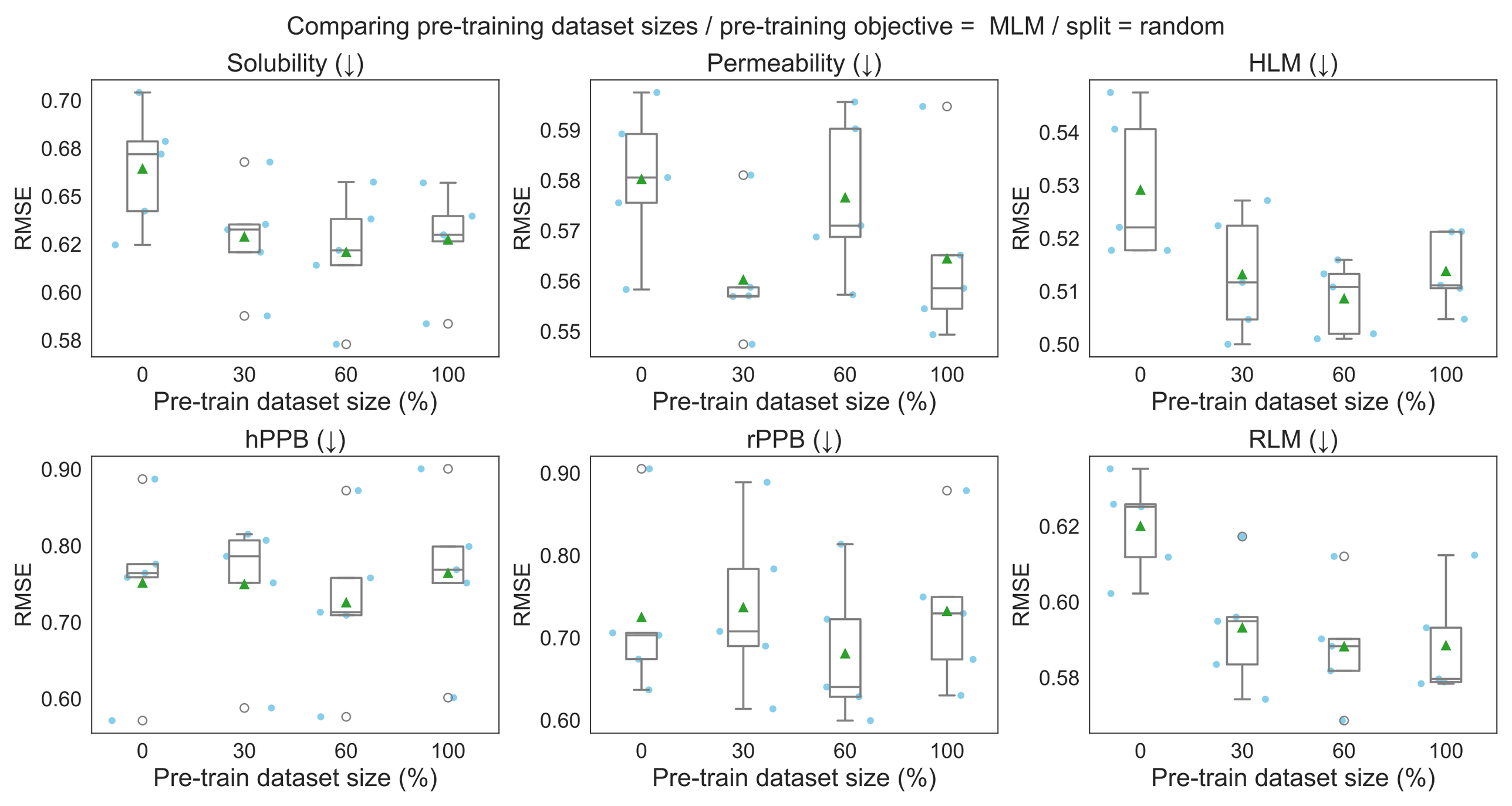
Comparing pre-training dataset sizes / pre-training objective = MLM / split = random
Solubility (↓)
Permeability (↓)
HLM (↓)
hPPB (↓)
rPPB (↓)
RLM (↓)
RMSE
Pre-train dataset size (%)
0
30
60
100
0.70
0.68
0.65
0.62
0.60
0.58
0.59
0.57
0.56
0.55
0.54
0.53
0.52
0.51
0.50
0.90
0.80
0.70
0.60