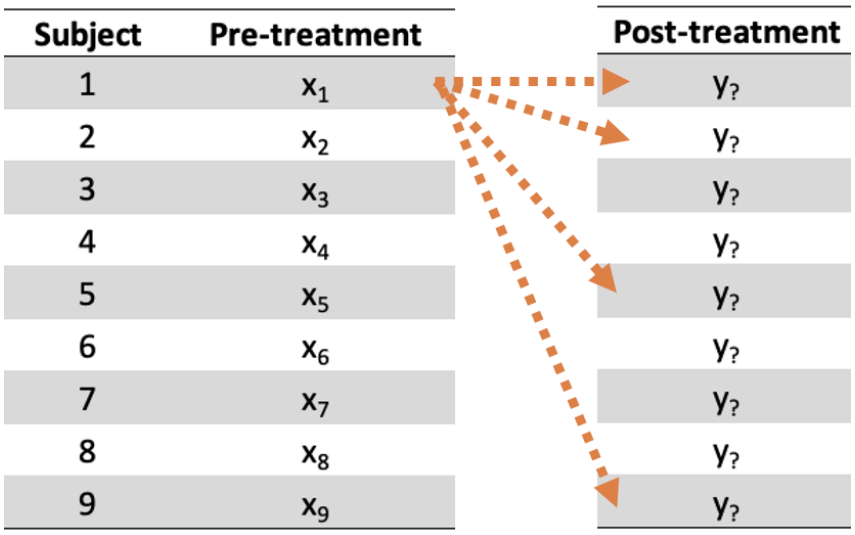
Figure 1: Illustration of datasets that are paired versus paired but unmatched.

A

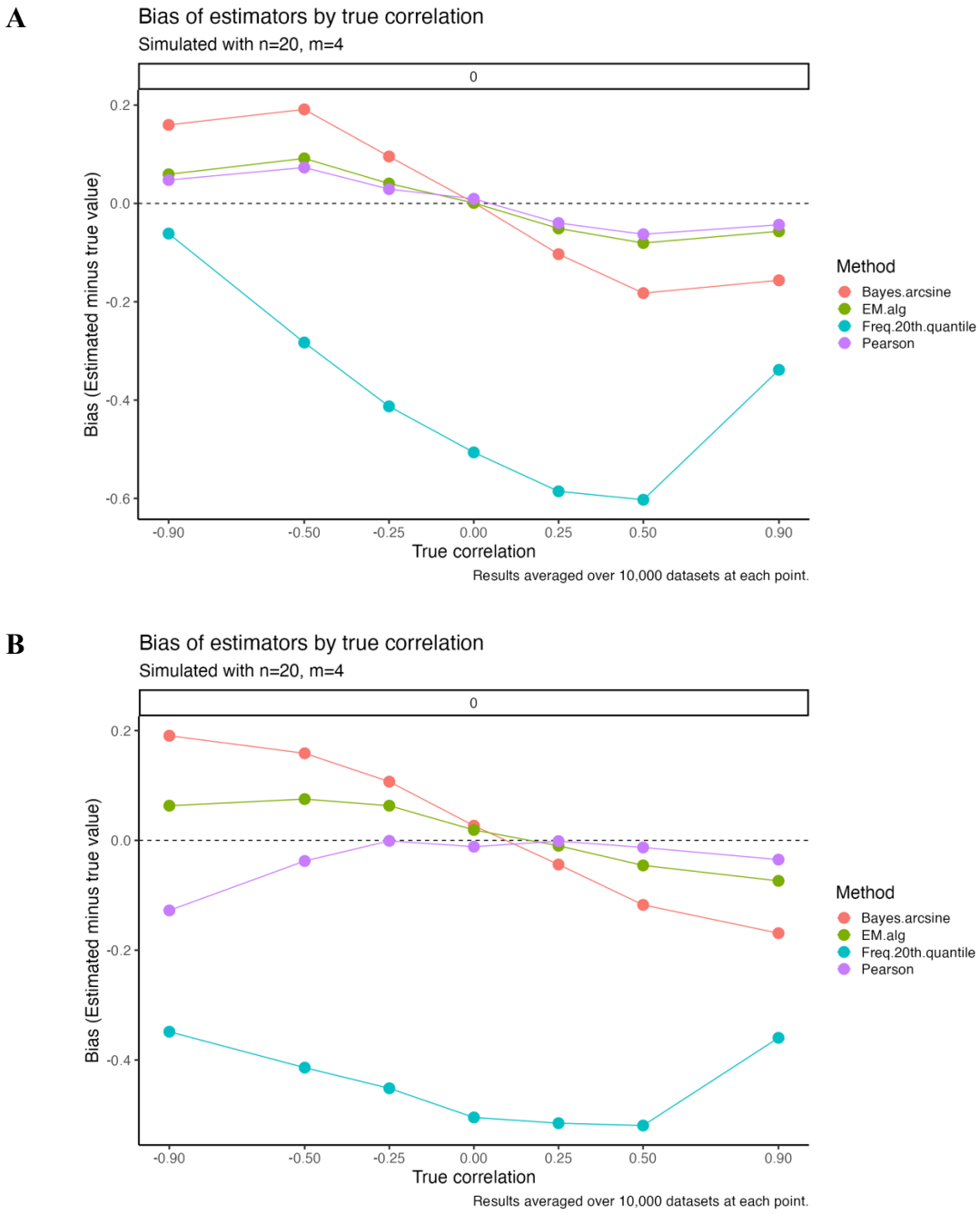
Subject	Pre-treatment	Post-treatment
1	x_1	y_1
2	x_2	y_2
3	x_3	y_3
4	x_4	y_4
5	x_5	y_5
6	x_6	y_6
7	x_7	y_7
8	x_8	y_8
9	x_9	y_9

