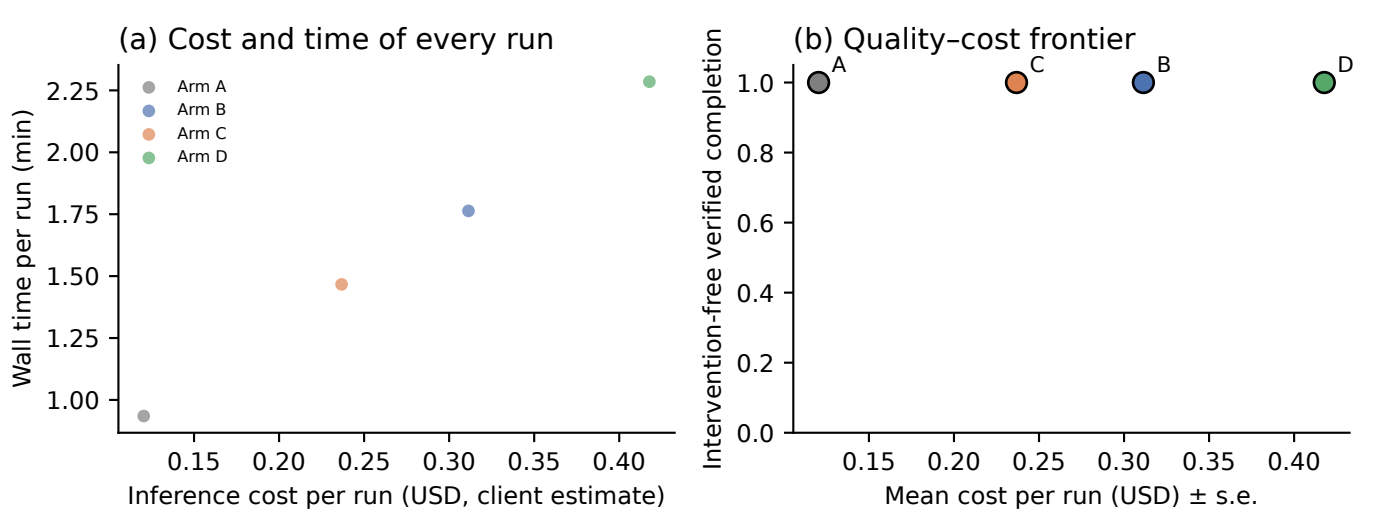

(a) Cost and time of every run
Arm A
Arm B
Arm C
Arm D
Wall time per run (min)
2.25
2.00
1.75
1.50
1.25
1.00
0.15
0.20
0.25
0.30
0.35
0.40
Inference cost per run (USD, client estimate)
(b) Quality–cost frontier
A
C
B
D
Intervention-free verified completion
1.0
0.8
0.6
0.4
0.2
0.0
0.15
0.20
0.25
0.30
0.35
0.40
Mean cost per run (USD) ± s.e.