# The Forecast of COVID-19 Spread Risk at The County Level

**Murtadha Hssayeni**
Florida Atlantic University College of Engineering and Computer Science https://orcid.org/0000-0002-8588-4639
**Arjuna Chala**
LexisNexis Risk Solutions
**Roger Dev**
LexisNexis Risk Solutions
**Lili Xu**
LexisNexis Risk Solutions
**Jesse Shaw**
LexisNexis Risk Solutions
**Borko Furht**
Florida Atlantic University College of Engineering and Computer Science
**Behnaz Ghoraani** (bghoraani@fau.edu)
Florida Atlantic University College of Engineering and Computer Science https://orcid.org/0000-0003-0075-7663

---

**Research**

**Keywords:** COVID-19 Forecast, Deep Learning, Mobility, County Demographics

**Posted Date:** April 19th, 2021

**DOI:** https://doi.org/10.21203/rs.3.rs-415377/v1



---

**Version of Record:** A version of this preprint was published at Journal of Big Data on July 7th, 2021. See the published version at https://doi.org/10.1186/s40537-021-00491-1.

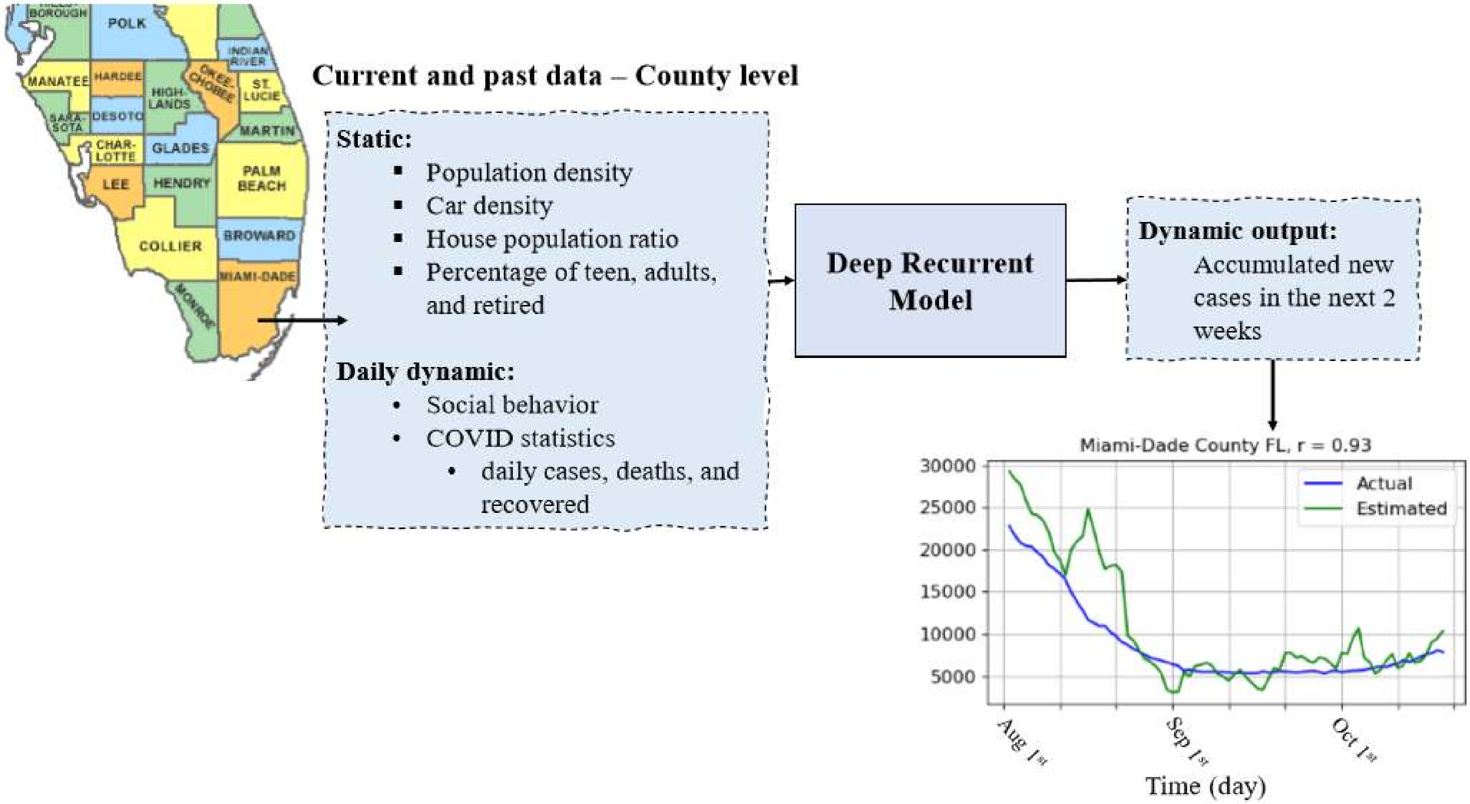
## Abstract

The early detection of the coronavirus disease 2019 (COVID-19) outbreak is important to save people's lives and restart the economy quickly and safely. People's social behavior as captured by their mobility data plays a role in spreading the disease. Therefore, we used the daily mobility data aggregated at the county level beside COVID-19 statistics and demographic information for short-term forecasting of COVID-19 outbreak in the United States. The daily data are fed to a deep model based on Long Short-Term Memory (LSTM) to predict the accumulated number of COVID-19 cases in the next two weeks. A significant average correlation was achieved (r=0.83 (p=0.005)) between the model prediction and the actual accumulated cases in the interval from August 1, 2020 until January 22, 2021. The model predictions had r > 0.7 for 87% of the counties across the United States. Lower correlation was reported for the counties with a total cases of <1,000 during the test interval. The average mean absolute error (MAE) was 605.4, and it was decreasing with the decrease in the total number of cases during the testing interval. The model was able to capture the effect of government responses on COVID-19 cases. Also, it was able to capture the effect of age demographics on the COVID-19 spread where average daily cases decrease with the decrease in retires percentage, and increase with the increase in young percentage. Lessons learned from this study not only can help with managing the COVID-19 pandemic but also could also help with early and effective management of possible future pandemics.