B



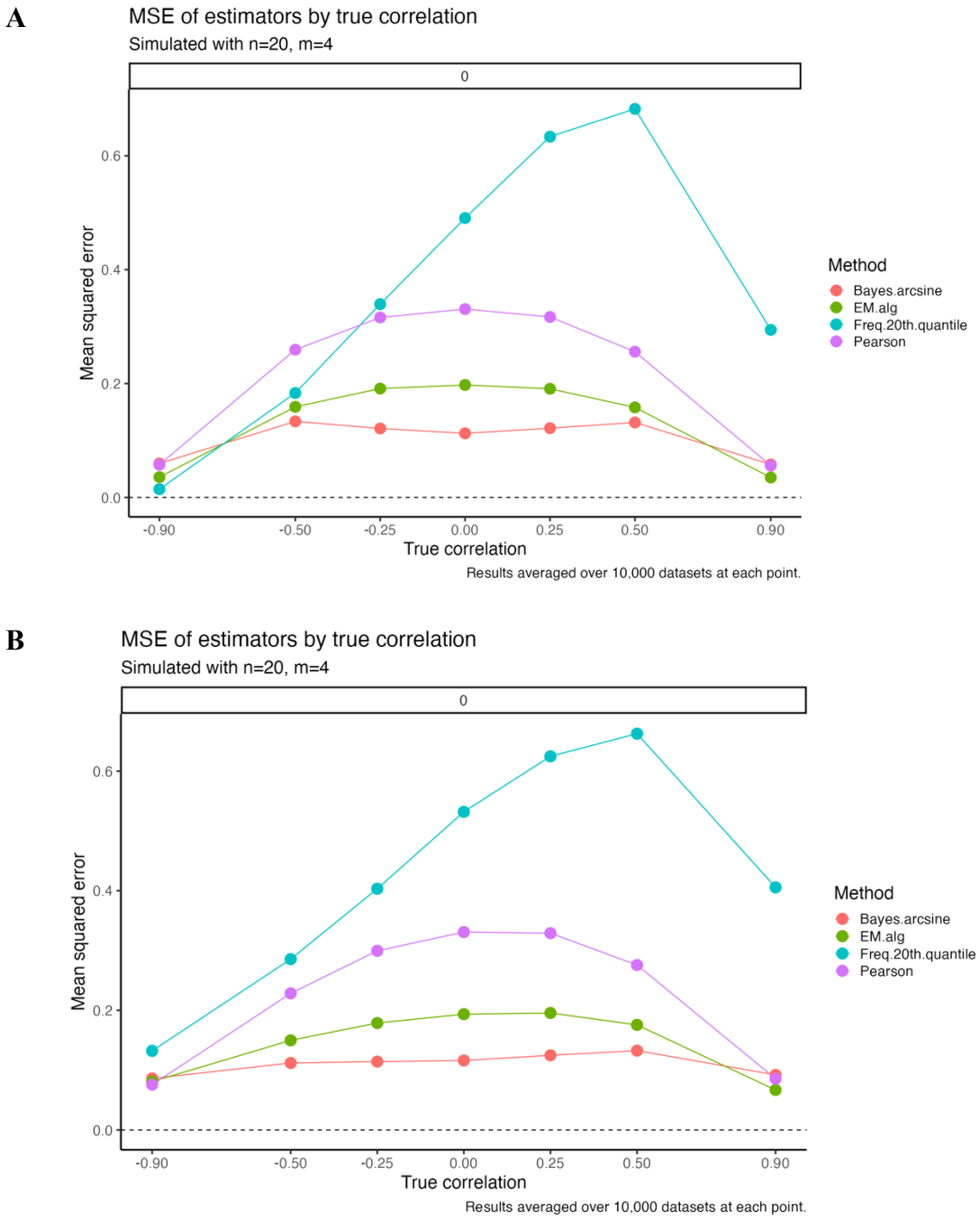
Caption: The dataset in panel (A) is paired and matched. The dataset in panel (B) is paired but unmatched since we cannot identify which observation in X corresponds to its pair in Y.

