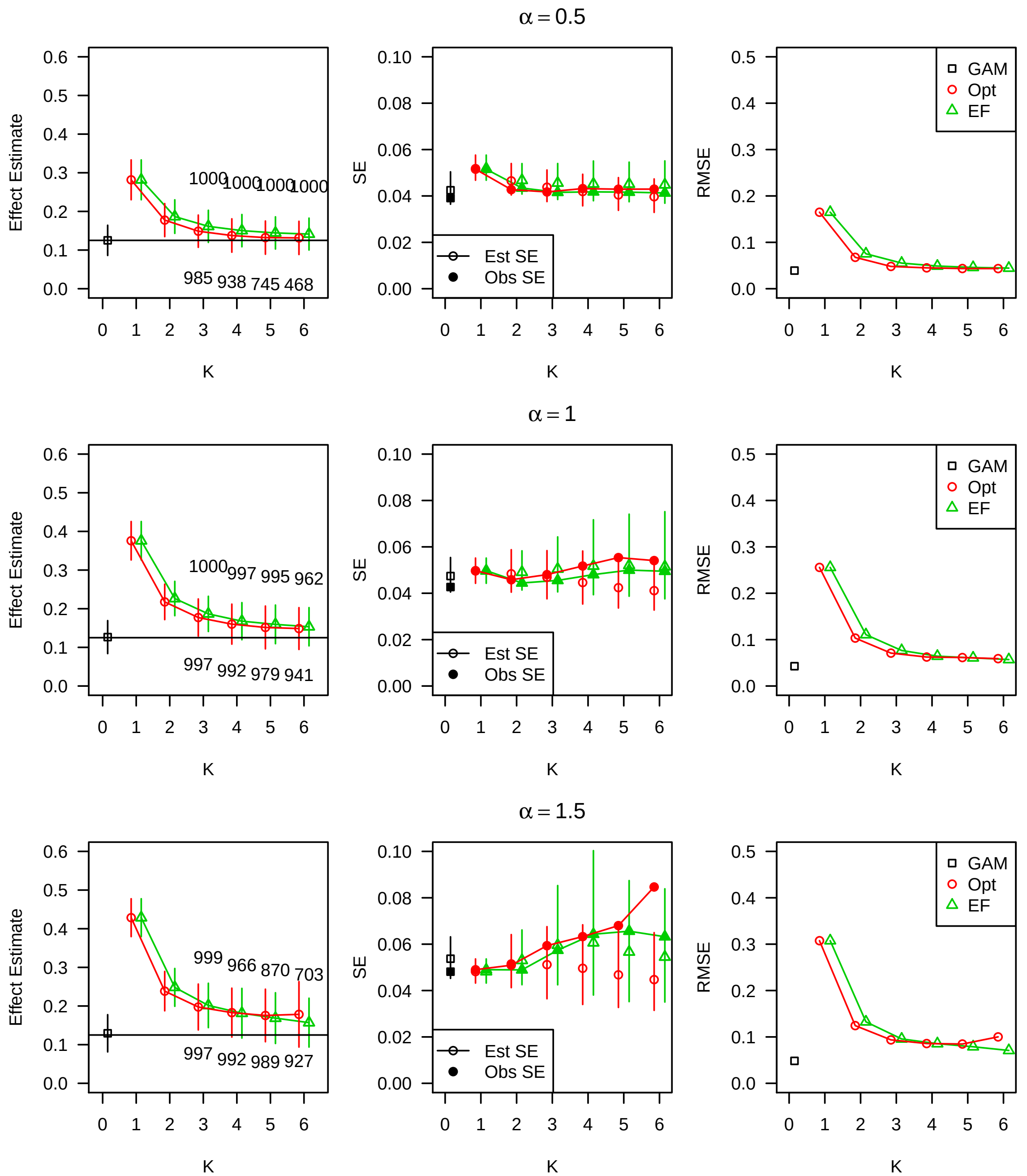

Figure 1: Simulation results for data simulated under the additive model with covariate-outcome relations (B), corresponding to $h_1(x) = 0.25x$ and $h_2(x) = 0.025(x + x^3)$.

