Investigating the infection rates in counties categorized under the CDC community transmission levels

Sina Mokhtar

The **AUROC** (Area Under the Receiver Operating Characteristic) is a metric to evaluate the prediction performance. A value closer to 1 suggests a better prediction ability, while a value close to 0.5 indicates random prediction.

To compare the next 21 days expected infection rates with the actual infection rates , we randomly sampled atleast %20 of counties from each UR category risk level for each day. We then compared their infection rates number with other counties of different risk levels in three weeks later.

The expected outcome (0 or 1) was determined based on the following revised criteria:

- If a county is categorized as low-risk, we expect its infection rate 3 weeks later to be lower than the other counties.

- If a county is categorized as medium-risk, we expect its infection rate 3 weeks later to be higher than low-risk counties and lower than substantial and high-risk counties.

- If a county is categorized as substantial-risk, we expect its infection rate 3 weeks later to be lower than high-risk counties and higher than low and medium-risk counties.

- If a county is categorized as high-risk, we expect its infection rate 3 weeks later to be higher than the other counties.

```
##  truePositive falsePositive  trueNegative falseNegative
##       2602464       6495197       2633498       6240380
```

## ROC Curve for Evaluating CDC Transmission Risk Level
AUROC: 0.29
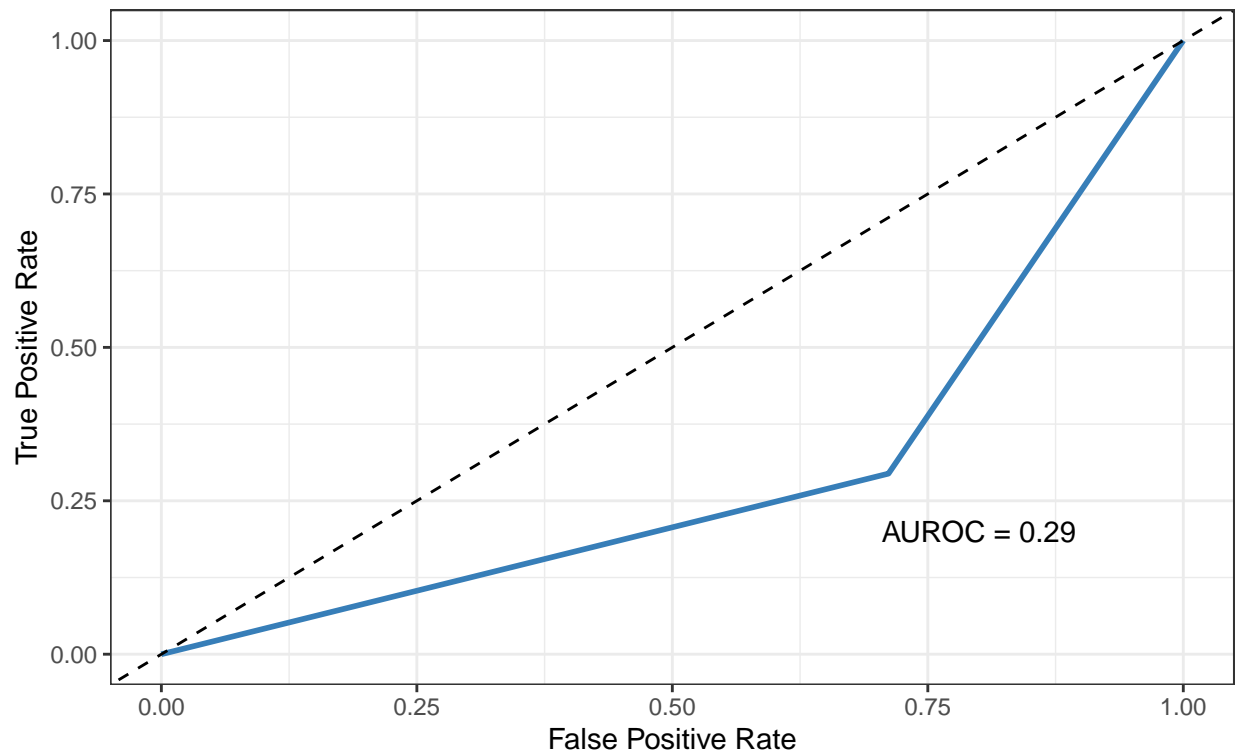
In this case an AUC of 29% indicates that the predictive model has limited power, as the AUC is relatively low. The AUC of 29% indicates that the predictive model is not even work as good as a random model. Based on this information, it can be concluded that the CDC Transmission Risk Level performance in predicting the infection rate is not accurate

Then we calculated the AUROC in each NCHS category.

