Research Question

- Unreported SARS-CoV-2 cases in Somerton
- Lag time between wastewater-based case numbers and clinical case numbers

Results

- 88% of data gathered supports the hypothesis clinical case numbers are less than a quarter of actual disease presence.
- Lag time can be approximated as a range between one and three months.

Methodology

- Filtration/concentration
- 20ul RT-QPCR reaction to quantify SARS-CoV-2 RNA
- 20ul RT-QPCR for PMMoV
- Application of equations to calculate estimated disease prevalence
- Normalization calculations

Conclusions

- Clinical data hugely underestimates case numbers, more than expected
- Lag time may be much longer than expected-- more research is necessary
- More research in normalization methods is necessary

Background

- The CDC estimates that only 1 in 4 SARS-CoV-2 cases are reported, and that even that number is an underestimate
- **Wastewater-based epidemiology (WWBE)** uses viral RNA found in sewage to track the incidence and distribution of disease
- It is being studied as a means of tracking and as a potential early warning system for infectious disease such as SARS-CoV-2
- Current wastewater sampling and PCR testing in Yuma covers about 70% of the city population and the majority of the population in the surrounding municipalities
- Determining the exact number of cases based on viral RNA fecal shedding is challenging due to the many correlating factors

Questions/Goals

1. Estimate the number of unreported SARS-CoV-2 cases in a well-defined Yuma municipality (Somerton) based on viral RNA quantities in wastewater samples compared to clinical data
2. Determine the approximate lag time between wastewater-based case numbers and clinical case numbers
    a. Create an epidemiological map and compare it to clinical data
3. Perform normalization of the SARS-CoV-2 RNA data to reduce noise and increase precision by estimating viral particles per person-- allowing for greater accuracy in answering questions
    a. Quantification of PMMoV in wastewater was used to estimate the number of individuals represented in each sample and thus scale the SARS-CoV-2 data to create more accurate comparisons between arrays

Hypotheses

- As stated, the CDC approximates that 1 in 4 cases are reported as of today, but claims that this estimate is likely low. Therefore, the hypothesis for the number of unreported cases in the Somerton area was some quantity greater than four times clinical case numbers at any given time.
    - However, it should be noted that formulating a well-evidenced hypothesis for this question was difficult, given that there is not yet a reliable method to approximate unreported case numbers
- Lag time between WWBE and clinical numbers is essentially the difference between when a person begins fecal shedding and when they become symptomatic to a degree significant enough that the individual gets tested. Based on the body of research available, lag time was hypothesized as somewhere within the range of 6-14 days.

Methods

Initial filtration/concentration

Methods: qPCR

- Primers and probes utilized:
  - SARS-CoV-2:
    - 2019 nCoV_N1 Forward Primer: 5’ - GAC CCC AAA ATC AGC GAA AT - 3’
    - 2019 nCoV_N1 Reverse Primer: 5’ - TCT GGT TAC TGC CAG TTG AAT CTG - 3’
    - 2019 nCoV_N1 Probe: 5’ - FAM-ACC CCG CAT TAC GTT TGG TGG ACC-BHQ1 - 3’
    - 2019-nCoV_N2 Forward Primer: 5’ - TTA CAA ACA TTG GCC GCA AA - 3’
    - 2019-nCoV_N2 Reverse Primer: 5’ - GCG CGA CAT TCC GAA GAA - 3’
    - 2019 nCoV_N2 Probe: 5’ - FAM-ACA ATT TGC CCC CAG CGC TTC AG-BHQ1 - 3’
  - PMMoV:
    - PMMV-FP1-rev: 5’ - GAG TGG TTT GAC CTT AAC GTT TGA - 3’
    - PMMV-RP1: 5’ - TTG TCG GTT GCA AAT GCA GT - 3’
    - PMMV-Probe1: 5’ - FAM-CCT ACC GAA GCA AAT G-BHQ1 - 3’
- 20ul reaction, variable number of samples, both positive and negative controls
- Bio-Rad CFX96 Touch Real-Time PCR Detection System thermocycler.

Thermocycler RT-qPCR Settings/Programs

N1 & N2	229E	PMMoV
1. RT (Reverse Transcriptase) 1: 10 min at 50C, 1 rep 2. RT 2: 10 min at 95C, 1 rep 3. PCR (Polymerase Chain Reaction) 3: 3 sec at 95C 4. PCR 4: 30 sec at 55C 5. PCR 5: Repeat steps 3 and 4 45x	1. RT 1: 10 min at 50C, 1 rep 2. RT 2: 10 min at 95C, 1 rep 3. PCR 3: 15 sec at 95C 4. PCR 4: 1 min at 60C 5. PCR 5: Repeat steps 3 and 4 45x	1. RT 1: 10 min at 50C, 1 rep 2. RT 2: 10 min at 95C, 1 rep 3. PCR 3: 5 sec at 95C 4. PCR 4: 1 min at 60C 5. PCR 5: Repeat steps 3 and 4 45x

Methods: Statistical Analysis

Application of equations to calculate approximate disease prevalence was modeled off of Curtis et al. (2020).

$$Load_{indiv} = C_{indiv} \times m$$

Equation 1 calculates viral load per individual.

