Test set metrics for optimizing a neural network observable (# bins = 20)

training strategy:
- $p_0$+CL$_s$
- BCE
- $\sigma_\mu$ (Fisher info)
- CL$_s$
- $p_0$