Number of training steps as a function of the learning rate 1000 •••• epoch size = 10 $\triangleleft \triangleleft \triangleleft \triangleleft$ epoch size = 50 Average number of training steps for 10 samples epoch size = 100 ▼▼▼ epoch size = 200 800 •••• epoch size = 500▶ ▶ ▶ ▶ epoch size = 1000 ◆◆◆◆ epoch size = 2000 600 400 200 9.00 0.05 0.10 0.15 0.20

Learning rate