Figure 2: Simulation results evaluating the bias in estimators of interest.



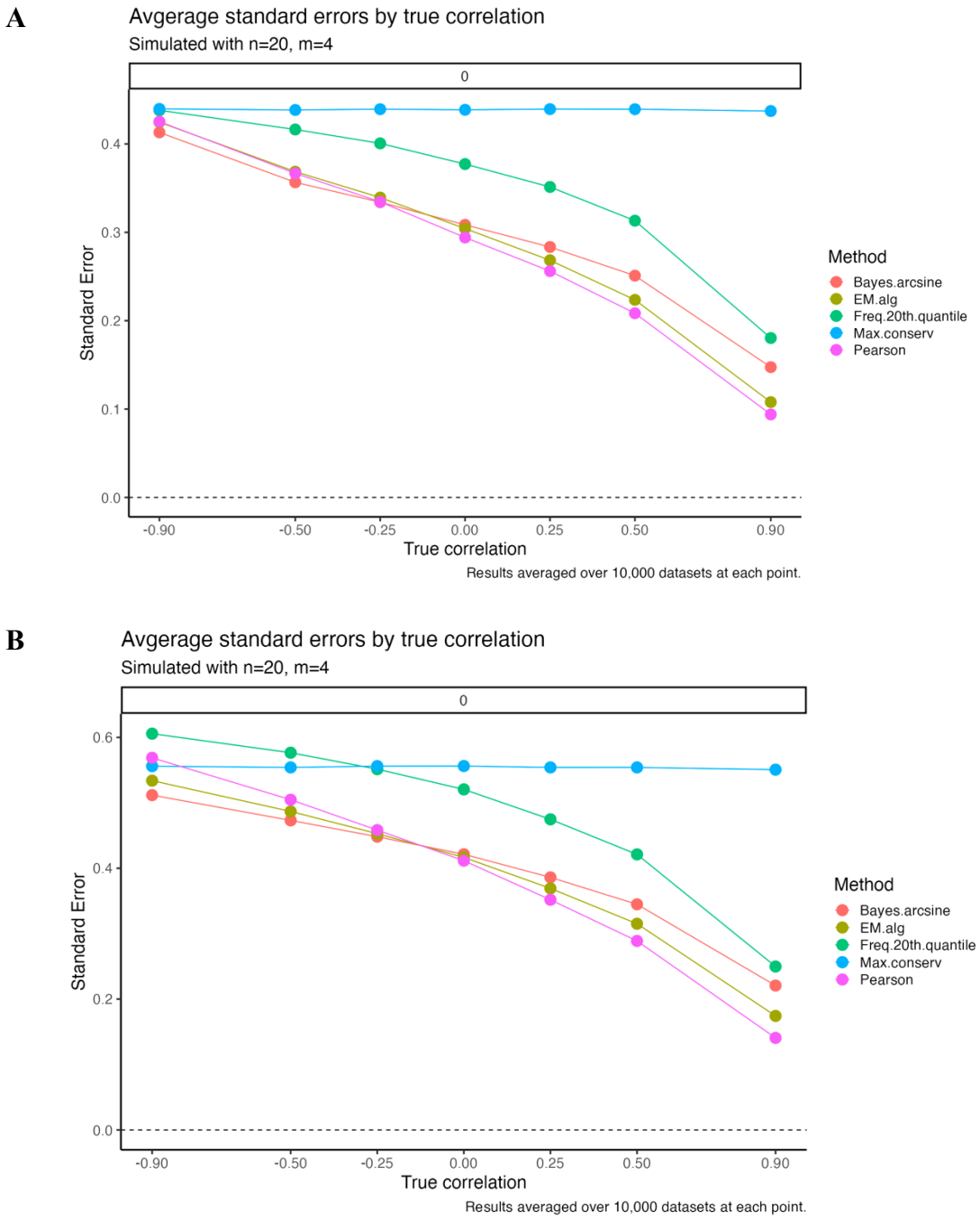
Caption: Datasets in panel (A) were simulated with bivariate normal distributions. Datasets in panel (B) were simulated with bivariate ordinal distributions. Note that the maximally conservative estimator was withheld from these plots to aid visualization.

