Appendix for Simulation Study
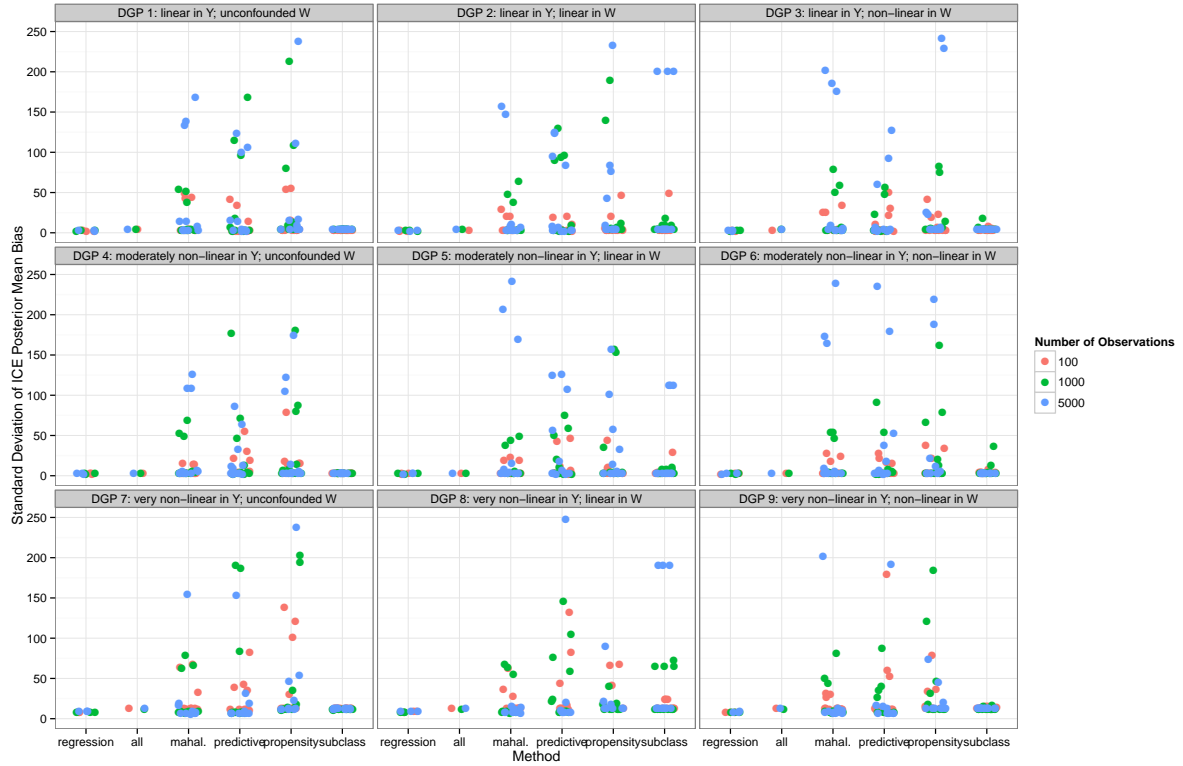
1 Comparing Methods for Continuous Outcomes

Figure 1: Comparing Standard Deviations of ICE Posterior Mean Bias for the Different Matching Methods (continuous outcome)

Figure 2: Comparing ATE Expected Error Loss for the Different Matching Methods (continuous outcome)
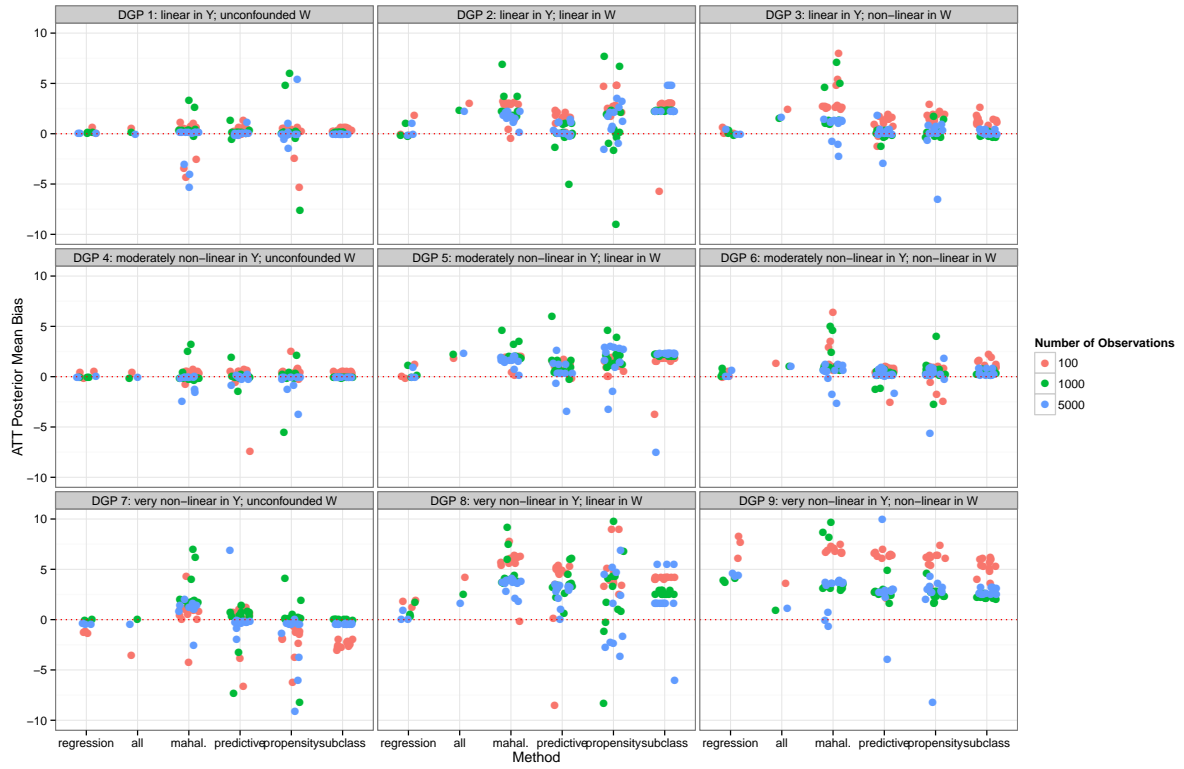
Figure 3: Comparing ATT Posterior Mean Bias for the Different Matching Methods (continuous outcome)
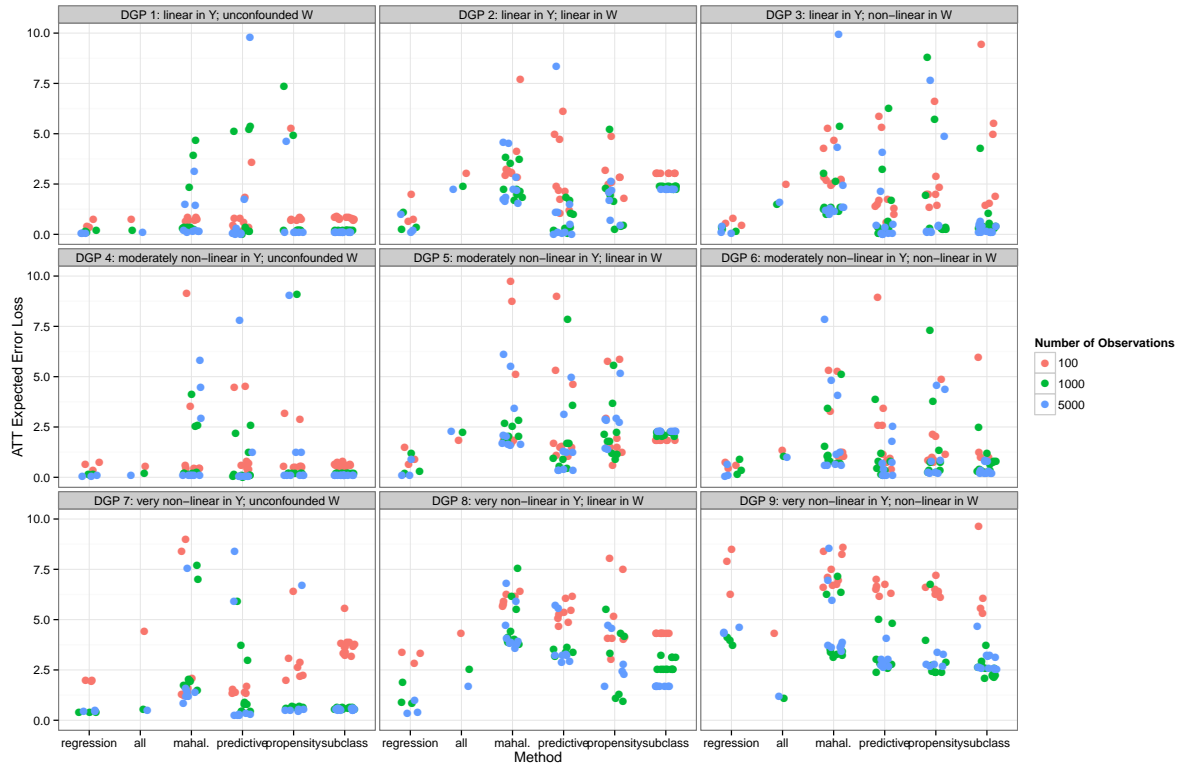
Figure 4: Comparing ATT Expected Error Loss for the Different Matching Methods (continuous outcome)
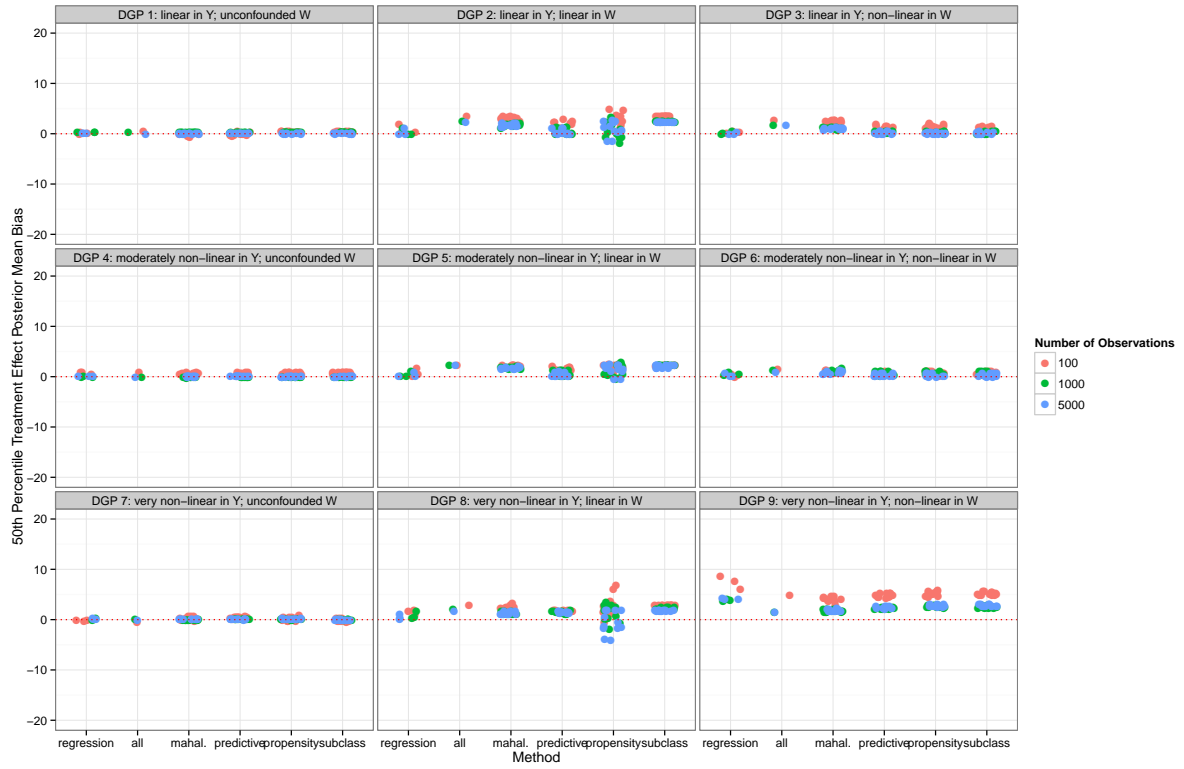
Figure 5: Comparing 50th Percentile Treatment Effect Posterior Mean Bias for the Different Matching Methods (continuous outcome)
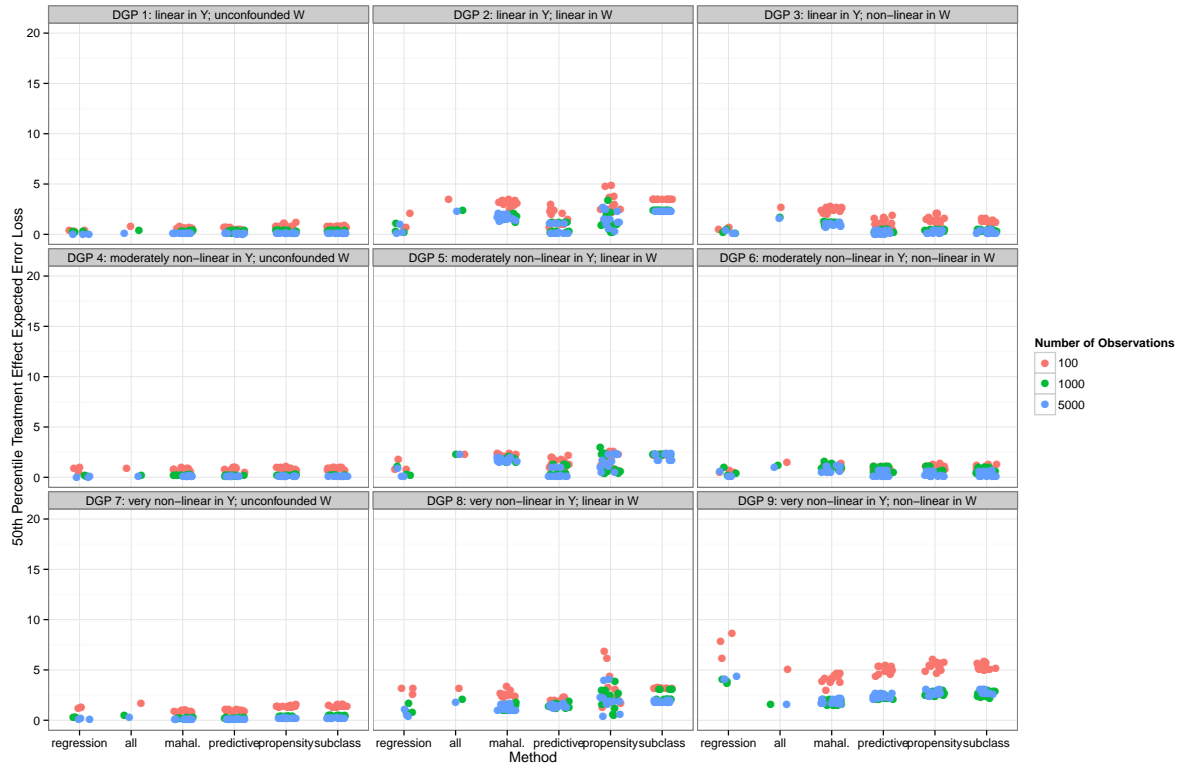
Figure 6: Comparing 50th Percentile Treatment Effect Expected Error Loss for the Different Matching Methods (continuous outcome)
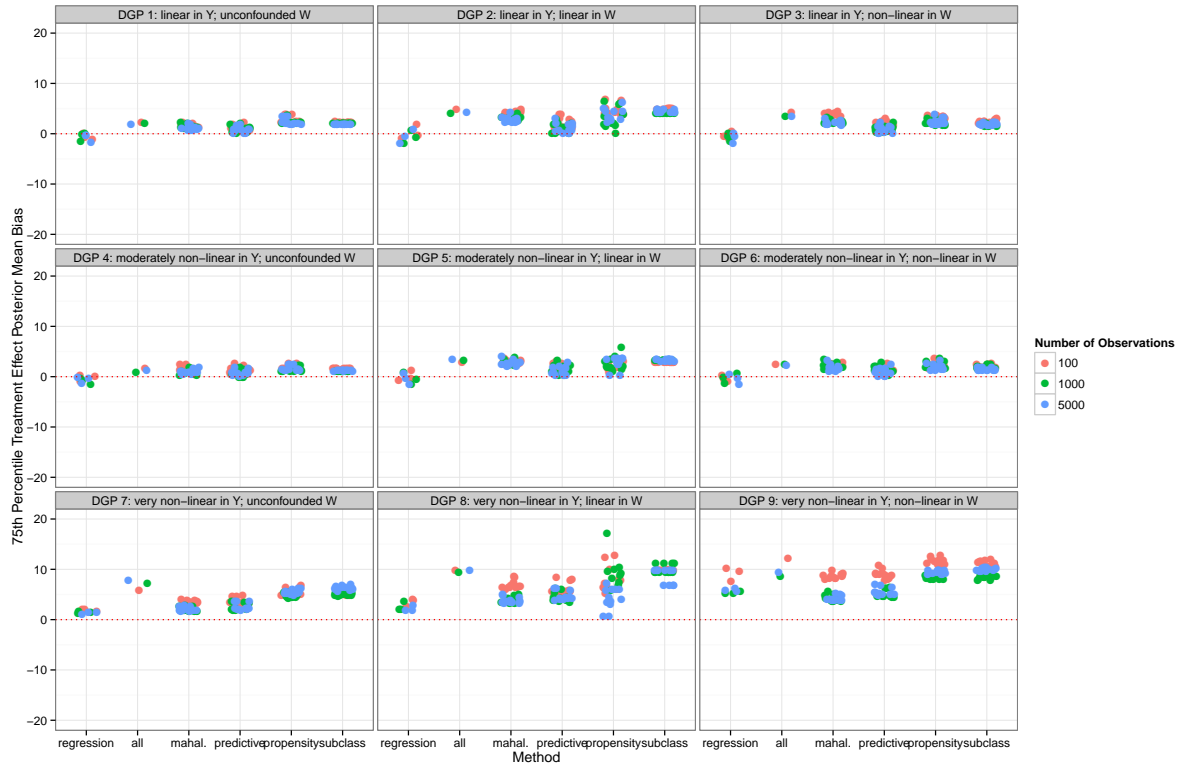
Figure 7: Comparing 75th Percentile Treatment Effect Posterior Mean Bias for the Different Matching Methods (continuous outcome)

Figure 8: Comparing 75th Percentile Treatment Effect Expected Error Loss for the Different Matching Methods (continuous outcome)

Figure 9: Comparing 95th Percentile Treatment Effect Posterior Mean Bias for the Different Matching Methods (continuous outcome)
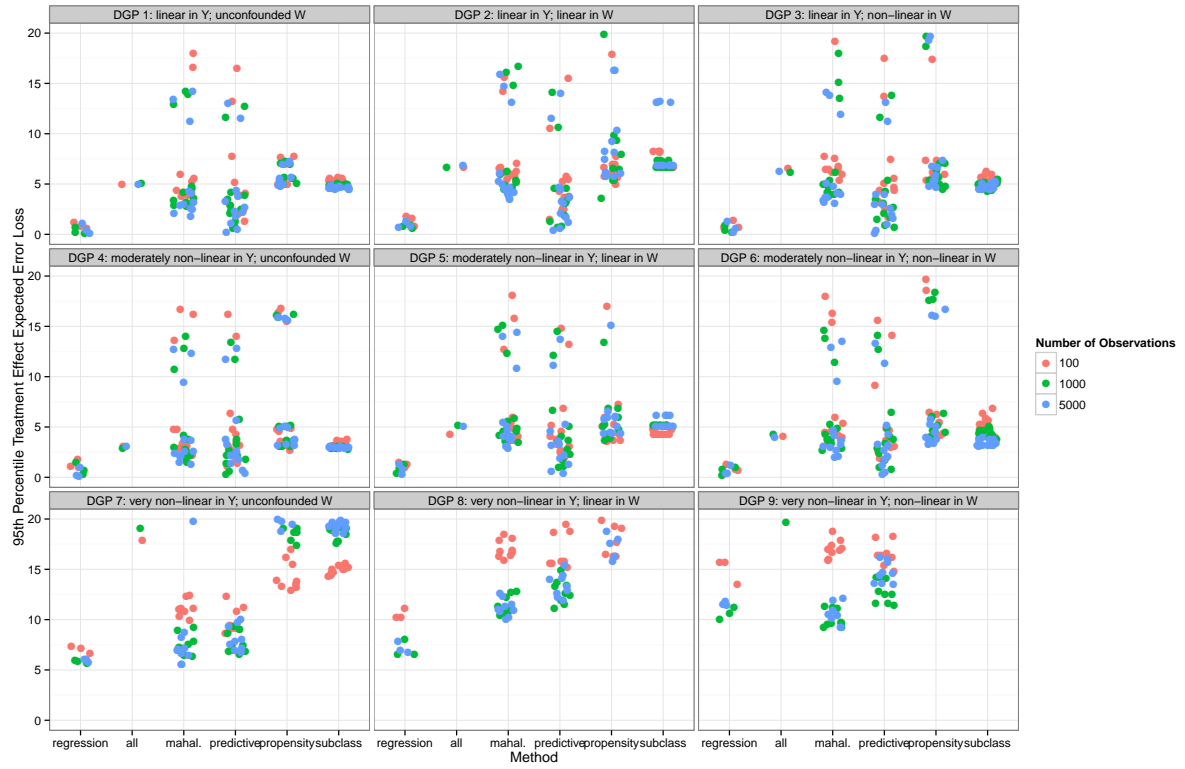
Figure 10: Comparing 95th Percentile Treatment Effect Expected Error Loss for the Different Matching Methods (continuous outcome)

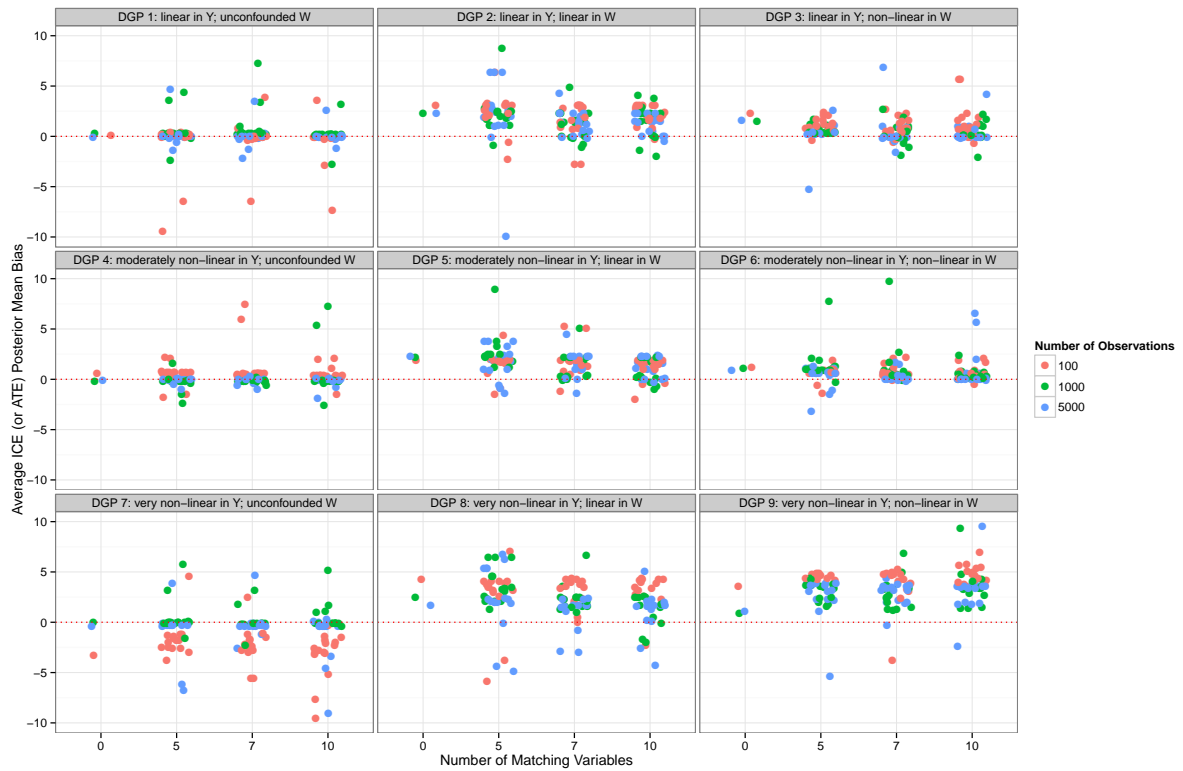## 2 Comparing Number of Conditioning Variables for Continuous Outcomes

Figure 11: Comparing Average ICE (or ATE) Posterior Mean Bias for Different Conditioning Sets (continuous outcome)

