Tracking COVID Locally and Adaptively

Researcher: Jessica Lim

University of Wisconsin – Eau Claire, Mathematics Department with Faculty advisor: Dr. Jessica Kraker

Abstract

In summer 2020, the project started with the faculty mentor created a dashboard to visualize and summarize information about local COVID data. Skills developed include learning a new programing package dplyr in R and hosting code on GitHub, which were then applied in preparatory work such as building new data frames and calculations. Thus, we will discuss a predictive time series model with lagged counts for future outcomes (such as hospitalizations), built on age-grouped case-counts to account for the disparities in outcomes observed for different ages in the COVID pandemic.

Principles and Process

This outline to the right summarizes the research, motivation, data management, and programming used to gather, clean, compute, organize, and evaluate the Wisconsin COVID data used in the dashboard (produced using RStudio and maintained on GitHub) and model-fitting. Communication was handled via numeric summaries and visualization and description of trends.

Who: – all people in Eau Claire county (summary visualizations, left columns) – all people in state of Wisconisin (modeling, right columns)

What: "Daily" and "Active" new-positive case summaries; Daily hospitalizations; Daily deaths. Generally all are reported as 7-day rolling averages. All data is publicly available, accessed from daily-updated Wisconsin Department of Health Services data page https://www.dhs.wisconsin.gov/covid-19/data.htm (last accessed on April 7, 2021), with multiple data-visualizations and provided data definitions.

When: County-level data is reported through April 7 (the cut-off point of the analytics), lag estimates are based on data through from August 2020-January 2021, and modeling uses data from August 2020 – March 2021.

How: Data is downloaded on an (at-least) weekly basis; data-analytics and visuals are programmed in (freeware) *R*; Dashboard is hosted on free server GitHub, with data and coding publicly available for transparency.

Why: The COVID-19 health information of the local community allows us to make better-informed decisions about how to safely interact. Data summaries were developed to address questions related to: local trends within geographic locations; differing patterns among age groups (informing guidelines about interactions with those age groups); predict trends in hospitalizations, informing health-care and community plan and act to slow the spread.

Data Principles and Communication

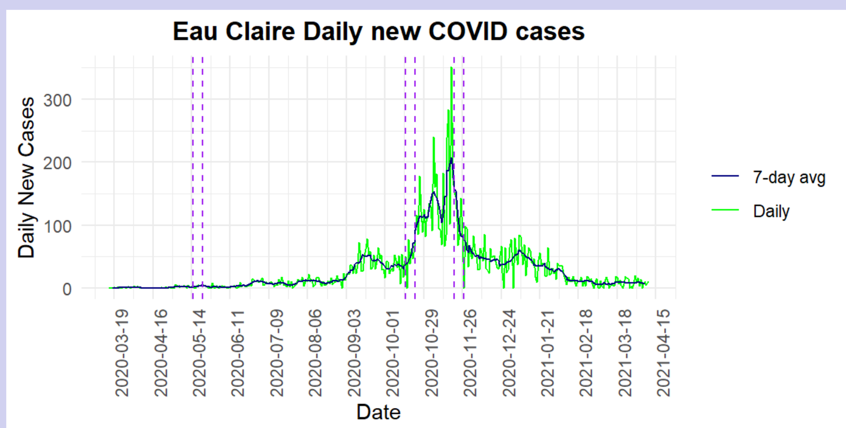
Notably, health and wellness data is carefully protected by HIPAA laws. The Wisconsin Department of Health Services provides COVID-related data publicly, stating "Our goal is to provide the highest quality COVID-19 data with full transparency." Data is used by numerous counties and individuals.
- No information about individuals are provided; counts under 5 are not included, so as to prevent individual identification.
- Both researchers completed full IRB training.
- Every summary and inference is assessed under the questions: Does this add something new? Is the benefit to the community clear? Is the information necessary to a decision-making process?

Lagged Effects

Noted in the initial data gathered in the COVID-19 pandemic (particularly those reported from Italy), hospitalization and death counts lagged by days to weeks behind the daily counts. This is reasonable, as many patients might be admitted to hospital only after getting sufficiently ill (at a time-lag from when they were initially diagnosed). That is, the *trends* we see in behaviors of the hospitalization and deaths daily counts also tend to exhibit a time-lag. The trend is not always precisely mirrored, since age-composition of new cases results in differing lagged counts of hospitalizations and deaths (noting the now well-known observation that COVID-19 affects older age groups much more harshly).

Daily Cases and 7-day rolling average – by County

We using centered 7- day moving average, which the average value is placed at middle of the range, for this 7-day average we placed the average value at period 4 and calculate moving average three days before and three days after then averaging the.