Figure 3: Simulation results evaluating the mean squared error (MSE) in estimators of interest.



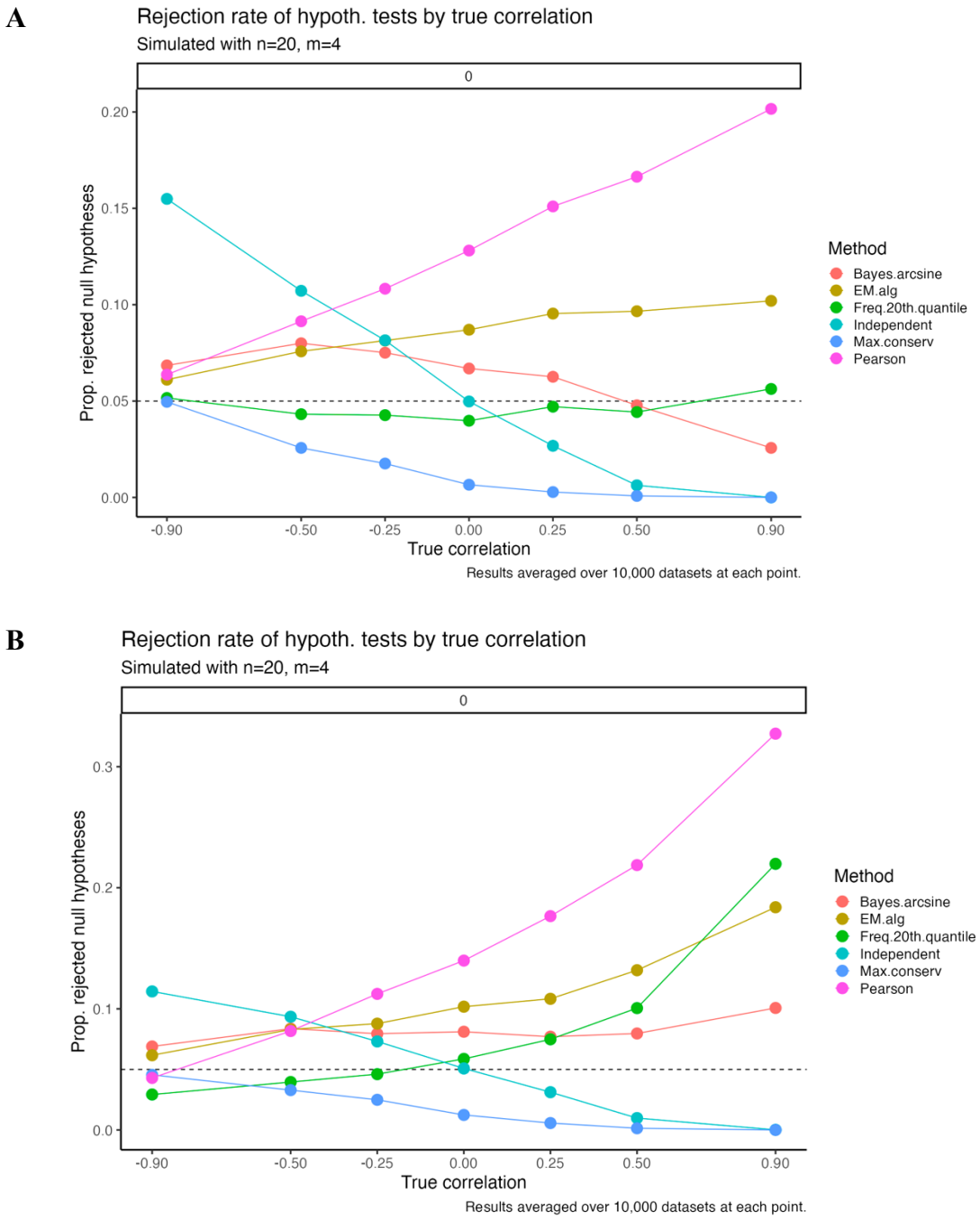
Caption: Datasets in panel (A) were simulated with bivariate normal distributions. Datasets in panel (B) were simulated with bivariate ordinal distributions. Note that the maximally conservative estimator was withheld from these plots to aid visualization.