Figure 12: Comparing Standard Deviations of ICE Posterior Mean Bias for Different Conditioning Sets (continuous outcome)

Figure 13: Comparing ICE "Power" for Different Conditioning Sets (continuous outcome)
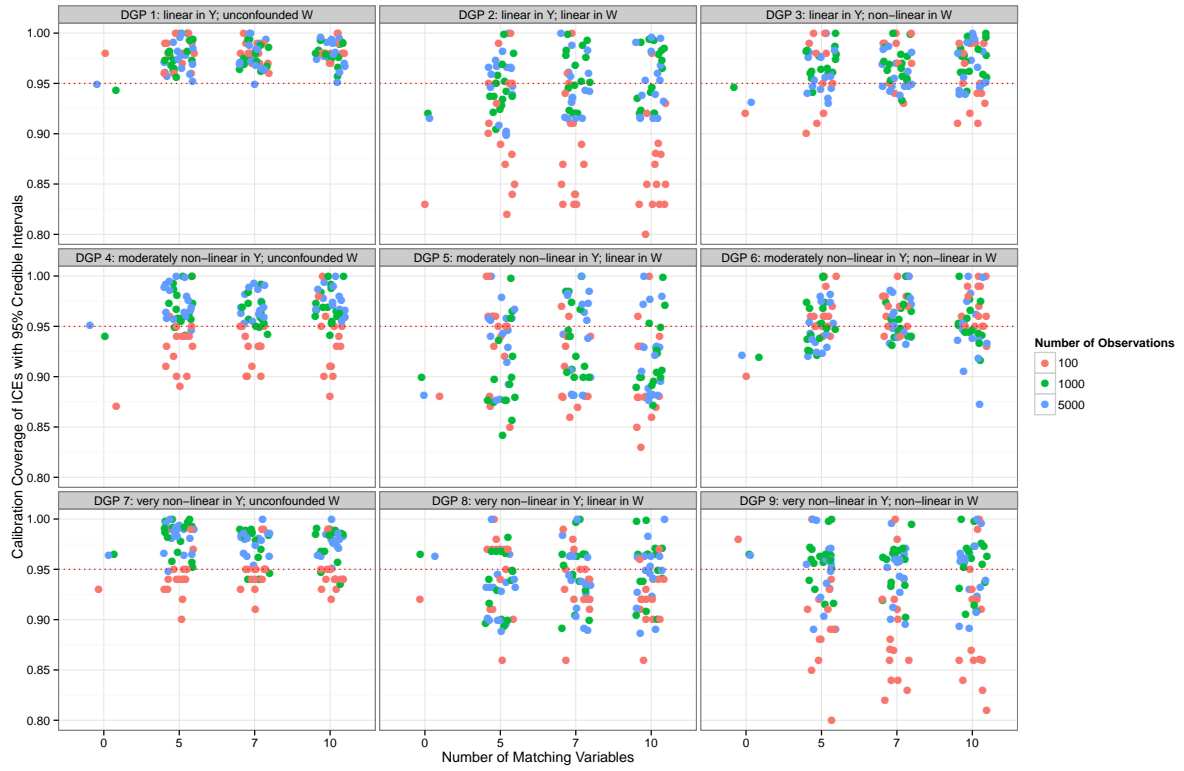
Figure 14: Comparing ICE Calibration Coverage for Different Conditioning Sets (continuous outcome)
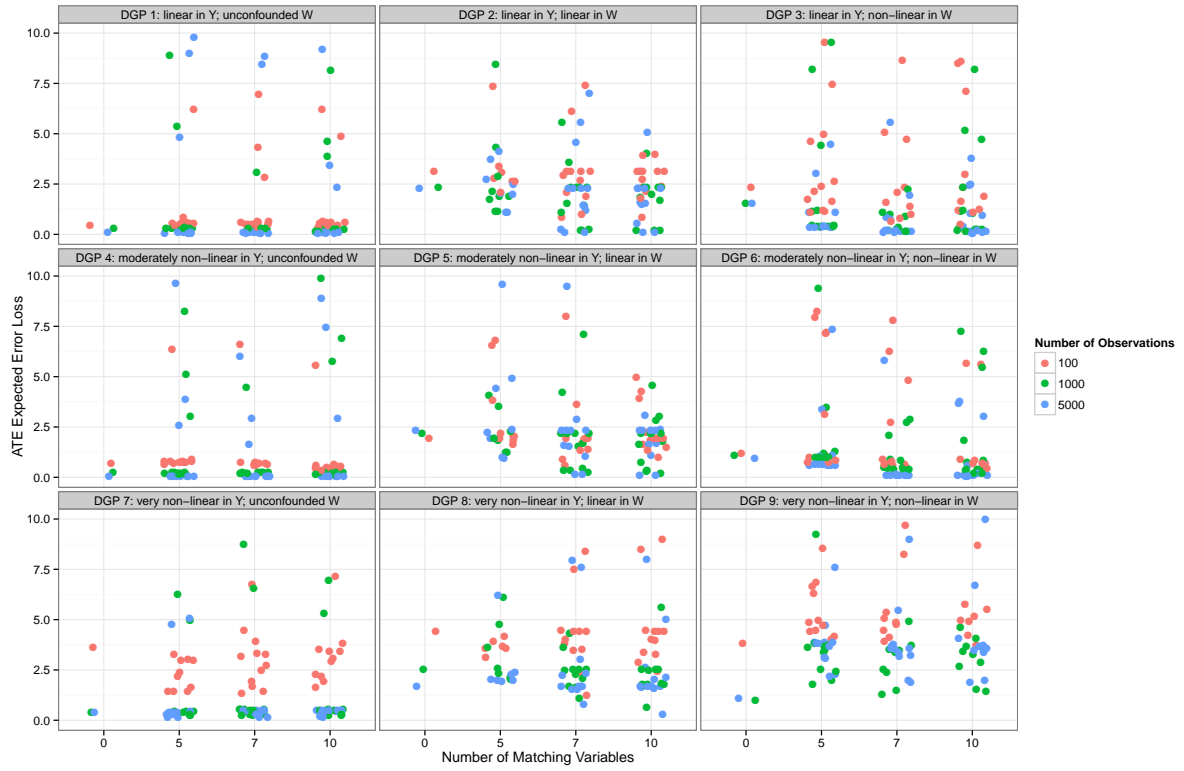
Figure 15: Comparing ATE Expected Error Loss for Different Conditioning Sets (continuous outcome)
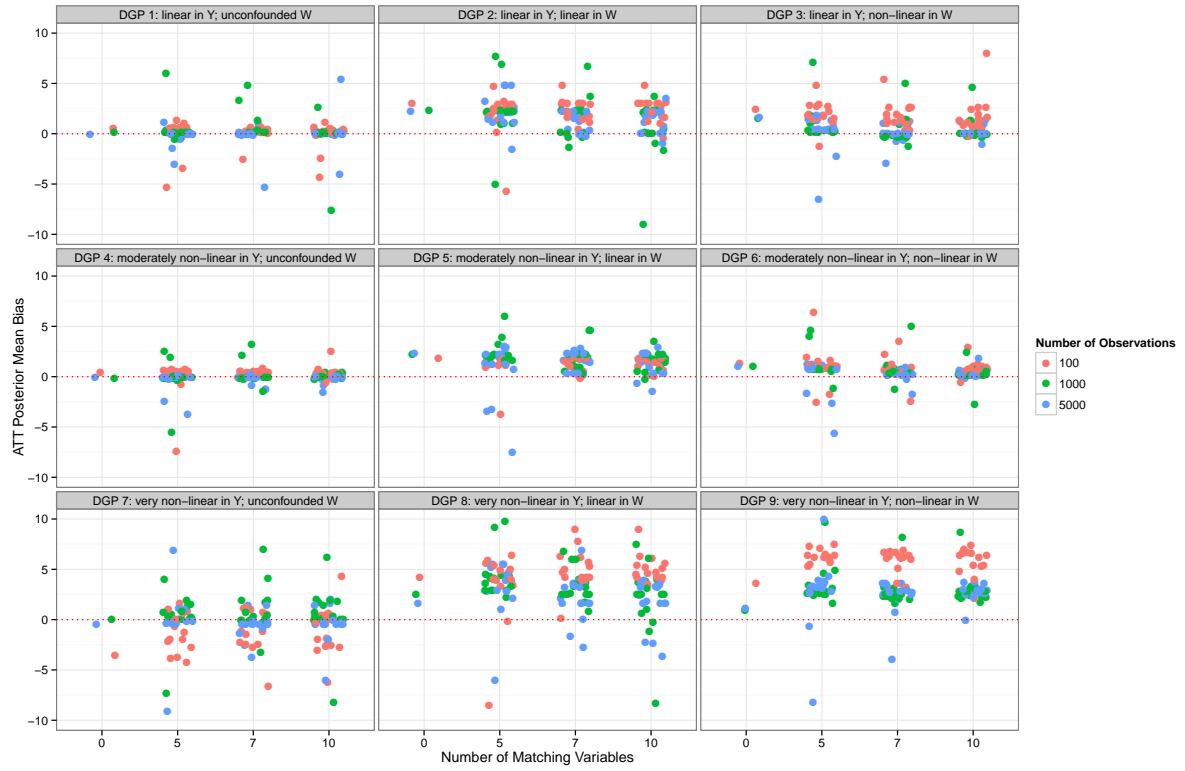
Figure 16: Comparing ATT Posterior Mean Bias for Different Conditioning Sets (continuous outcome)
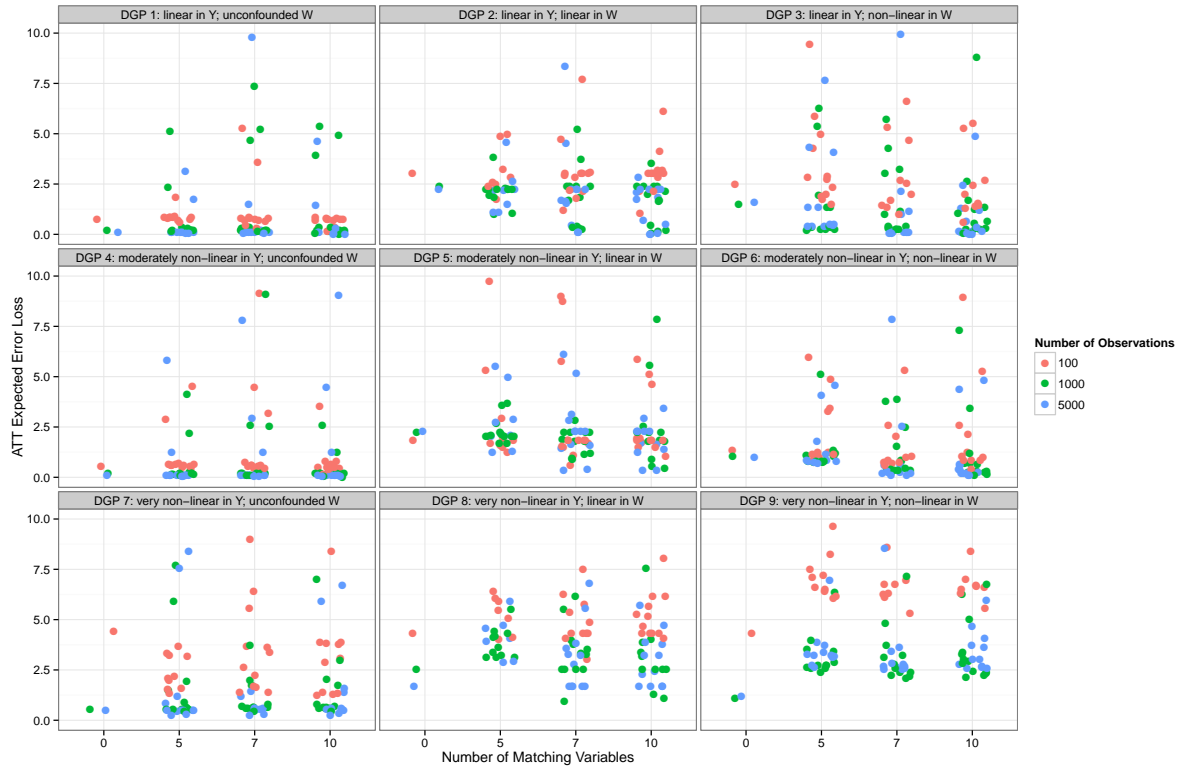
Figure 17: Comparing ATT Expected Error Loss for Different Conditioning Sets (continuous outcome)
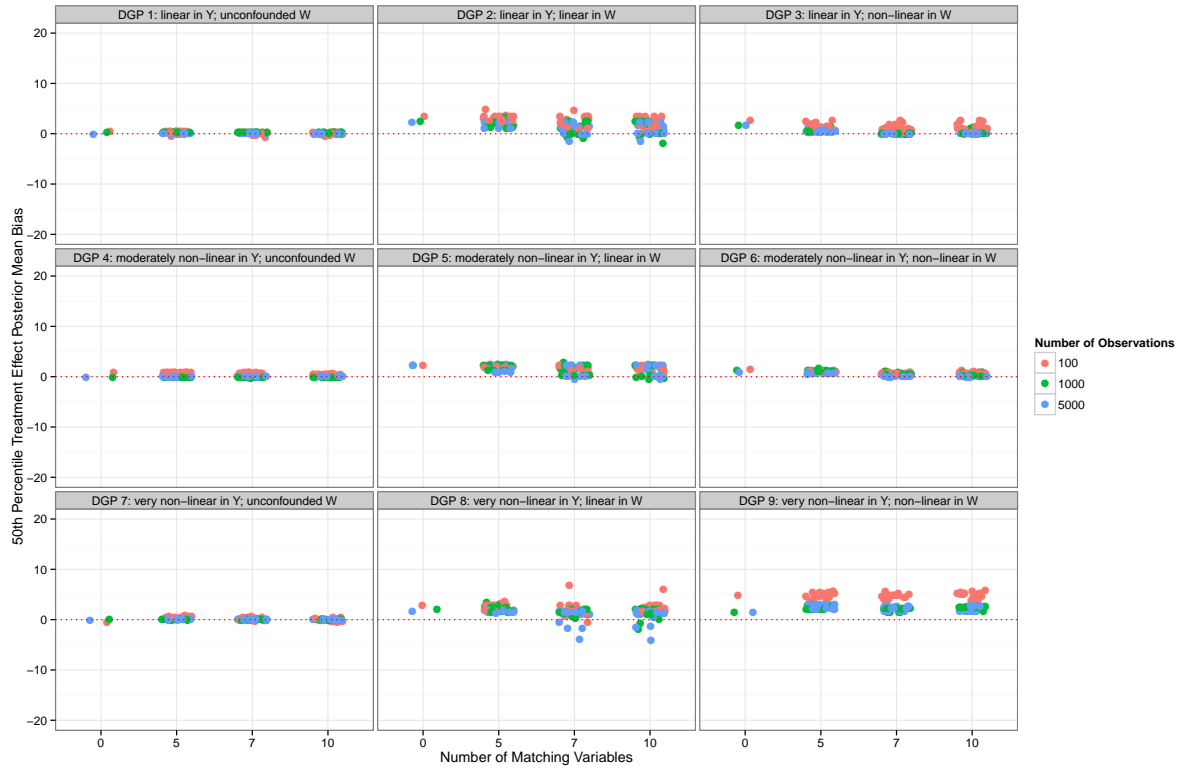
Figure 18: Comparing 50th Percentile Treatment Effect Posterior Mean Bias for Different Conditioning Sets (continuous outcome)

Figure 19: Comparing 50th Percentile Treatment Effect Expected Error Loss for Different Conditioning Sets (continuous outcome)

Figure 20: Comparing 75th Percentile Treatment Effect Posterior Mean Bias for Different Conditioning Sets (continuous outcome)
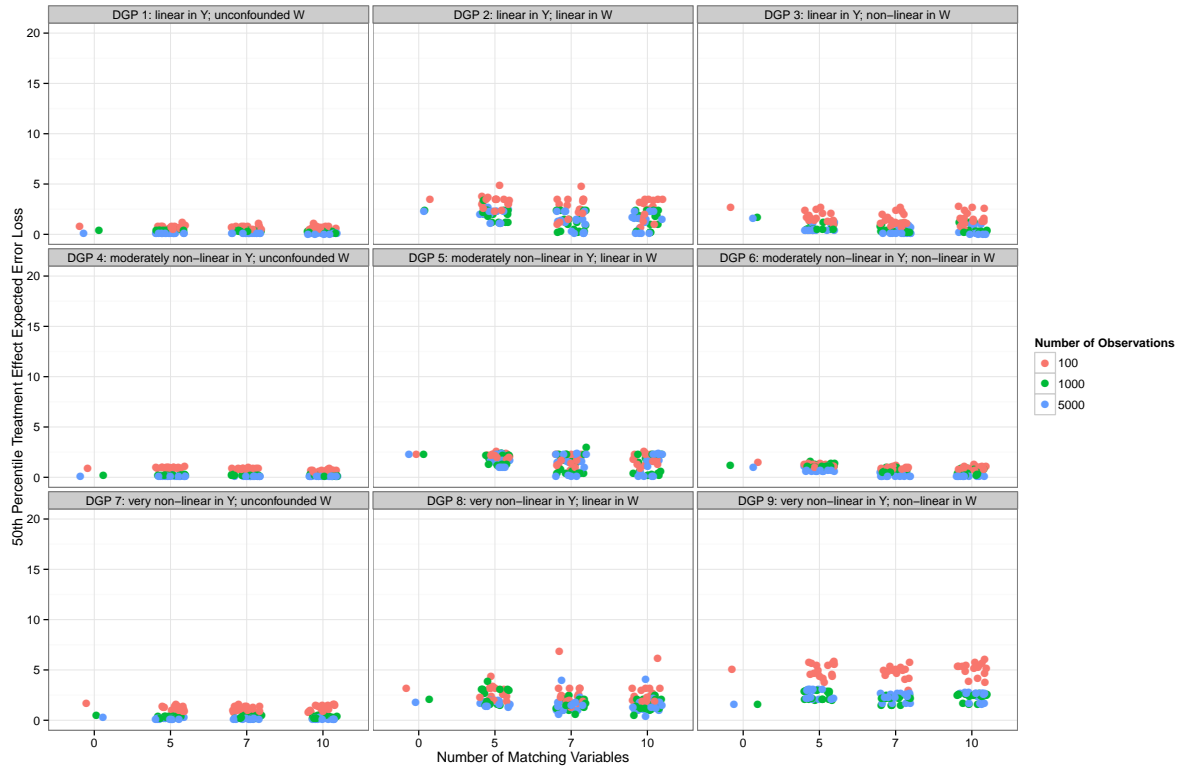
Figure 21: Comparing 75th Percentile Treatment Effect Expected Error Loss for Different Conditioning Sets (continuous outcome)
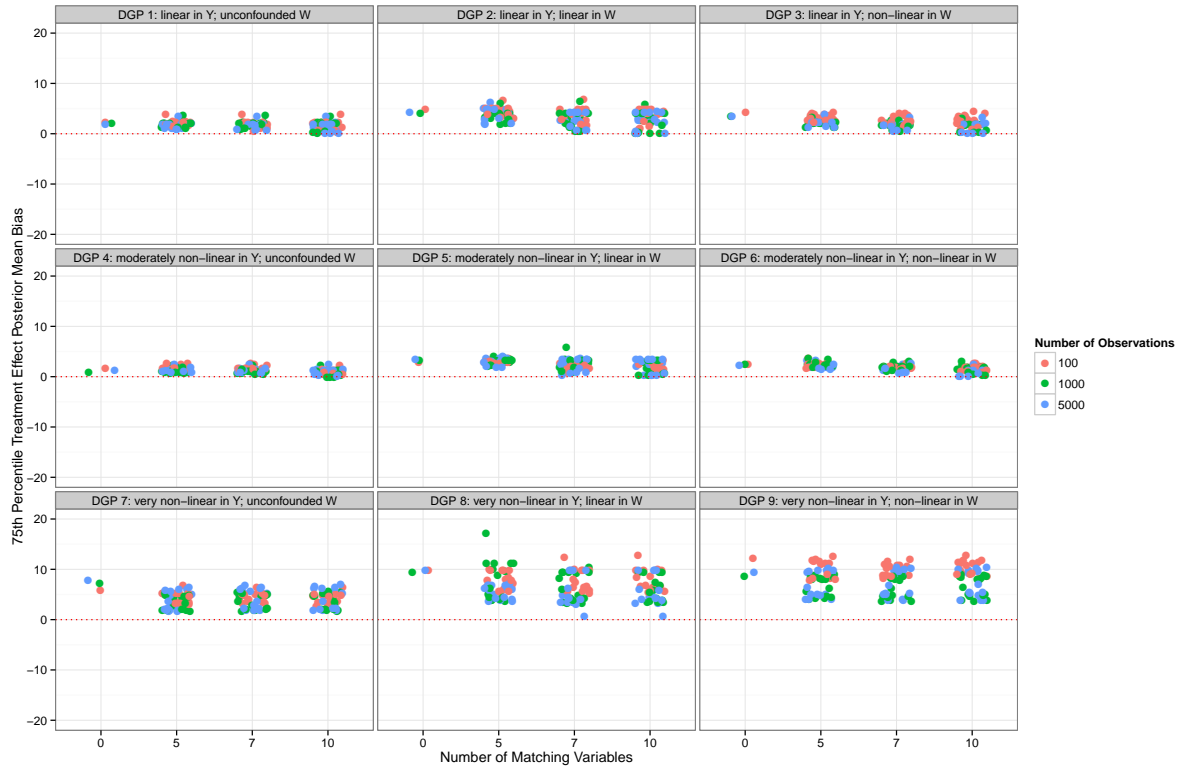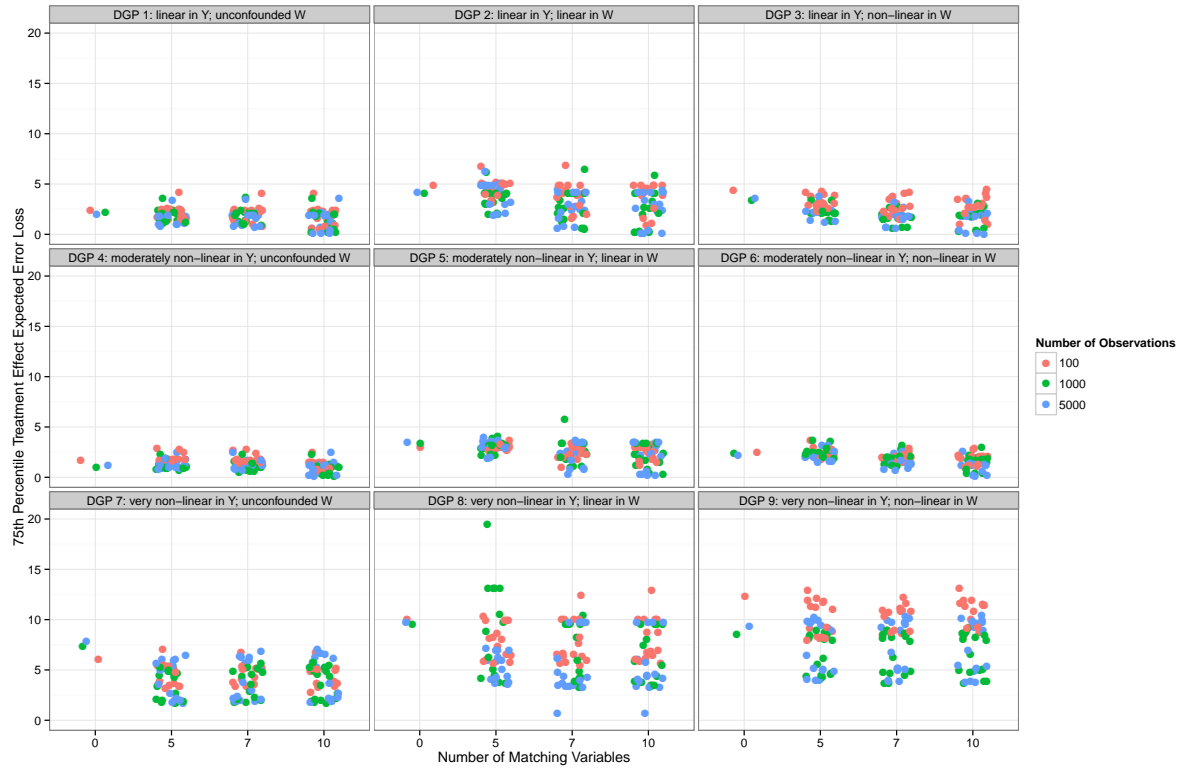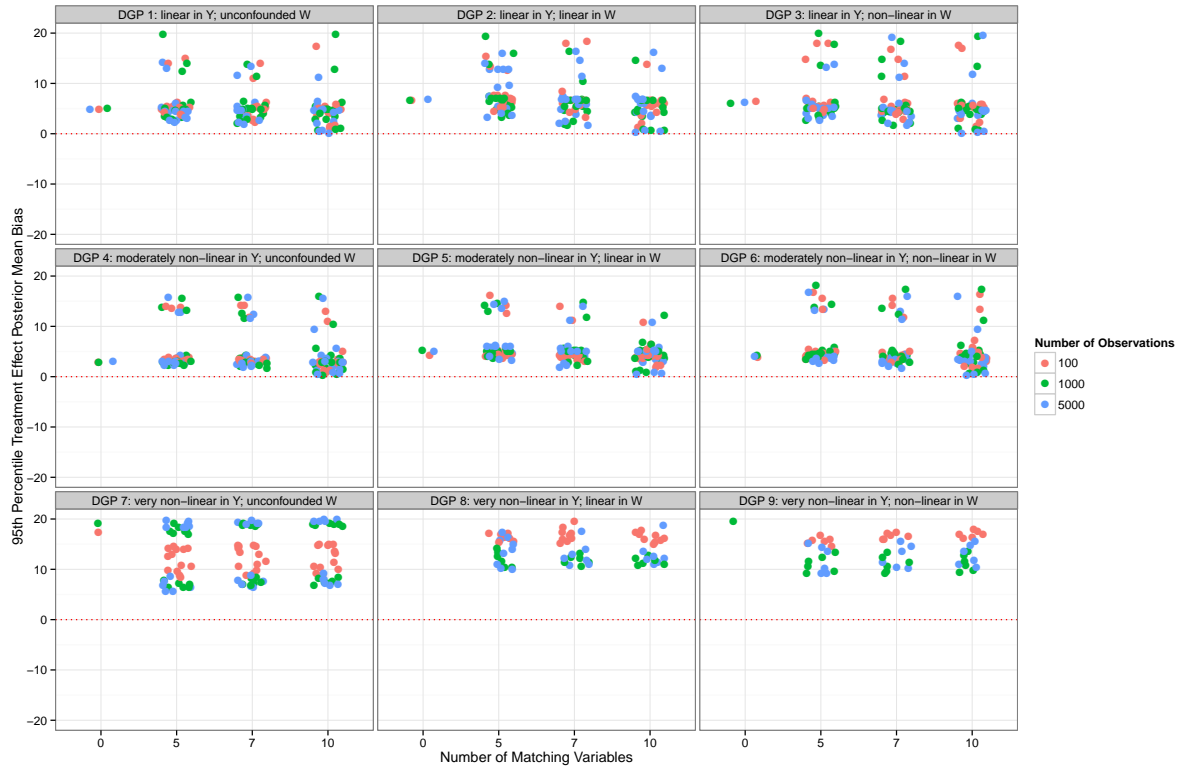
Figure 22: Comparing 95th Percentile Treatment Effect Posterior Mean Bias for Different Conditioning Sets (continuous outcome)
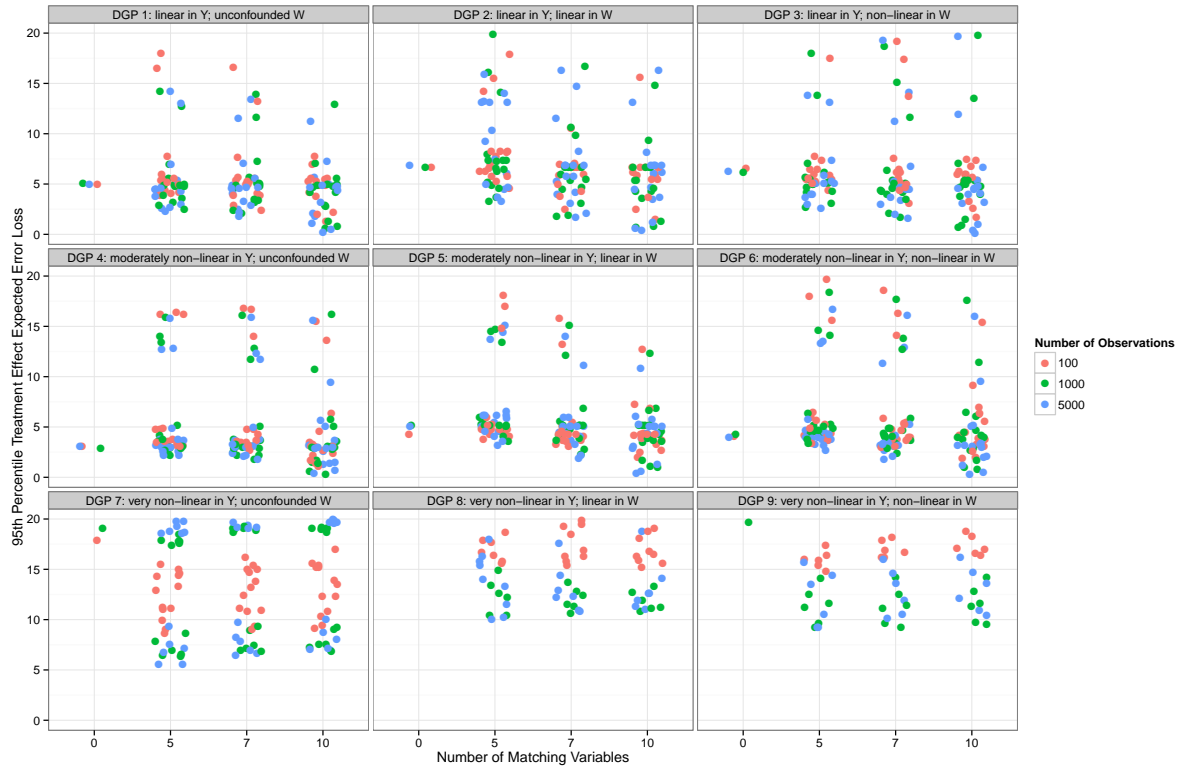
Figure 23: Comparing 95th Percentile Treatment Effect Expected Error Loss for Different Conditioning Sets (continuous outcome)

