EPI 945: Practicum & Culminating Experience for the MPH in Epidemiology

Student name and email address:

David Hachuel (dhachuel@hsph.harvard.edu)

Harvard practicum mentor name and email address:

Professor Meredith Rosenthal, Harvard School of Public Health, mrosenth@hsph.harvard.edu

Proposed practicum title:

A Framework for Stress Testing Hospital Bed Utilization

Background and public health relevance

The COVID-19 global pandemic has exposed many weaknesses in health systems when it comes to shortages in medical supplies and other critical hospital resources. To help health administrators avoid the severity of this problem in the future, health systems should adopt stress testing or simulation practices analogous to those the U.S. Federal Reserve or the European Central Bank imposed to large financial institutions after the 2008 economic recession. These exercises could help health administrators identify vulnerabilities and address them early on in supervisory dialogues.

In this study, we decided to focus on hospital bed utilization as we deem it one of the most critical resources. The literature shows that a number of external factors can be used to explain variations in hospital bed utilization. From local demand factors like weather[5] to macroeconomic factors like unemployment[3] have been shown to signal hospital resource utilization.

Specific aims/hypotheses

The overall goal of this work is to demostrate that hospital resource utilization, specifically overnight beds, is reasonably predictable based on macro and local factors. Once a historical model is constructed, we aim to propose a stress testing framework that can be used as a decision-support tool by regulators and health administrators to calibrate between lean and protective resource management in the event of not only severe events such as COVID-19 but also milder events like seasonal weather changes.

Study design and population

It is important to note that given the predictive nature of this study, we are interested in selecting the best possible set of predictors that are able to explain variations in hospital bed utilization using historical data. With that said, the population of our study are England's National Health Service (NHS) health centers or organizations as uniquely identified by their Organization Data Service (ODS) code. The study uses retrospective cohort data which combines NHS overnight bed occupancy data[2] (our primary outcome) with alternative data sources for predictors including the NHS and the Office for National Statistics[6] (ONS).

Primary and other exposure(s): (~1-2 paragraphs)

The goal of this project is to understand from a historical perspective the drivers of overnight hospital bed utilization. We will examine the predictive power of the following indicators:

- Demand factors:

    - Population changes (age, sex and income distribution)
    - Seasonal effects (weather, calendar, natural disasters)
    - Epidemiological changes such as disease prevalence
    - Region characteristics (urban vs. rural)
- Supply factors:

    - Medical and technological advances
    - Hospital efficiency
    - Alternatives to hospital care
- External factors:

    - Political pressures
    - Policy changes
    - Macro-economic context (unemployment rate, gross demestic product)

Primary outcome(s)

The primary outcome is total overnight bed occupancy rate given how critical it has been during the COVID-19 pandemic[1]. We use data from England's NHS[2] collected on a quarterly basis from June 2010 to September 2020. These data is provided at different granularity levels for each NHS organization (total, general acute, learning disabilities, maternity, and mental illness).

Secondary outcome(s)

Overnight bed occupancy data is provided at different granularity levels for each NHS organization (total, general acute, learning disabilities, maternity, and mental illness). As secondary outcomes, we will explore general acute care and mental illness separately in their relationship with our exposure variables.

Other variables: (~1 paragraph)

> Brief description of any other variables, in addition to the exposure(s) and outcome(s), that will be included (how and when they were or will be collected, and how they were or will be defined in your study).

Statistical analysis plan: (~3-4 paragraphs, plus list or shells of possible tables/figures)

> Please be sure to address all four of the points noted below.

- Describe all analyses that you plan to conduct to examine your specific aims(s). These should include descriptive, crude, and adjusted analyses.
- Be sure to identify potential confounders, covariates, effect modifiers, and any sensitivity analyses (if applicable).
- Identify any analytical challenges (technical, structural, etc.)
- Include a list or shells of possible tables and/or figures that will be included.

