

Virginia Wastewater Surveillance:
Can SARS-CoV-2 Once-Weekly Sampling Predict Imminent Rises in Community COVID-19
Disease Burden?

Michelle Marie Yancey

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in
partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Individual Interdisciplinary

Kerry Redican, Chair
Korine Kolivras
Kris Wernstedt
Chris Zobel
William Huckle, Advisor

July 21, 2025
Blacksburg, Virginia

Keywords: wastewater, surveillance, epidemiology, COVID-19, SARS-CoV-2, alert

Copyright 2025, Michelle Marie Yancey

Virginia Wastewater Surveillance: Can SARS-CoV-2 Once-Weekly Sampling Predict Imminent Rises in Community COVID-19 Disease Burden?

Michelle Marie Yancey

ABSTRACT

While wastewater surveillance (WWS) has been a tool arguably used since the days of John Snow, its use has increased sharply – and far more publicly – since the beginning of the COVID-19 pandemic. WWS typically involves sample collection of community sewage at the wastewater treatment plant (WWTP), where it is then analyzed for viral particle counts reflecting the entire WWTP’s service area (i.e., *sewershed*). This dataset is independent of other public health data, helping contextualize community disease burden. This is especially important in the absence of reliable case reporting. However, with WWS being so new, many public health practitioners struggle to understand and apply its nuance.

There is a need to provide easy, actionable interpretation for SARS-CoV-2 WWS in Virginia. To support the Virginia Department of Health (VDH), this PhD dissertation explored the SARS-CoV-2 WWS VDH Sentinel Monitoring Program (SMP) data and correlated with community COVID-19 burden indicators of cases and hospitalizations. I then used Yang’s (2017) framework for Early Warning Systems to build an alert system, utilizing Shewhart’s (1931) Simple Control Charts with moving ranges as the base.

The chosen variable for the alert system was by-site change in viral load (VL), with a dependent testing variable of by-site change in weekly cases. All data was assessed by individual site (n=24), as inter-site variability was extensive. Baselines for establishing normal site variance for VL and case change were determined by finding periods when case change and SARS-CoV-2

variant shifts were minimized. Once baselines were determined, the standard deviation (SD) of change was calculated for each site, with a multiplier of 3 to set the Upper Control Limit (UCL) of acceptable variance. The alert system indicated which changes of VL and cases were sufficiently out of normal variance by assigning flags. The flags for both VL and cases were compared to each other, and performance metrics calculated.

When using the SD multiplier threshold of 3 for both changing VL and cases (daily cases for next 7 days), the alert flags matched each other 88% of the time. If calculated for case flags of that week or following week (accounting for possible VL leading indicator effects), matching increased to 92%. For every positive VL alert, 49% had an associated positive case alert that week; 52% for the next 1-2 weeks. For every negative VL alert, 94% had an associated negative case alert that week; 97% for the next 1-2 weeks. Sensitivity was calculated at 39% for the same week, or 52% for the next 1-2 weeks. Specificity was calculated at 94% for the same week, or 99% for the next 1-2 weeks. These metrics were explored using other thresholds, which did improve certain performance indicators (e.g., sensitivity), but usually at the cost of other indicators. Major limitations include nuances of the datasets themselves, especially with the case counts, which were subject to changing reporting standards and methods throughout the pandemic.

This dissertation provides a proposed and evaluated alert system for WWS programs, with specific guidance and tips in Chapters 11-12 for readers. It follows up with a mock scenario retrospective test using the Omicron spike of December 2021. The methods and principles used here can be used by public health practitioners for SARS-CoV-2, as well as potentially other WWS targets.

Virginia Wastewater Surveillance: Can SARS-CoV-2 Once-Weekly Sampling Predict Imminent Rises in Community COVID-19 Disease Burden?

Michelle Marie Yancey

GENERAL AUDIENCE ABSTRACT

Wastewater Surveillance (WWS) is a relatively new tool to public health, finding footing during the COVID-19 pandemic as a snapshot viewing of a community's disease burden. However, due to this novelty, it is often difficult for the public -- and public health -- to accurately understand its nuance. The lack of gold standard for interpretation often leaves public health practitioners asking how meaningful these WWS *viral loads* are. The Virginia Department of Health's (VDH) WWS Sentinel Monitoring Program (SMP) tests at least 25 sites weekly for SARS-CoV-2 viral load, but there is no alert system in place to help inform viewers of its meaning. This dissertation takes a deep dive into the SMP data and addresses its nuances. While the viral loads themselves were found to be highly correlated to cases and hospitalizations, the most significant finding was the extensive variation found between and within the collection sites themselves. It is for this reason that any interpretation and developed alert system must be focused on the individual sites. This dissertation proposed a framework for such an alert system, tested the performance, and then applied it to the Omicron peak of 2021-2022. The alert system was determined to have excellent matching between viral load spikes and case spikes, and could be used to inform users if there is a likelihood of case rises on the horizon. This viral load alert framework can theoretically be used for any WWS project, providing a useful tool for WWS interpretation for VDH and others moving forward.

Dedications

I'd like to thank my Committee for being the exact kind of committee I needed for success -- one that was helpful and far away. And also, the kind that calls when I'm particularly unresponsive.

In the same vein, I'd like to give my appreciation to the Virginia Tech Individualized Interdisciplinary PhD program itself. It allowed me to play in a giant intellectual sandbox, getting absorbed where winds and interest took me. In my case, right into the heart of COVID-19 public health. There aren't a lot of PhD programs out there that would let me dance so wildly between the seemingly disparate fields of public health, data analytics, and disaster preparedness. Though, in our COVID-19 world, these fields are now more connected than ever. And, considering that I've been asked to guest lecture for MPH students at several universities, I wouldn't be surprised if public health itself has/is shifting towards inclusion of those disciplines.

I'd also like to thank the Virginia Department of Health. Not only did they pay me a good salary to build their wastewater surveillance system -- as part of an interdisciplinary team of environmental engineers and epidemiologists -- they gave me the freedom to do it my way. I got to write the book, so to speak. Any future employees in their wastewater surveillance program will likely use the data templates I've created, reference the reports I've written, and perhaps even use the code I've generated. I've become a builder of systems.

I also want to thank my public health mentor, Virginia regional epidemiologist Paige Bordwine, who recruited me into the Medical Reserve Corps during my time as a VT MPH graduate student. I still want to be her when I grow up.

And, of course, my deepest appreciation goes to my spouse and cats, who provided immeasurable support throughout the entire PhD process. One of them even read this dissertation in its entirety.

Acknowledgements

Much of the data used in this dissertation was pulled from the Virginia Department of Health's Sentinel Monitoring Program (SMP). While I was the one who created and sent this dataset (to myself), I must acknowledge the extensive, inter-agency historical development of this data.

This network includes:

- CDC NWSS – for funding the SMP program in its entirety, including the sampling, testing, and staff salary costs.
- VDH – for housing the SMP, as well as all affiliated assistance from adjoining departments (e.g., Epidemiology, Environmental Health, etc.).
- Wastewater utilities – who volunteered their time and resources to support the program by collecting WWTP samples.
- Laboratory partners – including HRSD and DCLS, who provided expertise and analysis for all the samples, in a format that was amenable for me to submit results to CDC DCIPHER.
- SMP itself – in particular my wonderful colleagues Haniyyah Majeed, Isaiah Omerhi, and Caroline Marti, who contributed to successful SMP data creation. And offered themselves as willing sacrifices for idea bouncing and troubleshooting in all things.

Table of Contents

Chapter 1: Introduction	1
1.1 Author Context.....	1
1.2 Wastewater Surveillance Defined	2
1.3 Dissertation Purpose.....	4
Chapter 2: Review of Literature	6
2.1 Methods.....	6
2.2 Historical WWS	9
2.3 Modern (Pre-Pandemic) WWS	14
2.4 COVID-19 WWS	15
Chapter 3: Variability and Limitations	23
3.1 Target	23
3.2 Viral Decay in Sewage.....	25
3.3 Sampling.....	26
3.4 Laboratory Analytical Methods	29
3.5 Interpretation & Data Analysis	31
3.6 Major Limitations.....	32
3.7 Major Advantages	34
Chapter 4: Alert System Fundamentals	35
4.1 Why Develop an Alert System?	35
4.2 What makes an Alert System?	36
4.3 Is the SMP an Alert System?.....	39
4.4 History of Infectious Disease Early Warning	41
4.5 Summary of Existing U.S. Alert Systems	43
4.6 On Selecting an Alert System Model.....	49
Chapter 5: Ethics.....	52
5.1 Practical Considerations.....	52
5.2 Applications	52
5.3 Ethical Frameworks.....	54

5.4 Drug Surveillance.....	58
5.5 Virginia Executive Order 26	59
5.6 EO-26 Ethical Issues	61
Chapter 6: Data & Analysis Methods	63
6.1 IRB Determination	63
6.2 Data Sources.....	63
6.3 About the WWS Data.....	65
Chapter 7: Initial Data Exploration.....	71
7.1 Variant Dominance	71
7.2 Sewersheds	79
7.3 Case Counts.....	80
7.4 WWS Analysis Data.....	82
Chapter 8: Case Correlation.....	94
8.1 VL & Daily Cases (general).....	94
8.2 VL & Weekly Cases (general)	95
8.3 VL & Daily Cases (by variant)	98
8.4 VL & Weekly Cases (by variant).....	103
8.5 Site-Specific VL & Daily Cases	108
8.6 Site-Specific VL & Weekly Cases	109
Chapter 9: Hospitalizations Correlation.....	111
9.1 Weekly VL & Daily Hospitalizations (general).....	111
9.2 Weekly VL & Weekly Hospitalizations (general)	112
9.3 Weekly VL & Daily Hospitalizations (by variant)	114
9.4 Weekly VL & Weekly Hospitalizations (by variant).....	118
9.5 Site-Specific VL & Daily Hospitalizations	121
9.6 Site-Specific VL & Weekly Hospitalizations	123
Chapter 10: Developing an ‘Alert’ Framework	124
10.1 Independent and Dependent Variables.....	124
10.2 Selecting a Base Model	125

10.3 Simple Control Charts Principles	127
10.4 Exploring Normality	129
10.5 Proposed Model.....	135
Chapter 11: Establishing Baseline & Applying Alerts	137
11.1 Establishing Baseline	138
11.2 Organizing WWS and Case Counts by Site	142
11.3 Assigning VL and Case Alerts	143
11.4 VL and Case Alert Matching	145
Chapter 12: Evaluating and Adjusting the Alert System	151
12.1 Summary Statistics.....	152
12.2 Adjusting XSD	154
Chapter 13: Omicron Retrospective Test.....	157
Chapter 14: Discussion & Conclusions	165
References	173
Appendix – 8.5 By Site VL & Daily Cases	183
Appendix – 8.6 By Site VL & Weekly Cases.....	189
Appendix – 9.5 By Site VL & Daily Hospitalizations.....	196
Appendix – 9.6 By Site VL & Weekly Hospitalizations	202

List of Figures and Tables

FIGURE 2.1: SCREENSHOT OF WBE PUBLICATION MAP (COVID19WBEC, 2023). ACCESSED MARCH 29, 2023.....	8
FIGURE 2.4.1: SCREENSHOTS FROM HRSD (2023), AGGREGATE VIRAL LOADS AND HOSPITALIZATIONS. ACCESSED JULY 13, 2023.....	19
FIGURE 2.4.2: SCREENSHOT FROM CDC COVID DATA TRACKER (CDC, 2023B). ACCESSED JULY 19, 2023.....	20
FIGURE 2.4.3: SCREENSHOT FROM CDC CLINICAL VARIANT PROPORTIONS (CDC, 2023A).ACCESSED JULY 19, 2023.....	21
FIGURE 2.4.4: SCREENSHOT FROM FDA WWS VARIANT ABUNDANCES (FDA, 2023). ACCESSED JULY 19, 2023.....	22
FIGURE 4.2: SUMMARIZED VERSION OF YANG’S CONCEPTUAL FRAMEWORK FOR AN EARLY WARNING SYSTEM.....	38
FIGURE 4.5: SCREENSHOT WITHIN NSSP ESSENCE 101 TRAINING (CDC, 2024A). ACCESSED NOV 17, 2024.....	45
TABLE 5.2: BARRIERS TO WWS (McCLARY-GUTIERREZ ET AL., 2021)	54
FIGURE 6.3: SCREENSHOT OF WASTEWATER METRIC MAP ON CDC COVID DATA TRACKER (CDC, 2023A)	66
TABLE 6.3.1: DCLS LAB ANALYSIS DATA DESCRIPTORS	67
TABLE 6.3.2: VDH WWS SM DATASET DESCRIPTORS	68
FIGURE 7.1.1: EXCEL SUMMARY OF FDA WW VARIANTS CONTAINING OVER 65% RELATIVE ABUNDANCE BY WEEK.....	72
TABLE 7.1.1: ALL VDH CLINICAL CASE VARIANT DESCRIPTIVE STATISTICS AND WHO LABEL.....	73
FIGURE 7.1.3: WEEKLY DELTA VARIANT CLINICAL CASE COUNTS OVER TIME	75
TABLE 7.1.2: VDH VARIANT CASE COUNT SUBSELECTIONS BY SUBLINEAGES, CASE COUNTS, AND DATE RANGES	75
FIGURE 7.1.4: WEEKLY OMICRON VARIANT CLINICAL CASE COUNTS OVER TIME.....	75
FIGURE 7.1.5: WEEKLY BA.1 VARIANT CLINICAL CASE COUNTS OVER TIME.....	76
FIGURE 7.1.6: WEEKLY BA.2 VARIANT CLINICAL CASE COUNTS OVER TIME.....	76
FIGURE 7.1.7: WEEKLY BA.5 VARIANT CLINICAL CASE COUNTS OVER TIME.....	76
FIGURE 7.1.8: WEEKLY BA.5 VARIANT CLINICAL CASE COUNTS OVER TIME.....	76
FIGURE 7.1.9: WEEKLY DELTA AND OMICRON VARIANT CLINICAL CASE COUNTS OVER TIME	77
FIGURE 7.1.10: WEEKLY DELTA AND BA.1 VARIANT CLINICAL CASE COUNTS OVER TIME.....	77
FIGURE 7.1.11: WEEKLY BA.1 AND BA.2 VARIANT CLINICAL CASE COUNTS OVER TIME.....	77
FIGURE 7.1.12: WEEKLY BA.2 AND BA.5 VARIANT CLINICAL CASE COUNTS OVER TIME.....	78
FIGURE 7.1.13: WEEKLY CLINICAL BA.5 AND XBB.1.5 VARIANT CLINICAL CASE COUNTS OVER TIME.....	78
FIGURE 7.1.14: WEEKLY VARIANT CLINICAL CASE COUNTS OVER TIME, LINE GRAPH	78
FIGURE 7.1.15: WEEKLY VARIANT CLINICAL CASE COUNTS OVER TIME, COLUMN GRAPH	79
TABLE 7.1.3: SYNTHESIZED DOMINANT VARIANT TIMELINE	79
FIGURE 7.3.1: DAILY VDH PUBLIC COVID-19 CASE COUNTS OVER TIME, STACKED AREA GRAPH	80
FIGURE 7.3.2: DAILY VDH PUBLIC COVID-19 HOSPITALIZATIONS OVER TIME, STACKED.....	80
FIGURE 7.3.3: DAILY VDH PUBLIC COVID-19 DEATHS OVER TIME, STACKED AREA GRAPH	81
FIGURE 7.3.4: DAILY VDH SMP CASE COUNTS OVER TIME, COLUMN GRAPH	81
FIGURE 7.3.5: VDH WEEKLY VARIANT CLINICAL CASE COUNTS AND DAILY VDH SM SEWERSHED CASE COUNTS OVER TIME.....	82
TABLE 7.4.1: SMP SITE INFORMATION INCLUDING #, NAME, POPULATION, MGD, AND ZIP	83
TABLE 7.4.2: SMP SITE COLLECTION AND METADATA STATISTICS	84
TABLE 7.4.3: SMP SITE ANALYTIC STATISTICS AND FLOW METADATA; LOWER (GREEN) VERSUS HIGHER (RED) VARIANCE.....	87
TABLE 7.4.4: SMP SITE VARIANCE STATISTICS AND RANKINGS; LOWER (GREEN) VERSUS HIGHER (RED) VARIANCE	89
TABLE 7.4.5: SMP 5 MOST VARIABLE SITES; LOWER (GREEN) VERSUS HIGHER (RED) VARIANCE.....	90
TABLE 7.4.6: SMP 5 LEAST VARIABLE SITES; LOWER (GREEN) VERSUS HIGHER (RED) VARIANCE.....	91
FIGURE 7.4.1: AGGREGATE VIRAL LOADS FOR ALL SMP SITES OVER TIME.....	92
FIGURE 7.4.2: AGGREGATE DAILY CASES FOR ALL SMP SITES OVER TIME	93
FIGURE 7.4.3: AGGREGATED SMP DAILY CASE COUNTS AND WEEKLY VIRAL LOADS OVER TIME	93
TABLE 8.1.1: SELECTED PEAK PERIOD FOR SM AGGREGATED VIRAL LOADS AND DAILY CASES CORRELATION WITH DAY SHIFTS	95
FIGURE 8.1.2: AGGREGATED VIRAL LOADS AND DAILY CASES CORRELATION COEFFICIENTS WITH DAY SHIFTS, LINE GRAPH.....	96

FIGURE 8.2.1: AGGREGATED SMP VIRAL LOADS AND WEEKLY (NEXT 7 DAYS') CASE COUNTS OVER TIME	97
TABLE 8.2.1: SELECTED PEAK PERIOD FOR SM AGGREGATED VIRAL LOADS AND WEEKLY CASES CORRELATION AND DAY SHIFTS	97
FIGURE 8.2.2: SM AGGREGATED VIRAL LOADS AND WEEKLY CASES CORRELATION COEFFICIENTS WITH DAY SHIFTS, LINE GRAPH.....	97
TABLE 8.3.1: SELECTED CORRELATION PERIOD FOR SM AGGREGATED VIRAL LOADS FOR DELTA VARIANT AND DAILY CASES CORRELATION COEFFICIENTS WITH DAY SHIFTS	99
FIGURE 8.3.1: SM AGGREGATED VIRAL LOADS FOR DELTA VARIANT AND DAILY CASES CORRELATION COEFFICIENTS WITH DAY SHIFTS	99
TABLE 8.3.2: SELECTED CORRELATION PERIOD FOR SM AGGREGATED VIRAL LOADS FOR BA.1 VARIANT AND DAILY CASES CORRELATION COEFFICIENTS WITH DAY SHIFTS	99
FIGURE 8.3.2: SM AGGREGATED VIRAL LOADS FOR BA.1 VARIANT AND DAILY CASES CORRELATION COEFFICIENTS WITH DAY SHIFTS	99
TABLE 8.3.3: SELECTED CORRELATION PERIOD FOR SM AGGREGATED VIRAL LOADS FOR BA.2 VARIANT AND DAILY CASES CORRELATION COEFFICIENTS WITH DAY SHIFTS	100
FIGURE 8.3.3: SM AGGREGATED VIRAL LOADS FOR BA.2 VARIANT AND DAILY CASES CORRELATION COEFFICIENTS WITH DAY SHIFTS	100
TABLE 8.3.4: SELECTED CORRELATION PERIOD FOR SM AGGREGATED VIRAL LOADS FOR BA.5 VARIANT AND DAILY CASES CORRELATION COEFFICIENTS WITH DAY SHIFTS	100
FIGURE 8.3.4: SM AGGREGATED VIRAL LOADS FOR BA.5 VARIANT AND DAILY CASES CORRELATION COEFFICIENTS WITH DAY SHIFTS	100
TABLE 8.3.5: SELECTED CORRELATION PERIOD FOR SM AGGREGATED VIRAL LOADS FOR XBB.1.5 VARIANT AND DAILY CASES CORRELATION COEFFICIENTS WITH DAY SHIFTS.....	101
FIGURE 8.3.5: SM AGGREGATED VIRAL LOADS FOR XBB.1.5 VARIANT AND DAILY CASES CORRELATION COEFFICIENTS WITH DAY SHIFTS	101
TABLE 8.3.6: VL - DAILY CASES CORRELATION BY VARIANT.....	102
TABLE 8.4.1: SELECTED CORRELATION PERIOD FOR SM AGGREGATED VIRAL LOADS FOR DELTA VARIANT AND WEEKLY CASES CORRELATION COEFFICIENTS WITH DAY SHIFTS	104
FIGURE 8.4.1: SM AGGREGATED VIRAL LOADS FOR DELTA VARIANT AND WEEKLY CASES CORRELATION COEFFICIENTS WITH DAY SHIFTS	104
TABLE 8.4.2: SELECTED CORRELATION PERIOD FOR SM AGGREGATED VIRAL LOADS FOR BA.1 VARIANT AND WEEKLY CASES CORRELATION COEFFICIENTS WITH DAY SHIFTS	104
FIGURE 8.4.2: SM AGGREGATED VIRAL LOADS FOR BA.1 VARIANT AND WEEKLY CASES CORRELATION COEFFICIENTS WITH DAY SHIFTS	104
TABLE 8.4.3: SELECTED CORRELATION PERIOD FOR SM AGGREGATED VIRAL LOADS FOR BA.2 VARIANT AND WEEKLY CASES CORRELATION COEFFICIENTS WITH DAY SHIFTS	105
FIGURE 8.4.3: SM AGGREGATED VIRAL LOADS FOR BA.2 VARIANT AND WEEKLY CASES CORRELATION COEFFICIENTS WITH DAY SHIFTS	105
TABLE 8.4.4: SELECTED CORRELATION PERIOD FOR SM AGGREGATED VIRAL LOADS FOR BA.5 VARIANT AND WEEKLY CASES CORRELATION COEFFICIENTS WITH DAY SHIFTS	105
FIGURE 8.4.4: SM AGGREGATED VIRAL LOADS FOR BA.5 VARIANT AND WEEKLY CASES CORRELATION COEFFICIENTS WITH DAY SHIFTS	105
TABLE 8.4.5: SELECTED CORRELATION PERIOD FOR SM AGGREGATED VIRAL LOADS FOR XBB.1.5 VARIANT AND WEEKLY CASES CORRELATION COEFFICIENTS WITH DAY SHIFTS	106
FIGURE 8.4.5: SM AGGREGATED VIRAL LOADS FOR XBB.1.5 VARIANT AND WEEKLY CASES CORRELATION COEFFICIENTS WITH DAY SHIFTS	106
TABLE 8.4.6 WEEKLY CASE CORRELATION BY VARIANT AND DAY SHIFT	107
TABLE 8.5.1 DAILY CASES CORRELATION COEFFICIENT BY SITE AND DAY SHIFT	109
TABLE 8.6.1 WEEKLY CASES CORRELATION BY SITE AND DAY SHIFT	110
FIGURE 9.1.1 AGGREGATED VDH DAILY HOSPITALIZATION AND SMP WEEKLY VIRAL LOADS OVER TIME, LINE GRAPH	111

TABLE 9.1.1 AGGREGATED WEEKLY VIRAL LOADS AND DAILY HOSPITALIZATIONS CORRELATION COEFFICIENTS WITH DAY SHIFTS	112
FIGURE 9.1.2 AGGREGATED WEEKLY VIRAL LOADS AND DAILY HOSPITALIZATIONS CORRELATION COEFFICIENTS WITH DAY SHIFTS, LINE GRAPH.....	112
FIGURE 9.2.1 AGGREGATED SMP WEEKLY VIRAL LOADS AND WEEKLY (NEXT 7 DAYS) HOSPITALIZATIONS OVER TIME	113
TABLE 9.2.1 AGGREGATED WEEKLY VIRAL LOADS AND WEEKLY HOSPITALIZATIONS CORRELATION COEFFICIENTS WITH DAY SHIFTS	113
FIGURE 9.2.2 AGGREGATED WEEKLY VIRAL LOADS AND WEEKLY HOSPITALIZATIONS CORRELATION COEFFICIENTS WITH DAY SHIFTS, LINE GRAPH.....	113
TABLE 9.3.1 AGGREGATED WEEKLY VIRAL LOADS AND DAILY HOSPITALIZATIONS CORRELATION COEFFICIENT WITH DAY SHIFTS FOR DELTA VARIANT	115
FIGURE 9.3.1 AGGREGATED WEEKLY VIRAL LOADS AND DAILY HOSPITALIZATIONS CORRELATION COEFFICIENT WITH DAY SHIFTS FOR DELTA VARIANT, LINE GRAPH.....	115
TABLE 9.3.2 AGGREGATED WEEKLY VIRAL LOADS AND DAILY HOSPITALIZATIONS CORRELATION COEFFICIENT WITH DAY SHIFTS FOR BA.1 VARIANT	115
FIGURE 9.3.2 AGGREGATED WEEKLY VIRAL LOADS AND DAILY HOSPITALIZATIONS CORRELATION COEFFICIENT WITH DAY SHIFTS FOR BA.1 VARIANT, LINE GRAPH	115
TABLE 9.3.3 AGGREGATED WEEKLY VIRAL LOADS AND DAILY HOSPITALIZATIONS CORRELATION COEFFICIENT WITH DAY SHIFTS FOR BA.2 VARIANT	116
FIGURE 9.3.3 AGGREGATED WEEKLY VIRAL LOADS AND DAILY HOSPITALIZATIONS CORRELATION COEFFICIENT WITH DAY SHIFTS FOR BA.2 VARIANT, LINE GRAPH	116
TABLE 9.3.4 AGGREGATED WEEKLY VIRAL LOADS AND DAILY HOSPITALIZATIONS CORRELATION COEFFICIENT WITH DAY SHIFTS FOR BA.5 VARIANT	116
FIGURE 9.3.4 AGGREGATED WEEKLY VIRAL LOADS AND DAILY HOSPITALIZATIONS CORRELATION COEFFICIENT WITH DAY SHIFTS FOR BA.5 VARIANT, LINE GRAPH	116
TABLE 9.3.5 AGGREGATED WEEKLY VIRAL LOADS AND DAILY HOSPITALIZATIONS CORRELATION COEFFICIENT WITH DAY SHIFTS FOR XBB.1.5 VARIANT	116
FIGURE 9.3.5 AGGREGATED WEEKLY VIRAL LOADS AND DAILY HOSPITALIZATIONS CORRELATION COEFFICIENT WITH DAY SHIFTS FOR XBB.1.5 VARIANT, LINE GRAPH.....	116
TABLE 9.3.6 SUMMARIZED WEEKLY VIRAL LOADS AND DAILY HOSPITALIZATIONS CORRELATION COEFFICIENT WITH DAY SHIFTS FOR ALL VARIANTS	117
TABLE 9.4.1 AGGREGATED WEEKLY VIRAL LOADS AND WEEKLY HOSPITALIZATIONS CORRELATION COEFFICIENT WITH DAY SHIFTS FOR DELTA VARIANT	118
FIGURE 9.4.1 AGGREGATED WEEKLY VIRAL LOADS AND WEEKLY HOSPITALIZATIONS CORRELATION COEFFICIENT WITH DAY SHIFTS FOR DELTA VARIANT, LINE GRAPH.....	118
TABLE 9.4.2 AGGREGATED WEEKLY VIRAL LOADS AND WEEKLY HOSPITALIZATIONS CORRELATION COEFFICIENT WITH DAY SHIFTS FOR BA.1 VARIANT	118
FIGURE 9.4.2 AGGREGATED WEEKLY VIRAL LOADS AND WEEKLY HOSPITALIZATIONS CORRELATION COEFFICIENT WITH DAY SHIFTS FOR BA.1 VARIANT, LINE GRAPH	118
TABLE 9.4.3 AGGREGATED WEEKLY VIRAL LOADS AND WEEKLY HOSPITALIZATIONS CORRELATION COEFFICIENT WITH DAY SHIFTS FOR BA.2 VARIANT	119
FIGURE 9.4.3 AGGREGATED WEEKLY VIRAL LOADS AND WEEKLY HOSPITALIZATIONS CORRELATION COEFFICIENT WITH DAY SHIFTS FOR BA.2 VARIANT, LINE GRAPH	119
TABLE 9.4.4 AGGREGATED WEEKLY VIRAL LOADS AND WEEKLY HOSPITALIZATIONS CORRELATION COEFFICIENT WITH DAY SHIFTS FOR BA.5 VARIANT	119
FIGURE 9.4.4 AGGREGATED WEEKLY VIRAL LOADS AND WEEKLY HOSPITALIZATIONS CORRELATION COEFFICIENT WITH DAY SHIFTS FOR BA.5 VARIANT, LINE GRAPH	119
TABLE 9.4.5 AGGREGATED WEEKLY VIRAL LOADS AND WEEKLY HOSPITALIZATIONS CORRELATION COEFFICIENT WITH DAY SHIFTS FOR XBB.1.5 VARIANT	119
FIGURE 9.4.5 AGGREGATED WEEKLY VIRAL LOADS AND WEEKLY HOSPITALIZATIONS CORRELATION COEFFICIENT WITH DAY SHIFTS FOR XBB.1.5 VARIANT, LINE GRAPH.....	119

TABLE 9.4.6 WEEKLY HOSPITALIZATIONS CORRELATION COEFFICIENT BY VARIANT	120
TABLE 9.5.1 BY SITE VL - DAILY HOSPITALIZATIONS CORRELATION COEFFICIENT WITH DAY SHIFTS	122
TABLE 9.6.1 BY SITE VL - WEEKLY HOSPITALIZATIONS CORRELATION COEFFICIENT WITH DAY SHIFTS	123
FIGURE 10.4.1 SITE 1 TOTAL DATASET DESCRIPTIVE STATISTICS AND SKEW HISTOGRAMS	130
FIGURE 10.4.2 SITE 1 TOTAL DATASET DESCRIPTIVE STATISTICS AND SKEW HISTOGRAMS	130
TABLE 10.4.1 BY-SITE DESCRIPTIVE STATISTICS FOR ALL SITES, INDIVIDUAL AND MOVING RANGE AVERAGE VIRAL LOADS	131
FIGURE 10.4.2 SITE 1 BASELINE DATASET DESCRIPTIVE STATISTICS AND SKEW HISTOGRAMS	133
TABLE 10.4.3 BY-SITE DESCRIPTIVE STATISTICS FOR ALL SITES, INDIVIDUAL AND MOVING RANGE AVERAGE VIRAL LOADS	134
FIGURE 11.1.1: DAILY CHANGES OF TOTAL SMP SEWERSHED CASE COUNTS OVER TIME	139
FIGURE 11.1.2: ISOLATION OF ACCEPTABLE PERIODS USING STRICT AND MODERATE CASE CHANGE THRESHOLDS ..	140
TABLE 11.1.1: ISOLATED POTENTIAL BASELINE PERIODS WITH RELATED DOMINANT VARIANTS, STRICT CASE CHANGE THRESHOLD	141
TABLE 11.1.2: ISOLATED POTENTIAL BASELINE PERIODS WITH RELATED DOMINANT VARIANTS, MODERATE CASE CHANGE THRESHOLD	141
FIGURE 11.2.1: EXCEL WORKBOOK SCREENSHOT WITH BY-SITE TABS FOR WWS-CASE COUNT PAIRS AND HIGHLIGHTED BASELINE	142
FIGURE 11.2.2: SITE 1 SCREENSHOT OF BASELINE STATISTICS FOR MOVING RANGES OF VL* AND CASE COUNT CHANGE, SD MULTIPLIER OF 3	143
TABLE 11.3.1: SMP BY-SITE MOVING RANGES AND ALERTS, SD 3X MULTIPLIER	144
FIGURE 11.3.1: SITE 1 VL* AND ASSIGNED VL ALERTS, SD 3X MULTIPLIER	145
FIGURE 11.3.2: SITE 1 ASSIGNED VL ALERTS AND CASE ALERTS, SD 3X MULTIPLIER	146
FIGURE 11.3.3: SITE 1 ASSIGNED VL ALERTS AND CASE ALERTS MATCHING AND SE-SP METRICS, SD 3X MULTIPLIER	147
FIGURE 12.1.1 EVALUATION SUMMARY STATISTICS FOR ALL SMP SITES	152
TABLE 12.2.1 SMP ALL-SITE AVERAGE EVALUATION METRICS FOR XSD INCREMENTS	155
FIGURE 13.1 DAILY VDH PUBLIC COVID-19 CASE COUNTS OVER TIME, STACKED AREA GRAPH, WITH START MARKER OF MOCK SCENARIO	157
TABLE 13.1 DELTA VL* AND CASE MOVING AVERAGE BASELINES OF SMP MOCK SCENARIO	159
TABLE 13.2 WEEKLY VL* ALERT FLAGS FOR ALL SMP SITES FOR MOCK SCENARIO, BINARY FLAGS (1=POSITIVE ALERT)	160
TABLE 13.3 WEEKLY CASE ALERT FLAGS FOR ALL SMP SITES FOR MOCK SCENARIO, BINARY FLAGS (1=POSITIVE ALERT)	162
FIGURE 13.2: WEEKLY AVERAGE VL* AND CASE ALERT PERCENT POSITIVITY SMOOTHED LINES IN TIME, MOCK SCENARIO	163
TABLE 13.4 BINARY AGREEMENT BETWEEN VL* AND CASE ALERTS FOR ALL SITES DURING MOCK SCENARIO	164
TABLE 14.1 SUMMARY DIFFERENCES BETWEEN CDC WVAL AND YANCEY VL ALERT	171

List of Abbreviations

Abbreviation	Definition
∂ or d or Δ	the change (i.e., difference between values)
ACE2	angiotensin converting enzyme 2
AMR	Antimicrobial resistance
APHL	Association of Public Health Laboratories
AVG	average
CDC	Centers for Disease Control and Prevention
COVID-19	Coronavirus Disease 2019
CUSUM	cumulative sums
DCLS	Division of Consolidated Laboratory Services
DCIPHER	NWSS Data Collation and Integration for Public Health Event Response
ddPCR	droplet digital PCR
DSA	data sharing agreement
ELC	CDC's Epidemiology and Laboratory Capacity grant
EO-26	Virginia Executive Order 26
EPA	Environmental Protection Agency
EWMA	exponentially weighted moving average
FDA	Food & Drug Administration
HRSD	Hampton Roads Sanitation District
I&I	Inflow and Infiltration
I-PhD	Individualized Interdisciplinary Doctorate of Philosophy
LOD	limit of detection
MAPE	mean absolute percentage error
mgd	millions of gallons per day
MPH	Master of Public Health
NCBI	NLM's National Center for Biotechnology Information
NLM	National Institutes of Health's National Library of Medicine
NWSS	National Wastewater Surveillance System
PCR	polymerase chain reaction
PEG	polyethylene glycol

PhD	Doctorate of Philosophy
qPCR	quantitative PCR
RMSE	root-mean-square error
RT-ddPCR	reverse transcription droplet digital PCR
RT-qPCR	reverse transcription quantitative PCR
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SD	standard deviation
Se-Sp	sensitivity and specificity
SM	Sentinel Monitoring
SMP	Sentinel Monitoring Program
VDH	Virginia Department of Health
Virginia Tech	Virginia Polytechnic Institute and State University
TMV	tobacco mosaic virus
UCL	upper control limit
U.S.	United States of America
VL*	viral load or viral load calculated with LOD concentration
VL Avg	viral load average
VL SD	Viral load standard deviation
WBE	Wastewater-Based Epidemiology
Wks	weeks
WHO	World Health Organization
WWS	wastewater surveillance
WWSP	VDH's Wastewater Surveillance Program
WWTP	wastewater treatment plant
XSD	threshold multiplier times the standard deviation

Chapter 1: Introduction

1.1 Author Context

This dissertation was developed to complete my Individualized Interdisciplinary Doctorate of Philosophy (I-PhD) degree at the Virginia Polytechnic Institute and State University (Virginia Tech). The main distinction of this PhD degree, as compared to others, is that it requires the creation of a unique interdisciplinary program and related dissertation. I developed my program in 2019, centered on the interspace between *public health*, *data analytics*, and *disaster preparedness*. The central idea was to provide research and expertise in the event of a public health challenge/disaster. Ironically, the COVID-19 pandemic started a few months into my program, which shaped the nature of study from theoretical to unfolding. As such, this dissertation reflects the subject matters studied, as well as the expertise earned while I worked for the Virginia Department of Health (VDH) during the pandemic.

My academic background includes undergraduate degrees in Biology and Psychology, as well as a Master of Public Health (MPH) from Virginia Tech. While I attended medical school at the Edward Via Virginia College of Osteopathic Medicine for three years, I withdrew due to illness. It was there, however, where I adopted a more concerned perspective about the ability of major health care institutions to sustain themselves during significant challenges or disasters. These educational pieces lent heavily to my interest and investment towards the public health and disaster preparedness disciplines.

The final discipline component, data analytics, I developed an appreciation for primarily through work experience. My career has generally revolved around being a research analyst --

for academic institutions, military contractors, and public health. Shortly after the pandemic began, I took a position with the VDH as their Data Analyst (and later Data Manager) to establish their first-ever Wastewater Surveillance Program (WWSP)(VDH, 2023). Our small but mighty team is tasked with tracking COVID-19 across Virginia, using wastewater as a new, supplemental surveillance tool. It is understood that case data has its limitations, and thus a potentially more objective and community-centered metric has value.

The WWSP is also interdisciplinary by its nature, as it consists of a team of environmental engineers and public health professionals. It is part of a much larger, national effort funded by the Centers for Disease Control and Prevention's (CDC) National Wastewater Surveillance System (NWSS) (CDC, 2023b). Virginia to date has received \$3.5 million in CDC wastewater surveillance (WWS) funding for the Epidemiology and Laboratory Capacity (ELC) grant for the 2nd, 3rd, and 4th budget periods. We are expected to continue to receive funding for the next several years.

1.2 Wastewater Surveillance Defined

WWS is the systematic detection and analysis of the constituents of wastewater (i.e., sewage) for public health.

Nomenclature Variations

A common term, *Wastewater-Based Epidemiology (WBE)* is generally used interchangeably with *Wastewater Surveillance (WWS)*. However, there are some minor, technical distinctions (WEF, 2023):

- *Epidemiology* implies the focus on health-related problems; the absence of this term may open the WWS field to other, non-health related foci.
- *Surveillance* implies the nature of systematic collection and evaluation; the absence of this term may open WBE to specific, short-term projects and research unrelated to the ongoing nature of public health surveillance.

Other, less common, terms of approximately similar meaning include: *Environmental*

Surveillance (Asghar et al., 2014) *Wastewater Analysis* (Hall et al., 2012), and *Sewer Epidemiology* (Hering, 2009).

Considering the purpose of this dissertation is to evaluate data collected from VDH's WWS Program, *WWS* is the preferred/default term when referring to any program or research utilizing wastewater analysis and monitoring.

WWS Process

WWS begins with the dumping of biological waste (and other materials) down the drains of toilets, sinks, showers, and other sources of sewage. This wastewater is carried through the connected sewer pipes to a wastewater treatment plant (WWTP), where it is then processed for safe return to the environment. Samples collected at the WWTP contain biological waste from all the people connected through those sewer pipes, an area otherwise known as a *sewershed*.

Sampling can take place anywhere along the wastewater pipeline, but for WWS it is commonly taken at either the WWTP influent (i.e., the raw sewage coming into the facility), or at the WWTP effluent (i.e., the processed sewage being released back into the environment). Whether it is taken at the influent or effluent depends on the practicality and purpose of the testing. For community health monitoring, influent testing is more appropriate, as it is subject to less processing and is a more accurate reflection of the community supplying that

waste. For environmental monitoring, effluent testing is more appropriate, as it is more reflective for what is being released back into the environment.

After sampling (and all other metadata collection) is complete, the samples are sent to a laboratory for analysis. Depending on the target of interest, the laboratory may run any series of tests ranging from light microscopy (e.g., for cyst detection) to polymerase chain reaction (PCR) (e.g., for viral particle detection).

WWS: A Community Urinalysis

While there are limitations and nuances to interpreting WWS data – especially regarding variations in sewer infrastructure and testing methods – the data collected represents a snapshot of the community’s targeted markers. Taken systemically, through time, this data can provide insights into changes of those markers. Much like an individual patient may contribute a fecal sample or urinalysis as part of their routine healthcare, in this sense the entire WWS sewershed community contributes to a pooled fecal sample or urinalysis.

Note: while time (or flow) weighted composite sampling can effectively collect small contributions of sewage over a 24-hr period for single samples, uncertainties in patient shedding and large community contributions may not identify sustained signals in the SARS-CoV-2 viral loads for communities with fewer than 10 infected people (Li et al., 2021).

1.3 Dissertation Purpose

WWS is a new tool being used to monitor emerging infections like SARS-CoV-2, though it is not well understood by the public or by public health practitioners. It exists in an interdisciplinary space between environmental engineering and public health, only recently being

thrust into the spotlight with the backing of pandemic funding and need. The purpose of this dissertation is to do a deep dive into what this data is and how it can present and inform public health practitioners. I will develop an ‘Alert System’ from the ground up, building on the basics of WWS is and its main inherent limitations. This framework will be tested with the current VDH WWS SMP data, but could easily be applied to other WWS targets, within and outside of VDH.

The main research question considered is: *“Can SARS-CoV-2 wastewater surveillance once-weekly sampling indicate imminent rises in community COVID-19 disease burden in Virginia?”*

This question will need some literature review and an initial data exploration of the dataset to assess the abilities and limitations of WWS, VDH SMP or otherwise. It will then need an exploration with correlation statistics with other disease burden datasets, such as case and hospitalization data. If a correlation can be established, the next step is to consider a framework in which an alert system could be created and tested.

The natural follow up question: *“Can an alert system framework be created and reasonably applied to VDH WWS data?”*

Using the information gained from the first question, this one will require designing a simple model for an alert system, applying it to the VDH WWS data, evaluating it, and then testing it. Since the VDH WWS SMP program covers years of data, appropriate periods to test the alert system should be available.

Chapter 2: Review of Literature

I performed a review of literature for my I-PhD preliminary exam, focusing on the field of WWS. While the preliminary exam focused on historical and seminal aspects of the field generally, this dissertation chapter expands that work and then focuses on the specified research questions.

2.1 Methods

The original preliminary exam literature review consisted of a combination of literature search sources, including:

- academic databases
- professional organization crawls (e.g., reviewing CDC's NWSS resources for information and relevant seminal work references)
- personal professional knowledge and resources (i.e., the knowledge and access to cite common references in the field)

The reasoning for this combined approach was that much of the WWS field – particularly in reference to the COVID-19 pandemic – was actively developing in real time during a public health crisis. This strain on time and resources is not conducive to building a thorough base of scientific inquiry. Additionally, in the interests of public health, funding and advocacy for WWS seemed to be heavily spear-headed by federal and state governments. As such, the literature and professional expertise resources were scattered between academic research, governmental bodies, and professional institutions.

As someone with the unique background of pursuing a relevant doctorate – while being a core member of the Virginia team building such a state program during a public health crisis – I had the unique opportunity to be privy to both professional and academic insights during a time

of crisis and change, affecting both public health and WWS. It would seem wasteful to limit a review of literature to just academic resources.

Academic Research

Initial academic review of several databases available through the Virginia Tech library was conducted between September and October 2022. The search terms of “Wastewater Surveillance” and “Wastewater Epidemiology” were used, sometimes with the qualifier of specifically excluding articles referencing “irrigation” and/or stratifying records by date. Databases included *Agriculture, Life & Natural Sciences* by ProQuest and *Google Scholar* by Google. Selection of databases and search terms was informed by a graduate library course I took during the I-PhD program, in which the semester’s assignment was to search dozens of relevant Virginia Tech databases for the most applicable literature set for my research topic. Those two databases proved the most relevant to WWS.

Results indicated only 382 observed articles in the ProQuest database during the 1900-1999 years. Many of the results were virus assessments in drinking water, which were not particularly insightful to the field of WWS. About 33 articles were deemed applicable enough to read beyond abstract, and 20 were added to the literature review. These articles, as well as articles contained within the linked citation trees and others referenced via organizational websites (e.g., the CDC), served as a basis for the “Historical WWS” section described later.

Upon opening the history to include recent years (2000-present), the number of articles exploded. Even when limiting the search to review articles only, Google Scholar noted 9,400 articles for the search terms. And while this database appeared to have overlapped with the 1900-1999 ProQuest search, a more strategic tack was required to synthesize recent history in the field.

Articles were collected from the database searches, as well as from organizational websites and other WWS professional circuits.

Professional Organizations

One of the most prominent collaborative organizations used for literature review included the *COVID-19 WBE Collaborative* (COVID19WBEC, 2023). They proclaimed partnerships with many other groups and referenced useful curated WWS review publications (McClary-Gutierrez et al., 2021), (Li et al., 2021), (Medema et al., 2020). They prominently showed a large interactive Google map, displaying 336 empirical studies with 15,318 views (as of March 29, 2023). This publication map is shown below, in Figure 2.1:

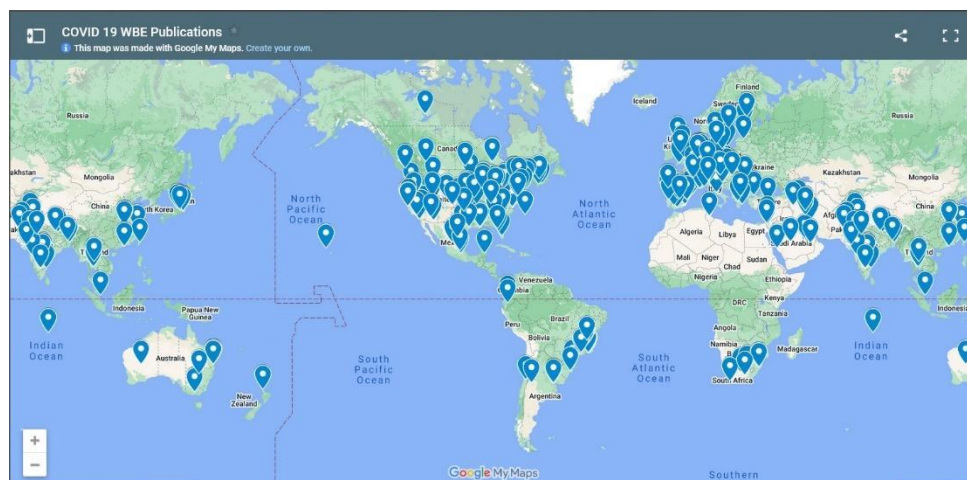


Figure 2.1: Screenshot of *WBE Publication Map* (COVID19WBEC, 2023). Accessed March 29, 2023.

Professional Knowledge

This includes all resources I use in my daily work with VDH WWS, to include items like CDC's NWSS guidance documents and Data Dictionary (CDC, 2023c), Association of Public Health Laboratories (APHL) WWS guidance documents, and SARS-CoV-2 Wastewater Surveillance Testing Guide for Public Health Laboratories (APHL, 2022).

2.2 Historical WWS

John Snow, 1854

The first historical use of early wastewater surveillance could be the start of epidemiology itself -- John Snow's 1854 work on the Broad Street Pump cholera epidemic. It was there that John Snow looked for well water contamination as the source of a London cholera epidemic. The pump handle was removed, and the epidemic ended. Contamination was later confirmed as sewage leaking into the Broad Street Pump.

For context, the prevailing disease theory of the day was miasma. The idea of fecal-oral contamination via drinking water was not yet readily accepted. Though the ability to 'see' and measure cholera in the water was not yet available, the premise of tracing infectious disease through water marks an important thought process towards the linking of sanitation systems and infectious disease control.

Robert Koch, 1876-1884

As germ theory gained acceptance, bacteriology entered its "golden era" in 1876 as Robert Koch published his research connecting a specific bacterium (anthrax bacillus) to its clinical disease (anthrax) (Blevins & Bronze, 2010). The plate techniques he invented (and published in 1881), allowed scientists the ability to 'see' bacteria and study their relationship to disease. He later applied his techniques to a cholera epidemic in India – tracing the infectious agent, comma-shaped bacilli, back to a local water tank. Water filtration was installed as a public health effort, and the epidemic ended (Blevins & Bronze, 2010).

Rise of the Water Engineer

For approximately 30 years since Koch's original 1876 publishing, bacteria were identified and tied to major diseases during this "golden era" of bacteriology (Blevins & Bronze, 2010). While water sanitation wasn't a new concept, the growing connection between fecal-oral diseases and water vectors helped frame sanitation as a public health imperative. Dr. James Wilson, in his research paper isolating enteric bacteria from sewage, stated plainly the importance of the environmental engineer:

"The control of typhoid was brought about by the water engineer and town surveyor, following the advice of the clinician. The bacteriologist had no part, since many decades elapsed before the typhoid bacillus was cultivated by Gaffky in 1884." (Wilson, 1933, p.560)

Wilson credited himself, and several colleagues during the 1928-1933 period, for isolating (often considerably high amounts of) typhoid bacilli (*B. typhosus* and *B. paratyphosus* B) in sewage and rivers. He attributed these contributions to advancements in plating mediums, namely use of a new bismuth sulphite medium (Wilson, 1933).

Virus Detection

While the concept of infectious agents smaller than bacteria and fungi have been known since Dimitri Ivanovsky's 1892 paper on the filtration of tobacco mosaic virus (TMV) disease using Chamberland filters, it wasn't until Ivanovsky's later work in 1903 where he visually noticed intracellular, crystal-like inclusion bodies in infected plants (now considered to be clusters of virus) (Lustig & Levine, 1992). X-ray crystallography studies officially visualized TMV in 1941 (Bernal & Fankuchen, 1941).

Prior to 1950, animal inoculation remained the gold standard for virus identification (Hsiung, 1984). It is through these methods that one of the earliest WWS viral studies took place -- a 1939 study where three Yale Medical Doctors tested for poliovirus in the wastewater of three U.S. cities suffering from polio outbreak (Paul et al., 1940). While poliovirus was recovered from two of the three wastewater systems, their methods required sewage inoculation into live monkeys. The subjects died by either polio-contracted illness, or by other diseases caused by direct sewage injection. Polio had specific clinical features that could be identified post-mortem, so the qualitative presence of disease indicated that transmission from wastewater had occurred. These methods were time-consuming, laborious, and generally not ideal for the monkeys.

By the late 1940's and 1950's, cell culture techniques had been developed by John Enders and others, allowing for *in vitro* virus production (Taylor, 2014). 1954 marked the publication of the technique using plaque assays to isolate and quantify the amount of virus (specifically, poliovirus) in cell cultures (Dulbecco & Vogt, 1954). No longer would animal subjects be considered the gold standard for virus isolation, identification, and quantification. This is not to say that tissue cultures, serous fluid, anti-sera and other tools of microbiology were not still heavily reliant on animal donation and experimentation.

Environmental Concern

The advent of cell culturing techniques brought about a new age of epidemiologic testing. Within a few months of the publication of Dulbecco & Vogt's 1954 paper, researchers had already begun applications of those techniques to wastewater. Citing colleagues' research in stool studies, Bloom et al. (1959) began a two-year sampling survey where they collected sewage from Michigan WWTPs. Their goal was to identify enteroviruses (e.g., polioviruses, Coxsackie

viruses, and echoviruses) in sewage, as well as to assess the removal effectiveness of the wastewater treatment. Their results showed successful identification (with tissue cultures out-performing mouse inoculations) and similar seasonal patterns to stool surveys. Samples noted a progressive – though certainly not complete – viral elimination from collection at the influent to collection at the unchlorinated effluent (33% to 10% percent positivity). Other studies also examined enteroviruses in the wastewater, sometimes with emphasis on polio and public health vaccination campaigns (Riordan, 1962). Methods were further explored (Hill Jr et al., 1971; LUND & HEDSTRÖM, 1966).

The research focus on environmental viral contamination – and the apparent lack of wastewater purification – was not unnoticed by larger scientific organizations. In 1979, a 50-page World Health Organization (WHO) scientific report opened with the following paragraph:

“Increasing attention is being paid to the contamination of water and soil by viruses. This problem, as a factor in the spread of viral diseases, has far-reaching implications which have yet to be fully appreciated by the medical and public health professions.” (WHO, 1979, p. 5)

While the report acknowledged that bacterial contamination had been well monitored and controlled by standard wastewater treatment methods, viruses presented a new, insidious potential for seepage into the environment and public spaces. Over 100 enteric viruses (e.g., poliovirus, coxsackievirus, and hepatitis A virus) were shown to survive treatment and persist for months in the environment. For WWS, the most typical method of viral detection included the following steps: concentration (with either ultrafiltration or adsorption/elution), inoculation to living cells, and identification by antisera (WHO, 1979).

In the United States (U.S.), Safferman (1982) of the Environmental Protection Agency (EPA) echoed this research need for human and ecological impacts of wastewater viruses, stating “There is no question that major public health significance of waterborne viruses stems from the failure to exercise full treatment control over domestic wastewaters” (p. 18). He further cited the passing of recent regulatory pressures and laws, to include the 1974 Safe Drinking Water Act and the 1977 Federal Water Pollution Control Act, which he considered reinforcing to the narrative of increased need for viral WWS.

Several studies followed this call to arms, exploring virus levels at different stages of wastewater treatment. These stages ranged from the raw influent to the processed effluents and sludges. Some studies looked at broad families of viruses and other pathogens (Sellwood et al., 1981), (Guskey, 1983), (Morris & Sharp, 1984), (Nashida et al., 1989), (Aulicino et al., 1998), and (Vantarakis & Papapetropoulou, 1999). Other studies narrowed in on specific diseases/targets, such as *Campylobacter jejuni* (Arimi et al., 1988), Hepatitis A (Divizia et al., 1998), Giardia (Sykora et al., 1991) and *Cryptosporidium parvum* (Mayer & Palmer, 1996), (Robertson et al., 1999).

Public Health Application

Viral/culture identification methods and study design often determined the kinds of targets researchers could investigate when conducting WWS studies. Some studies targeted a multitude of enteric virus families, such as echoviruses, coxsackieviruses, and reoviruses (Riordan, 1962), (Martins et al., 1983), (Guttman-Bass et al., 1987), (Grabow et al., 1999). Other studies narrowed in on a specific pathogen, often trying to link the wastewater detection rates to community case prevalence rates. These included a broad range of targets and associated

diseases, including Cholera (Martins et al., 1991), Giardia (Jakubowski et al., 1991), Campylobacter (Dousse et al., 1993), Polio (Tambini et al., 1993), rotavirus (Mehnert & Stewien, 1993) (Dubois et al., 1997), Hepatitis A (Divizia et al., 1998), and Hepatitis E (Pina et al., 1998).

2.3 Modern (Pre-Pandemic) WWS

Most modern pre-pandemic WWS efforts center around polio eradication, but other efforts include: antimicrobial resistance (AMR) (Hendriksen et al., 2019), drug use (Medema et al., 2020), and others potential directions (Du et al., 2021).

Polio

Since 1988, the Global Polio Eradication Initiative has had the noble goal of eliminating polio from the world. While they currently tout a 99.9% reduction success (WHO, 2023), they have -- and continue to use -- several strategies to achieve this goal. In addition to their gold standard of clinical surveillance and stool sample confirmation, an additional cornerstone complementary strategy includes WWS -- coined as *environmental surveillance* -- used wastewater to detect asymptomatic community wild and vaccine-derived poliovirus transmission. Starting as early as the program itself, ES for polio has been used throughout much of the program and across several countries (e.g., Egypt, India, Pakistan, Afghanistan, Nigeria, and Israel) (Asghar et al., 2014). It has been used in recent years to enhance sensitivity of poliovirus surveillance in priority regions (Lickness, 2020).

2.4 COVID-19 WWS

SARS-CoV-2: Identification and Symptoms

In December 2019, the novel coronavirus Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) was identified in Wuhan, China and quickly spread across the globe. SARS-CoV-2 is the infectious agent responsible for Coronavirus Disease 2019 (COVID-19). When COVID-19 symptoms present in infected people, they typically appear 2-14 days after exposure and can include: fever/chills, cough, shortness of breath, fatigue, muscle/body aches, headache, loss of taste/smell, sore throat, congestion/runny nose, nausea/vomiting, diarrhea and more (CDC, 2022). Not all those who become infected present with symptoms (i.e., they're asymptomatic). Symptoms can be mild to severe, even fatal.

As of July 16th, 2023, the United States reported over 6.2 million COVID-19 hospitalizations and 1.1 million deaths (CDC, 2023b).

SARS-CoV-2: Intestinal Pathogenesis

SARS-CoV-2 is an enveloped, single-stranded RNA virus with surface spike proteins. Infection is initiated by the binding of the SARS-CoV-2 spike proteins to the host angiotensin converting enzyme 2 (ACE2) receptors, alongside the transmembrane serine protease. Zhang et al. (2020) explored which human cells co-expressed ACE2 and transmembrane serine protease, to discover that co-expression was found in: lung alveolar type 2 cells, esophageal upper epithelial and gland cells, and ileal/colon absorptive enterocytes. The absorptive expression suggested a potential digestive system infection pathway, in addition to the already known respiratory pathway. Digestive pathways would also explain some of the enteric symptoms associated with COVID-19 (e.g., diarrhea).

Lamers et al. (2020) showed evidence of this digestive system infection pathway by using in vitro experimentation to successfully infect human small intestinal organoids with SARS-CoV-2. The organoids demonstrated infection processes and replicated viral particles, providing evidence that infection and replication could occur in the gut. Xu et al. (2020) furthered evidence for this pathway by finding that not only did eight out of the ten pediatric COVID-19 cases in their study flag positive on rectal swabs for SARS-CoV-2, but that those swabs continued to persistently show positive tests for weeks after their nasopharyngeal testing had become negative. All ten of the pediatric cases used were identified by screening tests and were found to be mild or asymptomatic cases. This shedding persistence indicates that SARS-CoV-2 viral shedding can occur even in individuals that would normally not present with severe enough symptoms to be typically identified as cases. Jiang et al. (2020) reported a pediatric asymptomatic case that never flagged positive on nasal swab, but had a positive anal swab for 42 days.

Early Pandemic

As the virus started to become endemic in U.S. communities, there was some concern that public health case data was not capturing true community infections – that there was an undercurrent of asymptomatic and undiagnosed flying under the radar. Gandhi et al. (2020) argued in the *New England Journal of Medicine* that a major driving factor in the high transmissibility of SARS-CoV-2 is community spread of the asymptomatic and pre-symptomatic infected. They pleaded for increased testing of asymptomatic, especially in high density living areas such as skilled nursing facilities and prisons. They argued that symptom-based testing is not nearly effective enough.

With suggestion to WWS, Thompson et al. (2020) in *Water Research* argued that WWS can provide a “valuable complement” (p. 1) to clinical testing, as it is an inexpensive, population-based tool that is independent of – and free of the biases associated with – case reporting. They advocated for building a framework of collaborative government agencies and wastewater utility partners to acquire “actionable community-level information needed to navigate pandemics like COVID-19” (p. 5). They did, however, warn of potential misuse and public backlash to such a program. They emphasized the need for ethical considerations such as the mitigation of privacy oversteps and exacerbation of social inequalities by over-surveillance (and the associated stigma/regulation sequelae).

To prove the concept of WWS, a group in the Netherlands provided correlation between WW viral copies and reported cases, at some points detecting community infection before clinical reporting (Medema, Heijnen, et al., 2020). As of October 23rd, 2023, the article is listed under ACS Publications as having 34,401 article views and 882 citations. Other nations, such as Australia (Ahmed et al., 2020), followed suit.

With the proof of concept attained, WWS continued to garner support for COVID-19 surveillance. Medema argued for the implementation of WWS to capture larger, sometimes asymptomatic, community infection, “So, only a fraction of infected individuals are tested. But everybody uses the toilet” (Medema, Been, et al., 2020, p. 51). Mallapaty (2020) argued in *Nature* that a main benefit of WWS would be in early detection of virus re-emergence. Granted, at the time the author thought that social distancing and other infection-control measures would suppress the pandemic. In retrospect, that might have been a bit hopeful. There was a realization that WWS could also be used as leading indicator of cases (Ahmed et al., 2021) and thus may offer a pathway for preventive or mitigative public health action.

COVID-19 WBE Collaborative

With consideration to the inherent variability present in wastewater systems – in particular the nature of urban centralized and rural decentralized systems – Bivins (2020) proposed a global effort for WWS researchers to share data and validate methodologies, in order to achieve “systematic harmonization” (p. 7755). They created a research hub, the *COVID-19 WBE Collaborative* (COVID19WBEC, 2023), which encourages collaboration with other WWS groups and the sharing of publications and data.

Bivins also recognized the need, and importance of, interdisciplinary collaboration:

"The ongoing COVID-19 pandemic requires engineers and scientists to collaborate with population-based scientists, including epidemiologists, mathematical modelers and public health agencies. A multi-disciplinary approach on a global scale is required for timely and high impact results to help society." (Bivins et al., 2020, p. 7755)

The research hub included several Virginia-specific collaborations: Virginia Tech, Hampton Roads Sanitation District (HRSD), and Harrisonburg-Rockingham Regional Sanitation Authority (HRRSA). These also included several Virginia publications: Gonzalez (Gonzalez et al., 2020), Curtis (Curtis et al., 2021), Ciesielski (Ciesielski et al., 2021).