## Full Text

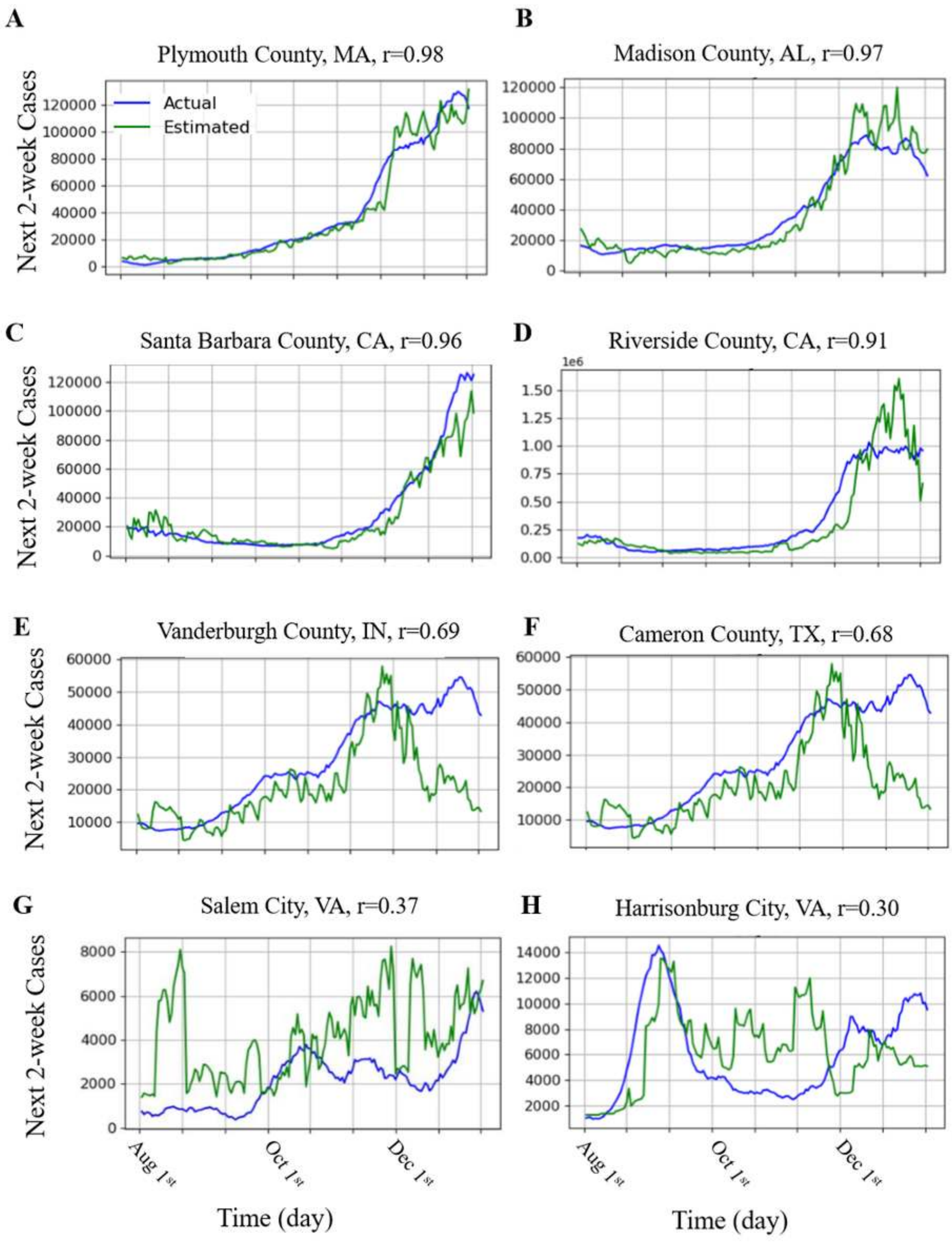
This preprint is available for download as a PDF.