3 Comparing Number of Matches for Continuous Outcomes

Figure 24: Comparing Standard Deviations of ICE Posterior Mean Bias for Different Numbers of Matches (continuous outcome)

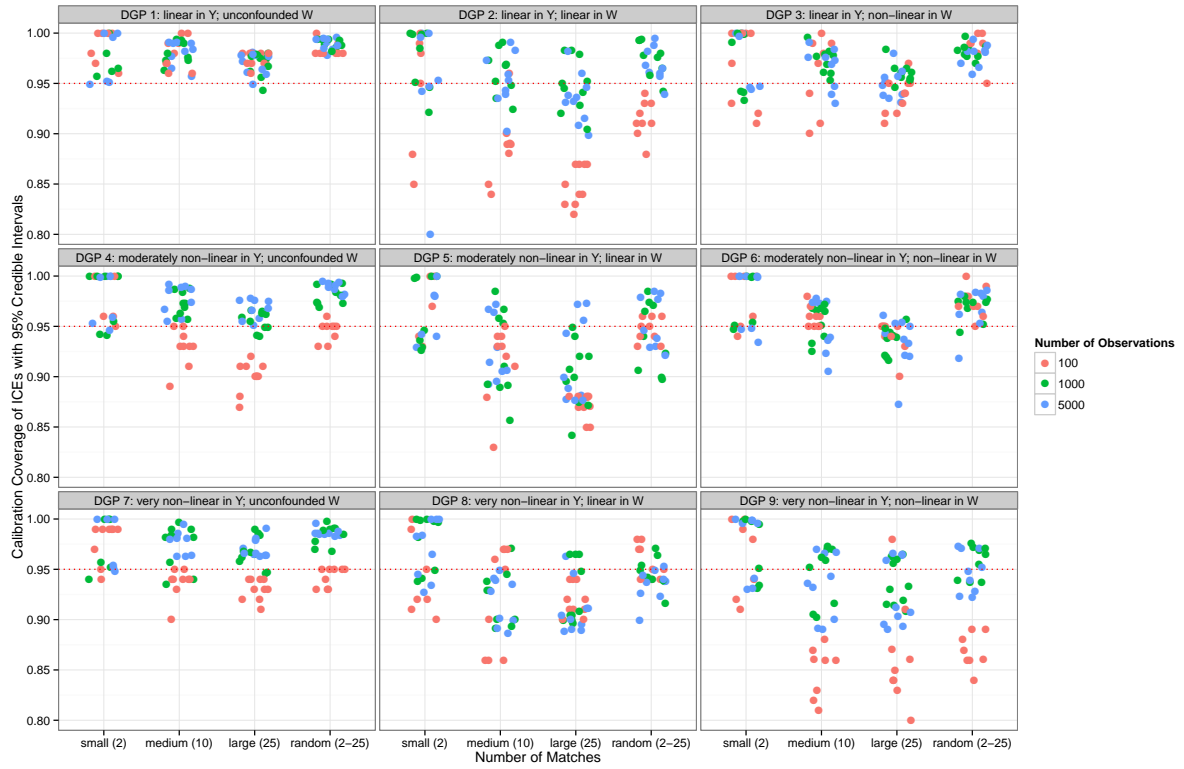Figure 25: Comparing ICE "Power" for Different Numbers of Matches (continuous outcome)

Figure 26: Comparing ICE Calibration Coverage for Different Numbers of Matches (continuous outcome)
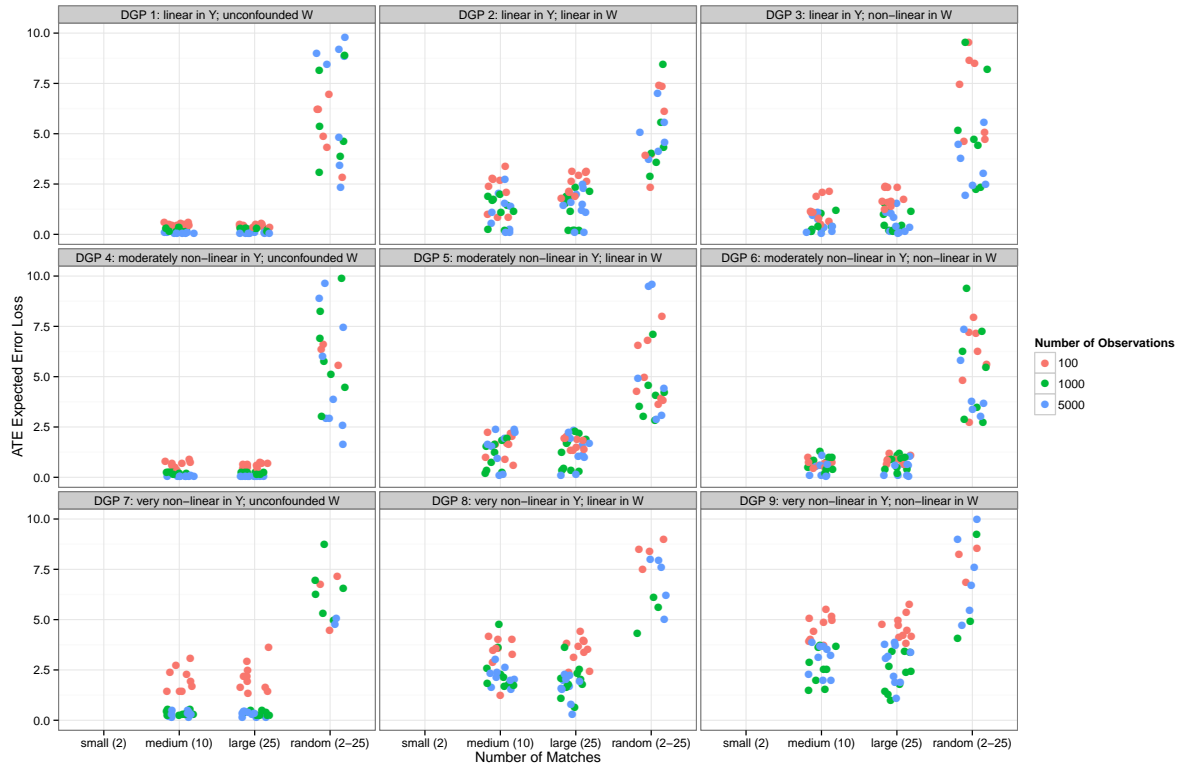
Figure 27: Comparing ATE Expected Error Loss for Different Numbers of Matches (continuous outcome)
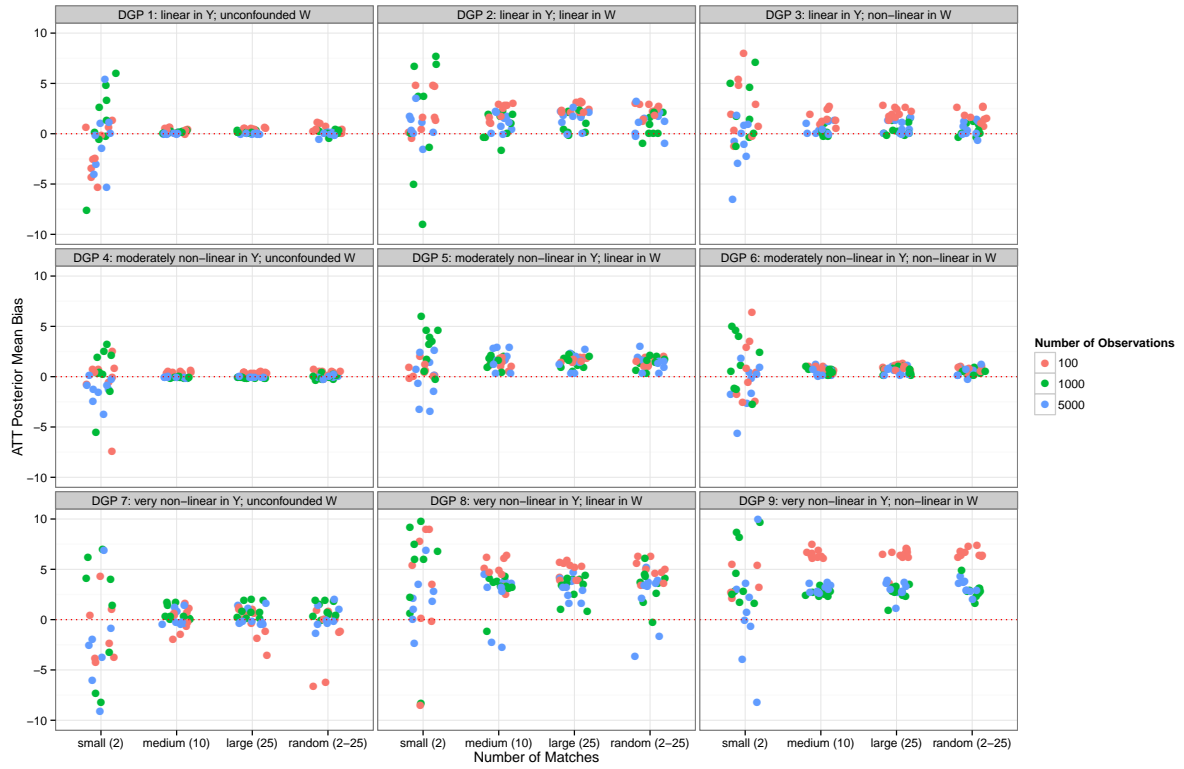
Figure 28: Comparing ATT Posterior Mean Bias for Different Numbers of Matches (continuous outcome)
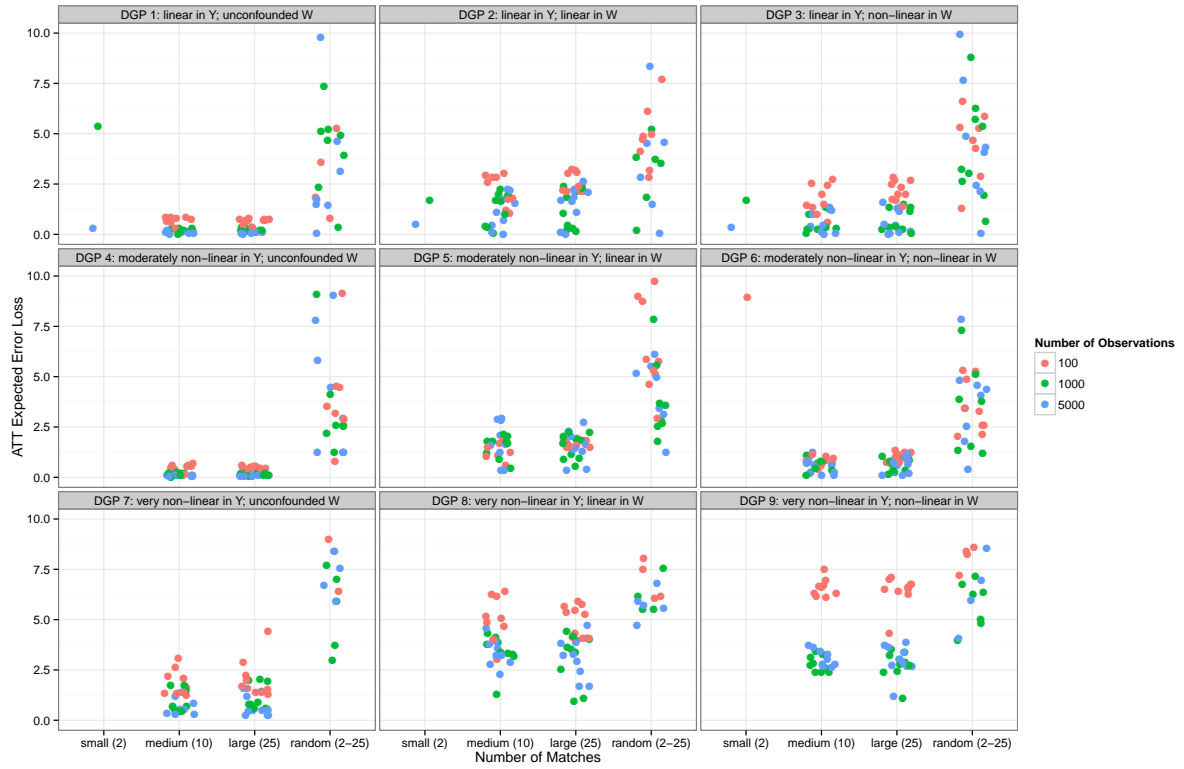
Figure 29: Comparing ATT Expected Error Loss for Different Numbers of Matches (continuous outcome)

Figure 30: Comparing 50th Percentile Treatment Effect Posterior Mean Bias for Different Numbers of Matches (continuous outcome)

Figure 31: Comparing 50th Percentile Treatment Effect Expected Error Loss for Different Numbers of Matches (continuous outcome)
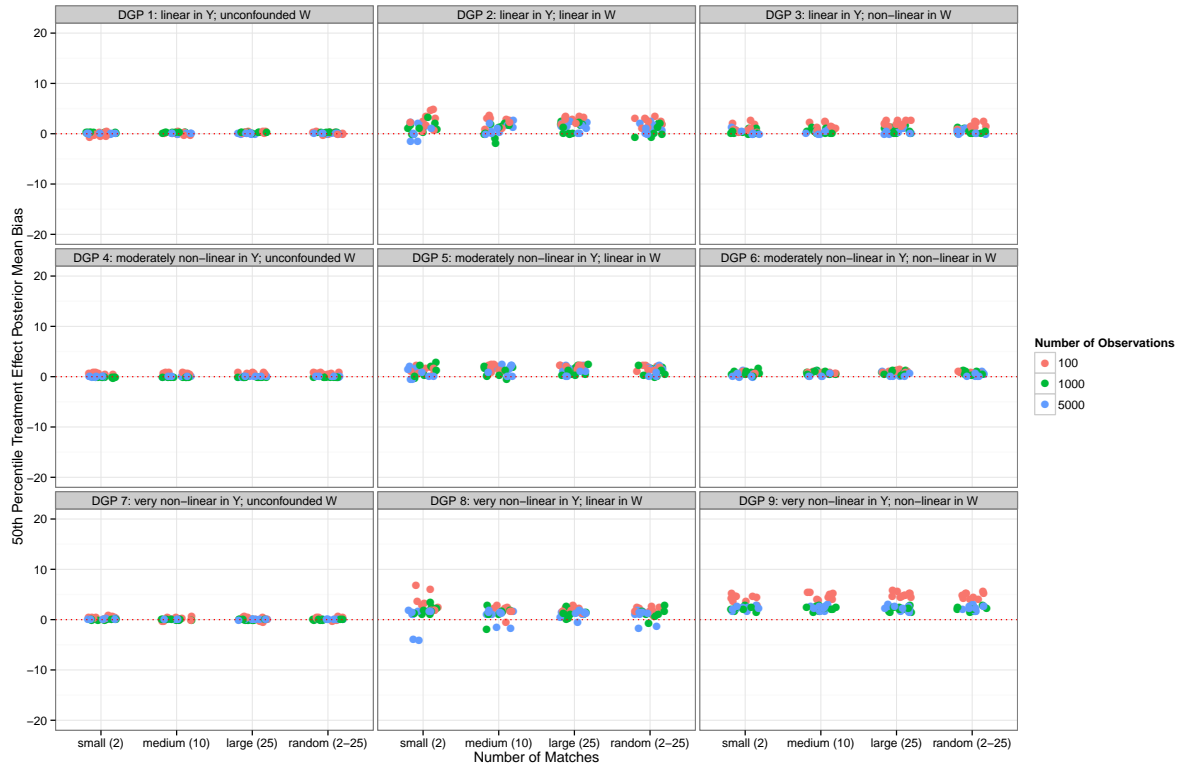
Figure 32: Comparing 75th Percentile Treatment Effect Posterior Mean Bias for Different Numbers of Matches (continuous outcome)

Figure 33: Comparing 75th Percentile Treatment Effect Expected Error Loss for Different Numbers of Matches (continuous outcome)