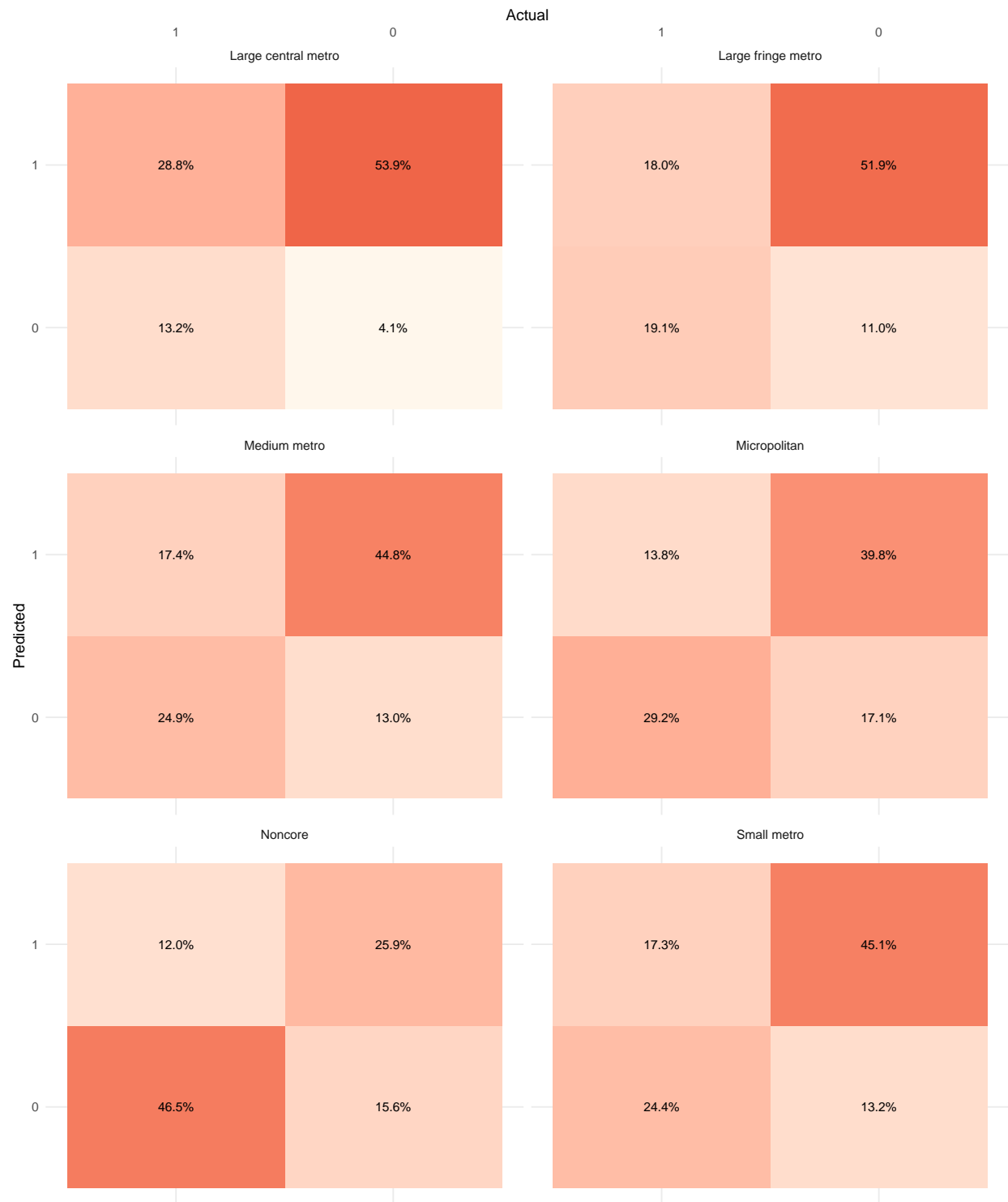## ROC Curve for Evaluating CDC Transmission Risk Level