## Figures



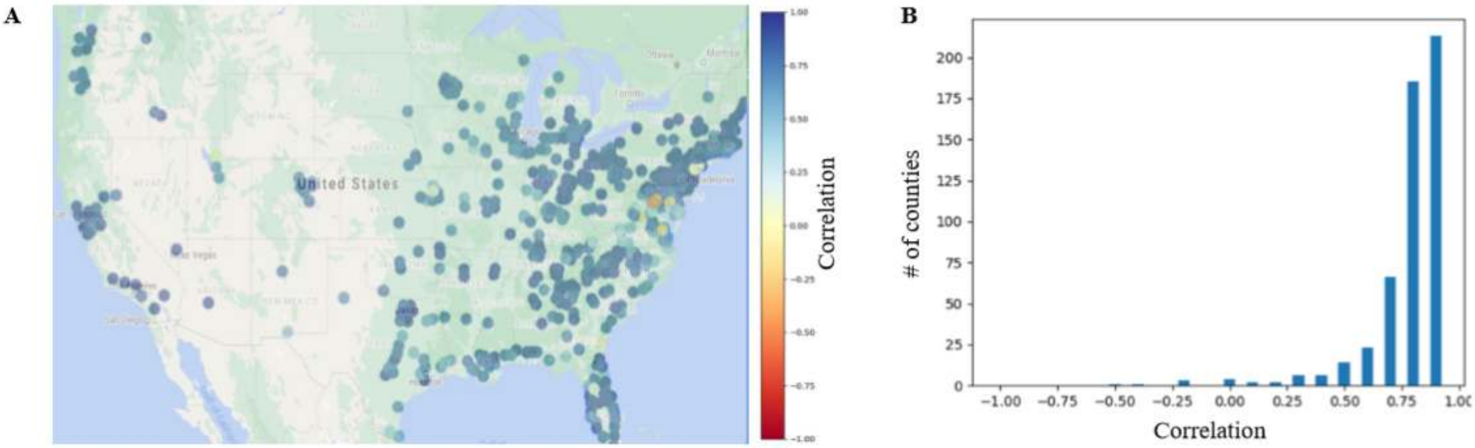

## Figure 1

The overall diagram of the proposed method to forecast accumulated new cases in the next two weeks. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.



