Y1: performance of single uncalibrated vs calibrated model on subgroup-by-sex: male uncalibrated-one-model-for-all calibrated-one-model-for-all 100 -93.6 94.2 88.9 80 70.4 Score(%) 60 · 52.8 50.8 42.2 40 35.6 20 f1 precision recall accuracy