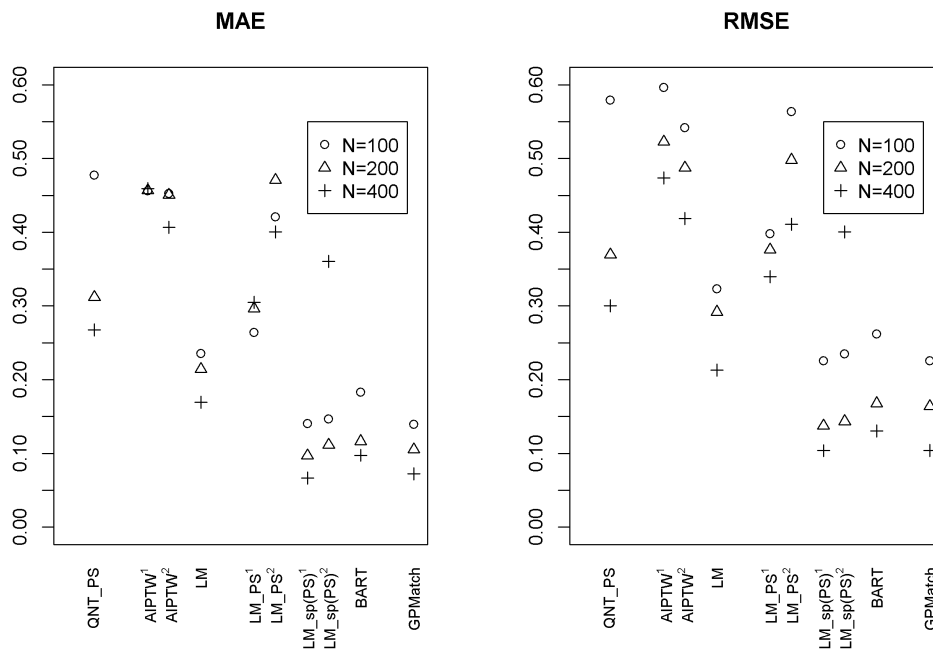**FIGURE S4** Comparisons of root mean square error (RMSE), and median absolute error (MAE) of the ATE Estimates by Different Methods Across Different Sample Sizes under the Simulation Setting 4: $\{\gamma_0, \gamma_1, \gamma_2, \gamma_3\} = \{1, 0.15, 0.7, 0\}$

[1] Propensity score estimated using logistic regression on $logitA \sim X$. [2] Propensity score estimated using logistic regression on $logitA \sim X^{1/3}$. GPMatch: Bayesian structural model with Gaussian process prior. QNT_PS: Propensity score sub-classification by quintiles. AIPTW: augmented inversed probability of treatment weighting; LM: linear regression modeling $Y \sim X$; LM_PS: linear regression modeling with propensity score adjustment. LM_sp(PS): linear regression modeling with spline fit propensity score adjustment. BART: Bayesian additive regression tree.