Figure 34: Comparing 95th Percentile Treatment Effect Posterior Mean Bias for Different Numbers of Matches (continuous outcome)
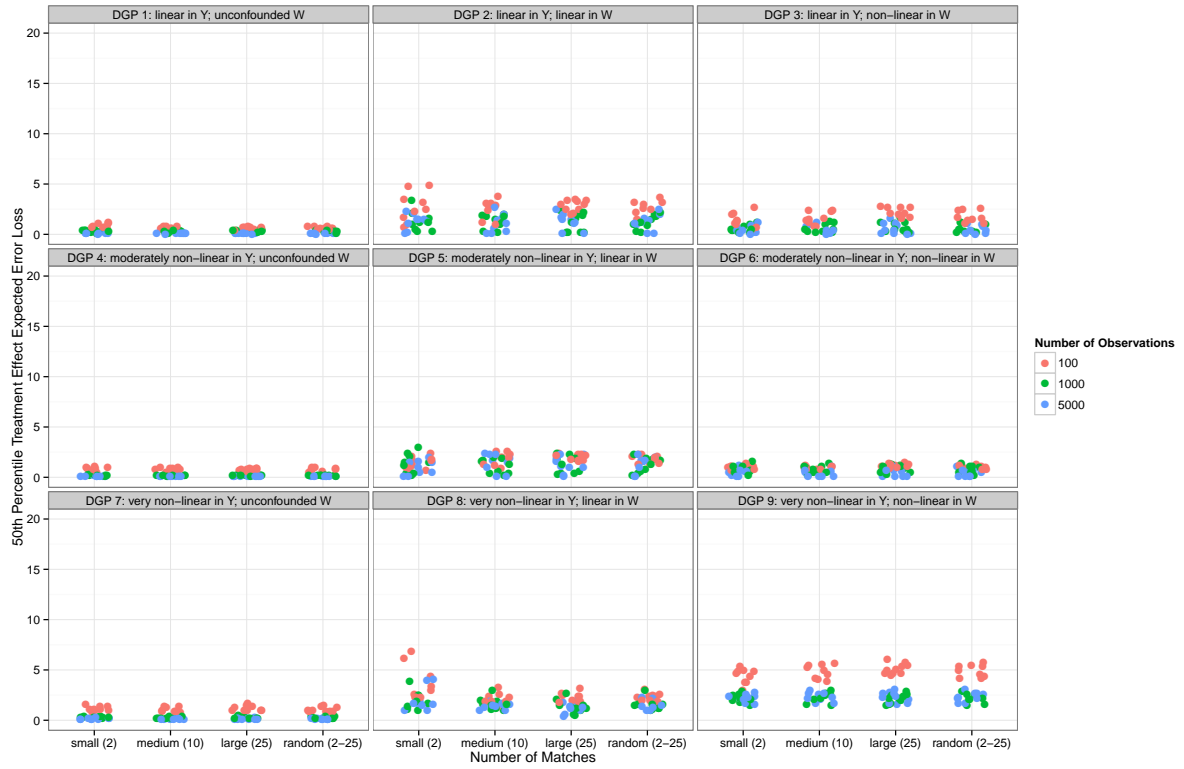
Figure 35: Comparing 95th Percentile Treatment Effect Expected Error Loss for Different Numbers of Matches (continuous outcome)
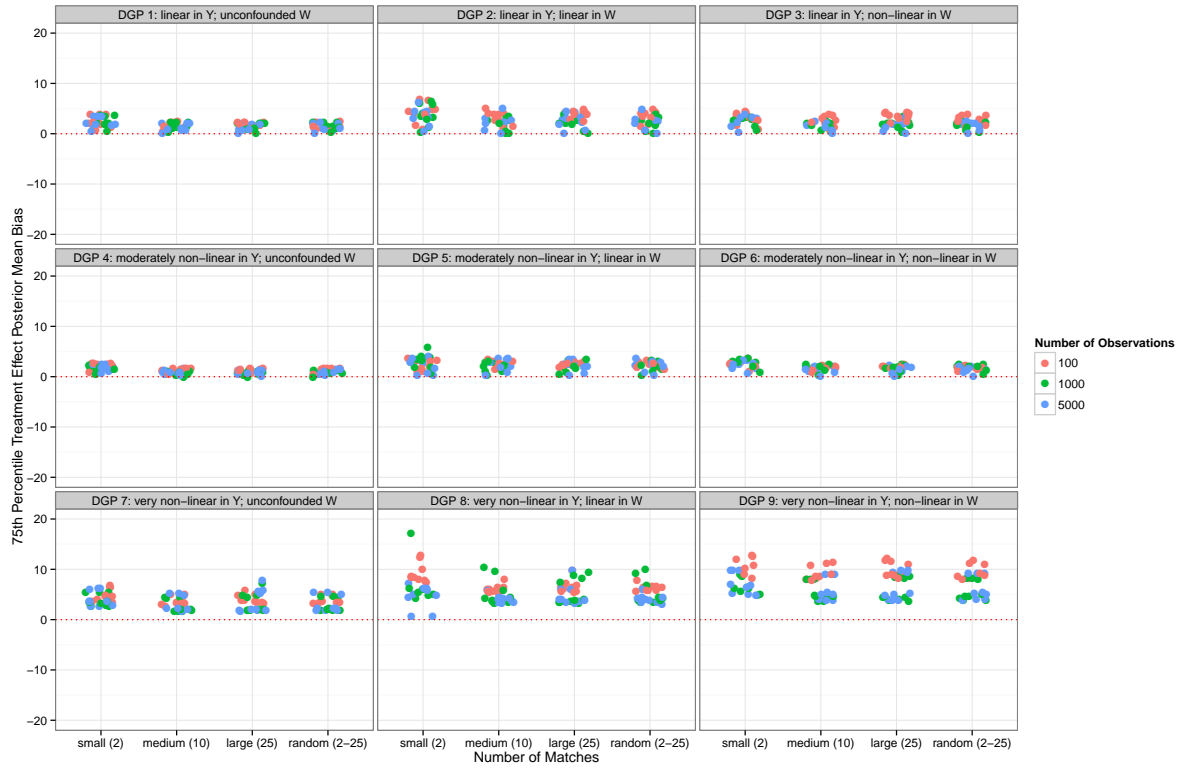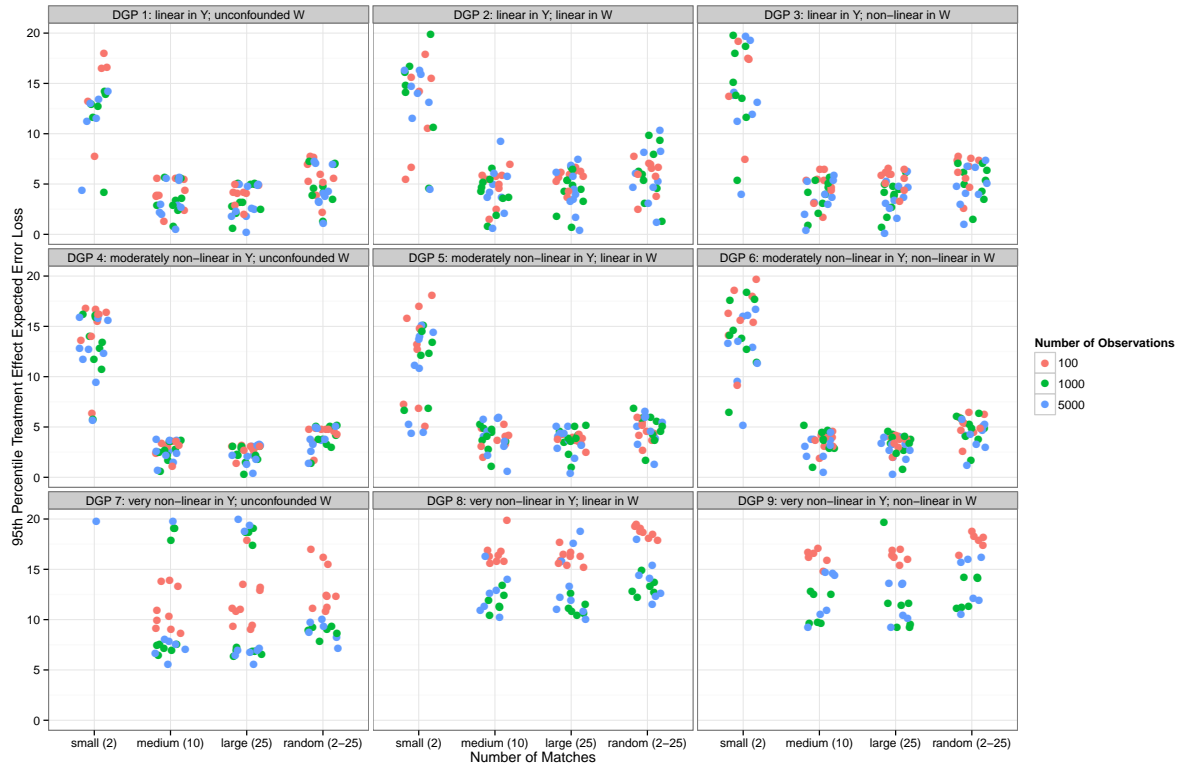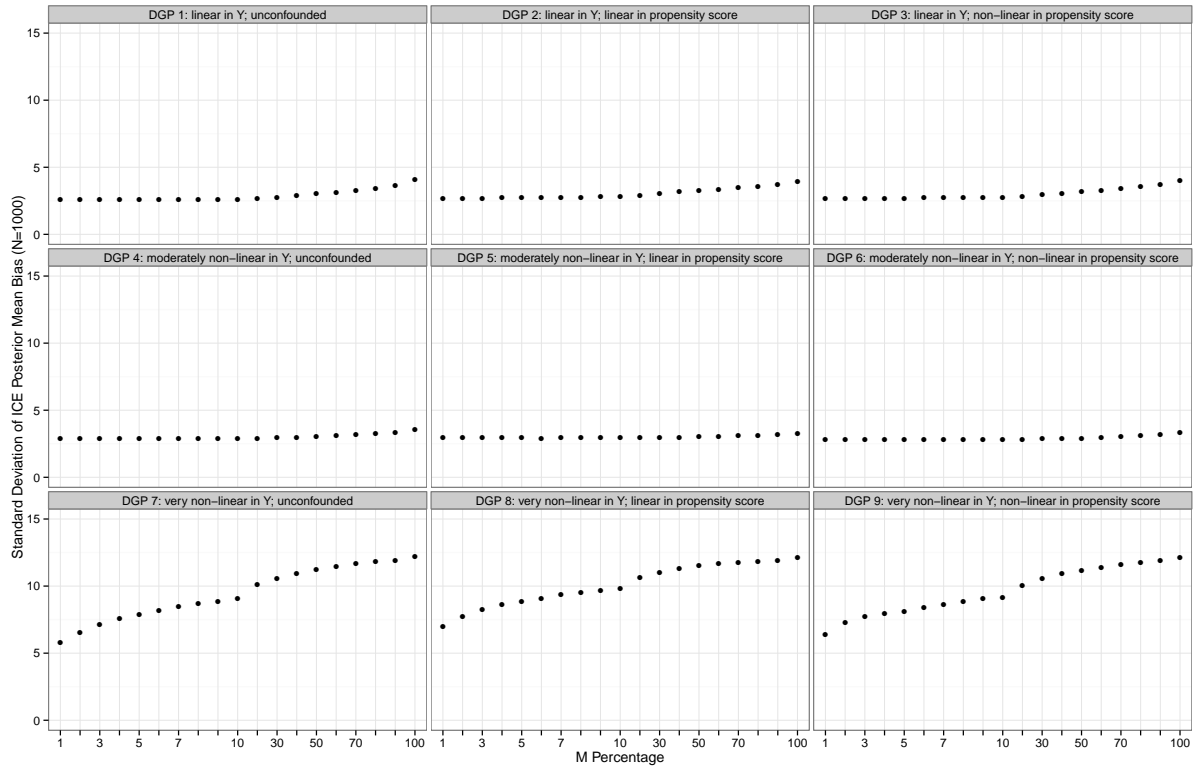
Figure 36: Comparing Standard Deviations of ICE Posterior Mean Bias for Different Match Percentages (continuous outcome)
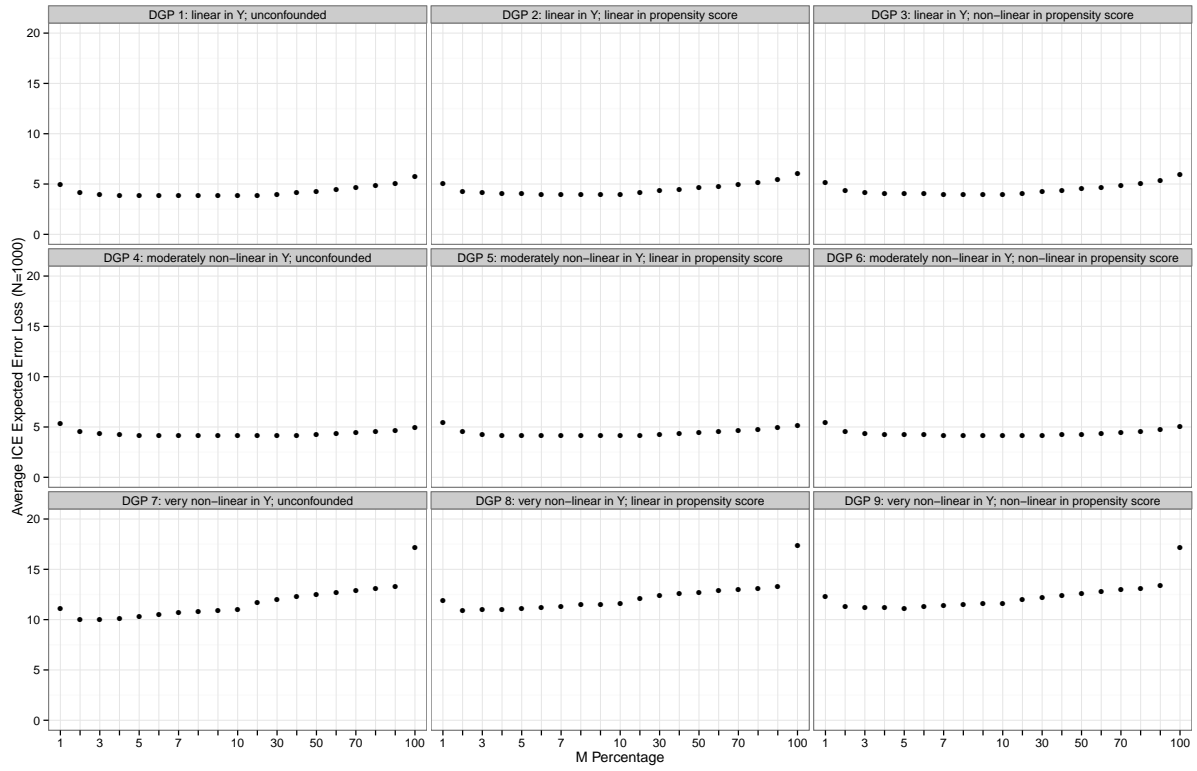
Figure 37: Comparing Average ICE Expected Error Loss for Different Match Percentages (continuous outcome)
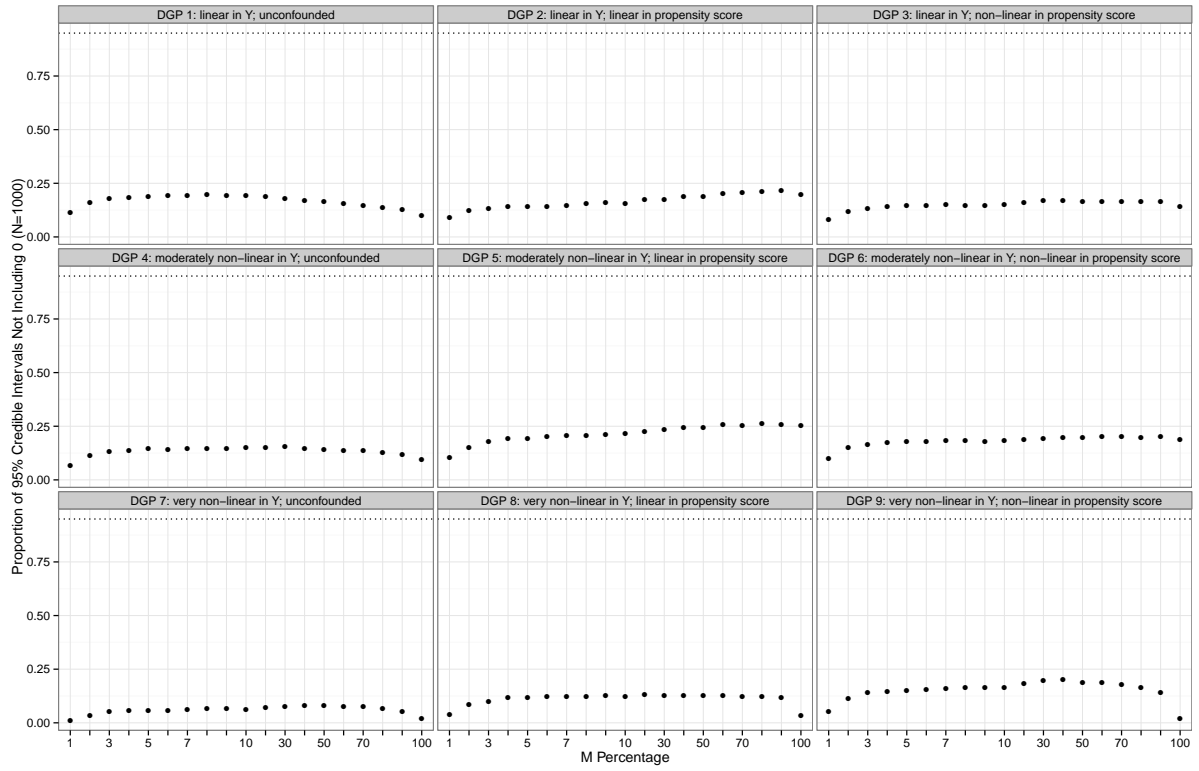
Figure 38: Comparing ICE "Power" for Different Match Percentages (continuous outcome)
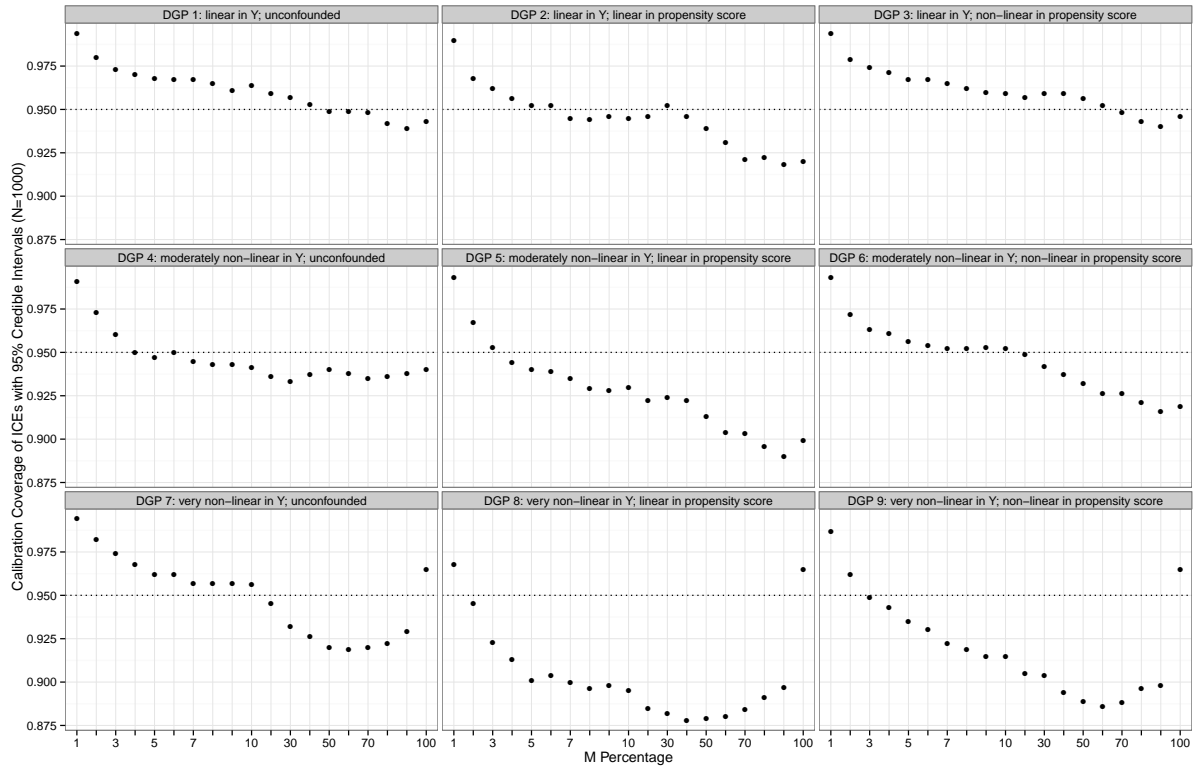
Figure 39: Comparing ICE Calibration Coverage for Different Match Percentages (continuous outcome)
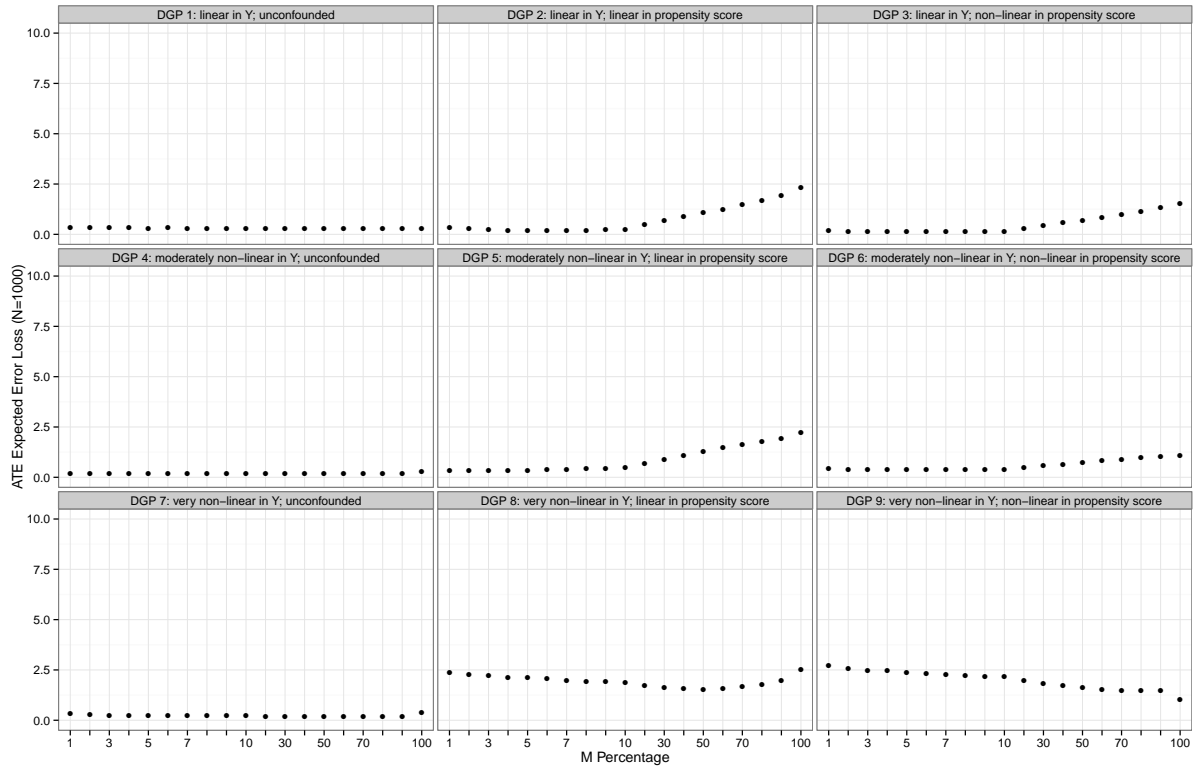
Figure 40: Comparing ATE Expected Error Loss for Different Match Percentages (continuous outcome)

Figure 41: Comparing ATT Posterior Mean Bias for Different Match Percentages (continuous outcome)

Figure 42: Comparing ATT Expected Error Loss for Different Match Percentages (continuous outcome)

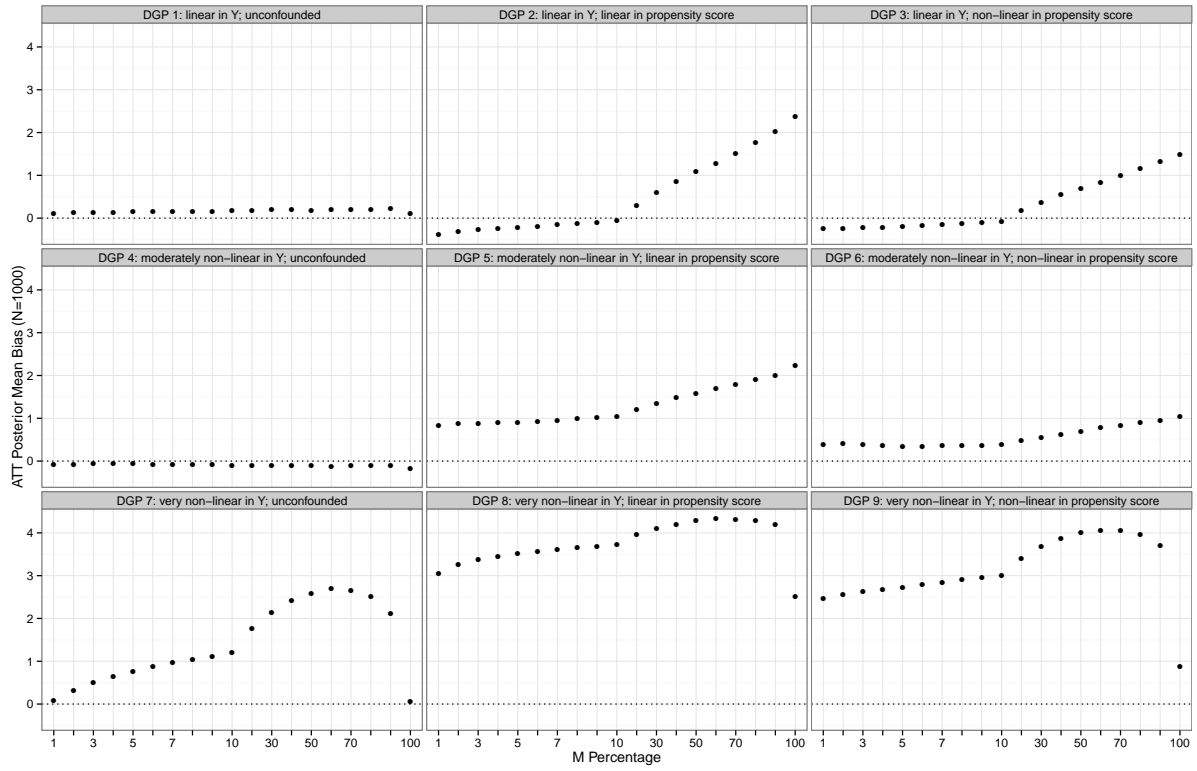Figure 43: Comparing 50th Percentile Treatment Effect Posterior Mean Bias for Different Match Percentages (continuous outcome)

Figure 44: Comparing 50th Percentile Treatment Effect Expected Error Loss for Different Match Percentages (continuous outcome)

Figure 45: Comparing 75th Percentile Treatment Effect Posterior Mean Bias for Different Match Percentages (continuous outcome)