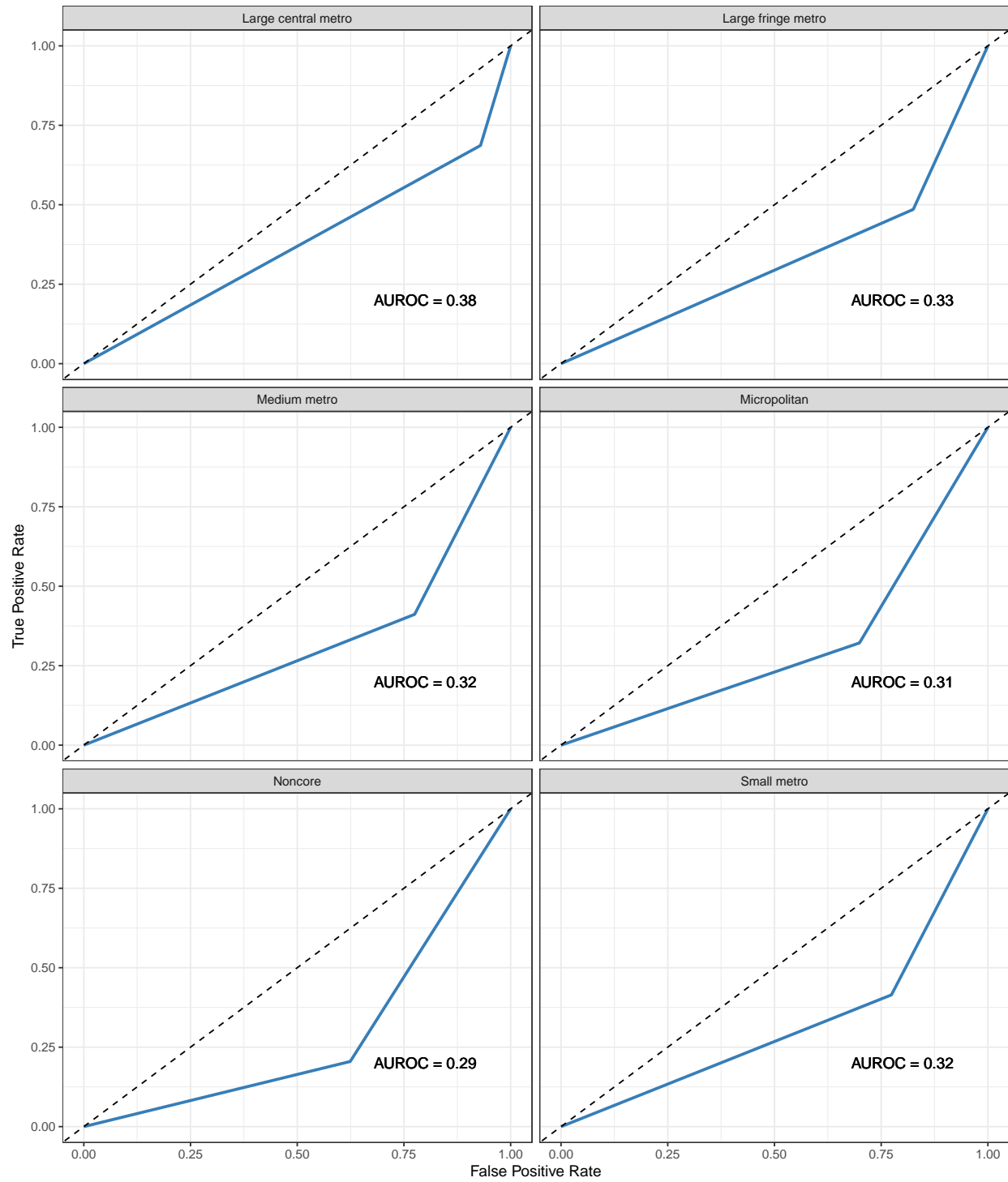
Confusion Matrix for all counties
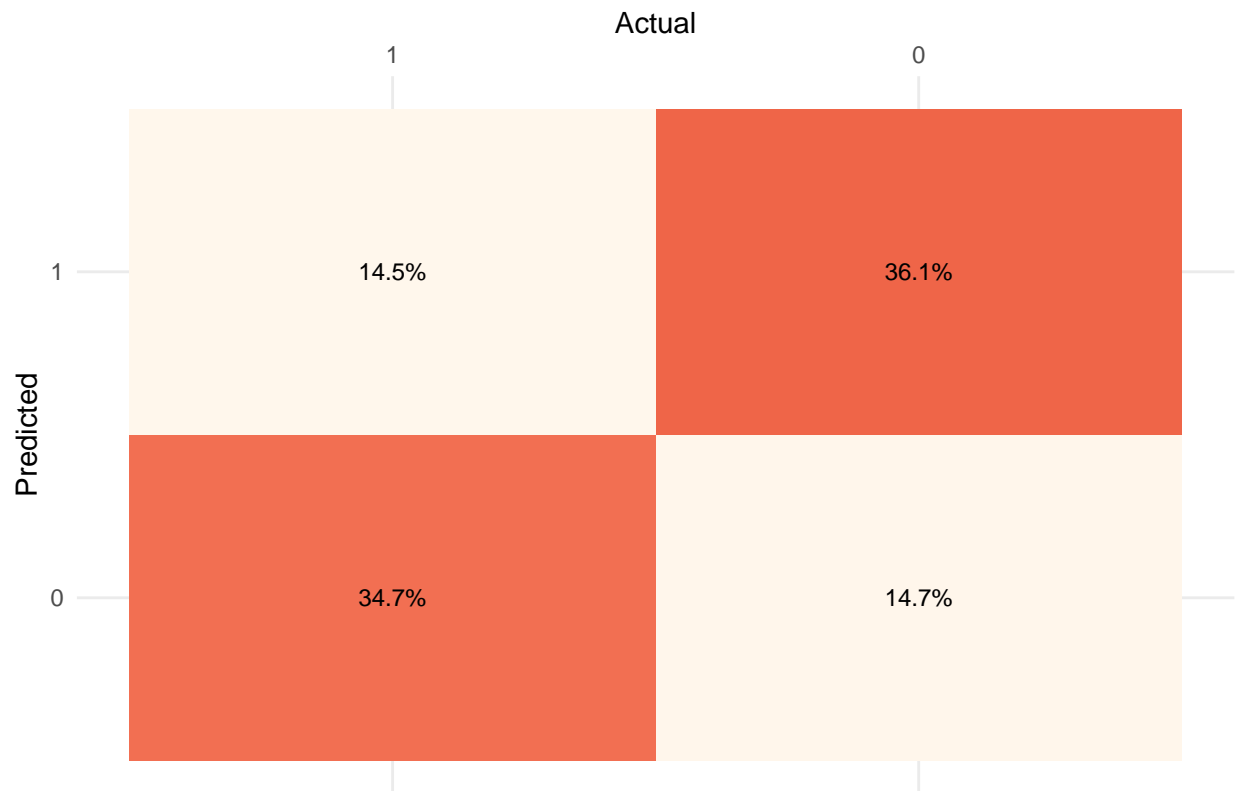
Actual