Figure 4: Simulation results evaluating the standard errors (SE) using estimators of interest.



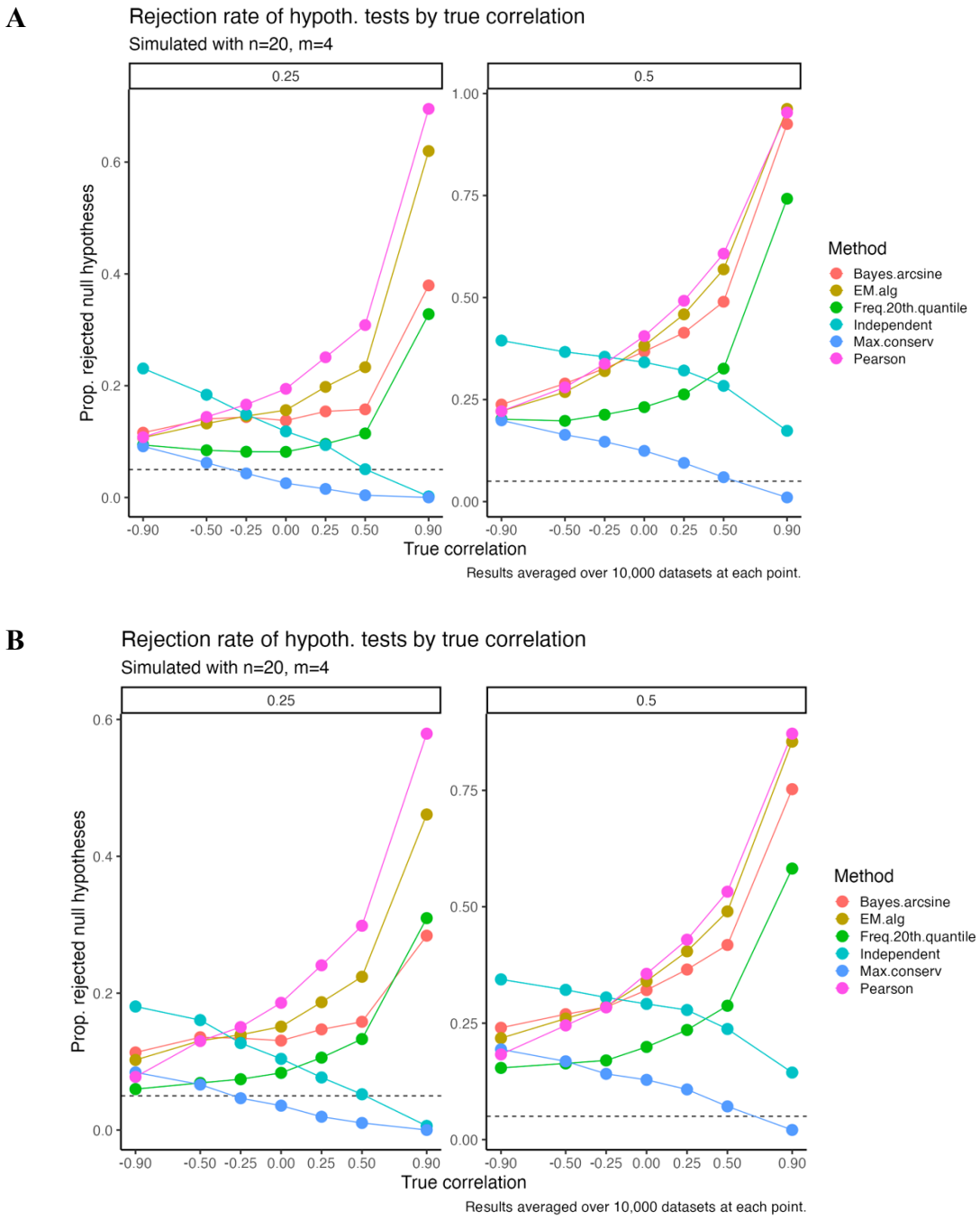
Caption: Datasets in panel (A) were simulated with bivariate normal distributions. Datasets in panel (B) were simulated with bivariate ordinal distributions.

