Table 1. **Illustration of the calculation of the proposed metrics to assess models predicting treatment effect based on matching patients.** The calibration metrics are calculated in the following manner calibration-in-the-large = mean(M)-mean(L) ≈ -0.267, E_{avg}-for-benefit = mean(abs(L-N)) ≈ 0.579, E_{50}-for-benefit = median(abs(L-N)) ≈ 0.624, and E_{90}-for-benefit = quantile(abs(L-N), 0.9) ≈ 0.904. The overall performance are calculated by Cross-entropy-for-benefit $= -\frac{1}{n_p}[I(M=1)\cdot\log[(1-K)J] + I(M=0)\log[(1-K)(1-J)+K\cdot J] + I(M=-1)\log[K(1-J)]] \approx 1.165$ and Brier-for-benefit $= \frac{1}{2n_p}[[(1-K)J - I(M=1)]^2 + [(1-K)(1-J)+K\cdot J - I(M=0)]^2 + [K(1-J) - I(M=-1)]^2] \approx 0.343$. Abbreviations: p_0 = P(Y=1│W=0); p_1=P(Y=1│W=1); LOESS curve is created by predict(loess(M ~ L)), which results in the same values as the observed pairwise treatment effect (M), when rounded to three decimals, due to a small number of observations.

	Patient assigned to treatment				**Patient assigned to control treatment**				**Matched pair**				
Matched patient pair (A)	$\boldsymbol{p_0}$ **(B)**	$\boldsymbol{p_1}$ **(C)**	**Predicted treatment effect (D=B-C)**	**Observed outcome (E)**	$\boldsymbol{p_0}$ **(F)**	$\boldsymbol{p_1}$ **(G)**	**Predicted treatment effect (H=F-G)**	**Observed outcome (I)**	$\boldsymbol{p_0}$ **(J=F)**	$\boldsymbol{p_1}$ **(K=C)**	**Predicted pairwise treatment effect (L=J-K)**	**Observed pairwise treatment effect (M=I-E)**	**LOESS curve (N)**
1	0.136	0.283	-0.147	1	0.162	0.307	-0.145	1	0.162	0.283	-0.121	0	0.412
2	0.246	0.343	-0.097	0	0.218	0.319	-0.101	1	0.218	0.343	-0.125	1	0.589
3	0.156	0.219	-0.063	1	0.142	0.203	-0.061	0	0.142	0.219	-0.077	-1	-0.901
4	0.081	0.083	0.002	0	0.098	0.062	0.036	0	0.098	0.083	0.015	0	0.081
5	0.345	0.212	0.133	1	0.299	0.171	0.128	0	0.299	0.212	0.087	-1	-0.937
6	0.421	0.390	0.031	1	0.561	0.255	0.306	1	0.561	0.390	0.171	0	-0.190
7	0.364	0.201	0.163	1	0.243	0.164	0.079	0	0.243	0.201	0.042	0	-0.217
8	0.264	0.199	0.065	1	0.345	0.278	0.067	0	0.345	0.199	0.146	-1	-0.707