$Load_{indiv}$ = viral load per individual (copies/day)

C_{indiv} = concentration of virus in fecal matter, i.e. shed rate (copies/g)

m = average mass of feces produced per individual per day (g/day)

Methods: Statistical Analysis

$$Load_{WWTP} = C_{WWTP} \times Q \times f$$

Equation 2 calculates total viral load of the wastewater treatment plant.

$Load_{WWTP}$= total viral load to WWTP (copies/day)

C_{WWTP}= concentration of virus in wastewater samples (copies/100mL)

Q= Plant flow (Millions of gallons [MG] per day)

f = conversion factor (100 mL to MG)

Methods: Statistical Analysis

$$I = \frac{Load_{WWTP}}{Load_{indiv}}$$

Equation 3 utilizes the products of Equations 1 and 2 to estimate disease prevalence.

I = approximate number of people in WWTP service area infected and at peak shedding rate

In addition to these equations, estimated infection per 10,000 people was calculated.

I / Population of WWTP service area * 10,000

Methods

Normalization

- Results were normalized to sample population via PMMoV quantification using the method modeled in Fuqing et al.
- One sample per month was normalized to depict general trends
- Deviation factor = deviation of PMMoV copies from median PMMoV copies

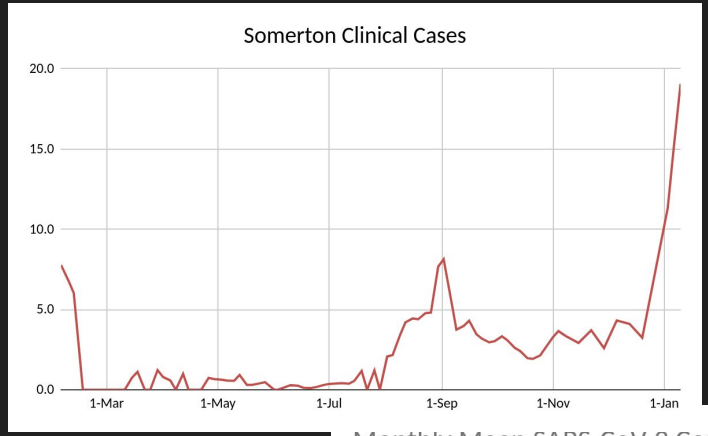
Deviation factor = 10 ^ [k x (sample CT - median CT)

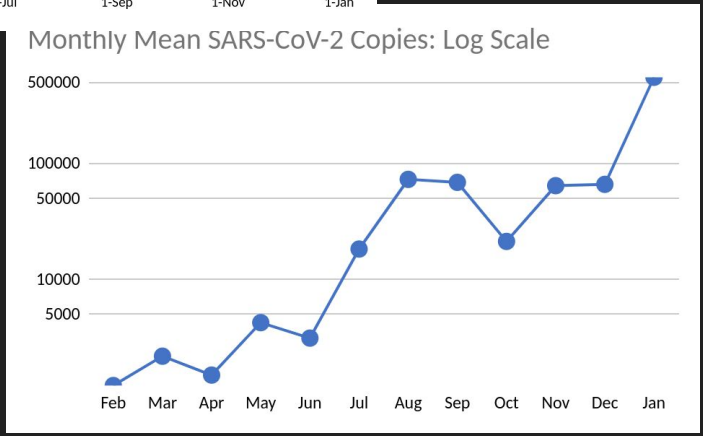
k is the slope of the standard curve (-3.058) and CT is the cycle threshold (also seen as cycle quantity, represented by CQ).