Data Collection

This works aims to combine several data sources into a master data set. Therefore this process requires careful consideration of data granularity as well as units when merging.

Data Cleaning

We will create all derived feature variables/covariates necessary for the analysis. We will also assess data quality in terms of missing values, outliers and possible human error.

Exploratory Data Analysis

We will then spend a decent amount of time exploring the data to build a deeper understanding of the available variables and their relationships. This exercise will be mainly graphical with perhaps some analytical components. During this process, we will also create a DAG to better visualize the theoretical associations between variables which will help us decide which modeling approaches could be suitable.
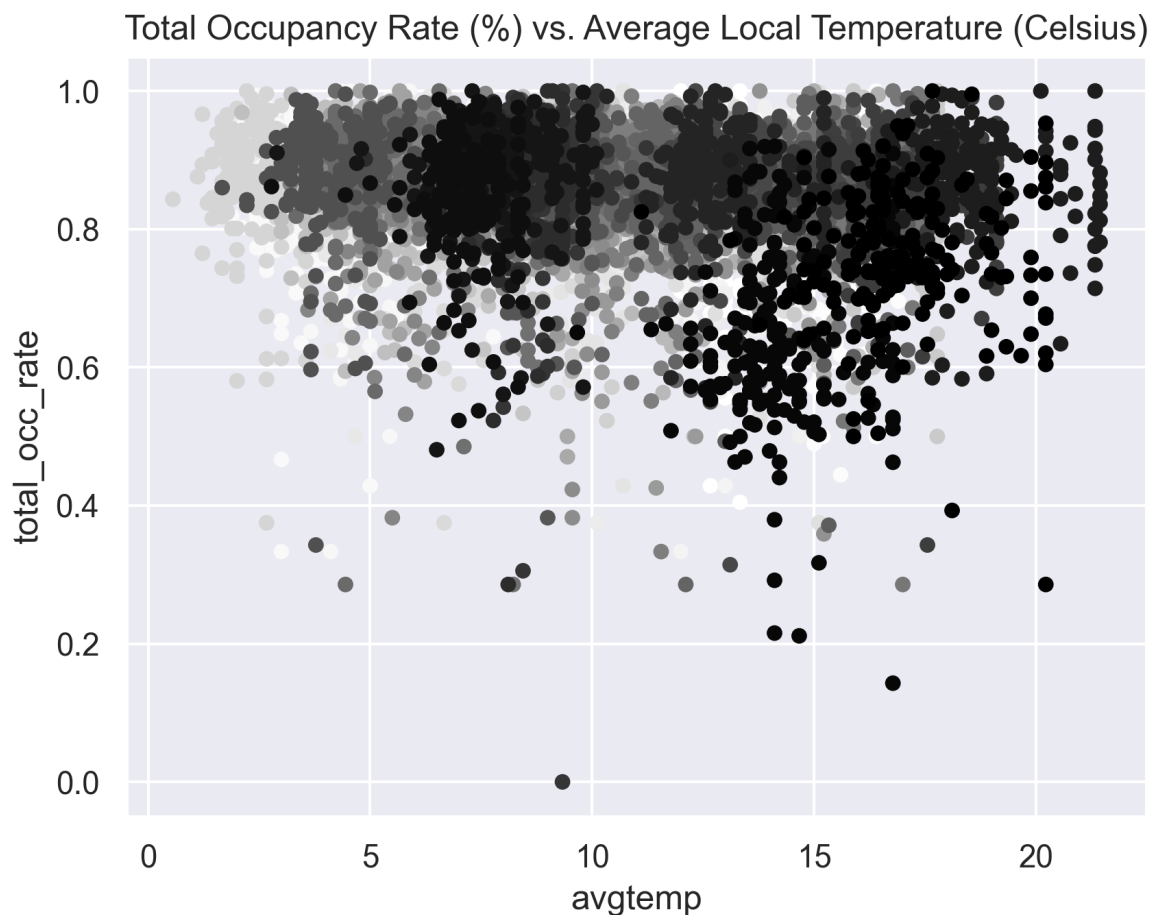
Model Building

