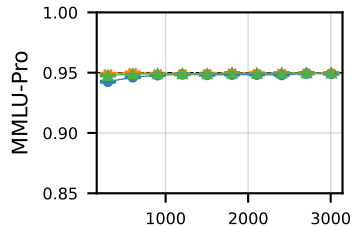
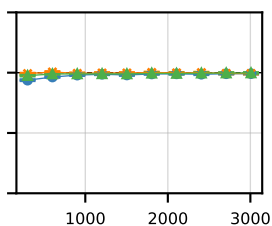


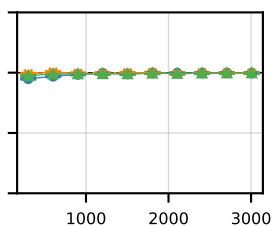
800 Fully-Obs. Rows
10% Obs. Entries



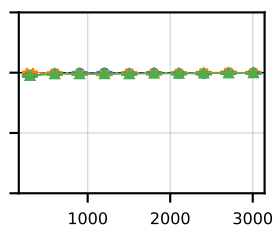
200 Fully-Obs. Rows
10% Obs. Entries



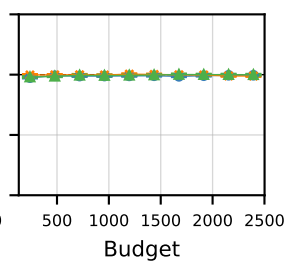
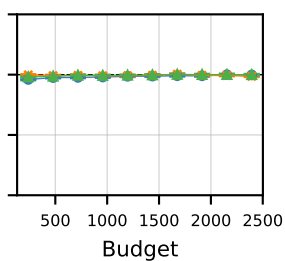
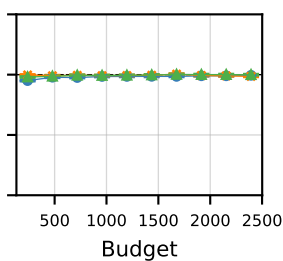
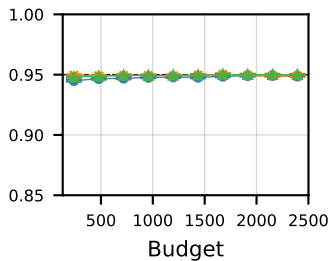
50 Fully-Obs. Rows
10% Obs. Entries



0 Fully-Obs. Rows
0.1% Obs. Entries



BBH+GPQA+IFEval
+MATH+MuSR



● FAQ ✕ Best Baseline Per Budget ▲ Uniform - - - - 95% Coverage (Bottom Subplots)