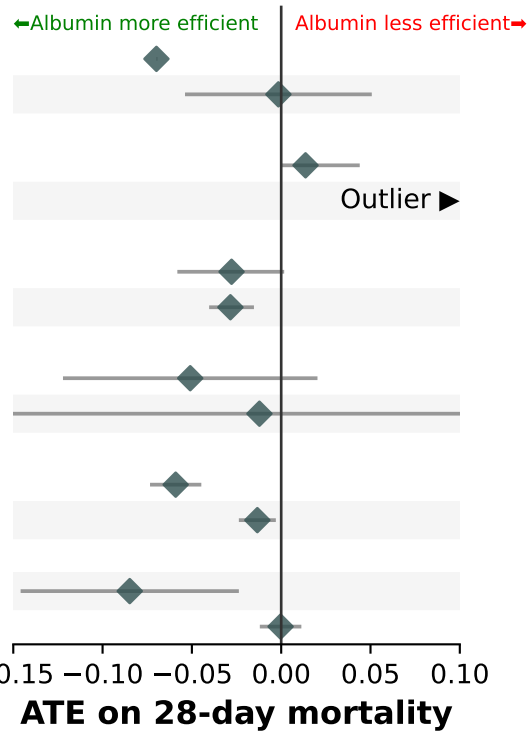


ATE (95% bootstrap confidence interval)



Difference in mean

RCT Gold Standard (Caironi et al. 2014)

-0.07(-0.07 to -0.07)

-0.00(-0.05 to 0.05)

Propensity Score Matching

Est=Regularized LR

0.01(-0.00 to 0.04)

Est=Forests

0.67(0.67 to 1.48)

Inverse Propensity Weighting

Est=Regularized LR

-0.03(-0.06 to 0.00)

Est=Forests

-0.03(-0.04 to -0.02)

Outcome model (Tlearner)

Est=Regularized LR

-0.05(-0.12 to 0.02)

Est=Forests

-0.01(-0.21 to 0.18)

Double Machine Learning

Est=Regularized LR

-0.06(-0.07 to -0.04)

Est=Forests

-0.01(-0.02 to -0.00)

Doubly Robust (AIPW)

Est=Regularized LR

-0.08(-0.15 to -0.02)

Est=Forests

-0.00(-0.01 to 0.01)