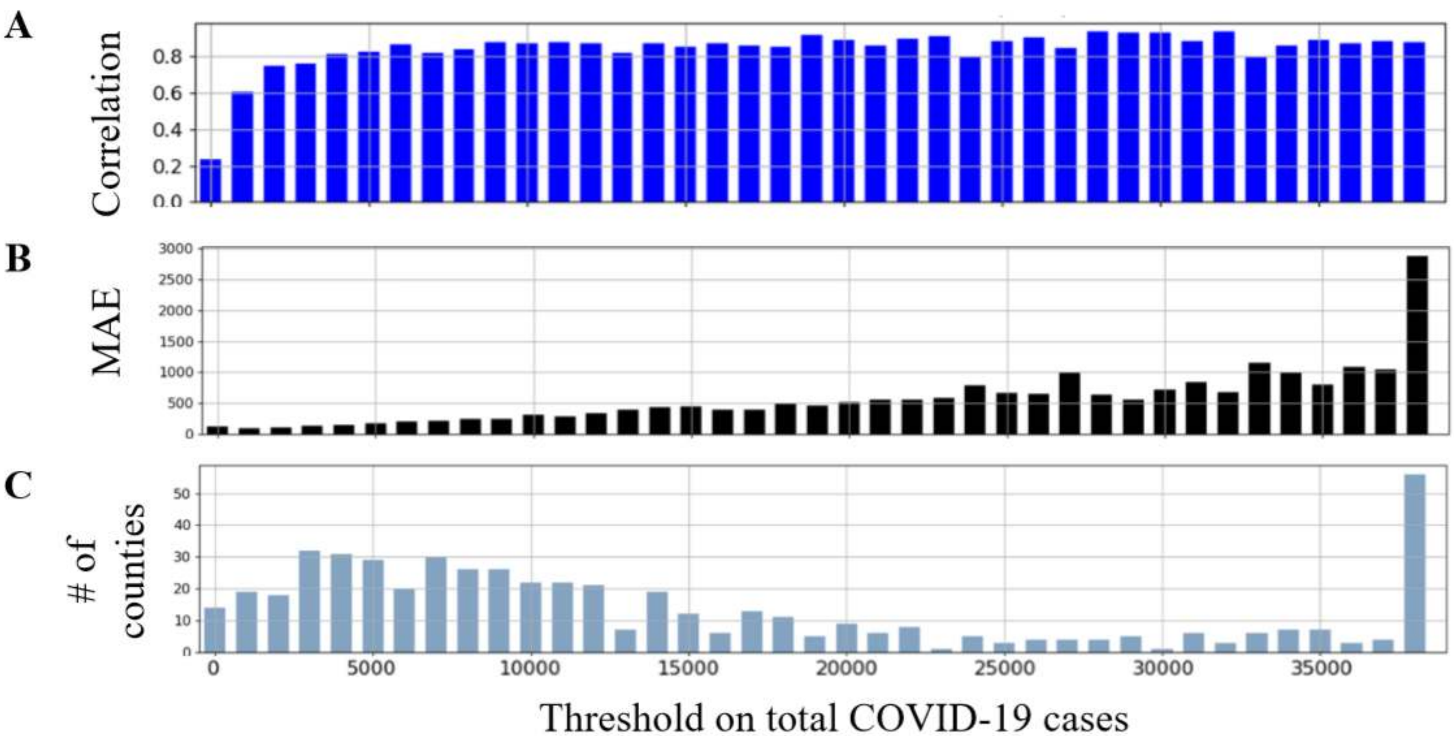

## Figure 2

The estimate and actual accumulated COVID-19 cases for eight counties. A, B, C and D shows accumulated cases of counties for which our model provides a high correlation > 0.9. A and C show increase in the number of cases v.s. B and D show decrease in the number of cases. E, F, G and H shows accumulated cases of counties for which our model provides a moderate and low correlation.



