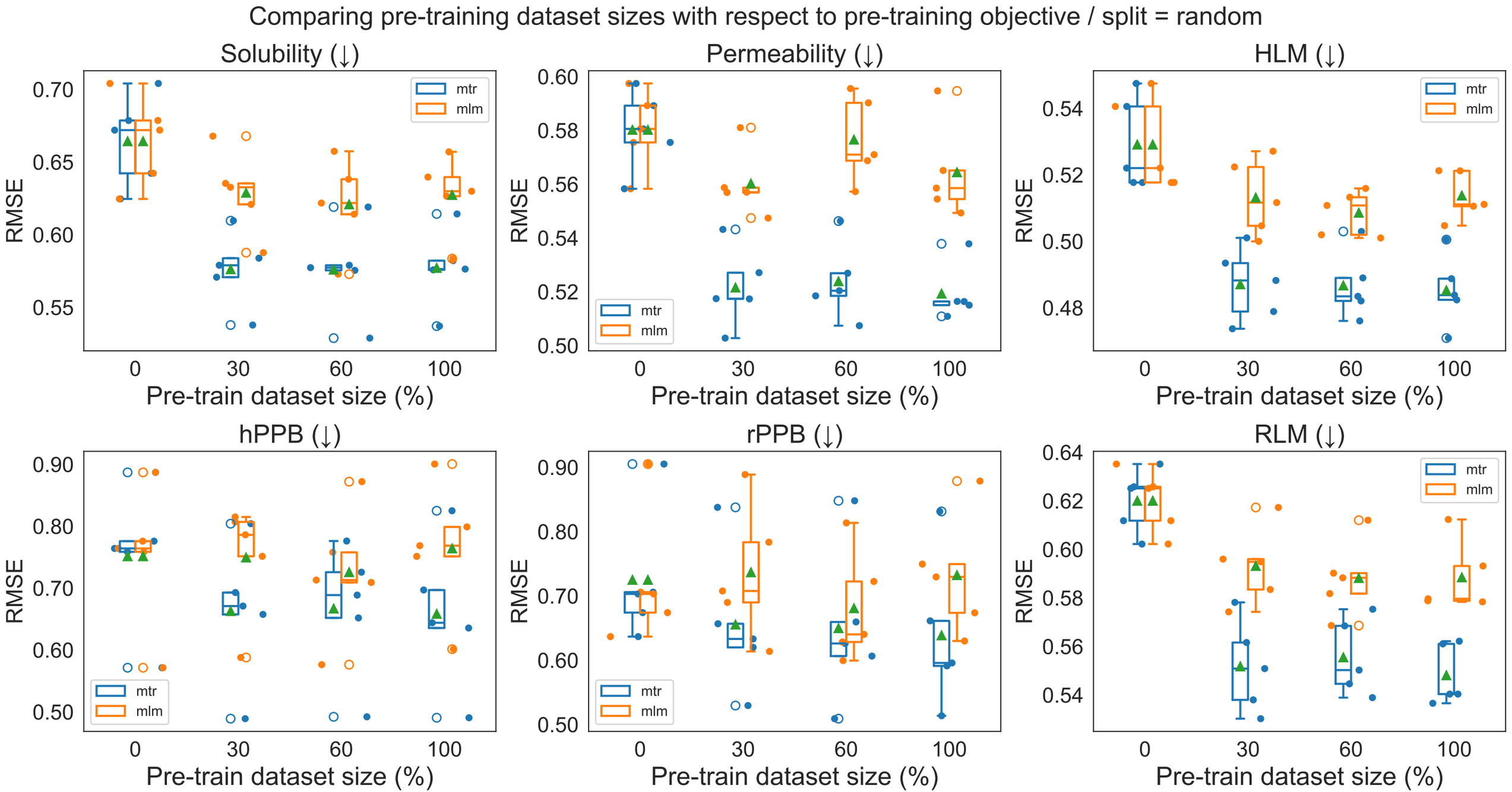
Comparing pre-training dataset sizes with respect to pre-training objective / split = random
Solubility (↓)
Permeability (↓)
HLM (↓)
hPPB (↓)
rPPB (↓)
RLM (↓)
RMSE
Pre-train dataset size (%)
0
30
60
100
mtr
mlm
0.70
0.65
0.60
0.55
0.58
0.56
0.54
0.52
0.50
0.48
0.90
0.80
0.70
0.64
0.62