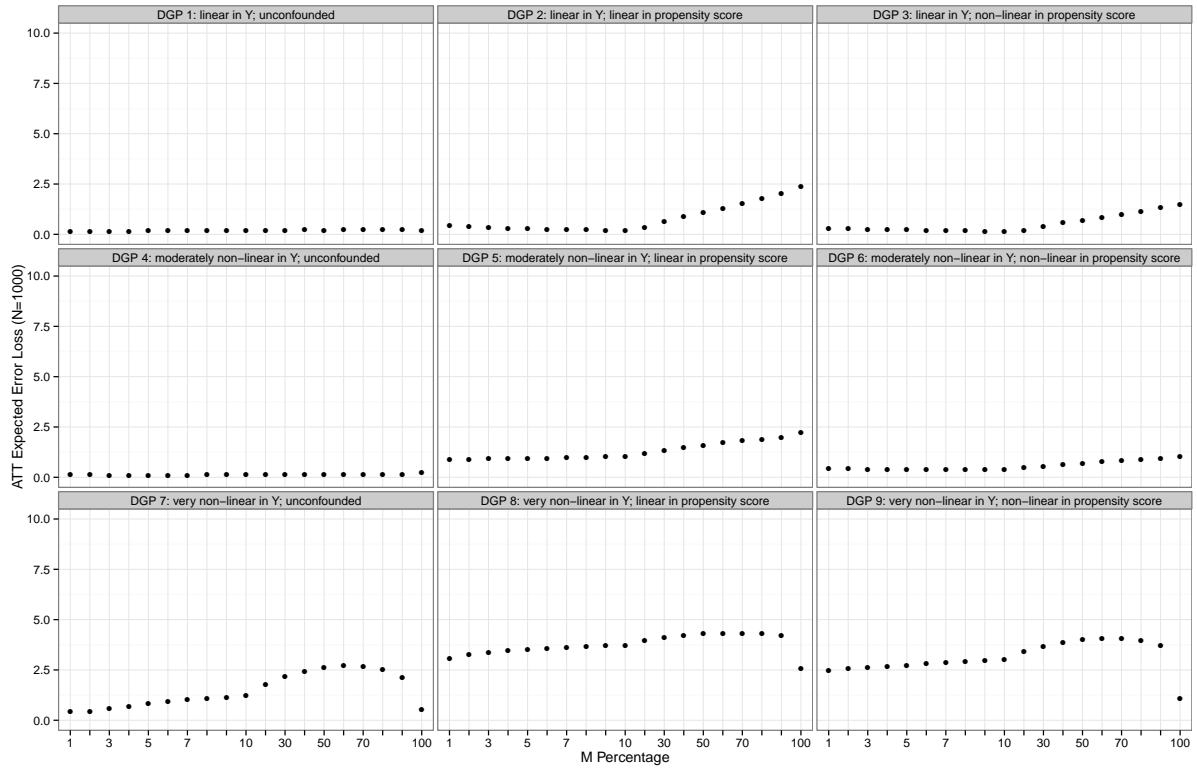Figure 46: Comparing 75th Percentile Treatment Effect Expected Error Loss for Different Match Percentages (continuous outcome)
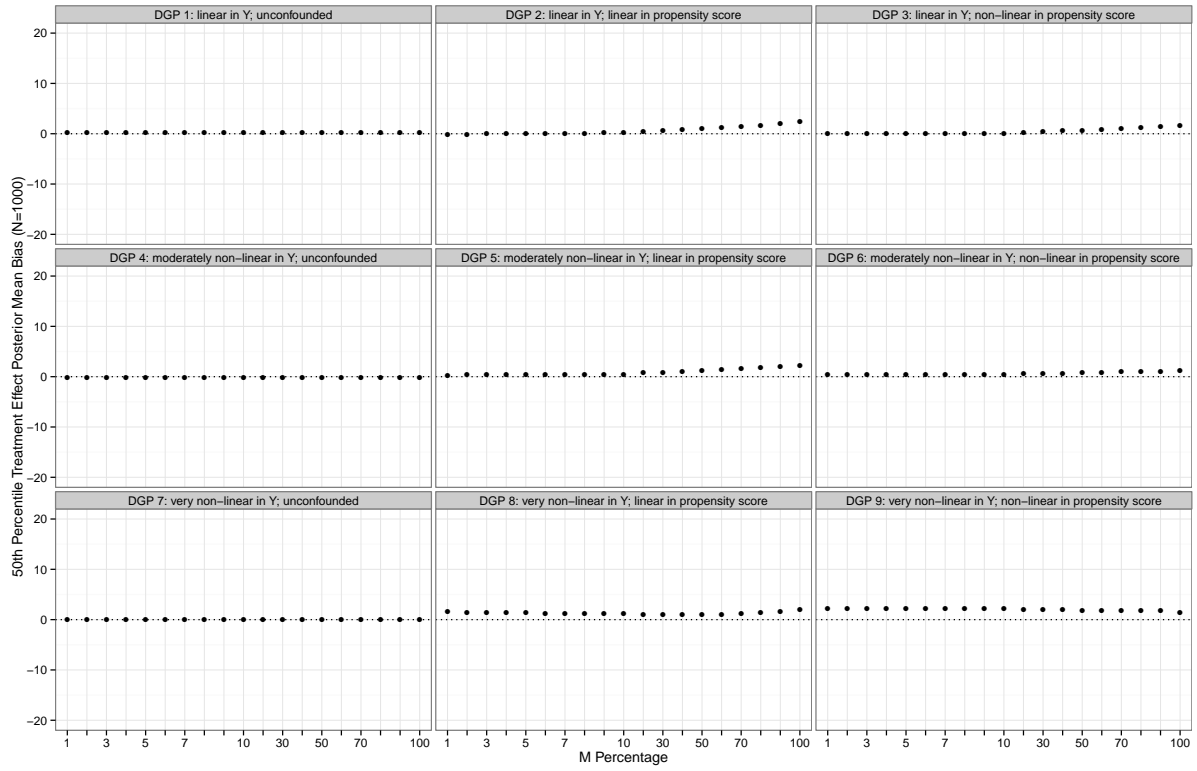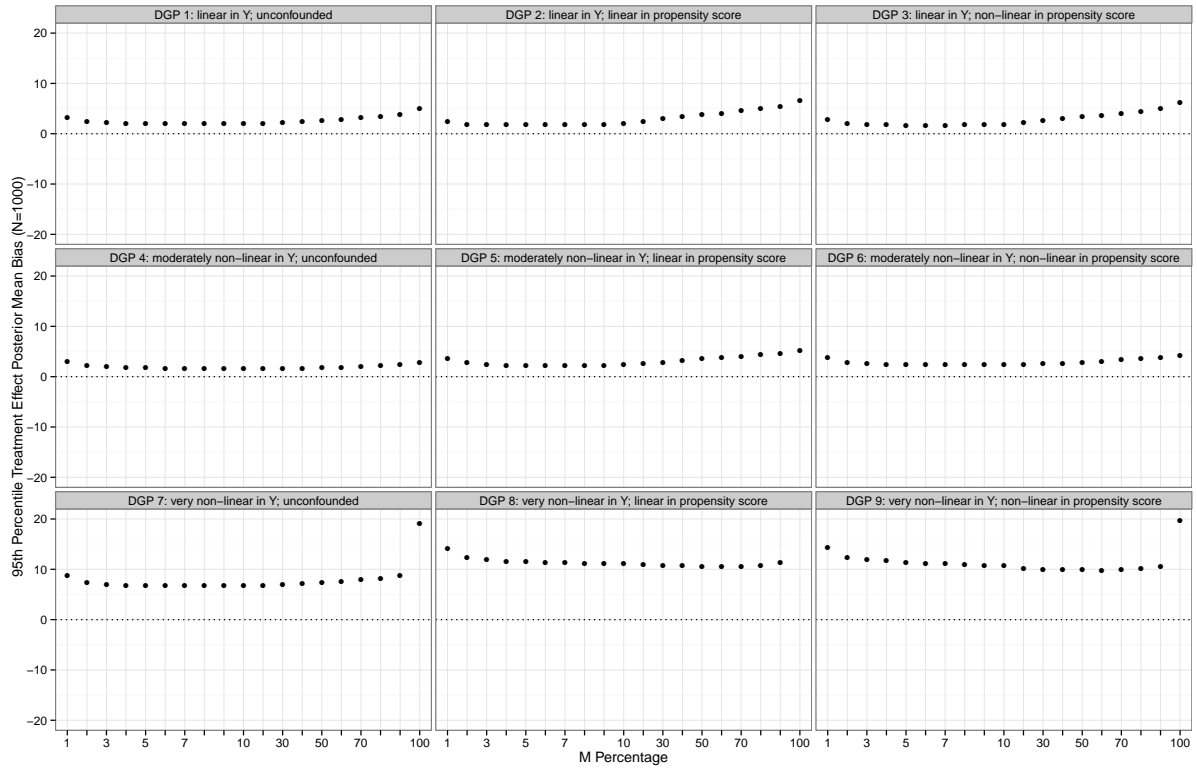
Figure 47: Comparing 95th Percentile Treatment Effect Posterior Mean Bias for Different Match Percentages (continuous outcome)
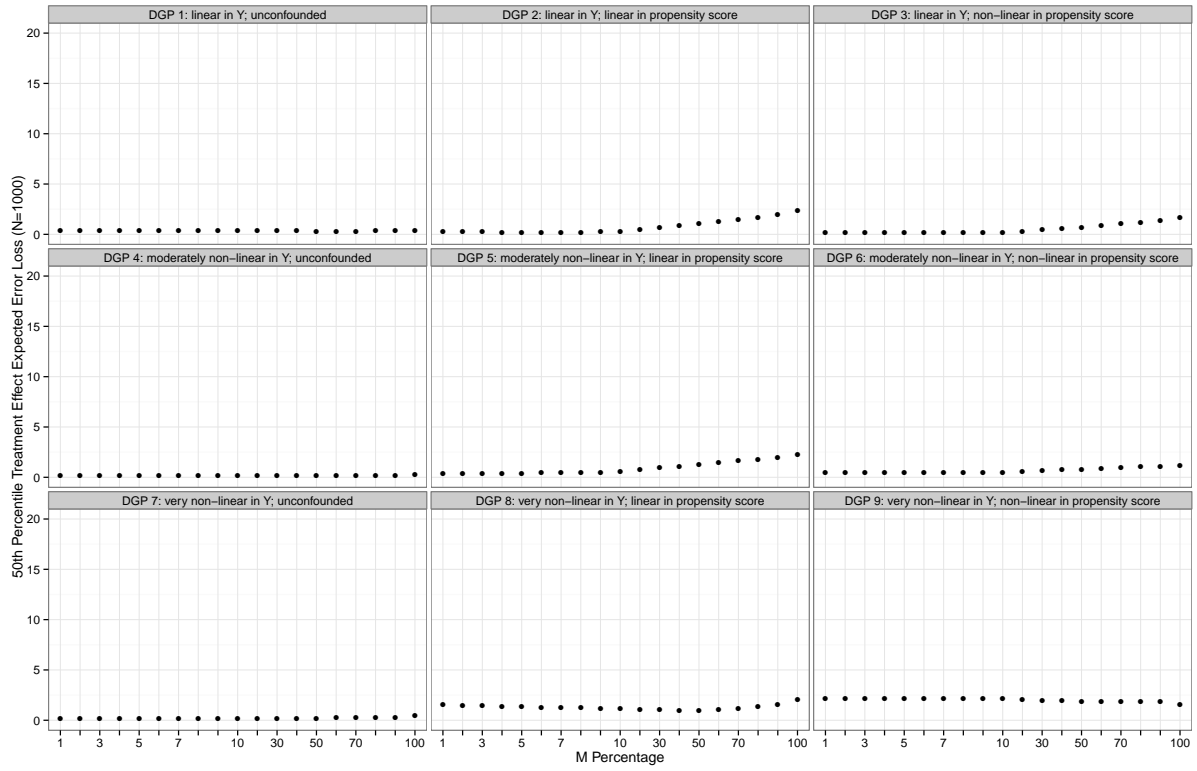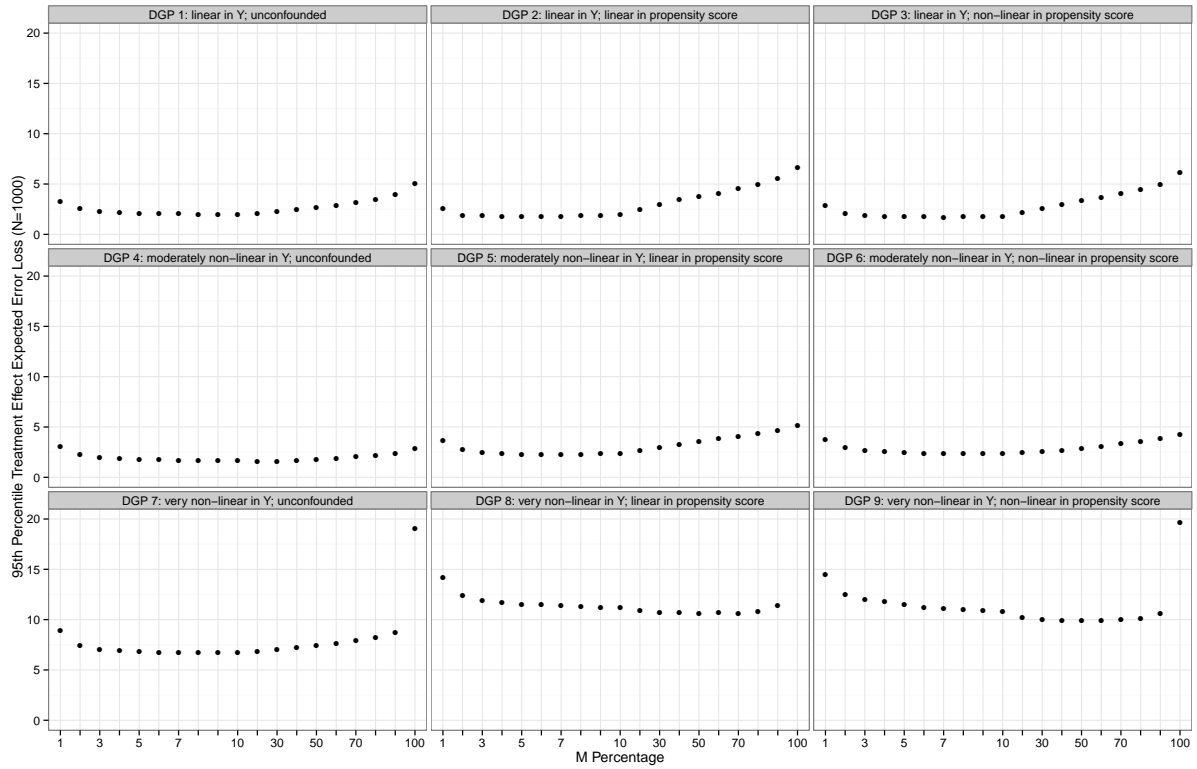
Figure 48: Comparing 95th Percentile Treatment Effect Expected Error Loss for Different Match Percentages (continuous outcome)
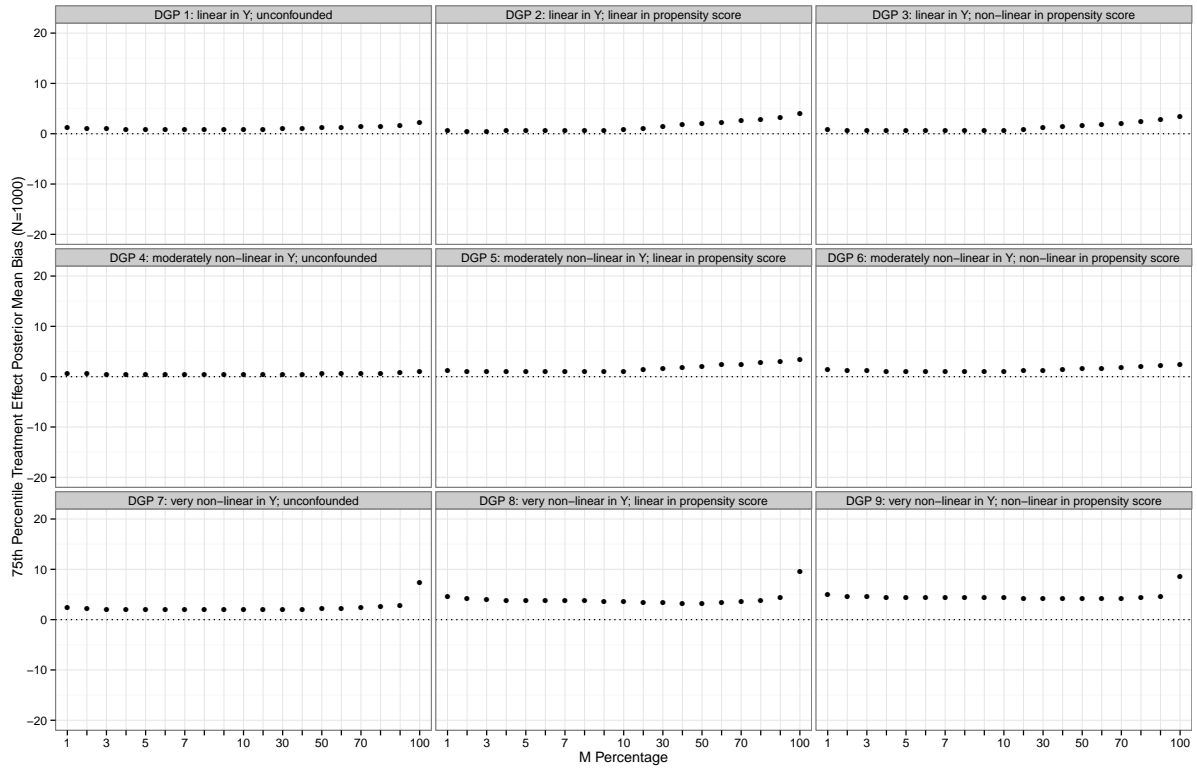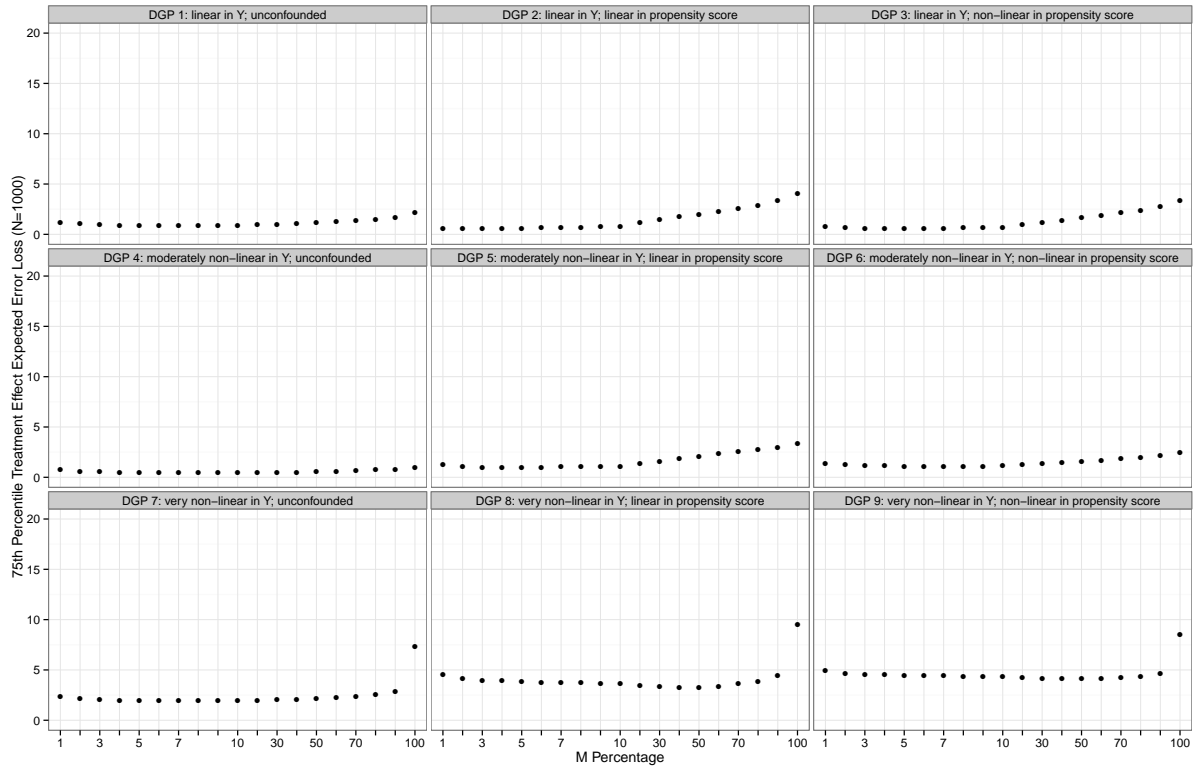
4 Comparing Different τ_i Distributions for Continuous Outcomes
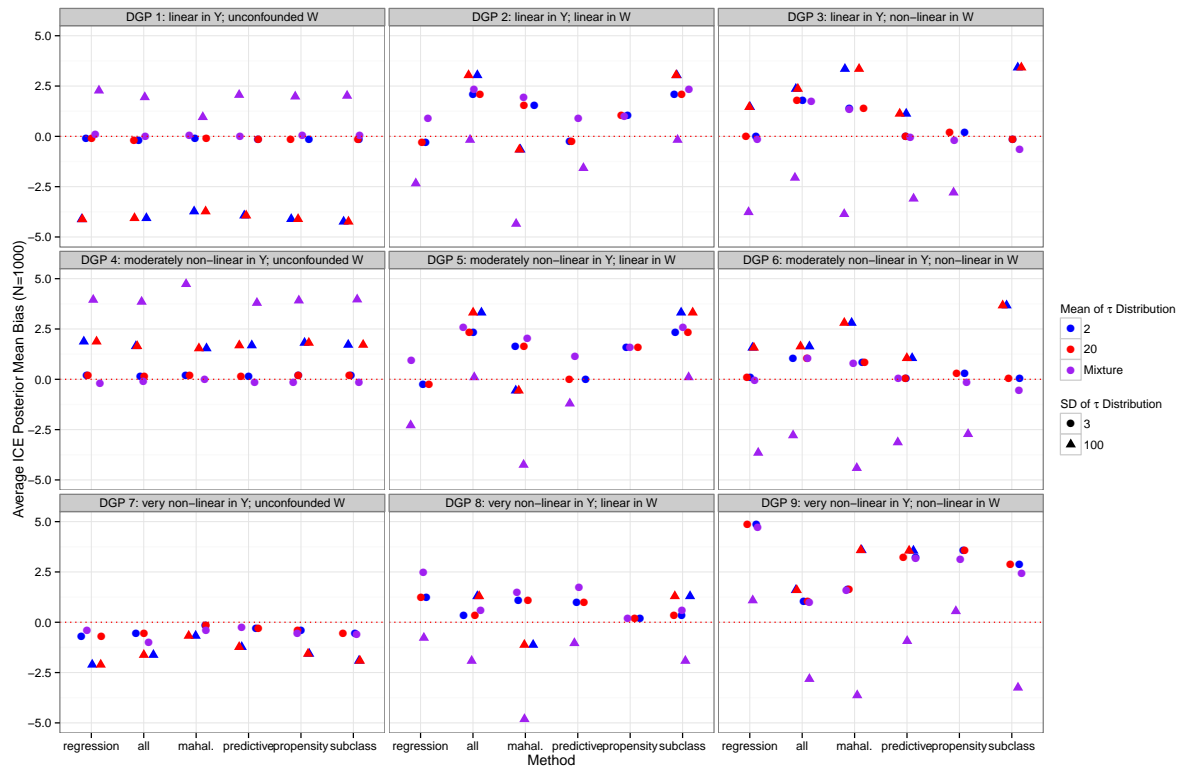
Figure 49: Comparing Average ICE (or ATE) Posterior Mean Bias with Different τ_i Distributions (continuous outcome)
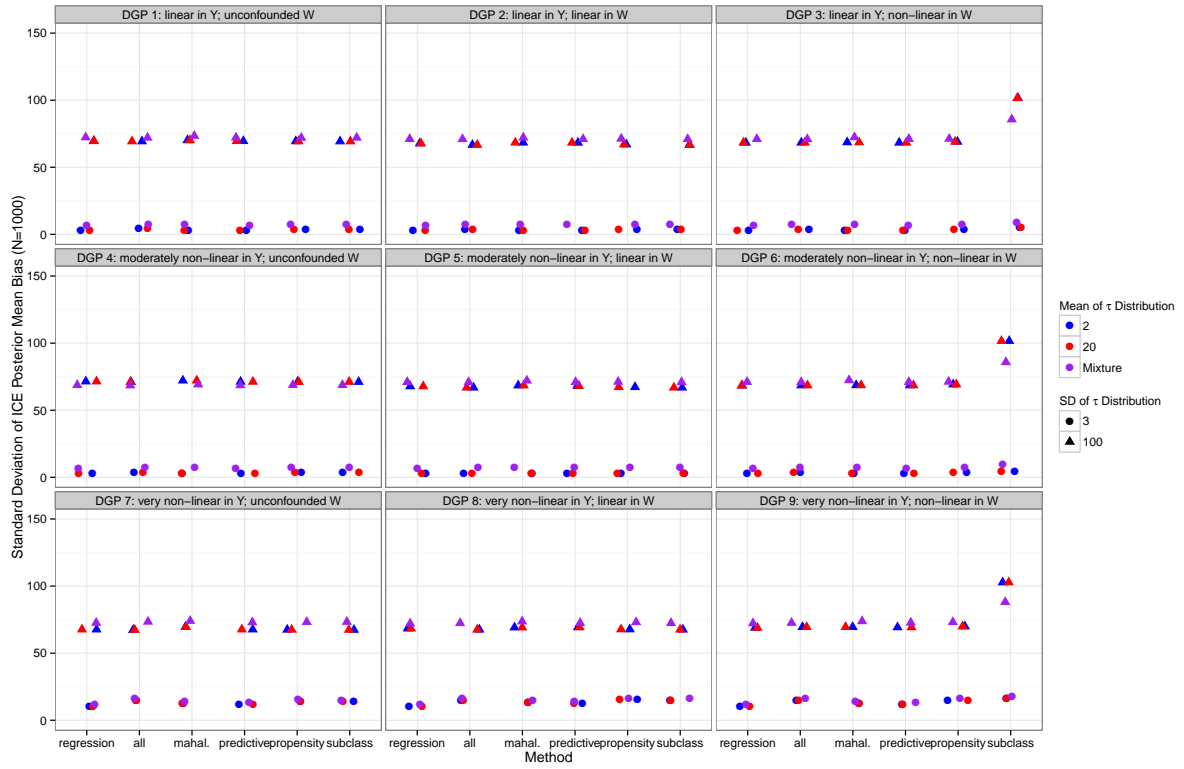
Figure 50: Comparing Standard Deviations of ICE Posterior Mean Bias with Different τ_i Distributions (continuous outcome)