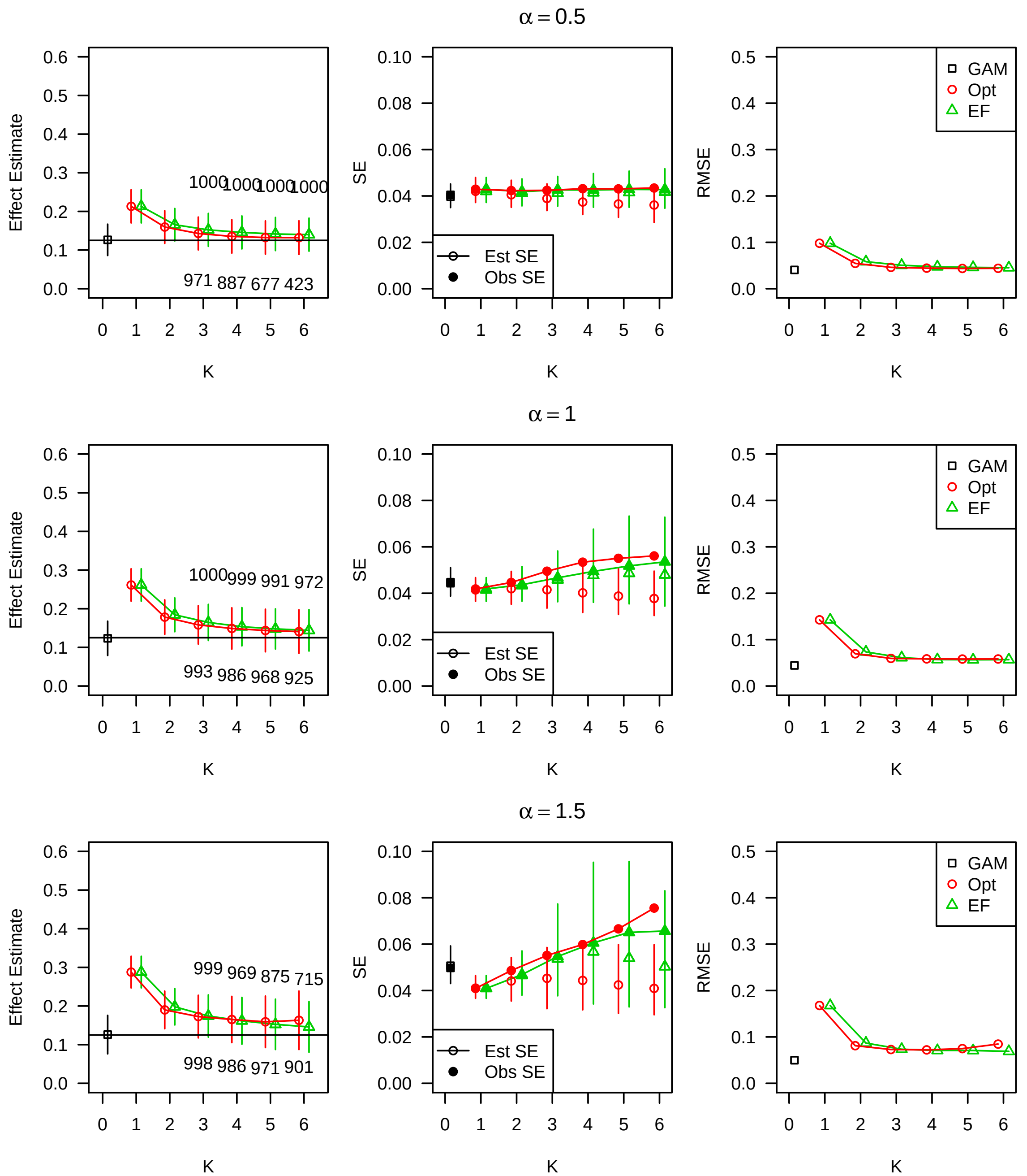

Figure 2: Simulation results for data simulated under the additive model with covariate-outcome relations (C), corresponding to $h_1(x) = h_2(x) = 0.025(x + x^3)$.