Why 7-day moving average?

-Make the result more stable. There are recording delay on somedays, especially on weekend. From the graph, we can see that there is fewer cases(almost 0 case) on weekend.

-Provide a smoother graph which help to reveals the underlying trends in our data.

-Increase the accuracy on predicting the future hospitalization cases. 7-day moving average may help to best time series forecasting model.

Rate of change

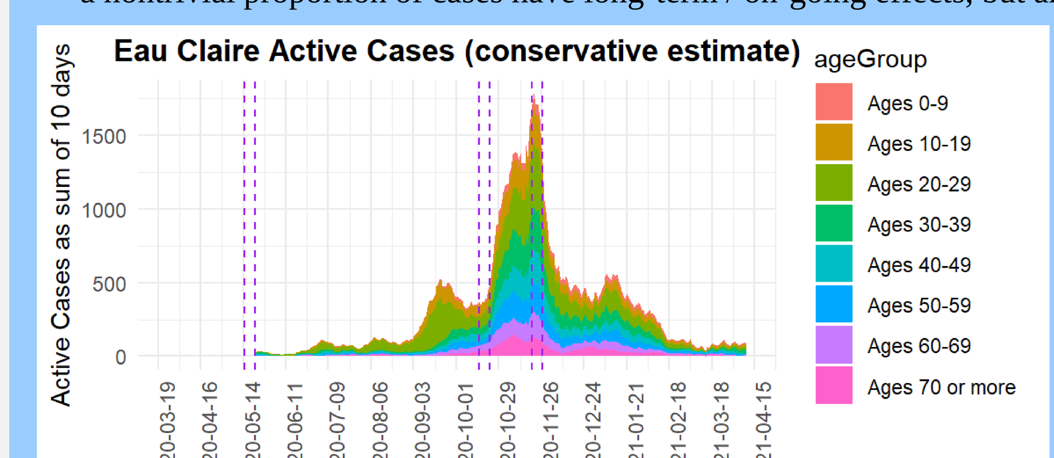
We are used to thinking about rate of change as a slope. We can effectively still think about this as "rise over run", where we are looking at the "run" across time. To explain this, let's take a look at "Friday to Friday" rates of change in Eau Claire county:

Friday dates	Change in Daily 7day Average	Interpret
"May 15 2020" to "May 22 2020"	1.86 to 4	115% increase, but only daily average of 4 case (not very many in a county of 104,000 people)
"October 16 2020" to "October 23 2020"	39.14 to 92.29	136%increase, this is a dramatic changed in cases. It is concerning and we need more precaution.
"November 20 2020" to "November 27 2020"	157.29 to 77.14	51% decrease, but there are 80 cases decreased (more cases than the example above)

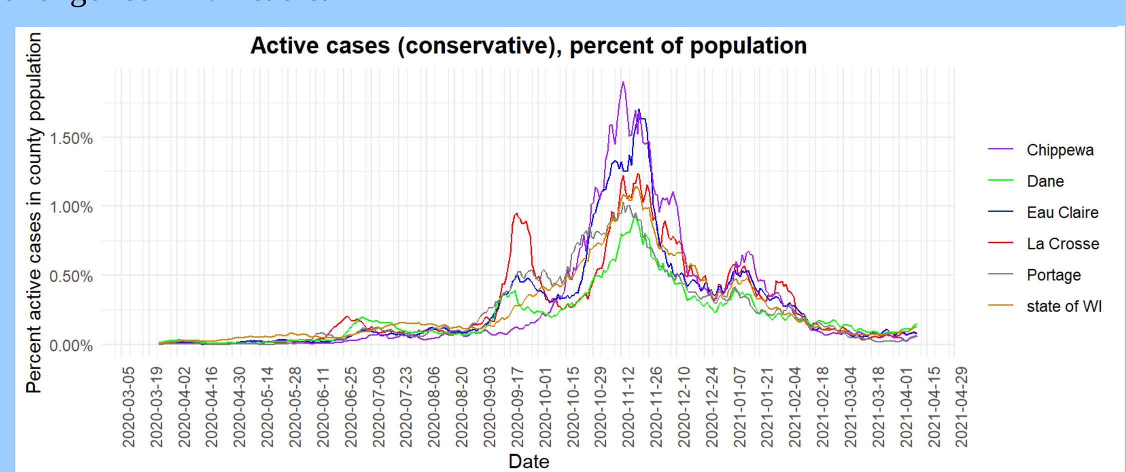
Summarize: Even though rate of change (%) that showed the changes between every 7day average is important to analysis the data, but we need to also assessed on daily 7day average cases.

