The Infection Rate Prediction Based on CDC Transmistion Risk Level

Sina Mokhtar

AUROC

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.

Method

To compare the next 21 days expected infection rates with the actual infection rates , we randomly sampled 10 counties from each CDC transmission 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 that of 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 that of other counties.

ROC Curves for Evaluate the CDC Transmission Risk Level
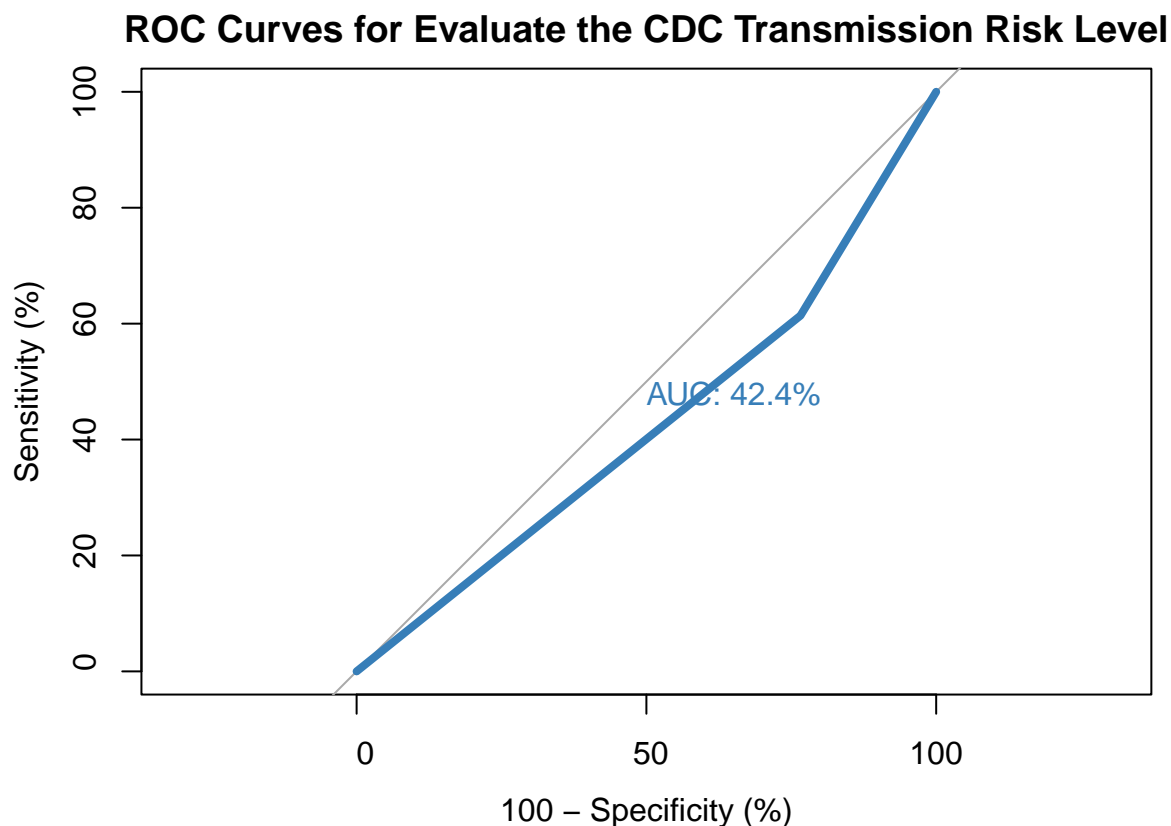

```
##
## Call:
## roc.data.frame(data = compared_counties, response = actual_higher_rt,    predictor = expected_highe
##
## Data: expected_higher_rt in 665396 controls (actual_higher_rt 0) < 516604 cases (actual_higher_rt 1)
## Area under the curve: 42.4%
```

Conclusion

In this case an AUC of 42.41% indicates that the predictive model has limited power, as the AUC is relatively low. The AUC of 42.41% 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

Confusion Matrix

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction      0      1
##          0 155791 199487
##          1 509605 317117
##
##               Accuracy : 0.4001
```