Figure 51: Comparing Average ICE Expected Error Loss with Different τ_i Distributions (continuous outcome)

Figure 52: Comparing ICE "Power" with Different τ_i Distributions (continuous outcome)
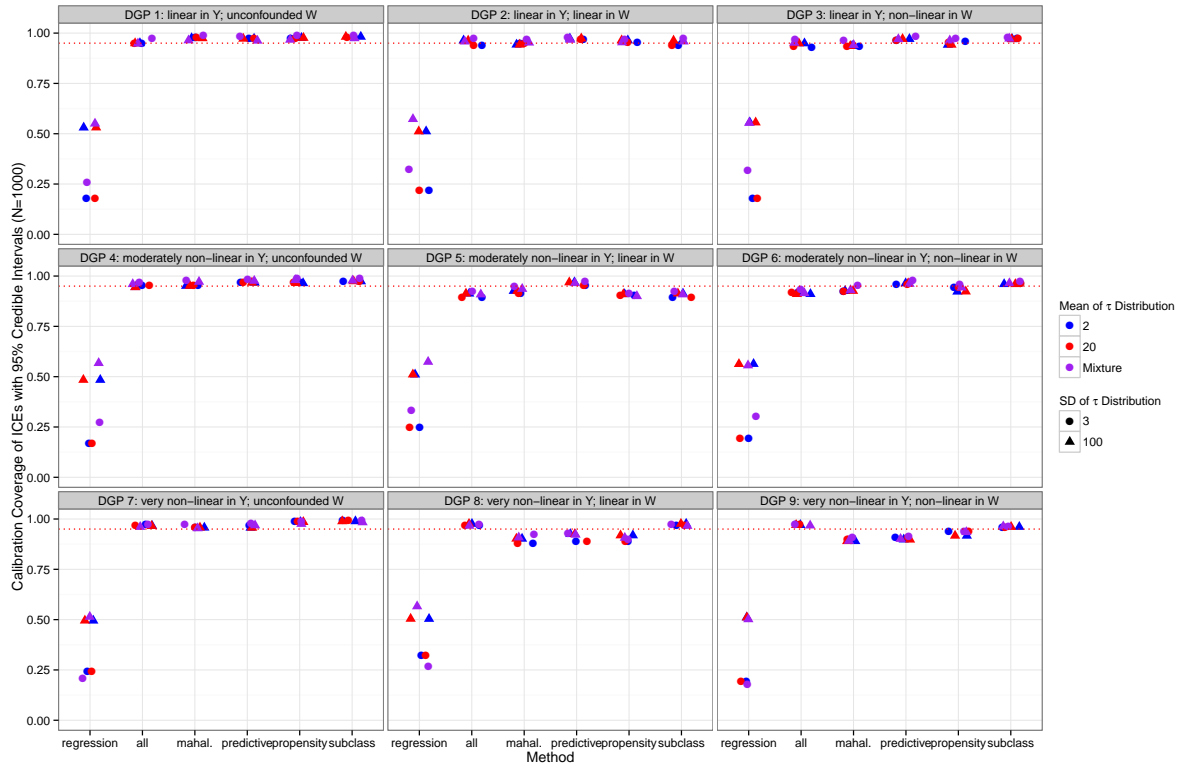
Figure 53: Comparing ICE Calibration Coverage with Different τ_i Distributions (continuous outcome)
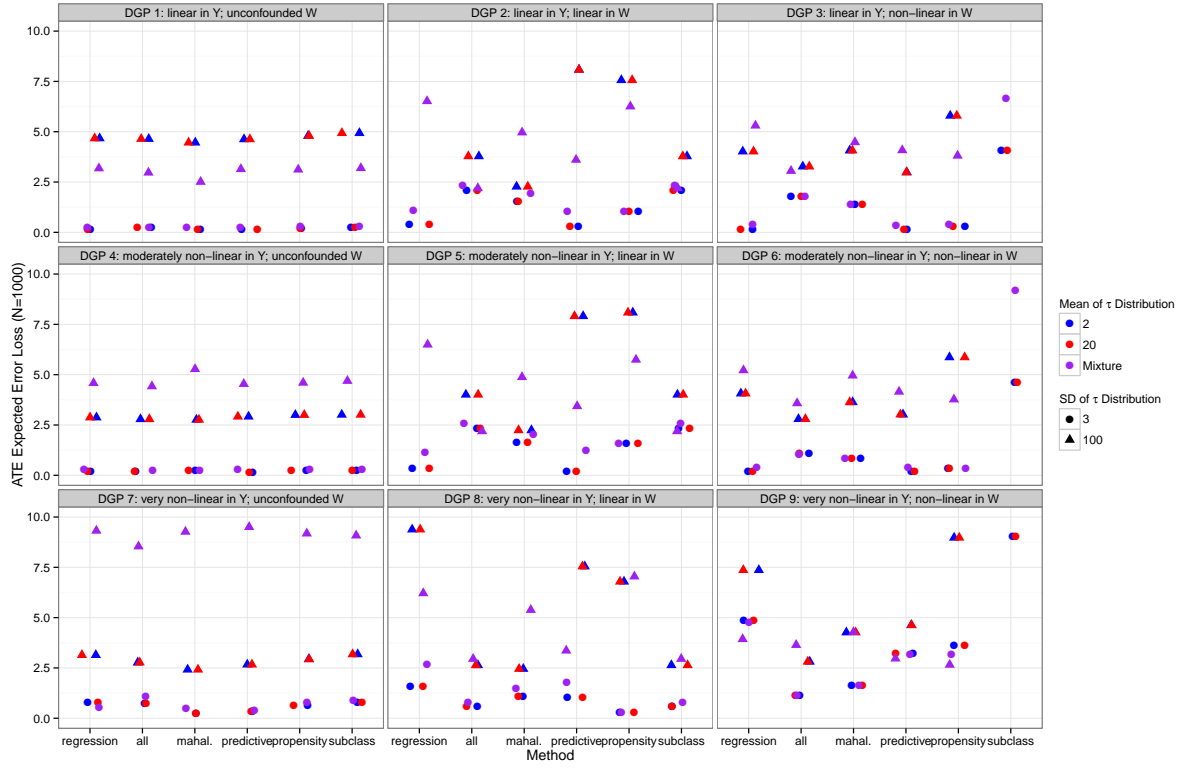
Figure 54: Comparing ATE Expected Error Loss with Different τ_i Distributions (continuous outcome)
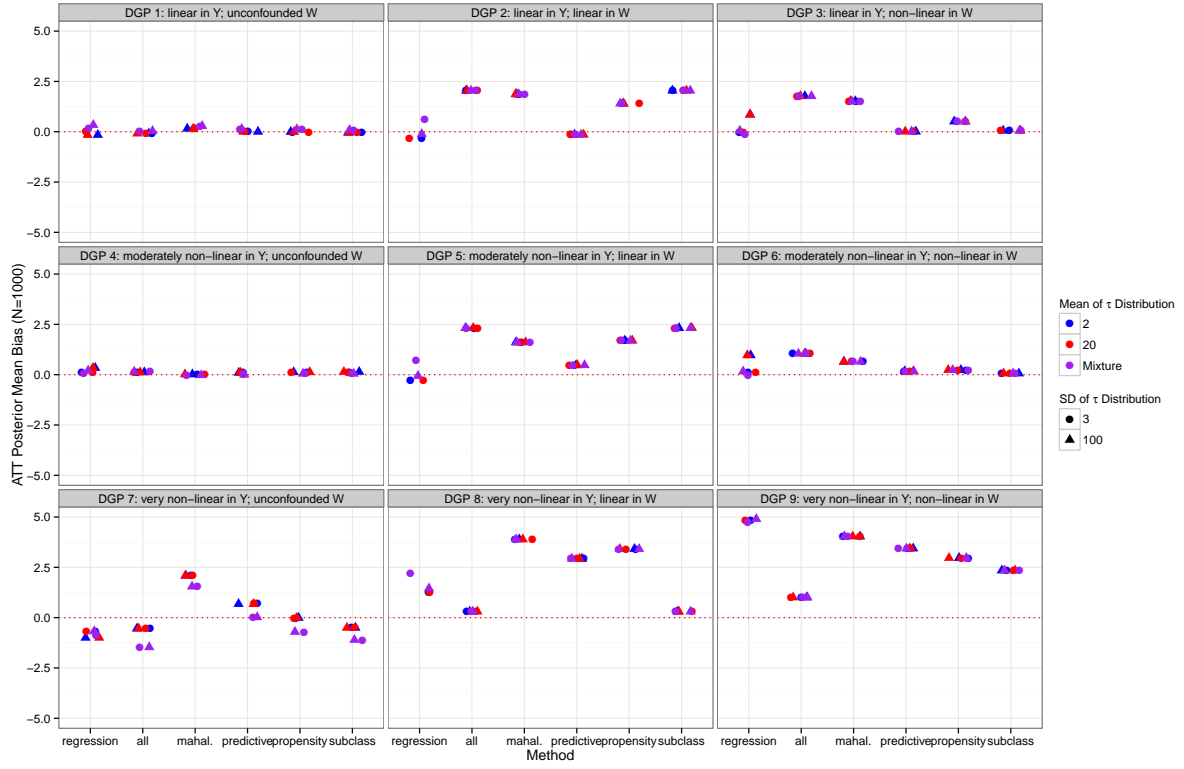
Figure 55: Comparing ATT Posterior Mean Bias with Different τ_i Distributions (continuous outcome)

Figure 56: Comparing ATT Expected Error Loss with Different τ_i Distributions (continuous outcome)

Figure 57: Comparing 50th Percentile Treatment Effect Posterior Mean Bias with Different τ_i Distributions (continuous outcome)
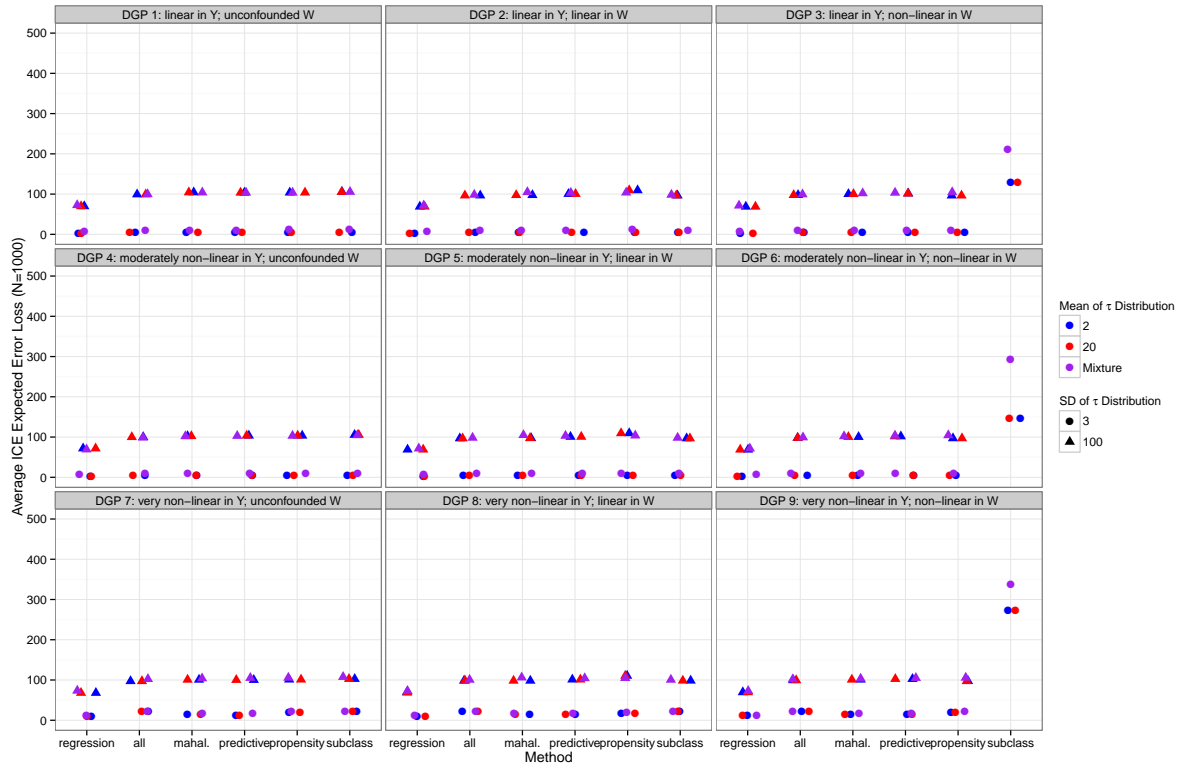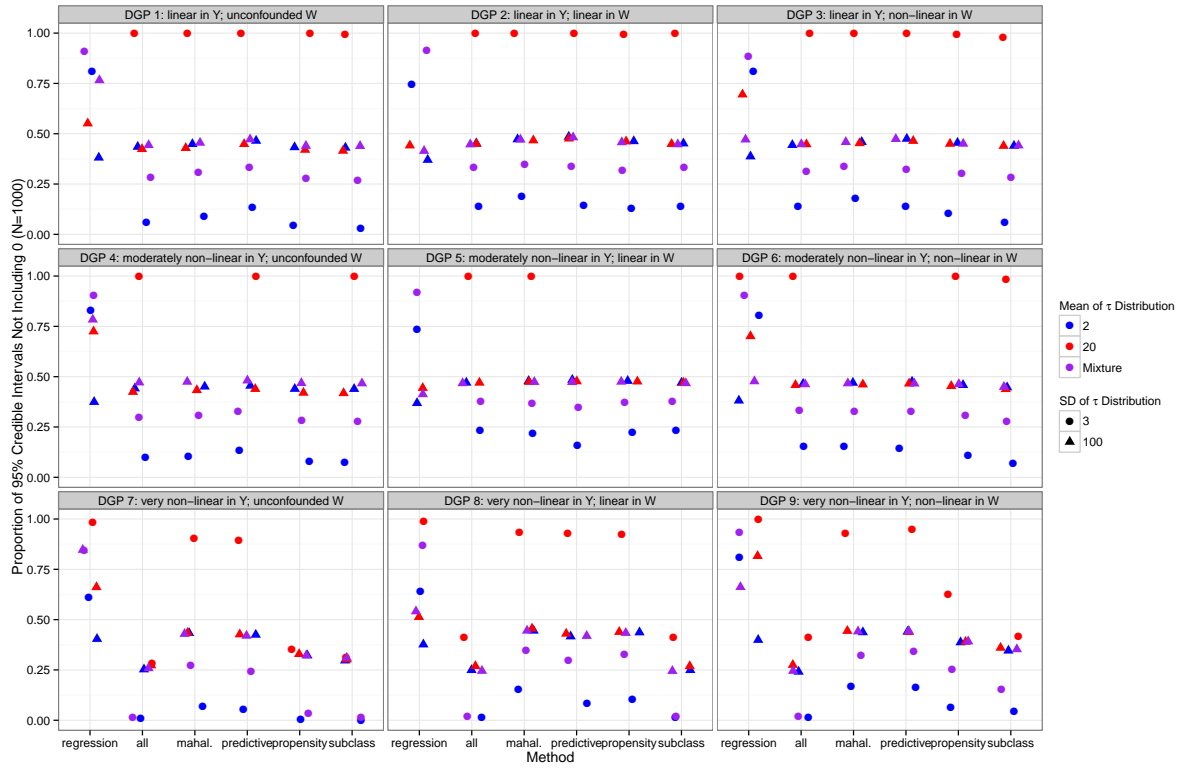
Figure 58: Comparing 50th Percentile Treatment Effect Expected Error Loss with Different τ_i Distributions (continuous outcome)

Figure 59: Comparing 75th Percentile Treatment Effect Posterior Mean Bias with Different τ_i Distributions (continuous outcome)

Figure 60: Comparing 75th Percentile Treatment Effect Expected Error Loss with Different τ_i Distributions (continuous outcome)