"Active" (Rolling-Total) Cases – by County

Why and What is "Active"? "Active" refers to those cases that currently may be able to be spread to other individuals.
- Estimated recovery time from health experts varied originally between 10-20 days, with progressively updated CDC recommendations allowing a clearance 10 days post-diagnosis.
- a nontrivial proportion of cases have long-term / on-going effects, but are no longer communicable.



Plot 1: 10-day running total ("active") cases, for Eau Claire county. Subdivided by age (notable both for recognizing sub-populations with higher active cases, and concerns for more severe effects as age increases).



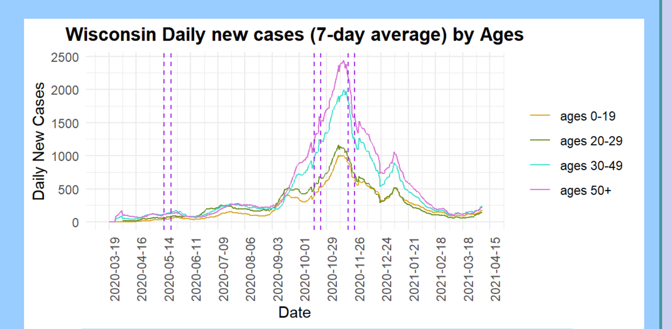
Plot 2: While simple counts of active cases show trends, including proportional composition of the population can help identify where active cases are denser within the population. Plot 2 cross-compares active cases across several Wisconsin counties as well as statewide.

Case Summarization by Age group and Lag Time - Statewide

Young	<30, 30s, 40s
Middle	50s, 60s
Elder	70s, 80+

This presumes that we are actually identifying active cases. However, this was not true throughout the course of the pandemic:
- day1 of recorded data = March 15, 2020 for WI data;
- cases were initially far underreported, due to lack of testing availability
- when should we actually start? day 140 = August 1
- for practical (testing ramped up) and empirical reasons (data patterns)
- data provided originally grouped by age-decade

Data through January: includes mostly prior to vaccination effects.				
New-hospitalization case 7day averages at the following lags				
Young		3	7	14
Middle		3	7	14
Elder	1	3		



Plot shows Wisconsin cases split into 4 age groups, with rolling average

How We Calculate The Lag Time

- We used sqrtMSE to find the best Lag Time for each group hospitalization and death. We focused on hospitalization in this project. The table showed the few lags that fit the model with lowest sqrtMSE.
- sqrtMSE is to measure the difference between predicted value in the model and the actual value we observed.
- The lower the sqrtMSE ,the closer predicted value and actual value.

Data Prep

- Analyze data into different age group which is < 20, 20s, 30s, 40s, 50s, 60s, 70s and 80+. Calculate 7-day average case for age group.
- The table shows that age group 20s has the highest case positive case counts and 60s has the lowest.
- Due to each age group has different lag time, we divide the age group into three groups with similar lag and has lowest sqrtMSE, which is Young, Middle and Old.

Age Group	Summaries: Positive 7day case through January			
	Min	Median	Mean	Max
<20	0	141.00	262.38	1005.86
20s	15.00	238.50	336.77	1163.86
30s	19.86	133.57	277.36	1033.29
40s	19.29	119.07	252.50	963.14
50s	27.14	121.50	270.73	1042.14
60s	15.71	80.79	193.57	744.57
70s	7.286	41.286	101.392	386.429
80+	0	22.29	66.46	269.43

Modeling Results – Planning for Space: Predicting Hospitalizations in the state of WI

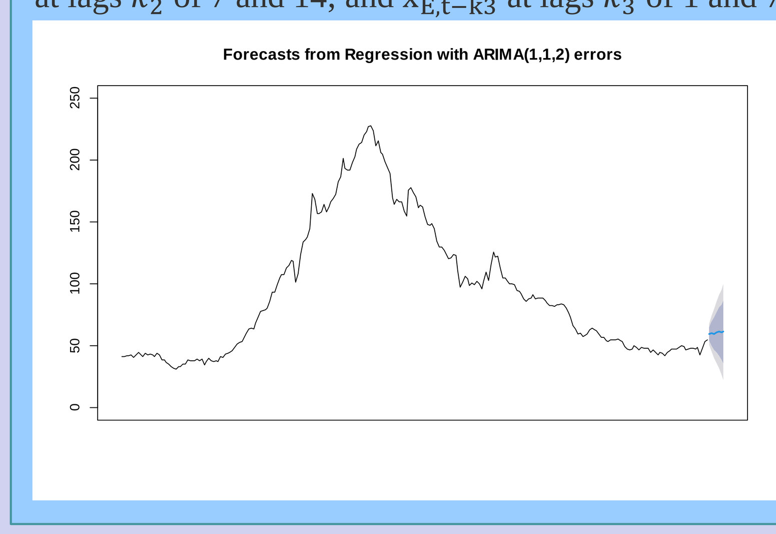
Statewide forecasting of Daily Hospitalizations