Virginia: Hampton Roads Sanitation District

The Hampton Roads Sanitation District (HRSD) is a political subdivision of Virginia, providing wastewater treatment services to approximately 1.9 million people in the coastal southeast (<https://www.hrsd.com/>). They were also one of the first to provide regional applications of COVID-19 WWS, covering their surveillance area as early as March 2020 (Gonzalez et al., 2020). Their 21-week study began the same week as the first confirmed case of

COVID-19 in southeastern Virginia and proceeded as confirmed cases in their service region expanded to nearly 15,000. The spatial analytics used for their study was published on their public-facing dashboard (HRSD, 2023).

HRSD continues to perform bi-weekly testing of COVID-19, MPOX, and influenza in their 8 major WWTPs. They post the results on their dashboard regularly, as seen below in Figure 2.4.1.

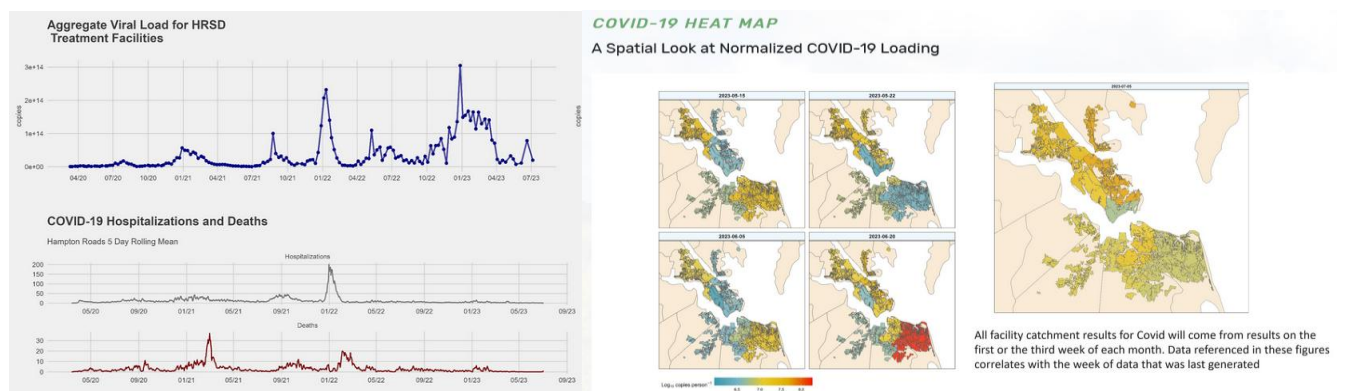


Figure 2.4.1: Screenshots from HRSD (2023), Aggregate Viral Loads and Hospitalizations. Accessed July 13, 2023.

CDC NWSS

In September 2020, the CDC launched the National Wastewater Surveillance System (NWSS). The purpose was to “build the nation’s capacity to track the presence of SARS-CoV-2” in wastewater samples nationally (CDC, 2023d). NWSS provides the WWS framework, structure, and general support for public health departments and other partners. Affiliated partners upload WWS data into the CDC NWSS portal – the Data Collection and Integration for Public Health Event Response (DCIPHER). Public health departments typically work with their state utility partners to collect samples, which are then sent to analytical laboratories, the data of which is then sent to the public health partners for interpretation and upload into DCIPHER.

DCIPHER data can currently be viewed internally by contributing partners, as well as publicly on their CDC COVID Data Tracker website (CDC, 2023b). A screenshot of their public-facing dashboard can be seen below in Figure 2.4.2:

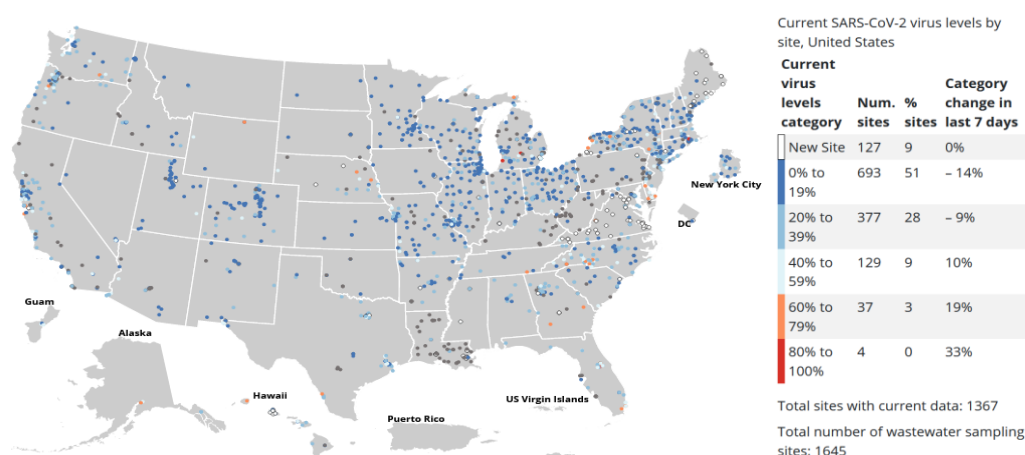


Figure 2.4.2: Screenshot from CDC COVID Data Tracker (CDC, 2023b). Accessed July 19, 2023.

Virginia Department of Health: WWS Program

The Virginia Department of Health’s Wastewater Surveillance Program (WWSP) (VDH, 2023d) represented Virginia as one of the first eight states to join (early adopters) CDC’s NWSS cooperative agreement. Initial funding was provided by CDC’s Epidemiology and Laboratory Capacity (ELC) grant, which was enough to hire three employees (including myself) to officially start the WWSP. Together we built the program and created a network of utility, lab, and research partners to build WWS in Virginia. Funding was provided by ELC yearly grants several times since then, and renewal is expected for several more years.

WWPS supports and participates in multiple communities of practice, to include the NWSS Utilities Community of Practice (NWBE, 2023), CDC’s Health Departments Community of Practice, Laboratories Community of Practice (APHL, 2023), and VDH’s Community of

Practice. We also attend the CDC NWSS Data Analytics Group and participate in Cohort #1 with other early adopter states.

VDH WWS produces several major sources of data, including:

- Sentinel Monitoring Program – 25+ utilities testing viral loads weekly (or twice weekly), with spatially-joined case count data
- Localized Projects - institution or subsewershed projects, usually spearheaded by interested partners (e.g., schools, local health departments, hospitals, prisons, etc.)
- Variant Sequencing - samples from Sentinel Monitoring are taken for genomic sequencing, and variant abundances identified

Variant Sequencing

While not the most significant WWPS product, or focus of this dissertation generally, genomic sequencing to determine the relative abundances of variants can be performed not just through clinical testing, but through wastewater as well. Over four million unique U.S. samples have been sent to the National Library of Medicine’s National Center for Biotechnology Information (NCBI) (NCBI, 2023). Their online dashboard shows which lineages are currently prominent. The CDC also has a version of national genomic surveillance, with their dashboard featuring Nowcast for future estimates of which variants are predicted to dominate (CDC, 2023a). A screenshot of their public-facing dashboard can be seen below in Figure 2.4.3:

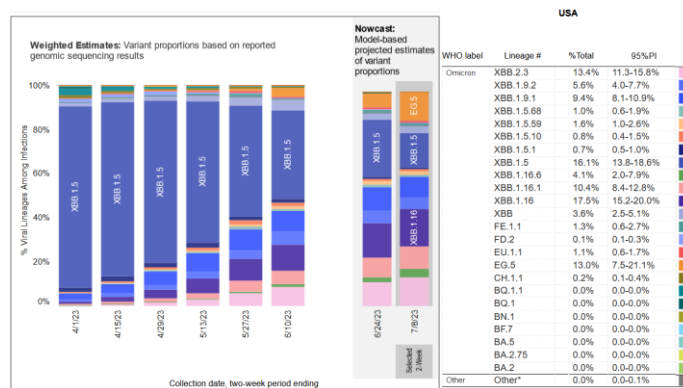


Figure 2.4.3: Screenshot from CDC Clinical Variant Proportions (CDC, 2023a). Accessed July 19, 2023.

The Food & Drug Administration (FDA) has specifically pulled WWS variant sequencing data from NCBI, and developed a software to share relative abundances (FDA, 2023). This would make it the leading WWS variant abundance site now, though CDC NWSS may adopt a similar feature on their dashboard in the future. However, with the closing of the COVID-19 public health emergency on May 11, 2023, the FDA stopped updating their WWS dashboard as of June 30, 2023. A screenshot of their public-facing dashboard can be seen below in Figure 2.4.4:

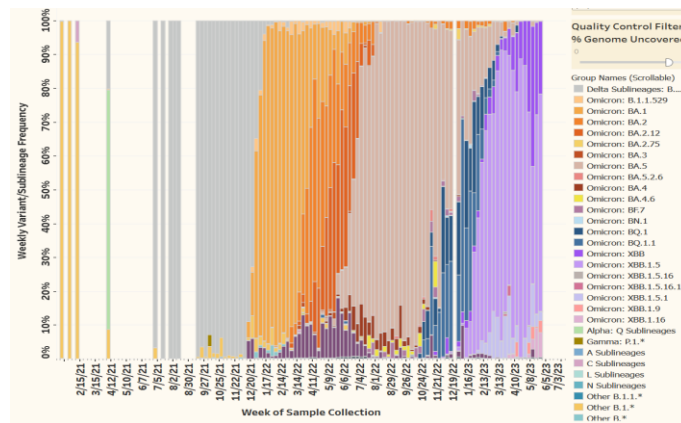


Figure 2.4.4: Screenshot from FDA WWS Variant Abundances (FDA, 2023). Accessed July 19, 2023.

Chapter 3: Variability and Limitations

There are many steps to conducting WWS, all of which have varying methods and associated limitations. The sections below will address major steps and associated method variations, as well as known uncertainty measures behind them. The CDC Data Dictionary (CDC, 2023c) will be referenced often, as it contains the standard for what kinds of data/options can be submitted to NWSS (and thus a good measure for what is currently standard in the WWS field). CDC NWSS focuses their WWS efforts mainly on SARS-CoV-2, but have extended to a few other targets (e.g., MPOX).

As for uncertainty estimation, Li et al. (2021) performed a systematic literature review to quantify WWS uncertainties into 5 categories: virus shedding, in-sewer transportation, sampling & storage, analysis, and prevalence back estimation. They explored major field method options and concluded that, by calculating COVID-19 prevalence using composite flow sampling (with populations over 10 infected people), uncertainty calculations were lowest at 20-40% relative standard deviation (RSD). This was deemed a reasonable amount of uncertainty for a complementary surveillance system (Li et al., 2021).

Elements of WWS uncertainty are explored below.

3.1 Target

A target, as well as the associated PCR probe set, needs to be identified before WWS can begin. This is generally a piece of RNA or DNA from a viral target, but there are non-infectious markers that can be detected via wastewater as well. When considering a viral target, laboratories

must first be able to identify and create a probe target set that is able to accurately identify the viral target of interest.

Once the COVID-19 pandemic began, the CDC used the SARS-CoV-2 genome to develop/validate a set of 3 real-time reverse transcription PCR (rRT-PCR) panels for Emergency Use Authorization. These panels targeted the nucleocapsid protein (N) gene of the virus, using several primer/probe target regions – N1, N2, and N3 (Lu et al., 2020). Gonzalez et al. (2020) found N2 target to be the most sensitive for their southeast Virginia region (Gonzalez et al., 2020).

Several targets are currently recognized by CDC NWSS Data Dictionary (CDC, 2023c):

Target variables (PCR targets, limits of detection, and references): PCR_target, PCR_gene_target, PCR_gene_target_ref, PCR_type, LOD_ref

PCR Target fields: SARS-CoV-2, Delta, Omicron, hMPXV, hMPXV Clade 1, hMPXV Clade II

PCR Gene Target fields:

SARS-CoV-2: n1, n2, n3, e_sarbeco, n_sarbeco, rdp_sarsr, niid_2019-ncov_n, rdp gene / ncov_ip2, rdp gene / ncov_ip4, taqpath n, taqpath s, orf1b, orf1ab, n1 and n2 combined, n, s, orf1a, ddcov_n, ddcov_e, ip2 and ip4 combined, CDC N1 (GT-Digital), CDC N2 (GT-Digital), N2 (PREvalence), E (PREvalence)

Delta: a1306s, p2046l, p2287s, t3646a, v2930l, g662s, p100l, a1918v, t19r, e156g, del 157/158, l452r, p681r, d950n, i82t, d63g, r203m, g215c, d377y, wt214, del156-157

Omicron: k856r, s2083i, del2084/2084, a2710t, p3395h, del3674/3676, i3758v, i1566v, a67v, del69/70, del143/145, n211i, del212/212, g339d, s371l, s373p, s375f, s477n, e484a, q493r, g496s, q498r, y505h, t547k, n679k, d796y, n856k, q954h, n969k, l981f, t9i, d3g, q19e, a63t, ins214epe, del142-144, del31-33, l24s, lppa24s

MPOX (hMPXV): E9L-NVAR, G2R_G, G2R_WA, C3L, E9L-OPX3, B6R, gtmol_hMPXV

PCR Type fields: qpcr, ddpcr, qiagen dpcr, fluidigm dpcr, life technologies dpcr, raindance dpcr, dpcr

After a target and probe set are created, there is a further question of how much shedding of that virus piece is dropped into the wastewater system by an infected individual. This is called the shedding rate and can be highly individualized to the infected person. It can also change by variant strain, as new variants may affect human physiology slightly differently.

Li et al. (2021) confirmed this individual variability and further estimated that mean stool shedding probability was at 54.5% with a standard deviation of 9.3%. The stool shedding magnitude mean was at 4.523 log₁₀ copies/g, with a standard deviation of 0.133 log₁₀ copies/g. The shedding period could last for weeks after an infected individual no longer tests COVID-19 positive with nasal swabs (Li et al., 2021).

3.2 Viral Decay in Sewage

Once the viral target is shed into the wastewater system, it is subject to denaturing forces, such as chemical and microbial interactions. The distance and travel time to the sample collection point have an inverse relationship to the amount of recoverable target.

Several denaturing water quality variables are currently recognized by CDC NWSS Data Dictionary (CDC, 2023c):

Water Quality Variables: sewage_travel_time, ph, conductivity, tss (total suspended solids), collection_water_temp

Li et al. (2021) found that in-sewer transportation reduced the infectivity of SARS-CoV-2 by 90% within minutes to days, while the RNA itself took days to weeks to degrade. They found that RNA decay is generally thought to follow the first-order decay equation of:

$$[\text{Conc}]_{\text{timeA}} = [\text{Conc}]_{\text{time0}} \times e^{-kt}$$

K represents a decay rate constant, but notice how the temperature (t) multiplies against the constant. Li et al. calculated that, at 4°C, k is 0.084; at 37°C, k is 0.286. At higher temperatures, decay is expected to play more of a factor (Li et al., 2021).

Further evidence is needed to elucidate the effects of other water factors, such as levels of dissolved oxygen, organic matter, solvents, etc.. The traveling time, which is inversely related to WWTP catchment size, ranges from minutes to 6-10 hours. Li et al. (2021) concluded that the uncertainty of decay was estimated as 5%-11% Relative Standard Deviation (RSD).

3.3 Sampling

Location

Sampling is typically done at the WWTP influent on raw wastewater. Other options include taking samples at other locations within the WWTP process (i.e., changing the sample matrix type), or at locations ‘upstream’ from the WWTP. Upstream sampling narrows the sewershed population only to the sewer line branches that feed into that point. Upstream sampling can narrow down the service area to a neighborhood or even building, if desired.

Several location variables are currently recognized by CDC NWSS Data Dictionary (CDC, 2023c):

Sample location variables: county_names, zipcode, population_served, sample_location, sample_location_specify, institution_type, epaid, wwtp_name, wwtp_jurisdiction, capacity_mgd, sample_collect_date, sample_collect_time, time_zone, flow_rate, sample_matrix

Sample Locations fields: WWTP, upstream

Sample Matrix fields: raw wastewater, post grit removal, primary sludge, primary effluent, secondary sludge, secondary effluent, septage, holding tank

Institution Type fields: not institution specific, correctional, long term care - nursing home, long term care - assisted living, other long term care, short stay acute care hospital, long term acute care hospital, child day care, k12, higher ed dorm, higher ed other, social services shelter, other residential building, ship, airplane, other worksite

Li et al. (2021) found that samples collected from sludge, sediment, and wastewater solids could result in higher viral concentrations (as compared to raw wastewater influent), due to the higher stool concentration.

Upstream Influences

Many factors upstream can influence viral concentration downstream. For example, some WWTP sewer systems are built as *combined systems*, which allow stormwater to mix with wastewater. This dilutes target concentrations, as heavy rainfall contributes directly to the wastewater flow. This is also referred to as Inflow and Infiltration (I&I).

Several sewage variables are currently recognized by CDC NWSS Data Dictionary (CDC, 2023c):

Sewage structural variables: industrial_input, stormwater_input, influent_equilibrated

Li et al. (2021) recommends consideration of unique hotspot facility areas, such as hospitals, which may treat their wastewater on site. Local wastewater treatment could result in the removal of their respective target contribution from downstream samples.

Collection

Wastewater use can fluctuate throughout the day, as people tend to have favored times when they shower and use toilets. As such, a grabbed sample at 9am may have very different concentration levels than a 3am grab of the same day. There may also be seasonal effects which contribute to wastewater over longer periods, such as the ebbs and flows of transient populations into regions (e.g., beach communities or college towns). If the sewer system is a combined system, rainy seasons may also contribute dilution to the wastewater.

It is generally estimated that using 24-hr flow-composite sampling – where small samples are automatically taken proportionally to flow throughout the day – is a more accurate way to measure the day’s viral load than by extrapolations from a single grab sample. Li et al. (2021) recommends using a composite sample (<10% uncertainty) instead of a grab sample (30% uncertainty), for reducing the daily fluctuation effects. However, Curtis et al. (2021) found minimal variation when comparing grab sampling to 24hr-flow-weighted composite sampling (Curtis et al., 2021).

Several collection variables are currently recognized by CDC NWSS Data Dictionary (CDC, 2023c):

Collection methods: sample_type, composite_freq

Sample Type fields: grab, (1-48)-hr flow-weighted composite, (1-48)-hr time-weighted composite, (1-48)-hr manual composite, (1-48)-hr passive sample

Storage

Once samples are collected, they need to be transported to an appropriate laboratory for analysis. The amount of time, as well as preservation techniques like chilling, may differ with different collection methods.

Several storage variables are currently recognized by CDC NWSS Data Dictionary (CDC, 2023c):

Storage: collection_storage_time, collection_storage_tem, pretreatment, pretreatment_specify

Li et al. (2021) found that having a storage time under 72 hours, as well as keeping the sample cool at 4⁰C (without freezing and unfreezing) and without direct exposure to UV light, is recommended to avoid RNA decay.

3.4 Laboratory Analytical Methods

Laboratories have different methodologies for every step of their process, from sample concentration to viral concentration calculations. As the APHL states in their public health laboratory guidance document, “There is no single method for SARS-CoV-2 testing for wastewater surveillance” (APHL, 2022, p. 4).

Li et al. (2021) found that different concentration and extraction methods led to different recovery efficiencies. The most efficient methods were polyethylene glycol (PEG) precipitation and aluminum hydroxide adsorption-precipitation (though still subject to inter-lab differences). Additionally, the wastewater matrix itself may play a role in the recovery efficiency. For detection, the techniques and primer-probe sets have differences between them. Droplet digital PCR (ddPCR) and reverse transcription droplet digital PCR (RT-ddPCR) were generally found to have higher precision and sensitivity to SARS-CoV-2, as compared to quantitative PCR (qPCR) and reverse transcription qPCR (RT-qPCR). Using spikes of surrogate virus, or other indicator viruses like pepper mild mottle virus (PMMoV), can be used to help establish recoveries. Uncertainty for recovery procedures was identified as 17-36%, presuming best practices (Li et al., 2021).

Processing

Several processing variables are currently recognized by CDC NWSS Data Dictionary (CDC, 2023c):

Processing variables: solids_separation, concentration_method, extraction_method, pre_conc_storage_time, pre_conc_storage_temp, pre_ext_storage_time, pre_ext_storage_temp, tot_conc_vol, ext_blank, rec_eff_target_name, rec_eff_spike_matrix, rec_eff_spike_conc, pasteurized, equiv_sewage_amt
Solids Separation fields: filtration, centrifugation, gravity settling, none

Concentration Method fields: membrane filtration with addition of mgcl2, membrane filtration with sample acidification, membrane filtration with acidification and mgcl2, membrane filtration with no amendment, membrane filtration with addition of mgcl2, membrane recombined with separated solids, membrane filtration with sample acidification, membrane recombined with separated solids, membrane filtration with acidification and mgcl2, membrane recombined with separated solids, membrane filtration with no amendment, membrane recombined with separated solids, peg precipitation, aloh3 precipitation, ultracentrifugation, skimmed milk flocculation, beef extract flocculation, promega wastewater large volume tna capture kit, centricon ultrafiltration, amicon ultrafiltration, hollow fiber dead end ultrafiltration, innovaprep ultrafiltration, no liquid concentration, liquid recombined with separated solids ceres nanotrap, zymo environ water rna kit/ zymo environ water rna kit (cat. r2042), membrane filtration with addition of mgcl3, none

Extraction Method fields: qiagen allprep powerviral dna/rna kit, qiagen allprep powerfecal dna/rna kit, qiagen allprep dna/rna kit, qiagen rneasy powermicrobiome kit, qiagen powerwater kit, qiagen rneasy kit, qiagen ez1 virus mini kit v2.0, promega ht tna kit, promega automated tna kit, promega manual tna kit, promega wastewater large volume tna capture kit, nuclisens automated magnetic bead extraction kit, nuclisens manual magnetic bead extraction kit, phenol chloroform, chemagic viral dna/rna 300 kit, trizol, zymo mag beads w/ zymo clean and concentrator, 4s method (<https://www.protocols.io/view/v-4-direct-wastewater-rna-capture-and-purification-bpdfmi3n>), qiagen qiaamp buffers with epoch columns, zymo quick-rna fungal/bacterial miniprep #r2014, thermo magmax microbiome ultra nucleic acid isolation kit, zymo environ water rna kit/ zymo environ water rna kit (cat. r2042), luminultra wastewater extraction kit, qiaamp viral rna kit, trizol and RNA purification kit, trizol, garnet bead beating, alcohol precipitation, zymo quick-rna viral 96 kit #r1041, zymo quick-rna viral kit #r1035, qiagen qiaamp dsp viral rna mini kit

Quantification

Using the target probe identified (see Target section above), the amount of virus needs to be quantified, and a concentration determined.

Several quantification variables are currently recognized by CDC NWSS Data Dictionary (CDC, 2023c):

Detecting Human Fecal Content variables: hum_frac_target_mic, hum_frac_target_mic_ref, hum_frac_target_chem, hum_frac_target_chem_ref, other_norm_name, other_norm_ref

Microbial Target for Estimating Human Fecal Content fields: pepper mild mottle virus, crassphage, hf183, f+ rna coliphage, f+ dna coliphage, PMMoV (GT-Digital)

Chemical Target for Estimating Human Fecal Content fields: caffeine, creatinine, sucralose, ibuprofen

Other Targets for Estimating Human Fecal Content fields: pepper mild mottle virus, crassphage, hf183, f+ rna coliphage, f+ dna coliphage, caffeine, creatinine, sucralose, ibuprofen, rnase p

Quantification Standards & Inhibition Methods variables: quant_stan_type, stan_ref, inhibition_method, num_no_target_control

Quantification Results

After human fecal content is identified, quantification can be performed.

Several quantification analysis results variables are currently recognized by CDC NWSS Data

Dictionary (CDC, 2023c):

Target concentration/statistics variables: test_results_date, PCR_target_units, PCR_target_avg_conc, PCR_target_std_error, PCR_target_cl_95_lo, PCR_target_cl_95_up, PCR_target_below_lod, lod_sewage, quality_flag
PCR Target Units fields: copies/L wastewater, log10 copies/L wastewater, copies/g wet sludge, log10 copies/g wet sludge, copies/g dry sludge, log10 copies/g dry sludge, micrograms/L wastewater, log10 micrograms/L wastewater, micrograms/g wet sludge, log10 micrograms/g wet sludge, micrograms/g dry sludge, log10 micrograms/g dry sludge
Amplification, Recoveries, Inhibition variables: ntc_amplify, rec_eff_percent, inhibition_detect, inhibition_adjust
Human Fecal Content Concentration variables: hum_frac_mic_conc, hum_frac_mic_unit, hum_frac_chem_conc, hum_frac_chem_unit, other_norm_conc, other_norm_unit
Microbial/Chemical Units fields: copies/L wastewater, log10 copies/L wastewater, copies/g wet sludge, log10 copies/g wet sludge, copies/g dry sludge, log10 copies/g dry sludge, micrograms/L wastewater, log10 micrograms/L wastewater, micrograms/g wet sludge, log10 micrograms/g wet sludge, micrograms/g dry sludge, log10 micrograms/g dry sludge

3.5 Interpretation & Data Analysis

Common analyses involve using the analyzed viral concentration by itself or using it to calculate a viral load (+/- normalization factors). The viral load – or amount of virus passing through the collection point throughout the day – can be calculated by taking the analyzed viral concentration, multiplied by the day’s flow, and then multiplied by a constant to adjust for units. This viral load can be *population normalized* by dividing the viral load by the population of the sewershed.

Ultimately, one of the major goals of WWS is to get a clear picture of how many people are actively shedding (i.e., infected) in a sewershed catchment area. Ideally, given a certain viral concentration number and some other underlying known variables, one would like to calculate

this metric. This is known as the *prevalence back estimate*. It would require a consistent and known population number, or failing that, a human fecal normalization (CDC, 2023f).

Li et al. (2021) found that, to calculate the prevalence back estimate, the sum of uncertainties included all those found in the viral load calculation, as well as several other underlying variables. Overall, they found that the viral load calculation uncertainties included: sampling uncertainty (5% for flow-proportional composite; 30% for grab), analytical uncertainty (17%-36%), and flow uncertainty (20% estimation). Other needed variables included the catchment population size (16%, due to population fluctuations) and water usage (10%, observational).

3.6 Major Limitations

While Li et al. (2021) summed quantitatively the uncertainties associated with different WWS methods – and defended its use as a supplemental surveillance strategy generally – there exist several glaring limitations of WWS. Several of these are discussed more thoroughly by the expert panel in McClary-Gutierrez et al. (2021).

Limitations:

1. **Predicting Cases:** uncertainties currently restrict the ability to use WWS to reliably predict the number of COVID-19 cases in a sewershed.
2. **Off-grid:** those persons, institutions, or communities not connected to (usually municipal WWTP) sewersheds will not be captured by WWS.
3. **Limit of Detection:** low levels of virus in the wastewater may not be accurately picked up by WWS. A small number of infected individuals may also be fairly variable in their shedding presentation, adding to unreliability at low numbers (Li et al., 2021).
4. **Lab & Collection Methods Differ Greatly:** inter-lab and collection methodologies can present dramatically differently, limiting the use of comparing one lab's results with another's.

5. **Site Variability:** not all WWTPs and sewersheds are created equally, and can vary greatly in population factors (e.g., transients, off-grid customers), sewer infrastructure (e.g., in-sewer travel time, stormwater combined systems), and variant-population dynamics (i.e., which variants are shedding in which populations). Site trend lines are generally more acceptable than the comparison of raw site concentrations or viral loads between sites.
6. **Variant Shifts:** as the pandemic continues and new variants arise, the accurate identification and detection of variants for PCR analysis is imperative. Using older PCR probes that do not capture current circulating viruses defeat the purpose of community testing.
7. **Data Underpinning:** WWS concentrations are rarely presented independently and usually require supporting data sources for proper visualization/analysis (e.g., flow data, sewershed maps, population and numbers). If flow data or population estimates are not reported accurately, viral load and normalization analytics will be less accurate. If COVID-19 case numbers (or change in) are not reliable – say, due to the rise of take-home test kits – case correlation metrics will become less reliable.

Additionally, there are currently several major human limitations when it comes to interpreting and using WWS:

- **Privacy & Ethics:** WWS exists in an interdisciplinary space between environmental surveillance and public health -- each of which treat their data with different levels of protections (McClary-Gutierrez et al., 2021). Beyond data ethics, there is a fundamental privacy concern when it comes to unconsented population surveillance. Unequal application of surveillance against already marginalized populations is a form of oppression that can create further abjection of those groups (Monahan, 2017). Upstream/granular WWS testing should be undertaken with caution, as to not exacerbate existing inequalities (Thompson et al., 2020).
- **Novelty:** WWS is a relatively new tool for public health. Introducing a new tool during a pandemic – while public health departments are already at capacity – creates some friction for epidemiologists, who may not have the time to understand and vet it appropriately. At present, there is a communication and comfort gap between producers of WWS data and public health practitioners (McClary-Gutierrez et al., 2021). The public may also struggle with using and understanding WWS. Establishing public legitimacy is paramount, else the public may get creative in finding ways to “opt-out” of WWS (Thompson et al., 2020, p. 4).

3.7 Major Advantages

WWS presents several uses and advantages over traditional metrics, especially during a pandemic when public health departments are at capacity.

Advantages:

- **Community-Centered:** unlike traditional case monitoring, which relies on infected individuals accessing healthcare (and subsequent reliable/timely reporting to public health), WWS measures communities. A WWS sample can provide a snapshot viral concentration of the sewershed's service area, therefore providing an objective indicator of community infection.
- **Independent Metric:** WWS does not rely on public health datasets, such as case reporting or hospitalization data. While it can be used in tandem with such datasets – and is advocated as a supplemental tool for them (CDC, 2023d) – it does not rely on existing metrics. This is particularly important as case data becomes less reliable with the rise of self take-home test kits.
- **Very Inexpensive:** WWS can test a community of hundreds of thousands with a single sample.
- **Detects Deceptive Viral Burdens:** viral shedding occurs in those infected with SARS-CoV-2, even if they have mild or no symptoms (Xu et al., 2020). These are cases that are unlikely to get flagged by healthcare systems, especially with the rise of unreported home test kits. As variants evolve and populations become increasingly immunized, these asymptomatic or mild cases may become even more prevalent. WWS detects community viral shedding, irrespective of symptom severity or official case counts.
- **Early Warning:** viral shedding can precede case presentation. Additionally, more granular use of WWS, particularly in areas that have a lower disease burden, may help public health detect and control outbreaks before they spread (Thompson et al., 2020).
- **Variant Abundance & Sequencing:** WWS can be used to detect and sequence variants, which can be useful for knowing the abundance of variants in a community, as well as the ability to detect cryptic variants or emerging strains.

Chapter 4: Alert System Fundamentals

This chapter outlines the basis and reasoning behind alert systems generally, as well as how they pertain to WWS. The need for an alert system will be discussed. Chapter 10 will discuss how these principles were used to develop this dissertation's Alert System, as well as how the developed system novelly contributes to VDH public health practice.

4.1 Why Develop an Alert System?

Much of my intent here was practical – I wanted my wastewater surveillance data to be useful. Our program has used millions of dollars to create a surveillance system collecting data on a weekly or biweekly cadence, which is entirely too much money and time to not be able to tell public health practitioners if it's warning us of something or not. See below for definitions and principles for what encompasses an alert system.

However, because state and nationally based WWS is relatively new, there isn't consensus on what system should be in place to display such collected data, let alone interpret it for immediate public health action. A summary of current federal and state directions is listed below.

Additionally, WWS is incredibly nuanced (see previous Chapter) and depends on contextual handling and interpretation of its data. The most pressing issue regarding this is that every sewershed is unique, with different geography, populations, and sewer systems. Some sewer systems have substantial industrial input, while others are open to rainwater and dilute their human fecal content concentration with every passing storm. Collection, analysis, and interpretation methods can all also vary with every sample, so comparability with one lab's results may be wildly different from another lab's results for exact same sample.

It is for this reason that the nature of the WWS surveillance system matters for any alert system development. What may work for one state's network may be inappropriate for another. This variation makes it extra difficult to develop a standard, universal alert system for wastewater – at least one that doesn't consider inherent variabilities between collection sites and laboratory methods.

4.2 What makes an Alert System?

“The early warning of infectious diseases is to analyze surveillance data with specialized technologies for early detection and warning of notable aberrations” (Yang, 2017, p. 3)

According to Weizhong Yang's 2017 textbook, *Early Warning for Infectious Disease Outbreak Theory and Practice*, an early warning system for infectious disease must have four qualities:

- 1) Information-based surveillance
- 2) Timely
- 3) Actionable & evidence-based
- 4) Exists in a space of blurry initial information (Yang, 2017, pp. 4-5)

In addition to their use as an aberration detector for infectious disease epidemics, Yang further explains that the main purpose is to contrast the actual value with the expected value. However, he notes that it is often unclear on how to determine case incidence, and that “there are no strict guidelines on setting a threshold” (Yang, 2017, p. 39). However, he does note that these systems usually compare observed cases with expected cases, with historical surveillance data encompassing a 3–5-year period to establish a “long history baseline comparison”, and less than that to be a “short history baseline comparison” (Yang, 2017, pp. 39-40).

Yang also emphasizes the need for simplicity in early warning:

“In theory, most prediction models can be used for early warning. The expected and threshold values can be obtained from prediction models. By comparing the difference between these two values, one can determine the need for early warning. However, in practice, early warning systems often use simple and clear models that do not have complex parameters. The primary considerations for early warning models are their stability; the availability, economic efficiency, and timeliness of data; and the simplicity of constructing a system.” (Yang, 2017, p. 40)

For the purposes of this dissertation, “Early Warning System” and “Alert System” will be considered interchangeable terms, as well as “system”, “model”, and “algorithm”. I have chosen to approach this dissertation with the assertion that an early warning system which generates an actual alert condition – to be able to be distributed to the necessary stakeholders, such as local epidemiologists – is the most direct way to handle a new public health surveillance early warning system for wastewater surveillance.

Yang’s Conceptual Framework

Yang’s conceptual framework, summarized below in Figure 4.2, for an Early Warning System consisted of 1) setting targets, 2) collecting data, 3) analyzing data, 4) disseminating information, and 5) responding with action (Yang, 2017, pp. 6-7).

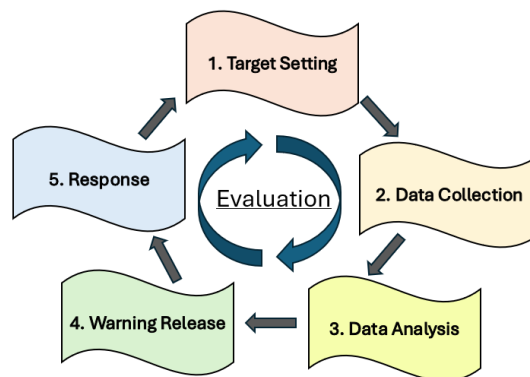


Figure 4.2: Summarized version of Yang’s conceptual framework for an Early Warning System

Prediction Vs. Alerts

There are a lot of similarities between “prediction” and “early warning.” Prediction is “characterized by speculation and the description of uncertain events in the future... basically the same meaning as the word ‘forecast’ (Yang, 2017, p. 5).

There is currently work being performed by VDH contract to the University of Virginia’s Biocomplexity Institute to examine VDH SMP wastewater data as it pertains to the greater landscape of infectious disease forecasting. Such an endeavor requires knowledge bases and skills far beyond mine to pursue. Instead, I will focus on taking what we do have and establishing an evidence-based framework to detect real-time aberrations with correlated imminent case rises – an alert system.

Types of Early Warning Models

According to Yang, early warning functions as an “information conversion process”, using models and statistical tools to provide aberration detection within surveillance systems (Yang, 2017, p. 35). Commonly used models and algorithms include *temporal*, *spatial*, *spatio-temporal*, *regression*, and *multifactor* analysis tools. Proper evaluation (e.g., *sensitivity-specificity model*) is required, usually using the evaluation methods of sensitivity-specificity, average run length, root-mean-square error (RMSE), and mean absolute percentage error (MAPE) (Yang, 2017, pp. 35-36).

The focus of this dissertation is the *Data Analysis* part of Yang’s framework, by ascertaining the best model for aberration detection of the Sentinel Monitoring data, implementing it, and then evaluating it.

4.3 Is the SMP an Alert System?

With the above definitional lens, it can be stated that VDH's WWS Sentinel Monitoring Program (SMP) is a surveillance system that captures many of the above stated qualities. It is *Information-based surveillance*, as it is "continuous, systematic collection, analysis, and interpretation of disease outbreaks and their related factors" (Yang, 2017, p.1). The community SARS-CoV-2 wastewater viral load quantitative surveillance data is collected and analyzed with consistent methodology every week. With this weekly testing and reporting cadence, it is also *timely*. Also, it exists in a space of *blurry initial information*, as clinical data relies on substantial reporting efforts often taking weeks to trickle in. Even more obfuscating, clinical case data may go through changes in case definitions, collection methods, and/or reporting shifts. For instance, when take-home diagnostic tests became available, many Virginians no longer needed to go to a clinic to get tested, thus reducing overall COVID-19 reporting.

Where the current SMP falls short, however, is the *actionable & evidence-based* criteria. There are currently no evidence-based thresholds for initiating action with the SMP data. While the SMP dashboard includes percentiles for historical highs and lows, there are no warnings, alerts, or other metrics to indicate when an aberration has occurred – one which might result in imminent community outbreak.

As the VDH WWSP Data Manager, I try to compensate for this deficit by visually inspecting each site's trend line as it is received, and then sending emails to any linked VDH local district epidemiologists whose site spiked up noticeably. However, this is currently done merely on my best guess as someone who works with the data and is hardly evidence-based. Or reproducible, should anyone else perform my work.

By Yang's Conceptual Framework

Using Yang's conceptual framework (Yang, 2017, pp. 6-7), the current SMP collects specific and agreed-upon targets for upload into CDC NWSS DCIPHER (1. *Target Setting*), which is then collected by the utilities (2. *Data Collection*), shipped to DCLS for analysis (3. *Data Analysis*) and ultimately cleaned and wrangled by the SMP for delivery to appropriate stakeholders. What is missing, however, is the specific early-warning model 3. *Data Analysis*, 4. *Warning Release*, and 5. *Response*.

Assuming that the target of a public health alert system should be the affected public health practitioners (i.e., local epidemiologists), an appropriate warning metric needs to be assessed and then appropriately distributed. One could assume that what's relevant to the public health practitioner is also relevant to the public, however, this dissertation does not examine the nuanced perspective that is public interpretation and messaging. I leave that perspective to all health departments and their own respective communities.

Once a warning has been released to the appropriate targets, they can decide on the appropriate response to that information, depending on their needs and abilities. Some examples of possible actions could include:

- distribution of that information to affected stakeholders (e.g., hospital staffing)
- increased public health personnel for case tracking and other related tasks
- public messaging & mitigation (e.g., mask recommendations)

The real, practical information relayed by an SMP warning is to alert practitioners that an imminent rise in cases highly probable. This dissertation seeks to ascertain what metrics are required to properly assess that question, how they can be evaluated using the surveillance data available, and what are the appropriate triggers to initiate those alerts. It assesses what analysis is required and establishes a framework for generating that warning consistently. The framework

should, theoretically, be applicable to any WWS jurisdiction looking to create their own warning system.

This dissertation's target of Yang's framework is Step 3, the data analysis step. Specifically, this dissertation will include the necessary steps, according to Yang, for the building of early warning data analysis (Yang, 2017, p. 9):

1. early warning model building
2. threshold setting
3. signal alert generation

By establishing this piece, WWS jurisdictions can apply the model to their systems to generate appropriate data analysis. Then they can use that to generate their own warning releases and responses.

4.4 History of Infectious Disease Early Warning

Yang (2017) states that the concept of early warning marked a stark shift in public health surveillance, attributing the global shift to an April 1990 CDC "Current/Past Experience Graph" using the national notifiable infectious disease surveillance system and a historical limit statistic metric. It was published in the Morbidity and Mortality Weekly Report (MMWR), where several other countries followed suit soon after (Yang, 2017, p. 16).

Yang (2017, pp. 16-18) goes on to cite several major developments in international cooperation and global approaches to disease surveillance, including the:

- Federation of American Scientists (FAS)'s and International Society of Infectious Diseases (ISID)'s ProMED-mail internet reporting system in 1994
- WHO's Global Public Health Intelligence Network (GPHIN) in 1997
- WHO's Global Outbreak Alert and Response Network (GOARN) in 2000
- US-Canada-Mexico's joint Early Warning Infectious Disease Surveillance (EWIDS) in 2004

- CDC's National Syndromic Surveillance Program (NSSP)'s BioSense platform in 2017

Some of the possible motivating factors they suggest behind these advances included post-9/11 terrorism concerns, internet technology facilitation, and international cooperation efforts. *Syndromic surveillance* – the public health data gathering which relies on clinical case features instead of actual case diagnosis (e.g., looking for influenza-like-illness symptoms within hospital admissions for potential bioterrorism anthrax attacks) – became a new discipline following the wake of the September 11th (Mandl et al., 2004).

For China's history specifically, Yang (2017) mentions that – in response to the rampant infectious disease in the 1950s – the government established epidemic reporting in 1953, where it has since grown over the decades. In 1990, the Public Health Information System was established. However, after China experienced their SARS outbreak in 2003, infrastructure and building capacity were greatly improved, spurring the Prevention and Treatment of Infectious Diseases of the People's Republic of China, as well as the Notifiable Infectious Disease Reporting Information System in 2004. In 2007, the Emergency Response Law of the People's Republic of China was created, and by 2008, the China Infectious Diseases Automated Alert and Response System was released (Yang, 2017, pp. 18-25).

Historically, in both the U.S. and China, it seems as though a combination of technology (i.e., internet) feasibility and bioterrorism/outbreak concern spurred major public health investments in infectious disease surveillance. The modern parallel to the building of NWSS and wastewater surveillance, in the wake of the SARS-CoV-2 pandemic, seems very much in line with this historical concept.

4.5 Summary of Existing U.S. Alert Systems

While WWS is relatively new, and therefore limited in alert surveillance application, other systems exist for other public health data. In particular, the use of Control Chart (see Ch. 10.2 for more about public health Control Charts).

A particular subset of public health alerts includes syndromic data, which is designed to pick up indirect indicators which might portend increased incidence of disease cases. However, much of the reasoning behind the syndromic systems (as described in the above section) is to detect early cases of bioterrorism agents (Burkom, 2003). It is important to tailor a surveillance system to the nature of the data collected, as well as the goals of the program.

VDH does not currently have a WWS alert system in place, but other public health surveillance systems may offer insight into possible options.

Early Aberration Reporting System (EARS)

Developed and supported by the National Center for Infectious Diseases' Bioterrorism Preparedness and Response Program, CDC's EARS provides alternative aberration methods for syndromic public health data (Hutwagner et al., 2003). Their recommendations include the following long-term implementation aberration detection methods:

- For 3+ years of baseline data:
 - Historical Limits
 - Log-Linear Regression
 - Quality Control Cumulative Sums
 - Quality Control Compound Smoothing Technique
 - Cyclical Regression Models
- For limited baseline data:
 - C1-MILD (CUSUM with flag for baseline mean +3SD; baseline last 1-7 days)
 - C2-MEDIUM (CUSUM with flag for baseline mean +3SD; baseline last 3-9 days)

- C3-ULTRA (CUSUM with flag for baseline mean +3SD; baseline last 3-9 days)
- For drop-in surveillance (about 30 days run length):
 - Quality Control Methods (e.g., P Chart, moving average, CUSUM)
 - 2x2 tables, chi square summary statistics

ESSENCE – hospital syndromic surveillance

CDC’s National Syndromic Surveillance Program (NSSP) supports the secure integrated electronic health information system, BioSense Platform (CDC, 2024a). The internet-based BioSense Platform collects data for detection and monitoring of public health issues, such as injuries, environmental disasters, and emerging diseases. The BioSense Platform regularly collects data from nearly seven thousand health care facilities across all U.S. states, including 80% of all U.S. emergency departments (CDC, 2024b).

In Virginia, ESSENCE pulls in data from over 170 emergency departments and urgent care centers. Some examples of pulled data categories include gastrointestinal illness, influenza-like-illness, and disaster-related health events (VDH, 2023e).

Public health professionals with appropriate access may visualize and monitor the data through the Electronic Surveillance System for the Early Notification of Community-based Epidemics (ESSENCE) (CDC, 2017). ESSENCE can examine specific data elements, or pre-generated commonly used queries.

Part of ESSENCE’s visualization includes running query-specific time series, using colored data points indicating the following data point statuses: Normal (blue), Warning (yellow), and Alert (red).

A screenshot of ESSENCE’s data warnings & alerts, captured from a training video (CDC, 2024a), is shown below in Figure 4.4.1.

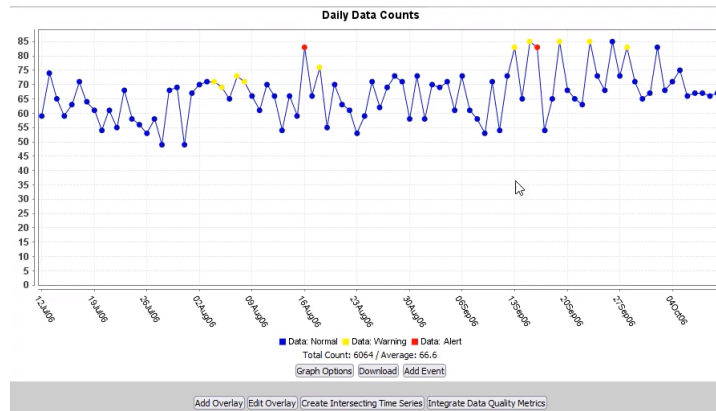


Figure 4.5: Screenshot within Nssp ESSENCE 101 Training (CDC, 2024a). Accessed Nov 17, 2024.

For the temporal visualizations, baseline is calculated from the last 30 days of data. The alert uses Poisson/Regression/Exponentially/Weight Moving Average switch algorithm to determine and calculate the best analysis (CDC, 2017).

ESSENCE’s algorithm models are described within the VDH intranet (VDH, 2024), and summarized below:

Adaptive Multiple Regression

- Baseline: last 30 days of counts in time series (not counting most recent two days) with outlier correction (to increase sensitivity, adjust for data problems or outbreaks)
- Trend Analysis: uses multiple least-squares regressions that account for systematic behavior in time series (e.g., Monday bumps)
- Aberration detection: regression multipliers calculated daily for expected versus observed counts with differences subtracted and scaled by the standard error of regression
- Alert statistic: Student’s T-test distribution significance (1% for red alerts; 5% for yellow alerts)
- Source: Brillman, et al. (2005) and Burkom (2003)
- Pros: seasonal and daily/weekly expected behaviors are accounted for
- Cons: must have baseline weekly and seasonal behavior

Adaptive Control Chart - EWMA

- Baseline: last 30 days of counts in time series (not counting most recent two days)
- Trend Analysis: exponential weighting to most recent values (e.g., within 3 days or other cutoff) and smoothing coefficients
- Aberration detection: weighted average of recent data to baseline expectation

- Alert statistic: Student's T-test distribution significance (1% for red alerts; 5% for yellow alerts)
- Source: Morton (2001)
- Pros: sensitivity to recent sudden outbreaks, as well as gradual outbreaks
- Cons: does not account for expected or seasonal behavior

Adaptive Switch Between Data Model & Control Chart – Poisson/Regression/EWMA

- Default algorithm
- Starts with regression (usually default for common respiratory/gastrointestinal illness), then flips to EWMA if test fails (usually for more rare illnesses)
- Poisson detection is used where there is less than a week of recent baseline data
- Regression Test: adjusted R² coefficient from data-derived threshold
- Poisson test: when there is less than a week of recent data, when mean and variance are equal to the mean
- Alert statistic: Poisson distribution significance (1% for red alerts; 5% for yellow alerts)
- Source: Burkom (2008)
- Pros: adjusts method to data
- Cons: occasional misclassification

Adaptive Control Chart – Algorithms C1, C2, and C3 (EARS)

- Uses EARS's C1, C2, and C3 algorithms (all use 3xSD against sliding baseline mean as alert flag)
- C1: moving baseline (ending previous day)
- C2: moving baseline (ending 3 days previously)
- C3: moving baseline (ending 3 days previously), but adds the previous two days' values
- Unlike EARS, the 7-day baseline is replaced with a 28-day baseline
- Source: Hutwagner et al. (1997), Tokars (2009)
- Pros: easy to understand
- Cons: no accounting for systematic data; only C3 has sensitivity to gradual outbreaks

Temporal Scan Statistic – GStat

- Like moving average chart, tests sum of cases in most recent time window (e.g., last 7 days)
- Test statistic: generalized likelihood ratio test (different from moving average statistic)
- Source: Wallenstein & Nause (2004), Wallenstein & Nause (2004-B)
- Pros: sensitive to brief signals/sparse cases
- Cons: complicated

CDC's WWS Mapping

CDC's Data Tracker for WWS (CDC, 2025) uses three main visualizations on their public Wastewater Metric: 1) current virus levels by site, 2) percent change in last 15 days, and 3) percent of samples with detectable virus (CDC, 2025). A recent addition to their site (CDC, 2025a) in late 2023 included Viral Activity Level.

All data is provisional and subject to change, with institution-specific and small sites (<3,000 people) suppressed. All data is lab method specific. None are alerts, per se, but each metric does confer relevant information about the state of the data.

Current Virus Levels by Site

- Normalized SARS-CoV-2 concentration
- Color-coded historical quintiles (collected after 12/1/2021)
- “New Sites” are for most recent 10 samples
- Most recent lab method only
- By-site selection: smoothed spline fit (last 4 months)

Percent Change in Last 15 Days

- Normalized SARS-CoV-2 concentration
- % change in the last 15 days (past to current data; at least 2 samples in the period required)
- Linear regression of log-transformed levels
- Does not show overall levels or history (after 15 days)

Percent of Samples with Detectable Virus

- % of positive samples within each site for the last 15 days (at least 1 sample)
- Does not show overall levels or history

Viral Activity Level

- Flow-population normalized SARS-CoV-2 concentration, log-transformed
- Samples with Z-scores above 4 dropped as outliers
- Baselines are re-calculated every 6 months (Jan 1st and July 1st) using the last 12 months of data (for sites with less history than 6 months, baselines are re-calculated weekly)
- Number of standard deviations per site/method calculate with every concentration value calculated against baseline, then converted back to linear scale (e^x)

- By-site weekly averaging
- National, regional, and state aggregation

None of the above metrics provide “alerts” as to whether a site’s viral concentration is at an actionable level, nor do any of the metrics consider contextual disease burden information like case or hospitalization data. The percentage change does not say whether large jumps in values are “normal” for a given site, though the WVAL can act as a snapshot to interpret whether aggregated levels are high compared to baselines. However, even WVAL does not specify an actionable level, nor bring real-world context (e.g., case and hospitalization correlation) into the metric. The WVAL also does not specify if those values have been increasing recently, or just at a seasonal high.

VDH WWS

Public Dashboard

The VDH WWS public dashboard (VDH, 2025) presents three major visualizations:

- Viral Load Trends – simple site increase/plateau/decrease/below detection/no trend linear regression trends from last 3-5 site data points (based on weekly or twice-weekly sampling)
- Percentiles – historical site percentiles divided into 5 categories
- Viral Load – by site historical plots for viral loads

None of the above metrics provide “alerts” as to whether a site’s viral concentration is at an actionable level, nor do any of the metrics consider contextual disease burden information like case or hospitalization data. The viral trends do not say whether large jumps in values are “normal” for a given site, though the percentiles can act as a snapshot to interpret whether site levels are historically high. The most descriptive visualization, the viral loads, can plot the by-site rises and falls of viral loads. However, this metric does not specify whether a recent rise is

“high enough” to warrant attention. None of the visualizations can really capture what is a normal variance jump versus one which portends a rise in cases.

Current Alert System

Since I’m the one who sends out alerts to all VDH epidemiologists regarding spikes in WWS viral loads each week, I can say with certainty that my ‘system’ is simply: if it ‘looks’ high, I send out an alert. I send emails to the local epidemiologists (and occasionally the regional epidemiologists) of whatever jurisdiction the unduly high site belongs to. No science, just opinion. When I eventually leave the program, my replacement will have to make that same guess. Improvements should be made.

4.6 On Selecting an Alert System Model

While Chapters 7-9 deal with the details of exploring the SMP surveillance data, the underlying nature of the data can already be used to lean into one of Yang’s early warning model ‘types’ over another.

Yang’s Early Warning Model Types (Yang, 2017, pp. 35-43):

- *Temporal*: time distribution of a specific region to detect significant increases or clustering
- *Spatial*: location distributions at one time point, or all time points, to detect clustering
- *Spatio-Temporal*: both temporal and spatial indicator analysis
- *Statistical Analysis Theories*: as defined by the analysis type (e.g., regression, statistical process control, and time series)
- *Multifactorial*: includes elements from multiple types

Yang’s Early Warning Evaluation Model Types (Yang, 2017, pp. 35-43):

- *Sensitivity-Specificity System*: uses indicators such as false positive rates, agreement rates, predictive value of positive and negative tests, receiver operating characteristic curve, area under the receiver curve, timeliness, active monitor operating characteristics, average run length, root-mean-square error, and mean absolute percentage error

The following chapters will demonstrate how site-specific the SMP data is, to include the fact that each site is unique in its sewage system/population, and as such, sites should not be compared to each other. Additionally, there are only a handful of pre-determined SMP sites across the state. As such, these static points are far less helpful than case data (with associated time and location markers) in identifying complex spatial or spatio-temporal patterning. Therefore, all spatial analyses comparing the sites are non-ideal.

As for statistical analyses approaches, many of these techniques lean into more advanced predictive modeling and go beyond the bounds of this dissertation. For example, the Serfling method (Serfling, 1963), a common regression-based technique, establishes baseline seasonality effects first, to gauge disease severity. Seasonality effects extend beyond the bounds of this dissertation, as they do not directly answer the question of whether, specifically, current wastewater levels indicate imminent rises in cases. Also, and perhaps more importantly, WWS is a relatively new tool and usually does not come into the picture with multiple years of data to draw upon.

As for multifactorial approaches to develop an integrated wastewater prediction tool using other data input streams, that is beyond this dissertation's scope. Emphasis here is on how valid WWS is to a user for predicting imminent disease burden only. There are other researchers attempting to integrate WWS generally into a holistic, predictive algorithm.

This leaves *temporal* analysis, within each site specifically, incorporating historical trends. This kind of warning would do well for local district epidemiologists, who wish to know if their region's wastewater is indicating an imminent increase in COVID-19 cases.

Temporal analysis models include (Yang, 2017, pp. 35):

- *Classic Time Series*: series changes and forecasting
- *Stationary Time Series* (Representative): specific trend moving average and weighted moving average
- *Stochastic Time Series*: autoregressive and autoregressive integrated moving average models

Time series, in general, uses statistical process controls to detect aberrant data points. These typically include control charts, moving average control charts, and moving percentile methods. Syndromic surveillance also leans into cumulative sums and exponentially weighted moving averages (Yang, 2017, pp. 36-37). Using the classic control chart with cumulative sums and exponentially weighted moving averages also has a history in hospital-acquired infections detection (Morton, 2001).

Chapter 5: Ethics

With consideration to how newly popularized WWS is to public health, a special section on ethical use is warranted. As with any surveillance system, especially one in which the public is opted in without informed consent, efforts must be maintained to balance privacy concerns with public good. This is especially true in the current political climate, where the Virginia Governor, on May 9th, 2023, mandated wastewater drug surveillance as part of his fentanyl law enforcement initiative (Youngkin, 2023).

5.1 Practical Considerations

WWS bears practical considerations, especially regarding the amount of variability involved in the collection and interpretation of WWS data. Ch. 3 explored these issues in greater detail, with the key points that WWS data collection is nuanced with many inherent variabilities. It is often difficult to get standardized, clear, and meaningful results from WWS, without a significant amount of research and experience first.

5.2 Applications

Current and future applications provide additional context to the directions and use of WWS. While only drug surveillance will be explored in more detail later in this chapter, some of the other applications, such as antibiotic resistance, are certainly worth a greater ethical exploration.

Current & Future Applications

Beyond the pandemic SARS-CoV-2 use WWS has garnered in recent years, there has been some research into a myriad of other applications. One comprehensive literature review (Choi et

al., 2018) explored current use and potential future applications for WWS, narrowing them down into several categories:

- Licit Drugs (e.g., caffeine, alcohol, nicotine)
- Illicit Drugs (e.g., psychoactive substances, performance enhancers)
- Pharmaceuticals and Personal Care Products (e.g., antihistamines, microplastics, UV filters)
- Population Markers (e.g., exogenous markers, endogenous markers)
- Industrial Chemical Markers (e.g., pesticides, flame retardants, plasticizers)
- Stress, Food, and Diet Markers (e.g., oxidative stress markers)
- Biologicals (e.g., pathogens, antibiotic resistance markers)

Generally, anything that could be measured by a urinalysis or stool sample could, most likely, be tested as a pooled WWS sample. The historical use section earlier qualified some of these options, but the practical understanding for what makes a good target is: 1) laboratories must have adequate PCR probes and a competent understanding in their methodology (which can take some time to develop, especially for novel pathogens), 2) there must be interest in testing for that target, and 3) the funding must be present for testing.

For example, polio has been of public interest lately, but WEF and CDC have advised against generalized WWS testing for polio. Since they will not support testing with their funds, and most labs themselves are not competent to perform polio testing, polio will most likely not be a target for testing soon. However, MPOX has public interest, adequate laboratory methods, and CDC funding available for ELC recipients.

Pushing through Barriers

McClary-Gutierrez et al. (2021) summarized the results of their WWS/public health expert panel into 4 major barriers that need addressing, shown below in Table 5.2.

Table 5.2: *Barriers to WWS (McClary-Gutierrez et al., 2021)*

Barrier	Recommended Best Practices	Future Needs/Strategies
Public health lack of comfort with WWS interpretation	Clear communication & collaboration with all parties	Research on WWS variability & uncertainties
Public health applications for increased confidence	Provide more case studies & retrospective analyses w/ data	Research, communication, and documentation
WWS systems require new knowledge & investment	Sharing of methods and co-developed programs	Investment in structure (lab, frameworks, etc.)
WWS ethics not yet established	Sample anonymity considerations & public engagement	Develop data use standards (policy & research)

These barriers summarized by McClary-Gutierrez et al. (2021) are very similar to the barriers we have been experiencing in the VDH WWS program. To move forward in WWS, public health practitioners will need to become more comfortable with WWS data generally – in addition to continued financial and methodological support from CDC.

5.3 Ethical Frameworks

The root principles of WWS ethics, though not well established per se, could be interpreted to derive from some of the disciplines which utilize it as a tool – namely, medical and public health, data governance, and federal/state laws and regulations.

Medical Ethics

The “four principles plus scope” is an ethical paradigm commonly used in the medical field (Gillon, 1994, pp. 184-188). It provides a common set of four moral commitments:

- *Autonomy* – respect for the ability to deliberate and make our own decisions (e.g., informed consent, confidentiality)
- *Beneficence* – the intent to provide benefits to people
- *Non-maleficence* – the intent to not harm people

- *Justice* (i.e., fairness) – the imperative to act fairly between competing claims; distributive justice (i.e., fair distribution of resources), rights-based justice, and legal justice
- *Scope* – deciding who/what we have moral duty to (e.g., one’s patients)

From this lens, one could argue that WWS treads within several of these principles. The concept of *autonomy* would argue for all participants of a study or surveillance to have informed consent, which is very much not the case with WWS. Participants need only be hooked up to a participating municipal water utility to be ‘opted-in’. To ‘opt out’, an individual would have to be competently informed and then remove themselves from the municipal sewage grid.

Obviously, this would incur significant personal cost, to include potentially moving residences.

The concepts of *beneficence* and *non-maleficence* would argue for providing more benefits than harm to the WWS participants. The perceived benefits is probably one of the most convincing advocates for WWS, as there is a lot of potential public health gain to surveilling large community sewersheds – WWS represents a non-invasive, cost-effective, and relatively objective insight into the health of a community, absent of other limitations that typically affect population health data (e.g., survey bias, healthcare access, reporting issues). However, the harm risk potentially increases as one reduces the sewershed population size and thusly diminishes the protections of anonymity. When anonymity is removed, WWS seeking stigmatized or punitive behaviors/conditions (e.g., illicit drug surveillance) could affect individuals or local communities.

The concept of *justice* is certainly interesting as well, since the idea of who benefits or is harmed by surveillance should be considered, especially regarding already disenfranchised populations. For example, WWS targeting a minority population runs the risk of further stigmatizing, or otherwise penalizing, this group.