The overall nature of the data is time-varying and therefore we will explore a number of multivariate time series modeling approaches applicable for the problem at hand. Namely, we will begin our analysis by implementing a vector autoregression model (VAR) as well as a distributed lag linear model (DLLM). These linear approaches are well suited to model associations between multiple time series. In particular, the DLLM approach is well applicable to describe associations of linear and delayed effects time series data[4]. In order to deal with potential non-linearities, we will implement an extension of the DLLM called distributed lag non-linear models (DLNM).
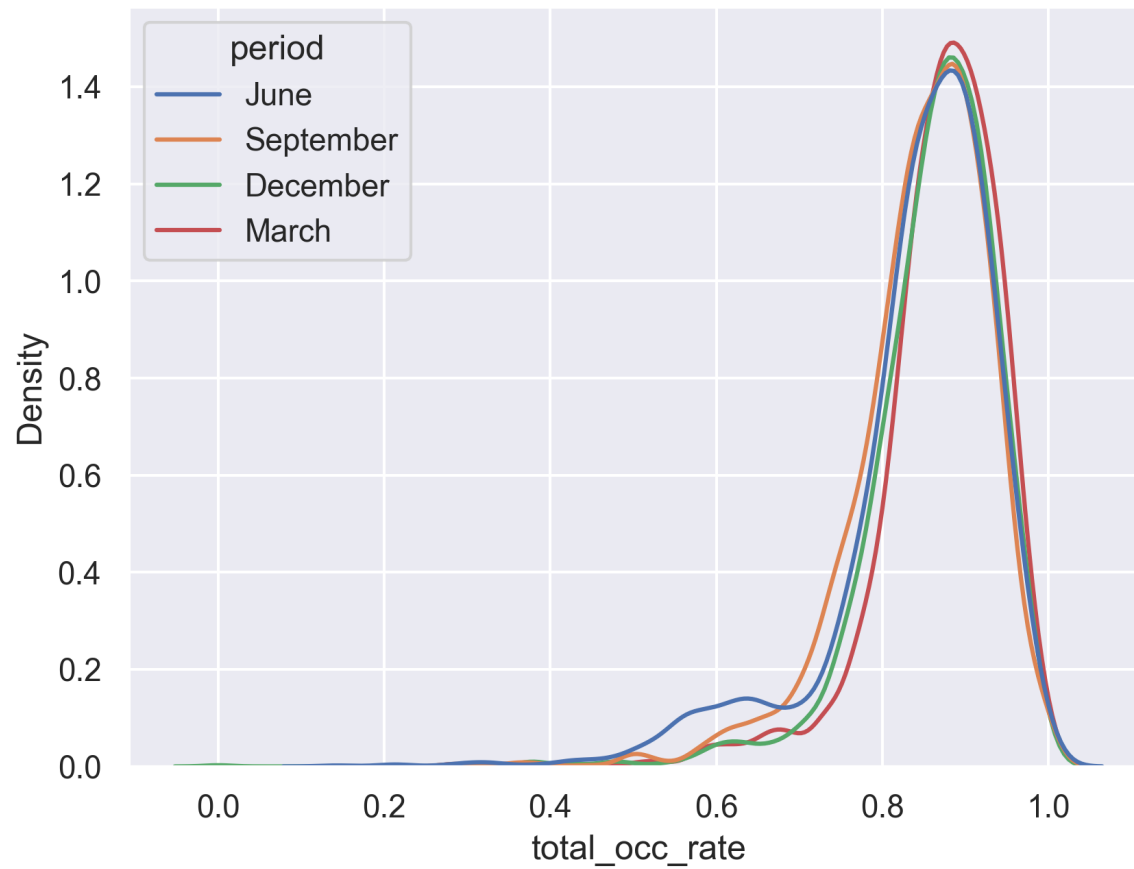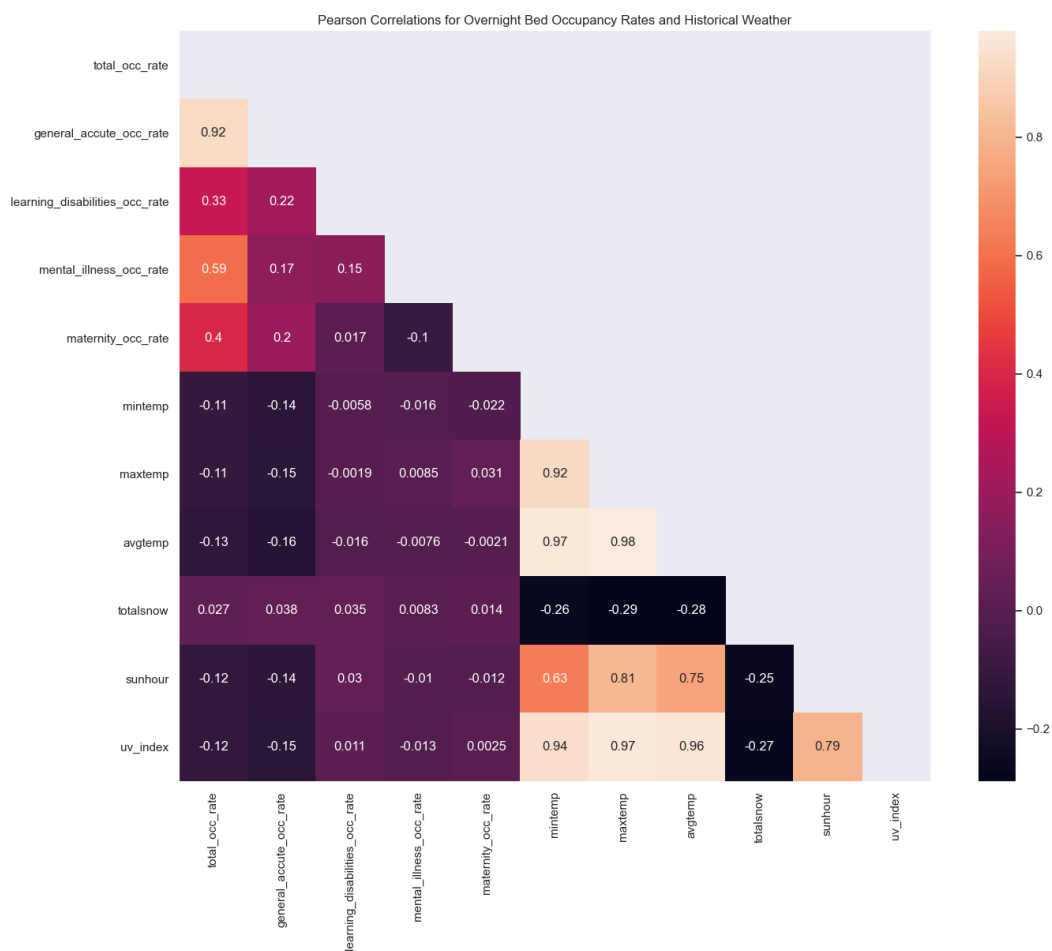
Summary of work completed to date: (~1-2 paragraphs)