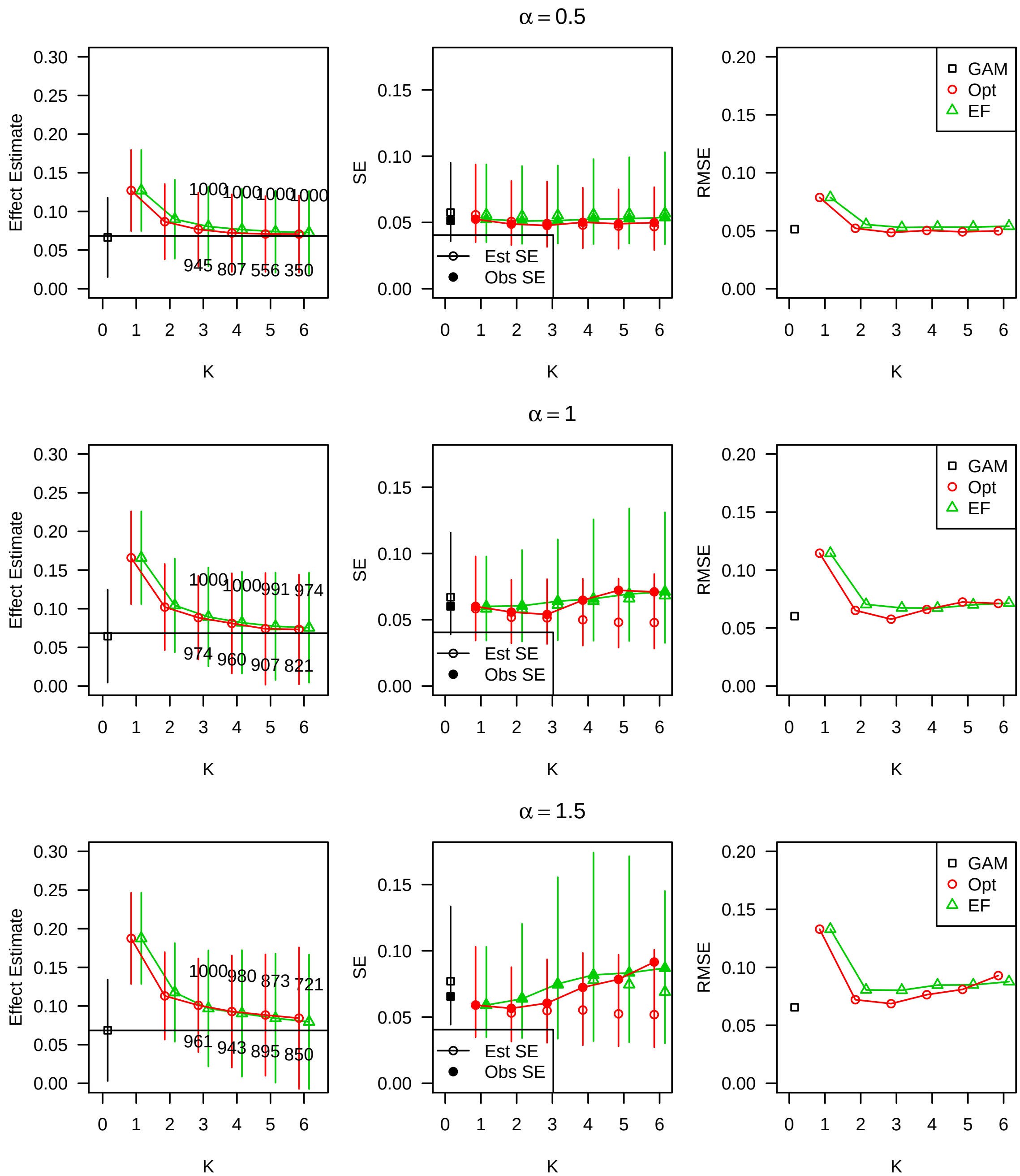

Figure 3: Simulation results for data simulated under the log-additive model with covariate-outcome relations (A), corresponding to $h_1(x) = h_2(x) = 0.25x$.

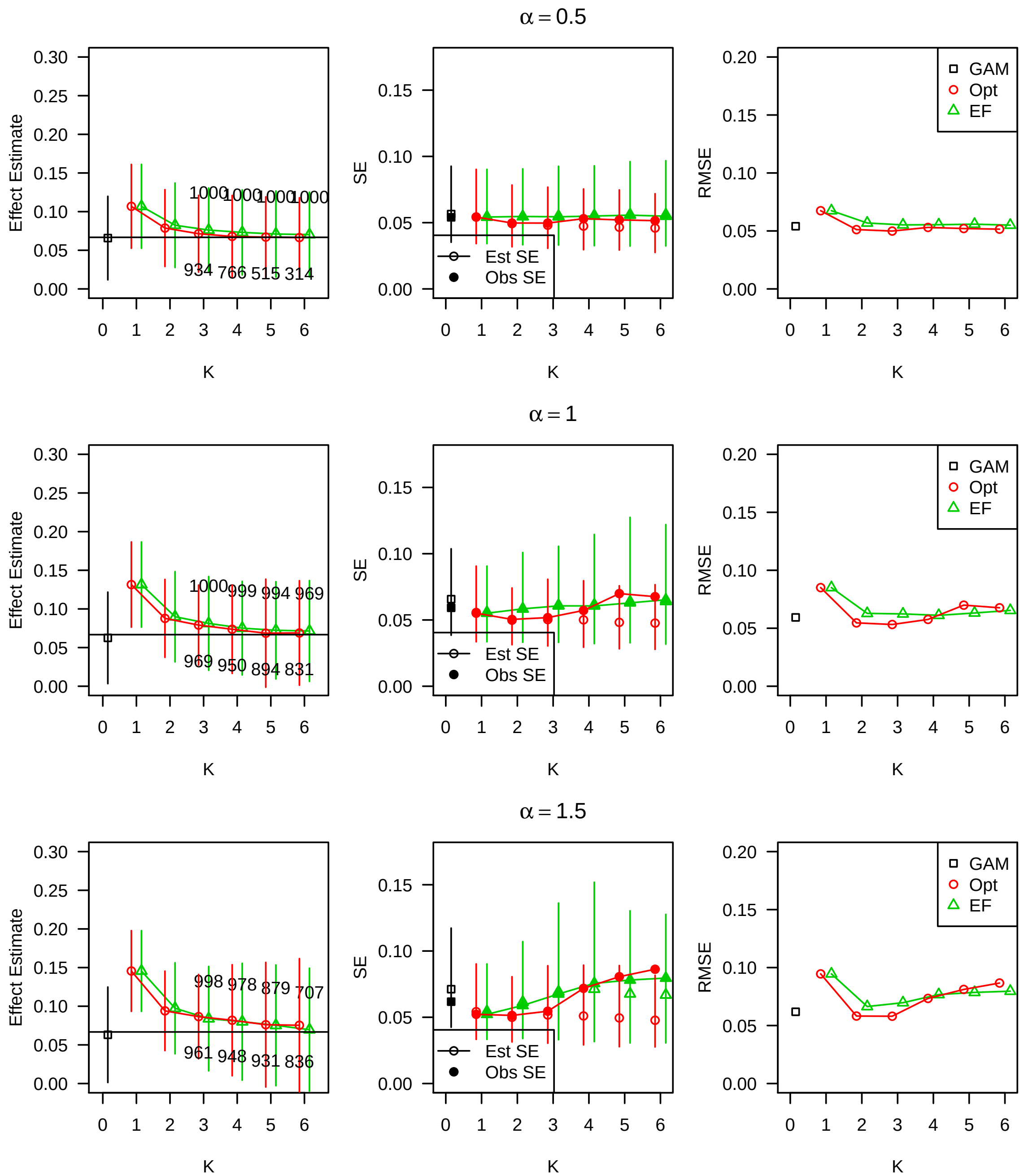

Figure 4: Simulation results for data simulated under the log-additive model with covariate-outcome relations (B), corresponding to $h_1(x) = 0.25x$ and $h_2(x) = 0.025(x + x^3)$.