Public Health Ethics

The twelve *Principles of the Ethical Practice of Public Health* (Thomas et al., 2002) were originally created and drawn from other known ethical concepts, including individual human rights, distributive justice, duty, human interdependence, and autonomy. The principles include:

“Principles of the Ethical Practice of Public Health

1. Public health should address principally the fundamental causes of disease and requirements for health, aiming to prevent adverse health outcomes.
 2. Public health should achieve community health in a way that respects the rights of individuals in the community.
 3. Public health policies, programs, and priorities should be developed and evaluated through processes that ensure an opportunity for input from community members.
 4. Public health should advocate for, or work for the empowerment of, disenfranchised community members, ensuring that the basic resources and conditions necessary for health are accessible to all people in the community.
 5. Public health should seek the information needed to implement effective policies and programs that protect and promote health.
 6. Public health institutions should provide communities with the information they have that is needed for decisions on policies or programs and should obtain the community’s consent for their implementation.
 7. Public health institutions should act in a timely manner on the information they have within the resources and the mandate given to them by the public.
 8. Public health programs and policies should incorporate a variety of approaches that anticipate and respect diverse values, beliefs, and cultures in the community.
 9. Public health programs and policies should be implemented in a manner that most enhances the physical and social environment.
 10. Public health institutions should protect the confidentiality of information that can bring harm to an individual or community if made public. Exceptions must be justified on the basis of the high likelihood of significant harm to the individual or others.
 11. Public health institutions should ensure the professional competence of their employees.
 12. Public health institutions and their employees should engage in collaborations and affiliations in ways that build the public’s trust and the institution’s effectiveness.”
- (Thomas et al., 2002, pp. 1057-1059)

WWS plays within several of these principles. Like Principle #2, WWS assesses community health without having to rely on datasets contributed from individuals -- so long as the sewersheds selected are not so granular as to invade individual privacy (i.e., populations less than 3,000). Principle #4 may be served by assessing communities that may not typically get served – so long as the information gathered does not stigmatize or penalize those communities disproportionately. Principle #10 deals with protecting the information of individuals and communities, becoming especially relevant if surveilling for stigmatizing or punitive health behaviors/conditions. Efforts in data collection, to include assuring anonymity by selecting large communities and targeting non-identifying information, are essential to mitigating individual harms. Community harm must also be addressed, as information can be harmful to reputation. For example, if a beach community saw a spike in SARS-CoV-2 viral loads on the state or national dashboard, that may adversely affect their tourism business.

Several of the other principles involve gathering and using information for the betterment of programs/policies, of which WWS could participate as an additional metric for assessing community health. Ideally, programs and policies would use this information appropriately.

Data Governance

Data ethics are also often regulated by higher institutions, including the state and federal government with laws including HIPAA and FOIA. Public health is often very reticent to release information to the public that could identify or harm an individual or community. Institutions such as the Department of Health have additional governance rules and responsibilities that must be followed by their employees and contractors.

Legal Issues

I am not a lawyer and will not provide a detailed analysis of legal issues involving WWS. One would expect, however, certain (state, federal, and local) constitutional and regulatory complications regarding WWS. The Fourth Amendment to the U.S. Constitution with its applied expectations of privacy to unwarranted search and seizure, may argue for or against WWS at the more granular level. Much of the argument may fall onto whether there is a societal expectation of privacy when flushing a toilet on municipal sewage.

5.4 Drug Surveillance

Even though WWS was not a public health staple decades ago, the inference of WWS as a potential (theoretical) tool for drug surveillance was present. Some academics (e.g., Hall et al., 2012) argued that, by examining drug WWS through the ethical lens of autonomy, non-maleficence, beneficence, and distributive justice, there was likely low potential for harm – when anonymity is safely guaranteed by large catchment populations, which in Hall et al.’s case was presumed to be at least 10,000 people. When examining smaller communities (e.g., schools and prisons), they argued that the risk of harms, such as stigmatization or punitive measures, increases and should be mitigated (Hall et al., 2012).

Additionally, some legal reviewers (Hering, 2009) have broached concerns about how wastewater drug surveillance could potentially challenge Fourth Amendment rights to privacy:

“Although scientists point out that technical flaws still exist with community urinalysis, civil libertarians harbor a far more ominous concern: that community urinalysis could be used by law enforcement to ‘tap’ plumbing in the same way they can tap telephone lines. Law enforcement could use the principles behind community urinalysis to test individual neighborhoods, or even individual homes, for evidence of drug activity.... This Note analyzes the possible Fourth Amendment implications of community urinalysis and

sewer epidemiology. It argues that the use of community urinalysis technology on an individual home is a search, requiring a warrant under the Fourth Amendment.” (Hering, 2009, p. 744)

5.5 Virginia Executive Order 26

On May 9th, 2023, Virginia Governor Glenn Youngkin signed Executive Order 26 (EO-26) mandating wastewater drug surveillance in his fentanyl law enforcement initiative, *Crushing the Fentanyl Epidemic: Strengthening Virginia’s Interdiction and Enforcement Response to Fentanyl Crisis* (Youngkin, 2023). The Executive Order outlines 10 directives, with the purported intention of reducing fentanyl overdoses and deaths. Of note to WWS, directives 3 and 9:

Directive 3:

“I direct the Department of Health to develop a cost-effective plan to utilize and fund wastewater surveillance to detect the frequency, potency, and occurrences of fentanyl use in specific locations. This plan shall include a response strategy that includes increased naloxone distribution, targeted public awareness campaigns, and other cost-effective strategies to reduce fentanyl’s prevalence in those communities where surveillance warrants increased response. The plan shall be sent to the Secretary of Health and Human Resources within 120 days of the date of this order.” (Youngkin, 2023, p. 2).

Directive 9:

“I direct the agencies listed below to send a plan to me to report, in accordance with federal and state law, their data to and work with the Virginia State Police’s fusion center and to participate in the Framework for Addiction Analysis and Community Transformation (FAACT), a secure data-sharing project led by the Department of Criminal Justice Services (DCJS) in collaboration with the Office of Data Governance and Analytics (ODGA), within 120 days of the date of this order. This effort will create the most cohesive and timely data on fatal and non-fatal overdoses in the Commonwealth

to allow for immediate interdiction, education, and abatement efforts in the neighborhoods where spikes in overdoses are observed.

- Department of Health
- Department of Health Professions
- Department of Social Services
- Department of Medical Assistance Services
- Department of Behavioral Health and Developmental Services” (Youngkin, 2023, pp. 3-4).

These directives mandate fentanyl drug WWS and require multiple agencies to participate and data share with Virginia State Police, the Department of Criminal Justice Services, and the Office of Data Governance and Analytics. No privacy or ethical issues regarding WWS are addressed, to include minimum population standards or other anonymity protections. The only indication to the intended WWS target population size are the keywords “specific locations” (Youngkin, 2023, p. 2) and “neighborhoods where spikes in overdoses are observed” (Youngkin, 2023, p. 3). These keywords suggest a more targeted approach than the general minimum population standards of around 3,000 people.

However, the Executive Order does, in fact, specify the intended target demographic, in its third paragraph:

“The scope of the problem has not only grown over the past decade, but the communities hit hardest have also changed. Historically, white males regularly had the highest rates of fatal opioid overdoses in Virginia. In 2019, however, black males had the highest rate of fatal opioid overdoses in the Commonwealth. From the latest available data, in 2022 black males fatally overdosed on fentanyl 1.9 times as often as white males (rate of 60.4 and 31.4 per 100,000, respectively). More must be done to reach these communities and connect them to a pathway to recovery and renewal.” (Youngkin, 2023, p. 1).

It is reasonable to infer that the targeted “specific locations” and “neighborhoods” for drug surveillance will disproportionately draw from the “communities” listed above, namely black males. As EO-26, even in its title, identifies primarily as an interdiction and enforcement order, pause should be taken when any kind of surveillance is added -- and used for law enforcement – against an already marginalized demographic group.

5.6 EO-26 Ethical Issues

With consideration to the ethical lens of autonomy, non-maleficence, beneficence, and distributive justice, several red flags appear in EO-26. Almost all these flags stem from EO-26’s lack of commitment to anonymization.

The first is Autonomy -- the respect and non-interference of rational persons. With an executive mandate being what it is, public health employees and other agency staff may be coerced into performing research that they may not feel comfortable conducting. Or feel comfortable reporting directly to law enforcement. Of course, most importantly, WWS is generally an automatic opt-in process, which means that all observed subjects will not be able to give proper consent for monitoring.

Second is Non-Maleficence – the avoidance of causing harm. When anonymity is not guaranteed, conducting drug surveillance has the risk of stigmatizing the studied populations, or worse, can lead to law enforcement punitive harms. This is even more pronounced in context to the recent Virginia terrorism definition change (HB 1682, 2023), charging all those found manufacturing or distributing fentanyl with a Class 4 felony.

Third is Beneficence – the reasonable assumption of the study benefiting the subjects. One could argue that, from a public health standpoint, gaining insights into community drug use

could help support public health practitioners to better serve their communities. However, there is also a risk of stigmatization or punitive harms on those populations specifically, especially if efforts are not taken to protect the identities of those groups.

Fourth is Distributive Justice – the fair distribution of risks/benefits to the subjects. EO-26 does not specify how the “specific locations” will be identified, nor how testing will ensure anonymization and demographically representative sampling to the greater community. However, EO-26 did mention, in its third paragraph, who the users of fentanyl are – black males. It stands to reason that black males will be disproportionately targeted for fentanyl surveillance and will be disproportionately subject to its interdiction and enforcement measures.

Chapter 6: Data & Analysis Methods

6.1 IRB Determination

I submitted a protocol of this research project to the Virginia Tech Institutional Review Board (IRB) before data collection. The classification of research was determined to be “Not Human Subjects Research.”

IRB Number: 23-282

6.2 Data Sources

Multiple data sets were used, many from VDH. It is important to note that all data, especially COVID-19 health department data, is provisional and subject to change. This doesn’t just include lags in data reporting but also includes things like case definitions and collection methods changes. The population pool the datasets drew from also changed, as people were either born or died over the years or changed their susceptibility status (via vaccination or actually acquiring COVID-19). COVID-19 ‘cases’ may, therefore, mean something very different in 2025, as compared to 2020. It is important to keep in mind the nuances of the datasets and take each in the context of a greater picture. This is true with WWS data, as it is with case or hospitalization data.

WWS data

VDH WWS Sentinel Monitoring Program data was requested through the VDH FOIA NextRequest system (<https://vdh.nextrequest.com>). Data requested from the WWS Sentinel Monitoring Program (VDH, 2023): Sentinel Monitoring WWS data, case counts, sewershed shapefiles, and miscellaneous dashboard/GIS screenshots.

FOIA request #: 23-1886 and 23-1888.

A VDH Single Researcher Agreement – a Data Sharing Agreement (DSA) -- was signed and submitted to VDH Data Governance on March 21, 2023. Terms of the data use were stipulated, to include only using the data for the expressed purpose and for the sharing of results with VDH. Data was shared periodically over the next few weeks, with the final datasets dated April 25th, 2023.

Variant Proportions

Two sources of information were used to determine SARS-CoV-2 variant proportions over time (see Initial Data Exploration chapter). The first dataset was the FDA WWS Variant Proportions visualization, found on their dashboard main page (FDA, 2023). This dataset was used to estimate general U.S. WW trends across time. The second variant proportion dataset used was a VDH public use dataset for SARS-CoV-2 Variants of Concern (VDH, 2023c). It included weekly cumulative numbers for Virginia COVID-19 infections, hospitalizations, and deaths for different COVID-19 Variants of Concern. It was last updated April 11, 2023.

VDH Case Data

While the VDH WWS data above does include spatially-linked case counts to sewersheds, VDH COVID-19 case data was also downloaded from a public use dataset (VDH, 2023b). It included daily cumulative numbers of COVID-19 cases, hospitalizations, and deaths by locality. It was last updated on October 10, 2023. In order to examine this VDH public use dataset by site, sites were bound to the closest local health district. This was done using the VDH Locality to health district chart, found at:

<https://www.vdh.virginia.gov/content/uploads/2020/05/Locality-to-HD-to-HPR.pdf>

6.3 About the WWS Data

Sentinel Monitoring Program

VDH's WWSP is funded by CDC's ELC grant. Funding is used to establish WWS in Virginia by paying the salaries of the VDH WWSP core team, building WWS networks, supporting CDC NWSS data upload, funding WWS testing, and facilitating localized/collaborative projects. The main testing avenue of the WWSP is its Sentinel Monitoring Program (SMP), a WWS program collecting samples from partnered utilities across the state. Sampling started September 13th, 2021, with an original 25 utility partners, and is still running to date.

SMP Data

SMP wastewater samples are collected by utility partners once- or twice-weekly at each of the WWTP's influents. The samples are collected at roughly the same time each week, by either grab or 24-hr composite sample techniques, and then shipped to state laboratory partners at the Division of Consolidated Laboratory Services (DCLS) (VDGS, 2023). DCLS analyzes the samples using PCR SARS-CoV-2 N and E gene targets and then combines their analytical data with submitted sample collection metadata (e.g., sample collect time, flow rate). Results are then submitted to the VDH WWSP for additional cleaning, analytics, and upload into CDC NWSS by the VDH WWSP Data Manager.

The WWS datasheet received from the VDH WWSP contains the SMP WWS submitted to CDC NWSS – specifically, through the NWSS Data Collation and Integration for Public Health Event Response (DCIPHER) portal. A suppressed version of this dataset is used for

CDC’s COVID Data Tracker (CDC, 2023a). A screenshot of their public-facing dashboard can be seen below in Figure 6.3:

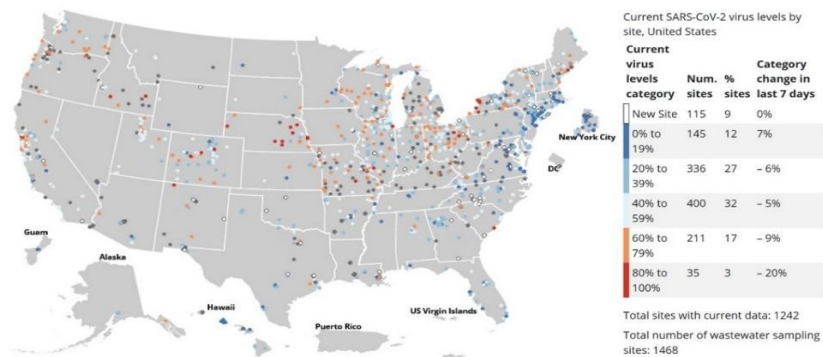


Figure 6.3: Screenshot of Wastewater Metric Map on CDC COVID Data Tracker (CDC, 2023a)

It is practice in NWSS to suppress WWS data with populations less than 3,000. This is done to enhance anonymity of cases. However, the VDH public dashboard (VDH, 2023a) does show a few participating sites below this threshold (with cases in the 1-4 range suppressed to 2).

Dataset Adjustments

Several data elements from the original dataset have been dropped, due to not being applicable to answer the dissertation’s main research focus. One variable was created.

Dropped elements:

- Any twice/weekly sampling – a few sites participated in twice/weekly sampling, and those that did tended to only do so only after about 45 weeks into the program.
- E gene analysis – confirmatory gene target only. N gene is considered the more appropriate standard for analysis (Gonzalez et al., 2020).
- Sites not included at the start of the SM program in September 2021.
- Sites which changed their service areas during the program – specifically, the Atlantic and Chesapeake Elizabeth sites, which combined their service areas in December 2021.

Created variables:

- Viral Load (copies per day) – the number of viral copies passing through a WWTP that day. Requires a calculation multiplying the recorded WWTP flow rate, the PCR target concentration, and a constant to adjust the units.

Viral Load Calculation:

Viral Load (copies per day) = flow_rate (million gallons per day) x pcr_target_avg_conc (copies per liter) x 3,785,411.8 (liters per day / million gallons per day)

DCLS Lab Analysis Data Descriptors

The following laboratory variable fields remained relatively consistent throughout the received dataset and have thus been removed from the dissertation analyses.

Table 6.3.1: DCLS Lab Analysis Data Descriptors

Variable	Field(s)
concentration_method	innovaprep ultrafiltration
ext_blank	yes
extraction_method	zymo quick-rna viral kit #r1035
inhibition_detect	not tested
inhibition_method	https://doi.org/10.1016/j.watres.2020.116296 ; https://doi.org/10.1016/j.scitotenv.2020.139960 ; https://doi.org/10.1016/j.scitotenv.2020.139961
lod_ref	https://doi.org/10.1101/2021.03.02.21252754
ntc_amplify	no
num_no_target_control	1; 2
pasteurized	no
pcr_gene_target	ddcov_n, ddcov_e
pcr_gene_target_name	Bcov vaccine; murine coronavirus
pcr_gene_target_ref	https://doi.org/10.1101/2021.03.02.21252754
pcr_target	sars-cov-2
pcr_target_units	copies/L wastewater
pcr_type	ddpcr
quant_stan_type	rna
rec_eff_spike_matrix	raw sample
solids_separation	centrifugation
stan_ref	https://doi.org/10.1016/j.watres.2020.116296 ; https://doi.org/10.1016/j.watres.2020.116297
tot_con_vol	40

Table 6.3.2: VDH WWS SM Dataset Descriptors

Sources	Utility partners (samples and metadata), DCLS (laboratory analysis, metadata collation), VDH WWSP cleaning/adjustments
VDH WWSP program	Sentinel Monitoring Program
Collection Start Date	September 13, 2021 (Week 1)
Last Collection Date	April 18 th , 2023 (Week 84)
Dropped Variables & Rows	E gene targets; the above consistent variables and other non-essential variables; Twice-weekly samples ('B' samples; applicable sites only); <u>Merged sites</u> : Atlantic, Chesapeake-Elizabeth; <u>Late-Joining SM sites</u> : Galax, Halifax, Henrico, Army Base, Boat Harbor, James River, York River, Independence, Low Moor, Lower Jackson, Richlands, Tazewell, Lebanon, Rocky Mount, Wytheville, Broad Run, Clintwood
Variables Kept (17)	sample_id, wwtp_name, zipcode, capacity_mgd, population_served, sample_type, sample_matrix, sample_collect_date, sample_collect_time, flow_rate, tset_result_date, pcr_target_avg_conc pcr_target_below_lod, lod sewage, rec_eff percent, quality_flag
Added Variables	Viral Load
Number of Rows	1,835 (samples)
Unique Sites	24
Site Name (wwtp_name) / Shortened Name	Alexandria Renew Enterprises / AlexRenew Aquia Wastewater Treatment Plant / Aquia Basham Simms Wastewater Treatment Facili / Basham Blacksburg-VPI Sanitation Authority / Blacksburg City of Danville Northside Wastewater Tr / Danville Coeburn Norton Wise (CNW) Regional Waste / Coeburn H.L. Mooney Advanced Water Reclamation F / HLM HRSD: Nansemond / NA HRSD: Onancock / ON HRSD: Virginia Initiative Plant / VIP HRSD: Williamsburg / WB Little Falls Run Wastewater Treatment Fa / LFR Lynchburg Water Resource Recovery Facili / LB Moores Creek Advanced Water Resource Rec / Moores North River Wastewater Treatment Facilit / NorthRiver Parkins Mill Wastewater Treatment Facili / Parkins Pepper's Ferry Regional Wastewater Treat / Peppers Pound Wastewater Treatment Plant / Pound Richmond Wastewater Treatment Plant / Richmond Roanoke Regional Water Pollution Control / Roanoke Rutledge Creek Wastewater Treatment Plan / Rutledge Town of Christiansburg Wastewater Treatm / CB Upper Occoquan Service Authority Regiona / UOSA Wolf Creek Water Reclamation Facility / Wolf

While the dataset used for this dissertation was received on April 25th, 2023, the VDH WWSP continues to collect Sentinel Monitoring data to date. Several more sites outside of the original 25 have since been added into the SM program.

Sewershed data

The WWSP has received -- or created, by the WWSP Data Manager -- sewershed spatial files for each participating site. This GIS shapefile shows the service region (e.g., the population feeding into) of each of the WWS collection points. The GIS shapefile package received on April 25, 2023 contained the sewershed boundaries of 39 sites in the Sentinel Monitoring Program. While 24 of the original 25 sites had a corresponding boundary shapefile, a few of the newer additions have yet to provide (or work with the VDH WWSP Data Manager to create) a shapefile of their service region.

It is practice, in both the VDH WWSP and NWSS, to suppress sewershed shapefile boundaries from public release. This is due to the proprietary nature of the shapefiles, as well as enhance anonymity for smaller sewersheds.

Case data

Using the sewershed shapefiles, the WWSP Technical Specialist then downloads VDH case data from the VDESS system. The case data is spatially linked to the appropriate sewersheds, and case count metrics are generated for each site. The case CSV dataset received on April 25, 2023 contained 4 variables: the date *event_date* (ranging from September 10, 2021 – April 23, 2023), site name *sewershed*, case count *count*, and the suppressed count *suppressed_count*. As all dates with a case count greater than 0 for each sewershed were given a row, there were 23,830 rows in total from 44 unique sites. It is practice in both the VDH WWSP

and in NWSS to suppress publicly available visualizations with case counts in the 1-4 range (suppressed to 2). This is done to enhance anonymity of cases.

Chapter 7: Initial Data Exploration

7.1 Variant Dominance

The WWS data analysis plan includes training a theoretical alert system on the premise of a *baseline*. Determining periods across the VDH Sentinel Monitoring Program collection period in which variables are as controlled as possible -- including the SARS-CoV-2 circulating variant strain -- will be necessary to establish normal, baseline variance within each sewershed.

Two sources of information were used to determine SARS-CoV-2 variant proportions over time. The first was the FDA WWS Variant Proportions visualization, found on their dashboard main page (FDA, 2023). The second was the VDH public use dataset for SARS-CoV-2 Variants of Concern (VDH, 2023c). See Methods chapter for details.

The FDA data was used to identify general U.S. WW trends across time, while the VDH data was used to identify the exact proportions of relevant variants in Virginia clinical cases, over time.

General WWS Trends

The FDA WWS Variant Proportions visualization, found on their dashboard main page (FDA, 2023) and seen before in Chapter 2's Figure 2.4.4, outlines general WWS trends across time. According to the FDA infographic (FDA, 2023), at the start of the VDH SM Program (September 2021), the overwhelming dominant variant abundances were the Delta sublineages. The week of December 12, 2021 marked the introduction of the BA.1 Omicron (6%) and X Recombinants (5%) to the Delta lineages (88%). Two weeks later (January 2, 2022), BA.1 Omicron stood at 77%. One week later, BA.1 Omicron was at 94%, and stayed there for the next two months.

BA.2 Omicron was introduced during the week of March 13, 2022 (21%). One month later (April 10, 2022), BA.2 Omicron stood at 87%. On May 22, 2022, BA.5 Omicron was introduced (6%), and seven weeks later (July 10, 2022) stood at 83%. Omicron BQ.1 was introduced on October 23, 2022 (5%), and peaked January 1, 2023 at 57%. Omicron XBB.1.5 was introduced on January 8th, 2023 (9%) and peaked March 23, 2023 (83%).

Simplified FDA Dominant Variant Timeline

To assess which variant was dominant during the SM collection weeks, the above FDA WWS Variant Abundances data was simplified to include only labels for weeks where a variant reached a dominance of at least 65% abundance. The abundance of 65% was selected as a threshold because it allowed some abundance above a simple majority, without requiring the variant to completely eclipse all others.

If a particular variant was indicated to be at least 65% in abundance, the variant was identified and then scatter plotted in Excel over time (see Figure 7.1.2 below). Points connected to adjacent weeks were drawn with a line between, indicating that the variant dominated during that duration of time.

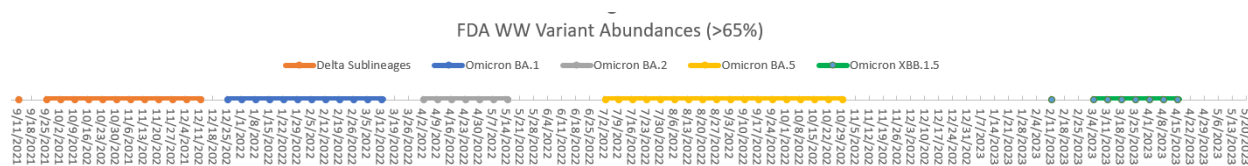


Figure 7.1.1: Excel summary of FDA WW variants containing over 65% relative abundance by week.

The dominant variant timeline indicated above by the FDA WW simplified dominant weekly abundance demonstrated a near-consistent continued Delta sublineage dominance from the start of early September through mid-December 2021, with the end week of December 11, 2021.

The dominant variants indicated above -- Delta sublineages, Omicron BA.1, Omicron BA.2, Omicron BA.5, and Omicron XBB.1.5 -- represented the chosen targets for VDH Clinical Variant Analysis (below).

VDH Clinical Variant Trends

The second variant proportion dataset used was a VDH public use dataset for SARS-CoV-2 Variants of Concern (VDH, 2023c). It included weekly cumulative numbers for Virginia COVID-19 infections, hospitalizations, and deaths for different COVID-19 Variants of Concern.

Each variant of VDH clinical variant case counts dataset was evaluated by exploring the cov-lineages.org (cov-lineages.org, 2023) lineage list for their Pango description and their CDC variant listing (CDC, 2023e) for WHO labels. If a WHO label (e.g., Omicron) was included in the CDC description, it was noted in the tables below.

As seen below in Table 7.1.1, all variants and associated descriptive statistics were included in the VDH clinical dataset by first observed date.

Table 7.1.1: *All VDH clinical case variant descriptive statistics and WHO label*

Variant	Total Case Counts	Date Range	WHO Label
Alpha	62	9/4/21 – 2/25/23	Alpha
Epsilon	43	9/4/21 – 2/25/23	Epsilon
NONE	2,547	9/4/21 – 3/18/23	-
[Blank]	420	9/4/21 – 12/31/22	-
Mu	11	9/4/21 – 7/23/22	Mu
Gamma	13	9/4/21 – 7/30/22	Gamma
Delta	23,058	9/4/21 – 9/10/22	Delta
Zeta	10	9/4/21 – 11/12/22	Zeta
Eta	3	9/18/21 – 9/17/22	Eta
Beta	69	9/25/21 – 4/1/23	Beta
Omicron	6,024	10/2/21 – 3/25/23	Omicron

XBB.1.5.1	6	11/6/21 – 3/25/23	Omicron
BA.1.1 Omicron	12,289	11/27/21– 11/24/22	Omicron (alias B.1.1.529.1.1)
Iota	7	1/1/22 – 12/31/22	Iota
BA.2 Omicron	16,025	1/8/22 – 2/18/23	Omicron (alias B.1.1.529.2)
XBB.1.5	2,902	1/15/22 – 4/1/23	Omicron
BA.5 Omicron	16,113	1/22/22 – 3/18/23	Omicron (alias B.1.1.529.5.1)
BA.4 Omicron	3,139	4/16/22 – 12/24/22	Omicron (alias B.1.1.529.4)
BA.2.12.1	27	5/14/22 – 8/6/22	Omicron (alias B.1.1.529.2.12.1)
BA.4.6 Omicron	1,903	5/14/22 – 2/18/23	Omicron (alias B.1.1.529.4.6)
BA.5.2.6 Omicron	164	6/25/22 – 2/18/23	Omicron (alias B.1.1.529.5.2.6)
BA.2.75 Omicron	104	7/2/22 – 3/18/23	Omicron (alias B.1.1.529.2.75)
BF.11	117	7/16/22 – 2/11/23	Omicron (alias B.1.1.529.5.2.1.1)
BF.7 omicron	432	7/23/22 – 2/18/23	Omicron (alias B.1.1.529.5.2.1.7)
BA.2.75.2	49	8/20/22 – 2/25/23	Omicron (alias B.1.1.529.2.75.2)
BN.1	162	9/10/22 – 3/4/23	Omicron (alias B.1.1.529.2.75.5.1)
BQ.1	850	9/10/22 – 3/18/23	Omicron (alias B.1.1.529.5.3.1.1.1.1.1)
BQ.1.1	3,060	9/17/22 – 3/25/23	Omicron (alias B.1.1.529.5.3.1.1.1.1.1.1)
XBB	504	10/15/22 – 4/1/23	Omicron
CH.1.1	149	10/29/22 – 3/25/23	Omicron (alias B.1.1.529.2.75.3.4.1.1.1.1)
Kappa	1	12/17/2022	Kappa
XBB.1.9.1	13	1/28/23 – 4/1/23	Omicron

Case counts were then subselected by the FDA WW dominant abundances (Delta sublineages, Omicron BA.1, Omicron BA.2, Omicron BA.5, and Omicron XBB.1.5), as noted above in the Simplified FDA Dominant Variant Timeline. Additionally, Omicron (including all sublineages) was viewed generally, with respect to Delta.

As seen below in Table 7.1.2, the VDH variant case count subselections included sublineages and descriptive statistics.

Table 7.1.2: *VDH variant case count subselections by sublineages, case counts, and date ranges.*

Variant	Variant Sublineages	Total Case Counts	Date Range
Delta	Delta	23,058	9/4/21 – 9/10/22
Omicron (generally)	BA.1.1 Omicron, BA.2 Omicron, BA.2.12.1, BA.2.75, BA.2.75.2, BA.4 Omicron, BA4.6 Omicron, BA.5 Omicron, BA.5.2.6, BF.11, BF.7 omicron, BN.1, BQ.1, BQ.1.1, CH.1.1, Omicron, XBB, XBB.1.5, XBB.1.5.1, XBB.1.9.1	64,032	10/2/21 – 4/1/23
Omicron BA.1	BA.1.1 Omicron	12,289	11/27/21 – 12/24/22
Omicron BA.2	BA.2 Omicron, BA.2.12.1, BA.2.75, BA.2.75.2	16,205	12/3/22 – 7/9/22
Omicron BA.5	BA.5 Omicron, BA.5.2.6	16,277	1/22/22 – 3/18/23
Omicron XBB.1.5	XBB.1.5, XBB.1.5.1	3,425	11/6/21 – 4/1/23

VDH Variants Over Time

In the following figures (Figures 7.1.3 – 7.1.8), line graphs were visualized in Excel for each of the selected VDH variant case counts over time.

Delta:

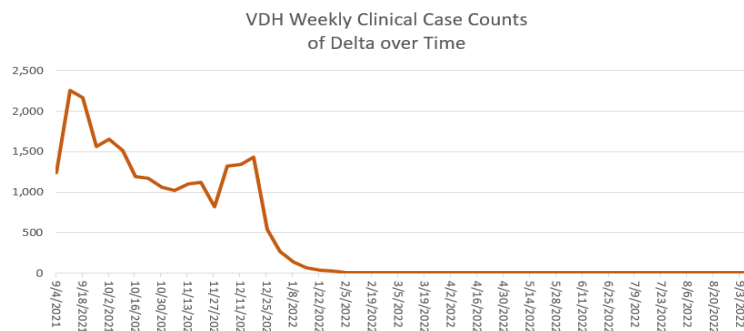


Figure 7.1.3: *Weekly Delta variant clinical case counts over time*

Omicron (generally):

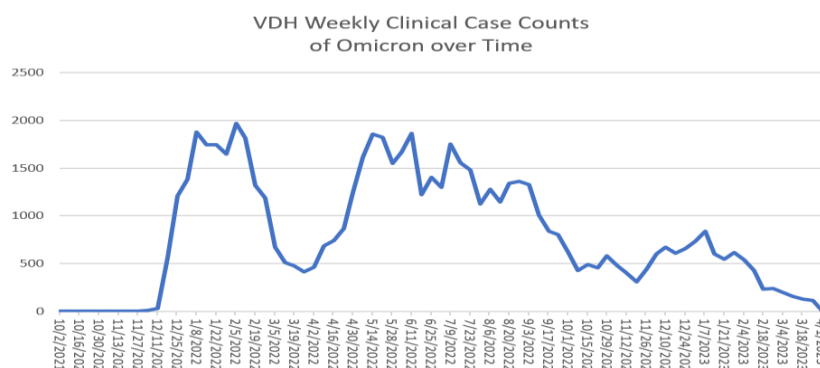


Figure 7.1.4: *Weekly Omicron variant clinical case counts over time*

BA.1:

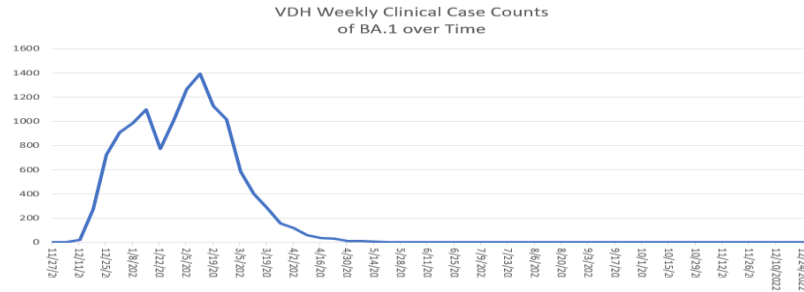


Figure 7.1.5: Weekly BA.1 variant clinical case counts over time

BA.2:

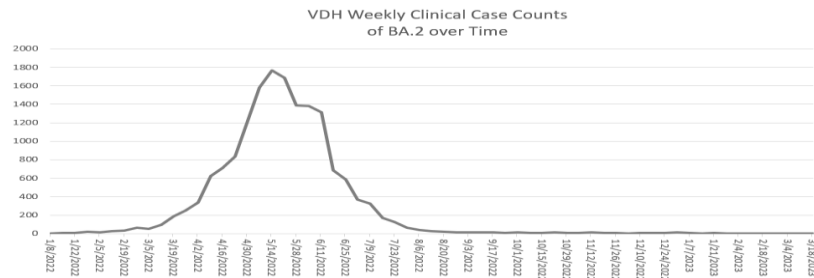


Figure 7.1.6: Weekly BA.2 variant clinical case counts over time

BA.5:

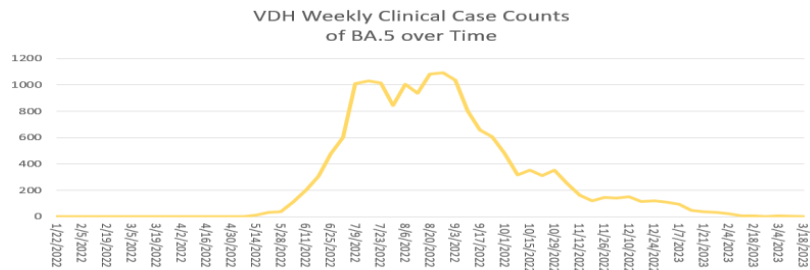


Figure 7.1.7: Weekly BA.5 variant clinical case counts over time

XBB.1.5:

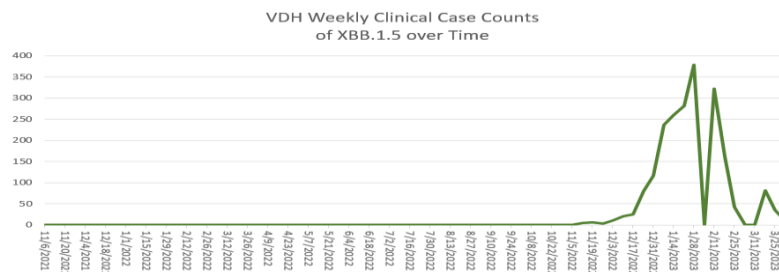


Figure 7.1.8: Weekly BA.5 variant clinical case counts over time

Bar graphs were then generated in Excel to contrast the transitions between each of the major dominant variants, as seen below in Figures 7.1.9 – 7.1.13.

Delta vs. Omicron (generally):

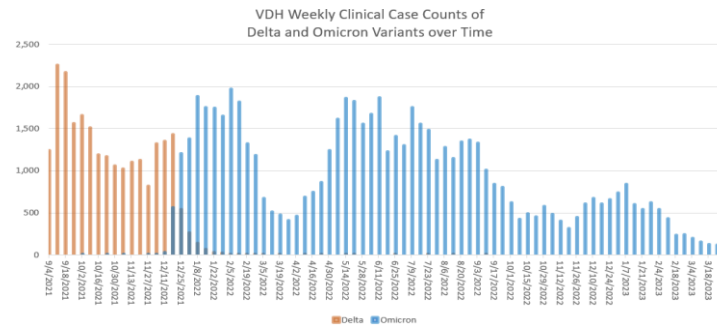


Figure 7.1.9: Weekly Delta and Omicron variant clinical case counts over

Delta vs. BA.1:

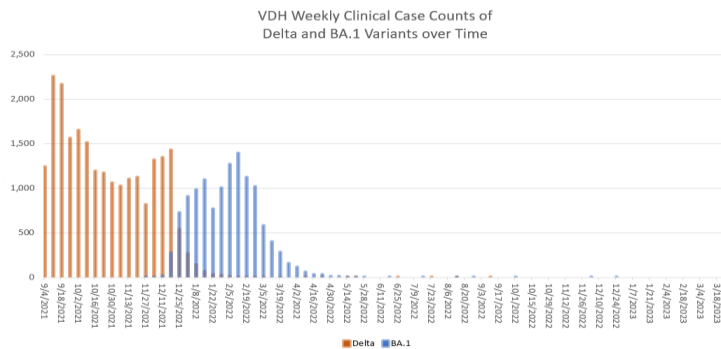


Figure 7.1.10: Weekly Delta and BA.1 variant clinical case counts over time

BA.1 vs BA.2:

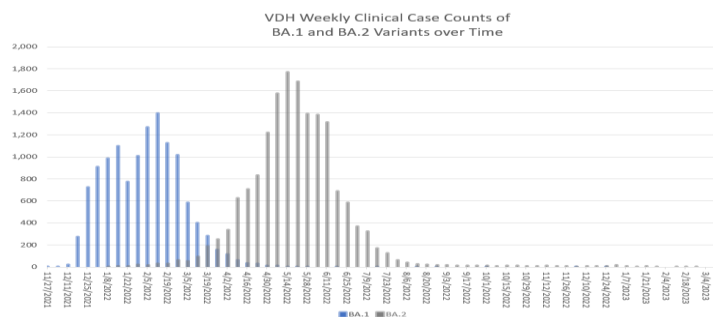


Figure 7.1.11: Weekly BA.1 and BA.2 variant clinical case counts over time

BA.2 vs BA.5:

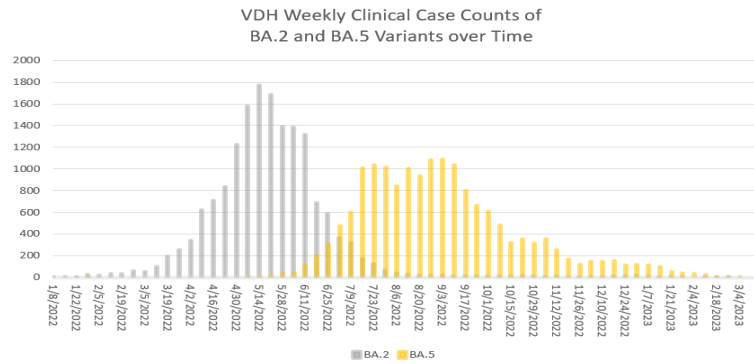


Figure 7.1.12: Weekly BA.2 and BA.5 variant clinical case counts over time

BA.5 vs XBB.1.5:

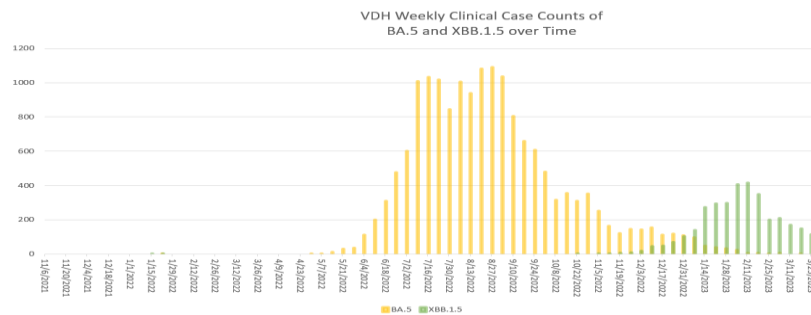


Figure 7.1.13: Weekly clinical BA.5 and XBB.1.5 variant clinical case counts over time

Synthesizing Variant Dominance Timeline

As seen below in Figures 7.1.14 and 7.1.15, the variants isolated above in line and column graphs were visualized to show the trends over time.

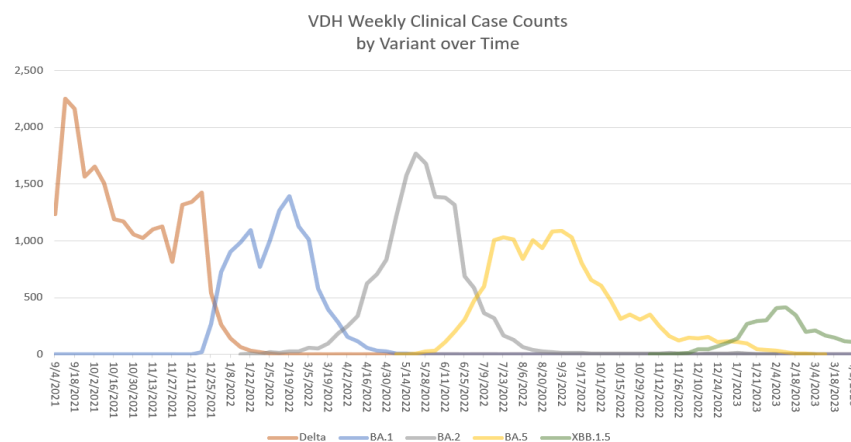


Figure 7.1.14: Weekly variant clinical case counts over time, line graph

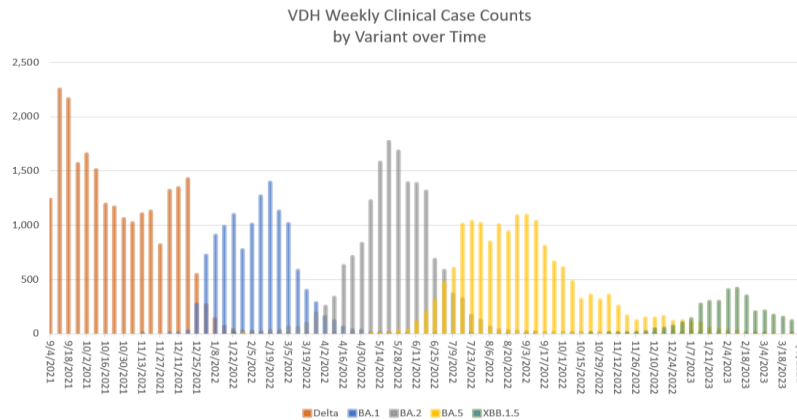


Figure 7.1.15: Weekly variant clinical case counts over time, column graph

By examining the VDH Clinical Variant trends above, through the frame of the FDA WW Variant Abundances, the following variant dominance timeline was deduced (Table 7.1.3).

Table 7.1.3: Synthesized dominant variant timeline

Dominant Variant	Start Date	End Date
Delta	September 4, 2021 (start date)	December 18, 2021
Omicron – BA.1	January 1, 2022	March 12, 2022
Omicron – BA.2	April 2, 2022	June 18, 2022
Omicron – BA.5	July 16, 2022	December 10, 2022
Omicron – XBB.1.5	Jan 14, 2023	April 1, 2023 (end date)

In instances where the FDA WW Variant Abundances timeline disagreed with VDH clinical data timeline, preference was given to VDH clinical data, as it should be a much better indicator of Virginia variant infection prevalence than federal wastewater data.

7.2 Sewersheds

While the VDH WWS program does collect the boundaries of the utility partner wastewater system service areas (i.e., sewersheds), the data is proprietary and not freely shared

for publication. It is for this reason that the actual boundaries files will not be visualized in this dissertation.

7.3 Case Counts

VDH COVID-19 Public Case Counts

VDH public use COVID-19 case data (VDH, 2023b) included cumulative daily COVID-19 case, hospitalization, and death counts by Virginia health district. All health districts were compiled together for a single count per day, as shown below in Figures 7.3.1 – 7.3.3. The 7/5/2023 data point in all three sets was removed, due it being an obvious error -- counts were about 10,000 times as high as adjacent points and were not considered in the next day's cumulative counts.

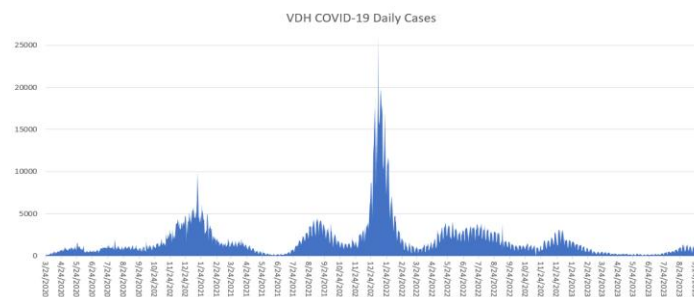


Figure 7.3.1: Daily VDH public COVID-19 case counts over time, stacked area graph

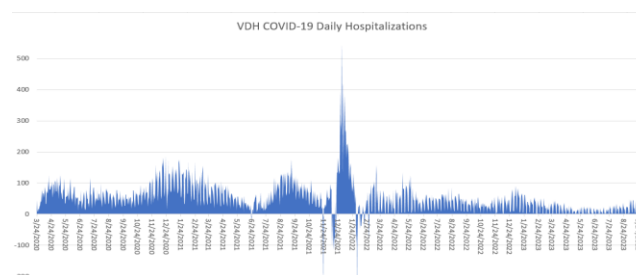


Figure 7.3.2: Daily VDH public COVID-19 hospitalizations over time, stacked

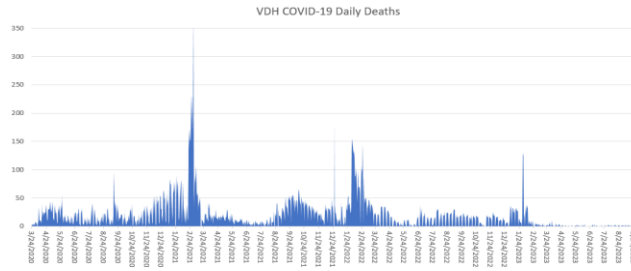


Figure 7.3.3: *Daily VDH public COVID-19 deaths over time, stacked area graph*

VDH WWS Sewershed Case Counts

Using the VDH WWS Case count data (seen below in Figure 7.3.4), the following 23 SM sewersheds were available to draw in daily case counts from Sept 10, 2021 to April 22, 2023: Alexandria Renew, Aquia, Basham Simms, Blacksburg, Christiansburg, Coeburn Norton Wise, HL Mooney, Little Falls Run, Lynchburg, Moores Creek, Nansemond, North River, Onancock, Parkins Mill, Pepper’s Ferry, Pound, Richmond, Roanoke, Rutledge Creek, Upper Occoquan, Virginia Initiative Plant, Williamsburg, Wolf Creek. The only selected SM site without a sewershed boundary or linked case counts was Danville.

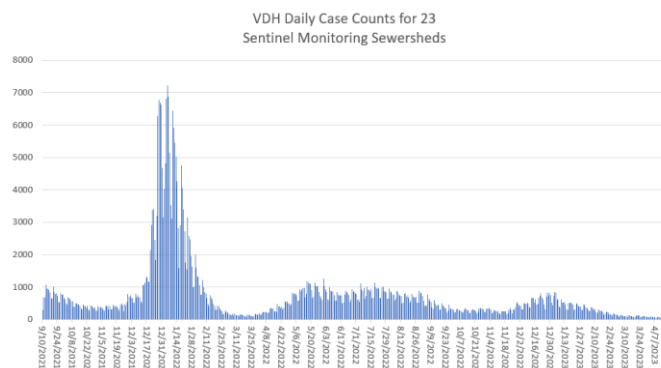


Figure 7.3.4: *Daily VDH SMP case counts over time, column graph*

When examining the above daily case count graph, it is important to know a key contextual fact – at-home COVID-19 tests became readily available to Virginians starting in the Spring of 2022. Due to this, the number of clinical (reported) COVID-19 infections probably

started to become artificially suppressed, as Virginians no longer needed to visit a clinical (reporting) facility to test themselves for COVID-19. In addition to this, case definitions at VDH changed on January 1, 2023 to no longer include positive at-home antigen tests as confirmatory. The result is that VDH case data is less reliable as a community indicator after Spring 2022.

Variant and sewershed cases over time are shown below in Figure 7.3.5.

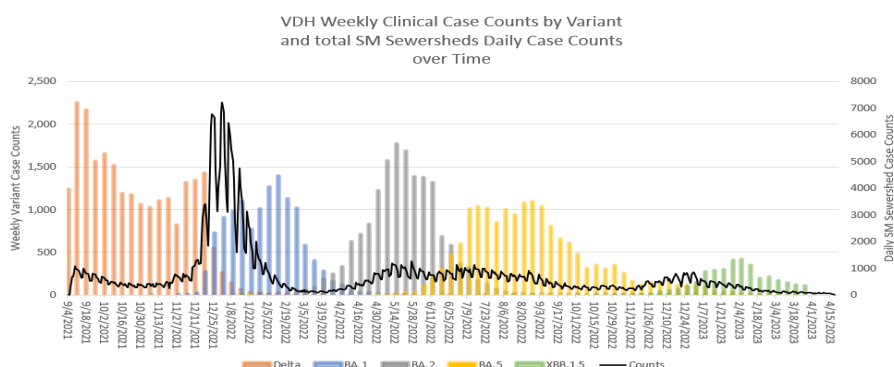


Figure 7.3.5: *VDH weekly variant clinical case counts and daily VDH SM sewershed case counts over time*

It is interesting to note that, for the weeks where daily case counts were comparatively high, also marked periods where variants transitioned. Summed weekly case counts were preferred over daily counts, as this nullified weekend dip effects.

7.4 WWS Analysis Data

WWTP Information

The 24 SMP sites (shortened names) were assigned numbers for visualization convenience. The sites' affiliated VDH health region, sewershed population, WWTP plant size (in millions of gallons per day), and WWTP zip code are listed in Table 7.4.1 below. The utility-reported populations were color coded by most populated (green) to least populated (red). The average number of people for all utilities was 108,343. The cumulative total was 2,600,241

people, approximately 30.0% of Virginia's 2022 estimated population of 8,683,619 people (Bureau).

Table 7.4.1: SMP Site Information including #, Name, Population, MGD, and ZIP

#	SITE	VDH Health Locality	Population	mgd	zip
1	AlexRenew	Alexandria	300,000	54	22314
2	Aquia	Stafford	92,000	10	22554
3	Basham	Loudoun	10,178	1.5	20132
4	Blacksburg	Montgomery	52,500	9	24060
5	Danville	Danville	47,000	20	24540
6	Coeburn	Wise	14,000	6.5	24230
7	HLM	Prince William	165,901	24	22191
8	NA	Suffolk	197,608	30	23435
9	ON	Accomack	1,263	0.3	23417
10	VIP	Norfolk	192,347	40	23508
11	WB	Williamsburg	69,059	22.5	23185
12	LFR	Fredericksburg	44,000	8	22405
13	LB	Lynchburg	220,000	22	24504
14	Moores	Charlottesville	118,266	15	22902
15	NorthRiver	Harrisonburg	81,000	23	22841
16	Parkins	Winchester	16,563	5	22602
17	Peppers	Montgomery	70,000	9	24141
18	Pound	Wise	1,090	0.5	24279
19	Richmond	Richmond City	225,000	75	23224
20	Roanoke	Botetourt	300,000	55	24014
21	Rutledge	Amherst	2,166	0.6	24521
22	CB	Montgomery	23,000	6	24073
23	UOSA	Fairfax	350,000	54	20121
24	Wolf	Washington	7,300	5.0	24211
AVG			108,343	20.7	

It is of note that three SMP sites are below the population threshold for NWSS (3,000 person minimum). When this data is submitted to CDC DCIPHER, their data is automatically suppressed. However, VDH WWSP has discretion to sample and post WWS data of any population size, electing to continue testing and publicly posting a few WWTPs in the 1,000-3,000 range.

Sample Information

Much of the sampling and analytical methodology was controlled for, based upon needs of DCLS, VDH WWSP, and CDC DCIPHER. However, some elements of sample collection did differ between sites. The number of collected samples, the start and ends dates, related quality control flags, sample day, lab turnaround time, and the sample matrix type were not always consistent between – and even within – each site, as seen below in Table 7.4.2. The highlighted red fields represent the more divergent sites.

Table 7.4.2: SMP site collection and metadata statistics

#	Samples		Collection Dates		Q.Flags		Sample Day		Lab TAT		Sample Matrix	
	#	Missed	Start	End	#	%	Day of Week	SD	AVG Days	SD	24-hr %	Grab %
1	84	0	2021-09-13	2023-04-17	3	4%	2.0	0.1	6.0	4.4	100%	0%
2	83	1	2021-09-13	2023-04-17	4	5%	2.0	0.1	5.6	4.2	100%	0%
3	44	40	2021-09-13	2022-07-18	4	9%	2.0	0.0	5.9	2.7	100%	0%
4	84	0	2021-09-13	2023-04-17	4	5%	2.0	0.1	5.9	4.3	100%	0%
5	46	38	2021-09-13	2022-07-25	3	7%	2.0	0.0	5.3	2.4	100%	0%
6	84	0	2021-09-13	2023-04-17	4	5%	2.0	0.2	5.7	4.3	100%	0%
7	84	0	2021-09-13	2023-04-17	3	4%	2.0	0.0	5.6	4.2	100%	0%
8	81	3	2021-09-14	2023-04-17	2	2%	2.3	0.6	5.5	4.2	0%	100%
9	80	4	2021-09-14	2023-04-21	1	1%	3.0	0.2	5.2	4.5	0%	100%
10	80	4	2021-09-13	2023-04-17	2	3%	2.2	0.5	5.6	4.2	6%	94%
11	80	4	2021-09-14	2023-04-17	3	4%	2.3	0.6	5.5	4.3	0%	100%
12	82	2	2021-09-13	2023-04-17	4	5%	2.0	0.2	5.6	4.2	100%	0%
13	83	1	2021-09-12	2023-04-16	4	5%	1.1	0.4	6.5	4.1	100%	0%
14	84	0	2021-09-13	2023-04-17	3	4%	2.0	0.0	5.7	4.2	100%	0%
15	83	1	2021-09-13	2023-04-17	4	5%	2.0	0.0	5.8	4.3	100%	0%
16	82	2	2021-09-13	2023-04-17	3	4%	2.0	0.2	5.6	4.2	100%	0%
17	84	0	2021-09-13	2023-04-17	4	5%	2.1	0.3	5.6	4.2	93%	7%
18	81	3	2021-09-14	2023-04-18	4	5%	3.1	0.3	4.9	4.2	0%	100%
19	82	2	2021-09-13	2023-04-17	4	5%	2.1	0.3	5.5	4.4	93%	7%
20	84	0	2021-09-13	2023-04-17	3	4%	2.0	0.1	5.8	4.4	100%	0%
21	45	39	2021-09-14	2022-07-26	3	7%	3.0	0.0	4.8	2.8	0%	100%
22	44	40	2021-09-13	2022-07-25	4	9%	2.0	0.0	5.3	2.3	100%	0%
23	82	2	2021-09-13	2023-04-17	2	2%	2.0	0.1	5.6	4.2	100%	0%
24	84	0	2021-09-13	2023-04-17	4	5%	2.0	0.0	5.6	4.2	100%	0%
AVG	76.3	7.8			3.3	5%	2.1	0.2	5.6	4.0	75%	25%

Sample Numbers & Dates:

While all SMP sites started sampling at the same time, not all sites stayed in the program until April 18th, 2023 (the request date of this dataset, Week #84). The Samples ‘#’ refers to the total number of samples taken, while the ‘Missed’ refers to the total number of samples not collected in those 84 weeks. The four dark red ‘Missed’ sites withdrew early, as confirmed by their ‘Collection Dates’ columns to their right.

The other sites’ missed numbers were due to a myriad of reasons, including utility failures to collect (e.g., holiday or staffing issues), courier failures, or DCLS laboratory failures. Considering the lengthy collection duration of 84 weeks total, it is relatively impressive that – of the 20 sites which stayed in the program – each site missed, at most, only 4 samples. The site average overall was 76 samples with 7.8 missed samples.

Note: most case-paired analyses in this dissertation used the cutoff point of SMP Wk53. This was due to several reasons, including that this is when many of the SMP sites started testing twice-weekly instead of once-weekly, and around the time that several others dropped out of the program. Additionally, case data lags for up to 6 weeks, which meant I’d have to cut off a good slice of WWS data anyway. It seemed like a clean breaking point for cleaner analysis.

Quality Flags:

Occasionally there were samples that DCLS marked with a quality flag. This was to indicate that the laboratory results may not be reflective of an intact sample. This could be due to issues with sample collection, shipping, or analysis. The average site had 3.3 quality flags, representing about 5% of their total samples.

Sampling Day:

Depending on several factors, including whether the sample matrix type was a grab sample (instantaneous collection) or a 24-hr composite collection (table date determined by the start date of the 24-hr process), the sampling 'Day of Week' may be slightly different within and between sites. The average start date for the sites was 2.1, representing Monday, with a standard deviation averaging 0.2 days.

Lab Turn-Around-Time:

The time between the sample collection and DCLS performing their analysis on the sample also varied. This is usually due to delays in either sample collection or laboratory analysis. The average turn-around-time (TAT) was 5.6 days, though the standard deviation was 4.0 days.

Of note: the 4 sites with lowest TAT were also the ones who cut early from the SMP, indicating a probable confounder.

Sampling Matrix:

On average, 75% of the sites used 24-hour composite sampling, as compared to the 25% using grab sampling. Most of the time the sites stayed consistent with their preferred method, though 3 sites switched around 6-7% of the time.

Flow & Analysis Information

The WWTP flow, reported in millions of gallons of sewage per day (mgd), for the sample collection day is reported to DCLS by the utility. However, not all utility sewer networks are created equal. Some sewage systems are subject to more I&I, responding with increased flow

during heavy rain periods. Some may have commercial sewage dumps periodically. The flow may therefore change radically between samplings and between sites.

Flow indirectly influences sample Viral Concentration and directly affects Viral Load calculations. As seen below in Table 4.4.3, the colored categories of Flow, Viral Concentration, and Viral Load indicate which sites have a higher (red) or lower (green) variance, indicated by the Standard Deviation (SD) as a percentage of the total site average.

The Less Than Limit of Detection (LOD) columns indicate the site numbers and percentages of samples that did not meet minimum viral concentration quantities needed for analysis. When this happens, the sample viral concentration quantities are not high enough to be reliably interpreted. The higher values are indicated by the red color gradient. While this is good for the sewershed communities having a reduced viral load burden, it is bad for reliable analysis interpretation.

Of note: confounding may occur with site #s 3, 5, 21, and 22, as they spent a longer relative duration in the Omicron case explosion beginning December 2021, due to leaving the SMP early.

Table 7.4.3: SMP site analytic statistics and flow metadata; lower (green) versus higher (red) variance

Flow				Viral Concentration			< LOD		Viral Load		
#	AVG Rate	SD	SD/AVG %	AVG	SD	SD/AVG %	#	%	AVG	SD	SD/AVG %
1	35.6	3.4	10%	6.7E+04	7.7E+04	116%	3	4%	8.9E+12	1.0E+13	115%
2	5.6	0.7	13%	8.9E+04	1.1E+05	122%	4	5%	1.9E+12	2.5E+12	130%
3	0.6	0.1	17%	6.4E+04	7.1E+04	111%	7	16%	1.4E+11	1.5E+11	107%
4	6.8	1.2	18%	4.5E+04	4.1E+04	90%	8	10%	1.2E+12	1.1E+12	98%
5	6.1	1.4	24%	1.5E+04	3.2E+04	216%	23	50%	3.1E+11	6.0E+11	196%
6	3.6	1.9	54%	1.8E+04	1.9E+04	103%	28	33%	2.1E+11	2.0E+11	95%
7	15.2	1.5	10%	9.0E+04	1.2E+05	134%	6	7%	5.2E+12	6.9E+12	134%
8	16.0	3.0	19%	5.7E+04	1.4E+05	242%	7	9%	3.5E+12	8.0E+12	229%
9	0.2	0.1	60%	7.1E+04	2.5E+05	345%	33	41%	7.1E+10	2.4E+11	345%
10	25.0	4.9	20%	3.9E+04	7.9E+04	203%	12	15%	3.5E+12	6.5E+12	187%
11	10.6	2.0	19%	4.2E+04	5.8E+04	140%	11	14%	1.7E+12	2.4E+12	143%

12	3.7	0.6	15%	7.0E+04	8.2E+04	117%	3	4%	9.7E+11	1.1E+12	112%
13	13.5	3.0	22%	3.4E+04	4.1E+04	120%	17	20%	1.8E+12	2.3E+12	129%
14	10.3	1.1	11%	4.3E+04	4.9E+04	114%	3	4%	1.6E+12	1.8E+12	110%
15	14.1	2.0	14%	1.7E+04	2.6E+04	148%	32	39%	8.8E+11	1.2E+12	140%
16	3.3	1.0	31%	2.6E+04	4.4E+04	166%	27	33%	3.3E+11	5.8E+11	173%
17	3.6	1.3	36%	3.8E+04	4.2E+04	111%	6	7%	4.9E+11	5.4E+11	111%
18	1.2	0.7	53%	4.9E+04	1.3E+05	261%	35	43%	1.8E+11	5.8E+11	316%
19	56.8	16.0	28%	2.1E+04	3.0E+04	143%	24	29%	4.7E+12	6.6E+12	140%
20	41.8	12.3	29%	3.1E+04	3.4E+04	110%	15	18%	4.7E+12	5.7E+12	121%
21	0.2	0.0	22%	2.4E+04	4.5E+04	187%	21	47%	1.9E+10	3.5E+10	179%
22	2.6	0.7	27%	6.1E+04	5.5E+04	90%	6	14%	6.0E+11	5.4E+11	91%
23	32.9	4.4	13%	6.9E+04	1.0E+05	149%	3	4%	8.9E+12	1.4E+13	157%
24	2.1	0.6	31%	6.8E+04	6.9E+04	101%	6	7%	4.9E+11	4.6E+11	94%
AVG	13.0	2.7	25%	4.8E+04	7.2E+04	152%	14.2	20%	2.2E+12	3.1E+12	152%

Flow:

The sites had very different reported flows, with averages ranging between 0.2 mgd to 56.8 mgd. This typically corresponds to service area population differences; the 0.2 mgd site serviced 2,166 people, while the 56.8 mgd site serviced 225,000 people. Variance calculated as the SD divided by the average flow indicated that the most variable sites were the ones with the smallest flow (i.e., the smallest service populations). However, this trend was not true in both directions, as there were several small flow sites with mid- (or even low-) tier variance. For example, Site #2 had an average of 5.6 mgd, but their SD was only 0.7 (13%).

