Forecasting COVID-19 hospital census: a multivariate time-series model based on local infection incidence

Hieu Nguyen, Philip Turk, Andrew McWilliams
Atrium Health

May 2021

COVID-19 pandemic

- A novel coronavirus causing infectious respiratory disease.
- Basic reproduction number (R_0): low-mid 2s, (SARS: 2; H1N1: 1.3 [1]).
- About 16% of COVID-19 positive hospital admissions require ICU (Feb 2021, Atrium Health).

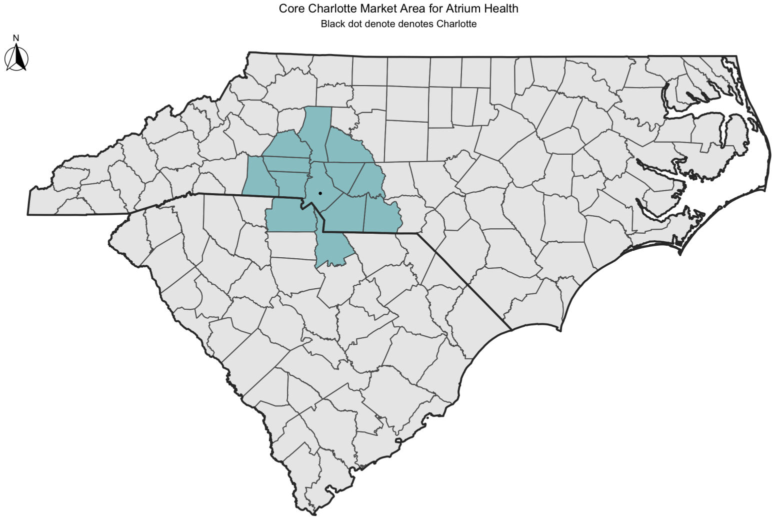
Forecasting COVID-19 hospital census

- Healthcare systems need to prepare for surges of hospital demands.
- COVID-19 hospital census (Census), the number of beds occupied by COVID-19 positive patients, is a central resource indicator in planning decisions.

Prior forecasting models

- Univariate time-series models: ARIMA, SARIMA, ETS [2-4].
- Multivariate time-series models, using leading indicators:
  - Hospital admissions [5]
  - Google-search trends of COVID-19 terms related to "testing" [6]
  - Number of people flagged by an Internet-based virtual health screen bot [6]

Local infection incidence as a leading indicator



Local infection incidence as a leading indicator (cont.)

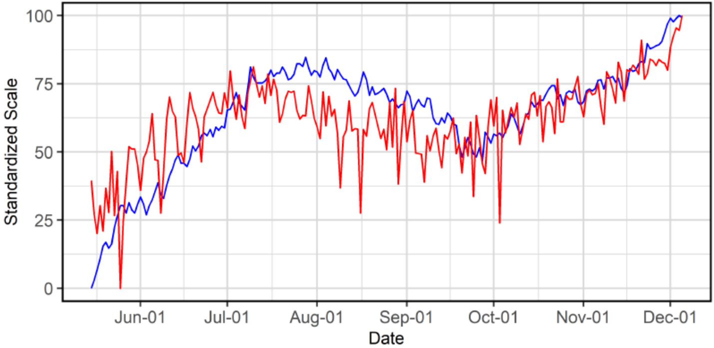


Fig 1. Scaled time-series for Census and Incidence for the period May 15, 2020 - December 5, 2020. Transformed Census (blue) and Incidence (red) are linearly standardized to the 0-100 scale.

Vector Error Correction model (VECM)

VECM is a vector autoregressive (VAR) model which accounts for cointegration, i.e., stable linear relationships. In its VAR representation, we have:

$$\boldsymbol{y}_t = \boldsymbol{\Pi}_1 \boldsymbol{y}_{t-1} + ... + \boldsymbol{\Pi}_p \boldsymbol{y}_{t-p} + \boldsymbol{\mu} + \boldsymbol{\Phi} \boldsymbol{D_t} + \boldsymbol{\epsilon}_t$$

for time $t = 1, ..., T$, where $\boldsymbol{\Pi}_i$ (for $i = 1, ..., p$) are k x k coefficient matrices of the lagged series at lag i, $\boldsymbol{\mu}$ is a k x 1 vector of constants, $\boldsymbol{D}_t$ is a 6 x 1 vector of day-of-the-week seasonal indicators, $\boldsymbol{\Phi}$ is a k x 6 coefficient matrix for seasonal indicators, and $\boldsymbol{\epsilon}_t$ is a k x 1 vector of random errors.

Vector Error Correction model (VECM) (cont.)