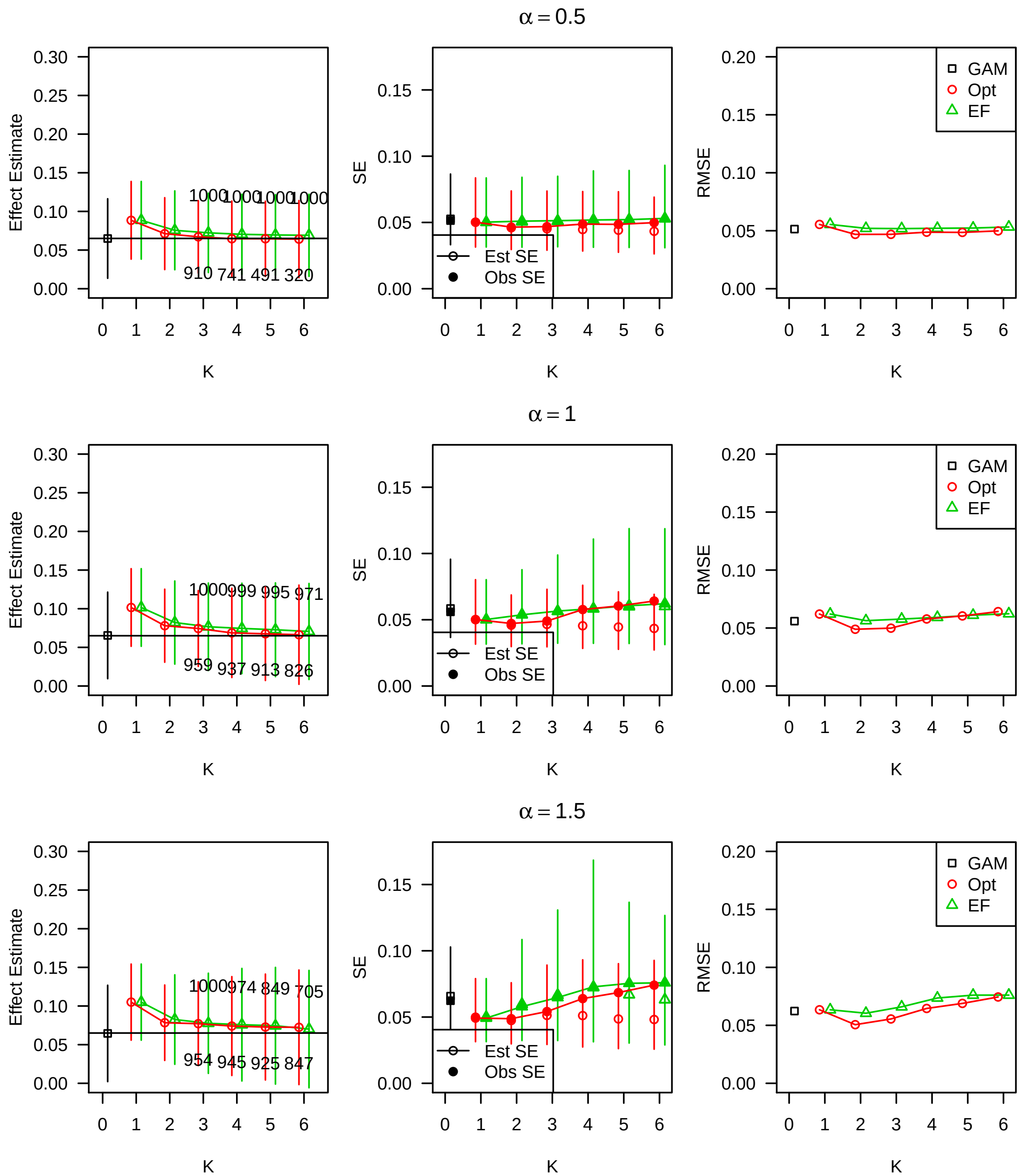

Figure 5: Simulation results for data simulated under the log-additive model with covariate-outcome relations (C), corresponding to $h_1(x) = h_2(x) = 0.025(x + x^3)$.