Viral Concentration:

Oddly, all sites reported average viral concentrations at the 10^4 factor level. However, the calculated variances (SD divided by site average) were widely distributed, with some sites experiencing as little as 90% variance (green) to as much as 345% variance (red).

Limit of Detection:

The percentage of site samples that were below the LOD ranged from 4% (white) to 50% (red). While many of the reddest sites corresponded to low service area populations, this was not always the case.

Viral Load:

Sites reported average viral loads in the 10^{10} to the 10^{14} factor levels. However, the calculated variances (SD divided by site average) were widely distributed, with some sites experiencing as little as 91% variance (green) to as much as 345% variance (red).

Identifying High & Low Variability Sites

Ranks were applied to the flow, viral concentration, and viral load variance categories, with the smallest rank indicating the highest variability (red), as seen below in Table 7.4.4. The three scores for each site were then averaged, for an AVG RANK SCORE. The highest variability sites had a lower score and red shading. The site numbers, population (also color-coded), and sample metrics were included on the left of the table to give context.

Table 7.4.4: SMP site variance statistics and rankings; lower (green) versus higher (red) variance

			Samples		Flow Variance		Viral Conc Variance		Viral Load Variance			
#	Population	mgd	#	Misssed	SD/AVG %	RANK	SD/AVG %	RANK	SD/AVG %	RANK		AVG RANK SCORE
1	300,000	54	84	0	10%	24	116%	16	115%	16		18.7
2	92,000	10	83	1	13%	20	122%	13	130%	13		15.3
3	10,178	1.5	44	40	17%	17	111%	18	107%	20		18.3
4	52,500	9	84	0	18%	16	90%	23	98%	21		20.0
5	47,000	20	46	38	24%	10	216%	4	196%	4		6.0
6	14,000	6.5	84	0	54%	2	103%	21	95%	22		15.0
7	165,901	24	84	0	10%	23	134%	12	134%	12		15.7
8	197,608	30	81	3	19%	15	242%	3	229%	3		7.0
9	1,263	0.3	80	4	60%	1	345%	1	345%	1		1.0

10	192,347	40	80	4	20%	13	203%	5	187%	5	7.7
11	69,059	22.5	80	4	19%	14	140%	11	143%	9	11.3
12	44,000	8	82	2	15%	18	117%	15	112%	17	16.7
13	220,000	22	83	1	22%	11	120%	14	129%	14	13.0
14	118,266	15	84	0	11%	22	114%	17	110%	19	19.3
15	81,000	23	83	1	14%	19	148%	9	140%	10	12.7
16	16,563	5	82	2	31%	6	166%	7	173%	7	6.7
17	70,000	9	84	0	36%	4	111%	19	111%	18	13.7
18	1,090	0.5	81	3	53%	3	261%	2	316%	2	2.3
19	225,000	75	82	2	28%	8	143%	10	140%	11	9.7
20	300,000	55	84	0	29%	7	110%	20	121%	15	14.0
21	2,166	0.6	45	39	22%	12	187%	6	179%	6	8.0
22	23,000	6	44	40	27%	9	90%	24	91%	24	19.0
23	350,000	54	82	2	13%	21	149%	8	157%	8	12.3
24	7,300	5.0	84	0	31%	5	101%	22	94%	23	16.7
AVG	108,343	20.7	76.3	7.8	25%	152%		152%			

The five highest variability ranked sites are seen below in Table 7.4.5.

Table 7.4.5: SMP 5 most variable sites; lower (green) versus higher (red) variance

			Samples		AVG RANK SCORE
#	Pop	mgd	#	Missed	
5	47,000	20	46	38	6.0
8	197,608	30	81	3	7.0
9	1,263	0.3	80	4	1.0
16	16,563	5	82	2	6.7
18	1,090	0.5	81	3	2.3
SMP AVG	108,343	20.7	76.3	7.8	

Most sites in the top 5 highest-variability ranks had relatively low populations, low capacity mgd WWTPs, and low missed numbers of samples. Though there were some exceptions.

The five lowest variability ranked sites are seen below in Table 7.4.6.

Table 7.4.6: SMP 5 least variable sites; lower (green) versus higher (red) variance

			Samples		AVG RANK SCORE
#	Pop	mgd	#	Missed	
1	300,000	54	84	0	18.7
3	10,178	1.5	44	40	18.3
4	52,500	9	84	0	20.0
14	118,266	15	84	0	19.3
22	23,000	6	44	40	19.0
SMP AVG	108,343	20.7	76.3	7.8	

Ironically, most sites in the top 5 lowest variability ranks also had relatively low populations, low capacity mgd WWTPs, and low missed numbers of samples. Though there were some exceptions.

These results suggest that the site variability scores are not strongly tied to population service area size. Each site should be examined individually for variance.

Viral Loads & Cases Over Time

The VDH WWS SM dataset has calculated viral loads for each sample analyzed. The viral loads can therefore be graphed over time within each site. This by-site exploration, along with cases by-site over time, will be explored more in the next chapter.

However, for context when viewing the next chapter, the aggregate of all viral loads over time is presented below in Figure 7.4.1. It represents, in general, how much the overall viral load changes across the collection period. It does not account for the change in status of the 4 sites which dropped from the program about halfway through. Since those sites were relatively small, a minimum drop in Aggregate Viral Loads should be expected at around the half-way point, through the end of the collection period. Visualization also does not account for missed samples, or distinctions/reassignments for viral loads that were below the limit of detection.

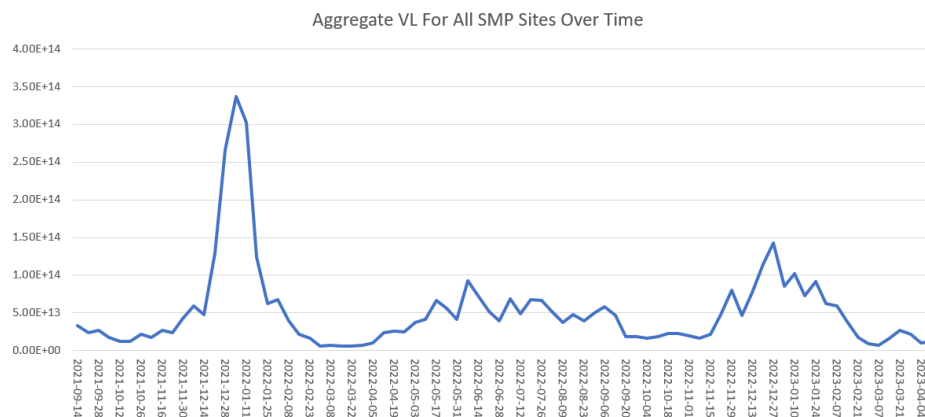


Figure 7.4.1: *Aggregate viral loads for all SMP sites over time*

When compared to the VDH WWS SMP aggregate case count data for the same time frame, with the appropriate dropping of the 4 early-withdraw sites around the halfway point, the similarities are striking. To be explored further in the next chapter.

The Case Count graph below in Figure 7.4.2 views cumulative daily cases counts for all sites, excluding Danville; without a Danville service region shapefile, spatially linked case counts could not be generated. When viewing this cumulative representation of cases, keep in mind that there is some very minor case count blurring at the start and end of each site's participation periods. This is due to the continuation of collecting case count data for anywhere from a few days before the first collection date, to 1-5 weeks past the end of the last viral load collection date. Often this is linked merely to the availability of count data, as well as the dissertation choice to include several additional weeks of counts to examine if there were leading indicator effects (see Ch.7 & Ch.8 for day shifts correlations measured a week before to 3 weeks after). While most sites generally tended to match the same dates for case count generation, the 3 early withdraw sites do represent a period, in August of 2022, where some case counts continued to be counted for sites which had already stopped generating viral loads. However, those 3 sites

represented very low populations and case counts generally, and most likely had only very minor effects on the aggregate case count visualizations.

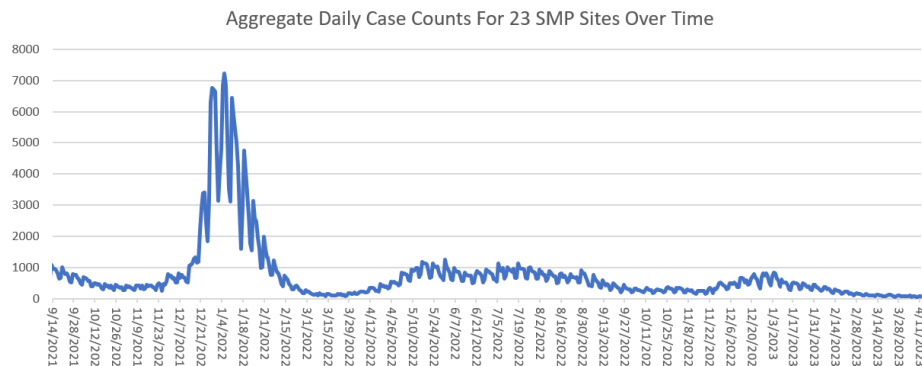


Figure 7.4.2: *Aggregate daily cases for all SMP sites over time*

When the aggregated cases and viral loads are placed together as seen below in Figure 7.4.3, the visual similarity is uncanny. It is particularly interesting to examine the time when state take-home testing kits became commonplace (Spring/Summer 2022). Between the unreported take-home tests -- and the redefinition of cases to not include take-home tests in January 2023 -- one might guess that cases started becoming under-reported around this time. If this is the case, it makes sense why the aggregated case counts visually started to drift below the expected viral load data.

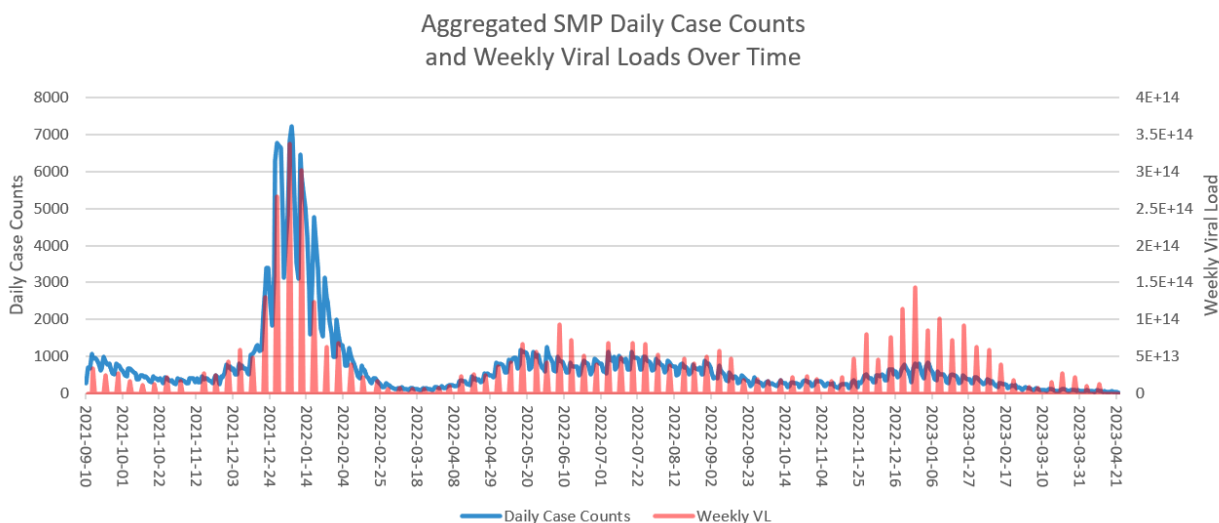


Figure 7.4.3: *Aggregated SMP daily case counts and weekly viral loads over time*

Chapter 8: Case Correlation

Other studies have examined correlation between WWS viral loads and community infection metrics (Hopkins et al., 2023). However, VDH WWS has two unique advantages when performing this research:

1. Only one lab is used, so no aggregation or adjustments are necessary to correct lab methodological or analytic differences.
2. The SMP represents a fair number of utility partners (25 to start), with a lengthy program duration covering several variant peaks over time (Sept 2021 to date).

The SMP dataset includes spatially linked case count data for 24 SMP sites, as described in Chapter 5. The SMP case count data was selected to compare against SMP viral load data, as compared to the VDH Public Use case dataset, due to its very specific spatially linked relationship with the SMP sites.

Note about “weekly” viral loads: while all viral loads (VL) are generated from once-weekly WWS sampling (see Ch. 5) in this dissertation, it is important to note that the VLs taken at that time represent a snapshot community view in that moment for grab samples, and a 24-hr view for 24-hr composite samples. So, while the VLs are from samples collected weekly, they do not represent sampling for an entire week (i.e., sampling daily for 7 days straight). Such daily testing is outside the scope of the VDH WWSP, as well as this dissertation.

8.1 VL & Daily Cases (general)

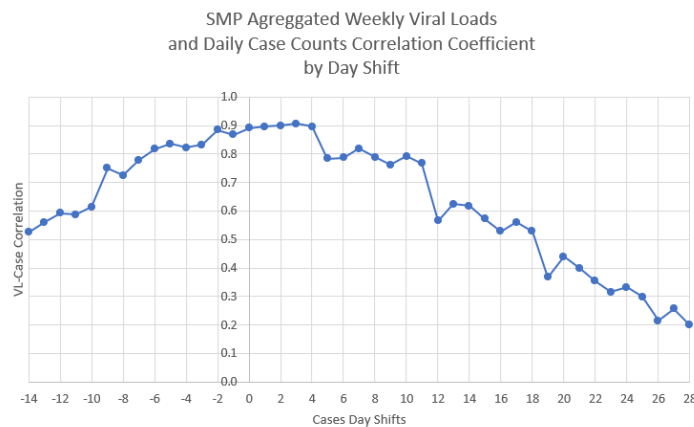
In Chapter 7, the aggregated SMP daily case counts and viral load data was visually explored and shown in Figure 7.4.3. The aggregates’ visual similarities are striking, especially when considering the possibility of case suppression becoming a factor about halfway through the dataset.

Using the Excel CORREL function (Microsoft, 2023), Pearson correlation was explored between the viral loads and daily cases (-1 to +1 coefficient). Additionally, the daily cases were shifted between -14 days and +28 days, to explore if there were leading indicator effects.

As seen below in Table 8.1.1 and Figure 8.1.2, the correlation between the SMP aggregated viral loads (n=82) and daily case counts was 0.890 at zero days shift (p<0.001), indicating a very strong relationship between aggregate VLs and Daily Cases. Shifting the days demonstrated a very strong correlation of approximately 0.9 between days -2 to +4.

Table 8.1.1: *Selected peak period for SM aggregated viral loads and daily cases correlation with day shifts*

Day Shift	Correlation Coefficient
-3	0.832
-2	0.885
-1	0.868
0	0.890
1	0.896
2	0.899
3	0.906
4	0.897
5	0.784



*n=82 (weekly aggregated samples); all p values < .05, except for +26 and +28 day shift

Figure 8.1.2: *Aggregated viral loads and daily cases correlation coefficients with day shifts, line graph*

8.2 VL & Weekly Cases (general)

While using viral loads and daily cases correlation gives us a snapshot moment of viral load and daily cases, it may be more practical to think of case counts in weekly terms. Then one (e.g., a local district epidemiologist) may view a once-weekly viral load value and ask how related the next weeks' worth of cases will be. However, it should still be noted that a weekly viral load is either a grab or 24-hr composite sample, taken once a week. It is itself a snapshot

analysis of a community and does not represent an entire week of testing. However, since viral load analysis simply isn't currently available consistently throughout the week, a once-weekly viral load will have to suffice.

Aggregated SMP viral loads and aggregated SMP case counts – as defined as the following 7 days' worth of cases – was explored visually below in Figure 8.2.1.

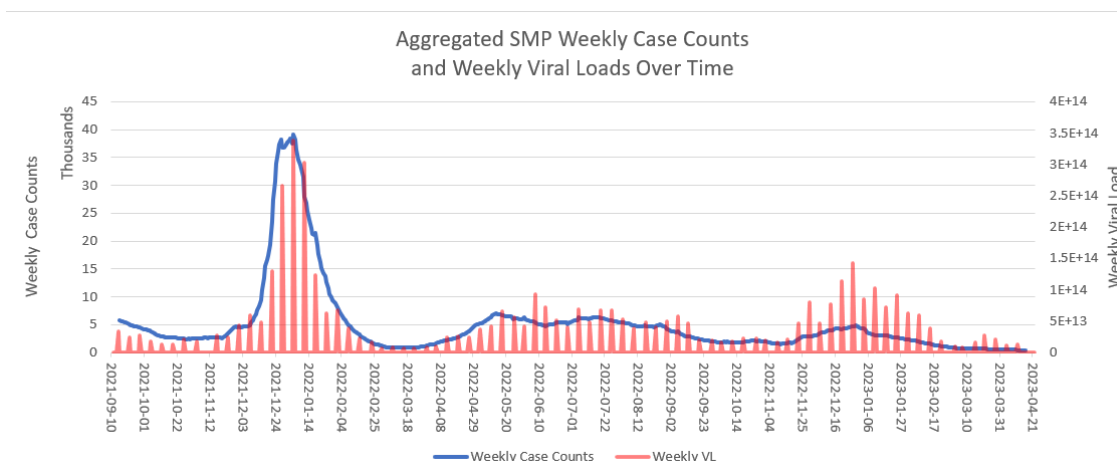
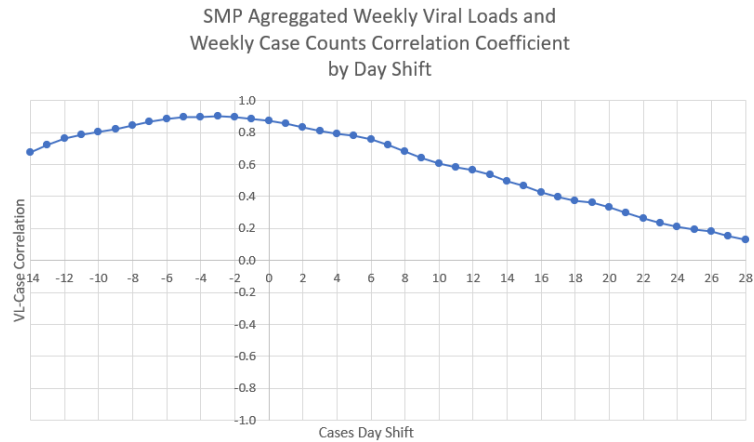


Figure 8.2.1: *Aggregated SMP viral loads and weekly (next 7 days') case counts over time*

For correlation analysis, the correlation coefficient was calculated for all sampling weeks where there was both a viral load and cumulative case count value (n=82). Then a day shift between -14 and +28 was applied, changing the week's cases timeframe start date from the next day to anywhere between -14 days to 28 days later, as seen below in Table 8.2.1 and Figure 8.2.2.

Table 8.2.1: *Selected peak period for SM aggregated viral loads and weekly cases correlation and day shifts*

Day Shift	Correlation Coefficient
-7	0.870
-6	0.885
-5	0.894
-4	0.898
-3	0.900
-2	0.896
-1	0.887
0	0.875
1	0.856
2	0.830



*n=82 (weekly aggregated samples); all p values < .05, except for +24 to +28 day shifts

Figure 8.2.2: *SM Aggregated viral loads and weekly cases correlation coefficients with day shifts, line graph*

The findings above indicate that a viral load, analyzed and interpreted the same day, is strongly correlated to the following 7 days' worth of cases (with a correlation coefficient of 0.875, $p < .001$). Furthermore, a similar (or higher) correlation coefficient is noted for the week preceding the viral load sample.

One could interpret this result to mean that there is a strong relationship between viral loads and weekly case counts, especially in the week's cases beginning a week before, to the week of, the viral load sample. For example, if a viral load sample were taken on the 15th of the month, there is a high correlation to the daily cases ranging in the days from the 9th to the 22nd. This makes sense, as one would assume a direct relationship between the number of community members with COVID-19 and the amount of SARS-CoV-2 viral particles shed into the wastewater.

8.3 VL & Daily Cases (by variant)

Several factors could be contributory to a dynamic relationship between VL and cases over time, to include:

- Changing variants with different infectious physiological effects
- Changing susceptibility within the host population (e.g., vaccination/exposure statuses, deaths)
- Case data collection changes (e.g., case definitions, take-home tests)
- WWS methodological changes (e.g., collection, courier, laboratory, analytic changes)

For this reason, the SMP aggregate viral loads and daily cases were broken up and analyzed by variant timelines, as determined in Chapter 6's Synthesizing Variant Dominance Timeline.

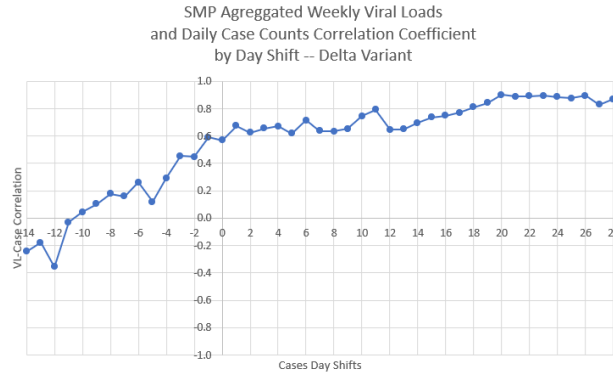
It is important to note here that daily case data does not smooth out natural variations found in case reporting. For example, case reporting on the weekends may drop below work week levels. Using a cumulative weekly summation, as seen in sections 7.2, 7.4, and 7.6, will by default smooth out work week fluctuations.

For each variant, the entire correlation coefficients for the -14 to 28 day shift period were graphed out below in Figures 8.3.1 to 8.3.5, and a relevant viewing subsection table placed adjacently in Tables 8.3.1 to 8.3.5.

Delta (9/4/21 – 12/18/21)

Table 8.3.1: Selected correlation period for SM aggregated viral loads for Delta variant and daily cases correlation coefficients with day shifts

Day Shift	Correlation Coefficient
-1	0.591
0	0.567
1	0.675
2	0.624
3	0.657
4	0.672
5	0.620
6	0.716



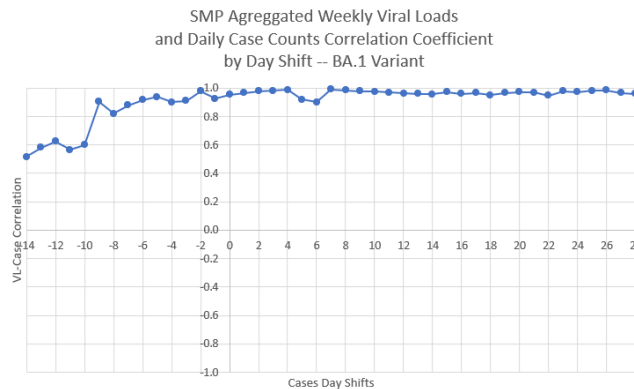
*n=13 (weekly aggregated samples); all p values < .05, except for -2 to -14 day shifts

Figure 8.3.1: SM Aggregated viral loads for Delta variant and daily cases correlation coefficients with day shifts

Omicron BA.1 (1/1/22 – 3/12/22)

Table 8.3.2: Selected correlation period for SM aggregated viral loads for BA.1 variant and daily cases correlation coefficients with day shifts

Day Shift	Correlation Coefficient
-2	0.979
-1	0.927
0	0.953
1	0.966
2	0.978
3	0.983
4	0.987
5	0.920



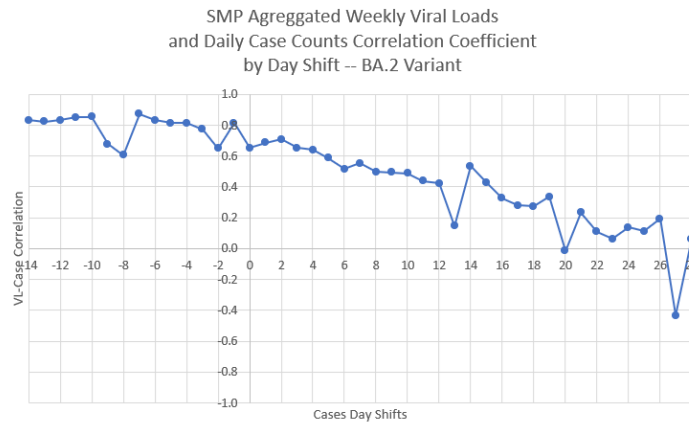
*n=10 (weekly aggregated samples); all p values < .05, except for -10 to -14 day shifts

Figure 8.3.2: SM Aggregated viral loads for BA.1 variant and daily cases correlation coefficients with day shifts

Omicron BA.2 (4/2/22 – 6/18/22)

Table 8.3.3: Selected correlation period for SM aggregated viral loads for BA.2 variant and daily cases correlation coefficients with day shifts

Day Shift	Correlation Coefficient
-4	0.814
-3	0.773
-2	0.651
-1	0.814
0	0.653
1	0.687
2	0.709
3	0.653
4	0.639



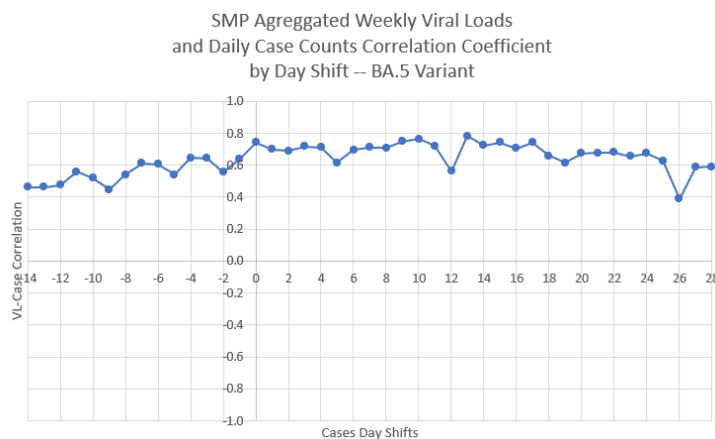
*n=11 (weekly aggregated samples); all p values < .05, except for +5 to +28 day shifts

Figure 8.3.3: SM Aggregated viral loads for BA.2 variant and daily cases correlation coefficients with day shifts

Omicron BA.5 (7/16/22 – 12/10/22)

Table 8.3.4: Selected correlation period for SM aggregated viral loads for BA.5 variant and daily cases correlation coefficients with day shifts

Day Shift	Correlation Coefficient
-2	0.557
-1	0.641
0	0.742
1	0.699
2	0.690
3	0.717
4	0.712
5	0.614
6	0.696



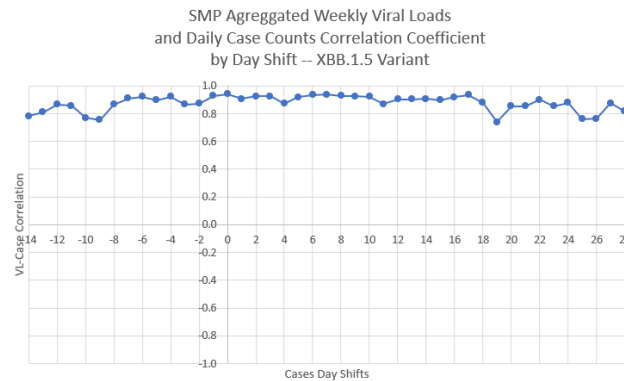
*n=21 (weekly aggregated samples); all p values < .05, except for +26 day shift

Figure 8.3.4: SM Aggregated viral loads for BA.5 variant and daily cases correlation coefficients with day shifts

Omicron XBB.1.5 (1/14/23 – 4/1/23)

Table 8.3.5: Selected correlation period for SM aggregated viral loads for XBB.1.5 variant and daily cases correlation coefficients with day shifts

Day Shift	Correlation Coefficient
-2	0.873
-1	0.930
0	0.940
1	0.907
2	0.926
3	0.925
4	0.874



n=11 (weekly aggregated samples); all p values < .05

Figure 8.3.5: SM Aggregated viral loads for XBB.1.5 variant and daily cases correlation coefficients with day shifts

All Variants Summarized

Daily cases reached their peak correlation with viral loads during the -2 to +4 day shift period, and specifically on the +3 peak day (0.91 correlation; n=82, p<0.001), as seen below in Table 8.3.6. Additionally, each variant visually demonstrated different peak periods and peak days -- Delta (+20 to +26 days period; +26 peak day), BA.1 (+2 to +9 days period; +7 peak day), BA.2 (-7 to -1 days; -7 peak day), BA.5 (+9 to +17 days; +13 peak day), XBB.1.5 (-1 to +17 days; +0 peak day). Bolded values have p values under 0.05, while italicized values are above 0.05. Stronger correlation coefficients are in green, weaker ones in red.

Table 8.3.6: VL - Daily Cases Correlation by Variant

Daily Case Correlation by Variant						
Day Shift	All	Delta	BA.1	BA.2	BA.5	XBB.1.5
-14	0.524	<i>-0.243</i>	<i>0.516</i>	0.832	0.462	0.784
-13	0.561	<i>-0.180</i>	<i>0.582</i>	0.823	0.462	0.810
-12	0.592	<i>-0.355</i>	<i>0.625</i>	0.832	0.475	0.867
-11	0.588	<i>-0.029</i>	<i>0.566</i>	0.852	0.561	0.857
-10	0.614	<i>0.043</i>	<i>0.599</i>	0.855	0.518	0.769
-9	0.750	<i>0.105</i>	0.904	0.679	0.446	0.754
-8	0.726	<i>0.177</i>	0.821	0.606	0.542	0.867
-7	0.779	<i>0.159</i>	0.879	0.874	0.612	0.911
-6	0.817	<i>0.260</i>	0.915	0.832	0.606	0.922
-5	0.836	<i>0.118</i>	0.938	0.815	0.542	0.897
-4	0.823	<i>0.291</i>	0.902	0.814	0.647	0.921
-3	0.832	<i>0.453</i>	0.910	0.773	0.644	0.868
-2	0.885	<i>0.449</i>	0.979	0.651	0.557	0.873
-1	0.868	0.591	0.927	0.814	0.641	0.930
0	0.890	0.567	0.953	0.653	0.742	0.940
1	0.896	0.675	0.966	0.687	0.699	0.907
2	0.899	0.624	0.978	0.709	0.690	0.926
3	0.906	0.657	0.983	0.653	0.717	0.925
4	0.897	0.672	0.987	0.639	0.712	0.874
5	0.784	0.620	0.920	<i>0.586</i>	0.614	0.920
6	0.788	0.716	0.902	<i>0.515</i>	0.696	0.936
7	0.818	0.637	0.991	<i>0.553</i>	0.713	0.937
8	0.788	0.633	0.986	<i>0.499</i>	0.708	0.930
9	0.762	0.653	0.980	<i>0.493</i>	0.750	0.925
10	0.791	0.747	0.974	<i>0.489</i>	0.765	0.922
11	0.768	0.791	0.971	<i>0.438</i>	0.720	0.870
12	0.566	0.646	0.963	<i>0.424</i>	0.562	0.904
13	0.624	0.650	0.961	<i>0.146</i>	0.783	0.904
14	0.618	0.698	0.957	<i>0.533</i>	0.723	0.906
15	0.573	0.737	0.972	<i>0.428</i>	0.743	0.897
16	0.528	0.748	0.961	<i>0.327</i>	0.706	0.919
17	0.560	0.769	0.966	<i>0.282</i>	0.742	0.936
18	0.528	0.810	0.951	<i>0.275</i>	0.658	0.880
19	0.368	0.842	0.966	<i>0.336</i>	0.614	0.739
20	0.439	0.900	0.972	<i>-0.015</i>	0.675	0.854
21	0.399	0.890	0.968	<i>0.234</i>	0.678	0.854
22	0.354	0.891	0.949	<i>0.109</i>	0.680	0.902
23	0.315	0.895	0.977	<i>0.062</i>	0.656	0.855
24	<i>0.333</i>	0.885	0.971	<i>0.137</i>	0.673	0.880
25	<i>0.298</i>	0.876	0.982	<i>0.113</i>	0.629	0.762
26	<i>0.215</i>	0.894	0.983	<i>0.192</i>	<i>0.390</i>	0.763
27	<i>0.256</i>	0.830	0.966	<i>-0.434</i>	0.587	0.877
28	<i>0.200</i>	0.870	0.959	<i>0.062</i>	0.589	0.816
AVG	0.636	0.550	0.909	0.492	0.636	0.877

*Strong correlations are green, weaker ones are red. Bolded values have p<0.05, italicized values are p>0.05.

There could be many reasons why the VL-case correlation varied so readily between variants, including:

- Physiological/infectious differences between variants (e.g., changes in viral shedding, symptoms)
- Variants entering the population during different phases of infection waves (e.g., a variant shift during a case lull versus a holiday exposure peak)
- Case definitions and/or reporting changes
- Shifting population susceptibility
- Viral Load changes (e.g., warmer temperatures increasing degradation)

The lowest peak correlation by variant was still a moderately-strong 0.783 correlation coefficient (BA.5 at 13 days), giving good support to the relationship between VL and daily cases. However, the range of peak correlation day shifts by variant gives pause to relying on the immediate predictive powers of VL. For example, a Delta VL would be most predictive with cases nearly 3 weeks later, while a BA.2 VL would be most predictive with cases in the two weeks leading up to VL collection. Section 8.5 examines site-specific differences between these daily case day shifts.

Exploring the underlying causes for day shift correlational differences between the variants in the sites would be an interesting follow-up to this dissertation.

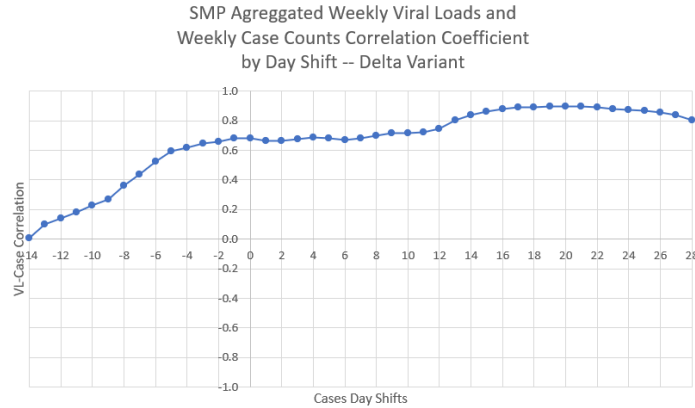
8.4 VL & Weekly Cases (by variant)

In the same way that section 7.2 reassessed daily cases for the sum of case counts for a week's time, this section will reassess 7.3's By-Variant correlation between VL and SMP aggregated daily case counts, as aggregated weekly case counts.

Delta (9/4/21 – 12/18/21)

Table 8.4.1: Selected correlation period for SM aggregated viral loads for Delta variant and weekly cases correlation coefficients with day shifts

Day Shift	Correlation Coefficient
-4	0.620
-3	0.645
-2	0.656
-1	0.680
0	0.682
1	0.667
2	0.662
3	0.674
4	0.688



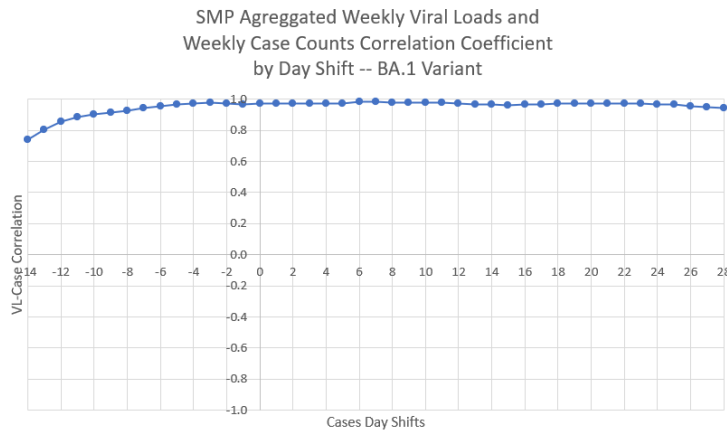
*n=13 (weekly aggregated samples); all p values < .05, except for -6 to -14 day shifts

Figure 8.4.1: SM Aggregated viral loads for Delta variant and weekly cases correlation coefficients with day shifts

Omicron BA.1 (1/1/22 – 3/12/22)

Table 8.4.2: Selected correlation period for SM aggregated viral loads for BA.1 variant and weekly cases correlation coefficients with day shifts

Day Shift	Correlation Coefficient
-4	0.973
-3	0.978
-2	0.975
-1	0.966
0	0.971
1	0.975
2	0.975
3	0.973
4	0.970



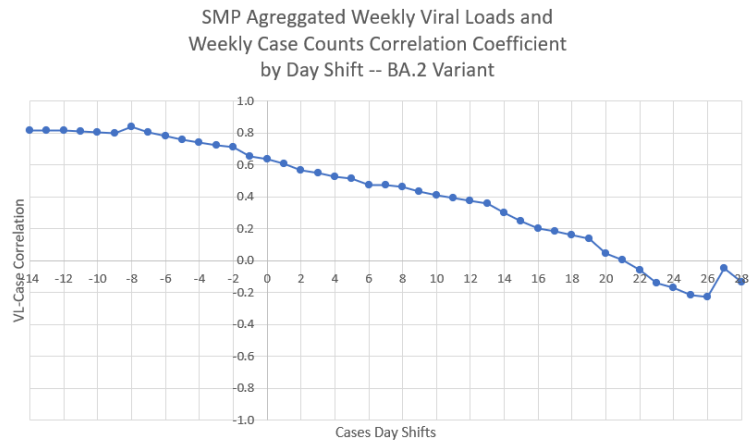
*n=10 (weekly aggregated samples); all p values < .05

Figure 8.4.2: SM Aggregated viral loads for BA.1 variant and weekly cases correlation coefficients with day shifts

Omicron BA.2 (4/2/22 – 6/18/22)

Table 8.4.3: Selected correlation period for SM aggregated viral loads for BA.2 variant and weekly cases correlation coefficients with day shifts

Day Shift	Correlation Coefficient
-8	0.839
-7	0.802
-6	0.778
-5	0.761
-4	0.738
-3	0.724
-2	0.711
-1	0.655
0	0.637
1	0.607



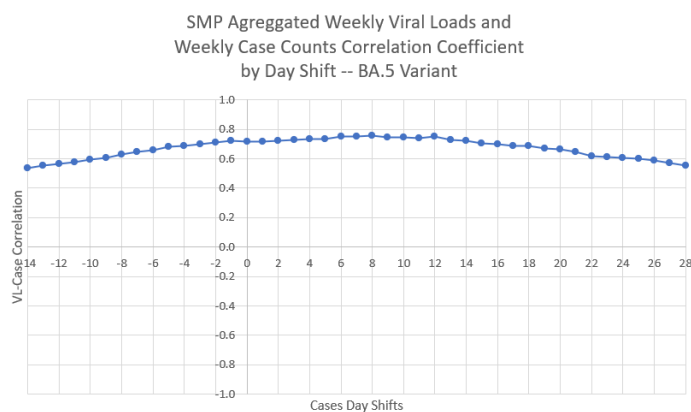
*n=11 (weekly aggregated samples); all p values < .05, except for +2 to +28 day shifts

Figure 8.4.3: SM Aggregated viral loads for BA.2 variant and weekly cases correlation coefficients with day shifts

Omicron BA.5 (7/16/22 – 12/10/22)

Table 8.4.4: Selected correlation period for SM aggregated viral loads for BA.5 variant and weekly cases correlation coefficients with day shifts

Day Shift	Correlation Coefficient
-2	0.708
-1	0.723
0	0.718
1	0.718
2	0.724
3	0.730
4	0.733



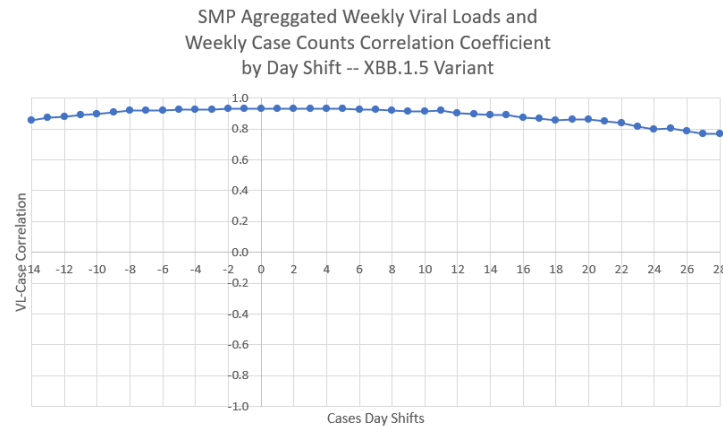
*n=21 (weekly aggregated samples); all p values < .05

Figure 8.4.4: SM Aggregated viral loads for BA.5 variant and weekly cases correlation coefficients with day shifts

Omicron XBB.1.5 (1/14/23 – 4/1/23)

Table 8.4.5: Selected correlation period for SM aggregated viral loads for XBB.1.5 variant and weekly cases correlation coefficients with day shifts

Day Shift	Correlation Coefficient
-2	0.930
-1	0.931
0	0.931
1	0.934
2	0.933
3	0.933
4	0.933
5	0.933
6	0.928



n=11 (weekly aggregated samples); all p values < .05

Figure 8.4.5: SM Aggregated viral loads for XBB.1.5 variant and weekly cases correlation coefficients with day shifts

All Variants Summarized

Weekly cases reached their peak correlation with viral loads during the -7 to +1 day shift period, and specifically on the -3 peak day (0.90 correlation; n=82, p<0.001), as seen below in Table 8.4.6. Additionally, each variant visually demonstrated different peak periods and peak days -- Delta (+16 to +23 days period; +19 peak day), BA.1 (+5 to +12 days period; +6 peak day), BA.2 (-14 to -5 days; -8 peak day), BA.5 (+6 to +12 days; +8 peak day), XBB.1.5 (-2 to +5 days; +1 peak day).

Weekly cases are defined as the VDH WWS SMP SARS-CoV-2 aggregated case count summation of the 7-day period, beginning on the noted shift day. Bolded values have p values under 0.05, while italicized values are above 0.05. Stronger correlation coefficients are in green, weaker ones in red.

Table 8.4.6 Weekly Case Correlation by Variant and day shift

Weekly Case Correlation Coefficient by Variant						
Day Shift	All	Delta	BA.1	BA.2	BA.5	XBB.1.5
-14	0.674	<i>0.008</i>	0.741	0.817	0.536	0.859
-13	0.722	<i>0.101</i>	0.804	0.818	0.557	0.875
-12	0.762	<i>0.139</i>	0.856	0.814	0.564	0.879
-11	0.787	<i>0.184</i>	0.887	0.812	0.579	0.890
-10	0.803	<i>0.228</i>	0.905	0.802	0.593	0.898
-9	0.820	<i>0.269</i>	0.916	0.799	0.607	0.908
-8	0.847	<i>0.361</i>	0.929	0.839	0.627	0.918
-7	0.870	<i>0.439</i>	0.943	0.802	0.647	0.922
-6	0.885	<i>0.522</i>	0.954	0.778	0.660	0.922
-5	0.894	0.596	0.964	0.761	0.680	0.927
-4	0.898	0.620	0.973	0.738	0.690	0.927
-3	0.900	0.645	0.978	0.724	0.700	0.929
-2	0.896	0.656	0.975	0.711	0.708	0.930
-1	0.887	0.680	0.966	0.655	0.723	0.931
0	0.875	0.682	0.971	0.637	0.718	0.931
1	0.856	0.667	0.975	0.607	0.718	0.934
2	0.830	0.662	0.975	<i>0.569</i>	0.724	0.933
3	0.810	0.674	0.973	<i>0.549</i>	0.730	0.933
4	0.795	0.688	0.970	<i>0.528</i>	0.733	0.933
5	0.778	0.684	0.975	<i>0.517</i>	0.732	0.933
6	0.755	0.671	0.986	<i>0.474</i>	0.749	0.928
7	0.720	0.682	0.982	<i>0.474</i>	0.751	0.924
8	0.681	0.700	0.980	<i>0.462</i>	0.756	0.919
9	0.639	0.714	0.978	<i>0.433</i>	0.746	0.917
10	0.605	0.719	0.978	<i>0.407</i>	0.744	0.917
11	0.581	0.724	0.977	<i>0.391</i>	0.741	0.917
12	0.564	0.747	0.974	<i>0.373</i>	0.752	0.904
13	0.536	0.803	0.965	<i>0.358</i>	0.729	0.897
14	0.497	0.840	0.967	<i>0.302</i>	0.720	0.889
15	0.464	0.864	0.964	<i>0.249</i>	0.706	0.888
16	0.427	0.882	0.966	<i>0.200</i>	0.698	0.876
17	0.396	0.889	0.967	<i>0.181</i>	0.689	0.866
18	0.374	0.891	0.970	<i>0.162</i>	0.689	0.856
19	0.363	0.900	0.972	<i>0.136</i>	0.670	0.860
20	0.331	0.896	0.971	<i>0.047</i>	0.662	0.862
21	0.296	0.896	0.970	<i>0.002</i>	0.648	0.853
22	0.263	0.892	0.973	<i>-0.060</i>	0.619	0.838
23	0.234	0.882	0.970	<i>-0.141</i>	0.612	0.815
24	<i>0.210</i>	0.874	0.968	<i>-0.172</i>	0.604	0.798
25	<i>0.194</i>	0.869	0.964	<i>-0.215</i>	0.600	0.802
26	<i>0.183</i>	0.854	0.957	<i>-0.228</i>	0.589	0.784
27	<i>0.155</i>	0.838	0.951	<i>-0.047</i>	0.572	0.768
28	<i>0.129</i>	0.806	0.945	<i>-0.135</i>	0.556	0.767
AVG	0.609	0.659	0.952	0.417	0.670	0.887

*Strong correlations are green, weaker ones are red. Bolded values have $p < 0.05$, italicized values are $p > 0.05$.

Overall, the weekly case correlations were very similar to the daily case correlations, with similar averages and general trends. Even the lowest peak correlation of the weekly cases (0.756 correlation of BA.5 at 8 days) mimicked that of the daily cases (0.783 correlation of BA.5 at 13 days). The main difference appears to be a general smoothing effect across the weekly cases.

8.5 Site-Specific VL & Daily Cases

Chapter 3 discussed inherent variability in any WWS system. VDH WWS offers a unique opportunity, in that the SMP:

- uses a single laboratory for all shipping and analysis (removing cross-laboratory variability)
- has been supported by CDC NWSS, financially and with good practice guidance
- has a wide network of participating utility partners, geographically covering much of Virginia
- has been collecting data with minimal changes to methodology/analysis for over 2 years
- has a top-notch Data Manager

While these factors are excellent for maintaining a proper dataset over time, they do not eliminate the core differences between the utility sites themselves. Each wastewater system is subject to different infrastructure, geography, population, and commercial factors which affect the time and degradation of viral particles across distance and time. It is for this reason that caution is heavily applied when discussing options for comparing sites.

As seen below in Table 8.5.1, each SMP site was analyzed by viral load and case counts, from the VDH WWS SMP datasets. See Appendix 8.5 By Site Viral Loads and Daily Cases. Note that four sites withdrew from the program early and are not included in the averaged sites columns. Bolded values have p values under 0.05, while italicized values are above 0.05. Stronger correlation coefficients are in green, weaker ones in red.

Table 8.5.1 Daily Cases Correlation Coefficient by Site and Day Shift

Daily Cases Correlation Coefficient by Site																									
Day Shift	AVG	1	2	3*	4	5*	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21*	22*	23	24
-14	0.36	0.46	0.55	0.57	0.35		0.38	0.50	0.25	-0.07	0.38	0.09	0.46	0.39	0.33	0.39	0.73	0.48	0.09	0.54	0.25	0.28	0.42	0.47	0.22
-13	0.40	0.44	0.56	0.43	0.37		0.42	0.54	0.26	0.10	0.41	0.11	0.47	0.36	0.37	0.58	0.73	0.65	0.00	0.59	0.22	0.05	0.44	0.51	0.29
-12	0.44	0.44	0.62	0.49	0.37		0.39	0.61	0.28	0.26	0.47	0.11	0.54	0.43	0.47	0.60	0.73	0.68	-0.04	0.64	0.26	0.22	0.53	0.57	0.29
-11	0.44	0.48	0.57	0.55	0.35		0.39	0.56	0.30	0.00	0.48	0.18	0.56	0.46	0.54	0.61	0.72	0.71	0.03	0.65	0.25	0.28	0.55	0.59	0.32
-10	0.45	0.42	0.58	0.35	0.46		0.33	0.64	0.31	-0.03	0.48	0.18	0.63	0.48	0.41	0.56	0.76	0.69	0.20	0.65	0.29	-0.10	0.52	0.55	0.33
-9	0.47	0.41	0.58	0.56	0.52		0.41	0.60	0.37	0.08	0.61	0.38	0.62	0.51	0.44	0.65	0.76	0.68	-0.02	0.70	0.27	0.36	0.61	0.58	0.31
-8	0.54	0.55	0.74	0.49	0.39		0.46	0.74	0.36	0.19	0.78	0.27	0.65	0.48	0.57	0.69	0.75	0.60	0.47	0.69	0.42	0.32	0.56	0.73	0.30
-7	0.51	0.42	0.41	0.47	0.49		0.46	0.81	0.37	0.09	0.69	0.35	0.58	0.71	0.53	0.62	0.79	0.67	-0.01	0.67	0.44	0.54	0.53	0.75	0.35
-6	0.55	0.49	0.43	0.57	0.57		0.49	0.86	0.39	0.23	0.72	0.37	0.60	0.66	0.58	0.73	0.83	0.73	0.09	0.74	0.46	0.10	0.61	0.78	0.35
-5	0.57	0.55	0.71	0.52	0.55		0.42	0.88	0.43	0.07	0.75	0.41	0.72	0.68	0.68	0.65	0.83	0.75	0.02	0.78	0.46	0.44	0.69	0.79	0.32
-4	0.61	0.58	0.83	0.62	0.61		0.49	0.85	0.46	0.22	0.76	0.44	0.72	0.70	0.74	0.74	0.80	0.75	-0.02	0.81	0.50	-0.04	0.71	0.80	0.40
-3	0.64	0.60	0.80	0.58	0.61		0.46	0.87	0.49	0.73	0.74	0.49	0.75	0.74	0.64	0.69	0.80	0.80	0.11	0.79	0.51	0.24	0.66	0.77	0.41
-2	0.63	0.62	0.87	0.59	0.64		0.38	0.83	0.64	0.08	0.92	0.74	0.75	0.73	0.66	0.69	0.77	0.76	-0.01	0.81	0.51	0.44	0.63	0.75	0.41
-1	0.63	0.54	0.80	0.63	0.48		0.41	0.90	0.65	0.27	0.94	0.72	0.73	0.76	0.77	0.72	0.72	0.57	0.05	0.82	0.71	0.49	0.64	0.81	0.31
0	0.63	0.55	0.87	0.65	0.53		0.48	0.70	0.62	0.26	0.91	0.63	0.79	0.70	0.62	0.62	0.78	0.76	0.06	0.87	0.71	0.46	0.65	0.73	0.43
1	0.65	0.60	0.83	0.56	0.61		0.42	0.80	0.63	0.45	0.93	0.64	0.81	0.68	0.76	0.66	0.67	0.73	0.02	0.87	0.72	0.31	0.67	0.83	0.42
2	0.66	0.59	0.86	0.72	0.62		0.40	0.81	0.70	0.14	0.93	0.75	0.81	0.80	0.75	0.65	0.75	0.75	0.24	0.86	0.68	0.36	0.68	0.84	0.35
3	0.67	0.58	0.79	0.57	0.65		0.41	0.85	0.68	0.60	0.93	0.75	0.80	0.78	0.76	0.67	0.68	0.77	-0.03	0.84	0.69	0.12	0.69	0.84	0.39
4	0.65	0.62	0.84	0.56	0.61		0.38	0.83	0.62	-0.01	0.87	0.73	0.81	0.82	0.78	0.72	0.75	0.78	0.05	0.84	0.69	0.61	0.74	0.82	0.41
5	0.66	0.56	0.78	0.68	0.57		0.28	0.80	0.78	0.45	0.91	0.84	0.75	0.81	0.81	0.70	0.65	0.64	0.05	0.83	0.75	-0.01	0.60	0.78	0.36
6	0.60	0.54	0.71	0.67	0.52		0.33	0.69	0.77	0.19	0.90	0.78	0.71	0.80	0.75	0.61	0.56	0.58	-0.01	0.77	0.59	0.30	0.39	0.71	0.46
7	0.57	0.62	0.53	0.51	0.51		0.40	0.68	0.77	0.12	0.86	0.83	0.63	0.62	0.64	0.49	0.62	0.56	-0.04	0.80	0.68	0.36	0.47	0.77	0.33
8	0.61	0.64	0.54	0.66	0.46		0.38	0.64	0.74	0.43	0.88	0.85	0.61	0.66	0.78	0.63	0.62	0.74	-0.04	0.82	0.75	0.49	0.69	0.72	0.30
9	0.61	0.56	0.62	0.56	0.47		0.25	0.57	0.71	0.63	0.83	0.86	0.72	0.69	0.73	0.66	0.59	0.72	0.07	0.79	0.72	0.34	0.66	0.65	0.36
10	0.58	0.54	0.69	0.55	0.47		0.30	0.63	0.70	0.13	0.75	0.81	0.70	0.72	0.66	0.58	0.69	0.71	-0.10	0.75	0.74	0.36	0.58	0.62	0.41
11	0.56	0.58	0.62	0.70	0.45		0.31	0.61	0.59	0.04	0.65	0.79	0.69	0.68	0.78	0.62	0.62	0.63	-0.04	0.77	0.77	0.38	0.69	0.66	0.38
12	0.55	0.51	0.67	0.43	0.43		0.19	0.66	0.73	0.16	0.67	0.77	0.70	0.64	0.73	0.62	0.64	0.53	0.07	0.77	0.67	-0.02	0.59	0.60	0.30
13	0.46	0.50	0.48	0.45	0.36		0.14	0.41	0.79	0.42	0.75	0.82	0.61	0.62	0.58	0.53	0.44	0.38	-0.06	0.58	0.33	0.18	0.58	0.27	0.26
14	0.44	0.33	0.51	0.52	0.35		0.26	0.43	0.61	0.24	0.62	0.76	0.59	0.43	0.56	0.62	0.60	0.51	-0.09	0.68	0.29	0.36	0.53	0.26	0.30
15	0.48	0.38	0.42	0.51	0.32		0.34	0.41	0.64	0.63	0.59	0.74	0.54	0.53	0.64	0.52	0.62	0.57	-0.06	0.62	0.55	0.55	0.59	0.29	0.33
16	0.44	0.41	0.47	0.63	0.29		0.18	0.35	0.56	0.25	0.51	0.74	0.56	0.51	0.59	0.61	0.59	0.56	-0.05	0.57	0.57	0.45	0.49	0.23	0.39
17	0.43	0.36	0.40	0.41	0.33		0.21	0.33	0.57	0.51	0.40	0.56	0.56	0.51	0.51	0.50	0.63	0.49	-0.01	0.53	0.62	0.30	0.51	0.23	0.27
18	0.42	0.44	0.46	0.51	0.33		0.26	0.37	0.40	0.41	0.39	0.52	0.54	0.42	0.58	0.53	0.57	0.49	-0.08	0.51	0.62	-0.14	0.48	0.27	0.30
19	0.39	0.42	0.43	0.47	0.32		0.21	0.39	0.40	0.29	0.35	0.50	0.62	0.40	0.50	0.50	0.50	0.50	-0.08	0.52	0.52	0.17	0.48	0.27	0.27
20	0.36	0.25	0.33	0.38	0.24		0.32	0.20	0.55	0.38	0.50	0.73	0.47	0.35	0.44	0.44	0.54	0.38	-0.08	0.38	0.26	0.18	0.49	0.11	0.32
21	0.37	0.25	0.36	0.23	0.25		0.28	0.26	0.42	0.52	0.35	0.53	0.53	0.43	0.50	0.46	0.59	0.47	-0.07	0.52	0.39	0.47	0.52	0.15	0.28
22	0.33	0.21	0.40	0.35	0.24		0.23	0.20	0.39	0.26	0.30	0.52	0.51	0.42	0.44	0.43	0.53	0.42	0.00	0.46	0.33	0.45	0.46	0.06	0.27
23	0.27	0.17	0.26	0.35	0.24		0.22	0.12	0.27	-0.04	0.23	0.43	0.46	0.29	0.40	0.45	0.55	0.39	-0.08	0.38	0.37	0.52	0.40	0.03	0.27
24	0.26	0.13	0.28	0.23	0.24		0.26	0.10	0.18	-0.03	0.17	0.33	0.44	0.30	0.39	0.43	0.52	0.37	-0.02	0.33	0.35	0.29	0.47	0.05	0.33
25	0.24	0.24	0.28	0.49	0.22		0.21	0.12	0.11	0.10	0.09	0.24	0.50	0.23	0.38	0.39	0.51	0.34	-0.06	0.33	0.40	0.04	0.42	0.04	0.20
26	0.26	0.19	0.25	0.32	0.16		0.23	0.16	0.28	0.33	0.18	0.32	0.47	0.22	0.31	0.31	0.40	0.27	0.00	0.36	0.42	0.12	0.38	0.05	0.21
27	0.25	0.06	0.20	0.28	0.08		0.26	0.06	0.39	0.20	0.30	0.64	0.42	0.19	0.35	0.32	0.48	0.27	-0.09	0.24	0.35	0.32	0.30	0.00	0.34
28	0.22	0.08	0.21	0.17	0.11		0.28	0.06	0.17	0.01	0.13	0.24	0.42	0.27	0.39	0.38	0.46	0.20	0.19	0.29	0.30	0.07	0.23	0.00	0.24
AVG	0.49	0.44	0.56	0.50	0.41		0.33	0.55	0.50	0.24	0.60	0.55	0.62	0.55	0.58	0.57	0.65	0.58	0.01	0.65	0.50	0.28	0.55	0.50	0.33

*early withdrawn sites not included in AVG in column 2; strong correlations are green, weaker ones are red; bolded values have p<0.05,

italicized values are p>0.05. n=84 maximum.

8.6 Site-Specific VL & Weekly Cases

As discussed in 8.2 and 8.4, weekly cases may provide a more practical, cumulative interpretation of the imminent case correlation. See Appendix 8.6 By Site Viral Loads and Weekly Cases. Note that four sites withdrew from the program early and are not included in the averaged sites columns. Bolded values have p values under 0.05, while italicized values are above 0.05. Stronger correlation coefficients are in green, weaker ones in red.

[illegible]

italicized values are $p>0.05$. n=84 maximum.

110

Chapter 9: Hospitalizations Correlation

The VDH Public Use dataset described in Chapter 5 contains Virginia COVID-19 case, hospitalization, and death counts on a cumulative daily cadence. Daily and Weekly hospitalization datasets are explored here in aggregate total, as well as by site and locality matched sets. The closest locality match to the respective SMP sites was selected for the by-site analyses, as determined by SMP site zip code and VDH locality listing.

9.1 Weekly VL & Daily Hospitalizations (general)

In Chapter 7, the aggregated VDH daily hospitalizations counts were visually explored in Figure 7.3.2. Combining the SMP weekly viral load data with the VDH daily hospitalization data yields Figure 9.1.1 below.

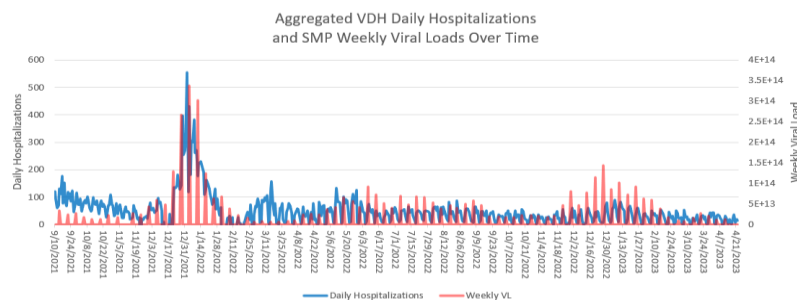


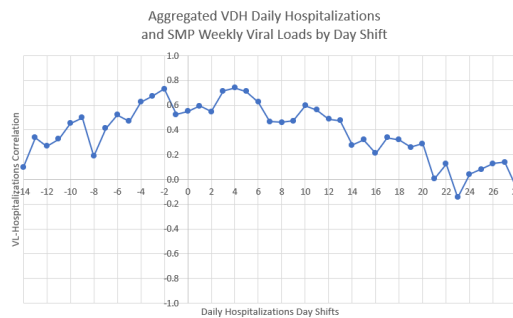
Figure 9.1.1 *Aggregated VDH daily hospitalization and SMP weekly viral loads over time, line graph*

Using the Excel CORREL function (Microsoft, 2023), correlation was explored between the weekly viral loads and daily hospitalizations (-1 to +1 correlation). Additionally, the hospitalizations were shifted between -14 days and +28 days, in order to explore if there were leading indicator effects.