VECM representation as a transformation of the VAR model:

$$\begin{aligned}\boldsymbol{\Delta y}_t &= \boldsymbol{\mu} + \boldsymbol{\Pi y}_{t-1} + \boldsymbol{\Gamma}_1 \boldsymbol{\Delta y}_{t-1} + \ldots + \boldsymbol{\Gamma}_{p-1} \boldsymbol{\Delta y}_{t-p+1} + \boldsymbol{\Phi D}_t + \boldsymbol{\epsilon}_t \\ &= \boldsymbol{\mu} + \alpha\beta^{\boldsymbol{T}} \boldsymbol{y}_{t-1} + \boldsymbol{\Gamma}_1 \boldsymbol{\Delta y}_{t-1} + \ldots + \boldsymbol{\Gamma}_{p-1} \boldsymbol{\Delta y}_{t-p+1} + \boldsymbol{\Phi D}_t + \boldsymbol{\epsilon}_t\end{aligned}$$

where $\boldsymbol{\Delta y}_t$ is a k x 1 vector of the differenced series $\boldsymbol{y}_t - \boldsymbol{y}_{t-1}$,
$\boldsymbol{\Pi}_i = -(\boldsymbol{\Pi}_{i+1} + \ldots + \boldsymbol{\Pi}_p)$ (for $i = 1, \ldots, p-1$),
$\boldsymbol{\Pi} = -(\boldsymbol{I} - \boldsymbol{\Pi}_1 - \ldots - \boldsymbol{\Pi}_p)$. α and β are k x r matrices, where $r = rank(\boldsymbol{\Pi})$.

$\beta^{\boldsymbol{T}} \boldsymbol{y}_{t-1}$ is a trend-stationary term showing the long-run cointegration relationship. Thus, α represents the long-run effects.
$\boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_{p-1}$ represent the short-run effects.

Long-range scenario-based forecasting

- High uncertainty near the peak of infection prevalance.
- Univariate time-series models may fail near the peak.
- With cointegration we can leverage subtle, but critical, changes in Incidence (e.g., concavity) suggesting the forecasting of Census under different pandemic scenarios.

Long-range scenario-based forecasting (cont.)

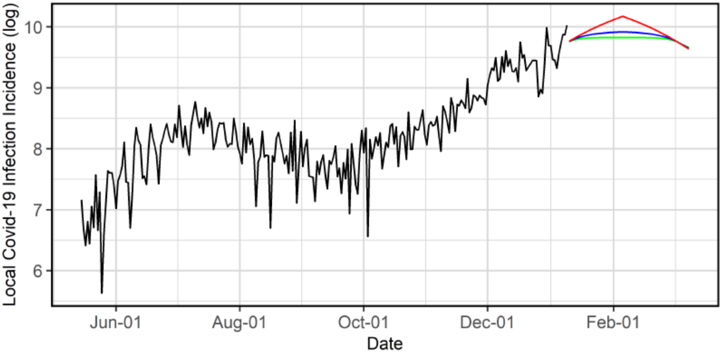


Fig 2. 60-day projected local Covid-19 infection incidence on the log scale, as of January 9, 2021. Past values (black), worst-case scenario (red), base-case scenario (blue), best-case scenario (green).

Main results

- The model was specified as a VECM with 7 lags in its VAR representation ($p = 7$), with 1 cointegration relationship (p-value < 0.01).

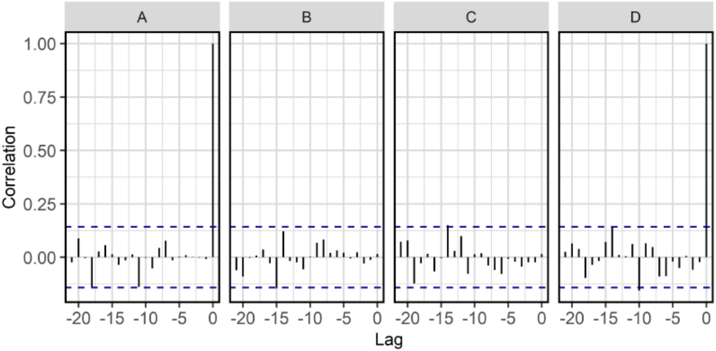


Fig 3. Autocorrelation functions and Cross-correlation functions of the residuals. (A) Census residuals, (B) Lagged Census residuals and Incidence residuals, (C) Census residuals and lagged Incidence residuals (D) Incidence residuals.

Main results (cont.)

- Cointegration relationship:

$$\boldsymbol{\beta}^{\boldsymbol{T}} \boldsymbol{y}_{t-1} = \text{Census}_{t-1} - 0.8013\,\text{Incidence}_{t-1}$$

- Long-run effects: $\boldsymbol{\beta}^{\boldsymbol{T}} \boldsymbol{y}_{t-1}$ had a negative and significant effect on Census change.
- Short-run effects on Census change from:
  - Past Census changes (lag 2).
  - Past Incidence changes (lag 1, 2, 4, 5 and 6).
  - Day-of-the-week: Census change was significantly higher on Monday compared to Thursday.