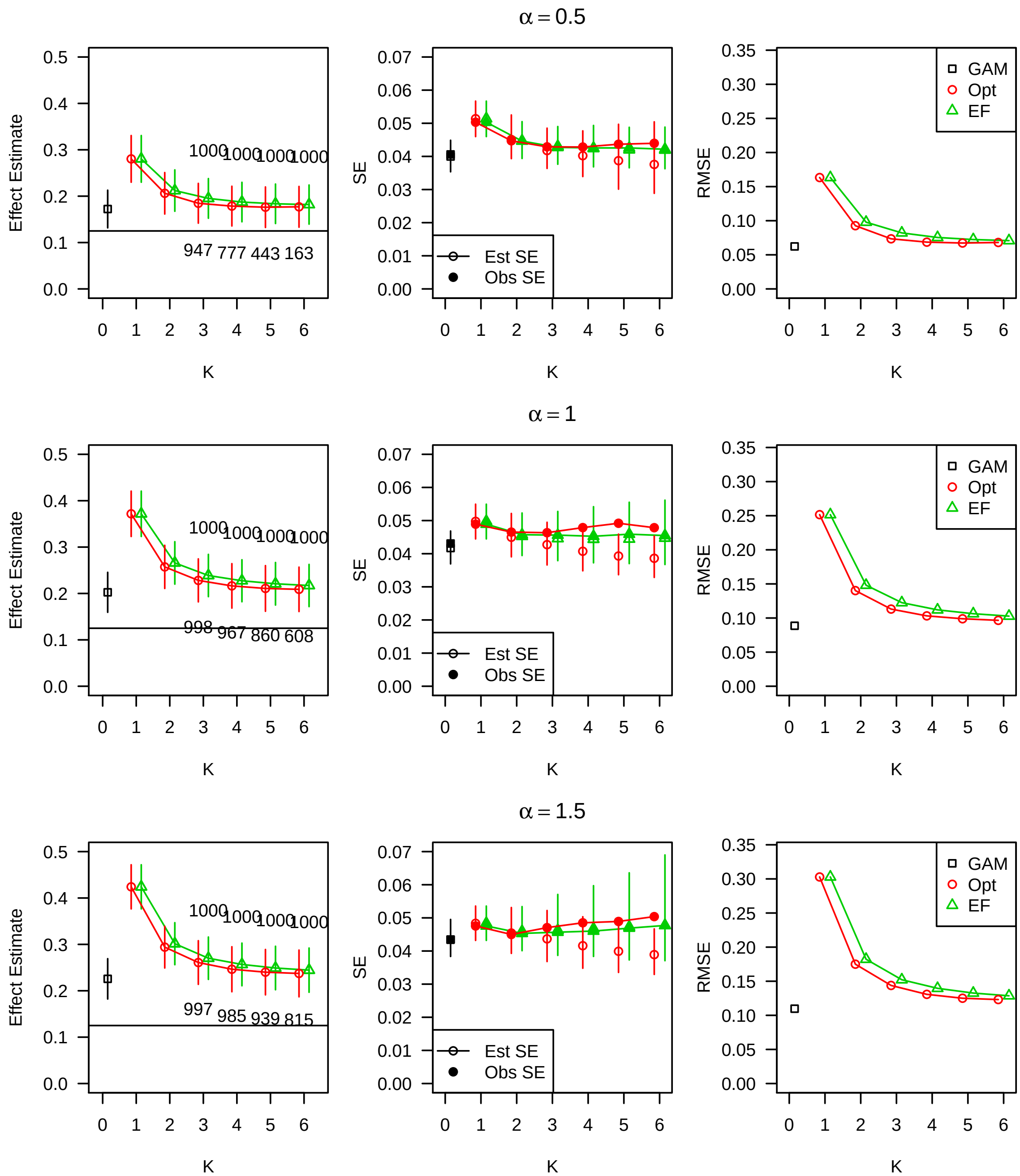

Figure 6: Simulation results for methods using a poorly-estimated propensity score. Outcomes are generated under an additive model with covariate-outcome relations (A), corresponding to $h_1(x) = 0.25x$ and $h_2(x) = 0.025(x + x^3)$.

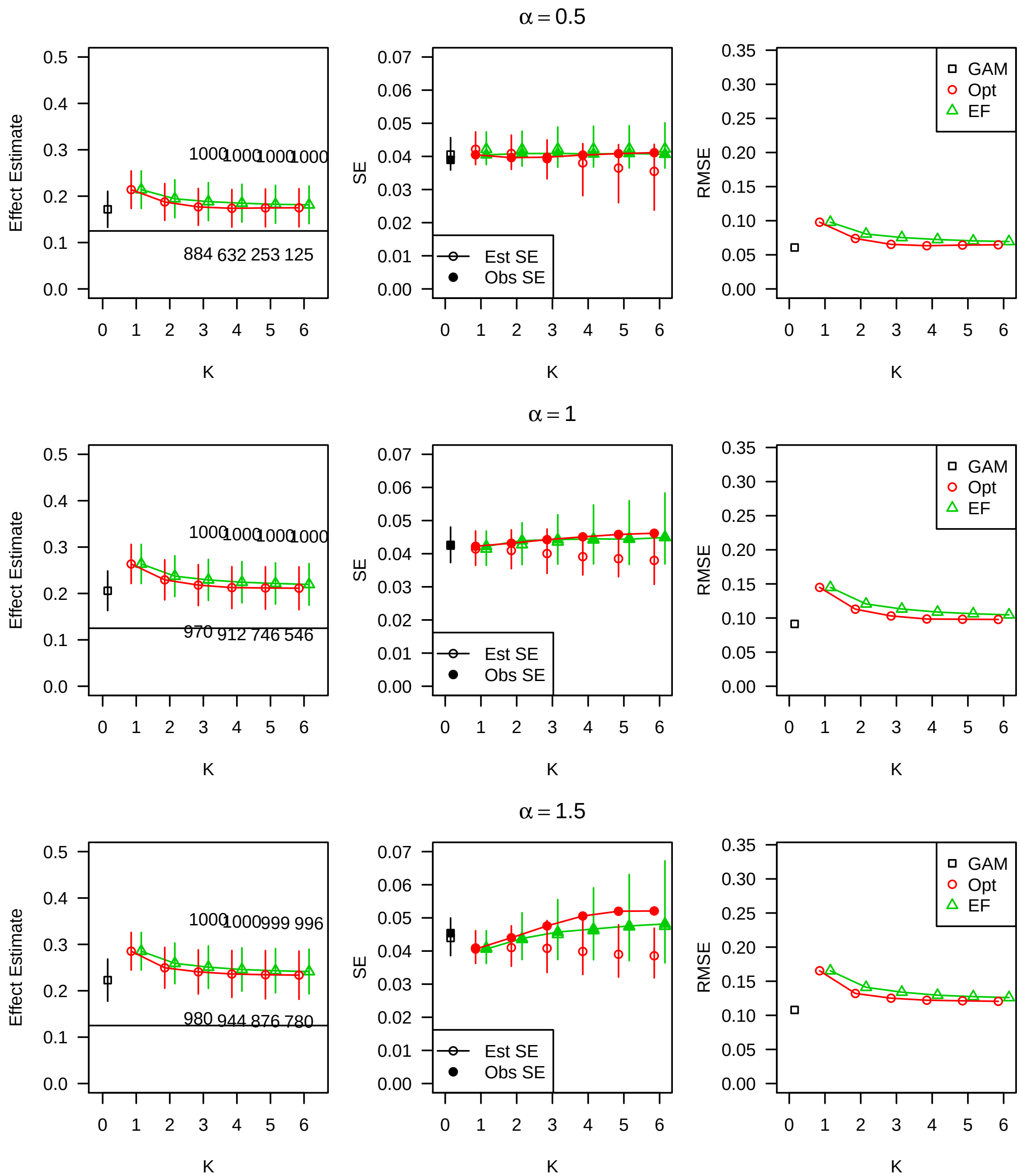

Figure 7: Simulation results for methods using a poorly-estimated propensity score. Outcomes are generated under an additive model with covariate-outcome relations (B), corresponding to $h_1(x) = h_2(x) = 0.025(x + x^3)$.