The correlation between the SMP aggregated weekly viral loads (n=82) and daily hospitalizations was 0.551 at zero days shift (p<0.001), as seen below in indicating a weak or moderate relationship between aggregate Viral Loads and Daily Hospitalizations. Shifting the days demonstrated a similar correlation of approximately 0.7 between days -3 to -2 and +3 to +5.

Table 9.1.1 *Aggregated weekly viral loads and daily hospitalizations correlation coefficients with day shifts*

Day Shift	Correlation Coefficient
-3	0.676
-2	0.734
-1	0.524
0	0.551
1	0.595
2	0.549
3	0.714
4	0.740



*n=82 (weekly aggregated samples); all p values < .05, except for days -14, -8, +16, +21 to +28, day shifts

Figure 9.1.2 *Aggregated weekly viral loads and daily hospitalizations correlation coefficients with day shifts, line graph*

9.2 Weekly VL & Weekly Hospitalizations (general)

While using weekly viral loads and daily hospitalizations correlation gives us a snapshot moment of viral load and hospitalizations, it may be more practical to think of hospitalizations in weekly terms. Then one (e.g., a local district epidemiologist) may view a once-weekly viral load value and ask how related the next weeks' worth of hospitalizations will be. However, it should still be noted that a weekly viral load is either a grab or 24-hr composite sample, taken once a week. It is itself a snapshot analysis of a community and does not represent an entire week of testing. However, since viral load analysis simply isn't currently available consistently throughout the week, a once-weekly viral load will have to suffice.

Aggregated SMP weekly viral loads and aggregated VDH hospitalizations – as defined as the following 7 days' worth of hospitalizations – was explored for correlation below in Figure

9.2.1. Then a day shift between -14 and +28 was applied, changing the week's hospitalizations timeframe to an entire week before to one-to-two weeks afterwards, as seen below in Table 9.2.1 and Figure 9.2.2.

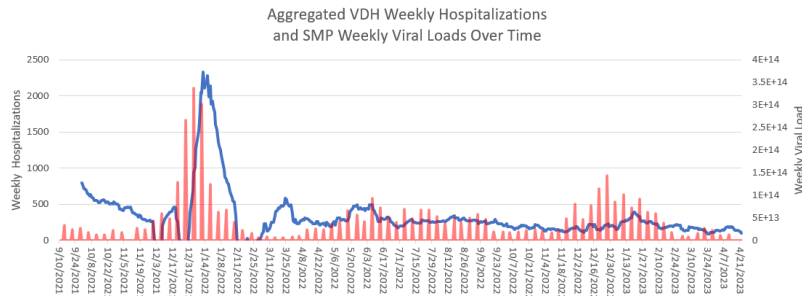
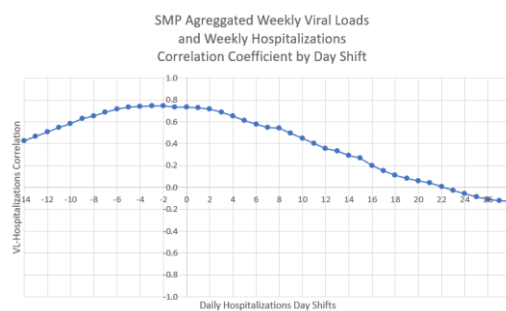


Figure 9.2.1 Aggregated SMP weekly viral loads and weekly (next 7 days) hospitalizations over time

Table 9.2.1 Aggregated weekly viral loads and weekly hospitalizations correlation coefficients with day shifts

Day Shift	Correlation Coefficient
-6	0.719
-5	0.736
-4	0.740
-3	0.745
-2	0.747
-1	0.736
0	0.732
1	0.729



*n=82 (weekly aggregated samples); all p values < .05, except for +16 to +28 day shifts

Figure 9.2.2 aggregated weekly viral loads and weekly hospitalizations correlation coefficients with day shifts, line graph

The findings above indicate that a viral load, analyzed and interpreted the same day, is moderately to strongly correlated to the following 7 days' worth of hospitalizations (with a correlation coefficient of 0.732, $p < 0.001$). Furthermore, a similar correlation coefficient is noted for the weeks starting in the days surrounding the viral load sampling.

One could interpret this result to mean that there is a moderate to strong relationship between weekly viral loads and weekly hospitalizations, especially in the week's hospitalizations beginning up to 6 days before to 2 days after. For example, if a viral load sample were taken on

the 15th of the month, there is a moderate correlation to the hospitalizations ranging in the days from the 9th to the 24th.

9.3 Weekly VL & Daily Hospitalizations (by variant)

Several factors could be contributory to a dynamic relationship between VL and hospitalizations over time, to include:

- Changing variants with different infectious physiological effects
- Changing susceptibility within the host population (e.g., vaccination/exposure statuses, deaths)
- Case data collection changes (e.g., case definitions, take-home tests)
- WWS methodological changes (e.g., collection, courier, laboratory, analytic changes)

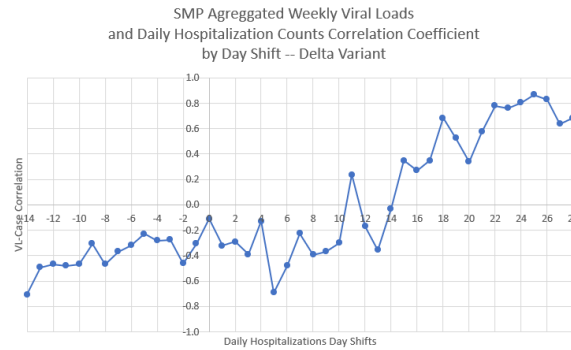
For this reason, the SMP aggregate weekly viral loads and daily hospitalizations were broken up and analyzed by variant timelines, as determined in Chapter 6's Synthesizing Variant Dominance Timeline. Seen below in Tables and Figures 9.3.1 to 9.3.5.

It is important to note here that daily case data does not smooth out natural variations found in case reporting. For example, hospitalization reporting on the weekends may drop below work week levels. Using a cumulative weekly summation, as seen in sections 8.2, 8.4, and 8.6, will by default smooth out work week fluctuations.

Delta (9/4/21 – 12/18/21)

Table 9.3.1 Aggregated weekly viral loads and daily hospitalizations correlation coefficient with day shifts for Delta variant

Day Shift	Correlation Coefficient
21	0.576
22	0.780
23	0.762
24	0.805
25	0.868
26	0.830
27	0.638
28	0.683



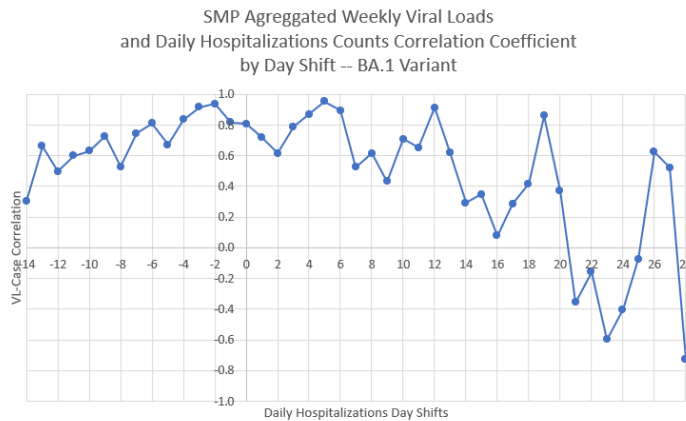
*n=13 (weekly aggregated samples); all p values < .05, except for -14 to +17 and +19 to +20 day shifts

Figure 9.3.1 Aggregated weekly viral loads and daily hospitalizations correlation coefficient with day shifts for Delta variant, line graph

Omicron BA.1 (1/1/22 – 3/12/22)

Table 9.3.2 Aggregated weekly viral loads and daily hospitalizations correlation coefficient with day shifts for BA.1 variant

Day Shift	Correlation Coefficient
-4	0.835
-3	0.916
-2	0.937
-1	0.817
0	0.808
1	0.718
2	0.616
3	0.788
4	0.870
5	0.953
6	0.894



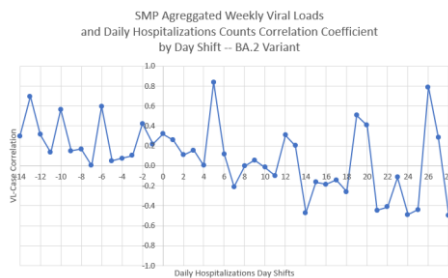
*n=10 (weekly aggregated samples); all p values < .05, except for -14, -12 to -10, -8, +2, +7 to +9, and +13 to +18, +20 to +28 day shifts

Figure 9.3.2 Aggregated weekly viral loads and daily hospitalizations correlation coefficient with day shifts for BA.1 variant, line graph

Omicron BA.2 (4/2/22 – 6/18/22)

Table 9.3.3 Aggregated weekly viral loads and daily hospitalizations correlation coefficient with day shifts for BA.2 variant

Day Shift	Correlation Coefficient
4	0.003
5	0.838
6	0.119
7	-0.210
8	-0.001



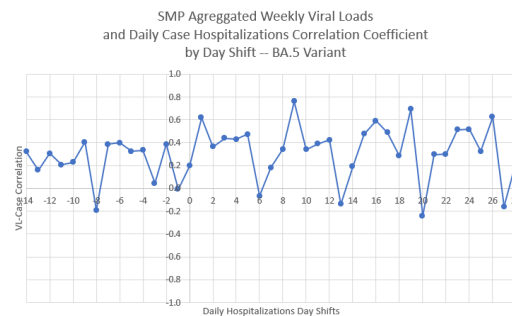
*n=11 (weekly aggregated samples); the only p values < .05 include -13, +5, and +26 day shifts

Figure 9.3.3 Aggregated weekly viral loads and daily hospitalizations correlation coefficient with day shifts for BA.2 variant, line graph

Omicron BA.5 (7/16/22 – 12/10/22)

Table 9.3.4 Aggregated weekly viral loads and daily hospitalizations correlation coefficient with day shifts for BA.5 variant

Day Shift	Correlation Coefficient
1	0.620
2	0.364
3	0.440
4	0.429
5	0.472
6	-0.066



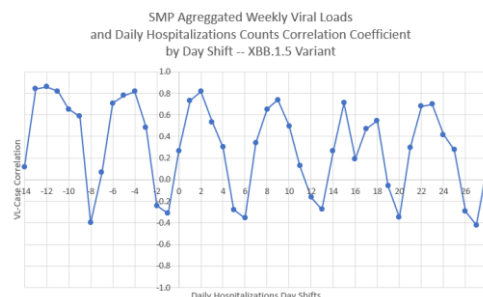
*n=21 (weekly aggregated samples); all p values < .05 except -14 to -12 and -10 to -5 day shifts

Figure 9.3.4 Aggregated weekly viral loads and daily hospitalizations correlation coefficient with day shifts for BA.5 variant, line graph

Omicron XBB.1.5 (1/14/23 – 4/1/23)

Table 9.3.5 Aggregated weekly viral loads and daily hospitalizations correlation coefficient with day shifts for XBB.1.5 variant

Day Shift	Correlation Coefficient
-1	-0.312
0	0.269
1	0.734
2	0.818
3	0.535
4	0.301



*n=11 (weekly aggregated samples); p values < .05 except +9 to +10 and +20 to +28 day shifts

Figure 9.3.5 Aggregated weekly viral loads and daily hospitalizations correlation coefficient with day shifts for XBB.1.5 variant, line graph

All Variants Summarized

Daily hospitalizations reached their peak correlation with viral loads during the -4 to -2 day shift period, as well as the +3 to +6 day shift period, as seen below in Table 9.3.6 (n=82). The most correlated day was on +4 day shift, with a 0.740 correlation coefficient (p<0.001). Additionally, each variant visually demonstrated different peak periods and peak days -- Delta (+18 to +28 days period; +25 peak day), BA.1 (-4 to +6 days period; +5 peak day), BA.2 (no discernable period; +5 peak day), BA.5 (no discernable period; +9 peak day), XBB.1.5 (no discernable period; -12 peak day).

Table 9.3.6 Summarized weekly viral loads and daily hospitalizations correlation coefficient with day shifts for all variants

Day Shift	All	Delta	BA.1	BA.2	BA.5	XBB.1.5
-14	0.098	-0.706	0.303	0.295	0.321	0.117
-13	0.338	-0.491	0.663	0.696	0.160	0.841
-12	0.270	-0.468	0.499	0.315	0.306	0.860
-11	0.326	-0.480	0.600	0.136	0.205	0.821
-10	0.453	-0.469	0.632	0.565	0.226	0.654
-9	0.499	-0.308	0.727	0.149	0.406	0.587
-8	0.188	-0.466	0.528	0.167	-0.191	-0.398
-7	0.412	-0.367	0.745	0.005	0.386	0.066
-6	0.521	-0.317	0.812	0.596	0.398	0.707
-5	0.471	-0.228	0.672	0.051	0.324	0.780
-4	0.625	-0.281	0.835	0.076	0.333	0.816
-3	0.676	-0.274	0.916	0.102	0.046	0.484
-2	0.734	-0.460	0.937	0.422	0.387	-0.242
-1	0.524	-0.306	0.817	0.217	-0.010	-0.312
0	0.551	-0.110	0.808	0.321	0.197	0.269
1	0.595	-0.321	0.718	0.258	0.620	0.734
2	0.549	-0.290	0.616	0.109	0.364	0.818
3	0.714	-0.391	0.788	0.155	0.440	0.535
4	0.740	-0.130	0.870	0.003	0.429	0.301
5	0.712	-0.691	0.953	0.838	0.472	-0.278
6	0.626	-0.479	0.894	0.119	-0.066	-0.355
7	0.467	-0.226	0.526	-0.210	0.182	0.340
8	0.460	-0.392	0.615	-0.001	0.342	0.654
9	0.475	-0.369	0.432	0.057	0.761	0.739
10	0.597	-0.299	0.709	-0.013	0.340	0.497
11	0.562	0.235	0.652	-0.097	0.388	0.127
12	0.488	-0.171	0.912	0.309	0.423	-0.162
13	0.478	-0.355	0.621	0.204	-0.138	-0.274
14	0.278	-0.029	0.291	-0.470	0.190	0.265
15	0.320	0.348	0.350	-0.165	0.475	0.710
16	0.212	0.271	0.078	-0.186	0.591	0.194
17	0.340	0.348	0.288	-0.143	0.490	0.471
18	0.321	0.684	0.416	-0.259	0.284	0.547
19	0.260	0.524	0.860	0.508	0.693	-0.057
20	0.288	0.339	0.369	0.408	-0.240	-0.350
21	0.004	0.576	-0.354	-0.445	0.296	0.297
22	0.125	0.780	-0.159	-0.412	0.299	0.680
23	-0.144	0.762	-0.597	-0.112	0.512	0.698
24	0.043	0.805	-0.401	-0.488	0.516	0.416

25	0.084	0.868	-0.072	-0.441	0.325	0.278
26	0.129	0.830	0.624	0.790	0.625	-0.293
27	0.140	0.638	0.523	0.284	-0.160	-0.422
28	-0.072	0.683	-0.725	-0.496	0.226	0.132

*Strong correlations are green, weaker ones are red. Bolded values have $p < 0.05$, italicized values are $p > 0.05$.

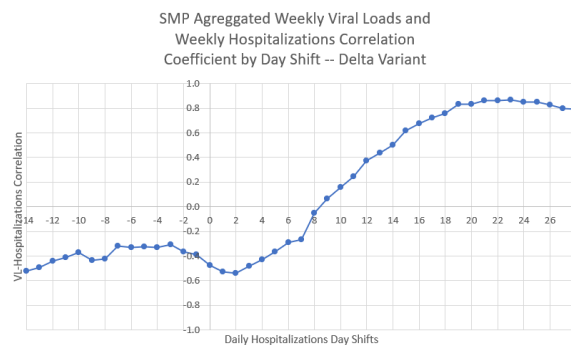
9.4 Weekly VL & Weekly Hospitalizations (by variant)

In the same way that section 8.2 reassessed daily hospitalizations for a week's time, this section will reassess 8.3's SMP weekly hospitalizations by variant in the following Tables and Figures 9.4.1 – 9.4.5.

Delta (9/4/21 – 12/18/21)

Table 9.4.1 Aggregated weekly viral loads and weekly hospitalizations correlation coefficient with day shifts for Delta variant

Day Shift	Correlation Coefficient
18	0.757
19	0.833
20	0.831
21	0.860
22	0.865
23	0.868
24	0.853
25	0.852



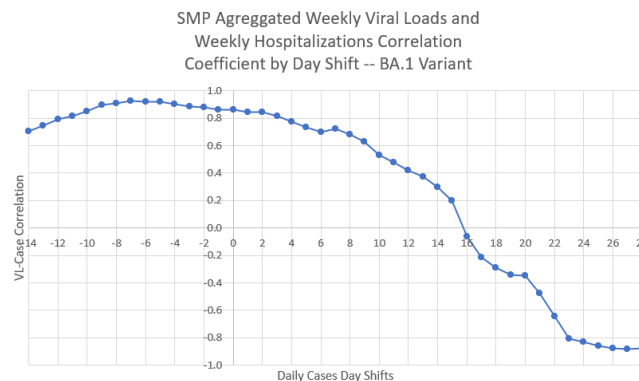
*n=13 (weekly aggregated samples); p values < .05 included +15 to +28 day shifts

Figure 9.4.1 Aggregated weekly viral loads and weekly hospitalizations correlation coefficient with day shifts for Delta variant, line

Omicron BA.1 (1/1/22 – 3/12/22)

Table 9.4.2 Aggregated weekly viral loads and weekly hospitalizations correlation coefficient with day shifts for BA.1 variant

Day Shift	Correlation Coefficient
-9	0.896
-8	0.909
-7	0.927
-6	0.921
-5	0.921
-4	0.902
-3	0.887
-2	0.880



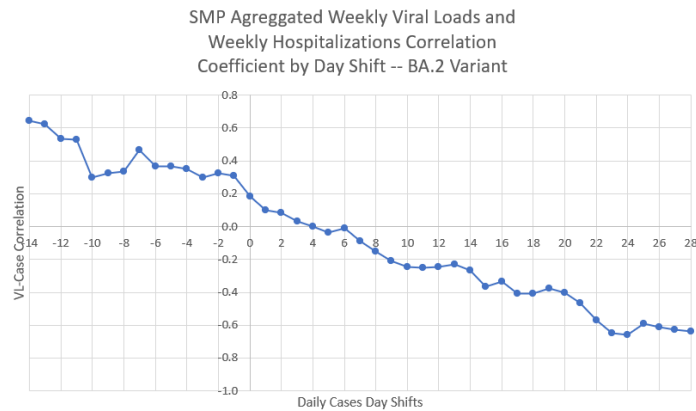
*n=10 (weekly aggregated samples); all p values < .05, except for +9 to +28 day shifts

Figure 9.4.2 Aggregated weekly viral loads and weekly hospitalizations correlation coefficient with day shifts for BA.1 variant, line graph

Omicron BA.2 (4/2/22 – 6/18/22)

Table 9.4.3 Aggregated weekly viral loads and weekly hospitalizations correlation coefficient with day shifts for BA.2 variant

Day Shift	Correlation Coefficient
-14	0.644
-13	0.622
-12	0.534
-11	0.529
-10	0.299
-9	0.325
-8	0.334
-7	0.465
-6	0.368
-5	0.365
-4	0.353
-3	0.300



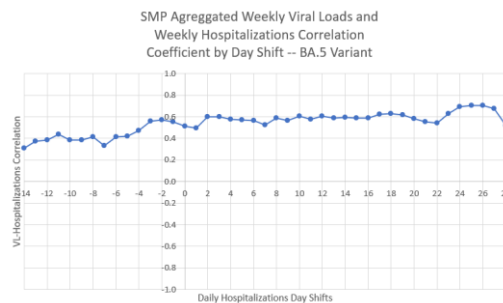
*n=11 (weekly aggregated samples); only p values < .05 included -14 to -13 day shifts

Figure 9.4.3 Aggregated weekly viral loads and weekly hospitalizations correlation coefficient with day shifts for BA.2 variant, line graph

Omicron BA.5 (7/16/22 – 12/10/22)

Table 9.4.4 Aggregated weekly viral loads and weekly hospitalizations correlation coefficient with day shifts for BA.5 variant

Day Shift	Correlation Coefficient
21	0.556
22	0.543
23	0.627
24	0.694
25	0.707
26	0.703
27	0.677
28	0.520



*n=21 (weekly aggregated samples); only p values < .05 included -11 and -4 to +28 day shifts

Figure 9.4.4 Aggregated weekly viral loads and weekly hospitalizations correlation coefficient with day shifts for BA.5 variant, line graph

Omicron XBB.1.5 (1/14/23 – 4/1/23)

Table 9.4.5 Aggregated weekly viral loads and weekly hospitalizations correlation coefficient with day shifts for XBB.1.5 variant

Day Shift	Correlation Coefficient
-8	0.674
-7	0.803
-6	0.808
-5	0.795
-4	0.743
-3	0.756
-2	0.758
-1	0.783
0	0.770
1	0.743

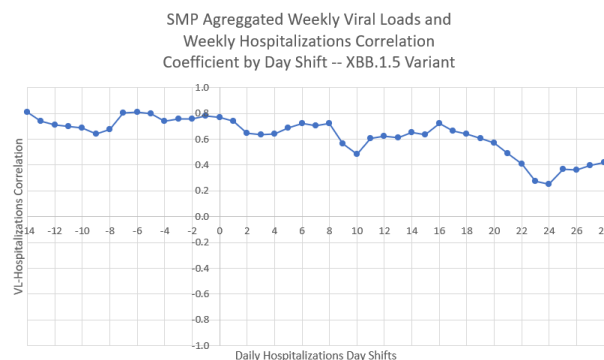


Figure 9.4.5 Aggregated weekly viral loads and weekly hospitalizations correlation coefficient with day shifts for XBB.1.5 variant, line graph

All Variants Summarized

Weekly hospitalizations reached their peak correlation with viral loads during the -7 to +3 day shift period, and specifically on the -3 peak day, as seen below in Table 9.4.5.

Additionally, each variant visually demonstrated different peak periods and peak days -- Delta (+19 to +26 days period; +23 peak day), BA.1 (-9 to -3 days period; -7 peak day), BA.2 (-14 to -11 days; -14 peak day), BA.5 (+23 to +27 days; +25 peak day), and XBB.1.5 (-14 to +1 days; -14 peak day). Weekly hospitalizations are defined as the VDH SARS-CoV-2 aggregated hospitalizations count summation of the 7-day period, beginning on the noted shift day.

Table 9.4.6 *Weekly hospitalizations correlation coefficient by variant*

Weekly Hospitalizations Correlation Coefficient by Variant						
Day Shift	All	Delta	BA.1	BA.2	BA.5	XBB.1.5
-14	0.425	-0.522	0.703	0.644	0.307	0.811
-13	0.468	-0.492	0.748	0.622	0.372	0.743
-12	0.508	-0.442	0.790	0.534	0.386	0.708
-11	0.547	-0.411	0.815	0.529	0.437	0.701
-10	0.580	-0.373	0.850	0.299	0.387	0.686
-9	0.632	-0.436	0.896	0.325	0.387	0.638
-8	0.654	-0.423	0.909	0.334	0.416	0.674
-7	0.689	-0.321	0.927	0.465	0.334	0.803
-6	0.719	-0.328	0.921	0.368	0.412	0.808
-5	0.736	-0.325	0.921	0.365	0.417	0.795
-4	0.740	-0.328	0.902	0.353	0.471	0.743
-3	0.745	-0.307	0.887	0.300	0.560	0.756
-2	0.747	-0.367	0.880	0.326	0.569	0.758
-1	0.736	-0.390	0.860	0.312	0.556	0.783
0	0.732	-0.474	0.862	0.184	0.516	0.770
1	0.729	-0.527	0.847	0.100	0.497	0.743
2	0.718	-0.540	0.844	0.086	0.603	0.648
3	0.687	-0.479	0.818	0.033	0.602	0.634
4	0.653	-0.428	0.775	-0.001	0.575	0.641
5	0.610	-0.364	0.736	-0.037	0.571	0.688
6	0.579	-0.292	0.701	-0.008	0.565	0.722
7	0.550	-0.264	0.720	-0.086	0.522	0.705
8	0.541	-0.050	0.684	-0.150	0.588	0.724
9	0.495	0.064	0.628	-0.211	0.563	0.567
10	0.448	0.160	0.533	-0.247	0.607	0.487
11	0.402	0.245	0.478	-0.253	0.576	0.609
12	0.359	0.376	0.422	-0.246	0.604	0.626
13	0.331	0.440	0.374	-0.230	0.590	0.612
14	0.291	0.502	0.296	-0.264	0.593	0.653
15	0.266	0.617	0.201	-0.367	0.587	0.638
16	0.201	0.677	-0.063	-0.333	0.590	0.723
17	0.152	0.725	-0.215	-0.407	0.623	0.664
18	0.111	0.757	-0.288	-0.409	0.629	0.640
19	0.085	0.833	-0.344	-0.375	0.618	0.604
20	0.058	0.831	-0.346	-0.403	0.585	0.569
21	0.045	0.860	-0.474	-0.467	0.556	0.492
22	0.006	0.865	-0.647	-0.570	0.543	0.409
23	-0.027	0.868	-0.804	-0.650	0.627	0.276
24	-0.057	0.853	-0.828	-0.658	0.694	0.254
25	-0.084	0.852	-0.862	-0.592	0.707	0.369
26	-0.111	0.825	-0.877	-0.613	0.703	0.360
27	-0.120	0.799	-0.880	-0.625	0.677	0.398
28	-0.129	0.792	-0.879	-0.635	0.520	0.423

*Strong correlations are green, weaker ones are red. Bolded values have $p < 0.05$, italicized values are $p > 0.05$.

9.5 Site-Specific VL & Daily Hospitalizations

Chapter 3 discussed inherent variability in any WWS system. VDH WWS offers a unique opportunity, in that the SMP:

- uses a single laboratory for all shipping and analysis (removing cross-laboratory variability)
- has been supported by CDC NWSS, financially and with good practice guidance
- has a wide network of participating utility partners, geographically covering much of Virginia
- has been collecting data with minimal changes to methodology/analysis for over 2 years
- has a top-notch Data Manager

While these factors are excellent for maintaining a proper dataset over time, they do not eliminate the core differences between the utility sites themselves. Each wastewater system is subject to different infrastructure, geography, population, and commercial factors which affect the time and degradation of viral particles across distance and time. It is for this reason that caution is heavily applied when discussing options for comparing sites.

Each SMP site was analyzed by SMP viral load and VDH hospitalizations datasets, as seen below in Table 9.5.1. See Appendix 9.5 By Site Viral Loads and Daily Hospitalizations. Bolded values have p values under 0.05, while italicized values are above 0.05. Stronger correlation coefficients are in green, weaker ones in red.

Table 9.5.1 By Site VL - Daily Hospitalizations Correlation Coefficient with Day Shifts

Daily Hospitalization Correlation Coefficient by Site																									
Day Shift	AVG	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
-14	0.03	0.17	0.17	0.05	-0.16		0.11	0.39	-0.40	-0.03	-0.02	-0.01	-0.13	-0.02	0.14	0.24	0.11	-0.01	-0.10	-0.02	ERROR	0.00	-0.19	0.24	0.22
-13	0.03	0.15	0.13	-0.25	0.02		-0.05	0.15	0.16	-0.09	-0.08	-0.16	0.17	0.07	0.24	0.17	-0.11	0.04	-0.10	0.02	-0.08	0.22	0.03	0.03	-0.10
-12	0.10	0.52	0.38	0.08	-0.06		0.19	0.39	0.32	0.06	0.31	-0.43	-0.02	0.08	0.16	0.19	-0.15	-0.03	-0.05	0.08	0.49	-0.07	-0.03	0.14	-0.17
-11	0.12	0.19	0.29	0.03	0.16		0.20	0.52	0.05	-0.07	0.13	0.11	0.09	0.17	-0.05	0.33	0.05	-0.04	0.24	-0.08	0.15	0.12	-0.09	0.30	-0.05
-10	0.10	0.12	0.00	-0.09	0.12		-0.01	0.60	0.18	0.14	0.17	0.12	0.13	-0.01	-0.13	0.39	-0.01	-0.08	0.02	0.00	0.19	0.14	0.11	0.15	-0.03
-9	0.05	-0.04	0.22	-0.08	-0.05		0.04	0.57	0.22	0.10	0.03	0.15	-0.01	0.09	-0.01	0.11	-0.05	-0.01	-0.05	-0.05	-0.01	-0.04	0.05	0.14	-0.11
-8	0.16	0.41	0.35	0.23	-0.09		0.18	0.70	0.24	0.01	0.33	0.02	-0.03	0.13	0.09	0.44	-0.10	0.02	-0.05	0.22	0.01	ERROR	-0.04	0.33	0.08
-7	0.12	0.26	0.10	0.37	-0.08		-0.07	0.71	0.19	0.15	0.05	-0.01	-0.03	0.14	0.21	0.52	-0.05	0.13	-0.07	-0.04	ERROR	0.05	-0.13	0.21	-0.02
-6	0.13	0.28	0.28	0.00	0.14		0.02	0.54	0.26	0.06	0.50	0.20	0.16	0.01	0.13	0.35	-0.17	0.04	-0.02	0.05	0.10	-0.01	0.14	0.03	-0.06
-5	0.14	0.37	0.37	0.31	-0.08		0.18	0.77	0.12	0.00	0.33	0.06	0.14	-0.02	0.09	0.46	-0.11	-0.10	-0.01	0.16	0.08	-0.11	-0.03	0.21	-0.05
-4	0.13	-0.01	0.21	0.10	0.15		0.19	0.73	0.12	0.40	0.08	0.14	0.21	0.01	0.03	0.38	0.12	-0.09	-0.03	0.02	0.14	-0.03	0.03	0.28	-0.10
-3	0.15	0.12	0.16	0.09	0.15		0.06	0.80	0.30	-0.01	0.61	0.24	0.20	0.24	0.03	0.30	0.07	-0.02	-0.02	0.00	-0.05	-0.08	0.10	0.17	-0.06
-2	0.12	0.14	0.19	-0.01	0.06		0.13	0.74	0.35	0.07	0.63	0.32	0.01	-0.04	0.03	0.20	-0.04	-0.08	0.01	0.02	-0.10	0.08	0.09	-0.02	0.02
-1	0.15	0.09	0.20	0.53	-0.07		0.15	0.81	0.18	0.02	-0.02	0.17	-0.08	0.14	0.22	0.33	-0.07	0.06	-0.06	0.13	0.08	ERROR	0.00	0.53	0.05
0	0.18	0.30	0.28	0.37	-0.02		0.04	0.70	0.12	0.30	0.33	0.27	-0.09	0.27	0.24	0.69	-0.03	0.21	0.12	-0.15	ERROR	-0.18	-0.02	0.22	0.09
1	0.17	0.37	0.06	-0.04	-0.04		0.16	0.77	0.57	0.20	0.61	0.15	0.31	0.07	0.22	0.20	-0.04	-0.06	0.11	-0.03	0.25	-0.04	0.15	0.07	-0.02
2	0.12	0.33	0.39	0.14	-0.20		-0.02	0.60	0.28	-0.20	0.56	0.16	0.22	0.00	-0.03	0.43	-0.10	-0.06	-0.05	0.18	-0.02	-0.04	0.03	0.16	-0.09
3	0.14	0.18	0.22	-0.12	0.19		0.07	0.67	0.14	0.15	0.37	0.43	0.20	-0.01	0.14	0.35	0.11	-0.07	-0.04	0.04	-0.04	0.24	-0.20	0.22	0.03
4	0.13	0.30	-0.03	-0.21	0.01		0.19	0.50	0.23	-0.05	0.84	0.32	0.20	0.26	0.05	0.55	0.14	0.00	-0.03	-0.07	-0.03	-0.17	-0.11	0.13	-0.03
5	0.16	0.11	0.38	-0.11	-0.08		0.08	0.70	0.71	0.50	0.73	0.42	-0.02	0.21	0.02	0.01	-0.13	-0.12	-0.05	0.03	0.02	0.06	-0.01	-0.04	0.15
6	0.21	0.15	0.43	0.22	-0.12		-0.08	0.54	0.16	0.61	0.21	0.13	0.12	0.05	0.30	0.47	-0.01	0.07	-0.06	0.40	0.39	ERROR	-0.01	0.66	-0.02
7	0.21	0.17	0.09	0.49	0.07		-0.04	0.68	0.53	0.20	0.76	0.50	-0.08	0.24	0.14	0.50	-0.09	0.09	-0.09	-0.03	ERROR	0.09	0.04	0.27	0.17
8	0.11	0.21	0.16	-0.06	-0.11		0.11	0.56	0.29	0.04	0.20	0.21	0.36	0.00	0.23	0.40	-0.13	-0.04	0.11	0.03	0.01	-0.01	-0.02	0.07	0.00
9	0.13	0.46	0.12	0.51	-0.10		0.17	0.46	0.31	0.26	0.46	0.02	0.16	0.06	-0.09	0.22	-0.12	-0.07	0.19	0.04	0.02	-0.18	-0.10	0.15	0.08
10	0.16	0.30	0.28	0.27	0.15		0.24	0.39	0.23	0.19	0.57	0.31	0.15	-0.05	0.27	0.35	0.09	0.02	-0.02	-0.02	0.08	-0.06	-0.12	0.23	-0.06
11	0.12	0.16	0.17	-0.06	-0.02		0.06	0.23	0.29	-0.03	0.62	0.41	-0.05	0.38	0.03	0.52	0.03	-0.04	0.02	0.02	-0.02	-0.11	-0.09	0.13	0.09
12	0.12	0.11	0.26	-0.16	-0.02		0.14	0.40	0.47	0.21	0.40	0.59	-0.11	0.01	0.09	0.08	-0.11	-0.08	-0.04	0.09	0.24	0.27	-0.10	0.03	-0.01
13	0.21	0.23	0.18	0.38	-0.09		0.22	0.34	0.22	0.38	0.27	0.28	0.29	0.16	0.27	0.48	-0.05	0.22	0.00	0.23	0.13	ERROR	0.06	0.32	0.06
14	0.21	0.26	0.10	0.43	0.06		0.05	0.36	0.33	0.36	0.16	0.76	-0.02	0.31	0.14	0.43	-0.02	0.04	0.09	-0.04	-0.05	0.28	0.10	0.35	0.25
15	0.11	0.33	0.15	-0.03	-0.09		0.24	0.31	0.21	0.31	0.26	0.34	0.26	0.03	0.13	0.44	-0.08	0.04	-0.11	-0.11	-0.05	-0.10	0.06	0.06	-0.05
16	0.07	0.23	0.11	0.18	-0.13		0.21	0.21	0.05	0.22	0.22	0.14	0.18	0.02	-0.09	0.14	-0.11	0.05	-0.09	-0.03	0.14	-0.01	-0.12	0.04	0.03
17	0.09	0.18	0.39	-0.04	0.02		0.05	0.23	0.28	0.10	0.35	0.05	0.12	0.04	0.37	0.12	-0.01	-0.04	-0.03	0.03	0.20	-0.21	-0.18	0.07	0.03
18	0.06	0.16	0.09	-0.05	0.07		0.10	0.05	0.14	-0.04	0.34	0.08	0.18	0.18	-0.03	0.36	0.21	-0.12	-0.27	-0.02	-0.01	-0.12	-0.12	0.07	0.07
19	0.10	0.18	0.43	-0.01	-0.01		0.14	0.18	0.09	0.30	0.17	0.40	0.01	0.16	0.00	-0.06	-0.11	0.07	0.27	0.02	0.17	0.03	-0.04	0.00	-0.07
20	0.09	0.01	0.13	0.34	-0.05		0.26	0.11	0.01	-0.05	0.19	0.68	-0.10	0.08	0.21	0.11	-0.04	-0.06	-0.07	-0.03	0.00	ERROR	0.19	0.20	-0.05
21	0.20	0.18	0.34	0.47	0.03		0.07	0.21	0.31	0.75	0.01	0.70	-0.07	0.14	0.14	0.22	0.01	0.12	-0.06	-0.07	ERROR	0.32	0.22	0.30	0.13
22	0.07	0.22	-0.10	0.06	-0.12		0.16	0.02	0.27	0.42	0.02	0.23	0.26	0.10	0.10	0.34	-0.03	0.01	-0.08	-0.04	0.02	-0.13	-0.19	0.05	0.01
23	0.07	0.17	0.14	0.14	-0.11		0.47	0.13	0.02	0.34	-0.05	0.34	0.36	0.03	-0.13	0.05	-0.15	0.01	-0.02	-0.04	0.11	-0.02	-0.02	-0.14	0.07
24	0.06	0.16	0.06	0.30	-0.03		0.00	0.11	0.04	0.04	-0.21	-0.13	0.23	0.01	0.32	0.23	-0.03	0.03	-0.02	0.05	0.03	0.31	-0.11	-0.06	0.04
25	0.02	0.26	0.18	0.07	0.03		0.10	-0.01	-0.07	0.11	-0.02	-0.11	0.18	0.13	-0.02	0.07	0.15	-0.13	-0.01	-0.14	-0.04	-0.01	-0.29	0.02	-0.01
26	0.03	0.11	0.21	-0.01	0.03		0.17	0.05	0.00	0.22	-0.15	-0.07	-0.05	0.02	-0.08	0.04	-0.04	0.10	-0.03	0.07	-0.01	0.30	-0.21	0.02	-0.04
27	0.05	0.13	0.06	0.13	-0.11		-0.02	0.02	0.06	-0.07	0.07	0.54	0.30	0.03	0.00	0.12	0.01	-0.13	0.01	0.18	0.03	ERROR	-0.07	-0.01	-0.07
28	0.10	-0.08	0.28	0.17	0.04		0.20	0.05	0.19	0.18	-0.17	0.60	0.20	-0.06	0.06	0.24	-0.01	0.26	-0.12	0.13	ERROR	-0.13	0.12	-0.07	0.02
AVG	0.12	0.20	0.20	0.12	-0.01		0.11	0.42	0.20	0.16	0.26	0.23	0.11	0.09	0.10	0.29	-0.03	0.00	-0.02	0.03	0.07	0.02	-0.03	0.15	0.01

*early withdrawn sites not included in AVG in column 2; strong correlations are green, weaker ones are red; bolded values have $p < 0.05$, italicized values are $p > 0.05$. n=84 maximum; ERROR values are division by zero errors.

In situations where an entire day shift has zero associated daily hospitalizations, the correlation reads as a divisible by zero error. This affected two of the sites.

9.6 Site-Specific VL & Weekly Hospitalizations

As discussed in 9.2 and 9.4, weekly hospitalizations may provide a more practical, cumulative interpretation on the imminent hospitalization correlation, as seen below in Table 9.6.1. See Appendix 9.6 for By Site Viral Loads and Weekly Hospitalizations.

Table 9.6.1 By Site VL - Weekly Hospitalizations Correlation Coefficient with Day Shifts

Daily Hospitalizations Correlation Coefficient by Site																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													</
--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	----

*early withdrawn sites not included in AVG in column 2; strong correlations are green, weaker ones are red; bolded values have p<0.05, italicized values are p>0.05. n=84 maximum.

Overall, the hospitalizations are less correlated with VL than cases.

Chapter 10: Developing an ‘Alert’ Framework

Chapter 4 explored Alert System fundamentals, with emphasis on the Early Warning framework by Yang (2017). Yang implored the need for a simple, timely system which can detect significant aberrations between expected and received data. Several base model types were discussed, though they depended heavily on the type of surveillance data received and resources available.

Chapter 3 discussed inherent WWS limitations and variabilities, and Chapters 6-9 provided an initial exploration and correlation testing for the VDH WWSP sites themselves. These chapters were in step with each other, as the by-site analyses indicated wide differences in site metadata, case correlation, and hospitalization correlation. It is for this reason that any selected alert framework WWS be very cognizant of the inherent limitations of WWS -- in particular the between-site variabilities.

10.1 Independent and Dependent Variables

Setting the alert system independent and dependent variables is crucial to aligning the alert system with the nature of the surveillance system, as well as the needs of its target audience. The goal in mind is to develop an alert system for public health practitioners to use WWS as a warning of imminent rise in disease burden. Specifically, the SARS-CoV-2 infectious disease burden detected by the SMP.

Dependent Variable – Case Rises

The analyses conducted within this dissertation discovered a stronger correlation with weekly cases (Ch. 8.2), with a peak average correlation coefficient of 0.90 at -3 days shift, than

weekly hospitalizations (Ch. 9.2), with a peak average correlation coefficient of 0.75 at -3 days shift. Therefore, the dependent variable for any developed alert system will rely on expected versus actual imminent rises in cases.

Independent Variable – Viral Load Rises

On first pass, the obvious independent variable should be the viral loads reported weekly by the WWS system. Viral loads are the metric used and reported by the SMP to provide concise, flow-controlled values for determining a sewershed's viral particle burden.

However, this may not be the most appropriate use of viral load. The nature of wastewater and disease physiology (explored in Chapter 2.4) showed that the dumping of viral particles into the wastewater is a long process which begins potentially weeks before case determination, and up to months after case resolution. In this way, wastewater viral loads are blunted across longer periods of time than the instantaneous nature of reported case incidence. If the viral load/cases for a site have been high for weeks, having an alert that it remains historically high is not new information for practitioners. The point of the alert is to warn the user that WWS portends a *rising change in cases*.

Therefore, the independent variable will be a *positive change* in viral load.

10.2 Selecting a Base Model

As WWS is taken in snapshot fashion, in participating sewersheds only, the lack of spatial and temporal granularity makes any spatial or multifactorial analyses complicated – and beyond the scope of this dissertation, which aims to arm any jurisdiction with a simple tool to interpret a new (and limited) WWS dataset. This would suggest a temporal-style model, with a “short history baseline comparison” (Yang, 2017, pp. 39-40), as an excellent starting category.

One of the earliest historical examples in Yang's text includes the Simple Control Chart (Yang, 2017, pp. 40-43). Referenced as early as 1931 in Shewhart's engineering article for practical statistical measuring (Shewhart, 1931), a simple literature search revealed a much wider historical use as an epidemic alert indicator (Waqas, 2020). When the Control Chart is applied to epidemics, case incidence is applied as a time series for a simple case counts over time, with a baseline created during a non-epidemic period and statistically significant threshold for case incidence to indicate an alert (Rich & Terry, 1946).

The use of Control Charts, sometimes with modifications for Cumulative Sums (CUSUM) and Exponentially Weighted Moving Average (EWMA), has been used in alert detection for hospital-acquired infections (Morton, 2001). Morton applauds the Control Chart generally for such epidemiologic outbreak identification, "... Shewhart Chart efficiently detects larger, more abrupt changes" (Morton, 2001, p. 112), though he credits the variations of CUSUM and EWMA as "particularly suitable for the detection of small sustained increases in uncommon events" (Morton, 2001, p. 112).

Oftentimes in examples as seen above, the actual case incidence is the variable under surveillance. In this dissertation, WWS is an indirect, correlated marker to case incidence (see Chapter 8). That is why the viral load-case count correlation (Ch. 8) was particularly relevant. However, by using the indirect viral load as the surveillance data input, there is an opportunity to explore WWS as a leading indicator. This may be quite the improvement to case surveillance, as it's more helpful to know if cases are imminent rather than already unfolding. Additionally, if case data is not reliable, secondary sources may be the only options to gain awareness of community disease burden.

Simple Control Charts were selected as a framework model for this dissertation, as they allow for simplistic aberration detection of a short historical time series with extremely limited granularity in spatial and temporal contexts.

10.3 Simple Control Charts Principles

Originally used in commercial quality control, Simple Control Charts are statistical process control tools found commonly in the U.S. to monitor epidemiology concerns, such as COVID-19 (Waqas, 2020). Their premise is to calculate the mean and standard deviation of a set of historical data to create a *baseline*, then establish the *Upper Control Limit* (UCL) and Lower Control Limit. These limits are calculated as a mean plus a multiplier-selected standard deviation of a normally distributed dataset. This is referred to as the *mean-standard deviation* chart, though there is an option for a *median-range* chart using the max and min points as upper and lower limits (Yang, 2017, pp. 40-43). Only Upper Control Limit lines will be used in this dissertation, as Lower Control Lines are not of interest to a public health practitioner looking for imminent rises in cases.

Several important factors must be addressed when dealing with Simple Control Charts (specifically the mean-standard deviation chart). Firstly, the “series process characteristics must be identified under normal circumstances” (Yang, 2017, p. 41). This is to say that a time series of data points must be used as a baseline, preferably during a time of minimal external interferences (e.g., epidemic, case definition change, or virus variant changes). Additionally, these numbers need to be normally distributed.

Once the mean and standard deviation are calculated, a warning can be issued for any values in the time series which go above the UCL (as defined as a selected threshold multiplier --

also known as a confidence coefficient -- against the standard deviation). For extra bells on this whistle, one could add an Alarm Level, which indicates the intensity of each given alert.

Calculations (Yang, 2017, pp. 41-42).

$$\text{Mean (center line): } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

$$\text{Standard deviation: } s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

$$\text{Upper Control Limit: } UCL = \bar{x} \pm k \cdot s$$

k represents a confidence coefficient (usually 2 or 3) to adjust sensitivity of early warning. Assuming normal distribution, a confidence coefficient of 2 represents a confidence level of 95%, while 3 represents a confidence level of 99%.

The above framework represents the Shewhart Control Chart, however, there are a few variations of the chart. A time series just using the data series numbers, just as they are, is referred to as an “Individuals” control chart. A time series using the difference between the individual data points is referred to as a *moving range* control chart. Using Individuals and Moving Range (I-MR) control charts can track process variability over time (Hessing & Ramana, 2024).

The framework selected for this dissertation alert system will primarily use the I-MR control chart to track the changes in viral loads and cases with each new viral load/case counts data point and calculate an Upper Limit for each set.

Calculations (NIST, 2024).

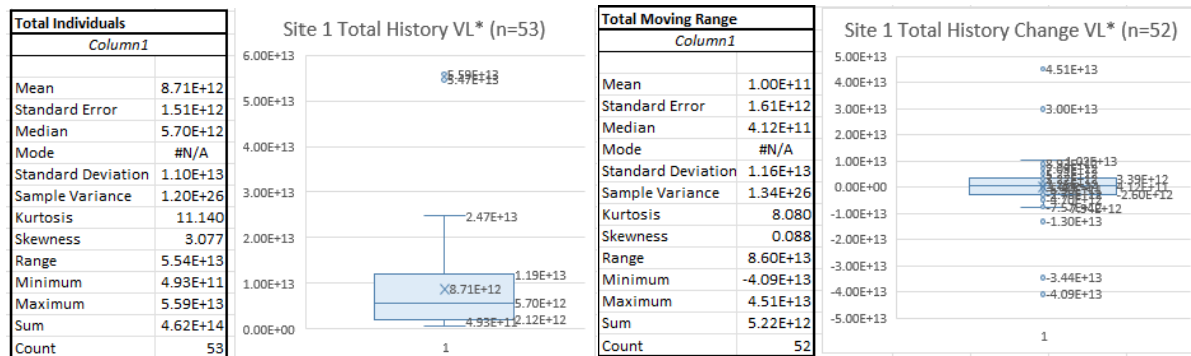
$$\text{Moving Range: } MR_i = |x_i - x_{i-1}|$$

The mean, standard deviation, and UCL are all calculated the same as *individuals* control chart, except the data points are changes in values instead of their raw values. Knowing the directionality and magnitude of the changes in viral load is particularly important when analyzing WWS, as viral shedding is a far more persistent environmental finding (occurring up to months after symptom abatement) than case incidence, which happens once per infection. However, a sudden rise in community viral load indicates a rise in community infection, and therefore a potential rise in case incidence.

10.4 Exploring Normality

One of the requirements to using Simple Control Charts – both Individuals and Moving Range – is that the unacted-upon data is normally distributed. In the context of epidemiologic Simple Control Charts, this means that the data points should be normally distributed (forming a bell curve) when they are not acted upon by other factors (e.g., outbreaks or viral/case changes). If the data series during these periods is normally distributed, it would indicate that only random chance (i.e., inherent variance) is acting on the system. The variance could then be assessed and control limits for acceptable thresholds set.

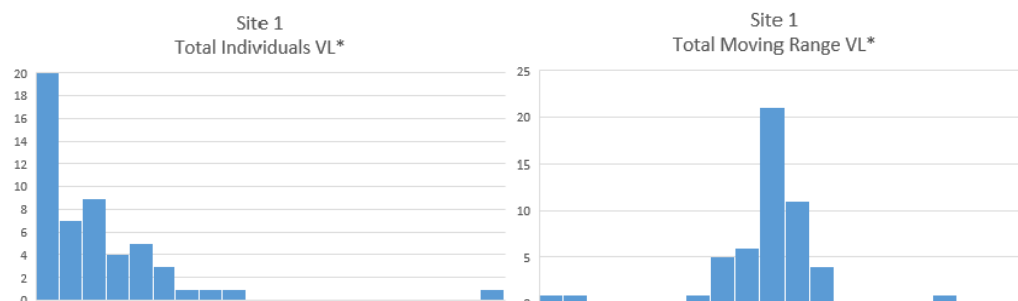
The mean, median, skewness, kurtosis, range divided by SD, and other normality metrics were calculated for each site's total viral load series, as well as the moving average viral load series. Site 1's descriptive statistics and histograms are shown below in Figure 10.4.1 as an example.



*Viral Load or Viral Load at LOD (if concentration is below LOD)

Figure 10.4.1 Site 1 total dataset descriptive statistics and skew histograms

For visual confirmation, the two sets above were visualized in two count histogram bars, dividing all points into 20 bins each. As seen below in Figure 10.4.2.



*Viral Load or Viral Load at LOD (if concentration is below LOD)

Figure 10.4.2 Site 1 total dataset descriptive statistics and skew histograms

The above histograms show the viral load skew to the left for all individuals, and a more symmetrical distribution for the moving range. When analyzing all data points of the Site 1's individual time series, there is a heavy skew of 3.077 and a high kurtosis of 11.140, indicating low symmetry and heavy tails/outliers. This would make sense, for three reasons:

1. Physiology of viral shedding is not instantaneous (i.e., an infected person may only be identified as a case once, but will continue to shed into the wastewater days to months before and after).
2. Viral loads are calculated in a way which would artificially stack the values to the left (i.e., when the viral concentration is below the LOD, the viral load is then calculated using the LOD concentration).
3. There were certainly COVID-19 outbreaks that occurred during that time frame which would inflict far heavier viral loads than would be expected by chance alone.

When analyzing all data points of the Site 1's moving range time series, there is a nearly non-existent skew of 0.088 and a slightly less-high kurtosis of 8.080, indicating high symmetry and heavy tails/outliers. The improvement makes sense, as the moving range reduces much of the first two above issues. These descriptive statistics were calculated for all SMP sites, as seen below in Table 10.4.1. Red highlighted cells indicate less normal values.

Table 10.4.1 By-site descriptive statistics for all sites, individual and moving range average viral loads

#	SITE	Samples	TOTAL					TOTAL				
			Individuals - Viral Load*					Moving Range - Viral Load* Change				
			mean	median	skew	kurtosis	range/SD	mean	median	skew	kurtosis	range/SD
1	AlexRenew	47	8.71E+12	5.70E+12	3.08	11.14	5.05	1.00E+11	4.12E+11	0.09	8.08	7.42
2	Aquia	51	2.15E+12	1.57E+12	3.95	17.04	5.72	-2.61E+10	4.81E+10	-2.39	17.10	8.46
3	Basham	41	1.43E+11	8.88E+10	1.39	1.42	3.78	-1.05E+09	-8.10E+08	-0.35	2.00	5.50
4	Blacksburg	45	1.32E+12	9.62E+11	1.85	4.13	4.79	2.93E+10	2.12E+10	0.62	4.40	6.77
5	Danville	43	3.60E+11	1.70E+11	3.65	15.08	5.60	-5.75E+08	-4.46E+09	-1.11	8.39	7.03
6	Coeburn	50	2.60E+11	2.00E+11	1.63	2.70	4.23	-2.17E+09	-2.18E+10	0.81	2.59	5.67
7	HLM	50	5.55E+12	3.48E+12	3.98	18.97	6.18	5.30E+10	-6.32E+10	-0.12	10.47	8.17
8	NA	40	4.41E+12	1.88E+12	5.86	37.59	6.85	-2.22E+10	-1.17E+10	1.08	22.84	9.56
9	ON	47	1.00E+11	8.33E+09	4.84	25.96	6.16	1.42E+09	1.19E+09	0.21	11.18	8.12
10	VIP	47	4.11E+12	1.83E+12	3.76	14.01	5.06	-2.08E+10	-4.46E+10	3.92	25.53	8.02
11	WB	47	2.12E+12	1.10E+12	3.25	12.50	5.58	1.06E+11	3.11E+10	1.28	8.96	7.35
12	LFR	49	1.14E+12	7.63E+11	2.14	4.86	4.63	-1.14E+10	-7.95E+09	0.08	2.35	5.91
13	LB	50	2.03E+12	1.23E+12	2.66	8.31	4.87	-8.51E+10	-2.89E+10	0.31	2.06	5.79
14	Moores	49	1.95E+12	1.27E+12	2.27	5.86	4.76	5.48E+10	1.47E+11	-0.33	1.39	5.17
15	NorthRiver	47	9.81E+11	4.31E+11	3.56	13.72	5.06	-2.84E+10	-2.82E+09	0.59	13.57	8.73
16	Parkins	50	4.56E+11	1.54E+11	2.55	7.23	5.05	-3.65E+08	1.06E+09	-0.11	5.62	6.64
17	Peppers	50	5.40E+11	2.91E+11	2.14	4.60	4.56	1.30E+09	-1.66E+10	0.76	2.27	5.60
18	Pound	47	2.62E+11	5.37E+10	4.41	20.19	5.72	2.75E+08	-1.59E+09	0.62	11.70	8.14
19	Richmond	50	5.49E+12	2.66E+12	2.75	7.90	4.86	1.10E+10	-7.16E+10	0.40	2.84	5.71
20	Roanoke	51	5.83E+12	4.49E+12	4.02	21.54	6.53	3.51E+10	2.52E+11	0.29	8.56	7.84
21	Rutledge	36	2.07E+10	5.37E+09	2.40	5.01	4.11	2.86E+08	3.16E+07	0.53	3.49	6.04
22	CB	40	6.01E+11	4.95E+11	1.49	2.56	4.60	-2.36E+09	4.13E+09	0.28	1.65	5.33
23	UOSA	43	8.68E+12	4.34E+12	4.49	22.83	6.22	1.17E+11	-1.56E+11	3.25	21.08	8.23
24	Wolf	50	5.00E+11	3.67E+11	1.78	3.49	4.75	4.43E+08	-7.17E+09	-0.53	2.53	5.94
AVG		46.7	2.40E+12	1.40E+12	3.079	12.028	5.197	1.29E+10	2.00E+10	0.424	8.360	6.964

*Viral Load or Viral Load at LOD (if concentration is below LOD)

**Red highlighted cells indicate above +/- 0.5 skew, 5 kurtosis, or a range/SD ratio score outside of 4-6

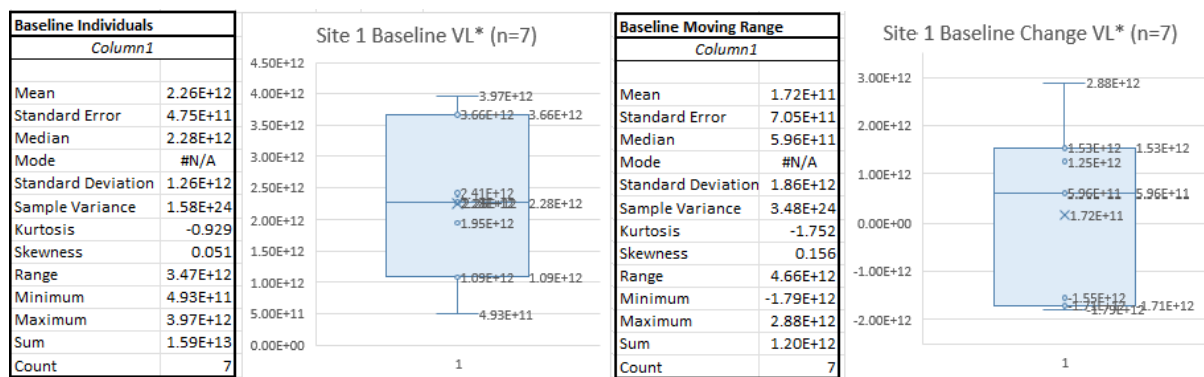
Examining the above table shows an average 3.079 skew, 12.028 kurtosis, and 5.197 range/SD for all site individuals. The all-sites moving range showed an average 0.424 skew, 8.360 kurtosis, and 6.964 range/SD. Red highlighted cells indicate poor indicators of normality, as defined by a greater than ± 0.5 skew, greater than 5.0 kurtosis, and a range/SD ratio score outside of 4-6.

As expected, the total site individuals are heavily skewed and have high kurtosis. This is most likely due to the reasons listed above in the Site 1 example. However, the viral load moving range values are much more in line with normality, with an average skew within tolerances and a drastic reduction of kurtosis. By site, 11 of the 24 sites were within ± 0.5 tolerances of skew, with an additional 7 sites within ± 1.0 skew. 11 of the 24 sites were within ± 5.0 tolerances of kurtosis. 12 sites had their range divided by SD within the 4-6 spectrum. 6 sites total passed by the skew, kurtosis, and range/SD test. The moving range values of the total dataset are much closer to normality than their viral load counterpart.

These descriptive statistics were also calculated for the proposed baselines for each site, the methods of which are described in the next Chapter. While I would normally be inclined to put the development of baseline methods before the normality results in any scientific report, I specifically gave the baseline establishment its own. Chapters 11 and 12 should be the chapters jurisdictions can jump to, should they attempt to replicate the proposed alert system.

Note: not all sites received the full 7 weeks of baseline data points, as some sites did not have perfect submission counts. No outliers were removed for analysis.

Site 1's baseline descriptive statistics and histograms are shown below in Figure 10.4.2 as an example.



*Viral Load or Viral Load at LOD (if concentration is below LOD)

Figure 10.4.2 Site 1 baseline dataset descriptive statistics and skew histograms

As shown above, when analyzing baseline data points of the Site 1's individual time series (n=7), there is a nearly non-existent skew of 0.051 and a low kurtosis of -0.929, indicating high symmetry and a normal bell curve. This shows that the Site 1 baseline viral load data meet normality standards.

When analyzing baseline data points of the Site 1's moving range time series, there is also a nearly non-existent skew of 0.156 and a slightly greater kurtosis of -1.752, indicating high symmetry and minimal tail irregularities/outliers. Both the individuals and moving range sets for Site 1 passed the normality test (less than +/- 0.5 skew and +/- 5.0 kurtosis), while both sets for their overall total dataset counterparts failed.

The baseline descriptive statistics were calculated for all sites, as seen below in Table 10.4.3.

