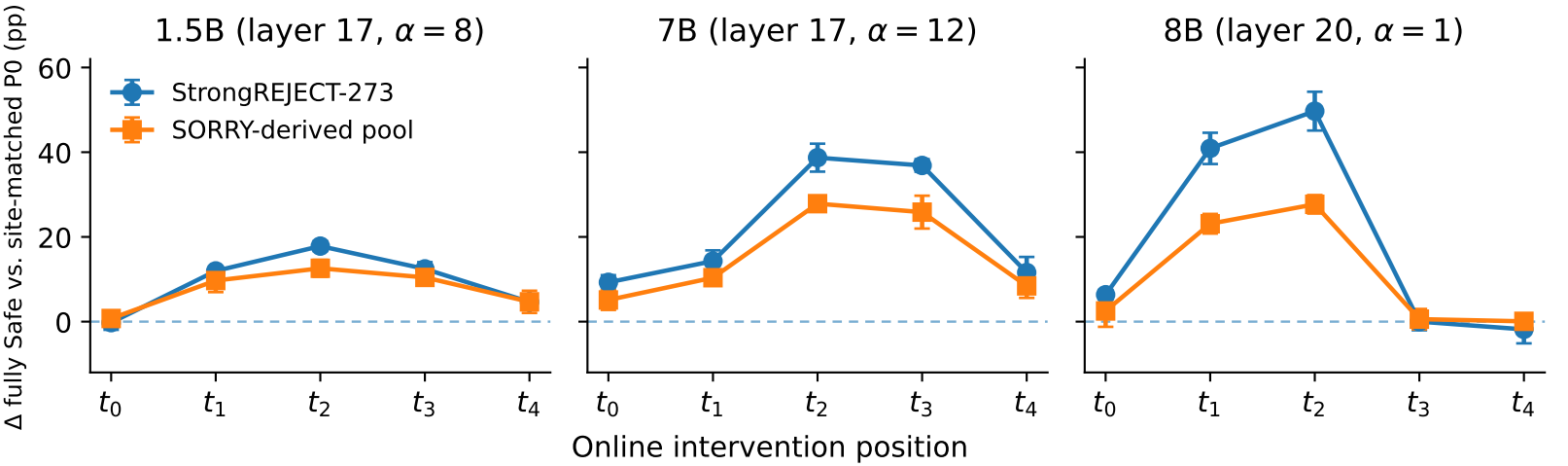

1.5B (layer 17, $\alpha = 8$)
7B (layer 17, $\alpha = 12$)
8B (layer 20, $\alpha = 1$)
StrongREJECT-273
SORRY-derived pool
Δ fully Safe vs. site-matched P0 (pp)
60
40
20
0
t_0
t_1
t_2
t_3
t_4
Online intervention position