Web Supplement C

In this section, we present information regarding the development and use of propensity scores in the analysis of insurance plan choice and asthma care from Section 4 of the main text. Web Figures 8 and 9 display the checks of association, controlling for treatment, between the outcome and each categorical and continuous covariate, respectively. Web Figures 10 and 11 show the checks of balance for categorical and continuous covariates, respectively, on the estimated propensity scores. These figures were checked to compare the balance achieved across several competing propensity score models. The figures in this section show the balance for the final propensity score model chosen. Web Figure 12 shows the estimated smooth term for propensity score from the GAM regression of satisfaction with asthma care on treatment and propensity score.

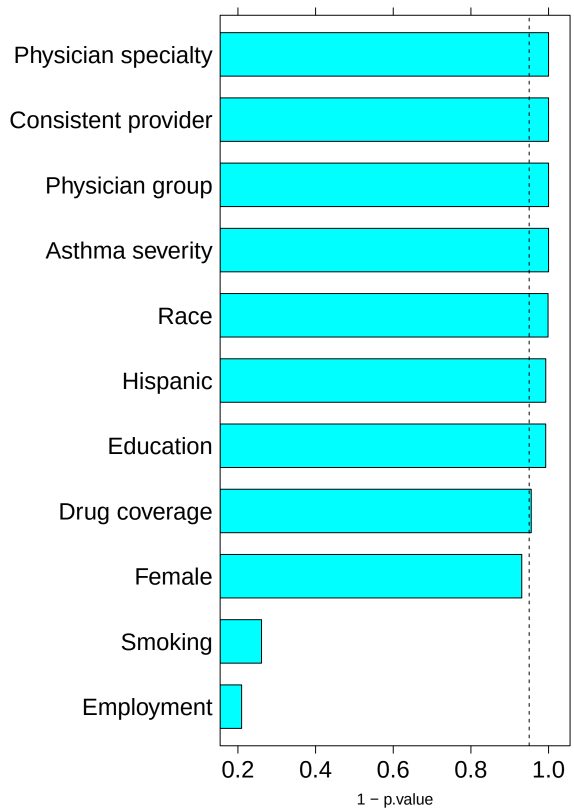

Figure 8: The p-value (subtracted from one) for the likelihood ratio test of each categorical input in a logistic regression model of satisfaction with asthma care on treatment and the covariate.