Goodness-of-fit

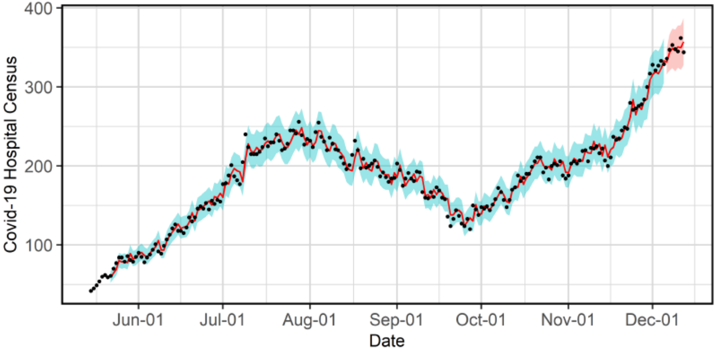


Fig 6. One-step-ahead in-sample and 7-day-ahead out-of-sample predictions. True values (black), in-sample and out-of-sample predictions (red line), 95% confidence intervals (blue band), 80% forecast intervals (red band). The model is fitted on data from May 15, 2020 to December 5, 2020.

7-day-ahead forecast performance

- 7-day-ahead Mean Absolute Percentage Error (MAPE) via time-series cross-validation has a median of 5.9% and a 95th percentile of 13.4%.

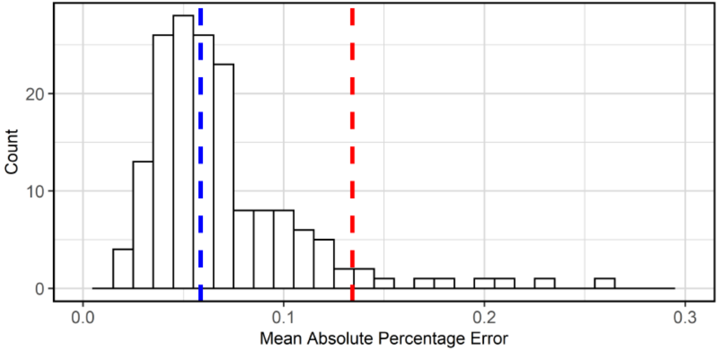


Fig 5. Distribution of the 7-day-ahead Mean Absolute Percentage Error from time-series cross-validation for the period June 16, 2020 - November 28, 2020. Median (blue), 95$^{\text{th}}$ percentile (red).

Worst-case scenario 60-day-ahead forecast

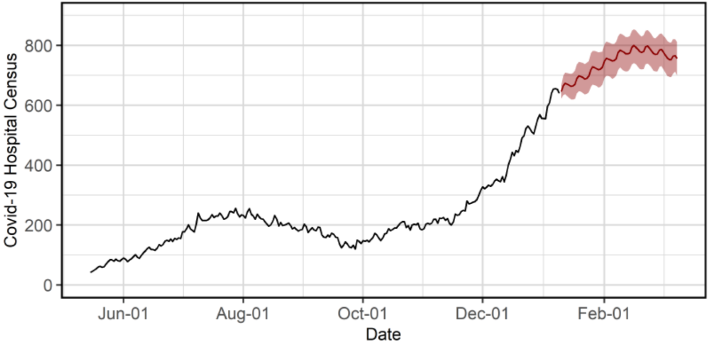


Fig 7. 60-day Census forecasts in the worst-case scenario, as of January 9, 2021. Past values (black), forecasts (red line), 80% forecast intervals (red band).

Conclusions

- Local infection incidence is cointegrated with Census and has important short-run and long-run effects on Census.
- The model offers competitive 7-day-ahead forecast performance.
- Long-range scenario-based forecasting has been successfully applied to determine the potential for resource capacity to be exceeded at Atrium Health.

References

1. Singhal T. A Review of Coronavirus Disease-2019 (COVID-19). Indian J Pediatr. 2020; 6. Available: https://doi.org/10.1007/s12098-020-03263-6
2. Earnest A, Chen MI, Ng D, Sin LY. Using autoregressive integrated moving average (ARIMA) models to predict and monitor the number of beds occupied during a SARS outbreak in a tertiary hospital in Singapore. BMC Health Services Research. 2005;5: 36. Available: https://doi.org/10.1186/1472-6963-5-36
3. Jones SS, Thomas A, Evans RS, Welch SJ, Haug PJ, Snow GL. Forecasting Daily Patient Volumes in the Emergency Department. Academic Emergency Medicine. 2008;15: 159–170. Available: https://doi.org/10.1111/j.1553-2712.2007.00032.x

4. Capan M, Hoover S, Jackson EV, Paul D, Locke R. Time Series Analysis for Forecasting Hospital Census: Application to the Neonatal Intensive Care Unit. Applied Clinical Informatics. 2016;7: 275–289. Available: https://doi.org/10.4338/ACI-2015-09-RA-0127
5. Berta P, Paruolo P, Verzillo S, Lovaglio PG. A bivariate prediction approach for adapting the health care system response to the spread of COVID-19. PLOS ONE. 2020;15: e0240150. Available: https://doi.org/10.1371/journal.pone.0240150
6. Turk P, Tran T, Rose G, McWilliams A. A Predictive Internet-Based Model for COVID-19 Hospitalization Census. medRxiv. 2020 [cited 5 Feb 2021]. Available: https://doi.org/10.1101/2020.11.15.20231845