```
##                  95% CI : (0.3992, 0.401)
##     No Information Rate : 0.5629
##     P-Value [Acc > NIR] : 1
##
##                   Kappa : -0.1425
##
##  Mcnemar's Test P-Value : <2e-16
##
##             Sensitivity : 0.2341
##             Specificity : 0.6138
##          Pos Pred Value : 0.4385
##          Neg Pred Value : 0.3836
##              Prevalence : 0.5629
##          Detection Rate : 0.1318
##    Detection Prevalence : 0.3006
##       Balanced Accuracy : 0.4240
##
##        'Positive' Class : 0
##
```

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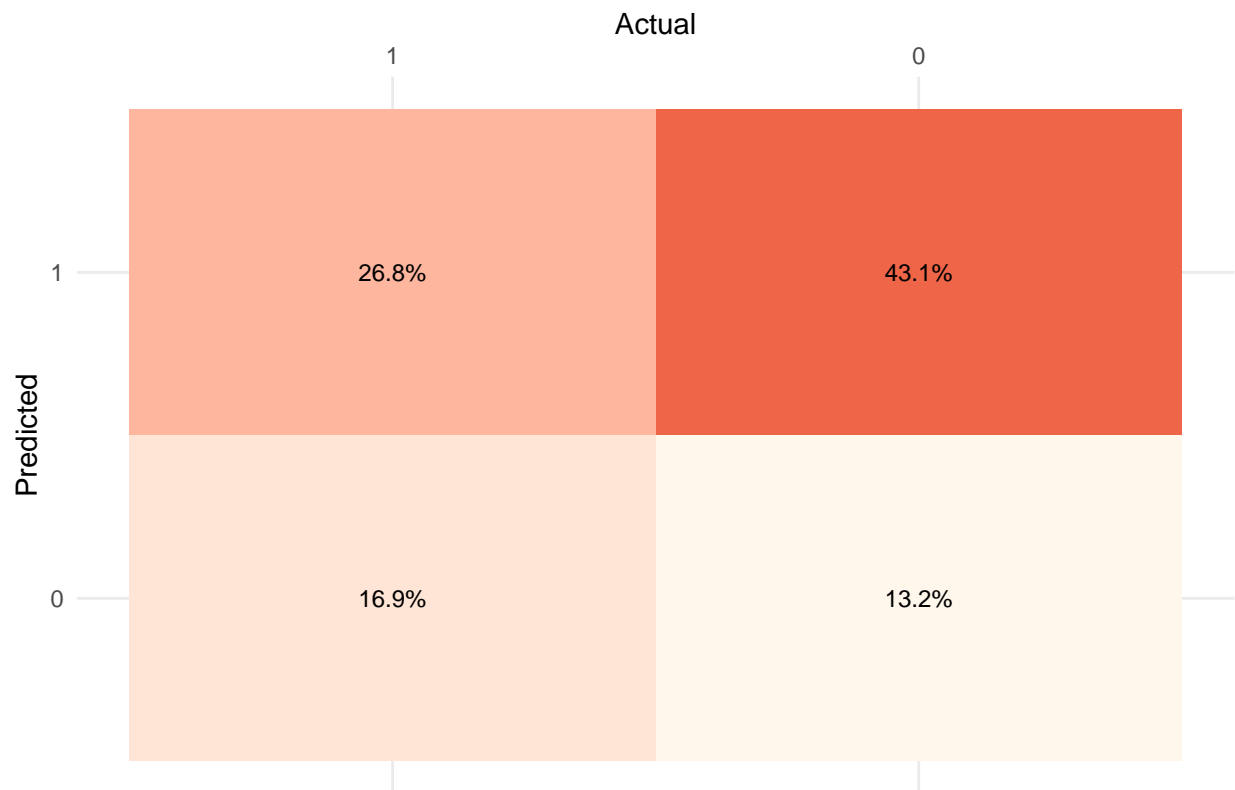
- **True Positives (TP):** In 317117 cases we expected a higher infection rate number in 3 weeks later compared to other counties, and the actual infection rate number was indeed higher in those cases.

- **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 0.4001, which means the model correctly identified 40.01% of the cases.