Estimated Covid Cases from Wastewater and the Reported Cases from Public and BU

1. Objective

In December 2021, Omicron has rapidly spread over US, and the daily new cases increased dramatically. The test results showed that the daily positive cases at BU increased higher than the public test results. Our hypothesis is the reason for higher number of positive cases is the test policy at BU, which asks students, faculties, and staff to have the test on campus. The test could detect the positive cases at the beginning and also the asymptomatic ones. Considering and asymptomatic and the self-test at home, we could assume that many of them would not report to the surveillance system, which made the estimated daily positive cases much less than the actual number.

However, the genome copies from wastewater shows that the positive cases are expected to be much more than the reported ones. Thus, we are wondering if we could use the data from wastewater to estimate the number of positive cases in a specific area and compare it with the report number and the estimated number from BU data. In this way, we might have a sense on the actual positive number.

The project could we divided into several small questions:

1. How to use the wastewater to estimate the positive cases in an area (6 counties in MA).
2. How to use the BU positive cases to estimate the positive cases around.
3. Comparison about the wastewater variance changes with BU changes.

1. Datasets we have
2. Wastewater by county data includes: the date (Wednesday of the week), normalized concentration rolling average (normalized genome copies per ml of wastewater), population, county name. Explanation of the value meaning is needed, can find the information from Wu, 2020, SARS-CoV-2 Titers in Wastewater Are Higher than Expected from Clinically Confirmed Cases.
3. Daily Wastewater: only have the Southern and Northern only, has variables of the number of copies/ml. Several missings.
4. Positive cases by county data includes: date (USA facts reported date), rolling average cases per 100k (centered 7-day rolling average), county name.
5. Covid-19 Death by county data includes: County name, daily death
6. BU Case data: may be pull from the covid dataset we want to create.
7. Weather Data: in progress (may not needed, since the PMMoV might already corrected it)
8. For question a. Data collections regarding to [1] [2]
9. Create a model with variables: such as precipitation (rainstorm, use PMMoV so might not need to be considered), death, population, degraded at different temperatures (be stored at 4C for more than 9-15 days without degradation) [1]. Not sure if the PMMoV also normalized based on the population.
10. Unknow: the timeline and load fecal shedding, loss of viral particles in sewage lines, the exact loss of RNA during experimental procedures [1]
11. Assumptions: typical stool size, average daily flow volume, population size, frequency of producing stool per day, no loss of viral RNA in processing, including collecting, excreted viruses. [1]
12. Consider the long incubation time, how long for the incubation time for different types of covid vires, does it have infectious impact during the incubation time?
13. Instead of daily incidence cases (7 -days rolling average), maybe daily prevalence data match better with the wastewater viral titers.
14. For question b.

References:

1. @article{wu2020sars,

title={SARS-CoV-2 titers in wastewater are higher than expected from clinically confirmed cases},

author={Wu, Fuqing and Zhang, Jianbo and Xiao, Amy and Gu, Xiaoqiong and Lee, Wei Lin and Armas, Federica and Kauffman, Kathryn and Hanage, William and Matus, Mariana and Ghaeli, Newsha and others},

journal={Msystems},

volume={5},

number={4},

pages={e00614--20},

year={2020},

publisher={Am Soc Microbiol}

}

[Nour.sharara@mail.harvard.edu](mailto:Nour.sharara@mail.harvard.edu) contact with wastewater and covid cases.