Table 10.4.3 By-site descriptive statistics for all sites, individual and moving range average viral loads

			BASELINE					BASELINE				
			Individuals - Viral Load*					Moving Range - Viral Load* Change				
#	SITE	n	mean	median	skew	kurtosis	range/SD	mean	median	skew	kurtosis	range/SD
1	AlexRenew	7	2.26E+12	2.28E+12	0.05	-0.93	2.76	1.72E+11	5.96E+11	0.16	-1.75	2.50
2	Aquia	7	1.27E+12	1.03E+12	0.35	-1.40	2.57	1.05E+11	2.65E+11	-0.04	-1.44	2.56
3	Basham	7	6.66E+10	1.57E+10	1.07	-0.93	2.31	2.21E+10	-1.29E+09	1.89	3.79	2.98
4	Blacksburg	7	2.92E+11	2.13E+11	0.92	-0.80	2.60	8.00E+09	4.81E+10	-0.08	-1.27	2.64
5	Danville	7	1.34E+11	2.80E+10	2.53	6.46	2.70	8.78E+10	1.59E+09	2.45	6.12	2.77
6	Coeburn	7	2.65E+11	2.43E+11	0.31	-1.12	2.62	1.29E+10	6.04E+09	-0.25	-1.31	2.67
7	HLM	7	1.11E+12	6.13E+11	1.82	3.13	2.81	-1.07E+11	-1.23E+11	-0.31	1.18	3.22
8	NA	5	1.80E+12	1.85E+12	-1.06	2.51	2.76	-1.15E+11	-7.45E+10	-0.20	-0.86	2.57
9	ON	6	1.27E+11	3.11E+09	2.45	6.00	2.46	2.44E+08	3.32E+08	0.00	2.50	3.16
10	VIP	5	1.41E+12	1.74E+12	-0.41	-3.01	2.13	-1.94E+11	-2.02E+11	-0.17	1.35	2.78
11	WB	5	4.38E+11	3.01E+11	0.28	-2.58	2.27	-7.71E+10	2.48E+11	-0.51	-3.06	2.10
12	LFR	7	6.03E+11	4.04E+11	0.31	-2.03	2.55	8.02E+10	2.05E+11	-1.54	2.90	2.95
13	LB	7	8.67E+11	3.05E+11	2.27	5.24	2.73	-6.33E+10	1.11E+10	0.04	1.55	3.29
14	Moores	7	4.40E+11	4.47E+11	1.56	2.89	2.94	-6.45E+10	-2.80E+11	1.46	2.30	2.98
15	NorthRiver	7	3.17E+11	3.40E+11	-1.07	1.31	3.06	-2.71E+10	-7.28E+10	0.50	-0.61	2.77
16	Parkins	7	4.79E+11	4.39E+11	0.86	-0.22	2.65	8.83E+10	-3.73E+09	0.77	1.70	3.24
17	Peppers	7	2.83E+11	1.63E+11	2.12	4.78	2.88	1.11E+11	7.58E+10	0.95	1.60	3.14
18	Pound	6	9.22E+10	7.64E+10	1.04	0.67	2.63	-4.56E+09	-6.21E+10	1.19	0.21	2.50
19	Richmond	7	2.19E+12	2.35E+12	0.14	-2.16	2.27	5.68E+11	2.22E+11	-0.95	1.04	2.96
20	Roanoke	7	3.67E+12	3.50E+12	0.53	-1.19	2.61	7.36E+10	1.37E+12	-1.40	1.29	2.84
21	Rutledge	7	2.41E+09	1.39E+09	0.44	-2.45	2.16	6.36E+07	-2.60E+08	0.33	-0.90	2.74
22	CB	6	2.14E+11	1.56E+11	1.47	1.69	2.60	6.23E+10	1.68E+10	0.77	1.85	3.03
23	UOSA	7	1.87E+12	2.02E+12	0.22	0.95	3.19	-5.40E+09	6.72E+10	-0.09	-2.17	2.46
24	Wolf	7	2.35E+11	2.42E+11	-0.85	2.02	3.23	-7.82E+10	-3.93E+10	-1.59	3.81	3.24
AVG		6.63	8.51E+11	7.81E+11	0.723	0.784	2.646	2.73E+10	9.49E+10	0.140	0.827	2.837

*Viral Load or Viral Load at LOD (if concentration is below LOD)

**Red highlighted cells indicate above +/- 0.5 skew, 5 kurtosis, or a range/SD score outside of 4-6

Examining the above table shows an average 0.723 skew, 0.784 kurtosis, and 2.646 range/SD for all baseline site individuals. The all-sites moving range showed an average 0.140 skew, 0.827 kurtosis, and 2.837 range/SD ratio. Red highlighted cells indicate poor indicators of normality, as defined by a greater than +/- 0.5 skew, greater than 5.0 kurtosis, and a range/SD ratio score outside of 4-6.

Both the baseline individuals and moving ranges had a drastic reduction in skew and kurtosis, as compared to the total dataset (Table 10.4.2). When comparing the number of skew and kurtosis passed sites, nine baseline individuals sites passed (as compared to zero from the total individuals set) and twelve baseline moving range sites passed (as compared to six from the total moving range set). This reduction in skew and kurtosis is a good sign that the identified

baseline is capturing normalized data, and therefore, data less affected by external factors. It shows the merits of both the baseline set, as well as the moving range set.

The range/SD ratio reduction by half would indicate that the baseline data points are more centered and clustered about the mean, which one would expect from a smaller and more tightly controlled dataset.

With baseline set normality confidence, especially the moving average baseline site, the use of the simple control charts (in particular, the I-MR variety) is justified. However, it is important to note that not all sites behaved the same. For instance, no Site 5 sets passed the normality tests.

10.5 Proposed Model

In summary, this dissertation established the following for argument underpinning for any proposed alert system:

- discussed need of a viral alert system (Chapter 1)
- reasoning for by-site variability analysis in WWS (Chapters 3 and 7)
- adoption of the Yang conceptual framework of an Early Warning System (Chapter 4)
- correlation between VDH WWSP viral load data and case counts (Chapter 8)
- independent and dependent target variables set as viral load change and case count change (Chapter 10.1)
- explored historical and current frameworks for alert systems (Chapters 10.2 and 10.3)
- baseline periods set using case counts and variant abundance (Chapters 11.1 and 11.2)
- viral load normality checks for all individuals and moving range data (Chapter 10.4)

In the interest of aligning the goals of a proposed alert system (e.g., being an easy-to-use and direct method for public health practitioners), with the style of surveillance data (e.g., short

history linear time series with site-specific variability), the style of a (site-specific) I-MR Control Chart was selected as the alert system base (Ch. 10.2).

The final proposed model includes the following principles:

1. Sites will be assessed independently from each other.
2. Viral Load* (or Viral Loads at LOD for all concentrations below LOD) change will be the base metric, used to indicate if there is a likelihood of an imminent change in cases.
3. A baseline period will be established, using a period where case counts are low/plateaued and variant abundance is stable.
4. An Upper Control Limit will be created for each site, i.e, a multiplier (usually set at 2 or 3) times the standard deviation of baseline viral load moving range. This indicates whether a site is *out of control*, given its baseline inherent variability.
5. If a site is *out of control* for a given week's data point, an anomaly detection flag will be assigned as an alert.
6. As data is collected, model evaluation will include calculating sensitivity and specificity. Thresholds for system variability multipliers can be adjusted to the needs of the user.

Chapter 11: Establishing Baseline & Applying Alerts

This dissertation chapter gives details on how the SARS-CoV 2 Wastewater Surveillance (WWS) alert system was created, as well as a few tips for prospective readers hoping to use the model for their own WWS. As such, this chapter contains the public health practitioner's need-to-know for creating this alert system. However, I also recommend reading the next chapter, as it provides the how-to on evaluating the Alert System. Creation and Evaluation should go together with data collection.

This chapter begins with how baselines for each site were created and then uses those baselines to create appropriate Upper Control Limits (UCL) for each site. The UCLs set the threshold for alert, so any time a data point goes above the UCL an alert flag is assigned.

Tip for readers: while I primarily used Excel for all calculations, the calculations are simple enough to be performed on standard statistical software.

Note: the paired WWS and case data used for the following analyses only use up to the SMP Wk53 data. This was due to several reasons, including accounting for case data lag at the time of initial data download, as well as drops of several sites from the program and switching sampling cadence for several others around that time.

Tip for readers: beware of major changes to the program that will complicate paired analyses.

11.1 Establishing Baseline

While ‘baselines’ can be open to interpretation, in this dissertation it will be considered as a period where minimal external factors are affecting the system. After all, “the engineer’s problem is to eliminate all assignable or findable causes of variation” (Shewhart, 1931, p. 266). Chapter 3 discussed the significant inherent variability found with WWS generally, but there are some clear external factors which contribute to the rise of viral particles in sewage. The most direct is an actual disease outbreak (the very thing this alert is trying to detect), but other external pressures can cause changes in viral load (e.g., sewage infrastructure or new virus variants).

The goal of the baseline is to find a period of minimal external changes, as a marker of when the data collected should only represent inherent variability in the WWS system itself. This means finding lull periods in the data where disease outbreak was at a minimum and few other changes were occurring.

Tip for readers: my WWS data is fortunate that the same participating utility sites and laboratory continued data collection and analysis with consistent methodology. This may not be true for you. Be aware of methodology or case definition changes.

VDH Case and Variant Data

To get a good overview of Virginia cases and variants, the VDH publicly-available case and variant data (VDH, 2023b) was downloaded and visualized (see Chapters 6 & 7). Case data showed several peaks throughout the entirety of the COVID-19 pandemic period (to date), with the strongest peak around December 2021. The major variants of interest were selected using the FDA WWS Variant Proportions national trend visualization (FDA, 2023) for context (see Chapter 7), and then pulling appropriate VDH variant clinical counts.

When combining in the VDH WWS sewershed case data, one can view specifically the major dominant variants during the data collection period. The overlay of total sewershed case counts over time reveals cases increased with variant shifts. The FDA WWS Variant Proportions visualization, found on their dashboard main page (FDA, 2023).

Figure 7.3.5 outlines general WWS trends across the 1.5-year spread of the WWS dataset covered, shifting into and out of five different SARS-CoV-2 variants. Of particular interest, the shift from Delta to Omicron-BA.1 saw a substantial increases in case counts. To identify an appropriate baseline period, case count spikes and variants need to be controlled.

Controlling Cases

In Chapter 7, I pulled daily case counts for all SMP sewersheds during the testing period shown in Figure 7.3.4. To get a slightly better view, and appreciate the changing nature of sudden rises and falls of cases, the change of total cases was visualized below in Figure 11.1.1.

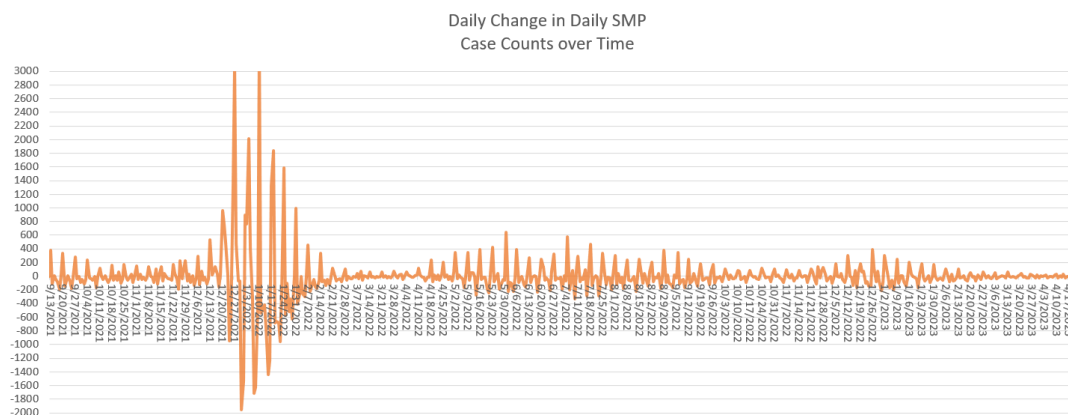


Figure 11.1.1: Daily changes of total SMP sewershed case counts over time

Identifying a ‘lull’ period requires choosing a threshold for maximum change and a threshold for minimum continuous time. For this dissertation, I selected two thresholds for maximum change: a stricter mobility of ± 100 cases change, and a more moderate mobility of ± 200 cases change. The threshold for a minimum continuous time was set at three consecutive weeks of behavior.

Tip for readers: the time frame and threshold minimums selected above were my personal judgement of the general trends of the datasets. You may choose different thresholds, depending on the data.

Applying the change and period thresholds above, the strict threshold of ± 100 cases change was marked in green boxes, while the moderate threshold of ± 200 cases change was marked in blue boxes, as shown below in Figure 11.1.2.

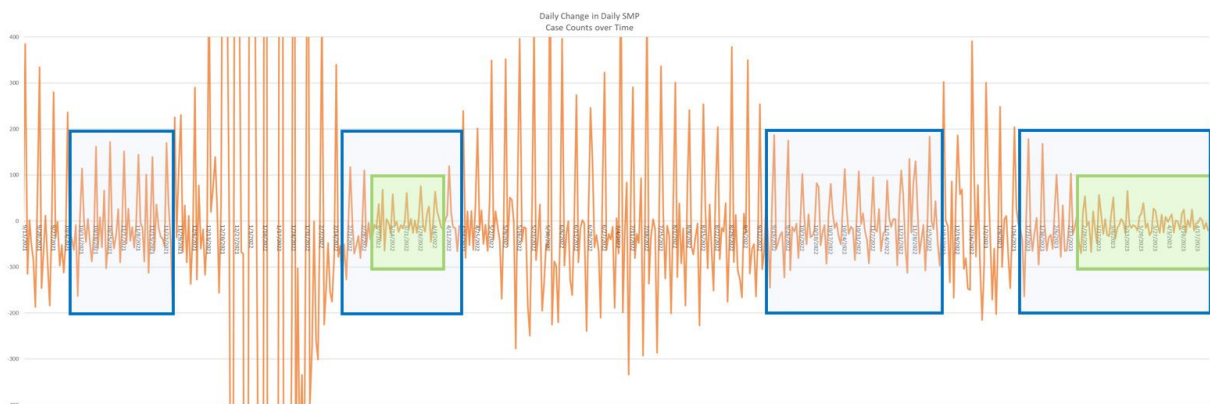


Figure 11.1.2: Isolation of acceptable periods using strict and moderate case change thresholds

The visualization above shows four potential time periods to train a baseline under the moderate case change threshold (blue boxes), and two under the strict case change threshold (green boxes). However, before deciding on the best baseline period, we should also remove any options which encompass changes in variant dominance.

Controlling Variants

The previously created Table 7.1.3 in Chapter 7 (informed by federal and state trends) showed the synthesized dominant variant timeline. Using this reference chart as a guide to when appropriate baselines can be pulled (i.e., when there are no dominant variant shifts), it can be crossed with the selected case change stability periods isolated in the above section.

The following Tables 11.1.2 and 11.1.3 cross the boxed daily change periods with their dominant variant status, for strict (± 100 daily case change) and moderate (± 200 daily case change) thresholds.

Table 11.1.1: *Isolated potential baseline periods with related dominant variants, strict case change threshold*

Time period	SMP Sampling Weeks	Dominant Variant (applicable Week #s)
3/7/22 – 4/4/22	Wk26 – 30	BA.1 (Wk26), none (Wks27 – 30)
2/20/23 – 4/17/23	Wk76 – 84*	XBB.1.5 (Wks77 – 81**)

*Wk84 is the last week of received SMP dataset

**Wk81 is the last week of received VDH variant case counts dataset

Table 11.1.2: *Isolated potential baseline periods with related dominant variants, moderate case change threshold*

Time period	SMP Sampling Weeks	Dominant Variant (% of weeks)
10/11/21 – 11/22/21	Wk5 – 11	Delta (Wks5 – 11)
2/21/22 – 4/11/22	Wk24 – 31	BA.1 (Wks24 – 26), None (Wks27 – 29), BA.2 (Wks30 – 31)
1/23/23 – 4/17/23	Wk72 – 84*	XBB.1.5 (Wks72 – 81**)

*Wk84 is the last week of received SMP dataset

**Wk81 is the last week of received VDH variant case counts dataset

The above results indicate that time period using **SMP Wks5 – 11** are ideal for baseline training, as those weeks experienced moderately minimal case changes, as well as consistent Delta dominance. There is a similar period of moderately lulled cases and variant dominance during the Wks72 – 81 timeframe, however, this period occurs at the end of the received collection data and would not be useful for predictive alert testing. A follow-up to this dissertation could include testing more recent data against a re-trained XBB.1.5 baseline.

Of note, baseline training periods were selected by total SMP case and VDH variant indicators. The training for baseline could also be performed by site-specific case and variant indicators, though variant indicators for individual sites will most likely be unavailable. While SMP site sequencing is currently being performed for WWS variant abundances at the site level, not all sites are tested each week and results have not been thoroughly evaluated in the context of state or national trends.

11.2 Organizing WWS and Case Counts by Site

If you don't already have a process in place for giving a viral load and weekly case count value for every site's weekly collection (by individual site), I recommend a simple Excel workbook, as seen in the screen of Figure 11.2.1 below.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S				
Sample #	sample_id	wsite_name	sample_collect_date	flow_rate	per_target_vir_c	one	VIRAL LOAD	below lod_sewa	LOD	LOD VIRAL LOAD	VL	DIFF	XSD	(-)/VL ALERT	WWS Sample	Case join	CASES NEXT 7 DAYS	Difference	XSD			
17	DCLS_SMP_WWS_WK1	Alexandria Renew Enterprises	2021-09-13	32.8	21875	2.72E+12	no	1350	1.68E+11	2.72E+12					2021-09-13	Alexandria Renew2021-09-13	973					
67	DCLS_SMP_WWS_WK2	Alexandria Renew Enterprises	2021-09-20	35.1	9500	1.26E+12	no	1350	1.79E+11	1.26E+12	-1.45E+12	5.18E+12	0	2021-09-20	Alexandria Renew2021-09-20	342	-35	136.1				
117	DCLS_SMP_WWS_WK3	Alexandria Renew Enterprises	2021-09-27	36.8	21875	3.05E+12	no	1350	1.88E+11	3.05E+12	1.78E+12	5.18E+12	0	2021-09-27	Alexandria Renew2021-09-27	316	-26	136.1				
159	DCLS_SMP_WWS_WK4	Alexandria Renew Enterprises	2021-10-04	34.8	5688	7.49E+11	no	1350	1.78E+11	7.49E+11	-2.30E+12	5.18E+12	0	2021-10-04	Alexandria Renew2021-10-04	316	0	136.1				
201	DCLS_SMP_WWS_WK5	Alexandria Renew Enterprises	2021-10-11	34.4	17500	2.28E+12	no	1350	1.76E+11	2.28E+12	1.53E+12	5.18E+12	0	2021-10-11	Alexandria Renew2021-10-11	263	-53	136.1				
247	DCLS_SMP_WWS_WK6	Alexandria Renew Enterprises	2021-10-18	33.1	3938	4.93E+11	no	1350	1.69E+11	4.93E+11	-1.79E+12	5.18E+12	0	2021-10-18	Alexandria Renew2021-10-18	228	-35	136.1				
297	DCLS_SMP_WWS_WK7	Alexandria Renew Enterprises	2021-10-25	40.3	7130	1.09E+12	no	1350	2.06E+11	1.09E+12	5.96E+11	5.18E+12	0	2021-10-25	Alexandria Renew2021-10-25	237	9	136.1				
347	DCLS_SMP_WWS_WK8	Alexandria Renew Enterprises	2021-11-01	35.6	29400	3.97E+12	no	1350	1.82E+11	3.97E+12	2.88E+12	5.18E+12	0	2021-11-01	Alexandria Renew2021-11-01	255	18	136.1				
395	DCLS_SMP_WWS_WK9	Alexandria Renew Enterprises	2021-11-08	32.9	19375	2.41E+12	no	1350	1.68E+11	2.41E+12	-1.55E+12	5.18E+12	0	2021-11-08	Alexandria Renew2021-11-08	303	48	136.1				
445	DCLS_SMP_WWS_WK10	Alexandria Renew Enterprises	2021-11-15	32.3	30000	3.66E+12	no	6200	7.57E+11	3.66E+12	1.25E+12	5.18E+12	0	2021-11-15	Alexandria Renew2021-11-15	306	3	136.1				
487	DCLS_SMP_WWS_WK11	Alexandria Renew Enterprises	2021-11-22	33.3	15475	1.95E+12	no	6200	7.82E+11	1.95E+12	-3.71E+12	5.18E+12	0	2021-11-22	Alexandria Renew2021-11-22	399	93	136.1				
533	DCLS_SMP_WWS_WK12	Alexandria Renew Enterprises	2021-11-29	32.9	72750	9.05E+12	no	6200	7.71E+11	9.05E+12	7.09E+12	5.18E+12	1	2021-11-29	Alexandria Renew2021-11-29	530	131	136.1				
583	DCLS_SMP_WWS_WK13	Alexandria Renew Enterprises	2021-12-06	32.4	12750	1.56E+12	no	6200	7.60E+11	1.56E+12	-7.48E+12	5.18E+12	0	2021-12-06	Alexandria Renew2021-12-06	685	155	136.1				
633	DCLS_SMP_WWS_WK14	Alexandria Renew Enterprises	2021-12-13	29.8	95625	1.08E+13	no	6200	6.99E+11	1.08E+13	9.22E+12	5.18E+12	1	2021-12-13	Alexandria Renew2021-12-13	2133	1450	136.1				
681	DCLS_SMP_WWS_WK15	Alexandria Renew Enterprises	2021-12-20	29.4	502182	5.59E+13	no	6200	6.90E+11	5.59E+13	4.51E+13	5.18E+12	1	2021-12-20	Alexandria Renew2021-12-20	5064	2929	136.1				
<	>	TOTAL	Se-Sp	CASES	WWS	AlexRenew	Aquia	Basham	BB	Darville	Coeburn	HLM	NA	+	-							
Ready		Accessibility	Investigate																			
															Display Settings							90

Figure 11.2.1: Excel workbook screenshot with by-site tabs for WWS-case count pairs and highlighted baseline

Tip for readers: viral loads are calculated (viral load = flow*concentration*3,785,411.8).

Consider using half LOD concentration as the base of the viral load calculation, when the concentration is below LOD.

When the baseline is determined (see above section – e.g., SMP Wks5-11), highlight the baseline dates, VL*, VL difference, weekly case counts, and weekly case count changes. Then run basic variance statistics, as seen below for Site 1 in Figure 11.2.2.

Site 1 Moving Range Baselines for VL* and Weekly Case Counts:

CHANGE VIRAL LOAD* BASELINE		CHANGE CASE BASELINE	
Samples #:	7	Samples #:	7
Average:	1.72E+11	Average:	11.9
Pop Variance:	2.98E+24	Pop Variance:	2.06E+03
Std Dev	1.73E+12	Std Dev	45.4
XSD	5.18E+12	XSD	136.1

*Viral Load or Viral Load at LOD (if concentration is below LOD)

**XSD is the SD multiplied by 3

Figure 11.2.2: Site 1 screenshot of baseline statistics for moving ranges of VL* and case count change, SD multiplier of 3

Note: “XSD” is the standard deviation chosen threshold multiplier (e.g., three times the standard deviation of the viral load moving range). This chapter’s examples all use three as the default multiplier for both viral load and case change thresholds, though changing this number allows for modifications for sensitivity and specificity later.

11.3 Assigning VL and Case Alerts

Using the XSD threshold (i.e., three), flags (alerts) can be assigned for every VL* and change of weekly case counts which rise above those values. This is our Upper Control Limits (UCL) for both changes of VL* and case counts. Now comparisons can be made regarding whether statistically significant changes of viral loads result in statistically significant changes over the next seven days of case counts. See Table 11.3.1 below for by-site baseline moving ranges and assigned alert flags.

Table 11.3.1: SMP by-site moving ranges and alerts, SD 3x multiplier

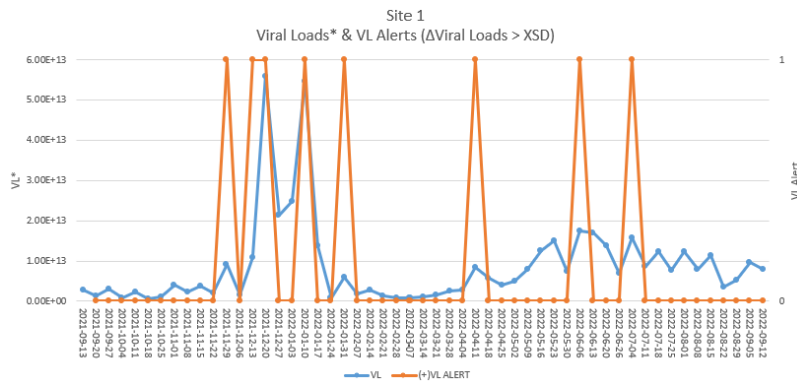
		Total	Baseline		Flags		
#	SITE	Samples	Samples	∂VL* SD	∂Case* SD	VL Flags	Case Flags
1	AlexRenew	53	7	1.73E+12	45.4	8	10
2	Aquia	52	7	9.38E+11	32.5	3	3
3	Basham	44	7	7.17E+10	4.9	3	4
4	Blacksburg	53	7	2.70E+11	5.1	6	15
5	Danville	46	7	1.92E+11	24.0	3	3
6	Coeburn	53	7	1.72E+11	13.5	2	4
7	HLM	53	7	1.49E+12	33.8	3	5
8	NA	49	5	2.83E+11	83.5	12	3
9	ON	48	6	4.30E+11	2.2	1	3
10	VIP	49	5	9.27E+11	84.2	3	2
11	WB	49	5	6.32E+11	31.5	3	4
12	LFR	52	7	5.17E+11	13.0	3	5
13	LB	52	7	1.74E+12	35.7	1	3
14	Moores	53	7	6.00E+11	16.4	7	10
15	NorthRiver	52	7	1.55E+11	27.4	7	4
16	Parkins	53	7	5.10E+11	9.6	1	4
17	Peppers	53	7	2.38E+11	12.7	5	6
18	Pound	51	6	1.29E+11	3.2	3	5
19	Richmond	53	7	1.72E+12	40.1	5	6
20	Roanoke	53	7	2.77E+12	77.9	3	3
21	Rutledge	45	7	2.17E+09	2.2	9	2
22	CB	44	6	1.84E+11	10.9	4	3
23	UOSA	51	7	1.02E+12	58.9	10	6
24	Wolf	53	7	2.09E+11	12.9	6	2
Site Average		50.58	6.63	7.05E+11	28.39	4.63	4.79

*Viral Load or Viral Load at LOD (if concentration is below LOD)

**XSD is the SD multiplied by 3

Ideally, the number of VL alerts should match the number of case alerts. The case alerts should follow the VL alert by same or next week (i.e., the next 7-14 days of cases). The addition of the second week of observation of any delayed case response is due to possible leading indicator effects of VL.

Graphing each site's VL* with their respective VL alerts reveals the relatedness of VL* increase to assigned alerts, as seen below in Figure 11.3.1.



*Viral Load or Viral Load at LOD (if concentration is below LOD)

**XSD is the SD multiplied by 3

Figure 11.3.1: Site 1 VL* and assigned VL alerts, SD 3x multiplier

The Site 1 example above shows the overlap of VL rises with assigned VL alerts. In general, an alert appears to have been issued any time there's a noticeable (3x SD) peak, capturing the Omicron peak of December 2021 and the summer peak of 2022. Now we need to check and see if the VL alerts coincide with assigned case rises (i.e., case alerts).

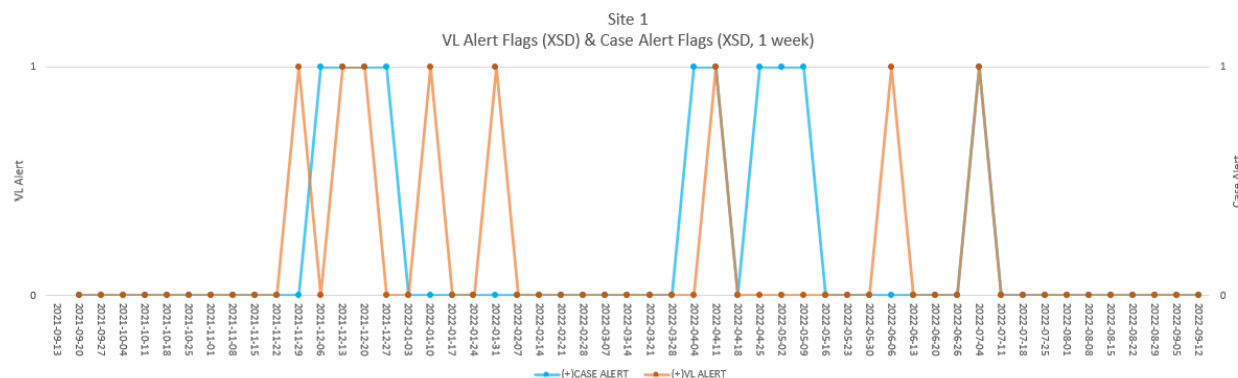
11.4 VL and Case Alert Matching

Determining if the VL alerts coincide with the case alerts is imperative for knowing how functional the alert system is. If cases don't imminently rise with VL alerts, the alert system simply doesn't work. Matching metrics must be explored, and the XSD thresholds can be adjusted to improve results. Sensitivity and specificity can be adjusted with XSD thresholds, depending on the need of the practitioner using this system.

Tip for readers: determine ahead of time if you want an overactive (i.e., sensitive) alert system which may deliver false positives, or a more reticent alert system (i.e., specific) that delivers fewer but truer positives.

The first step is to check the relationship visually, by plotting the alert flags together in time.

Figure 11.3.2 below shows the Site 1 assigned VL alerts and case alerts.



*XSD is the SD multiplied by 3

Figure 11.3.2: Site 1 assigned VL alerts and case alerts, SD 3x multiplier

The Site 1 example above shows the overlap of assigned VL alerts (orange) with case alerts (blue). The December 2021 exchanges show VL alerts coinciding with case alerts. However, the January 2022 exchanges showed two VL alerts which did not result threshold-level differences in cases. Conversely, the April 2022 exchanges showed more sensitivity to case alerts than VL alerts. Additionally, there was one errant VL alert on June 6h, 2022, and a perfectly timed VL-Case alert on July 4th, 2022. Otherwise, both alerts were quiet.

Using simple flags to determine whether the alerts match each other, additional metrics like sensitivity and specificity (Se-Sp) can also be generated. The statistical terms of sensitivity and specificity are common in medicine, as a means of determining how good a test is at detecting percentage true positives (sensitivity) and true negatives (specificity). Excel alert metrics for Site 1 are shown below in Figure 11.3.3.

MATCHING: VL Alert(XSD) & CASE Alert(XSD)	#	%
Samples #:	52	
VL Alert(XSD) & Case Alert (XSD, 1week) MATCH	42	80.8%
VL Alert(XSD) & Case Alert(XSD, 1-OR-2weeks) MATCH	45	86.5%

(+)VL ALERT		
# (+)VL ALERTS(XSD)	8	
# (+)VL Alerts(XSD) w/ (+)Case Alerts(XSD,1week)	4	50.0%
# (+)VL Alerts(XSD) w/ (+)Case Alerts(XSD,1-OR-2weeks)	5	62.5%

NO WARNING		
# VL SAFE	44	
#(+)VL SAFE(XSD) w/ (+)Case SAFE(XSD,1week)	38	86.4%
#(+)VL SAFE(XSD) w/ (+)Case SAFE(XSD,1-OR-2weeks)	40	90.9%

Sensitivity (true positives; of all case warns, % VL alert)		
# Case Alerts	10	
# Case Alerts + VL Alerts (-1 week)	4	40.0%
# Case Alerts + VL Alerts (-1-2 weeks)	6	60.0%

Specificity (true negatives; of all case safes, % VL safe)		
# Case SAFEs	42	
# Case Safes + VL SAFE (-1 week)	38	90.5%
# Case Safes + VL SAFES (-1-2 weeks)	42	100.0%

**"Safe" indicates a lack of respective VL/Case alert; "Warning" indicates an alert

**XSD is the SD multiplied by 3

Figure 11.3.3: Site 1 assigned VL alerts and case alerts matching and Se-Sp metrics, SD 3x multiplier

The Site 1 example above contains several useful metrics, including matching, sensitivity, and specificity. See descriptions of each below.

Matching

The matching metrics shown above show the overlap in agreement between VL and case alerts. I.e., if a VL alert status was either positive or negative, what was the percentage of times in which the same was true of the case alert status that same week? Or perhaps either that week or the following week? The extended time option allows for VL leading indicator effects. Both calculations are performed.

For example, if a positive VL alert was issued on Wk9, a same-week match would require a case alert to have been issued that same week (i.e., a flag that the next 7 days of case counts was higher than the UCL for the case baseline). Or, as with the 2-week match, a case alert could happen on either Wk9 or Wk10 for a positive match flag.

In the above Site 1 example (XSD=3), the same-week matching was 81%. The 2-week matching was 87%.

Positive VL Alert

This section considers just the positive VL alerts. It asks the question of how often the positive VL alerts are associated with positive case alerts. If issued VL alerts often result with nonplused case increases, that would not be a good alert system. The agreement of a positive VL alert with both its same week, as well as either/next week, case alerts were calculated.

Decreasing the case alert threshold limit will increase the sensitivity of picking up case increases.

In the above Site 1 example (XSD=3), the same-week positive VL to positive case alert matching was 50%. The 2-week matching was 63%. Which is to say that, whenever a VL alert was flagged, 63% of the time there was a matched rise in cases that week (or next) for that site.

Negative VL Alert

This section considers just the negative VL alerts, i.e., when there is no VL alert issued. It asks the question of how often the negative VL alerts are associated with negative case alerts (i.e., no VL or case alerts). If the lack of having VL alerts (i.e., VL “safe”) are often associated with case increases anyway, that would not be a good alert system. The agreement of a negative VL alert with both its same week, as well as either/next week, case alerts were calculated.

Increasing the case alert threshold limit will decrease the sensitivity of picking up case increases and therefore increase the negative VL alert matching.

In the above Site 1 example (XSD=3), the same week positive VL to positive case alert matching was 86%. The 2-week matching was 91%. Which is to say that, whenever a VL alert

was not issued, 91% of the time there was a matched “safe” status for cases either that week (or next) for that site.

Sensitivity

For sensitivity, the number of true positives (i.e., positive case increases) are backchecked with the associated tests (i.e., VL alerts at the 1-2 previous weeks) to see how many were correctly identified.

True Positives are positive case alerts which also had positive VL alerts.

False Negatives are positive case alerts which incorrectly had negative VL alerts.

$$\text{Sensitivity} = \frac{\text{number of True Positives}}{\text{number of True Positives} + \text{number of False Negatives}}$$

The higher the sensitivity, the lower the chance of making a Type II Error.

In the above Site 1 example (XSD=3), the number of positive case increases which were correctly identified (i.e., sensitivity) was 40% for the same week and 60% for either the same or previous week. Decreasing the threshold for VL alerts, or increasing the threshold for case alerts, will increase sensitivity.

Specificity

For specificity, the number of true negatives (i.e., negative case increases) are backchecked with the associated tests (i.e., VL alerts at the 1-2 previous weeks) to see how many were correctly identified.

True Negatives are negative case alerts which also had negative VL alerts.

False Positives are negative case alerts which incorrectly had positive VL alerts.

$$\text{Specificity} = \frac{\text{number of True Negatives}}{\text{number of True Negatives} + \text{number of False Positives}}$$

The higher the specificity, the lower the chance of making a Type I Error.

In the above Site 1 example (XSD=3), the number of negative case increases which were correctly identified (i.e., specificity) was 91% for the same week and 100% for either the same or previous week. Increasing the threshold for VL alerts will increase specificity.

Chapter 12: Evaluating and Adjusting the Alert System

This chapter provides the evaluation and adjustment need-to-know for potential health care practitioners wishing to apply their own WWS system. It describes what I did for my dissertation, as well as giving tips for readers.

Evaluation involves summarizing the data and calibrating the Upper Control Limit (UCL) -- as determined by the standard deviation multipliers of the VL and cases baseline moving ranges (XSD). As collection data is gathered over time, this evaluation can be performed again for recalibration. Also, if over time, better baselines become available, the baselines can be reformed to better train the system (see Ch.11).

Now, since you've organized and set by-site your data – complete with identified baselines and issued VL and case alerts (see Ch. 11) – it's time to evaluate how well the alerts are working. The metrics used in Ch.11.4 should be applied to each site individually and then sent to a summary sheet for easy evaluation.

Some adjustments with the thresholds may be required to meet the needs of your WWS program.

Tip for readers: whether you're using Excel or any other statistical program to accomplish this summary sheet task, it's a good idea to include an easy VL and case alert threshold input box which can allow you to change the summary stats over all sheets/calculations at will. That way, changing any of the thresholds and viewing those changes on the summary sheet will be much easier than changing everything by hand.

12.1 Summary Statistics

Create a summary sheet marking the Ch.11.4 metrics, including matching rates and Sensitivities and Specificities (Se-Sp). See Figure 12.1.1 below for an example of an evaluation summary sheet.

#	SITE	Thresholds		By-Site Alerts		By-Site Validity										Matching/Se-Sp
		XSD		Alerts		Matching		Positive VL Alert		Negative VL Alert		Sensitivity		Specificity		
		ΔVL*	ΔCase	VL	Case	1 week	1-2 weeks	1 week	1-2 weeks	1 week	1-2 weeks	1 week	1-2 weeks	1 week	1-2 weeks	
1	AlexRenew	3	3	8	10	80.8%	86.5%	50.0%	62.5%	86.4%	90.9%	40.0%	60.0%	90.5%	100.0%	76.3%
2	Aquia	3	3	3	3	96.2%	96.2%	66.7%	66.7%	98.0%	98.0%	66.7%	66.7%	98.0%	98.0%	86.9%
3	Basham	3	3	3	4	88.6%	93.2%	33.3%	33.3%	93.9%	98.0%	25.0%	50.0%	95.8%	100.0%	75.4%
4	Blacksburg	3	3	6	15	75.0%	82.7%	66.7%	66.7%	76.1%	84.8%	26.7%	40.0%	94.6%	97.3%	69.4%
5	Danville	3	3	3	3	95.6%	95.6%	66.7%	66.7%	98.0%	98.0%	66.7%	100.0%	98.0%	98.0%	92.3%
6	Coeburn	3	3	2	4	88.5%	94.2%	0.0%	0.0%	92.0%	98.0%	0.0%	0.0%	95.8%	100.0%	63.1%
7	HLM	3	3	3	5	92.3%	96.2%	66.7%	66.7%	93.9%	98.0%	40.0%	40.0%	97.9%	97.9%	77.4%
8	NA	3	3	12	3	73.5%	75.5%	8.3%	8.3%	95.0%	97.5%	33.3%	66.7%	77.6%	95.9%	70.4%
9	ON	3	3	1	3	95.7%	100.0%	100.0%	100.0%	96.1%	100.0%	33.3%	33.3%	100.0%	100.0%	77.1%
10	VIP	3	3	3	2	93.8%	93.8%	33.3%	33.3%	98.0%	98.0%	50.0%	50.0%	96.0%	100.0%	80.6%
11	WB	3	3	3	4	93.8%	95.8%	66.7%	66.7%	95.9%	98.0%	50.0%	75.0%	97.9%	100.0%	85.4%
12	LFR	3	3	3	5	92.3%	94.2%	66.7%	66.7%	93.9%	95.9%	40.0%	40.0%	97.9%	100.0%	77.4%
13	LB	3	3	1	3	96.2%	96.2%	100.0%	100.0%	96.1%	96.1%	33.3%	33.3%	100.0%	100.0%	76.5%
14	Moore	3	3	7	10	82.7%	92.3%	57.1%	71.4%	86.7%	95.6%	40.0%	40.0%	92.9%	100.0%	74.6%
15	NorthRiver	3	3	7	4	86.5%	90.4%	28.6%	42.9%	95.6%	97.8%	50.0%	100.0%	89.6%	100.0%	86.1%
16	Parkins	3	3	1	4	94.2%	96.2%	100.0%	100.0%	94.1%	96.1%	25.0%	25.0%	100.0%	100.0%	73.4%
17	Peppers	3	3	5	6	86.5%	94.2%	40.0%	60.0%	91.5%	97.9%	33.3%	66.7%	93.5%	97.8%	78.7%
18	Pound	3	3	3	5	84.0%	92.0%	0.0%	0.0%	89.8%	98.0%	0.0%	0.0%	93.6%	100.0%	61.6%
19	Richmond	3	3	5	6	90.4%	94.2%	60.0%	60.0%	93.6%	97.9%	50.0%	83.3%	95.7%	100.0%	85.6%
20	Roanoke	3	3	3	3	96.2%	96.2%	66.7%	66.7%	98.0%	98.0%	66.7%	66.7%	98.0%	100.0%	87.3%
21	Rutledge	3	3	9	2	77.3%	79.5%	0.0%	0.0%	95.3%	100.0%	0.0%	0.0%	82.0%	98.0%	56.1%
22	CB	3	3	4	3	93.0%	93.0%	50.0%	50.0%	97.9%	97.9%	66.7%	100.0%	95.9%	98.0%	91.1%
23	UOSA	3	3	10	6	80.0%	82.0%	30.0%	30.0%	92.9%	95.2%	50.0%	50.0%	84.8%	100.0%	74.5%
24	Wolf	3	3	6	2	88.5%	90.4%	16.7%	16.7%	97.8%	100.0%	50.0%	50.0%	90.0%	100.0%	78.1%
AVG		3	3	4.63	4.79	88.4%	91.7%	48.9%	51.5%	93.6%	96.9%	39.0%	51.5%	94.0%	99.2%	77.3%

Viral Load or Viral Load at LOD (if concentration is below LOD), expressed as moving range (i.e., change of viral load or cases)

**Red highlighted cells indicate values below 50%

Figure 12.1.1 Evaluation summary statistics for all SMP sites

The summary sheet above shows all by-site SMP site evaluation metrics, as well as the total averages at the bottom. Validity cells with less than 50% are highlighted. The rightmost column averages all the by-site validity values into one metric, providing an *over-simplified* metric for better or worse validity.

Of the 24 sites, all were calibrated at a default XSD multiplier threshold of three (for both VL and case moving range baselines) and shown in the ‘XSD’ columns. Using those XSDs, the number of VL and case alerts issued are shown in the ‘Alerts’ columns. Average VL alerts were

at 4.6, and alert alerts were at 4.8. Ideally, the alert numbers should be approximately equal. However, when looking at the site breakdowns, there was substantial variability in alert counts between sites. For instance, Site 8 had 12 VL alerts, and Site 4 had 15 case alerts. Of important note, the since the number of VL and case alerts were as low as 1-2 per site during the tested time frame, the positive VL alerts and sensitivity success percentage markers are subject to wide jumps up and down.

When examining the validity statistics, average matching was 88.4% for one week, and 91.7% for up to two weeks. Average positive VL alert success was 48.9% for one week, and 51.5% for up to two weeks. Average negative VL alert success was 93.6% for one week, and 96.9% for up to two weeks. Average sensitivity was 39.0% for one week, and 51.5% for up to two weeks. Average specificity was 94.0% for one week, and 99.2% for up to two weeks. The average for all validities for all sites (from matching to specificities) was 77.3%.

This summary table shows an alert system with high matching, low-mid sensitivity, and high specificity. This is to say that there is a strong relationship between VL alerts and impending Upper Control Limit (UCL) case rises (matching of 92%). Even though the accuracy of correctly identifying no UCL case rises is very high (specificity of 99%), the VL alerts may only detect about half of imminent UCL case rises (sensitivity of 52%). Of the VL alerts issued, half resulted in UCL case rises (positive VL alert correctness of 52%). Of all the negative VL alerts issued, nearly all resulted in no UCL case rises (negative VL alert correctness of 99%).

All in all, as the thresholds stand here, the alert system is very good at saying whether there will *not* be imminent case rise. It's about evenly split on how successful positive alerts are.

To adjust these values (i.e., changing the VL and case XSD thresholds to adjust the UCL), see the next section.

12.2 Adjusting XSD

The moving range baseline variance we assessed in 11.1 for both VL and cases allowed us to select a working threshold multiplier (XSD) for creation of an UCL. However, both XSDs for VL and cases can be adjusted. Decreasing the VL XSD will allow for a more sensitive alert detection, while an increased case XSD will increase the amount of case change we want to detect. For instance, if we only care about very large imminent increases in cases, we can set the threshold to 4 or higher.

The following table is the all-site average evaluation metrics for all 0.5 increments of XSD, ranging from 0.5 to 4.0 XSD. The red-to-green highlights indicate in-column spectrum from lowest (red) to highest (green) percentages. Ideally, the best alert system should have high percentage values in all columns. However, choices will have to be made about whether to prioritize certain functionalities over others. This prioritization heavily depends on the needs of the public health practitioners using this alert system.

The goal of this dissertation is to create a system with relatively high-performance metrics for detection of a fairly substantial rise in cases. For this reason, the XSD of three was used for UCL case alerts for initial exploration. All site evaluation metrics and their XSD increments are shown below in Table 12.2.1. Color gradients indicate which values are stronger (green) and weaker (red) in each column.

Table 12.2.1 SMP all-site average evaluation metrics for XSD increments

Thresholds		By-Site Alerts				By-Site Validity											
XSD		Alerts				Matching		Positive VL Alert		Negative VL Alert		Sensitivity		Specificity		General	
ΔVL**	ΔCase	VL	Case			1 wk	1-2 wks	1 wks	1-2 wks	1 wk	1-2 wks	1 wk	1-2 wks	1 wk	1-2 wks	Average	
0.5	0.5	15.3	16.1			63.5%	78.0%	43.8%	59.1%	73.8%	87.5%	40.6%	64.4%	75.4%	94.4%	68.0%	
0.5	1.0	15.3	11.6			67.9%	78.7%	36.3%	50.1%	82.9%	92.3%	46.6%	71.9%	75.3%	94.7%	69.7%	
0.5	1.5	15.3	8.5			69.7%	77.3%	29.4%	40.4%	88.4%	94.5%	50.8%	75.9%	74.6%	94.5%	69.6%	
0.5	2.0	15.3	6.9			69.6%	76.1%	23.8%	32.6%	90.5%	95.9%	49.9%	73.9%	73.6%	94.3%	68.0%	
0.5	2.5	15.3	6.0			70.1%	75.5%	21.9%	30.0%	92.1%	96.3%	52.1%	76.6%	73.5%	94.3%	68.2%	
0.5	3.0	15.3	4.8			70.7%	74.8%	18.6%	25.3%	94.1%	97.0%	55.9%	80.3%	73.1%	94.1%	68.4%	
0.5	3.5	15.3	4.0			70.7%	74.1%	16.6%	21.5%	95.3%	97.9%	56.0%	80.2%	72.9%	94.0%	67.9%	
0.5	4.0	15.3	3.5			71.0%	73.5%	15.5%	18.9%	96.3%	98.3%	58.6%	78.7%	72.8%	93.9%	67.8%	
1.0	0.5	11.4	16.1			65.7%	80.0%	46.2%	59.9%	72.8%	86.7%	31.6%	51.8%	82.4%	96.8%	67.4%	
1.0	1.0	11.4	11.6			71.6%	82.3%	39.2%	51.1%	81.9%	92.0%	37.2%	59.4%	82.4%	97.1%	69.4%	
1.0	1.5	11.4	8.5			74.4%	81.8%	31.5%	41.0%	87.5%	94.3%	41.3%	63.2%	81.8%	97.3%	69.4%	
1.0	2.0	11.4	6.9			75.6%	81.7%	26.8%	34.5%	90.2%	95.8%	43.9%	63.4%	81.3%	97.2%	69.0%	
1.0	2.5	11.4	6.0			76.3%	81.4%	24.6%	32.2%	91.8%	96.3%	45.7%	65.8%	81.1%	97.3%	69.2%	
1.0	3.0	11.4	4.8			77.4%	81.3%	22.1%	28.6%	93.9%	97.0%	50.2%	70.0%	80.8%	97.4%	69.9%	
1.0	3.5	11.4	4.0			77.7%	81.0%	20.1%	25.5%	95.1%	97.8%	50.9%	71.7%	80.6%	97.3%	69.8%	
1.0	4.0	11.4	3.5			78.2%	80.5%	19.0%	22.6%	96.1%	98.1%	53.6%	70.3%	80.6%	97.1%	69.6%	
1.5	0.5	8.8	16.1			67.1%	81.4%	50.2%	60.2%	72.4%	86.7%	26.3%	41.4%	87.2%	97.9%	67.1%	
1.5	1.0	8.8	11.6			73.6%	84.2%	41.6%	50.8%	81.3%	91.8%	30.8%	47.0%	87.0%	98.1%	68.6%	
1.5	1.5	8.8	8.5			77.4%	84.6%	34.5%	42.0%	87.1%	94.1%	35.2%	51.5%	86.5%	98.1%	69.1%	
1.5	2.0	8.8	6.9			79.4%	85.1%	30.9%	36.8%	90.1%	95.6%	39.7%	54.7%	86.4%	98.1%	69.7%	
1.5	2.5	8.8	6.0			80.3%	84.9%	28.7%	33.9%	91.7%	96.1%	41.2%	56.3%	86.2%	98.2%	69.7%	
1.5	3.0	8.8	4.8			81.9%	85.3%	26.7%	31.5%	93.9%	97.1%	48.0%	64.4%	86.0%	98.2%	71.3%	
1.5	3.5	8.8	4.0			82.2%	85.1%	24.4%	28.3%	95.1%	97.8%	48.2%	65.0%	85.7%	98.2%	71.0%	
1.5	4.0	8.8	3.5			82.7%	84.9%	22.8%	25.6%	96.0%	98.1%	50.9%	65.8%	85.6%	98.1%	71.0%	
2.0	0.5	7.1	16.1			68.5%	82.4%	55.9%	61.6%	72.3%	86.7%	23.1%	34.2%	90.5%	98.4%	67.4%	
2.0	1.0	7.1	11.6			75.8%	86.1%	49.0%	53.5%	81.4%	92.1%	28.3%	41.0%	90.4%	98.6%	69.6%	
2.0	1.5	7.1	8.5			79.8%	86.9%	41.7%	45.0%	87.0%	94.4%	32.7%	45.2%	89.8%	98.7%	70.1%	
2.0	2.0	7.1	6.9			82.0%	87.5%	37.9%	40.7%	90.0%	95.8%	37.4%	48.2%	89.6%	98.6%	70.8%	
2.0	2.5	7.1	6.0			82.8%	87.6%	35.0%	38.6%	91.5%	96.3%	38.3%	51.0%	89.3%	98.6%	70.9%	
2.0	3.0	7.1	4.8			84.5%	88.1%	33.3%	36.6%	93.8%	97.2%	45.0%	58.5%	89.2%	98.7%	72.5%	
2.0	3.5	7.1	4.0			85.0%	87.9%	31.0%	33.2%	95.0%	97.9%	45.4%	58.8%	88.9%	98.6%	72.2%	
2.0	4.0	7.1	3.5			85.7%	87.9%	29.9%	31.7%	95.9%	98.2%	49.0%	61.5%	88.8%	98.5%	72.7%	
2.5	0.5	5.9	16.1			68.5%	82.6%	56.0%	60.5%	71.8%	86.3%	19.6%	28.3%	92.1%	98.6%	66.4%	
2.5	1.0	5.9	11.6			76.0%	86.6%	50.8%	54.7%	80.7%	91.5%	24.4%	33.9%	92.1%	98.8%	69.0%	
2.5	1.5	5.9	8.5			81.2%	88.3%	47.4%	49.1%	86.8%	94.2%	30.4%	40.9%	92.1%	98.9%	70.9%	
2.5	2.0	5.9	6.9			83.9%	89.3%	45.8%	47.9%	90.0%	95.6%	36.2%	46.0%	92.1%	98.8%	72.5%	
2.5	2.5	5.9	6.0			84.9%	89.4%	42.9%	45.6%	91.6%	96.2%	37.2%	48.8%	91.8%	98.8%	72.7%	
2.5	3.0	5.9	4.8			86.6%	90.0%	40.9%	44.0%	93.8%	97.1%	43.9%	56.7%	91.6%	98.9%	74.3%	
2.5	3.5	5.9	4.0			87.1%	90.0%	38.1%	40.5%	94.9%	97.9%	43.9%	56.9%	91.3%	98.9%	73.9%	
2.5	4.0	5.9	3.5			87.8%	90.0%	36.7%	38.5%	95.8%	98.2%	47.3%	59.4%	91.2%	98.8%	74.4%	
3.0	0.5	4.6	16.1			68.8%	83.3%	61.3%	64.3%	71.4%	86.2%	16.2%	23.8%	94.3%	99.1%	66.9%	
3.0	1.0	4.6	11.6			77.3%	87.8%	58.8%	61.0%	80.6%	91.4%	21.4%	30.2%	94.5%	99.2%	70.2%	
3.0	1.5	4.6	8.5			82.7%	89.8%	55.2%	56.6%	86.7%	94.0%	27.2%	37.5%	94.4%	99.2%	72.3%	
3.0	2.0	4.6	6.9			85.3%	90.8%	53.5%	55.3%	89.7%	95.4%	31.4%	40.9%	94.3%	99.2%	73.6%	
3.0	2.5	4.6	6.0			86.5%	91.0%	51.2%	53.6%	91.3%	95.9%	32.9%	44.2%	94.1%	99.2%	74.0%	
3.0	3.0	4.6	4.8			88.4%	91.7%	48.9%	51.5%	93.6%	96.9%	39.0%	51.5%	94.0%	99.2%	75.5%	
3.0	3.5	4.6	4.0			88.9%	91.8%	45.6%	47.9%	94.7%	97.6%	38.9%	51.9%	93.6%	99.2%	75.0%	
3.0	4.0	4.6	3.5			89.5%	91.8%	43.6%	45.3%	95.6%	97.9%	41.8%	53.9%	93.5%	99.1%	75.2%	
3.5*	0.5	3.9	16.1			69.0%	83.8%	64.5%	66.3%	71.3%	86.3%	14.4%	21.4%	95.5%	99.5%	67.2%	
3.5*	1.0	3.9	11.6			77.7%	88.4%	62.4%	63.3%	80.4%	91.4%	19.3%	27.2%	95.7%	99.5%	70.6%	
3.5*	1.5	3.9	8.5			83.0%	90.5%	58.1%	59.0%	86.3%	94.0%	23.9%	33.3%	95.5%	99.5%	72.4%	
3.5*	2.0	3.9	6.9			85.9%	91.5%	56.7%	57.9%	89.5%	95.3%	28.3%	36.9%	95.5%	99.5%	73.8%	
3.5*	2.5	3.9	6.0			87.1%	91.8%	53.7%	55.8%	91.1%	95.8%	29.4%	39.4%	95.3%	99.5%	74.0%	
3.5*	3.0	3.9	4.8			89.0%	92.4%	51.0%	53.4%	93.3%	96.8%	34.9%	46.0%	95.1%	99.5%	75.3%	
3.5*	3.5	3.9	4.0			89.6%	92.6%	47.2%	49.2%	94.4%	97.5%	33.7%	44.3%	94.7%	99.5%	74.5%	
3.5*	4.0	3.9	3.5			90.3%	92.7%	44.9%	46.8%	95.3%	97.8%	36.6%	46.4%	94.6%	99.5%	74.7%	
4.0*	0.5	3.5	16.1			69.0%	84.2%	65.6%	68.0%	71.0%	86.3%	12.7%	19.5%	96.1%	99.5%	67.2%	
4.0*	1.0	3.5	11.6			78.0%	89.0%	65.1%	66.1%	80.2%	91.5%	17.8%	25.6%	96.4%	99.5%	70.9%	
4.0*	1.5	3.5	8.5			83.6%	91.2%	62.5%	63.6%	86.3%	94.0%	22.7%	32.1%	96.3%	99.5%	73.3%	
4.0*	2.0	3.5	6.9			86.5%	92.3%	60.9%	62.5%	89.4%	95.3%	27.0%	35.6%	96.3%	99.5%	74.6%	
4.0*	2.5	3.5	6.0			87.8%	92.6%	58.3%	60.9%	91.0%	95.9%	28.9%	38.9%	96.1%	99.5%	75.1%	
4.0*	3.0	3.5	4.8			89.7%	93.3%	54.7%	57.6%	93.3%	96.8%	34.0%	45.2%	96.0%	99.5%	76.1%	
4.0*	3.5	3.5	4.0			90.3%	93.4%	50.6%	52.8%	94.4%	97.5%	32.9%	43.5%	95.6%	99.5%	75.2%	
4.0*	4.0	3.5	3.5			91.2%	93.6%	49.7%	51.9%	95.3%	97.8%	36.6%	46.4%	95.5%	99.5%	75.9%	

*Viral Load or Viral Load at LOD (if concentration is below LOD), expressed as moving range (i.e., change of viral load or cases)

**VL XSDs which had to drop the Site 16 Positive VL Alert metric from analysis, due to 0 division error

Color gradients indicate higher values (green) and lower values (red) in each column.

The above table demonstrates that the original selection of three for both VL and case XSD was an excellent choice, as it provided well-rounded performance values across the board. The worst-performing metric was sensitivity, which was in the middle. If I wanted to increase sensitivity in the future, I could up the case XSD to 4 and drop the VL XSD to 2.5. This new setting would keep the general average metric (74.4% versus 75.5%) but increases the sensitivity from 51.5% to 59.4%.

Chapter 13: Omicron Retrospective Test

Chapter 12 conducted performance evaluation on the alert system, training it with the SMP Wks5-11 baselines and testing against the UCL XSDs of three (see Ch.12). However, the whole alert system itself still feels very abstract, and as such, this penultimate chapter is indulging a more realistic application.

Here we will conduct a retrospective test of the alert system, pretending that we were back in December of 2021. The VDH SMP just started collecting data twelve weeks prior, and everyone was blithely unaware of the impending Omicron arrival. Let's also imagine that this SMP team has an incredible Data Manager, who set up this exact alert system.

We're going to walk through each week of the Omicron spike, seeing what the alert system would have shown, versus what came.

As a reminder, this test occurs at the following point in time (see Figure 13.1 below):

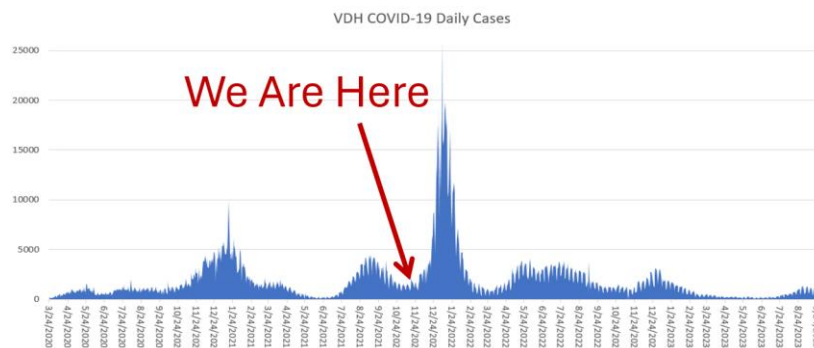


Figure 13.1 Daily VDH public COVID-19 case counts over time, stacked area graph, with start marker of mock scenario

Setting the Stage

The date is November 22nd, 2021. The SMP has been collecting data from 24 sites for 11 weeks. A proactive Data Manager reads this dissertation from the ether and uses it to organize her data, establish baselines, and apply the VL* alert system (see Ch.11) to run every time she receives and cleans new data. While she is not compelled to draw UCL for case counts presently, she does so because she knows she'll want to evaluate the system later. She is ready to alert any health districts as indicated.

Of note, the case alerts are truly retrospective. The lag time with case reporting is substantial, with up to a month of smudging applied to the VDH data dashboards. For this reason, consider case alerts to be more in line with the physical reality of locally infected people getting tested in their hospitals/other health care centers, while the district and state-level epidemiologists take longer to catch up to what's happening on the ground. Even though it takes a few days for WWS data to come in, it is most likely faster than case reporting.

Baselines

The baselines used for this scenario are also ones used for the creation and evaluation of this alert system generally (see Chapters 11-12). The baselines were trained on SMP Wks5-11, with XSDs of three for VL* and case moving averages to establish the Upper Control Limits (UCL). See Table 13.1 below for applied by-site baselines of VL* and case changes. Minor color coding is applied to emphasize the UCL for VL* (blue) and case changes (pink).

Table 13.1 Delta VL* and case moving average baselines of SMP mock scenario

VDH SMP Delta Baseline ΔVL* Alert System					
#	SITE	VL* SD	VL UCL	Case SD	Case UCL
1	AlexRenew	1.73E+12	5.18E+12	45	136
2	Aquia	9.38E+11	2.81E+12	32	97
3	Basham	7.17E+10	2.15E+11	5	15
4	Blacksburg	2.70E+11	8.09E+11	5	15
5	Danville	1.92E+11	5.75E+11	24	72
6	Coeburn	1.72E+11	5.17E+11	14	41
7	HLM	1.49E+12	4.46E+12	34	101
8	NA	2.83E+11	8.50E+11	84	251
9	ON	4.30E+11	1.29E+12	2	6
10	VIP	9.27E+11	2.78E+12	84	252
11	WB	6.32E+11	1.90E+12	32	95
12	LFR	5.17E+11	1.55E+12	13	39
13	LB	1.74E+12	5.22E+12	36	107
14	Moores	6.00E+11	1.80E+12	16	49
15	NorthRiver	1.55E+11	4.65E+11	27	82
16	Parkins	5.10E+11	1.53E+12	10	29
17	Peppers	2.38E+11	7.13E+11	13	38
18	Pound	1.29E+11	3.87E+11	3	10
19	Richmond	1.72E+12	5.16E+12	40	120
20	Roanoke	2.77E+12	8.31E+12	78	234
21	Rutledge	2.17E+09	6.51E+09	2	6
22	CB	1.84E+11	5.52E+11	11	33
23	UOSA	1.02E+12	3.06E+12	59	177
24	Wolf	2.09E+11	6.27E+11	13	39
Site Average		7.05E+11	2.12E+12	28	85

*Baselines: SMP Wks 7-11

**UCLs are generated from SD variance times a threshold multiplier (3)

Applying the VL* Alerts

A simple flag system is in place, so that anytime a site's increase in VL* is higher than the UCL allowance, an alert is generated for that site. For the 11 weeks prior, none of the sites experienced a VL* alert. However, most of those data points included baseline testing, so that is to be expected.

The Next Nine Weeks...

Table 13.2 below shows the issued by-site VL* flags for the following nine weeks and total average flag percentage. Minor color coding on the binary flags indicate a positive (1, pink),

negative (0, white), and no VL* data (grey). The average columns have some color gradients to indicate high (red) to low (green) percentage of flags by site and week.