Figure 61: Comparing 95th Percentile Treatment Effect Posterior Mean Bias with Different τ_i Distributions (continuous outcome)
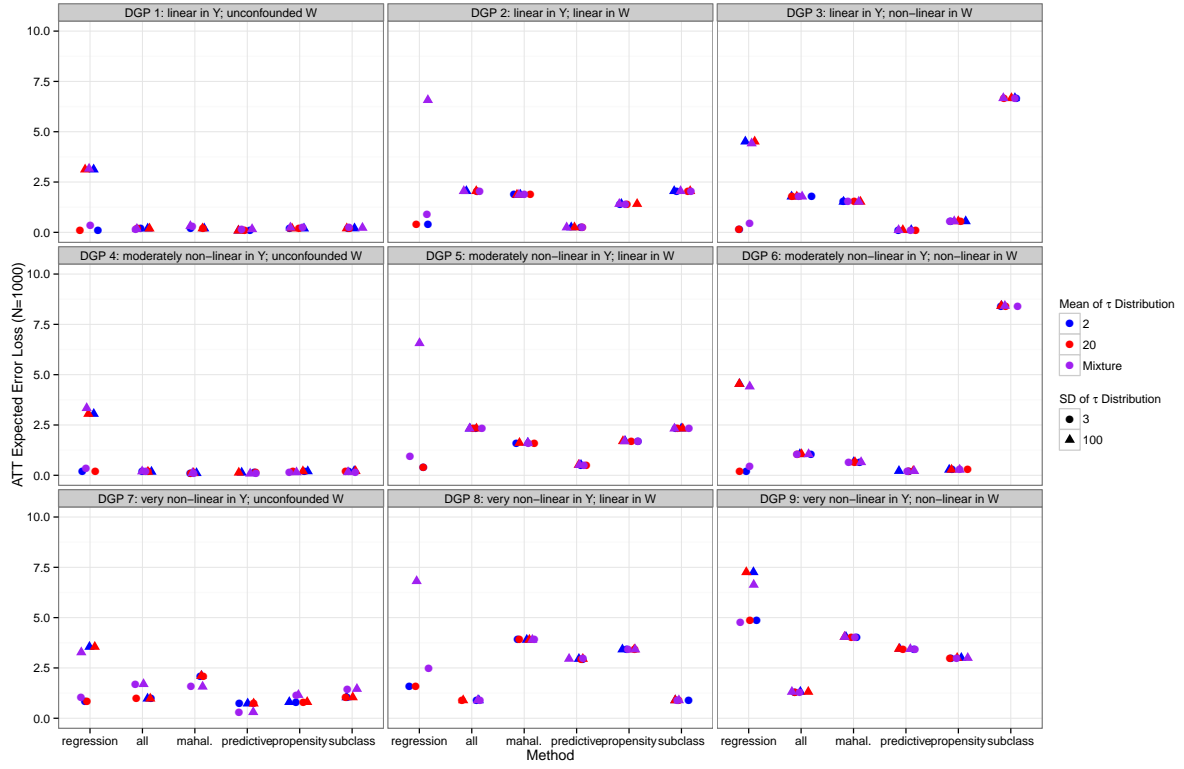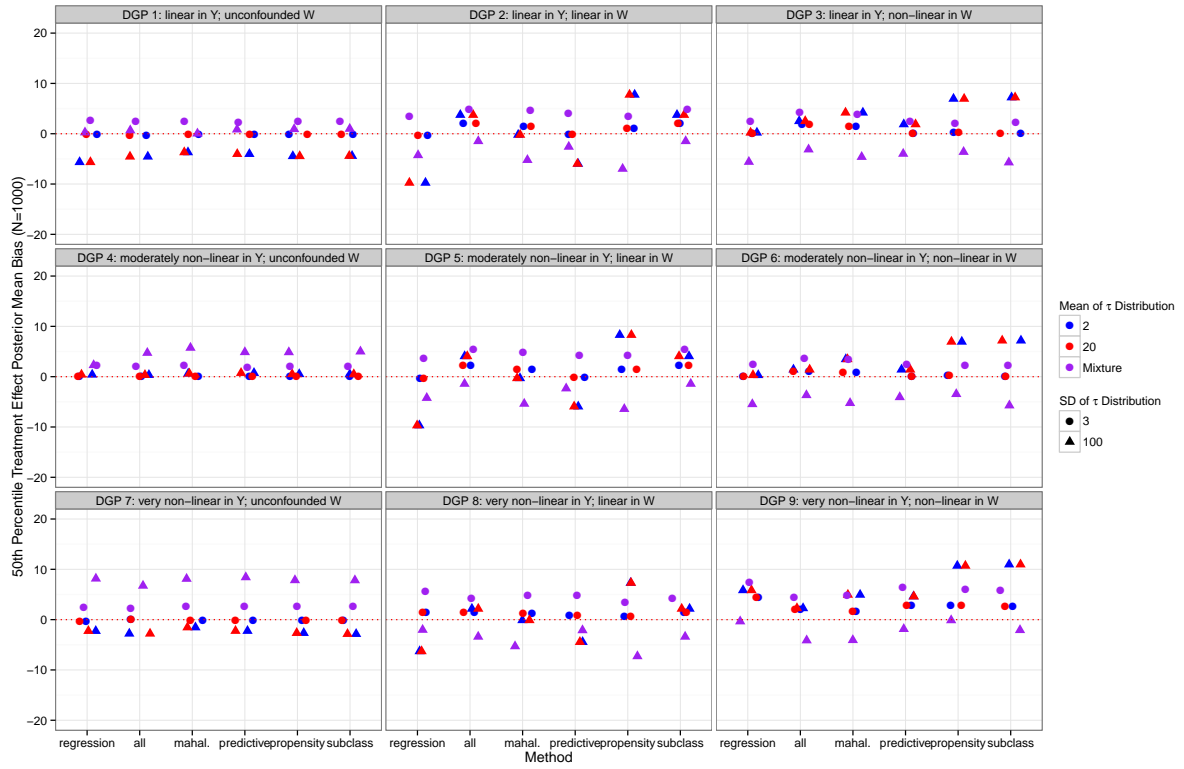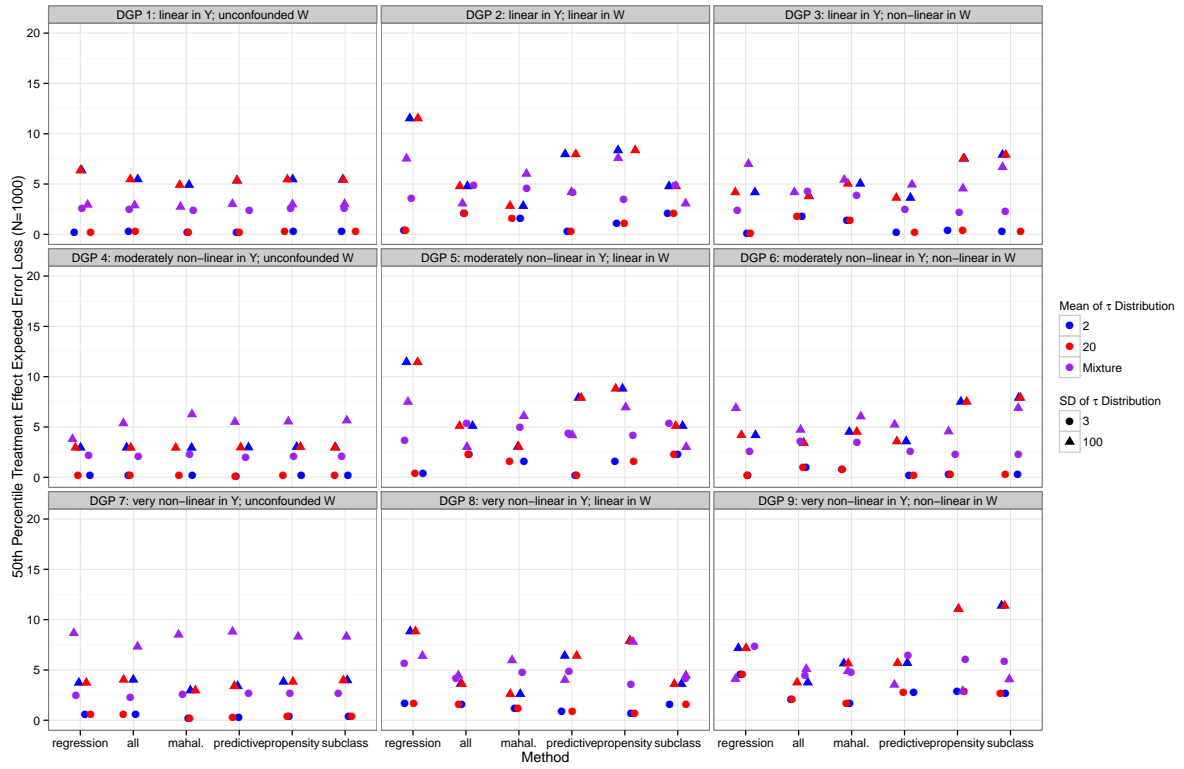
Figure 62: Comparing 95th Percentile Treatment Effect Expected Error Loss with Different τ_i Distributions (continuous outcome)
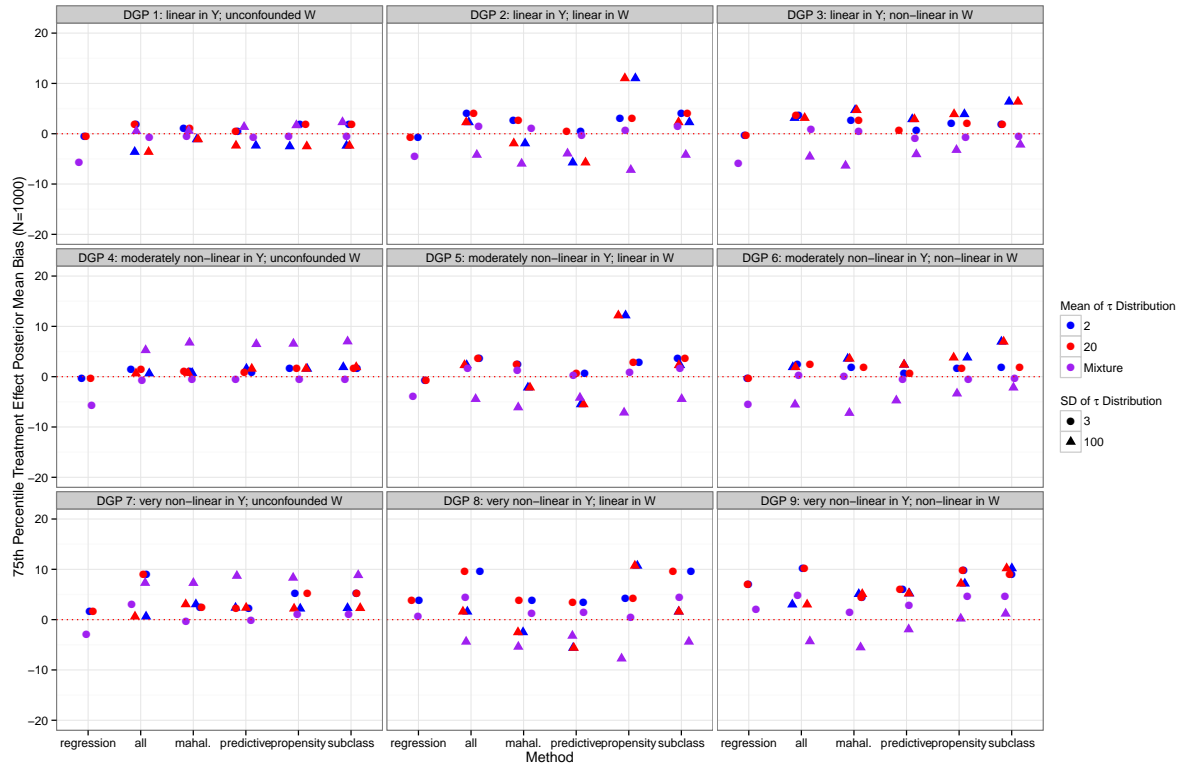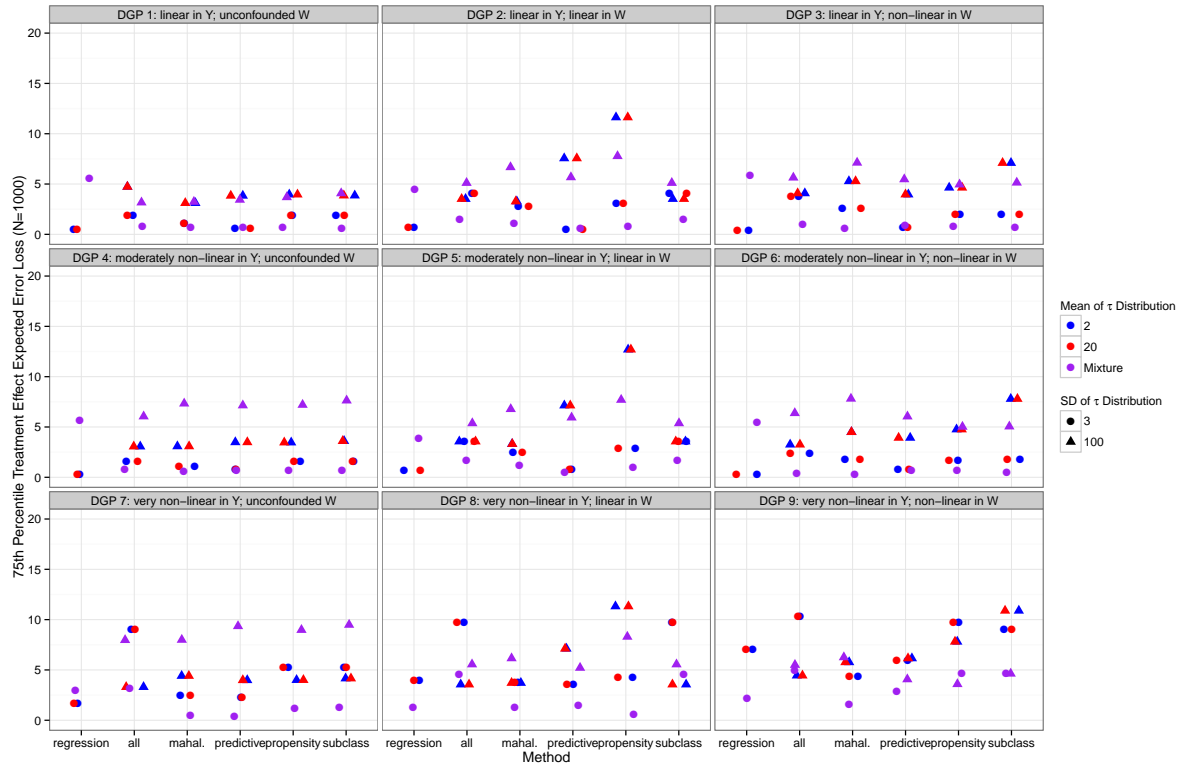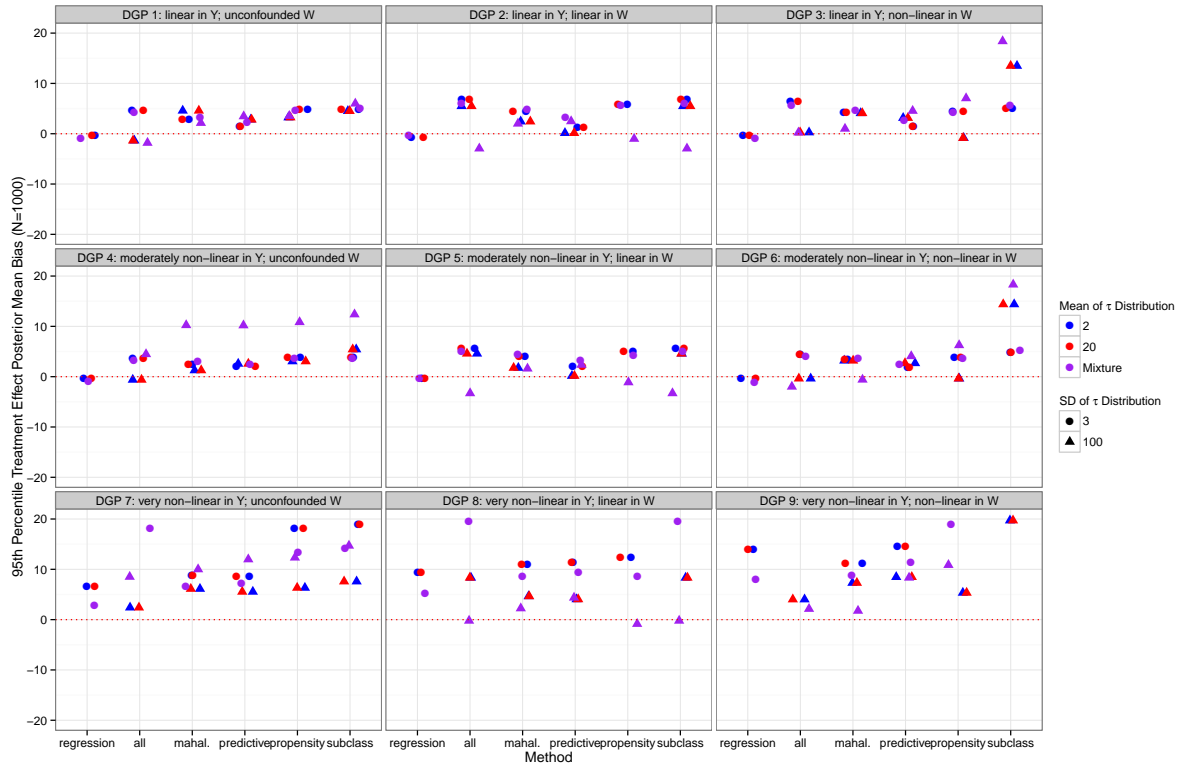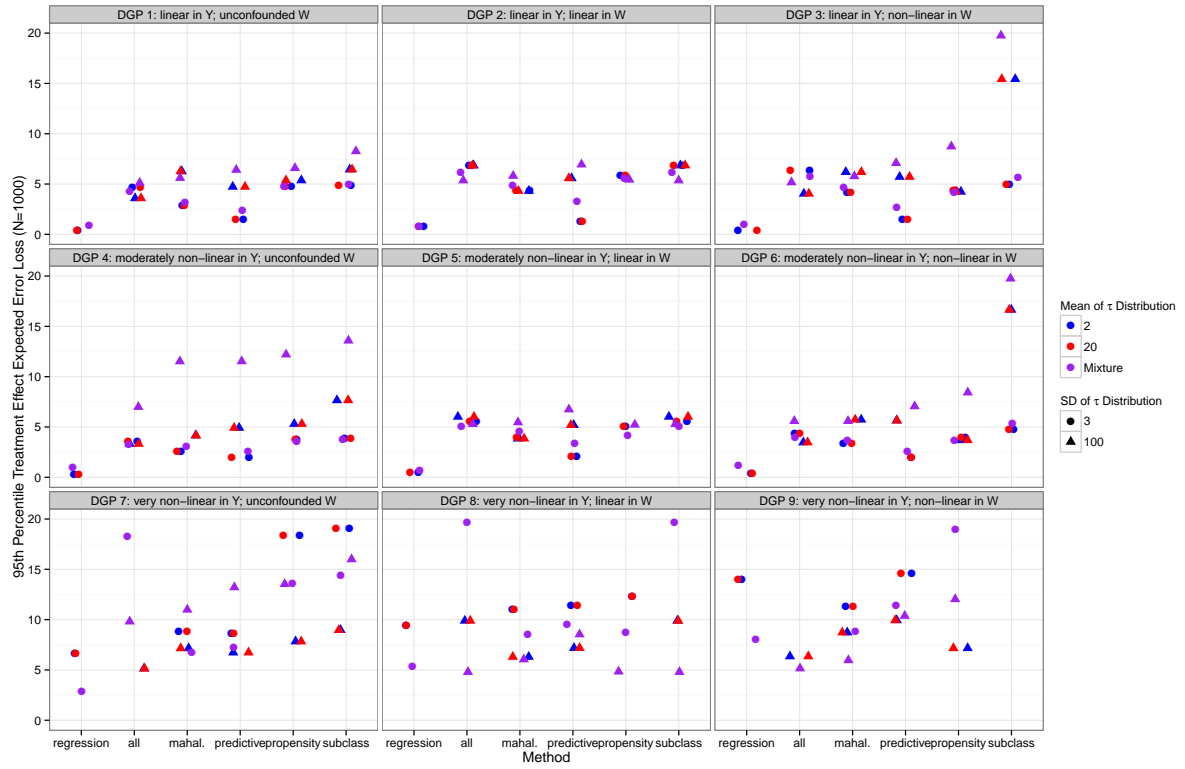
5 Simulations for Binary Outcomes

The simulations for binary outcomes test the same methods and ideas as the simulations for the continuous outcomes. Because of the nature of ICEs for binary outcomes with binary treatment, where each τ_i can only take on a value of -1, 0, 1, it is tough to develop the ICE proportion of 95% credible intervals including 0 ("power") or the ICE calibration coverage ("coverage") metrics. The posterior draws for each ICE consist of values of 0 and 1 or -1, depending on the treatment assignment and observed outcomes. Since it is unlikely that 95% or more of the posterior draws are of the same value, the 95% credible intervals almost certainly contain both possible values, so the two metrics are meaningless. Therefore, I only present the metrics of posterior mean bias ("bias") and expected error loss ("root mse"). For similar reasons, I only calculate and present results for two causal estimands, the ATE and the ATT. The simulation testing capabilities for binary outcomes are much more limited, so I rely mostly on the simulations for continuous outcomes to reach my conclusions. However, the results that I do calculate for the simulations for binary outcome variables are very similar to the results for continuous outcomes.

The data generating process for binary outcomes is also very similar to that of continuous outcomes. I use the same covariates generated before. Once again, there are nine different data generating processes with three different sample sizes. I first take the continuous outcomes for $Y(0)$ from before:

1. $Y(0) = x_1 + x_2 + x_3 - x_4 + x_5 + x_6 + x_7 - x_8 + x_9 - x_{10}$

2. $Y(0) = x_1 + x_2 + 0.2x_3x_4 - \sqrt{x_5} + x_7 + x_8 - x_9 + x_{10}$

3. $Y(0) = (x_1 + x_2 + x_5)^2 + x_7 - x_8 + x_9 - x_{10}$

To generate a binary outcome, I simply assign $Y_i(0) = 1$ if the continuous outcome is greater than the mean of all the $Y(0)$ and $Y_i(0) = 0$ if the continuous outcome is less than the mean. The treatment assignment generating process stays the same as before.

1. $p(W = 1) = 0.5$

2. $\eta = x_1 + 2x_2 - 2x_3 - x_4 - 0.5x_5 + x_6 + x_7$
   $W = 1$ if $\eta > 0$; otherwise $W = 0$

3. $\eta = 0.5x_1 + 2x_1x_2 + x_3^2 - x_4 - 0.5\sqrt{x5} - x_5x_6 + x_7$
   $W = 1$ if $\eta > 0$; otherwise $W = 0$

To generate τ_i, I use the following formula:

$$
\begin{array}{l}
\text{If } Y_i(0) = 0, \\
\\
\qquad \begin{aligned}
P(\tau_i = 1) &= 0.75 \\
P(\tau_i = 0) &= 0.25
\end{aligned} \\
\\
\text{If } Y_i(0) = 1, \\
\\
\qquad \begin{aligned}
P(\tau_i = -1) &= 0.4 \\
P(\tau_i = 0) &= 0.6
\end{aligned}
\end{array}
$$

Given $Y_i(0)$, W_i, and τ_i, then

$$
\begin{aligned}
Y_i(1) &= Y_i(0) + \tau_i \\
Y_i &= W_iY_i(1) + (1 - W_i)Y_i(0)
\end{aligned}
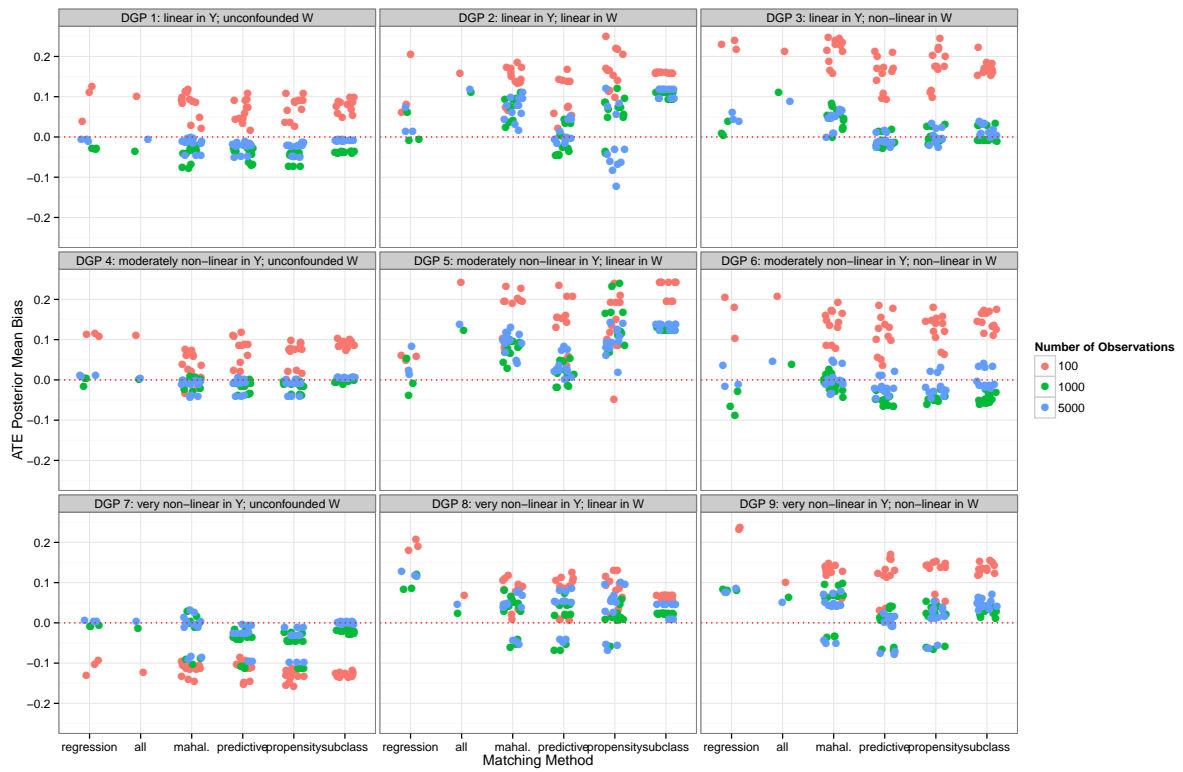$$

6 Comparing Methods for Binary Outcomes

Figure 63: Comparing ATE Posterior Mean Bias for Different Matching Methods (binary outcome)
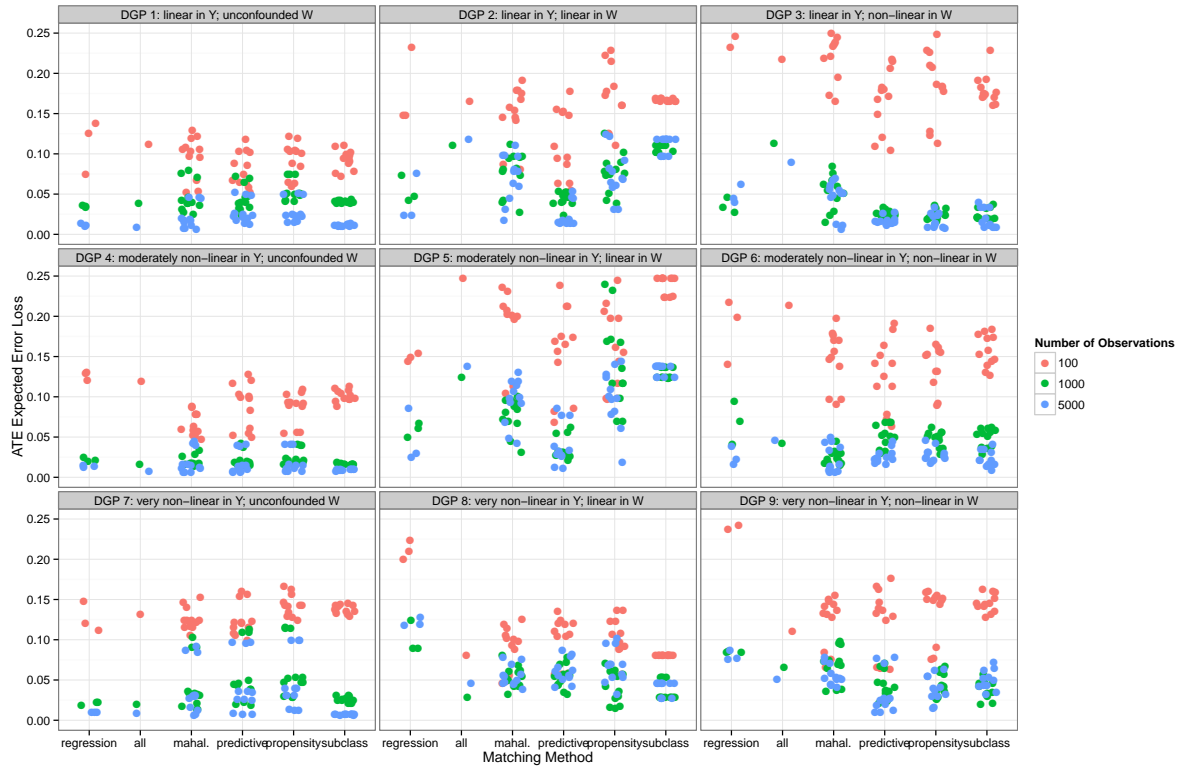
Figure 64: Comparing ATE Expected Error Loss for Different Matching Methods (binary outcome)
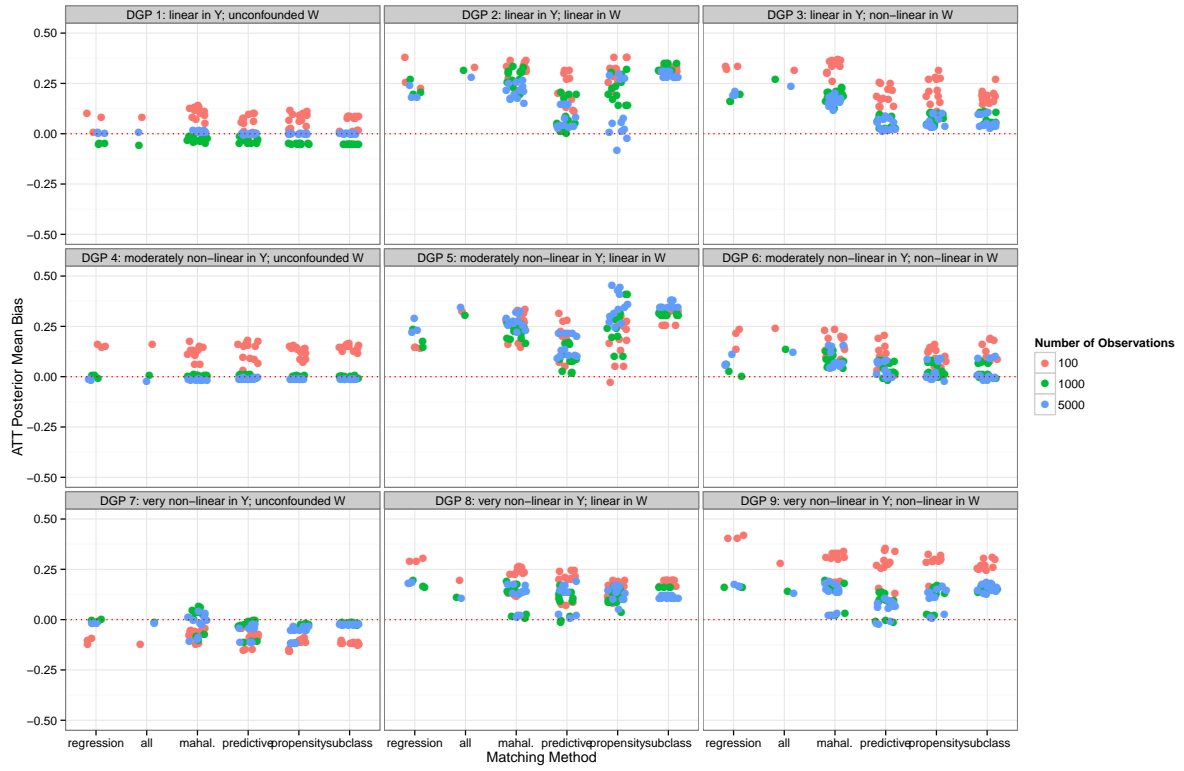
Figure 65: Comparing ATT Posterior Mean Bias for Different Matching Methods (binary outcome)
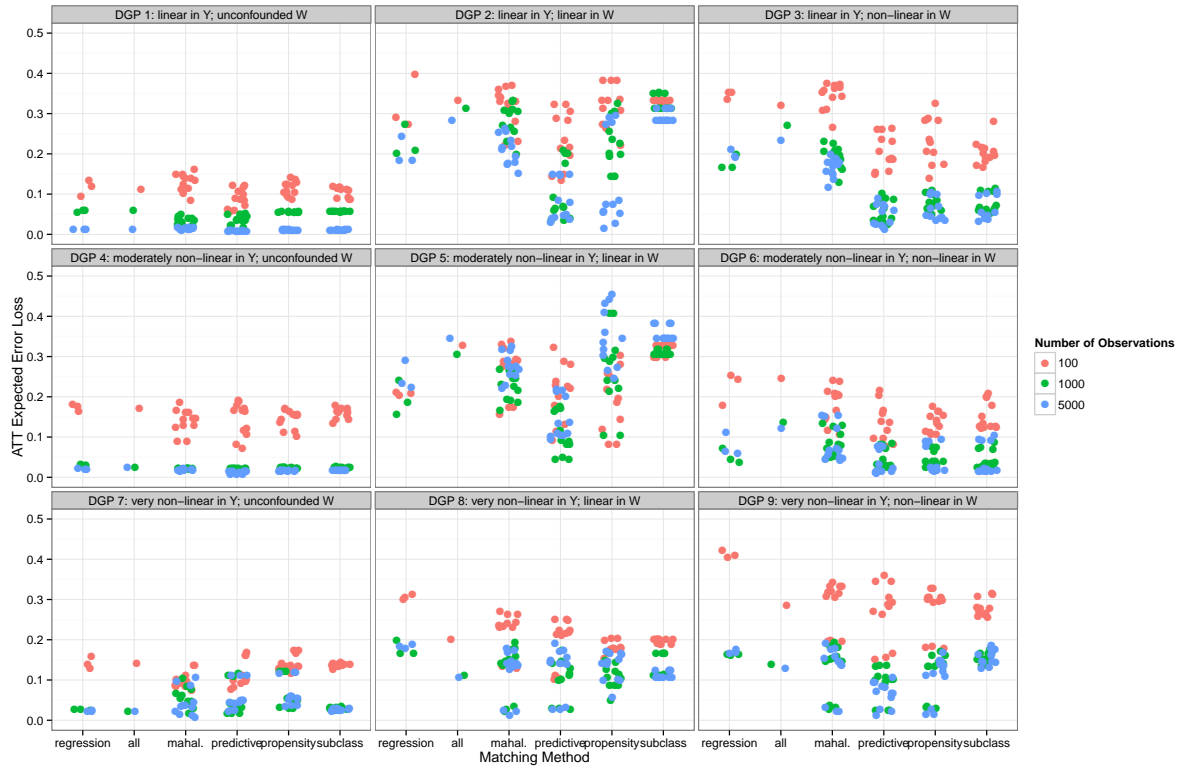
Figure 66: Comparing ATT Expected Error Loss for Different Matching Methods (binary outcome)