Figure 5: Simulation results evaluating the Type I error for candidate methods of interest.



Caption: Datasets in panel (A) were simulated with bivariate normal distributions. Datasets in panel (B) were simulated with bivariate ordinal distributions. Note that the true difference in means was zero. The line labelled 'Independent' corresponds to Student's t-test.

Figure 6: Simulation results evaluating power for candidate methods of interest.



Caption: Datasets in panel (A) were simulated with bivariate normal distributions. Datasets in panel (B) were simulated with bivariate ordinal distributions. On the left of both panels, the true difference in means was 0.25 standardized units; on the right, the difference was 0.5 standardized unites. Note that the line labelled 'Independent' corresponds to Student's t-test.

Table 1: Results of tests for equality of means in the Intervention Group.

Method	**Correlation**	**T Statistic**	**Std. Error**	**95% CI**		**p-value**
Max. Conservative	-0.906	1.364	0.409	-0.243	1.358	0.175
Pearson	0.418	2.470	0.226	0.115	1.000	**0.015**
20th Quantile	0.102	1.987	0.281	0.008	1.107	0.050
Bayesian	0.427	2.487	0.224	0.118	0.997	**0.014**
EM Algorithm						
Independence (Student's *t*-test)	0	1.883	0.296	-0.023	1.138	0.062

Caption: The mean difference between the pre- and post- intervention outcomes was 0.556 (pre- minus post-), suggesting average decline in outcomes following the intervention. Using the Bayesian estimator of correlation, we would infer a 95% CI of (0.118, 0.997) and conclude a significant change following intervention. With Student's t-test, i.e., assuming no correlation, we would infer a 95% CI of (-0.023, 2.238) and conclude no significant change.

Table 2: Results of tests for equality of means in the Control Group.

Method	**Correlation**	**T Statistic**	**Std. Error**	**95% CI**		**p-value**
Max. Conservative	-0.933	-1.674	0.320	-1.164	0.091	0.097
Pearson	0.059	-2.400	0.223	-0.974	-0.098	**0.018**
20th Quantile	-0.253	-2.079	0.258	-1.042	-0.031	**0.040**
Bayesian	0.058	-2.399	0.224	-0.974	-0.098	**0.018**
EM Algorithm						
Independence (Student's *t*-test)	0	-2.328	0.230	-0.988	-0.085	**0.022**

Caption: The mean difference between the pre- and post- intervention outcomes was -0.536 (pre- minus post-), suggesting average improvement in outcomes following the intervention. Using the Bayesian estimator of correlation, we would infer a 95% CI of (-0.974, -0.098) and conclude a significant change following intervention. With Student's t-test, i.e., assuming no correlation, we would infer a 95% CI of (-0.988, -0.085) and reach the same conclusion.