Table 13.2 Weekly VL alert flags for all SMP sites for mock scenario, Binary flags (1=positive alert)*

Weekly Δ VL* Alert Flags										
	Wk 12	Wk 13	Wk 14	Wk 15	Wk 16	Wk 17	Wk 18	Wk 19	Wk 20	AVG
# SITE	11/29/2021	12/6/2021	12/13/2021	12/20/2021	12/27/2021	1/3/2022	1/10/2022	1/17/2022	1/24/2022	
1 AlexRenew	1	0	1	1	0	0	1	0	0	44%
2 Aquia	0	0	0	1	1		1	0	0	38%
3 Basham	0	0	0	1	0	1	0	0	0	22%
4 Blacksburg	0	0	0	0	0	1	0	1	0	22%
5 Danville	0	0	0	1	0	1	1	0	0	33%
6 Coeburn	0	0	0	0	0	0	0	0	0	0%
7 HLM	0	0	0	1	1	1	0	0	0	33%
8 NA	1	1	0	0	1	0	1	0	0	44%
9 ON		0	0	0	1	0	0	0	0	13%
10 VIP	0	0	0	0	1	0	0		0	13%
11 WB	0	0	0	1	1	0	0	0	0	22%
12 LFR	1	0	0	0	1		0	0	0	25%
13 LB	0	0	0	0	0	1	0	0	0	11%
14 Moores	0	0	0	0	1	1	0	0	0	22%
15 NorthRiver	1	0	1	0	1	0	1	0	0	44%
16 Parkins	0	0	0	0	0	1	0	0	0	11%
17 Peppers	0	0	0	1	0	1	1	0	0	33%
18 Pound		0	0	0	0	0	0	0	0	0%
19 Richmond	0	1	0	1	1	0	1	0	0	44%
20 Roanoke	0	1	0	0	1	1	0	0	0	33%
21 Rutledge	0	1	0	0	1	0	0	0	1	33%
22 CB	0	0	0	1	0	1	1		0	38%
23 UOSA	1	0	0			1	0	0	0	29%
24 Wolf	0	1	0	0	0	0	1	0	0	22%
Sites Average	23%	21%	8%	39%	52%	50%	38%	5%	4%	26%

*Baselines: SMP Wks 7-11

**UCLs are generated from SD variance times a threshold multiplier (3)

Color coding and binary flags explained in text.

Weeks 12-13: 5 sites flagged positive on Wk12 (23%) and Wk13 (21%). This seemed like a lot, considering the previous eleven weeks had none. The Data Manager sends alerts to the five site health districts each week but is worried that the alert system is too sensitive. What she

doesn't know (but will in 54 weeks when she runs performance evaluation metrics), is that the average flag % across the entire timeline is 9.2%. This alert positivity is actually *much too high*.

Case alerts are minimal (14% and 8%) -- see Table 13.3 below for case alert comparison.

Week 14: only sites 2 flagged positive (8%). The Data Manager sends alerts to the 2 site health districts. Case alerts pick up a touch (21%).

Weeks 15-17: VL* alerts explode, with 9 (39%) on Wk15, 12 (52%) on Wk16, and 11 (50%) on Wk17. Case alerts explode at 61%, 74%, and 64% respectively.

Weeks 18-20: VL* alerts drop back down, with 9 (38%) on Wk18, 1 (5%) on Wk17, and 1 (4%) on Wk18. Case alerts also come back down at 21%, 14%, and 0% respectively.

Table 13.3 below shows the issued by-site case change flags for the nine weeks and total average flag percentage. Minor color coding on the binary flags indicate a positive, (1, pink) negative (0, white), and no VL* data (grey). The average columns have some color gradients to indicate high (red) to low (green) percentage of flags by site and week.

Table 13.3 Weekly case alert flags for all SMP sites for mock scenario, Binary flags (1=positive alert)

Weekly δ Case* Alert Flags											
		Wk 12	Wk 13	Wk 14	Wk 15	Wk 16	Wk 17	Wk 18	Wk 19	Wk 20	AVG
#	SITE	11/29/2021	12/6/2021	12/13/2021	12/20/2021	12/27/2021	1/3/2022	1/10/2022	1/17/2022	1/24/2022	
1	AlexRenew	0	0	0	0	0	0	0	0	0	0%
2	Aquia	0	0	0	0	0		0	0	0	0%
3	Basham	0	0	0	0	0	0	0	0	0	0%
4	Blacksburg	0	0	0	0	0	1	1	0	0	22%
5	Danville	0	0	0	1	1	1	0	0	0	33%
6	Coeburn	0	0	0	0	0	1	1	0	0	22%
7	HLM	0	0	1	1	1	0	0	0	0	33%
8	NA	0	0	0	1	1	1	0	0	0	33%
9	ON		0	0	0	1	0	1	0	0	25%
10	VIP	0	0	0	1	1	0	0		0	25%
11	WB	0	0	1	1	1	1	0	0	0	44%
12	LFR	1	0	1	1	1		0	0	0	50%
13	LB	0	0	0	1	1	1	0	0	0	33%
14	Moores	0	1	0	1	1	1	0	1	0	56%
15	NorthRiver	0	0	0	1	1	1	0	0	0	33%
16	Parkins	1	0	0	1	1	1	0	0	0	44%
17	Peppers	0	0	0	1	1	1	0	1	0	44%
18	Pound		0	0	0	1	0	1	1	0	38%
19	Richmond	0	1	1	1	1	1	0	0	0	56%
20	Roanoke	0	0	0	1	1	1	0	0	0	33%
21	Rutledge	0	0	0	0	0	0	0	0	0	0%
22	CB	0	0	0	1	1	1	0		0	38%
23	UOSA	1	0	1			1	0	0	0	43%
24	Wolf	0	0	0	0	1	0	1	0	0	22%
Sites Average		14%	8%	21%	61%	74%	64%	21%	14%	0%	30%

*Baselines: SMP Wks 7-11

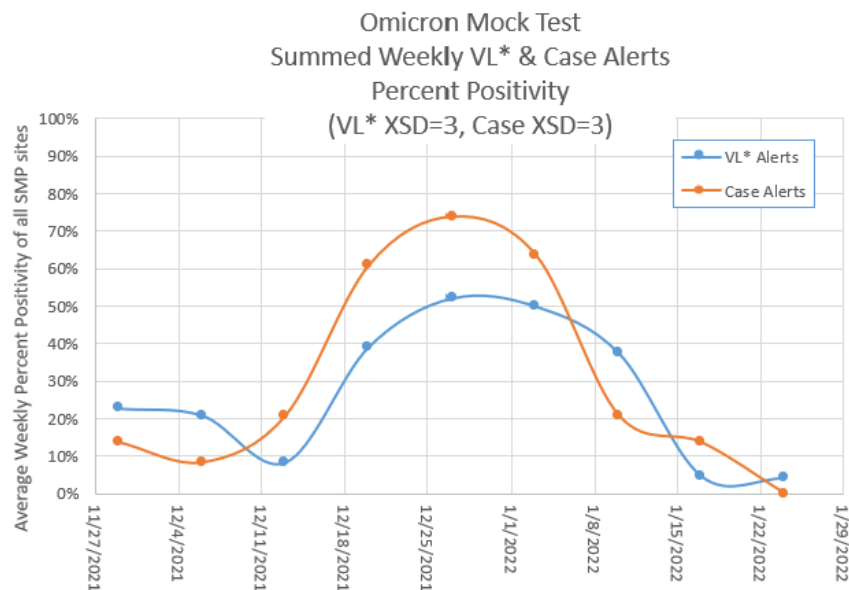
**UCLs are generated from SD variance times a threshold multiplier (3)

Color coding and binary flags explained in text.

Overlaying the Big Picture

In general, the VL* alerts warned of imminent case increase within the first two weeks of data collection and proceeded to march along with case alerts until the end. The Data Manager may not have fully appreciated this warning, as she was unlucky to have such an incident occur immediately after baseline testing. But, by having her send out so many alerts, she may have helped public health practitioners interpret what they were hearing from ground level.

See Figure 13.2 below for smoothed line graphs of the two-alert percent positivity in time.



*Viral Load or Viral Load at LOD (if concentration is below LOD), expressed as moving range (i.e., change of viral load or cases)

Figure 13.2: Weekly average VL* and case alert percent positivity smoothed lines in time, mock scenario

The agreement between the VL* alert and (retrospective) case alert flags during the same week was approximately 72% overall. By-site agreement was highly variable, ranging from 44% to 89%, as seen below in Table 13.4. Minor color coding on the binary flags indicate positive (1, green) and negative (0, yellow) flags. The average columns have some color gradients to indicate high (green) to low (yellow) percentage of flags by site and week.

Table 13.4 Binary agreement between VL* and case alerts for all sites during mock scenario

Weekly VL & Case Agreement (same week)											
		Wk 12	Wk 13	Wk 14	Wk 15	Wk 16	Wk 17	Wk 18	Wk 19	Wk 20	AVG
#	SITE	11/29/2021	12/6/2021	12/13/2021	12/20/2021	12/27/2021	1/3/2022	1/10/2022	1/17/2022	1/24/2022	
1	AlexRenew	0	1	0	0	1	1	0	1	1	56%
2	Aquia	1	1	1	0	0	1	0	1	1	67%
3	Basham	1	1	1	0	1	0	1	1	1	78%
4	Blacksburg	1	1	1	1	1	1	0	0	1	78%
5	Danville	1	1	1	1	0	1	0	1	1	78%
6	Coeburn	1	1	1	1	1	0	0	1	1	78%
7	HLM	1	1	0	1	1	0	1	1	1	78%
8	NA	0	0	1	0	1	0	0	1	1	44%
9	ON	1	1	1	1	1	1	0	1	1	89%
10	VIP	1	1	1	0	1	1	1	1	1	89%
11	WB	1	1	0	1	1	0	1	1	1	78%
12	LFR	1	1	0	0	1	1	1	1	1	78%
13	LB	1	1	1	0	0	1	1	1	1	78%
14	Moore's	1	0	1	0	1	1	1	0	1	67%
15	NorthRiver	0	1	0	0	1	0	0	1	1	44%
16	Parkins	0	1	1	0	0	1	1	1	1	67%
17	Peppers	1	1	1	1	0	1	0	0	1	67%
18	Pound	1	1	1	1	0	1	0	0	1	67%
19	Richmond	1	1	0	1	1	0	0	1	1	67%
20	Roanoke	1	0	1	0	1	1	1	1	1	78%
21	Rutledge	1	0	1	1	0	1	1	1	0	67%
22	CB	1	1	1	1	0	1	0	1	1	78%
23	UOSA	1	1	0	1	1	1	1	1	1	89%
24	Wolf	1	0	1	1	0	1	1	1	1	78%
Sites Average		83%	79%	71%	54%	63%	71%	50%	83%	96%	72%

*Baselines: SMP Wks 7-11

**UCLs are generated from SD variance times a threshold multiplier (3)

***1 is agreement between VL* & case alerts for same week, 0 is disagreement
Color coding and binary flags explained in text.

The 44% to 89% range demonstrates the by-site variability in case-VL* alignment, whether that's due to VL* sewage factors, population case reporting, or other factors that might delay the same-week relationship. The alert system may therefore be better suited for some sites over others.

Chapter 14: Discussion & Conclusions

Dissertation Overview

This dissertation asked the question of whether WWS could predict imminent rises in community COVID-19 disease burden (Ch.1).

A review of literature (Ch.2) and WWS variability exploration (Ch.3) gave context to the nature of the WWS data, guiding the kinds of data analysis and early warning frameworks/systems which could be entertained. Chapter 4 explored the fundamentals of alert systems and selected Yang's (2017) framework model for building an early warning system. This includes focus on the data analysis phase and using the following steps to build an alert system: 1) early warning model building, 2) threshold setting, and 3) signal alert generation. As well as performance evaluation.

Chapters 6 and 7 laid the groundwork for analysis methods and identified controllable aspects like variant dominance and case count changes. It also revealed substantial variabilities between the sites themselves, setting the guidepost of only using metrics that were *by site*.

Chapters 8 and 9 explored viral load (VL) and case and hospital correlation, revealing a stronger relationship between VL and case counts. The correlation coefficient of VL and the next 7 days of cases was 0.875, with an increase to 0.900 if there was a 3-day shift backward (i.e., the previous 3 days and the next 4 days). This high correlation created confidence in setting up a system which uses VL to predict case counts.

Chapter 10 used the Yang model, correlation testing, and general analysis findings to select the independent and dependent variables for an alert system – the change of viral load (or

viral load of LOD if below LOD) to the change of weekly cases by site. Base models were explored, ultimately landing on the by-site I-MR Control Chart as the alert system base. This provided the simplicity and directness, which were key element goals set from Ch. 1. To see if the data normality was agreeable to such a chart, normality tests for baselines and total datasets were explored and ultimately approved. The proposed model included the I-MR Control Chart for VL and testing against a similar control chart for case increases. This completed Yang's first step: 1) early warning model building.

To complete the next of Yang's steps, 2) threshold setting, the baseline for system training was selected in Ch.11. The controllable aspects of variant dominance and case change (Ch. 6) to isolate the period of SMP Wks5-11 as acceptable baselines for establishing the UCLs for both VL and case changes. The standard deviation multiplier threshold of 3 for both VL and case change was used as default, though Ch. 12 evaluated the performance effects of 0.5 increments from 0.5 to 4.0 for both VL and case change thresholds. This completed Yang's second step.

And finally, Yang's final step, 3) signal alert generation. Ch.11 used a simple Excel flag system to assign VL and case alerts to the SMP data, based on whether the VL and case changes rose above their UCL. However, performance evaluation was still required, so summary statistics were generated by site and total average. When using the standard deviation multiplier threshold of 3 for both changing VL and cases (daily cases for next 7 days), the alert flags matched each other 88% of the time. If calculated for case flags of that week or following week (accounting for possible VL leading indicator effects), matching increased to 92%. For every positive VL alert, 49% had an associated positive case alert that week; 52% for next 1-2 weeks. For every negative VL alert, 94% had an associated negative case alert that week; 97% for next 1-2 weeks.

Sensitivity was calculated at 39% for same week, or 52% for next 1-2 weeks. Specificity was calculated at 94% for same week, or 99% for next 1-2 weeks. These metrics were explored using other thresholds, which did improve certain performance indicators (e.g., sensitivity), but usually at the cost of other indicators.

To make this model more accessible to public health practitioners who would like to use it for their own WWS, Chapters 11-13 were written in a more targeted voice, complete with tips to the readers. Chapter 13 even provided a mock scenario, effectively a retrospective analysis of the system as it pertained to the Omicron spike of December 2021. This scenario allowed readers to visualize the system in application, being able to place themselves in an impending case spike scenario.

Discussion

Overall, I am pleased with the building of the alert system and performance metrics. The matching of the VL and case alerts (92%) indicate a strong relationship between an acceptable threshold VL change and an imminent threshold-derived rise of cases. This answers the question of whether WWS could predict imminent rises in community COVID-19 disease: yes, as a VL change alert to case change rise.

The alert sensitivity of 52%, meaning that only about half the imminent rises will be detected at that threshold, could be improved. Specificity is, however, excellent at 99%, meaning that the system is very good at knowing when a case increase will *not* happen. And while changing VL and case thresholds can adjust the performance evaluation metrics, they usually come at the cost of other metrics. To decide on the value of each would require a conversation with the users of the system. For my needs, I would rather have a system that is more discerning

and underreport, than be overly sensitive and overreport. However, this does answer the follow-up question in Ch.1 about whether a framework can be reasonably applied to WWS data: yes, with a system as simple as an Excel workbook, in fact.

The final follow-up Ch.1 question, referring to application to other non-VDH WWS data, was outside the scope of this dissertation and unaddressed. However, the data used here relied mostly on case data and viral loads. Provided that those pieces of data are available – and it is standard practice in NWSS to collect the pieces that make up viral load (viral concentration and flow rate) – I don't see why any other jurisdiction could not replicate this alert system for SARS-CoV-2 or other WWS targets. The biggest limitation in my mind would be the lack of good, consistent WWS collection and analysis. I am lucky to be working with data that maintained collection and analysis methods throughout, performed by a single state lab. Perhaps other jurisdictions may have even more inter-site and inter-lab variability to contend with.

Which takes me to my final take-home theme that seemed to pop up in every chapter – the by site variability. It was beyond the scope of my dissertation to identify which sites were better suited to my alert system, but it isn't hard to see that not all sites had the same consistency in data reporting or viral load variability. A follow-up to this dissertation could include a deeper dive into discerning how and why some sites have better evaluation outcomes than others. Perhaps there is something the SMP can do to facilitate better data collection, whether that means working with volunteering utility partners or lab staff to ensure the highest quality data, or simply not including sites with certain metadata qualities that have higher likelihood of more variable data.

Limitations

There were several major limitations to this dissertation, all centered on the data itself. The first of which was the case data, which used VDH public download and VDH WWS spatially linked case counts. The source of both those sets is ultimately VDH case reporting. It's no secret that there are major limitations around this provisional case reporting data, including reporting lag and other issues. The reporting lag was controlled for in this dissertation with the cut-off of analyzed datasets set several weeks before the most recent data, where the likelihood of most reporting being in is higher. However, at-home COVID-19 tests became readily available to Virginians starting in the Spring of 2022. Due to this, the number of clinical (reported) COVID-19 infections probably started to become artificially suppressed, as Virginians no longer needed to visit a clinical (reporting) facility to test themselves for COVID-19. In addition to this, case definitions at VDH changed on January 1, 2023 to no longer include positive at-home antigen tests as confirmatory. The result is that VDH case data is less reliable as a community indicator after Spring 2022. With the start of 2025, COVID-19 case reporting will no longer happen within VDH.

Additionally, site metadata like sewershed boundaries are not created equally. I know, because I made many of them. Some sites have professional quality, detailed mapping of their service regions, while others provided a generalized 'gist' of service area. This means that spatially mapped case counts may not be an accurate representation of that region.

Lastly, the variant reporting is very limited. Here I used federal and state data points to narrow in on the best controlled time for variant dominance, but it was surely lacking in granularity. Hence why the same baseline was used for all sites, and not by-site. This is

something which may be improved upon as VDH WWS continues doing their own sequencing of their sites.

Importance & Practical Application

As summarized in Chapter 4.5, the current landscape of mainstream WWS alert systems is painfully inadequate. This alert system model provides a framework for creating not only actionable flags – which assists public and public health interpretation of WWS data – but also ties in contextual relationships of real-world disease burden (i.e., cases). This answers the question of whether a sample’s viral load is “high enough” to be worried about imminent case rise.

This proposed alert system offers substantial differences and value when compared to CDC NWSS’s most closely alert-related metric, the Wastewater Viral Activity Level (WVAL)(CDC, 2025a). The two systems are contrasted below in Table 14.1.

Table 14.1 *Summary Differences between CDC WVAL and Yancey VL Alert*

Summary: CDC WVAL vs Yancey VL Alert		
	WVAL**	Yancey VL Alert
Provides “Alerts”?	No, but provides current Z-score/category [“What is the current level of viral activity?”]	Yes (if input > XSD Upper Control Line) [“Are cases going to rise imminently?”]
Provides contextual relationship to cases or other health outcomes?	No	Yes (built for case change, but also considered hospitalization correlates)
Relationship to imminent case rise?	No	Yes
Provides future modeling?	No	No
Provides validation metrics?	Not stated on dashboard	Yes (matching, Se-Sp)
Based on academic work?	Not stated on dashboard	This dissertation
Assesses sites individually?	Yes	Yes
Input metric	Weekly averaged by-site log FPNC (flow/population normalized concentration) [FPNC calculates the same as VL*/population]	Each by-site VL* change
Uses historical baseline data?	Yes, last 12 months (recalculated 2x/yearly) at 10 th percentile	Yes, VL* change measured during most recent state-level period of low cases & variant stability
Output metric	E^X for $(\ln(\text{FPNC weekly}) - \ln(\text{FPNC baseline})) / (\text{FPNC baseline SD})$ [logged and unlogged FPNC z-score]	Positive or negative alert flag per site; positive when VL* change is greater than XSD threshold
User customizable outputs?	No	Yes, case & VL* thresholds can be adjusted
Ability to aggregate regionally?	Yes, state/regional/national categories on dashboard	Currently by-site alerts only
Can be used for other PCR targets?	Yes, currently used for SARS-CoV-2, flu A, RSV	Theoretically, but not validated
Model currently used by:	CDC NWSS	VDH WWS

*Viral Load or Viral Load at LOD (if concentration is below LOD), expressed as moving range (i.e., change of viral load or cases)

** CDC NWSS Viral Activity Level (CDC, 2025a)

The most distinguishing difference between the two is that the primary purpose of the WVAL is not to give assign alerts or tie in cases, though one would likely be alarmed if they saw their geographic region sporting a high category. The metric’s primary purpose, as distinguished from percentiles, is to be able to aggregate different sites and regions together under a common metric. It is not designed to share whether the recent site value is higher than the last, or whether that portends real disease burden changes.

The purpose of this proposed alert system is to be able to ask whether an individual site’s value is high enough to elicit concern with a simple flag. It is meant to inform the public/public health practitioners that there has been a statistically notable increase in community viral shedding, which could warn of increased cases or hospitalizations in the coming days. Hospitals and case workers are most likely already aware of their current patient burden, but having insights into potential surges over the next few days can give them precious foresight for

preparation. Similarly, interested members of the public may appreciate some heads-up regarding imminent increases in community illness. If everyone is going to get sick over the next few days, perhaps social distancing planning may be appropriate.

While the above practical applications for the public and public health practitioners may be useful, the real value of this dissertation is the groundwork it lays for future WWS. With the extensive monetary investment in NWSS infrastructure across the country over the last several years, and other emerging infectious diseases looming upon the horizon (e.g., H5N1, MPOX), we could again begin to worry about overwhelming our healthcare system and look to WWS for surveillance support. This will be especially true if case data is limited or unreliable. And if that were to happen, doing dissertation-level deep dives on variability analysis, early warning systems, correlational analyses, and developing/evaluating alert systems will be inconvenient at best.

The answer to the primary research question of whether once-weekly WWS sampling can predict imminent rises in COVID-19 community disease burden is a resounding ‘yes’. With a tool as simple as Excel. Yet, with the novelty and inherent variability associated with WWS, the proposed framework required a thorough evaluation of data variability, literature review, data analysis, alert system building, and evaluation. It required interdisciplinary expertise (e.g., environmental health, public health, data sciences, and disaster management), professional-level insights into the building and management of this burgeoning field, as well as the time, creativity, and interest of a PhD student to sustain a multi-year research project. Being in this place, at this time, has allowed me to be uniquely of service during a public health crisis. It is my hope that the research presented here helps shape the way we use and share WWS in this state and beyond.

References

- Ahmed, W., Angel, N., Edson, J., Bibby, K., Bivins, A., O'Brien, J. W., Choi, P. M., Kitajima, M., Simpson, S. L., Li, J., Tschärke, B., Verhagen, R., Smith, W. J. M., Zaugg, J., Dierens, L., Hugenholtz, P., Thomas, K. V., & Mueller, J. F. (2020). First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: A proof of concept for the wastewater surveillance of COVID-19 in the community. *Science of The Total Environment*, 728, 138764. <https://doi.org/10.1016/j.scitotenv.2020.138764>
- Ahmed, W., Tschärke, B., Bertsch, P.M., Bibby, K., Bivins, A., Choi, P., Clarke, L., Dwyer, J., Edson, J., Nguyen, T.M.H., O'Brien, J.W., Simpson, S.L., Sherman, P., Thomas, K.V., Verhagen, R., Zaugg, J., Mueller, J.F. (2021). SARS-CoV-2 RNA monitoring in wastewater as a potential early warning system for COVID-19 transmission in the community: A temporal case study. *Science of The Total Environment*, 761. <https://www.sciencedirect.com/science/article/pii/S0048969720377470>
- APHL. (2022). *SARS-CoV-2 Wastewater Surveillance Testing Guide for Public Health Laboratories*. APHL.org: Association of Public Health Laboratories. Retrieved from <https://www.aphl.org/aboutAPHL/publications/Documents/EH-2022-SARSCoV2-Wastewater-Surveillance-Testing-Guide.pdf>
- APHL. (2023). *Association of Public Health Laboratories*. Retrieved July 19, 2023 from <https://www.aphl.org/>
- Arimi, S. M., Fricker, C. R., & Park, R. W. A. (1988). Occurrence of "thermophilic" campylobacters in sewage and their removal by treatment processes. *Epidemiology and infection*. London, New York NY, 101(2), 279-286. <https://www.proquest.com/scholarly-journals/occurrence-thermophilic-campylobacters-sewage/docview/15162834/se-2?accountid=14826>
- Asghar, H., Diop, O. M., Weldegebriel, G., Malik, F., Shetty, S., El Bassioni, L., Akande, A. O., Al Maamoun, E., Zaidi, S., Adeniji, A. J., Burns, C. C., Deshpande, J., Oberste, M. S., & Lowther, S. A. (2014). Environmental Surveillance for Polioviruses in the Global Polio Eradication Initiative. *The Journal of Infectious Diseases*, 210(suppl_1), S294-S303. <https://doi.org/10.1093/infdis/jiu384>
- Aulicino, F. A., Colombi, A., Calcaterra, E., Carere, M., Mastrantonio, A., & Orsini, P. (1998). Microbiological and chemical quality of sludges from domestic wastewater plants. *International Journal of Environmental Health Research*, 8(2), 137-144. <https://doi.org/10.1080/09603129873561>
- Bernal, J. D., & Fankuchen, I. (1941). X-RAY AND CRYSTALLOGRAPHIC STUDIES OF PLANT VIRUS PREPARATIONS : I. INTRODUCTION AND PREPARATION OF SPECIMENS II. MODES OF AGGREGATION OF THE VIRUS PARTICLES. *J Gen Physiol*, 25(1), 111-146. <https://doi.org/10.1085/jgp.25.1.111>
- Bivins, A., North, D., Ahmad, A., Ahmed, W., Alm, E., Been, F., Bhattacharya, P., Bijlsma, L., Boehm, A. B., Brown, J., Buttiglieri, G., Calabro, V., Carducci, A., Castiglioni, S., Cetecioglu Gurol, Z., Chakraborty, S., Costa, F., Curcio, S., de los Reyes, F. L., . . . Bibby, K. (2020). Wastewater-Based Epidemiology: Global Collaborative to Maximize

- Contributions in the Fight Against COVID-19. *Environmental Science & Technology*, 54(13), 7754-7757. <https://doi.org/10.1021/acs.est.0c02388>
- Blevins, S. M., & Bronze, M. S. (2010). Robert Koch and the 'golden age' of bacteriology. *Int J Infect Dis*, 14(9), e744-751. <https://doi.org/10.1016/j.ijid.2009.12.003>
- Bloom, H. H., MacK, W. N., Krueger, B. J., & Mallmann, W. L. (1959). Identification of Enteroviruses in Sewage. *The Journal of Infectious Diseases*, 105(1), 61-68. <http://www.jstor.org.ezproxy.lib.vt.edu/stable/30102963>
- Brillman, J.C., Burr, T., Forslund, D., Joyce, E., Picard, R., & Umland, E. (2005). Modeling emergency department visit patterns for infectious disease complaints: results and application to disease surveillance. *BMC Medical Informatics and Decision Making*, 2005, 5(4). Retrieved from <https://bmcmmedinformdecismak.biomedcentral.com/counter/pdf/10.1186/1472-6947-5-4.pdf>
- Bureau, U. S. C. *QuickFacts Virginia*. Retrieved October 28, 2023 from <https://www.census.gov/quickfacts/fact/table/VA/PST045222>
- Burkom, H.S. (2005). Development, Adaptation, and Assessment of Alerting Algorithms for Biosurveillance. *Johns Hopkins APL Technical Digest*, 24(4), 335-342. Retrieved from <https://secwww.jhuapl.edu/techdigest/content/techdigest/pdf/v24-n04/24-04-burkom.pdf>
- Burkom, H.S., Elbert, Y., Magruder, S.F., Najmi, A.H., Peter, W., Thompson, M.W. (2008). Developments in the Roles, Features, and Evaluation of Alerting Algorithms for Disease Outbreak Monitoring. *Johns Hopkins APL Technical Digest*, 27(4). <https://secwww.jhuapl.edu/techdigest/Content/techdigest/pdf/V27-N04/27-04-BurkomDevelopments.pdf>
- CDC. (2017). *BioSense Platform Quick Start Guide to Using ESSENCE*. Retrieved Nov 17, 2024 from <https://www.cdc.gov/nssp/biosense/docs/BioSense-Platform-Quick-Start-Guide-for-ESSENCE.pdf>
- CDC. (2022, October 26, 2022). *Symptoms of COVID-19*. Retrieved July 16, 2023 from <https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html>
- CDC. (2023a). *CDC COVID Data Tracker: Variant Proportions*. Retrieved July 19, 2023 from <https://covid.cdc.gov/covid-data-tracker/#variant-proportions>
- CDC. (2023a). *COVID Data Tracker*. Centers for Disease Control and Prevention. Retrieved May 5, 2023 from <https://covid.cdc.gov/covid-data-tracker/#wastewater-surveillance>
- CDC. (2023b). *COVID Data Tracker*. Centers for Disease Control and Prevention. Retrieved May 5, 2023 from <https://covid.cdc.gov/covid-data-tracker/#wastewater-surveillance>
- CDC. (2023b). *National Wastewater Surveillance System (NWSS)*. Centers for Disease Control and Prevention. Retrieved March 19, 2023 from <https://www.cdc.gov/nwss/wastewater-surveillance/index.html>
- CDC. (2023c). *Data Dictionary (Excel File)*. Wastewater Surveillance Data Reporting and Analytics: CDC Retrieved from https://www.cdc.gov/nwss/files/NWSS-Data-Dictionary_v5.0.0_2023-07-10.xlsx

- CDC. (2023d). *National Wastewater Surveillance System (NWSS)*. Centers for Disease Control and Prevention. Retrieved March 19, 2023 from <https://www.cdc.gov/nwss/wastewater-surveillance/index.html>
- CDC. (2023e, September 1, 2023). *SARS-CoV-2 Variant Classifications and Definitions*. Retrieved September 23, 2023 from <https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html>
- CDC. (2023f). *Wastewater Surveillance Testing Methods*. Retrieved October 8, 2023 from <https://www.cdc.gov/nwss/testing.html>
- CDC. (2024a). *National Syndromic Surveillance Program (NSSP): BioSense Platform*. Centers for Disease Control and Prevention. Retrieved Nov 17, 2024 from <https://www.cdc.gov/nssp/php/onboarding-toolkits/essence.html>
- CDC. (2024b). *National Syndromic Surveillance Program*. Centers for Disease Control and Prevention. Retrieved Nov 17, 2024 from <https://www.cdc.gov/nssp/documents/NSSP-Infographic.pdf>
- CDC. (2025). *COVID Data Tracker*. Centers for Disease Control and Prevention. Retrieved January 28, 2025 from <https://covid.cdc.gov/covid-data-tracker/#wastewater-surveillance>
- CDC. (2025a). *COVID-19 Current Wastewater Viral Activity Levels Map*. Centers for Disease Control and Prevention. Retrieved February 19, 2025 from <https://www.cdc.gov/nwss/rv/COVID19-currentlevels.html>
- Choi, P. M., Tschärke, B. J., Donner, E., O'Brien, J. W., Grant, S. C., Kaserzon, S. L., Mackie, R., O'Malley, E., Crosbie, N. D., Thomas, K. V., & Mueller, J. F. (2018). Wastewater-based epidemiology biomarkers: Past, present and future. *TrAC Trends in Analytical Chemistry*, 105, 453-469. <https://doi.org/https://doi.org/10.1016/j.trac.2018.06.004>
- Ciesielski, M., Blackwood, D., Clerkin, T., Gonzalez, R., Thompson, H., Larson, A., & Noble, R. (2021). Assessing sensitivity and reproducibility of RT-ddPCR and RT-qPCR for the quantification of SARS-CoV-2 in wastewater. *Journal of Virological Methods*, 297, 114230. <https://doi.org/https://doi.org/10.1016/j.jviromet.2021.114230>
- cov-lineages.org. (2023). *Lineage List*. Retrieved September 24, 2023 from https://cov-lineages.org/lineage_list.html
- COVID19WBEC. (2023). *COVID-19 WBE Collaborative*. COVID19WBEC.ORG. Retrieved March 29, 2023 from <https://www.covid19wbec.org/>
- Curtis, K., Keeling, D., Yetka, K., Larson, A., & Gonzalez, R. (2021). Wastewater SARS-CoV-2 RNA Concentration and Loading Variability from Grab and 24-Hour Composite Samples. *medRxiv*, 2020.2007.2010.20150607. <https://doi.org/10.1101/2020.07.10.20150607>
- Divizia, M., Ruscio, V., Degener, A. M., & Panà, A. (1998). Hepatitis A virus detection in wastewater by PCR and hybridization. *The new microbiologica*, 21(2), 161-167. <http://europepmc.org/abstract/MED/9579341>
- Dousse, C., Schneider, E., & Schwartzbrod, J. (1993). Comparison of Methods for Isolating *Campylobacter* SPP from Sewage. *Water Science and Technology*, 27(3-4), 275-278.

- <https://www.proquest.com/scholarly-journals/comparison-methods-isolating-campylobacter-spp/docview/1943275873/se-2?accountid=14826>
- Du, P., Zhou, Z., Wang, Z., Xu, Z., Zheng, Q., Li, X., He, J., Li, X., Cheng, H., & Thai, P. K. (2021). Analysing wastewater to estimate fentanyl and tramadol use in major Chinese cities. *Science of The Total Environment*, 795, 148838. <https://doi.org/https://doi.org/10.1016/j.scitotenv.2021.148838>
- Dubois, E., Le Guyader, F., Haugarreau, L., Kopecka, H., Cormier, M., & Pommepuy, M. (1997). Molecular epidemiological survey of rotaviruses in sewage by reverse transcriptase seminested PCR and restriction fragment length polymorphism assay. *Applied and Environmental Microbiology*, 63(5), 1794-1800. <https://www.proquest.com/scholarly-journals/molecular-epidemiological-survey-rotaviruses/docview/16052223/se-2>
- Dulbecco, R., & Vogt, M. (1954). PLAQUE FORMATION AND ISOLATION OF PURE LINES WITH POLIOMYELITIS VIRUSES. *Journal of Experimental Medicine*, 99(2), 167-182. <https://doi.org/10.1084/jem.99.2.167>
- FDA. (2023). *FDA: Wastewater Surveillance for SARS-CoV-2 Variants*. Retrieved July 19, 2023 from <https://www.fda.gov/food/whole-genome-sequencing-wgs-program/wastewater-surveillance-sars-cov-2-variants>
- Gandhi, M., Yokoe, D. S., & Havlir, D. V. (2020). Asymptomatic Transmission, the Achilles' Heel of Current Strategies to Control Covid-19. *New England Journal of Medicine*, 382(22), 2158-2160. <https://doi.org/10.1056/NEJMe2009758>
- Gillon, R. (1994). Medical ethics: four principles plus attention to scope. *Bmj*, 309(6948), 184.
- Gonzalez, R., Curtis, K., Bivins, A., Bibby, K., Weir, M. H., Yetka, K., Thompson, H., Keeling, D., Mitchell, J., & Gonzalez, D. (2020). COVID-19 surveillance in Southeastern Virginia using wastewater-based epidemiology. *Water Research*, 186, 116296. <https://doi.org/https://doi.org/10.1016/j.watres.2020.116296>
- Grabow, W. O., Botma, K. L., de Villiers, J. C., Clay, C. G., & Erasmus, B. (1999). Assessment of cell culture and polymerase chain reaction procedures for the detection of polioviruses in wastewater. *Bull World Health Organ*, 77(12), 973-980.
- Guskey, L. E. (1983). Detection of particle-associated enteroviruses from raw and treated sewage and Lake Michigan Harbour water. *Journal of Great Lakes Research*, 9(4), 468-473. <https://www.proquest.com/scholarly-journals/detection-particle-associated-enteroviruses-raw/docview/13734537/se-2?accountid=14826>
- Guttman-Bass, N., Tchorsh, Y., & Marva, E. (1987). Comparison of methods for rotavirus detection in water and results of a survey of Jerusalem wastewater. *Applied and Environmental Microbiology*, 53(4), 761-767. <https://www.proquest.com/scholarly-journals/comparison-methods-rotavirus-detection-water/docview/14642725/se-2?accountid=14826>
- Hall, W., Prichard, J., Kirkbride, P., Bruno, R., Thai, P. K., Gartner, C., Lai, F. Y., Ort, C., & Mueller, J. F. (2012). An analysis of ethical issues in using wastewater analysis to

- monitor illicit drug use. *Addiction*, 107(10), 1767-1773. <https://doi.org/10.1111/j.1360-0443.2012.03887.x>
- HB 1682 (2023). Weapon of terrorism; definition, penalty. , Virginia House (2023). <https://lis.virginia.gov/cgi-bin/legp604.exe?231+sum+HB1682>
- Hendriksen, R. S., Munk, P., Njage, P., van Bunnik, B., McNally, L., Lukjancenko, O., Röder, T., Nieuwenhuijse, D., Pedersen, S. K., Kjeldgaard, J., Kaas, R. S., Clausen, P. T. L. C., Vogt, J. K., Leekitcharoenphon, P., van de Schans, M. G. M., Zuidema, T., de Roda Husman, A. M., Rasmussen, S., Petersen, B., . . . The Global Sewage Surveillance project, c. (2019). Global monitoring of antimicrobial resistance based on metagenomics analyses of urban sewage. *Nature Communications*, 10(1), 1124. <https://doi.org/10.1038/s41467-019-08853-3>
- Hering, C. L. (2009). Flushing the fourth amendment down the toilet: how community urinalysis threatens individuals privacy. *Ariz. L. Rev.*, 51, 741.
- Hessing, T., & Ramana, P.V. (2024). *I-MR Chart*. Six Sigma Study Guide. Retrieved November 22, 2024 from <https://sixsigmastudyguide.com/i-mr-chart/>
- Hill Jr, W. F., Akin, E. W., & Benton, W. H. (1971). Detection of viruses in water: a review of methods and application. *Water Research*, 5(11), 967-995.
- Hopkins, L., Persse, D., Caton, K., Ensor, K., Schneider, R., McCall, C., & Stadler, L. B. (2023). Citywide wastewater SARS-CoV-2 levels strongly correlated with multiple disease surveillance indicators and outcomes over three COVID-19 waves. *Science of The Total Environment*, 855, 158967. <https://doi.org/https://doi.org/10.1016/j.scitotenv.2022.158967>
- HRSD. (2023). *HRSD Wastewater Surveillance*. Retrieved July 18, 2023 from <https://www.hrsd.com/HRSD-Wastewater-Surveillance>
- Hsiung, G. D. (1984). Diagnostic virology: from animals to automation. *Yale J Biol Med*, 57(5), 727-733.
- Hutwagner, L.C., Maloney, E.K., Bean, N.H., Slutsker, L., Martin, S.M. (1997). Using laboratory-based surveillance data for prevention: an algorithm for detecting Salmonella outbreaks. *Emerg Infect Dis*, 3(3). https://wwwnc.cdc.gov/eid/article/3/3/97-0322_article
- Hutwagner, L., Thompson, W., Seeman, G.M., and Treadwell, T. (2003). The Bioterrorism Preparedness and Response Early Aberration Reporting System (EARS). *Journal of Urban Health*, 80(2).
- Jakubowski, W., Sykora, J. L., Sorber, C. A., Casson, L. W., & Gavaghan, P. D. (1991). Determining Giardiasis Prevalence by Examination of Sewage. *Water Science and Technology*, 24(2), 173-178. <https://www.proquest.com/scholarly-journals/determining-giardiasis-prevalence-examination/docview/1943295195/se-2?accountid=14826>
- Jiang, X., Luo, M., Zou, Z., Wang, X., Chen, C., & Qiu, J. (2020). Asymptomatic SARS-CoV-2 infected case with viral detection positive in stool but negative in nasopharyngeal samples lasts for 42 days. *Journal of Medical Virology*, 92(10), 1807-1809. <https://doi.org/https://doi.org/10.1002/jmv.25941>
- Lamers, M. M., Beumer, J., van der Vaart, J., Knoops, K., Puschhof, J., Breugem, T. I., Ravelli, R. B. G., Paul van Schayck, J., Mykityn, A. Z., Duimel, H. Q., van Donselaar, E.,

- Riesebosch, S., Kuijpers, H. J. H., Schipper, D., van de Wetering, W. J., de Graaf, M., Koopmans, M., Cuppen, E., Peters, P. J., . . . Clevers, H. (2020). SARS-CoV-2 productively infects human gut enterocytes. *Science*, 369(6499), 50-54. <https://doi.org/10.1126/science.abc1669>
- Li, X., Zhang, S., Shi, J., Luby, S. P., & Jiang, G. (2021). Uncertainties in estimating SARS-CoV-2 prevalence by wastewater-based epidemiology. *Chemical Engineering Journal*, 415, 129039. <https://doi.org/10.1016/j.cej.2021.129039>
- Lickness, J. S. (2020). Surveillance to track progress toward polio eradication—worldwide, 2018–2019. *MMWR. Morbidity and Mortality Weekly Report*, 69.
- LUND, E., & HEDSTRÖM, C. E. (1966). THE USE OF AN AQUEOUS POLYMER PHASE SYSTEM FOR ENTEROVIRUS ISOLATIONS FROM SEWAGE12. *American Journal of Epidemiology*, 84(2), 287-291. <https://doi.org/10.1093/oxfordjournals.aje.a120642>
- Lustig, A., & Levine, A. J. (1992). One hundred years of virology. *Journal of virology*, 66(8), 4629-4631.
- Mallapaty, S. (2020). How sewage could reveal true scale of coronavirus outbreak. *Nature*, 580(7802), 176-177.
- Mandl, K. D., et al. (2004). Implementing syndromic surveillance: a practical guide informed by the early experience. *Journal of the American Medical Informatics Association : JAMIA*, 11(2), 141–150. <https://doi.org/10.1197/jamia.M1356>
- Martins, M. T., Pessoa, G. V. A., Sanchez, P. S., Sato, M. I. Z., Coimbra, C. A., Monteiro, C. K., & Marques, E. (1991). Occurrence of *V. cholerae* O:1 non-toxicogenic in wastewaters from Sao Paulo, Brazil. <https://www.proquest.com/books/occurrence-v-cholerae-o-1-non-toxicogenic/docview/15885436/se-2?accountid=14826>
- Martins, M. T., Soares, L. A., Marques, E., & Molina, A. G. (1983). Human Enteric Viruses Isolated from Influent of Sewage Treatment Plants in S. Paulo, Brazil. *Water Science and Technology*, 15(5), 69-73. <https://www.proquest.com/scholarly-journals/human-enteric-viruses-isolated-influent-sewage/docview/1943164126/se-2?accountid=14826>
- Mayer, C. L., & Palmer, C. J. (1996). Evaluation of PCR, nested PCR, and fluorescent antibodies for detection of *Giardia* and *Cryptosporidium* species in wastewater. *Applied and Environmental Microbiology*, 62(6), 2081-2085. <https://doi.org/doi:10.1128/aem.62.6.2081-2085.1996>
- McClary-Gutierrez, J., Mattioli, M., Marcenac, P., Silverman, A., Boehm, A., Bibby, K., Balliet, M., de los Reyes, F., Gerrity, D., Griffith, J., Holden, P., Katehis, D., Kester, G., LaCross, N., Lipp, E., Meiman, J., Noble, R., Brossard, D., & McLellan, S. (2021). SARS-CoV-2 Wastewater Surveillance for Public Health Action. *Emerging Infectious Disease journal*, 27(9), 1. <https://doi.org/10.3201/eid2709.210753>
- Medema, G., Been, F., Heijnen, L., & Pettersson, S. (2020). Implementation of environmental surveillance for SARS-CoV-2 virus to support public health decisions: Opportunities and challenges. *Current Opinion in Environmental Science & Health*, 17, 49-71. <https://doi.org/10.1016/j.coesh.2020.09.006>

- Medema, G., Heijnen, L., Elsinga, G., Italiaander, R., & Brouwer, A. (2020). Presence of SARS-Coronavirus-2 RNA in Sewage and Correlation with Reported COVID-19 Prevalence in the Early Stage of the Epidemic in The Netherlands. *Environmental Science & Technology Letters*, 7(7), 511-516. <https://doi.org/10.1021/acs.estlett.0c00357>
- Mehnert, D. U., & Stewien, K. E. (1993). Detection and distribution of rotavirus in raw sewage and creeks in Sau Paulo, Brazil. *Applied and Environmental Microbiology*, 59(1), 140-143. <https://doi.org/doi:10.1128/aem.59.1.140-143.1993>
- Monahan, T. (2017). Regulating belonging: surveillance, inequality, and the cultural production of abjection. *Journal of Cultural Economy*, 10(2), 191-206. <https://doi.org/10.1080/17530350.2016.1273843>
- Morris, R., & Sharp, D. N. (1984). Enteric virus levels in wastewater effluents and surface waters in the Severn Trent Water Authority 1979-1981. *Water Research*, 18(8), 935-939. <https://www.proquest.com/scholarly-journals/enteric-virus-levels-wastewater-effluents-surface/docview/13964801/se-2?accountid=14826>
- Morton, A. P., Whitby, M., McLaws, M.-L., Dobson, A., McElwain, S., Looke, D., Stackelroth, J., & Sartor, A. (2001). The application of statistical process control charts to the detection and monitoring of hospital-acquired infections. *Journal of Quality in Clinical Practice*, 21(4), 112-117. <https://doi-org.ezproxy.lib.vt.edu/10.1046/j.1440-1762.2001.00423.x>
- Nashida, M., Shimohara, E., Sugishima, S., & Kaneko, M. (1989). Viruses Isolated from Activated Sludges, 1984-1986. *Water Science and Technology*, 21(3), 279-282. <https://www.proquest.com/scholarly-journals/viruses-isolated-activated-sludges-1984-1986/docview/1943174364/se-2?accountid=14826>
- NCBI. (2023). *NCBI Virus: SARS-CoV-2 Variants Overview*. Retrieved July 19, 2023 from <https://www.ncbi.nlm.nih.gov/activ>
- NWBE. (2023). *WEF Network of Wastewater-Based Epidemiology*. Retrieved July 19, 2023 from <https://nwbe.org/>
- NIST. (2024). 6.3.2.2. *Individuals Control Charts*. National Institute of Standards and Technology. Retrieved November 21, 2024 from <https://www.itl.nist.gov/div898/handbook/pmc/section3/pmc322.htm>
- Paul, J. R., Trask, J. D., & Gard, S. (1940). II. POLIOMYELITIC VIRUS IN URBAN SEWAGE. *Journal of Experimental Medicine*, 71(6), 765-777. <https://doi.org/10.1084/jem.71.6.765>
- Pina, S., Jofre, J., Emerson, S. U., Purcell, R. H., & Girones, R. (1998). Characterization of a Strain of Infectious Hepatitis E Virus Isolated from Sewage in an Area where Hepatitis E Is Not Endemic. *Applied and Environmental Microbiology*, 64(11), 4485-4488. <https://doi.org/doi:10.1128/AEM.64.11.4485-4488.1998>
- Rich, W.H. & Terry M.C. (1946). Public Health Weekly Reports for OCTOBER 18, 1946. *Public Health Rep (1896)*. 1946 Oct 18;61(42):1501-1528. PMID: 19316100; PMCID: PMC1976172.

- Riordan, J. T. (1962). The 1961 Middletown Oral Poliovirus Vaccine Program. IX. Isolation of enteroviruses from sewage before and after vaccine administration. *Yale J Biol Med*, 34(5), 512-521.
- Robertson, L. J., Smith, P. G., Grimason, A. T., & Smith, H. V. (1999). Removal and destruction of intestinal parasitic protozoans by sewage treatment processes. *International Journal of Environmental Health Research*, 9(2), 85-96. <https://doi.org/10.1080/09603129973227>
- Safferman, R. S. (1982). Viruses in wastewaters. *Environment International [ENVIRON. INT.]*. Vol. 7, no. 1. 1982., 7(1). <https://www.proquest.com/scholarly-journals/viruses-wastewaters/docview/13755840/se-2?accountid=14826>
- Sellwood, J., Dadswell, J. V., & Slade, J. S. (1981). Viruses in sewage as an indicator of their presence in the community. *Journal of Hygiene*, 86(2), 217-225. <https://doi.org/10.1017/S0022172400068947>
- Serfling, R. E. (1963). A statistical algorithm for the early detection of outbreaks of infectious diseases. *Journal of the American Statistical Association*, 58(301), 947-954. <https://pmc.ncbi.nlm.nih.gov/articles/PMC1915276/>
- Shewhart, W.A. (1931). Statistical Method From an Engineering Viewpoint. *Journal of the American Statistical Association*, 26(175), 62-269.
- Microsoft Support (2023). *CORREL function*. Retrieved November 1, 2023 from <https://support.microsoft.com/en-gb/office/correl-function-995dcef7-0c0a-4bed-a3fb-239d7b68ca92>
- Sykora, J. L., Sorber, C. A., Jakubowski, W., Casson, L. W., Gavaghan, P. D., Shapiro, M. A., & Schott, M. J. (1991). Distribution of Giardia Cysts in Wastewater. *Water Science and Technology*, 24(2), 187-192. <https://www.proquest.com/scholarly-journals/distribution-giardia-cysts-wastewater/docview/1943314307/se-2?accountid=14826>
- Tambini, G., Andrus, J. K., Marques, E., Boshell, J., Pallansch, M., De Quadros, G. A., & Kew, O. (1993). Direct Detection of Wild Poliovirus Circulation by Stool Surveys of Healthy Children and Analysis of Community Wastewater. *The Journal of Infectious Diseases*, 168(6), 1510-1514. <http://www.jstor.org.ezproxy.lib.vt.edu/stable/30113705>
- Taylor, M. W. (2014). A History of Cell Culture. In M. W. Taylor (Ed.), *Viruses and Man: A History of Interactions* (pp. 41-52). Springer International Publishing. https://doi.org/10.1007/978-3-319-07758-1_3
- Thomas, J. C., Sage, M., Dillenberg, J., & Guillory, V. J. (2002). A Code of Ethics for Public Health. *American Journal of Public Health*, 92(7), 1057-1059. <https://doi.org/10.2105/ajph.92.7.1057>
- Thompson, J. R., Nanchaiah, Y. V., Gu, X., Lee, W. L., Rajal, V. B., Haines, M. B., Girones, R., Ng, L. C., Alm, E. J., & Wuertz, S. (2020). Making waves: Wastewater surveillance of SARS-CoV-2 for population-based health management. *Water Research*, 184, 116181. <https://doi.org/https://doi.org/10.1016/j.watres.2020.116181>
- Tokars, J. I., Burkom, H., Xing, J., English, R., Bloom, S., Cox, K., & Pavlin, J. A. (2009). Enhancing Time-Series Detection Algorithms for Automated Biosurveillance. *Emerging infectious diseases*, 15(4), 533-539. <https://doi.org/10.3201/eid1504.080616>

- Vantarakis, A., & Papapetropoulou, M. (1999). Detection of enteroviruses, adenoviruses and Hepatitis A viruses in raw sewage and treated effluents by nested-PCR. *Water, Air, & Soil Pollution*, 114(1-2), 85-93. <https://doi.org/https://doi.org/10.1023/A:1005065326395>
- VDGS. (2023). *Division of Consolidated Laboratory Services*. Virginia Department of General Services. Retrieved May 5, 2023 from <https://dgs.virginia.gov/division-of-consolidated-laboratory-services>
- VDH. (2023a). *SARS-CoV-2 in Wastewater*. Retrieved December 8th, 2023 from <https://www.vdh.virginia.gov/coronavirus/see-the-numbers/covid-19-data-insights/sars-cov-2-in-wastewater/>
- VDH. (2023b, October 10, 2023). *VDH COVID-19 Public Use Dataset: Cases*. Retrieved October 14, 2023 from <https://data.virginia.gov/Government/VDH-COVID-19-PublicUseDataset-Cases/bre9-aqqr>
- VDH. (2023c, April 11, 2023). *VDH COVID-19 Public Use Dataset: Variants of Concern*. Retrieved July 19, 2023 from https://data.virginia.gov/Government/VDH-COVID-19-PublicUseDataset-Variants_of_Concern/ji2y-fake
- VDH. (2023d). *VDH Wastewater Surveillance for COVID-19*. Retrieved July 19, 2023 from <https://www.vdh.virginia.gov/environmental-health/wastewater-surveillance-for-covid-19/>
- VDH. (2023e). *What is Syndromic Surveillance?*. Retrieved Nov 16, 2024 from <https://www.vdh.virginia.gov/surveillance-and-investigation/syndromic-surveillance/about/>
- VDH. (2024). *Wastewater Surveillance Sentinel Monitoring Program*. Virginia Department of Health. Retrieved November 19, 2024 from https://www.vdh.virginia.gov/virginia_essence/servlet/DetectorAlgorithmsServlet#
- VDH. (2024). *ESSENCE Algorithms for Univariate Temporal Alerting*. Virginia Department of Health. Retrieved March 15, 2023 from <https://www.vdh.virginia.gov/environmental-health/wastewater-surveillance-for-covid-19/>
- VDH. (2025). *VDH Wastewater Surveillance for COVID-19*. Retrieved February 26, 2025 from <https://www.vdh.virginia.gov/coronavirus/sars-cov-2-in-wastewater/>
- Wallenstein, S., & Naus, J. (2004). Scan Statistics for Temporal Surveillance for Biologic Terrorism. *MMWR. Morbidity and Mortality Weekly Report*, 53, 74–78. <https://www.jstor.org/stable/23315694>
- Waqas, M. Xu, S.H., Usman Aslam, M., Hussain, S., Shahzad, K., & Masengo, G. (2020). Global contribution of statistical control charts to epidemiology monitoring: A 23-year analysis with optimized EWMA real-life application on COVID-19. *Medicine* 103(27):p e38766. Retrieved from https://journals.lww.com/md-journal/fulltext/2024/07050/global_contribution_of_statistical_control_charts.51.aspx#T1
- WEF. (2023). *Network of Wastewater-Based Epidemiology (NWBE): Frequently Asked Questions*. Water Environment Federation (WEF). Retrieved March 19, 2023 from https://nwbe.org/?page_id=77

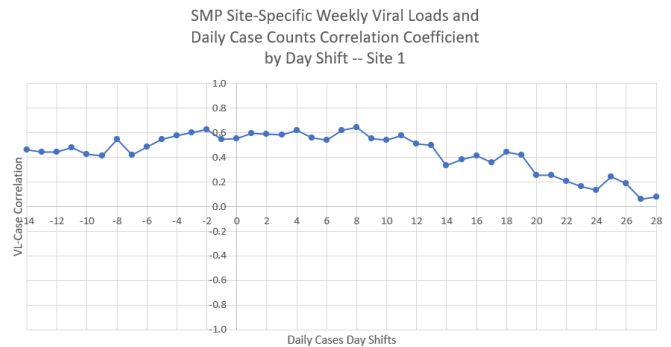
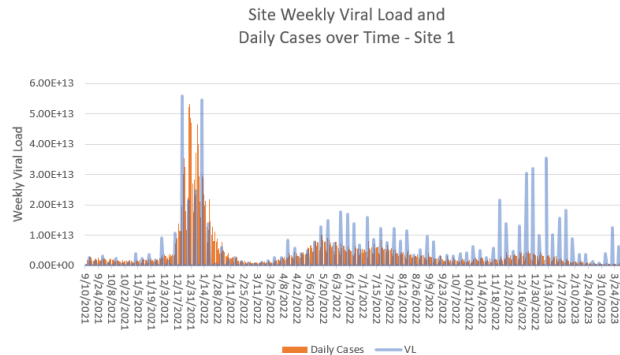
- WHO. (1979). *Human viruses in water, wastewater and soil: report of a WHO scientific group [meeting held in Geneva from 23 to 27 October 1978]*. World Health Organization.
- WHO. (2023). *Global Polio Eradication Initiative*. World Health Organization. Retrieved May 13, 2023 from <https://polioeradication.org/>
- Wilson, W. J. (1933). ISOLATION OF ENTERIC BACILLI FROM SEWAGE AND WATER AND ITS BEARING ON EPIDEMIOLOGY. *Br Med J*, 2(3794), 560-562. <https://doi.org/10.1136/bmj.2.3794.560>
- Xu, Y., Li, X., Zhu, B., Liang, H., Fang, C., Gong, Y., Guo, Q., Sun, X., Zhao, D., Shen, J., Zhang, H., Liu, H., Xia, H., Tang, J., Zhang, K., & Gong, S. (2020). Characteristics of pediatric SARS-CoV-2 infection and potential evidence for persistent fecal viral shedding. *Nature Medicine*, 26(4), 502-505. <https://doi.org/10.1038/s41591-020-0817-4>
- Yang, Weizhong (2017). *Early Warning for Infectious Disease Outbreak: Theory and Practice*. Academic Press.
- Youngkin, G. G. (2023). *Executive Order 26*. Commonwealth of Virginia. Retrieved from <https://www.governor.virginia.gov/media/governorvirginiagov/governor-of-virginia/pdf/eo/EO-26-Crushing-the-Fentanyl-Epidemic-Strengthening-Virginia's-Interdiction-and-Enforcement-Response-to-Fentanyl-Crisis.pdf>
- Zhang, H., Kang, Z., Gong, H., Xu, D., Wang, J., Li, Z., Li, Z., Cui, X., Xiao, J., Zhan, J., Meng, T., Zhou, W., Liu, J., & Xu, H. (2020). Digestive system is a potential route of COVID-19: an analysis of single-cell coexpression pattern of key proteins in viral entry process. *Gut*, 69(6), 1010-1018. <https://doi.org/10.1136/gutjnl-2020-320953>

Appendix – 8.5 By Site VL & Daily Cases

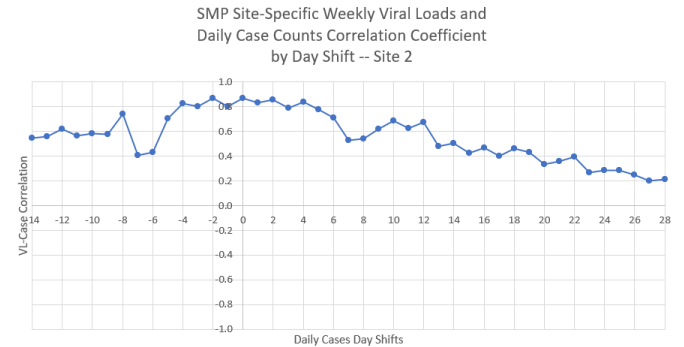
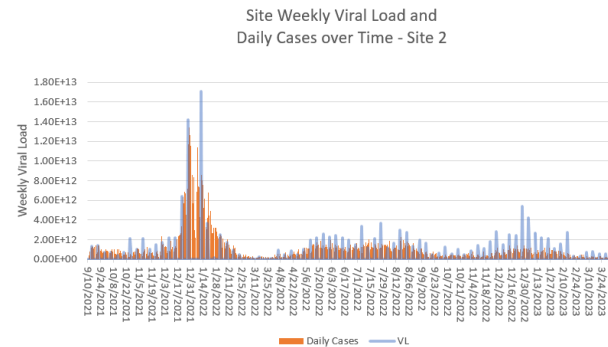
Some sites withdrew from SMP early (*).

Note: secondary axes are intentionally not labeled, to protect case anonymity.

