E. Strobl[1], T. Lasko[2]

Synthesized difference in differences

Departments of [1]Psychiatry and Behavioral Sciences and
[2]Biomedical Informatics, VUMC

Randomized clinical trials (RCTs) eliminate confounding but impose strict exclusion criteria that prevent sampling of the entire clinical population. Observational (OBS) datasets are more inclusive but suffer from confounding. Difference in Differences (DD) eliminates confounding from OBS data by comparing outcomes before and after treatment administration. The algorithm however requires a parallel slope assumption that may not apply in practice when confounding shifts across time. In this paper, we propose Synthesized Difference in Differences (SDD) that infers the correct (possibly non-parallel) slope by linearly adjusting DD using additional RCT data. The algorithm achieves state of the art performance across multiple synthetic and real datasets even when the RCT excludes the majority of patients.