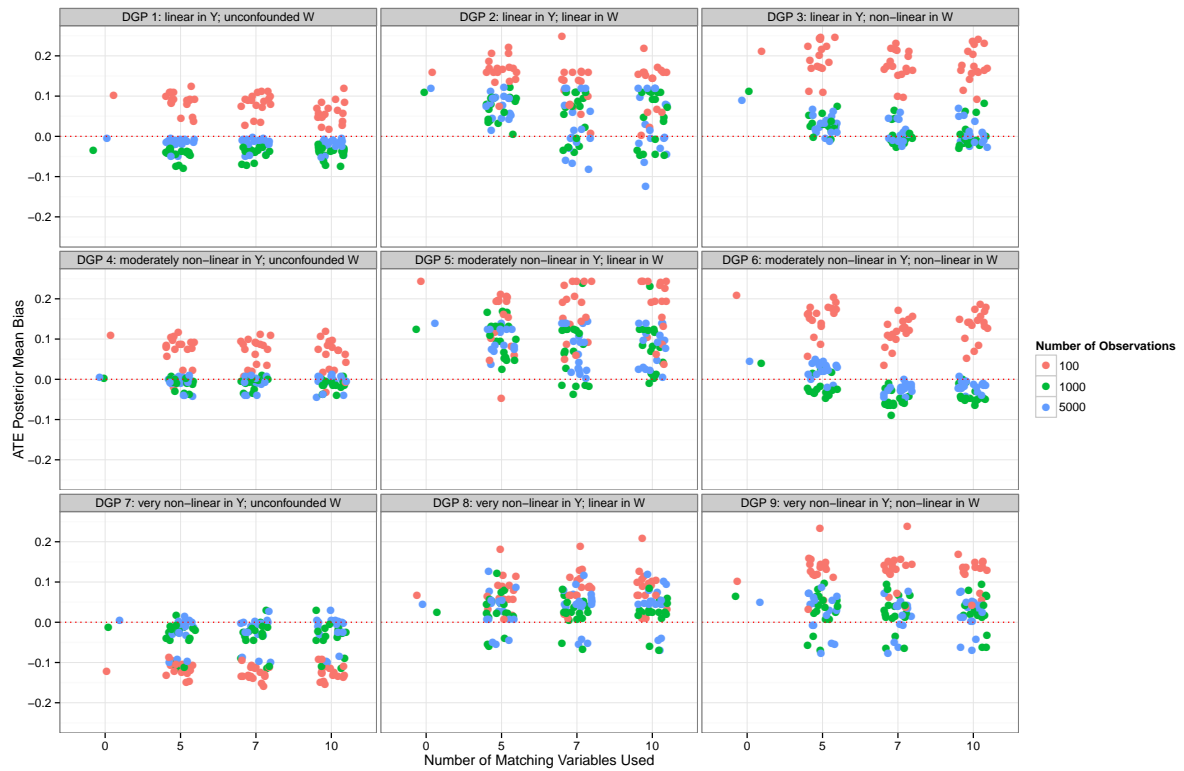# 7 Comparing Number of Conditioning Variables for Binary Outcomes

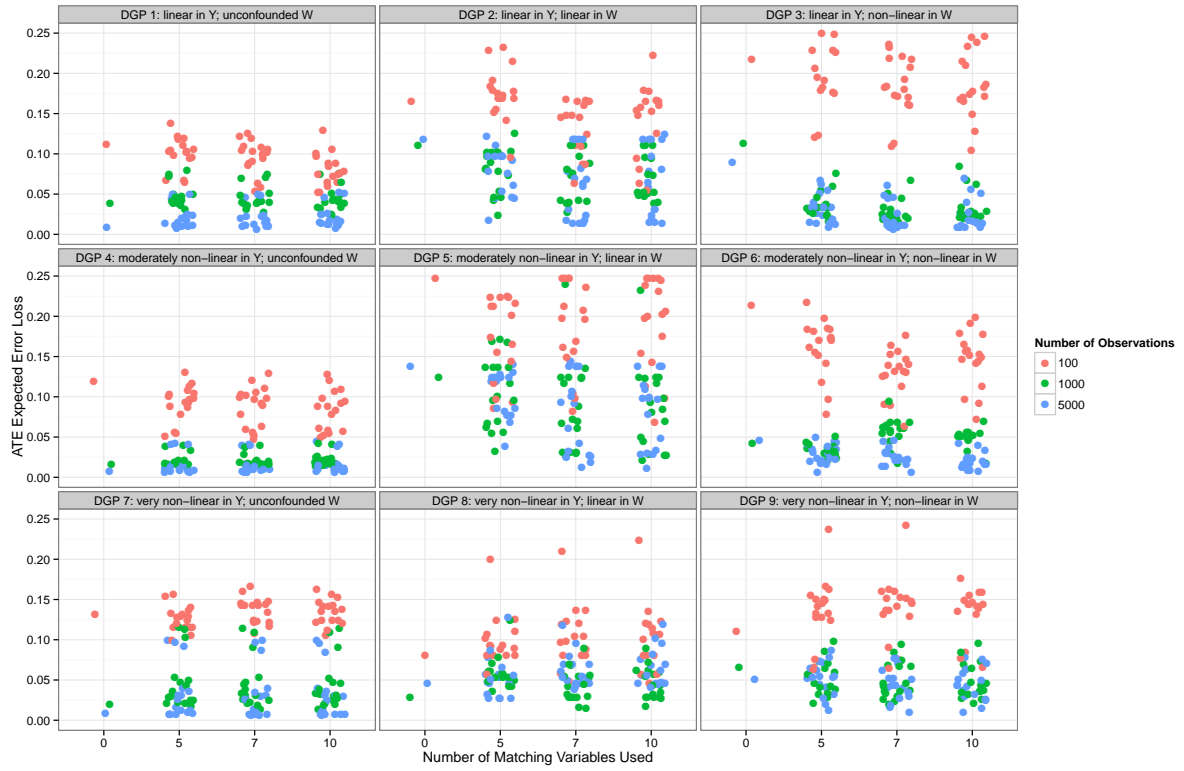Figure 67: Comparing ATE Posterior Mean Bias for Different Conditioning Sets (binary outcome)

Figure 68: Comparing ATE Expected Error Loss for Different Conditioning Sets (binary outcome)

Figure 69: Comparing ATT Posterior Mean Bias for Different Conditioning Sets (binary outcome)
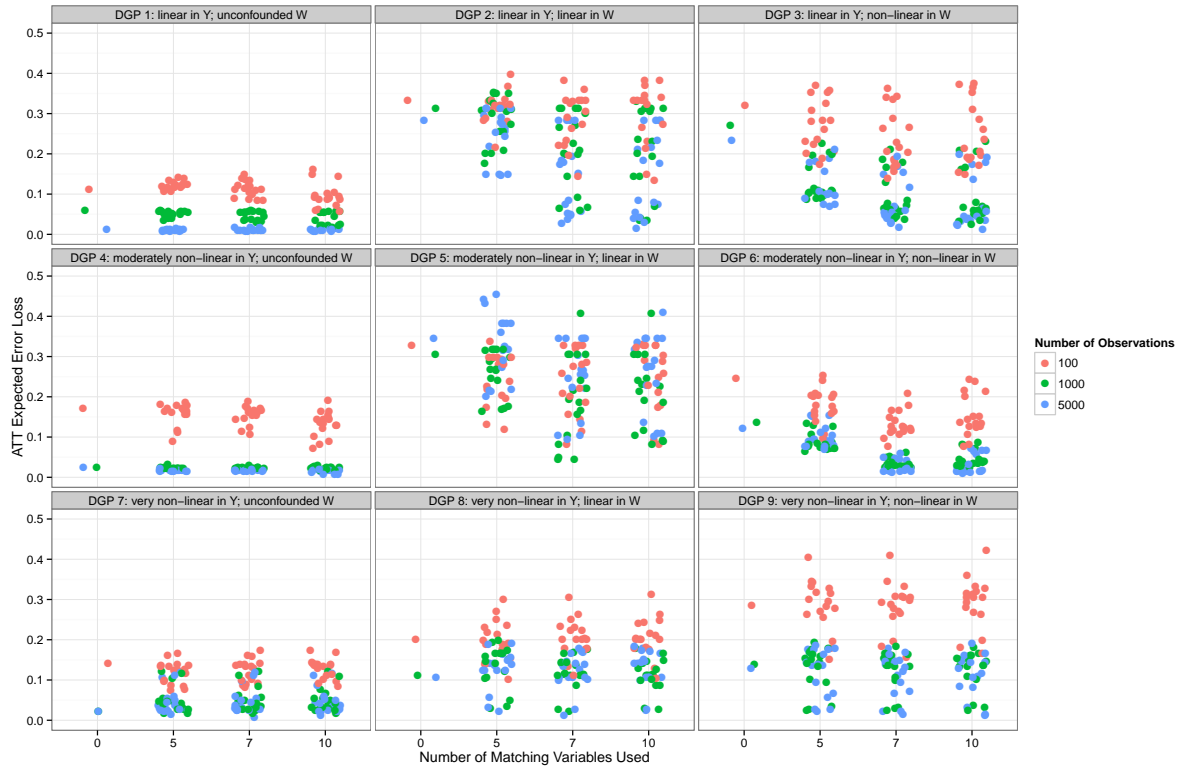
Figure 70: Comparing ATT Expected Error Loss for Different Conditioning Sets (binary outcome)

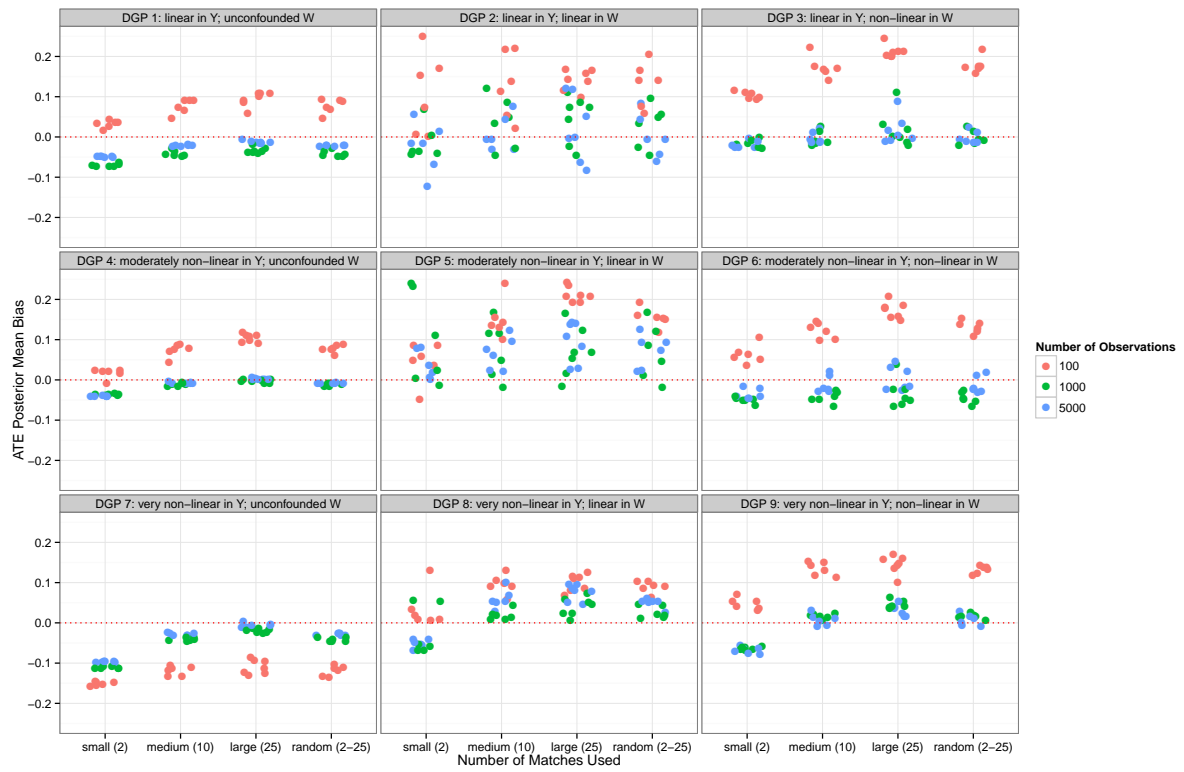# 8 Comparing Number of Matches for Binary Outcomes

Figure 71: Comparing ATE Posterior Mean Bias for Different Numbers of Matches (binary outcome)
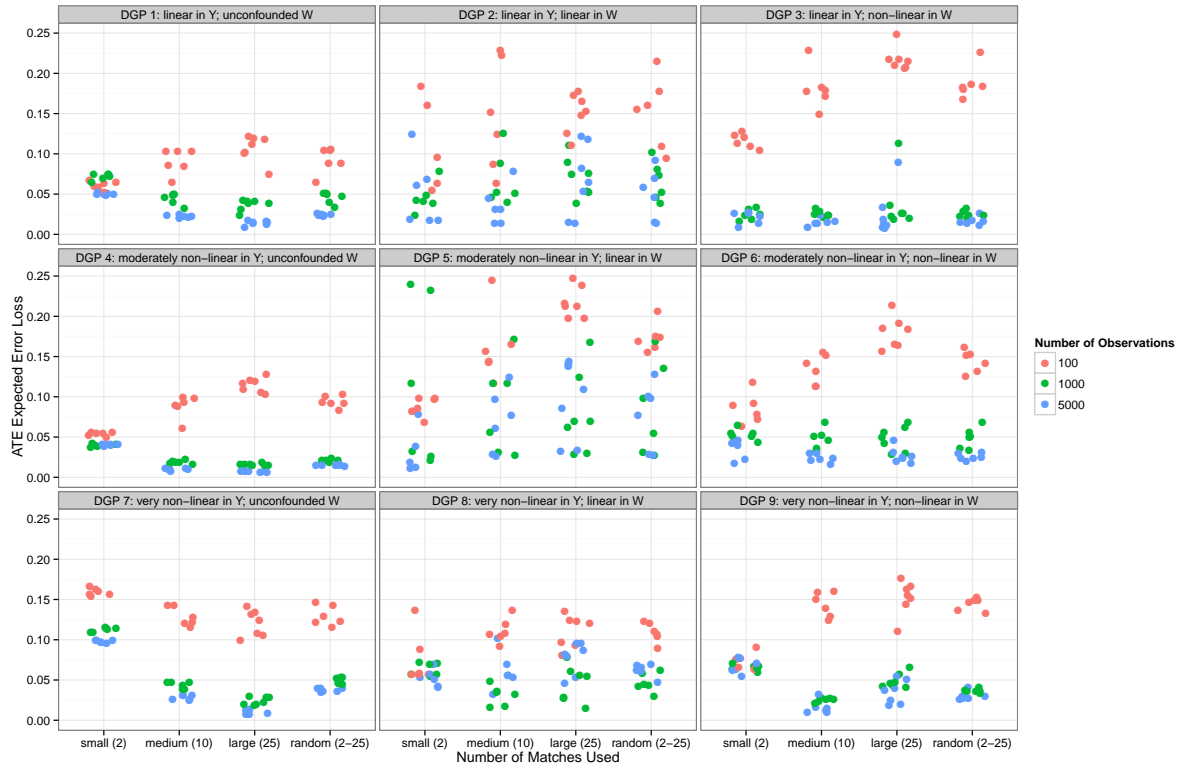
Figure 72: Comparing ATE Expected Error Loss for Different Numbers of Matches (binary outcome)
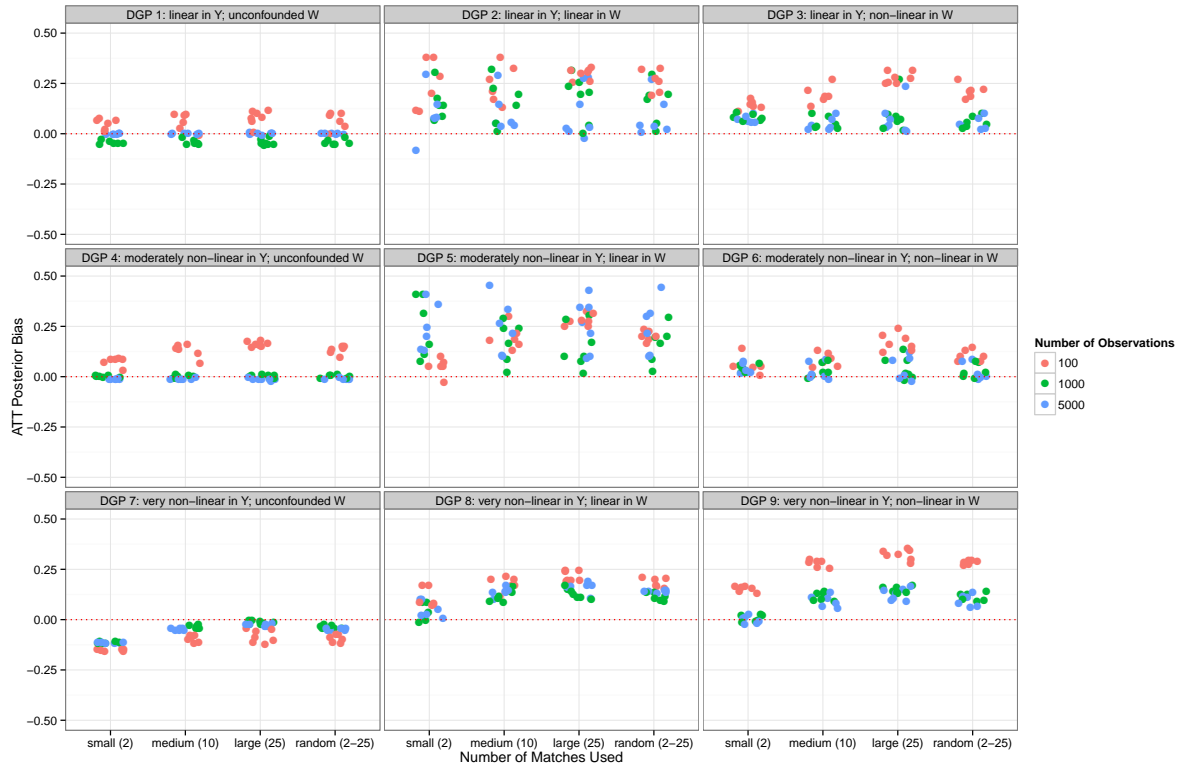
Figure 73: Comparing ATT Posterior Mean Bias for Different Numbers of Matches (binary outcome)
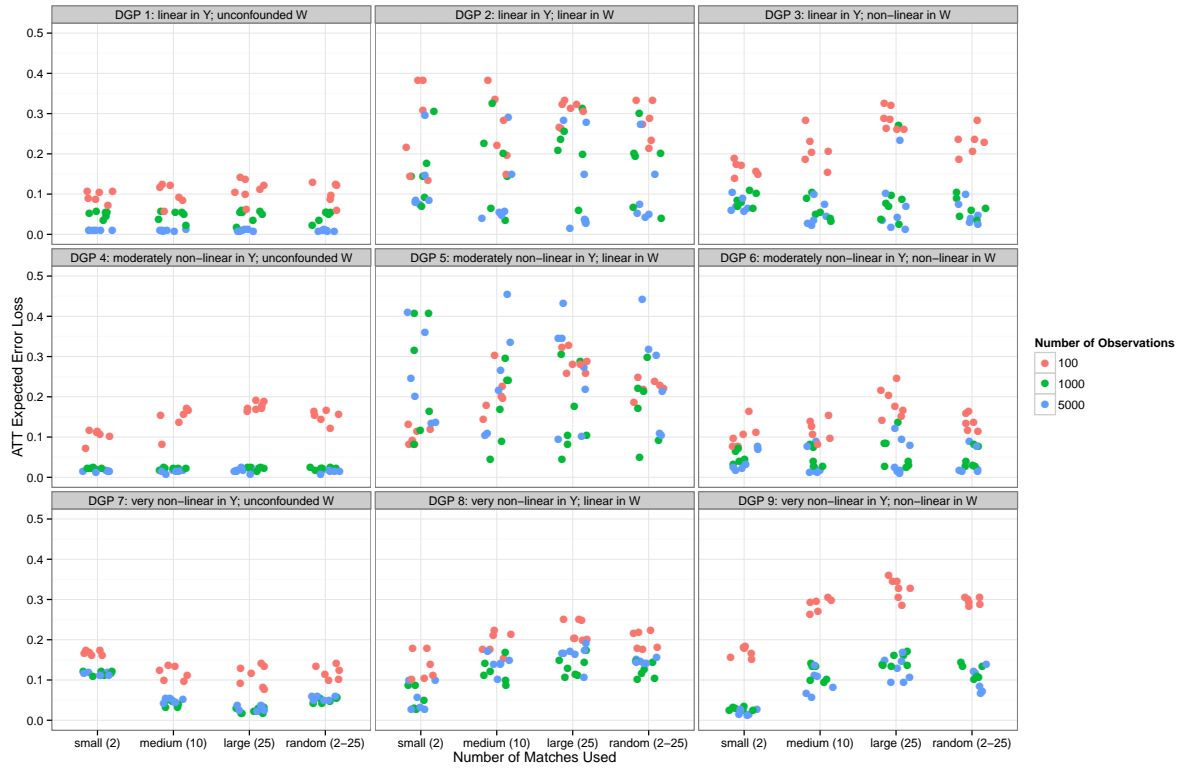
Figure 74: Comparing ATT Expected Error Loss for Different Numbers of Matches (binary outcome)