Y2: performance of single uncalibrated vs calibrated model on subgroup-by-sex: male uncalibrated-one-model-for-all calibrated-one-model-for-all 100 89.8 90.3 85.5 86.2 81.6 82.5 79.3 80.4 80 Score(%) 60 40 20 f1 precision recall accuracy