# An Investigation into the Temporal Lag Between Covid-19 Cases and Wastewater Measurements

Kyllan Wunder, Marlin Lee, Abe Megahed

University of Wisconsin Data Science Institute - December 2022

Introduction: In this analysis, we examine the correlation between wastewater data and case data, focusing on the temporal lag between the two. This lag is to be expected since we know there is a lag of several days between when shedding begins and when someone is tested.

Data: For this analysis, we used case data from the Department of Health Services (DHS) and wastewater data given to us by the State Lab of Hygiene (SLH). This data ranges from August 2020 to April 2022. This data contains the daily case data and the measured wastewater concentrations in 79 sites across Wisconsin. We took the mean of the data across every site to get a minimal variance signal. This minimizes natural variance and helps to capture the overall trend. A per-site analysis fails due to a lower sample rate and significant variance.
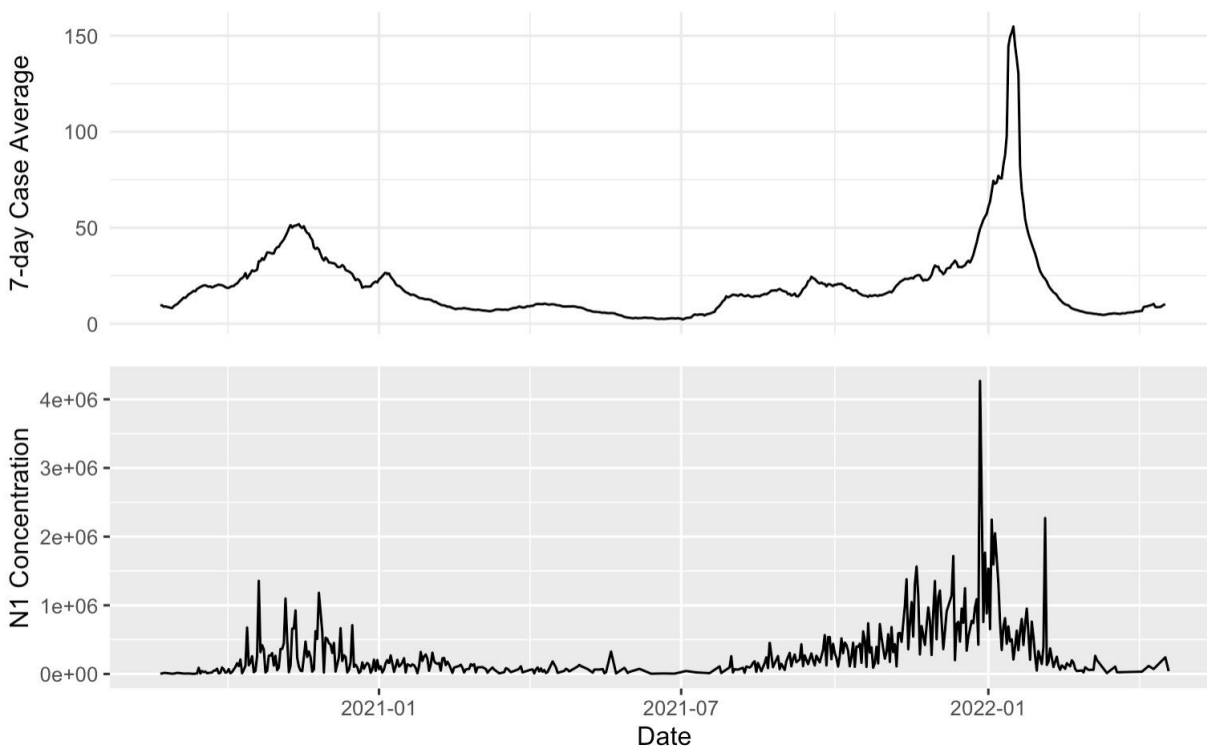
Methods: By shifting the wastewater data one day at a time and matching it up with case data, then using a correlation test, we can find how well the shifted (or lagged) wastewater data matches with case data. We chose the Kendall Correlation method as it makes no assumptions about the data distribution. The wastewater-to-case ratio is between the n1 (the concentration of measured virus particles) and the 7-day rolling average of cases. It is used to find a correlation between n1 and cases as the ratio will remain most constant when n1 and cases mirror each other. If one lags behind the other, the ratio will be less consistent.

Results:
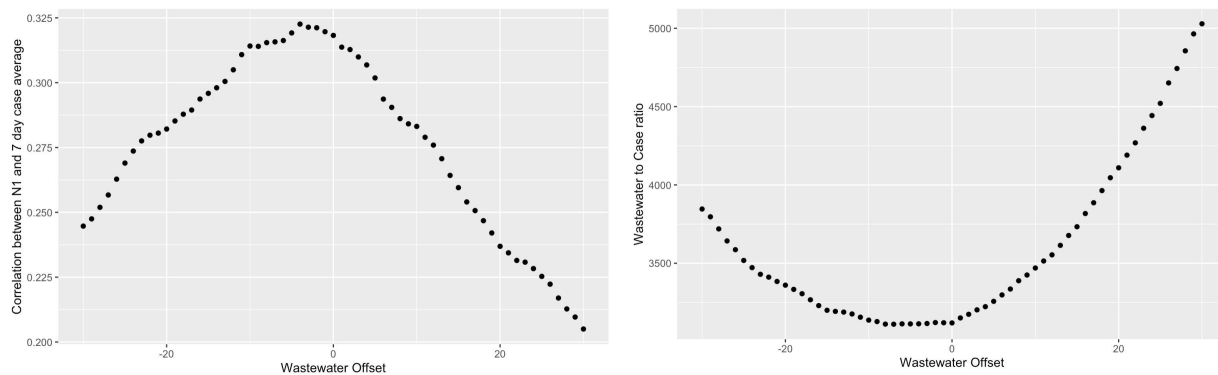
Conclusion: The Kendall method finds that wastewater precedes case data by approximately four days. This is also confirmed by the wastewater-to-case ratio, which is lowest between -8 to 0 days offset. Our findings are consistent with other findings that the waste-to-case lag time is between 0 to 6 days offset.

Future direction: A more robust correlation could be made between wastewater and total cases (i.e., not just new cases as we have here), as wastewater samples represent everyone who is shedding, not just new cases. Another future analysis could examine whether the lag time changes over time, by geographic region throughout the pandemic, or if it is constant. We speculate that various factors could influence the lag time, such as ease of access to testing, local attitudes toward testing, etc. This would require additional investigation to determine if these factors are visible within the data.