Normalized copies = Non-normalized monthly mean copies / deviation factor

Results/Discussion: Clinical vs. Monthly Mean

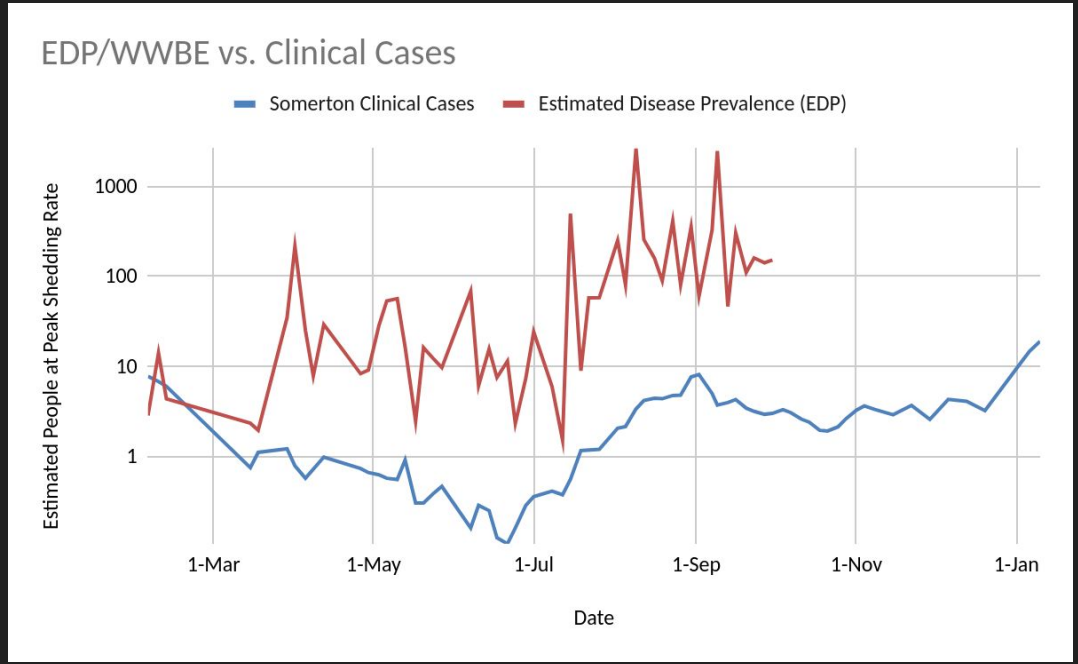
- To reduce noise and smooth out the trendline, the viral copies/L for each month was averaged.
- Three peaks were defined in which lag between data sets was clear. (May vs. August, August, vs. September, early December vs. late December)
- “Lag time” can be approximated as a range between one and three months. Based on that conclusion, the lag time hypothesis was an underestimation.





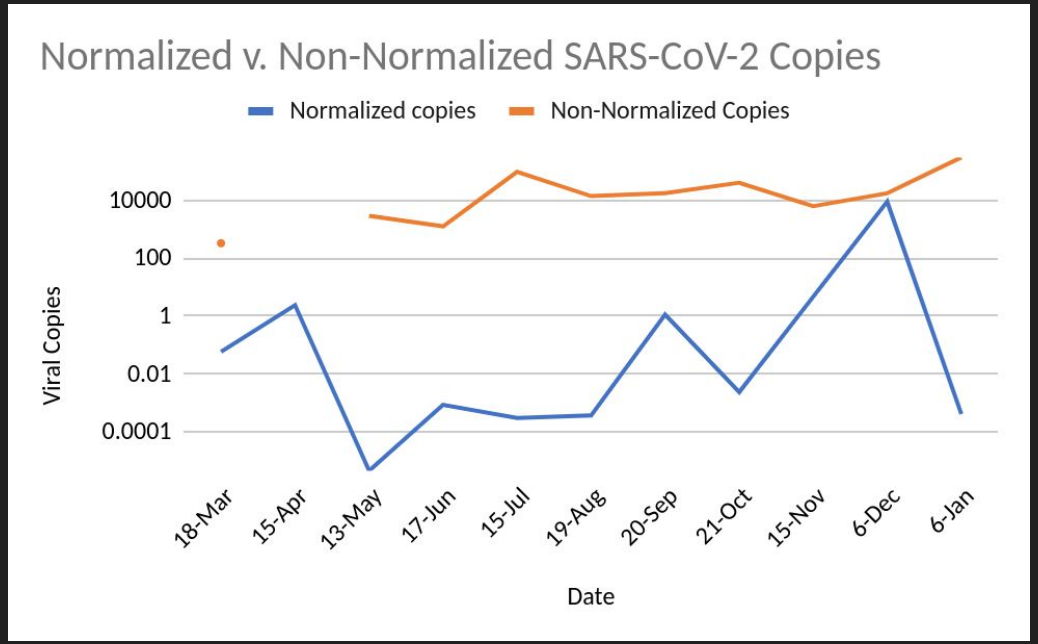
Results/Discussion: EDP vs. Clinical

- The average difference was that EDP case numbers were 178 greater than clinical case numbers.
- Out of 51 data points, only 6 EDP values were less than four times clinical case numbers. So, 88% of data gathered supports the hypothesis clinical case numbers are less than a quarter of actual disease presence.



Results/Discussion: Normalized vs. Non-Normalized

- Normalized values were significantly lower and had more variation than raw data
- Normalized conflicts with monthly mean graph: April spike
- Contour of normalized graph aligns with clinical, yet lags behind
- Conflict with lag time hypothesis-- normalized determined to be less reliable than monthly mean



Limitations, Future Research, and Acknowledgements

- Limitations
  - Unknowns and outside factors-- individual fecal shedding rate, variable m, etc
  - Availability of data collection locations and clinical data
- Future research
  - Fecal shedding differences based on demographic
  - Optimal normalization methods
- Acknowledgements
  - STAR Labs program
    - Research director Margaret Wilch
    - Facilitator Evy Nguyen
  - Mentor Dr. Christian Roessler
  - Dr. Stephanie Slinski
  - Dr. Bradley Schmitz