While propensity score (PS) matching and reweighting are widely used in many disciplines to balance characteristics across comparison groups, there is a lack of guidance regarding their use with single arm trials, where datasets are typically smaller than those used in observational studies.

The goal of this analysis is to illustrate a general framework for the use of PS methods for indirect treatment comparisons and to compare the performance of common and emerging PS algorithms using Monte Carlo simulations.

A general framework for propensity score analysis with single-arm trials

Propensity score methods evaluated in Monte Carlo simulations

Monte Carlo studies with scenarios defined by treatment arms of varying population sizes and overlap in baseline characteristics.

Simulation Results

Heat map of numeric results reported left for Bias (smaller is better), coverage (proximity to 95% is better), and Mean Squared Error (smaller is better). The top-ranked algorithm for each case is reported above, with the second-ranked algorithm listed in parentheses.

The target estimand was the average treatment effect on the treated (ATET). The "true" treatment effect that was used to simulate the data had an ATET of 0.5 on a mean difference scale. Results are based on 1000 simulations.

CONCLUSIONS

Based on these simulations, entropy balancing had superior bias reduction and ranked highly on MSE and coverage. IPTW also yielded favorable bias reduction in all scenarios. NNM had favorable MSE but high bias in most scenarios; its coverage was 100% in all scenarios, suggesting that caliper matching may reduce statistical power when comparing single-arm trials.