All work is publicly available here: https://github.com/dhachuel/a-framework-for-stress-testing-hospital-bed-utilization

- literature research to identify potential predictive factors for our outcome
- collected and combined overnight hospital bed utilization data from the NHS (approx. 10,000 observations, 321 centers across 42 quarters from 2010 to 2020)
- collected and combined historical (10-day intervals) weather data[8] for each center using geographical coordinates (approx. 120,000 observations)
- collected unemployment data from the Office of National Statistics (quarterly resolution from June 2010 to September 2020)
- performed preliminary data exploration

Total Occupancy Rate (%) vs. Average Local Temperature (Celsius)

Total Occupancy Rate (%) Density vs. Time of Year

Pearson Correlations for Overnight Bed Occupancy Rates and Historical Weather

	total_occ_rate	general_accute_occ_rate	learning_disabilities_occ_rate	mental_illness_occ_rate	maternity_occ_rate	mintemp	maxtemp	avgtemp	totalsnow	sunhour	uv_index
total_occ_rate											
general_accute_occ_rate	0.92										
learning_disabilities_occ_rate	0.33	0.22									
mental_illness_occ_rate	0.59	0.17	0.15								
maternity_occ_rate	0.4	0.2	0.017	-0.1							
mintemp	-0.11	-0.14	-0.0058	-0.016	-0.022						
maxtemp	-0.11	-0.15	-0.0019	0.0085	0.031	0.92					
avgtemp	-0.13	-0.16	-0.016	-0.0076	-0.0021	0.97	0.98				
totalsnow	0.027	0.038	0.035	0.0083	0.014	-0.26	-0.29	-0.28			
sunhour	-0.12	-0.14	0.03	-0.01	-0.012	0.63	0.81	0.75	-0.25		
uv_index	-0.12	-0.15	0.011	-0.013	0.0025	0.94	0.97	0.96	-0.27	0.79	

Summary of results to date: (~1-2 paragraphs)

> If you do not have any results yet, you can leave this section blank. If you have results, describe them briefly here. You may include tables/figures as appropriate, but you should also include text that summarizes your results. Please do not include Stata output.

Next steps: (~1-2 paragraphs)

- finish collecting all exposure variables
- research proposed modeling approaches to understand their correct applicability
- meet with advisor to review progress to date
- continue exploratory data analysis
- implement modeling approaches and proceed to result interpretation
- meet with advisor to review results to date

Any challenges/problems encountered so far: (~1 paragraph)

References

1. Rees, E.M., Nightingale, E.S., Jafari, Y. et al. COVID-19 length of hospital stay: a systematic review and data synthesis. BMC Med 18, 270 (2020). https://doi.org/10.1186/s12916-020-01726-3
2. England, NHS. "Bed Availability and Occupancy Data – Overnight." *NHS Statistics*, NHS, www.england.nhs.uk/statistics/statistical-work-areas/bed-availability-and-occupancy/bed-data-overnight/.
3. Bidargaddi, Niranjan, et al. "Changes in Monthly Unemployment Rates May Predict Changes in the Number of Psychiatric Presentations to Emergency Services in South Australia." *BMC Emergency Medicine*, vol. 15, no. 1, 2015, doi:10.1186/s12873-015-0042-5.
4. Gasparrini, Antonio. "Distributed Lag Linear and Non-Linear Models In R: The Package dlnm." *Journal of Statistical Software*, vol. 43, no. 8, 2011, doi:10.18637/jss.v043.i08.
5. Giang, Pham Ngan, et al. "The Effect of Temperature on Cardiovascular Disease Hospital Admissions among Elderly People in Thai Nguyen Province, Vietnam." *Global Health Action*, vol. 7, no. 1, 2014, p. 23649., doi:10.3402/gha.v7.23649.
6. Leaker, Debra. "Unemployment Rate (Aged 16 and over, Seasonally Adjusted)." *Unemployment Rate (Aged 16 and over, Seasonally Adjusted) - Office for National Statistics*, Office for National Statistics, 10 Nov. 2020, www.ons.gov.uk/employmentandlabourmarket/peoplenotinwork/unemployment/timeseries/mgsx/lms.
7. "Open Postcode Geo." *Open Postcode Geo*, Calderdale Metropolitan Borough Council, 17 Oct. 2016, data.gov.uk/dataset/091feb1c-aea6-45c9-82bf-768a15c65307/open-postcode-geo.
8. "Real-Time World Weather REST API." *Weather Stack*, weatherstack.com/.