- True Positives (TP): In 317117 cases we expected a higher infection rate number in 3 weeks later compared to other counties, and the actual infection rate number was indeed higher in those cases.

- False Positives (FP): In 509605 cases we expected a higher infection rate number in 3 weeks later compared to other counties, but the actual infection rate number was lower in those cases.

- True Negatives (TN): In 155791 cases we expected a lower infection rate number in 3 weeks later compared to other counties, and the actual infection rate number was indeed lower in those cases.

- False Negatives (FN): In 199487 cases we expected a lower infection rate number in 3 weeks later compared to other counties, but the actual infection rate number was higher in those cases.

Accuracy: The overall accuracy of the model is 0.29, which means the model correctly identified 29% of the cases.

Then we calculated the confusion matrix in each NCHS category.

Actual

	1	0		1	0

Large central metro

	1	28.8%	53.9%
	0	13.2%	4.1%

Large fringe metro

	1	18.0%	51.9%
	0	19.1%	11.0%

Predicted

Medium metro

	1	17.4%	44.8%
	0	24.9%	13.0%

Micropolitan

	1	13.8%	39.8%
	0	29.2%	17.1%

Noncore

	1	12.0%	25.9%
	0	46.5%	15.6%

Small metro

	1	17.3%	45.1%
	0	24.4%	13.2%

RESULTS

References