Data Studio

1:30–3:00pm, Wednesday, 8 April 2020
Videoconference: https://stanford.zoom.us/j/118966261

Investigator: **Bibek Paudel** Biomedical Data Science

Title: **Spatiotemporal Modeling of Change in Healthcare Utilization due to Exposure to Wildfire Smoke**

Summary.

Previous studies have shown strong relation between exposure to air pollution and several health conditions, including cardiovascular and respiratory ailments. Air pollution due to wildfires have several additional complexities due to its intense and sudden nature. Therefore, as the severity and frequency of wildfires increase, understanding their effect on healthcare utilization is becoming a pressing concern. For public health officials, it is important to know how exposure to wildfire changes healthcare utilization at a population level in order to improve resource planning, preparation and decision making. We are building a predictive model that takes historical healthcare utilization and smoke exposure for individual geographical regions as inputs to predict the short-term change in healthcare utilization.

Objective. To leverage spatial and temporal environmental measurements of smoke and PM2.5 exposure for predicting short-term change in emergent and unplanned inpatient healthcare utilization.

Setting. A large de-identified US commercial healthcare claims database (Optum Clinformatics Datamart), intensity and geographical boundaries of smoke plumes, air pollution measurements (PM2.5) from different ground stations.

Participants. All emergent and unplanned inpatient healthcare claims in California, 2010-2019 (n = 11 million), divided into multiple panels, each covering a two-year period. Condition for inclusion: continuous enrolment for the two-year analysis period, emergency or urgent care claims, including unplanned inpatient stays.

Main Outcome Measures. Absolute and relative change in healthcare utilization for different regions covered in the study, for disease groups in the Clinical Classification Software (CCS) scheme.

Data sources/measurements. Healthcare utilization for each geographical region and health condition, Zipcodes of patient residences, PM2.5 levels from the nearest measurement station, smoke levels from NOAA.

Bias. Seasonal and annual effects, spatial properties.

Statistical methods. Recurrent neural network (nonlinear deep learning) based prediction algorithm, controlling for spatial, and temporal variables.

Questions:

1. For baseline models (e.g., ARIMA), what is the best way to make the time-series non-stationary, and to account for seasonal effects?
2. How to incorporate data on smoke exposure and PM2.5?
3. What are good measures of statistical significance for this setting?
4. For a future study, how to best modify the study to quantify causal effects?

Zoom Meeting Information

```
Topic: Workshop: Data Studio
Time: Apr 8, 2020 13:30 Pacific Time (US and Canada)

Join from PC, Mac, Linux, iOS or Android: https://stanford.zoom.us/j/118966261

Or iPhone one-tap (US Toll):
+18333021536,,118966261#
+16507249799,,118966261#

Or Telephone:
    Dial:
    +1 650 724 9799 (US, Canada, Caribbean Toll)
    +1 833 302 1536 (US, Canada, Caribbean Toll Free)

    Meeting ID: 118 966 261
    International numbers available: https://stanford.zoom.us/u/ab6IwuAEdQ

    Meeting ID: 118 966 261
    SIP: 118966261@zoomcrc.com
```

For more information about Data Studio:

https://med.stanford.edu/dbds/programs/data-studio.html