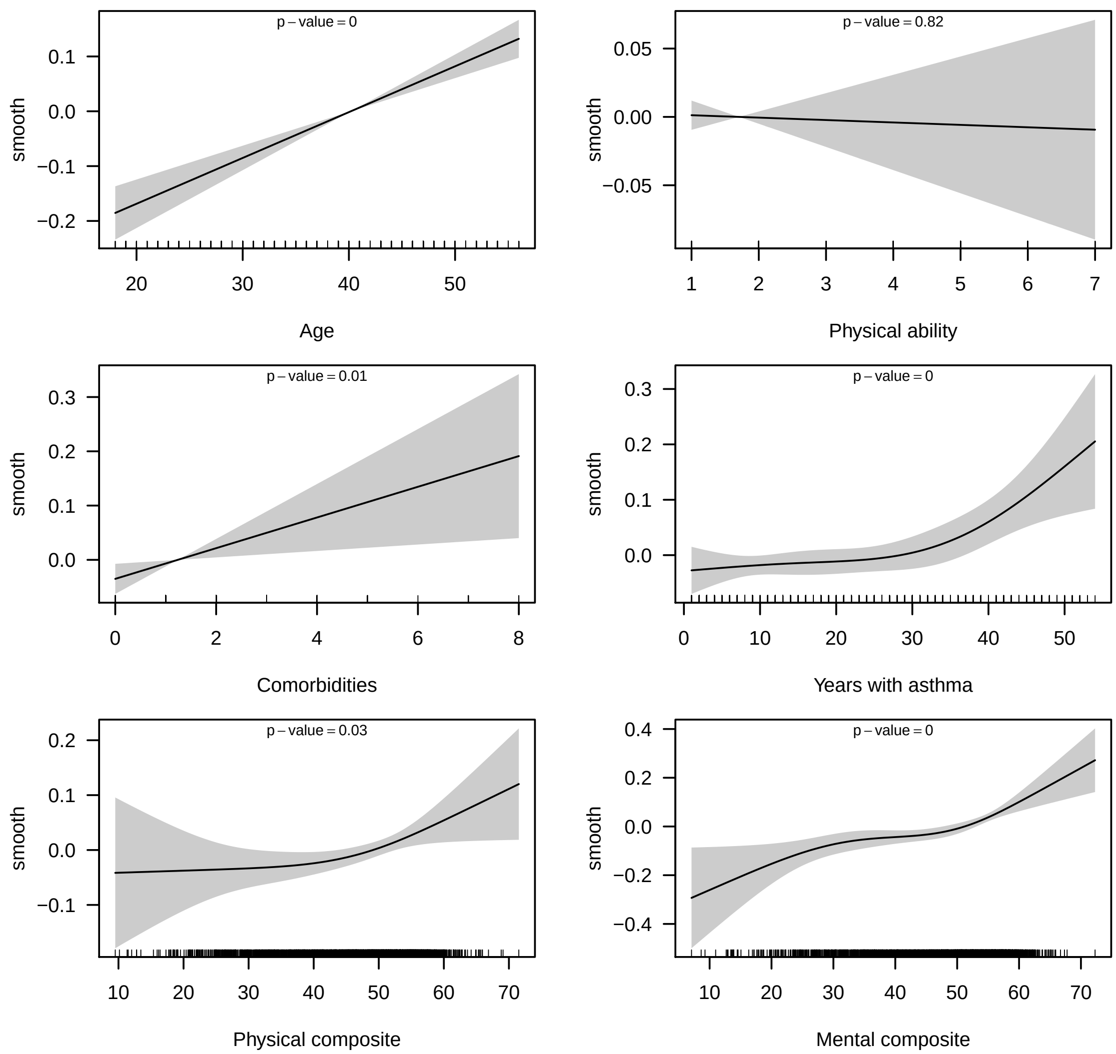

Figure 9: Estimated relation between each continuous covariate and treatment, as estimated in a logistic GAM of satisfaction with asthma care on treatment and the covariate.

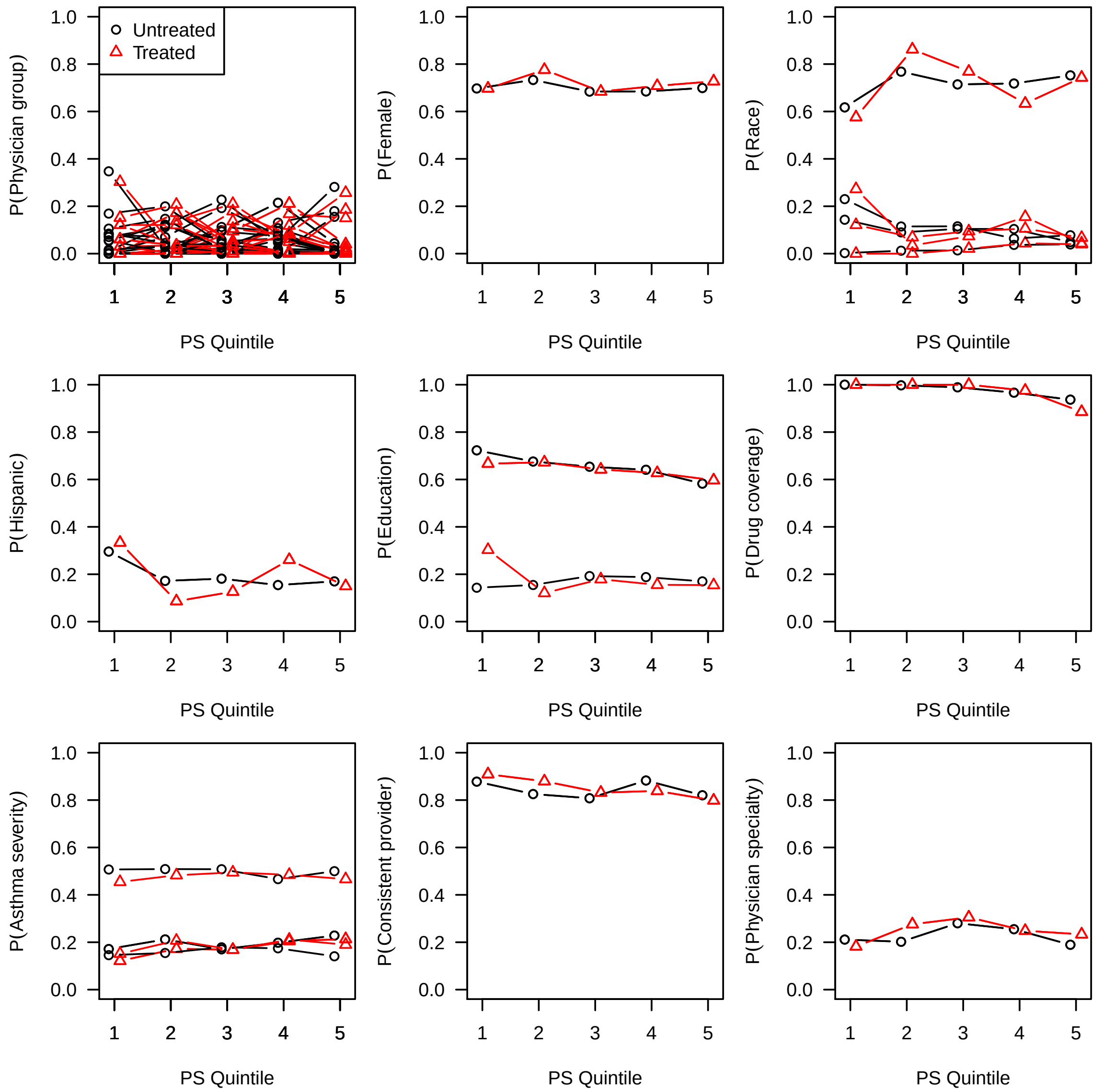

Figure 10: For each covariate with m categories, the proportions of individuals in each of $m-1$ of the categories, stratified by treatment and propensity score quintile. Lines connect proportions across quintiles within category and treatment group. These plots were checked for several competing propensity score models; the plots shown here are for the final model chosen.