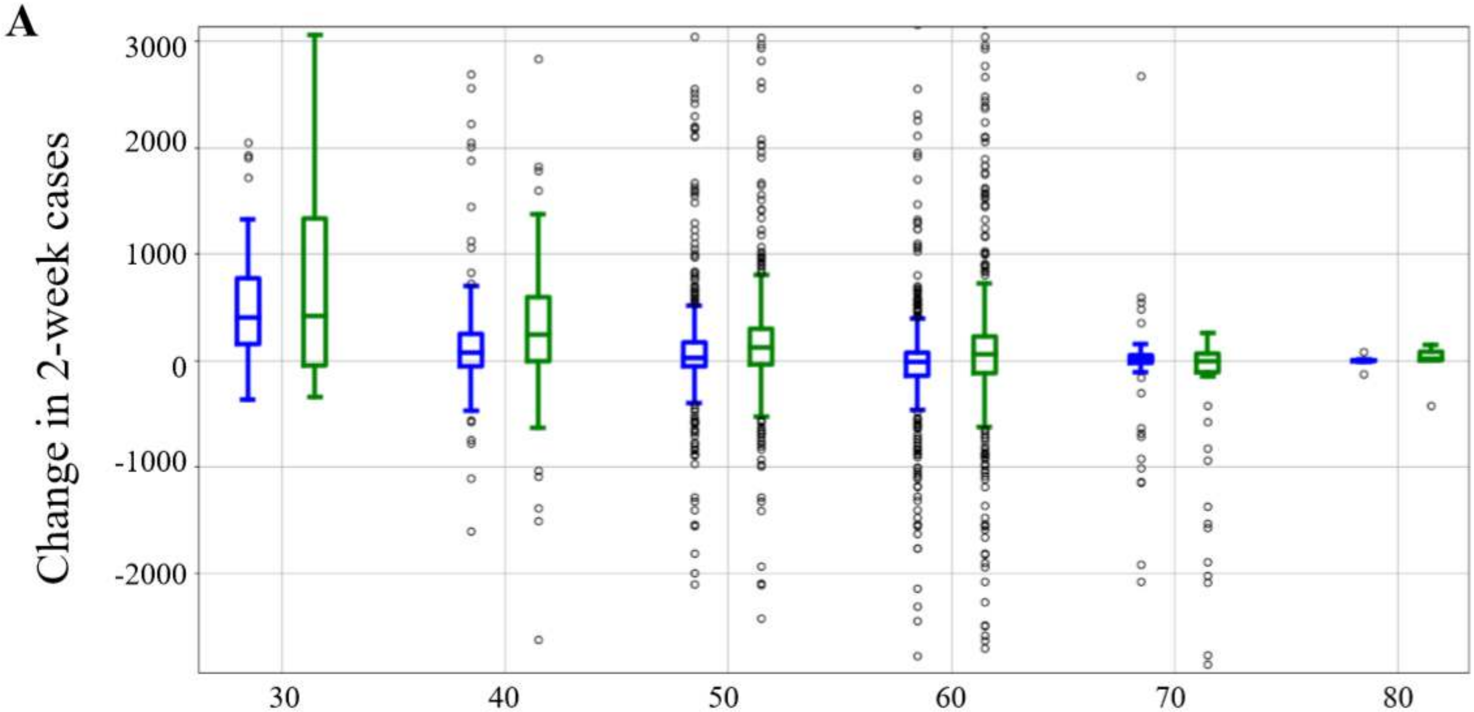

## Figure 3

The correlation between the predicted and actual accumulated COVID-19 cases in the testing data. A. Spatial distribution of the testing correlation per county. B. Histogram of the correlation for the US counties. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.