We use the lagged predictors model, with an ARIMA process for the errors [Section 9.6 Hyndman and Athanasopoulos]:

$$y_t = \beta_0 + \beta_Y x_{Y,t-k1} + \beta_M x_{M,t-k_2} + \beta_E x_{E,t-k_3} + \eta_t$$

where η_t follows an ARIMA process, and $x_{Y,t-14}$ represents the 7-day average Positive cases in one of the Young age groups at time *t-14* days (lagged 14 days prior).

Using the forecast package in R, we fit the response **7-day average Hospitalized cases**. The best model (assessed using AIC criterion) used ARIMA(1,1,2) errors and included $x_{Y,t-k1}$ at lags k_1 of 7 and 14; $x_{M,t-k2}$ at lags k_2 of 7 and 14; and $x_{E,t-k3}$ at lags k_3 of 1 and 7:



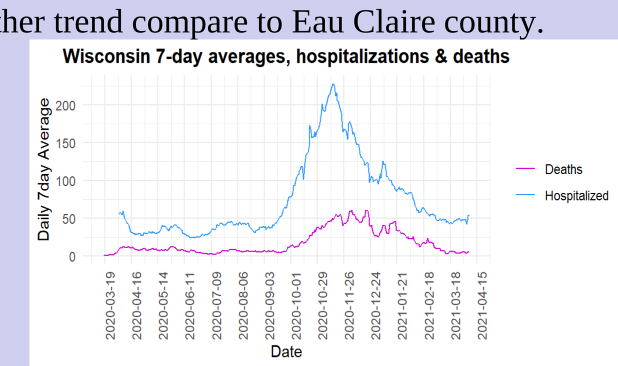
Forecasts shown for a 1-week future look past March 31

Why so wide? Vaccination effects have not yet been integrated into model

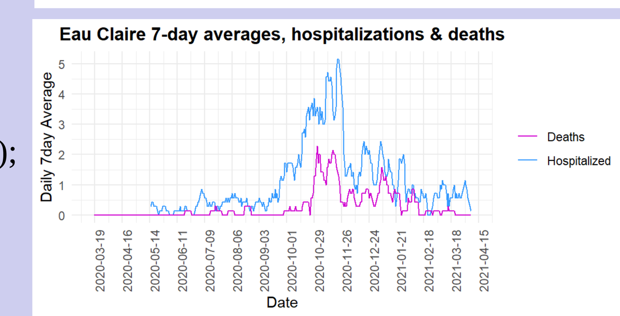
Why statewide?

- Bigger population provided more dataset which help to increase the accuracy of models to fit the data.
- WI statewide data has smother trend compare to Eau Claire county.

Plot 1: WI 7-day averages; Clear peaks and dips; Notable pattern (lag) between hospitalizations and deaths



Plot 2: WI 7-day averages; Patterns more abrupt due to lower counts (less population); Still discernable lag



Why hospitalizations?

- Help medical industry and government to predict the future health care needs in this COVID-19 pandemic.
- Communication of when hospital facilities are reaching limits.

Future Work

- Continue disseminating information via dashboard and communication with Eau Claire County health department
- Revising functions to be updatable to new situations, especially incorporating vaccination information (lag-time is important here)
- Embedding forecasting in dashboard

Acknowledgements

- This research project was funded by the Office of Research and Sponsored Programs (ORSP) during the 2020-21 academic year.
- Computing support and research space (both physical and virtual) from the Mathematics Department and UW-Eau Claire.

Data

Data gathered daily (or nearly so) from https://www.dhs.wisconsin.gov/covid-19/data.htm, open and supported for public access.

References – updated visuals may be found at source http://jjkraker.github.io/EauClaireCOVIDsummary/TabbedDashboard.html

- Forecasting: Principles and Practice (2nd ed), Rob J Hyndman and George Athanasopoulos, available online at https://otexts.com/fpp2/.
- Minnesota's Safe Learning Plan: https://education.mn.gov/MDE/dse/health/covid19/
- CDC guidance on home isolation: https://www.cdc.gov/coronavirus/2019-ncov/if-you-are-sick/end-home-isolation.html
- Statistics and Important Guidance from Eau Claire County: https://coronavirus-and-covid-19-information-hub-eccounty.hub.arcgis.com/pages/local-statistics