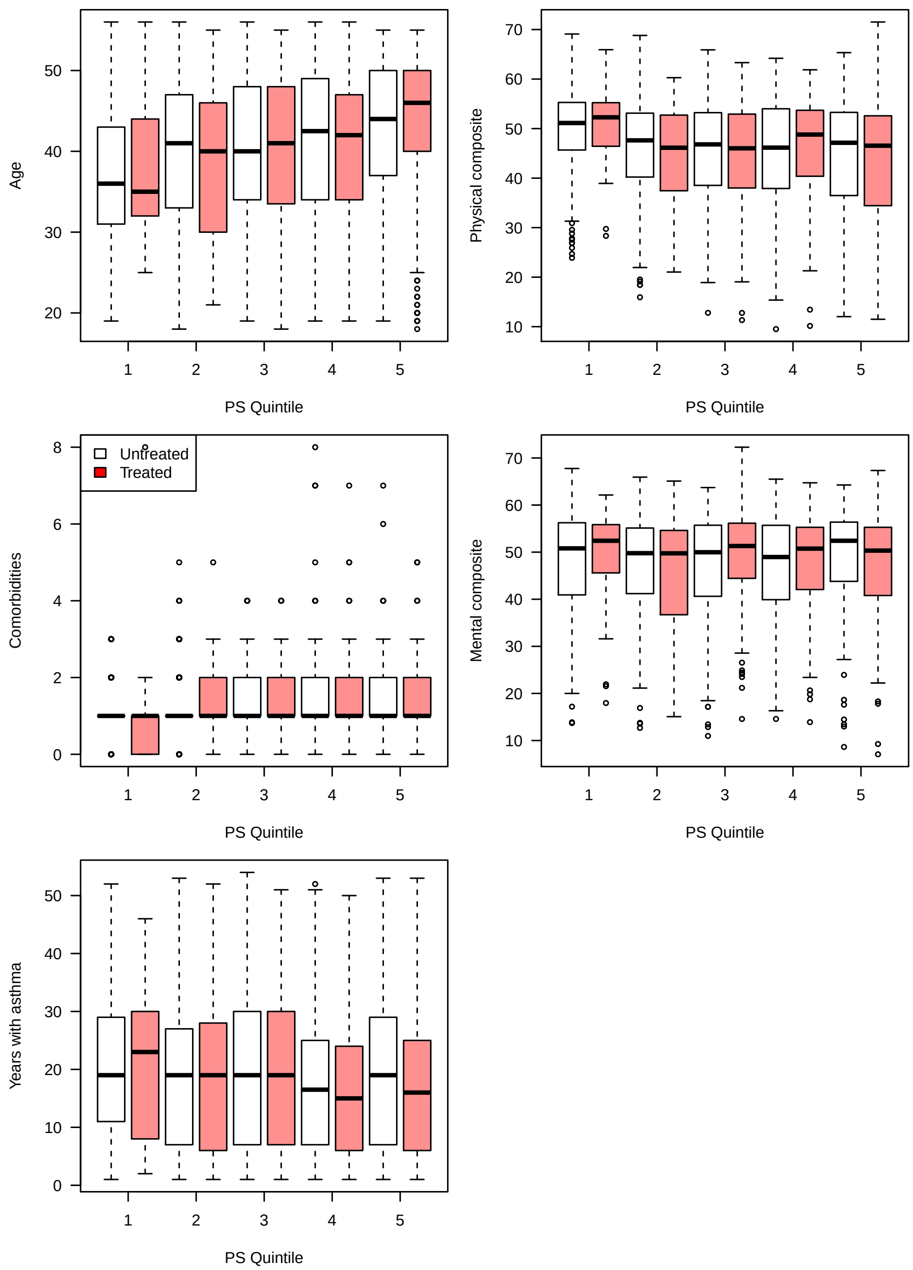

Figure 11: Boxplots of each continuous covariate, stratified by treatment and propensity score quintile. These plots were checked for several competing propensity score models; the plots shown here are for the final model chosen.

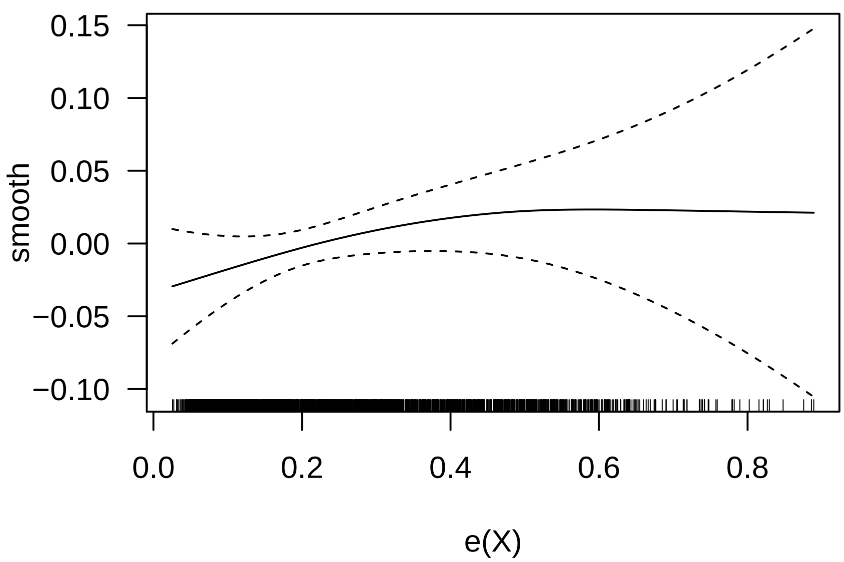

Figure 12: Estimated relation between propensity score and satisfaction with asthma care, adjusting for treatment.