## Figure 4

The model performance for counties with specific ranges of total cases during the testing interval. The correlation and MAE are shown in part A and B, and the number of counties for each range of total cases is shown in part C. The last bar represents the counties with total cases of more than 40,000.



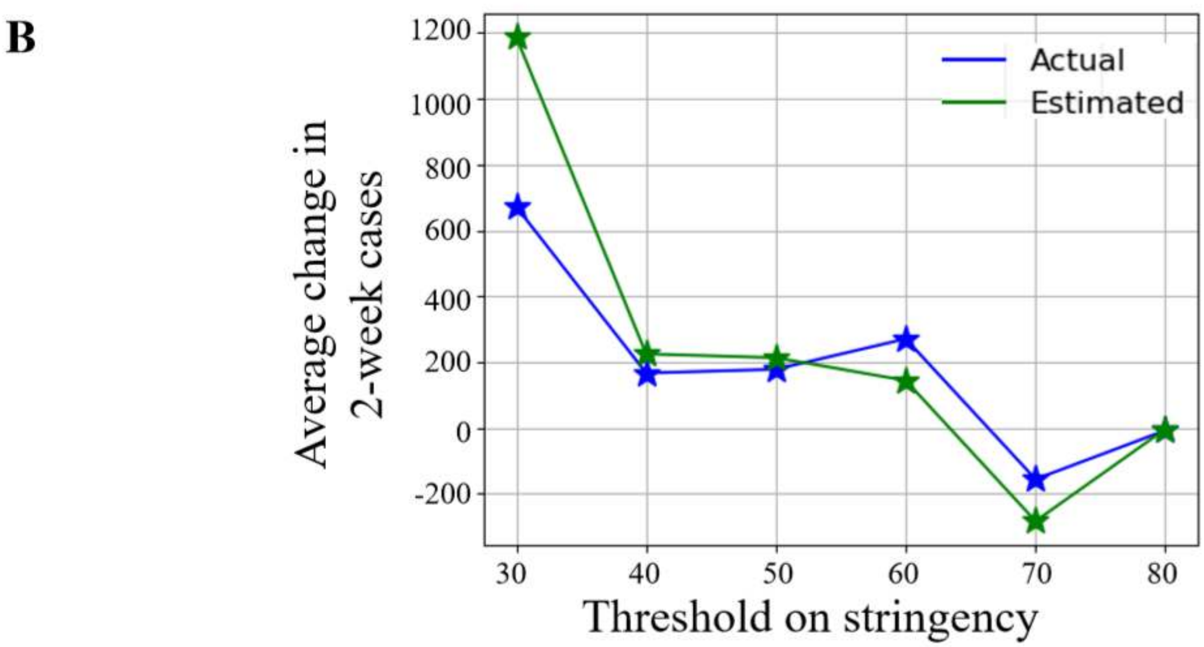
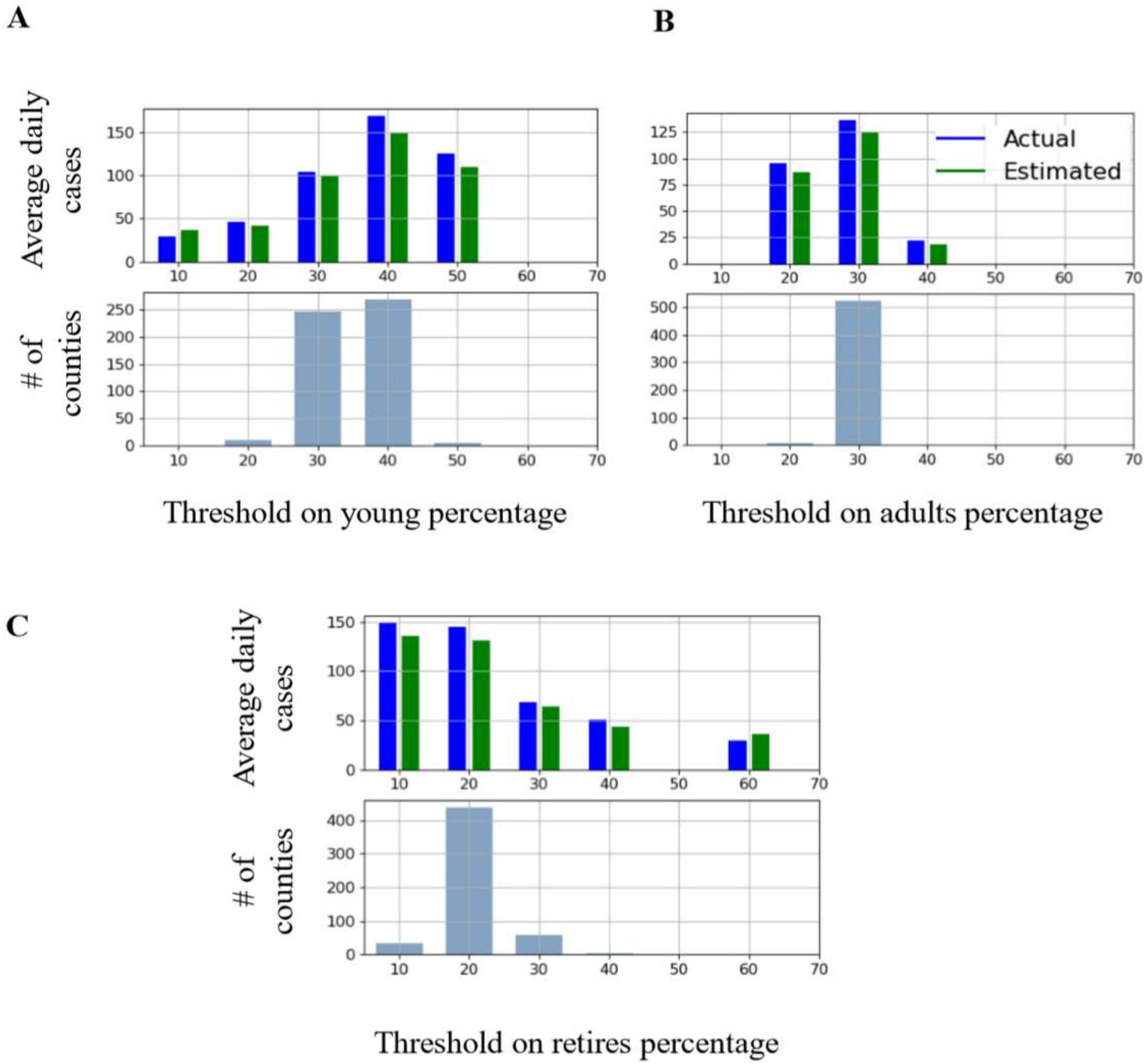




**Figure 5**

The change in accumulated daily cases for two weeks as estimated by the model and the actual cases two weeks after change in the stringency level. Part A shows the change box-plots for all counties during testing interval for each stringency range, and part B shows the averaged changes.




## Figure 6

The actual and the predicated average daily cases during testing interval distributed based on the percentage of young, adults and retires in each county. The trend of the average daily cases based on young, adults, and retires demographics are shown in part A, B, and C, respectively.