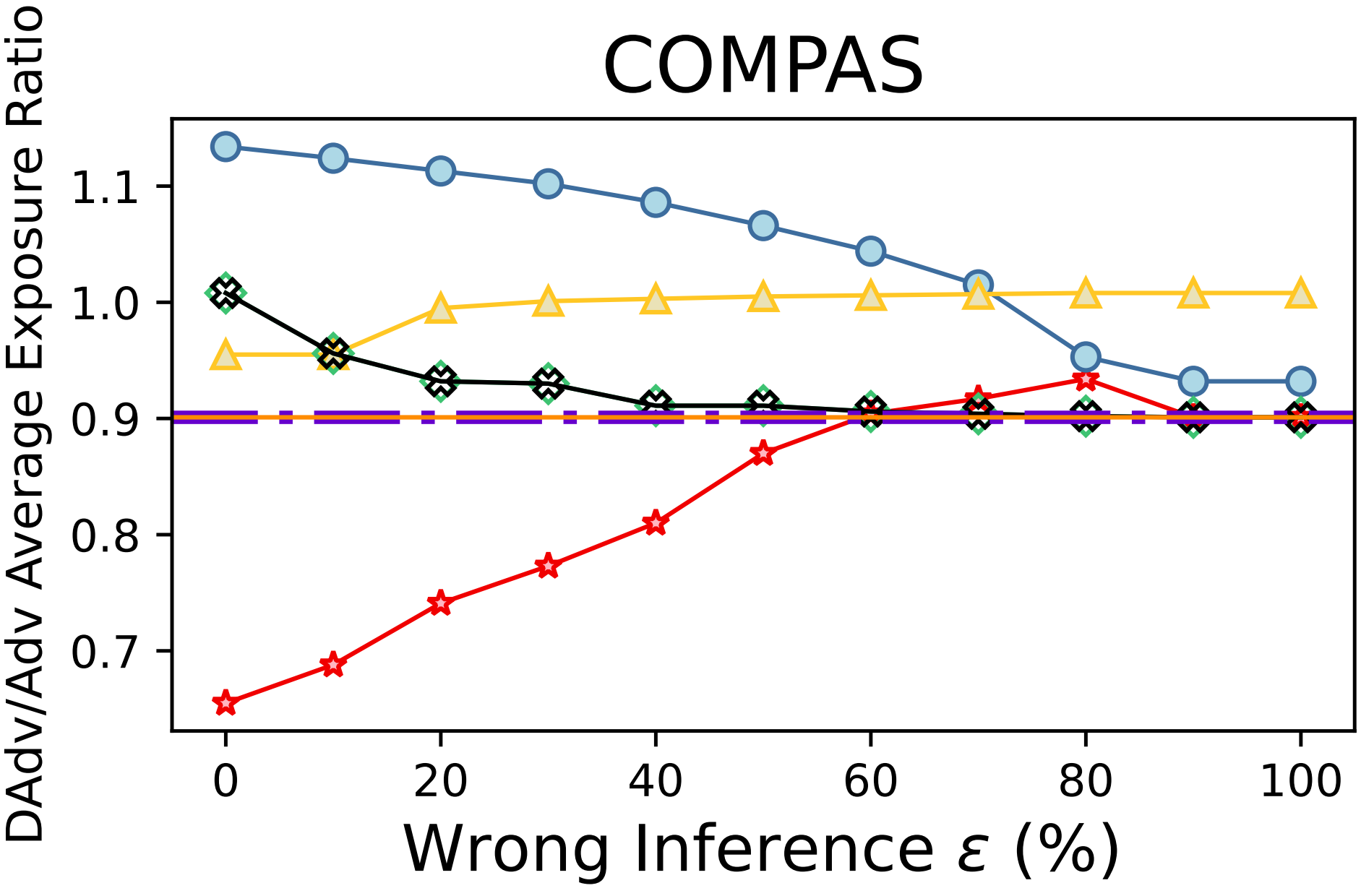
COMPAS
DAdv/Adv Average Exposure Ratio
1.1
1.0
0.9
0.8
0.7
0
20
40
60
80
100
Wrong Inference ε (%)