Introduction

Background
Our work is motivated by the need from hospital leaders to have timely and accurate forecasts to guide planning for surges in hospital census, i.e., bed capacity, due to the COVID-19 pandemic. Adequate preparation can help prevent or mitigate strains on hospital resources COVID-19 that result when hospitals exceed their historical capacity.

Objective
We want to explore whether the local COVID-19 infection incidence and the COVID-19 hospital census can be successfully incorporated within a multivariate time-series model to deliver satisfactory 7-day-ahead forecast performance and examine the application of this model to scenario-based long-term forecasting.

Study data
The study data are aggregated daily COVID-19 hospital census across 11 Atrium Health hospitals plus a virtual hospital in the greater Charlotte metropolitan area of North Carolina, as well as the total daily infection incidence across the same region during the May 15, 2020 - December 5, 2020 period (Figure 1). The data was applied to appropriate transformations to linearize their relationship.

Methods

Model
A Vector Error Correction model (VECM) is a vector autoregressive (VAR) model used for nonstationary multivariate time-series and accounts for stable long-run relationships, i.e., cointegration, between the time series. A time-series vector is said to be cointegrated if there is at least a linear combination of the vector that is trend-stationary.

Following [1], we first describe the VAR representation of the model, i.e., the level equation:

\[ y_t = \Pi_1 y_{t-1} + \Pi_2 y_{t-2} + \ldots + \Pi_p y_{t-p} + \mu + \Phi D_t + \epsilon_t \]

for time \( t = 1, \ldots, T \), where \( \Pi_i \) (for \( i = 1, \ldots, p \)) are \( k \times k \) coefficient matrices of the lagged series at lag \( i \), \( \Pi_0 \) is a \( k \times 1 \) vector of constants, \( D_t \) is a \( 6 \times 1 \) vector of weekly seasonal indicators, \( \Phi \) is a \( k \times k \) coefficient matrix for seasonal indicators, and \( \epsilon_t \) is a \( k \times 1 \) vector of random errors.

The VECM representation, i.e., difference equation, can be derived from above:

\[ \Delta y_t = \Delta y_{t-1} + \Gamma_1 \Delta y_{t-2} + \ldots + \Gamma_p \Delta y_{t-p} + \mu + \Phi D_t + \epsilon_t \]

where \( \Delta y_t \) is a \( k \times 1 \) vector of the differenced series \( y_t, y_{t-1}, \ldots, y_{t-p} \), \( \Pi = -(\Pi_1 + \ldots + \Pi_p) \), and \( \Gamma_i = -(\Pi_{i+1} + \ldots + \Pi_p) \) (for \( i = 1, \ldots, p - 1 \)).

The model has the following assumptions:

- Assumption 1: The components of \( y_t \) are at most I(1), i.e., integrated of order 1.
- Assumption 2: \( 0 \leq r = \text{rank}(\Pi) \leq k \)
- Assumption 3: \( \Sigma \), which is a vector of random error, is a stationary, depending on the scenario. In the worst-case scenario, alleviated resource demand against our resource capacity, we assessed the long-range scenario-based forecasting model, used for COVID-19 hospital census, and the worst-case scenario, the hospital census was projected to peak on February 16, 2021 (11 days later than incidence) with approximately 850 patients at the 80% forecast value upper bound (Figure 4).

Results
Model estimation
The level equation requires 7 lags (p = 7) to capture all temporal dependencies. Strong evidence for a cointegration relationship (P<0.01):

\[ \Delta y_{t-1} = \text{Census}_{t-1} - 0.8013 \text{Incidence}_{t-1} + 7.8266 \]

where \( e_{t-1} \) was the (lagged) error correction term.

- Long-run effect: the error correction term had a negative and a statistically significant effect on Census change (P<0.01).
- Short-run effects: past incidence changes at lags 1, 2, 4, 5, and 6, as well as past Census change at lag 2, had significant effects on Census change.

Forecast performance
The total value (median) of MAPE was 5.9% and the 95th percentile of MAPE was 13.4% (Figure 3). For the sake of comparison, the corresponding values from an Autoregressive Integrated Moving Average (ARIMA) model using the COVID-19 hospital census only were 6.6% and 14.3%.

Approach
We used Mean Absolute Percentage Error (MAPE) to evaluate the 7-day-ahead forecasts of Census:

\[ \text{MAPE} = \frac{100}{T} \sum_{t=1}^{T} \left| \frac{\text{Forecast}_t - \text{Actual}_t}{\text{Actual}_t} \right| \]

where \( \text{Forecast}_t \) is the forecast value and \( \text{Actual}_t \) is the actual value.

The sampling distribution of out-of-sample MAPE is obtained by 1000 bootstrapping cross-validation sets.

Long-range scenario-based forecasting
In all scenarios, due to cointegration, the hospital census followed corresponding converge trajectories with peaks occurring at approximately the same time as the incidence, depending on the scenario. In the worst-case scenario, the hospital census was projected to peak on February 16, 2021 (11 days later than incidence) with approximately 850 patients at the 80% forecast value upper bound (Figure 4).

Discussion
We have ascertained the long-term stable relationship between local infection incidence and COVID-19 hospital census. Whereas, current models, e.g., the COVID-19 Hospital Impact Model for Epidemics (CHIME) [4], rely on simplified assumptions about the relationships.

Local infection shows to be an effective leading indicator for COVID-19 hospital census, through both short-run and long-run effects, and as demonstrated by very good forecast performance against the traditional ARIMA model.

- In hindsight, by evaluating different scenarios of peak resource demand against our resource capacity, we have correctly assisted our leaders of our capability to handle even the worst-case scenario, alleviated uncertainty, and effectively guided long-term planning of adequate staffing, bed capacity, and equipment supplies through the pandemic.

Resources
1. Pfaff B. Analysis of integrated and cointegrated time series with R. 2nd ed. New York: Springer; 2008.
2. An epidemiological forecast model and software assessing interventions on COVID-19 epidemic in China. Journal of Data Science. 2020;18(3): 409-432. [doi: 10.6339/JDS.20200718C(00003)
3. Gardner ES, McKenzie ED. Forecasting Trends in Time Series. Management Science. 1985;31: 1237-1246. [doi: 10.1287/mnsc.31.10.1237]
4. Weissman G, Crane-Boesche A, Chivers C, Luong T, Hanish A, Levy MZ, et al. Locally Informed Simulation to Predict Hospital Capacity Needs During the COVID-19 Pandemic. Annals of Internal Medicine. 2020;173: 21-28. [doi: 10.7326/M20-1260] [Medline: 32529197]

Acknowledgements
This research protocol was submitted to the Atrium Health Institutional Review Board (IRB) prior to execution and the study was deemed exempt from IRB oversight. In compliance with HIPAA regulations, individual patient information is not disclosed, all data have been deidentified and reported as aggregates.

Contact Info
Hieu.Nguyen@atriumhealth.org