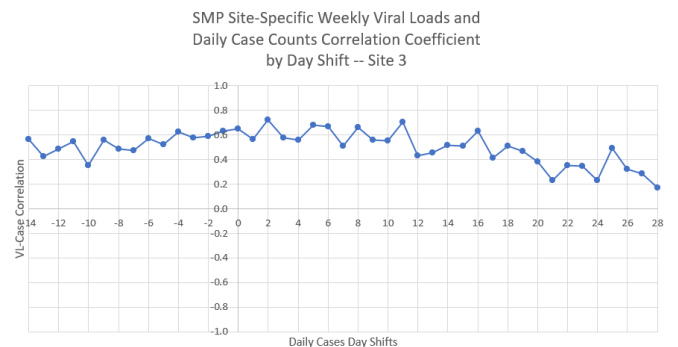
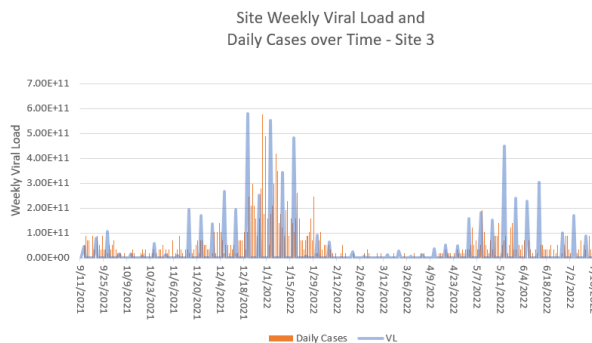
Site 1 – AlexRenew



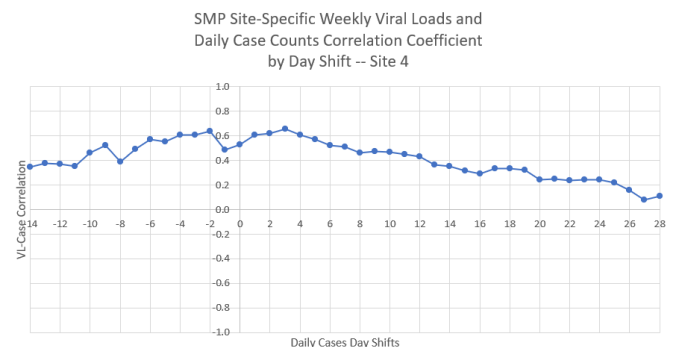
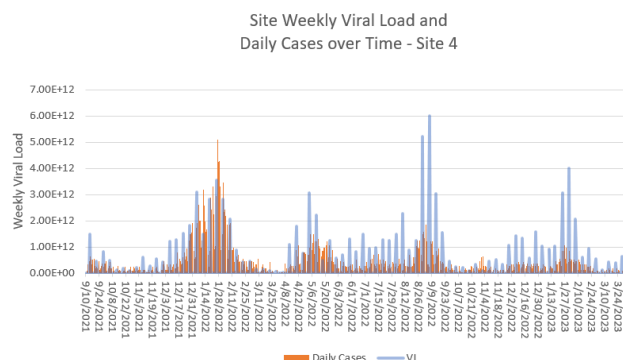
Site 2 – Aquia



Site 3 – Basham*



Site 4 – Blacksburg

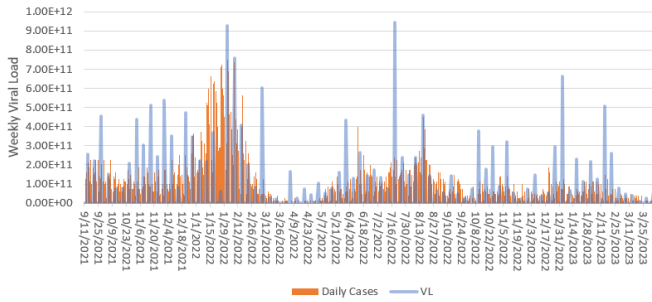


Site 5 – Danville*

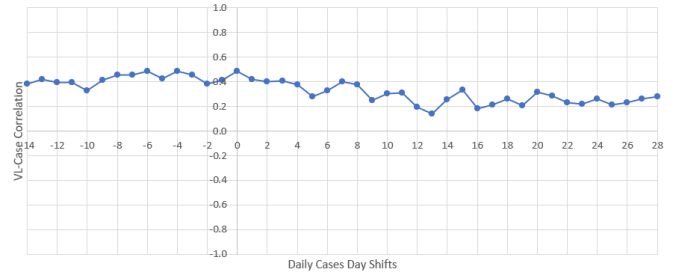
[sewershed unavailable]

Site 6 – Coeburn

Site Weekly Viral Load and
Daily Cases over Time - Site 6

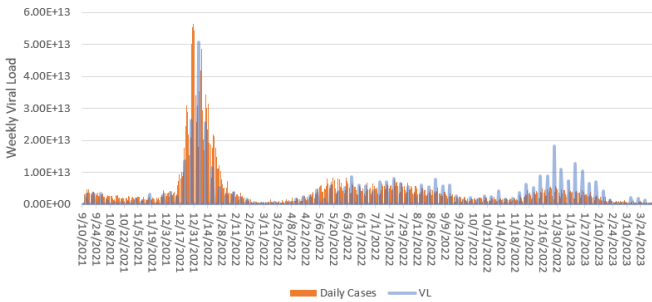


SMP Site-Specific Weekly Viral Loads and
Daily Case Counts Correlation Coefficient
by Day Shift -- Site 6

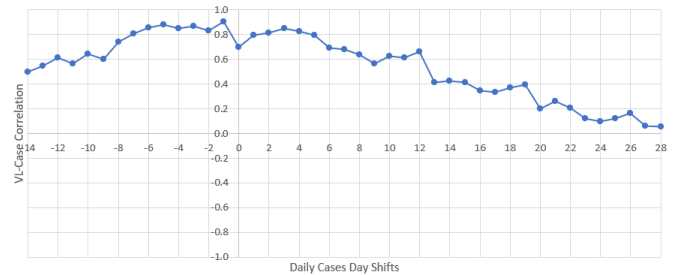


Site 7 – HLM

Site Weekly Viral Load and
Daily Cases over Time - Site 7

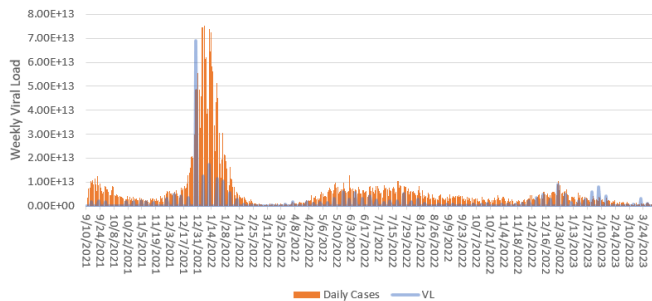


SMP Site-Specific Weekly Viral Loads and
Daily Case Counts Correlation Coefficient
by Day Shift -- Site 7

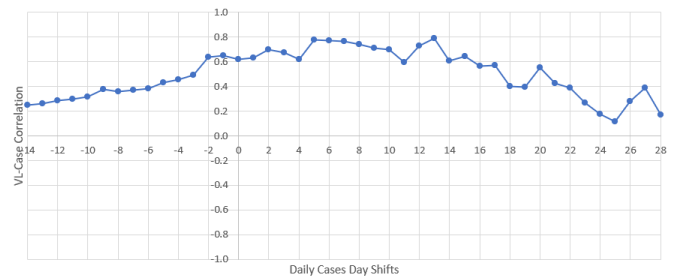


Site 8 – NA

Site Weekly Viral Load and
Daily Cases over Time - Site 8

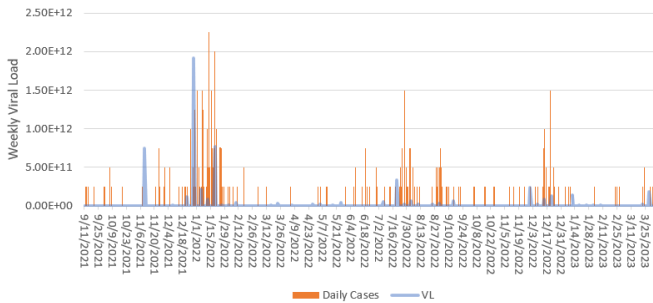


SMP Site-Specific Weekly Viral Loads and
Daily Case Counts Correlation Coefficient
by Day Shift -- Site 8

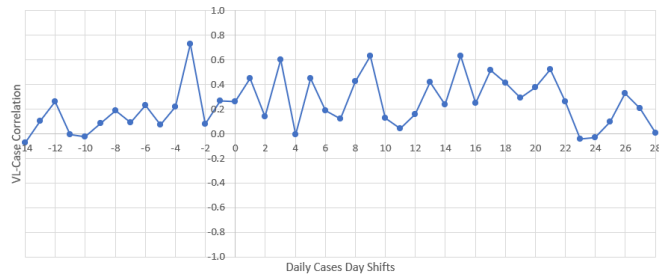


Site 9 – ON

Site Weekly Viral Load and
Daily Cases over Time - Site 9

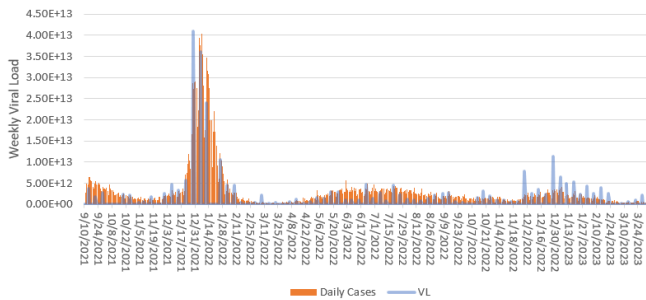


SMP Site-Specific Weekly Viral Loads and
Daily Case Counts Correlation Coefficient
by Day Shift -- Site 9

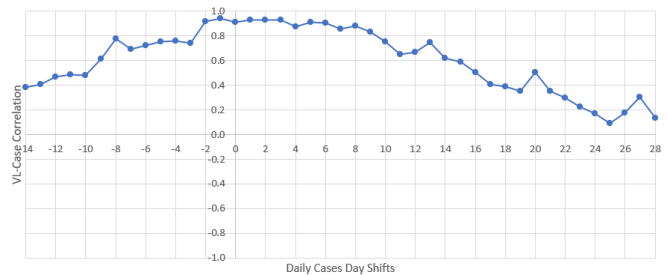


Site 10 – VIP

Site Weekly Viral Load and
Daily Cases over Time - Site 10

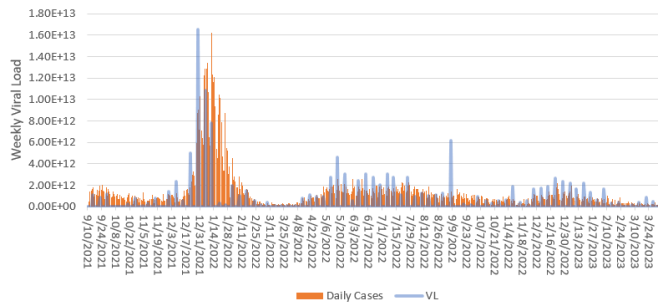


SMP Site-Specific Weekly Viral Loads and
Daily Case Counts Correlation Coefficient
by Day Shift -- Site 10

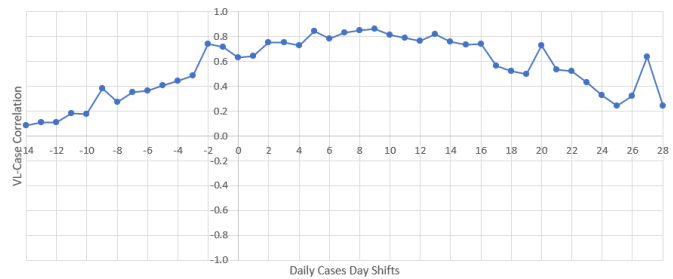


Site 11 – WB

Site Weekly Viral Load and
Daily Cases over Time - Site 11

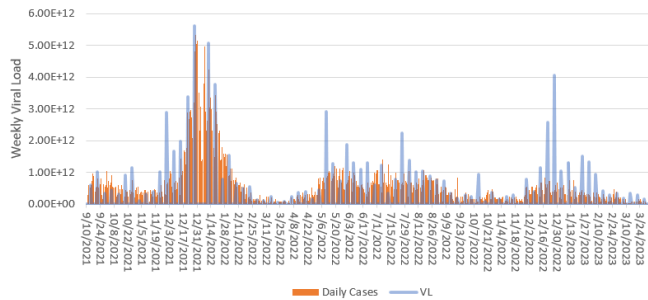


SMP Site-Specific Weekly Viral Loads and
Daily Case Counts Correlation Coefficient
by Day Shift -- Site 11

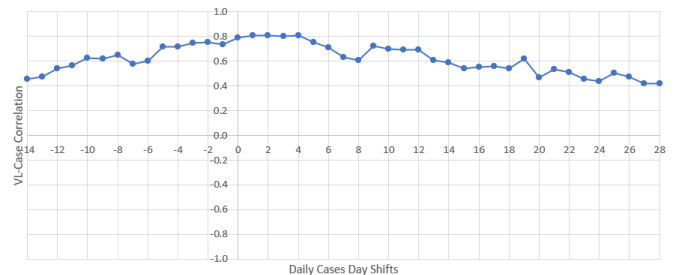


Site 12 – LFR

Site Weekly Viral Load and
Daily Cases over Time - Site 12

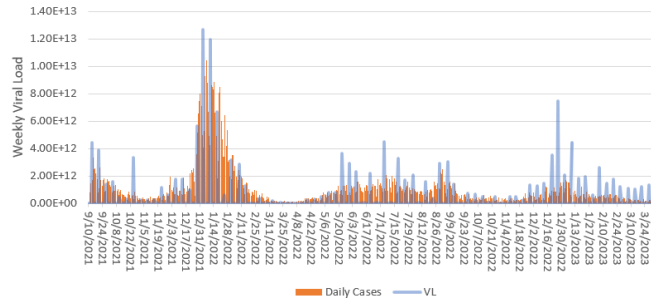


SMP Site-Specific Weekly Viral Loads and
Daily Case Counts Correlation Coefficient
by Day Shift -- Site 12

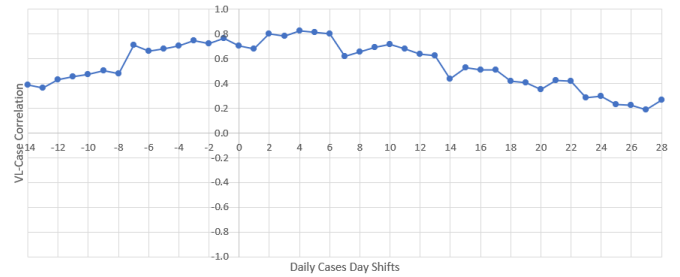


Site 13 – LB

Site Weekly Viral Load and
Daily Cases over Time - Site 13

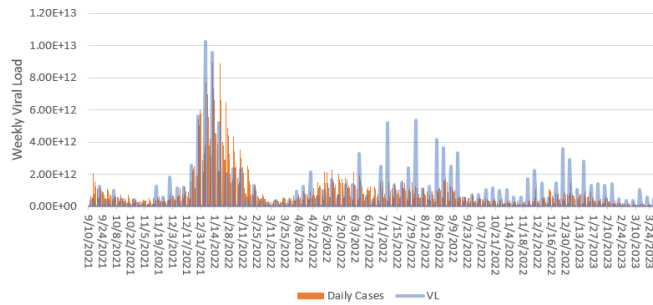


SMP Site-Specific Weekly Viral Loads and
Daily Case Counts Correlation Coefficient
by Day Shift -- Site 13

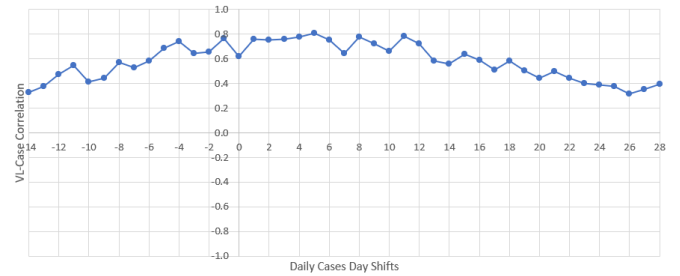


Site 14 – Moores

Site Weekly Viral Load and
Daily Cases over Time - Site 14

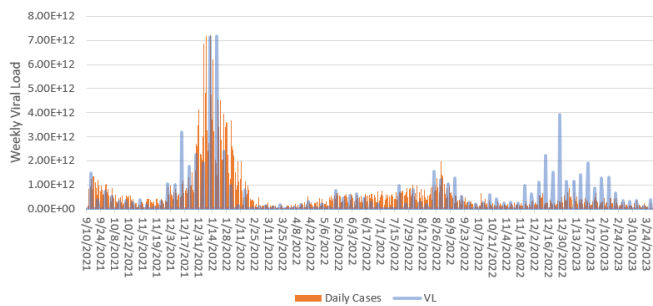


SMP Site-Specific Weekly Viral Loads and
Daily Case Counts Correlation Coefficient
by Day Shift -- Site 14

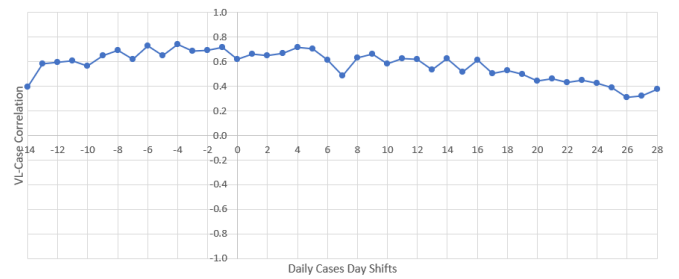


Site 15 – NorthRiver

Site Weekly Viral Load and
Daily Cases over Time - Site 15

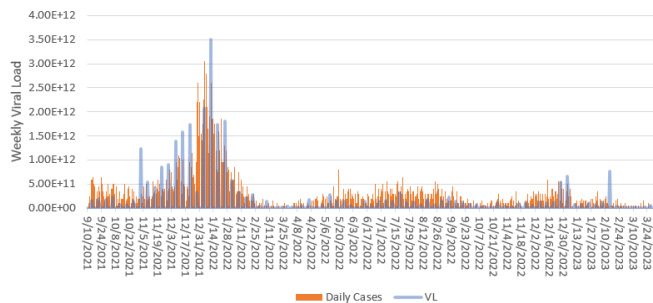


SMP Site-Specific Weekly Viral Loads and
Daily Case Counts Correlation Coefficient
by Day Shift -- Site 15

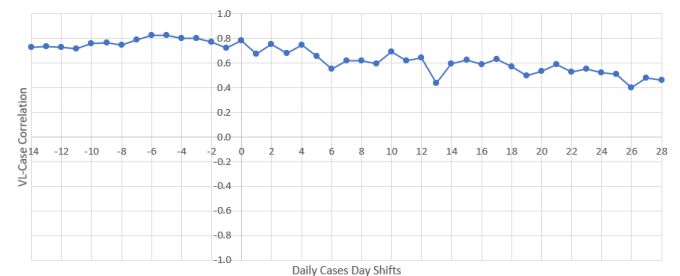


Site 16 – Parkins

Site Weekly Viral Load and
Daily Cases over Time - Site 16

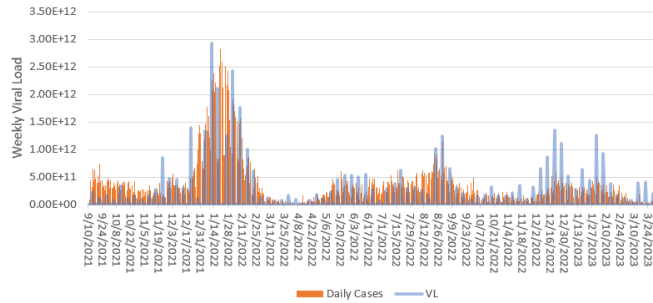


SMP Site-Specific Weekly Viral Loads and
Daily Case Counts Correlation Coefficient
by Day Shift -- Site 16

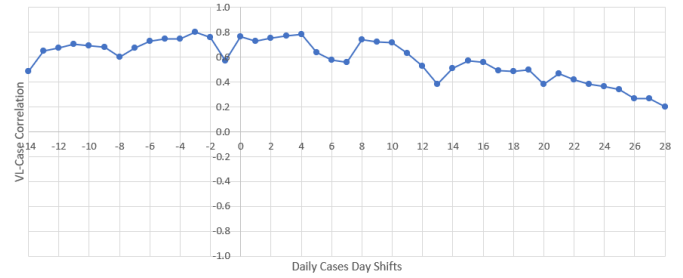


Site 17 – Peppers

Site Weekly Viral Load and Daily Cases over Time - Site 17

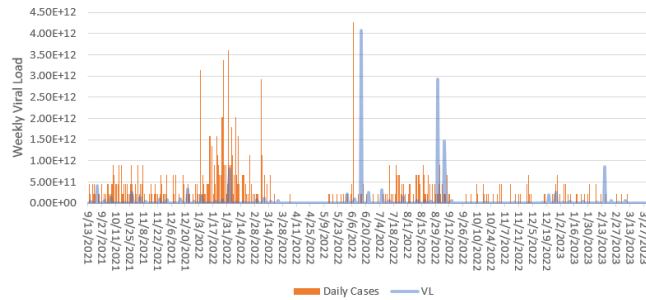


SMP Site-Specific Weekly Viral Loads and Daily Case Counts Correlation Coefficient by Day Shift -- Site 17

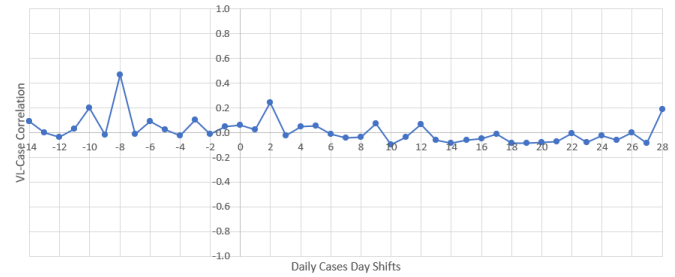


Site 18 – Pound

Site Weekly Viral Load and Daily Cases over Time - Site 18

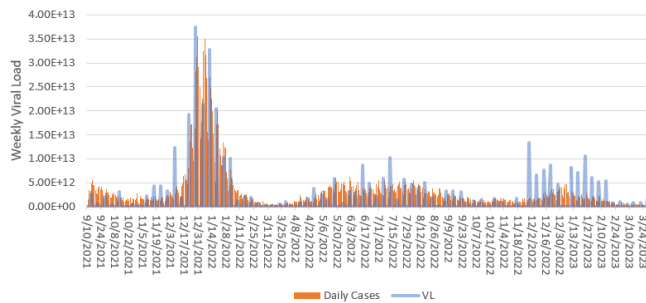


SMP Site-Specific Weekly Viral Loads and Daily Case Counts Correlation Coefficient by Day Shift -- Site 18

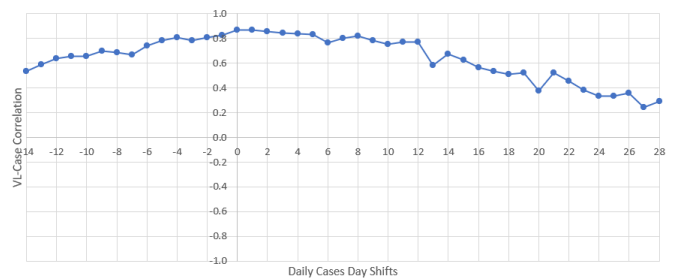


Site 19 – Richmond

Site Weekly Viral Load and Daily Cases over Time - Site 19

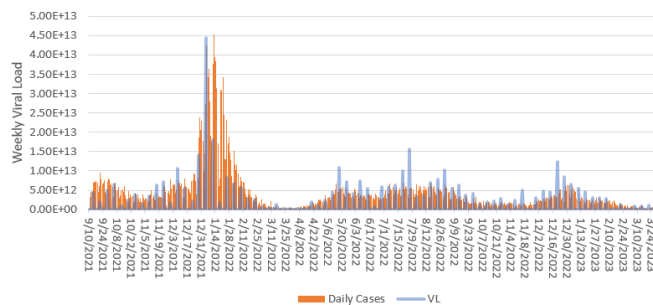


SMP Site-Specific Weekly Viral Loads and Daily Case Counts Correlation Coefficient by Day Shift -- Site 19

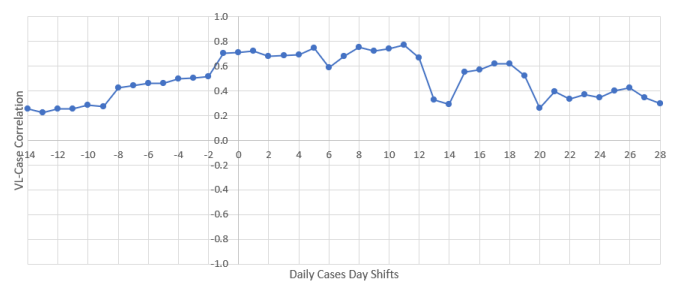


Site 20 – Roanoke

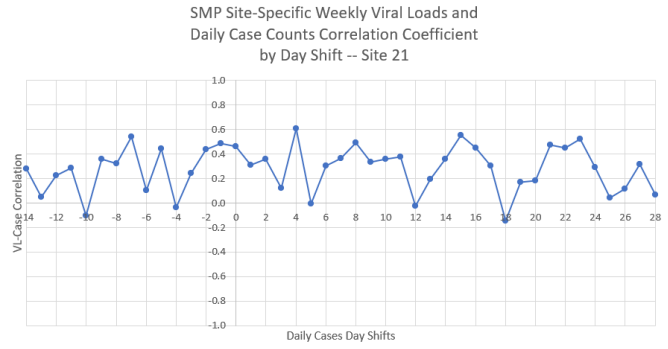
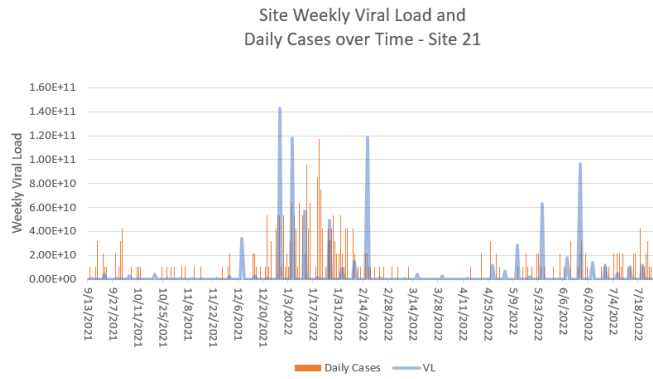
Site Weekly Viral Load and Daily Cases over Time - Site 20



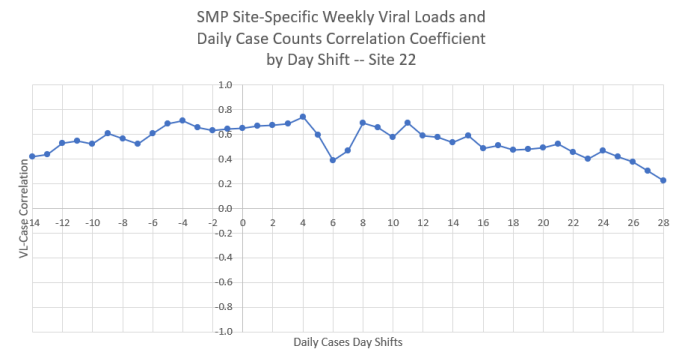
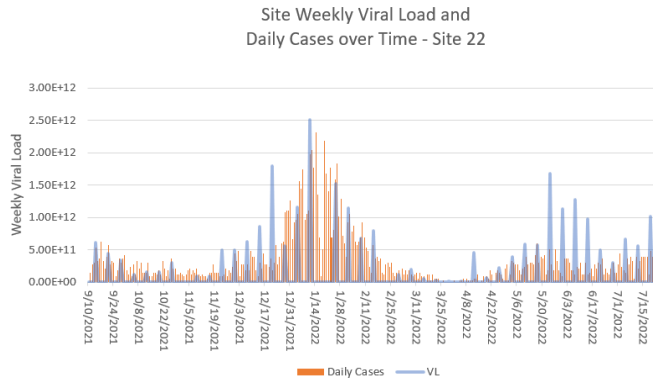
SMP Site-Specific Weekly Viral Loads and Daily Case Counts Correlation Coefficient by Day Shift -- Site 20



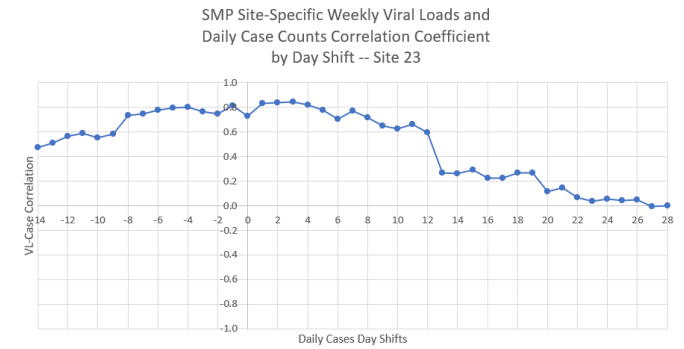
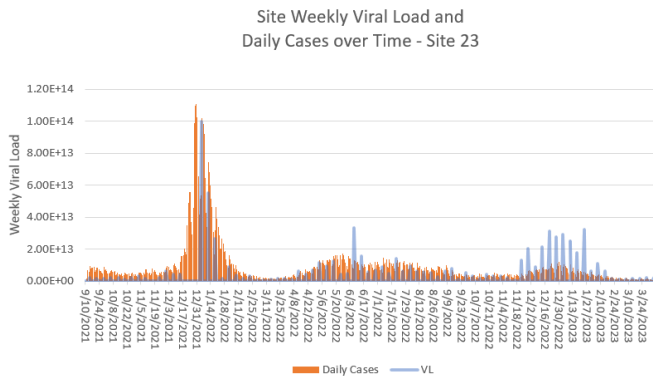
Site 21 – Rutledge*



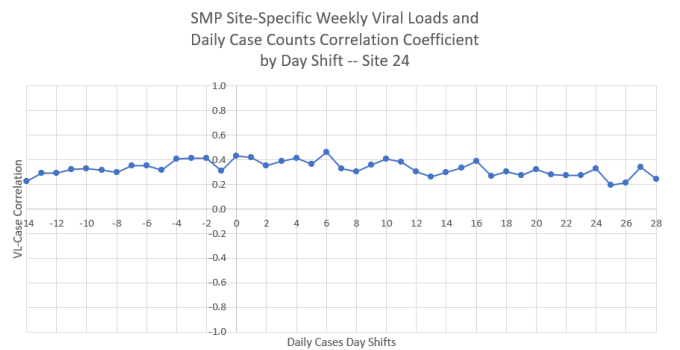
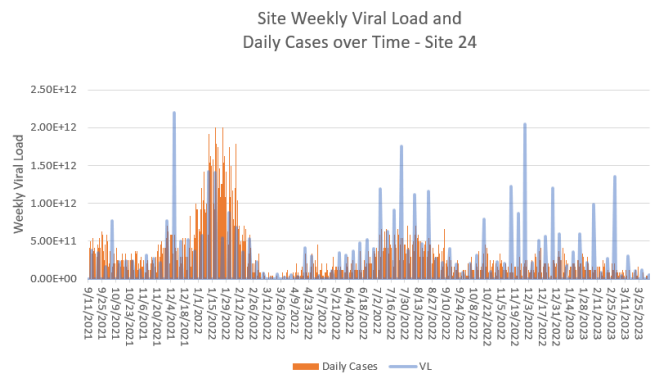
Site 22 – CB*



Site 23 – UOSA



Site 24 – Wolf

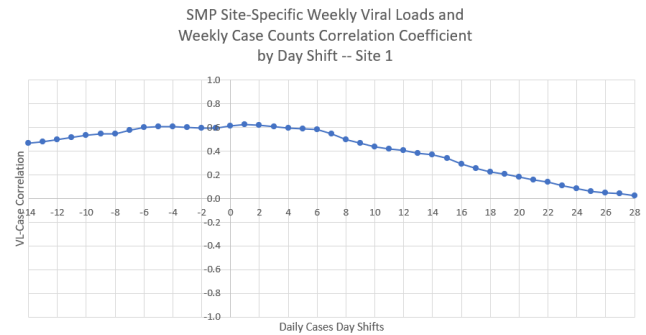
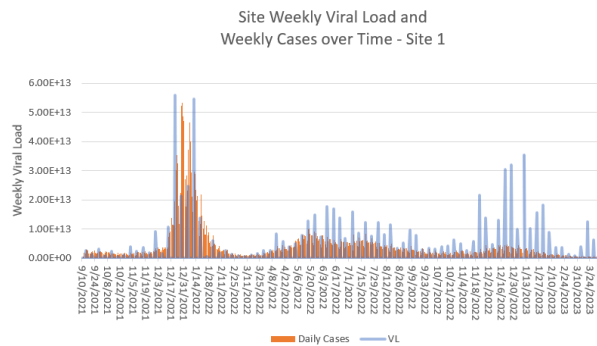


Appendix – 8.6 By Site VL & Weekly Cases

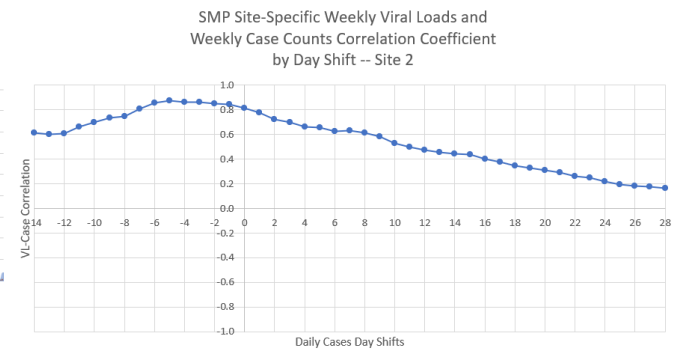
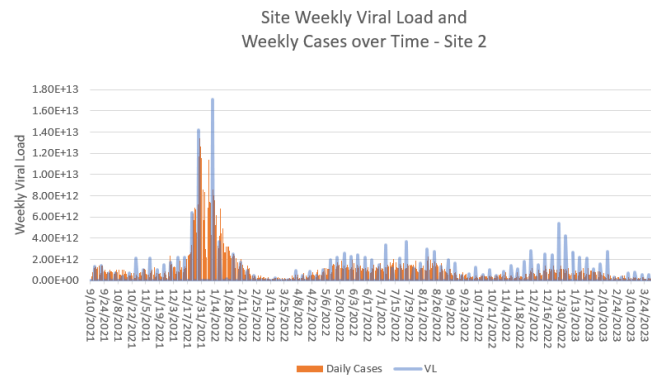
Some sites withdrew from SMP early (*).

Note: secondary axes are intentionally not labeled, to protect case anonymity.

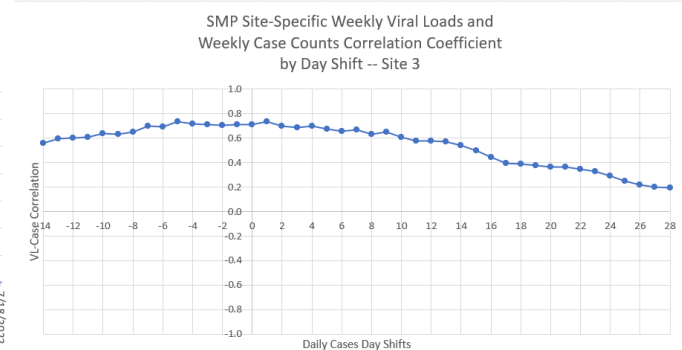
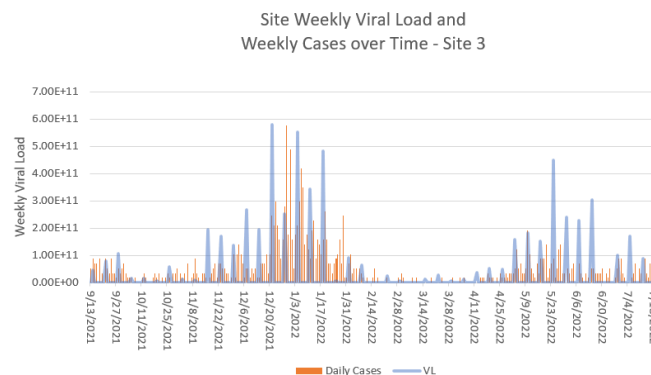
Site 1 – AlexRenew



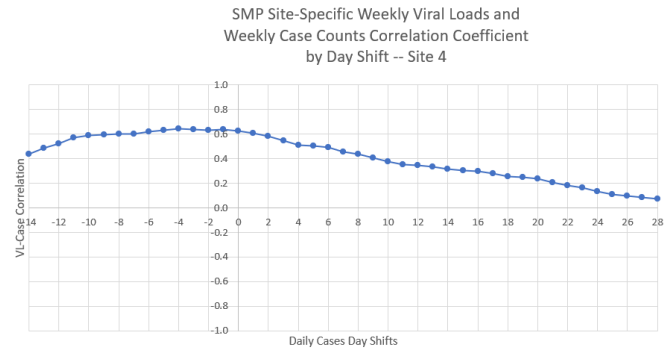
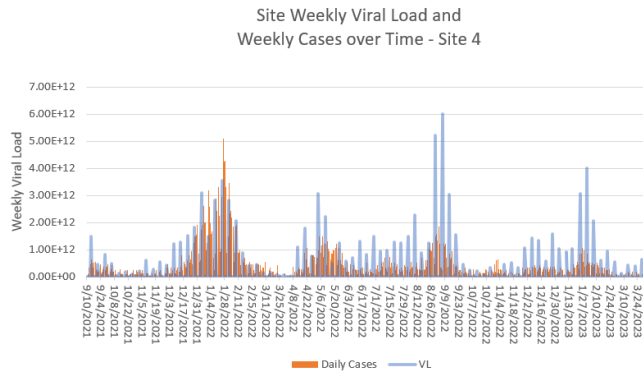
Site 2 – Aquia



Site 3 – Basham*

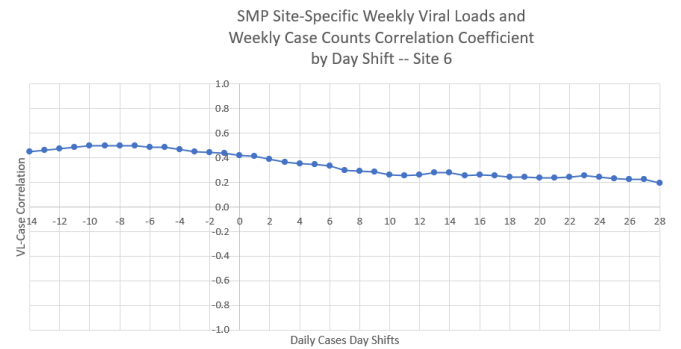
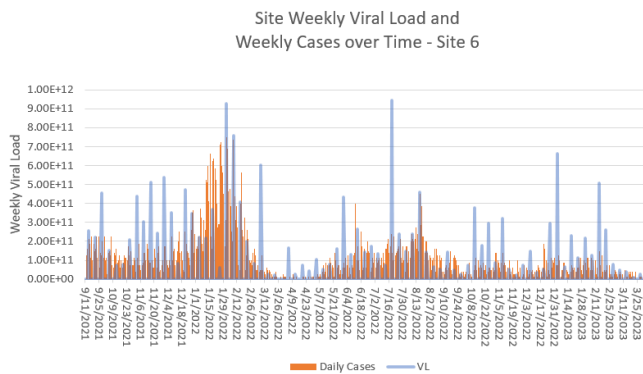


Site 4 – Blacksburg

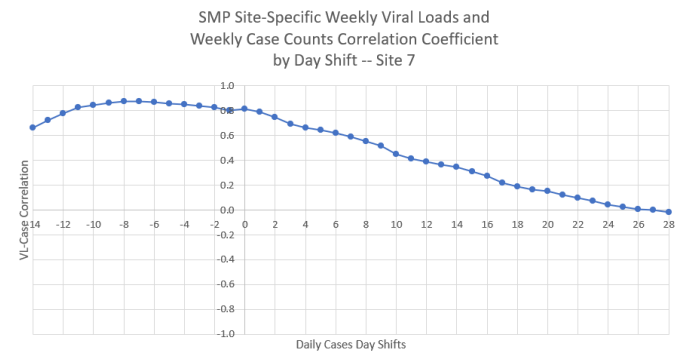
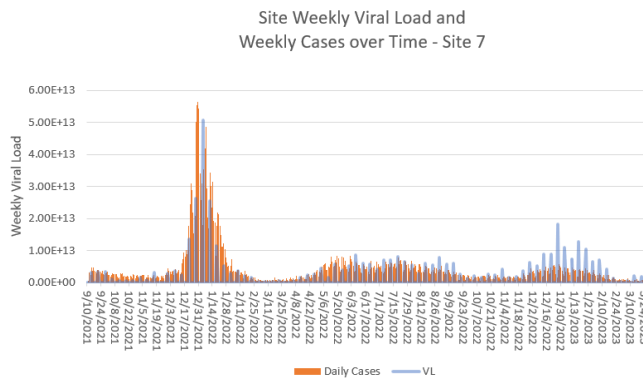


Site 5 – Danville* [sewershed unavailable]

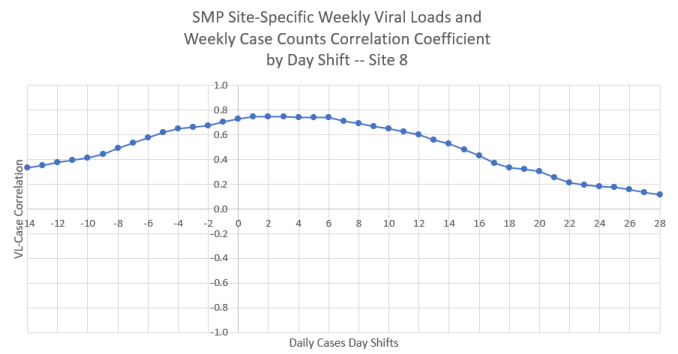
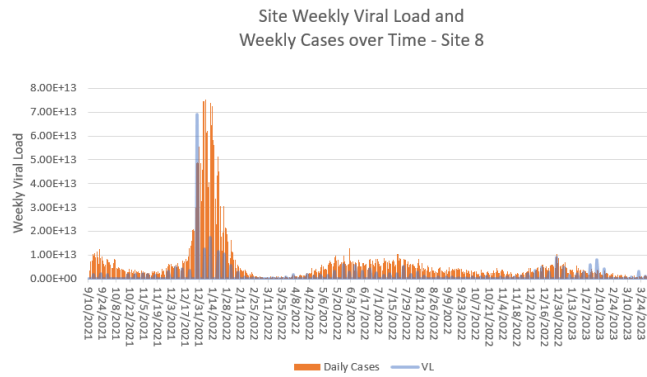
Site 6 – Coeburn



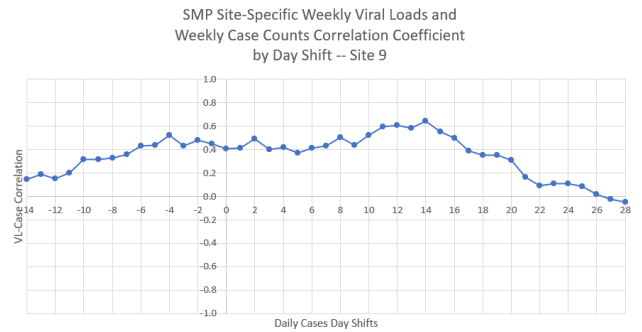
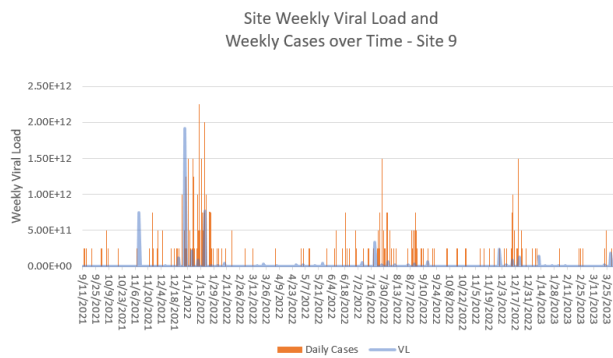
Site 7 – HLM



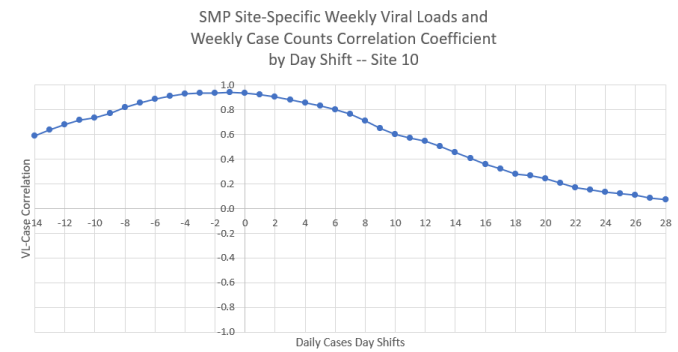
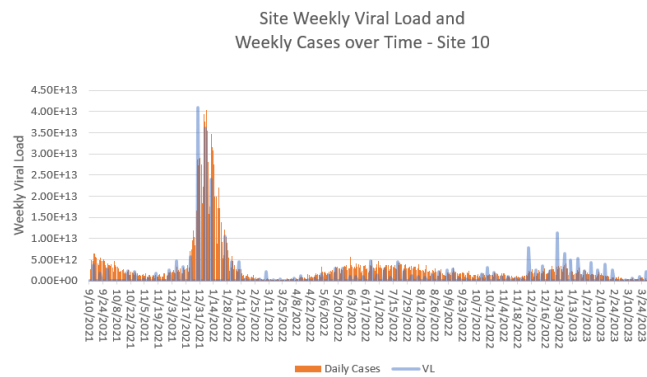
Site 8 – NA



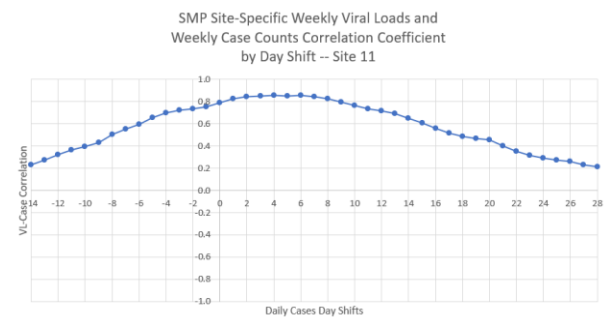
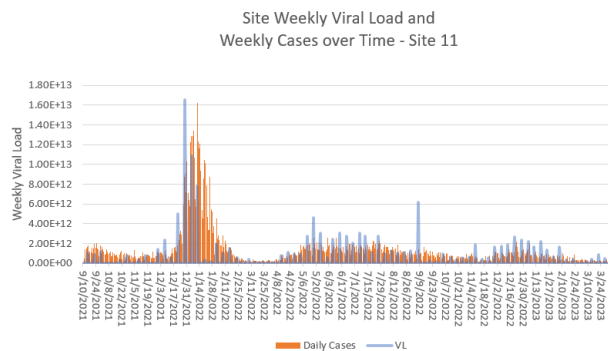
Site 9 – ON



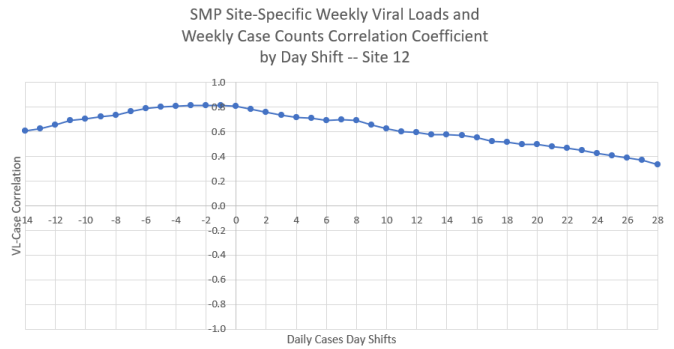
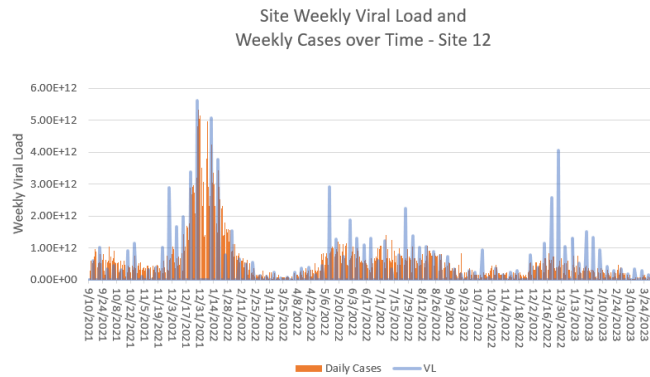
Site 10 – VIP



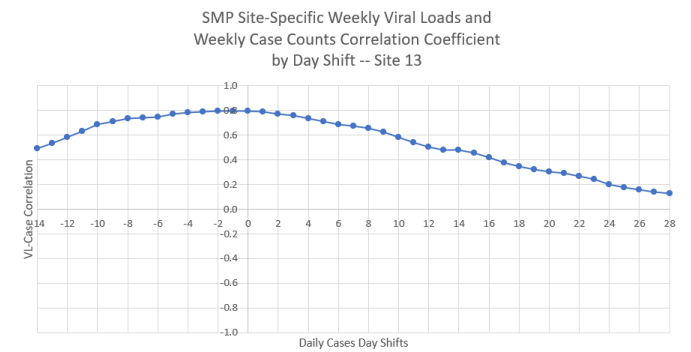
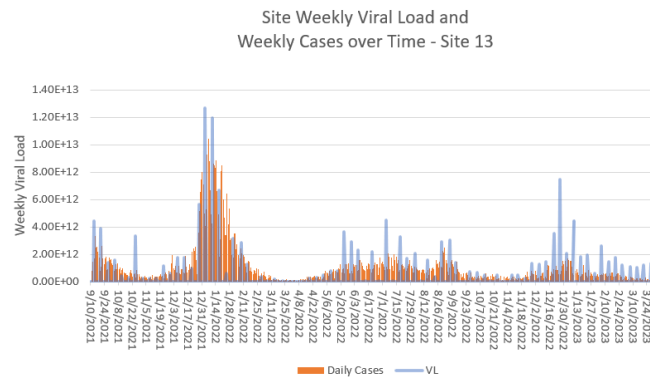
Site 11 – WB



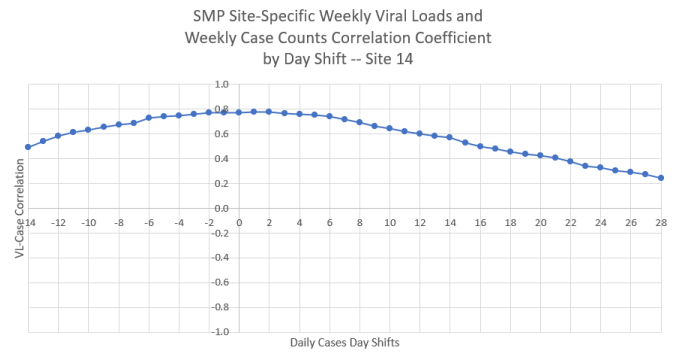
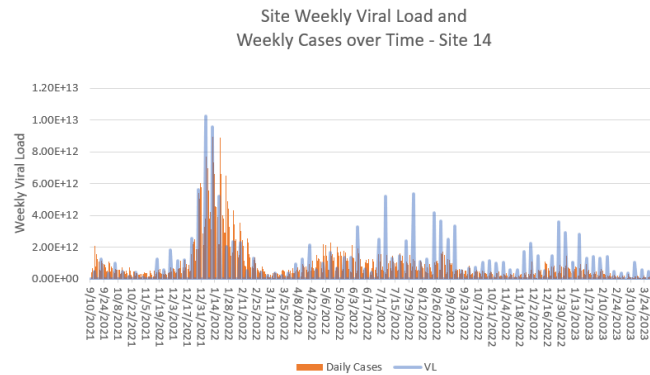
Site 12 – LFR



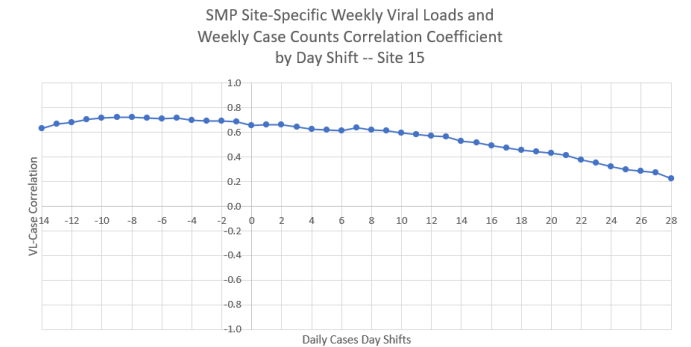
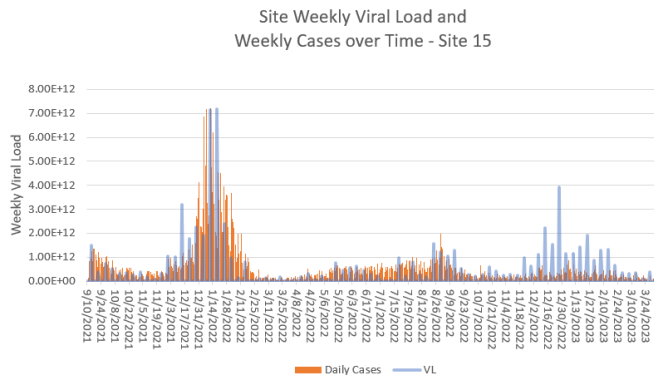
Site 13 – LB



Site 14 – Moores

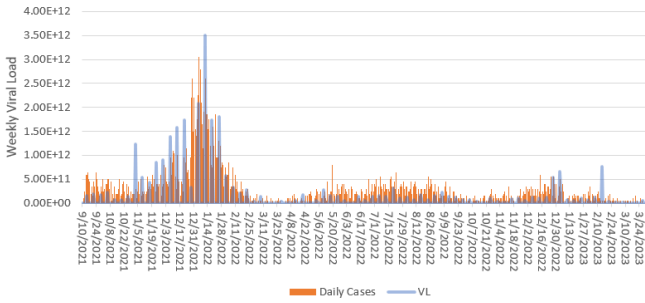


Site 15 – NorthRiver

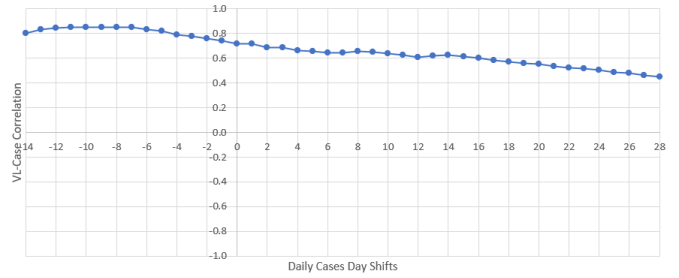


Site 16 – Parkins

Site Weekly Viral Load and Weekly Cases over Time - Site 16

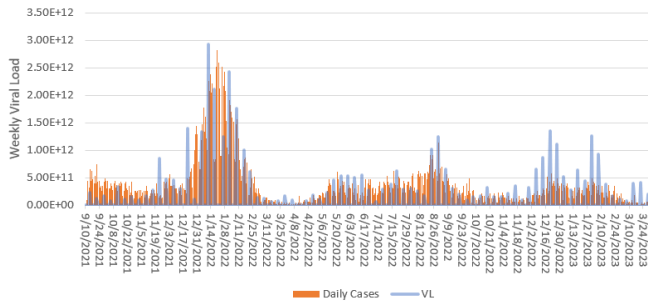


SMP Site-Specific Weekly Viral Loads and Weekly Case Counts Correlation Coefficient by Day Shift -- Site 16

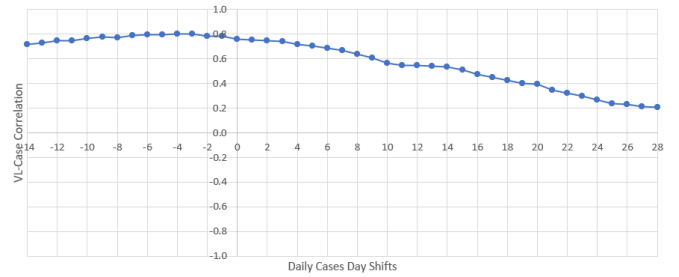


Site 17 – Peppers

Site Weekly Viral Load and Weekly Cases over Time - Site 17

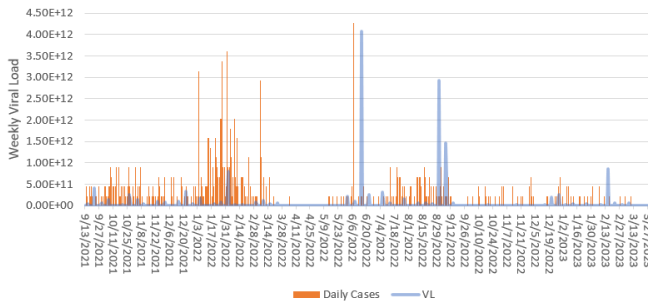


SMP Site-Specific Weekly Viral Loads and Weekly Case Counts Correlation Coefficient by Day Shift -- Site 17

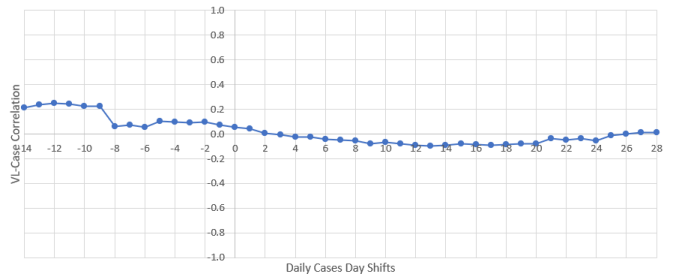


Site 18 – Pound

Site Weekly Viral Load and Weekly Cases over Time - Site 18

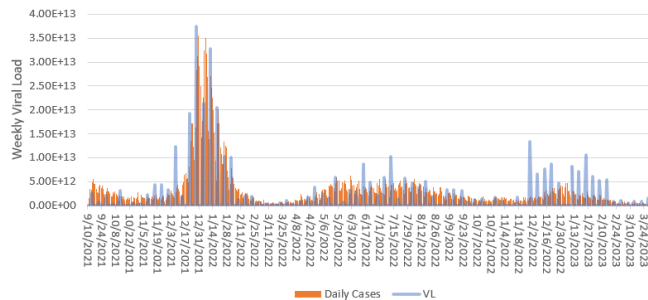


SMP Site-Specific Weekly Viral Loads and Weekly Case Counts Correlation Coefficient by Day Shift -- Site 18

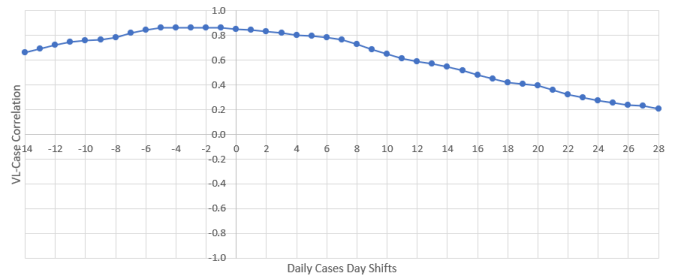


Site 19 – Richmond

Site Weekly Viral Load and Weekly Cases over Time - Site 19

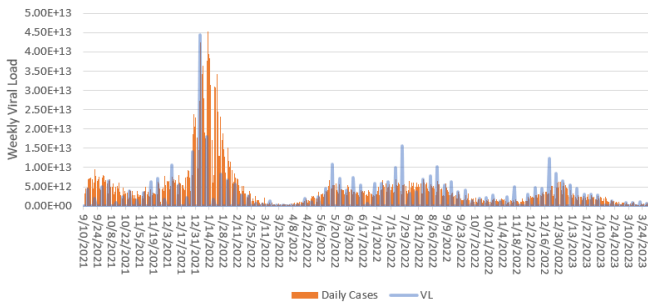


SMP Site-Specific Weekly Viral Loads and Weekly Case Counts Correlation Coefficient by Day Shift -- Site 19

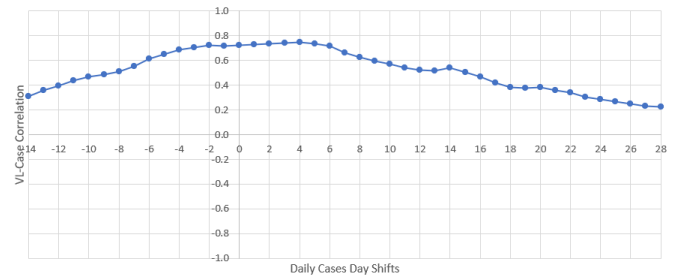


Site 20 – Roanoke

Site Weekly Viral Load and
Weekly Cases over Time - Site 20

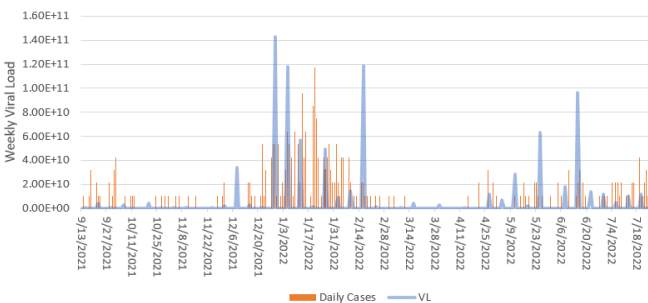


SMP Site-Specific Weekly Viral Loads and
Weekly Case Counts Correlation Coefficient
by Day Shift -- Site 20

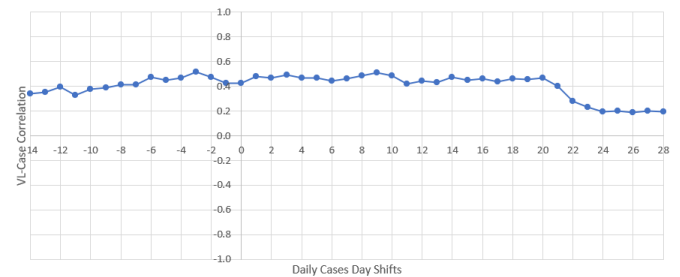


Site 21 – Rutledge*

Site Weekly Viral Load and
Weekly Cases over Time - Site 21

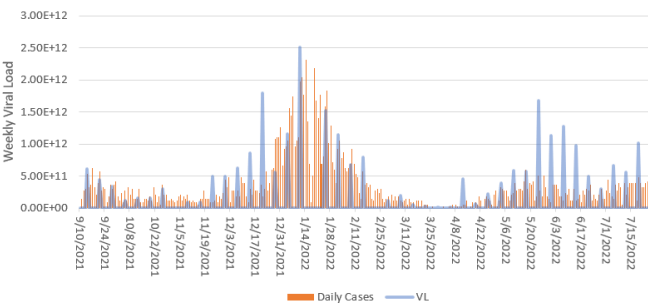


SMP Site-Specific Weekly Viral Loads and
Weekly Case Counts Correlation Coefficient
by Day Shift -- Site 21

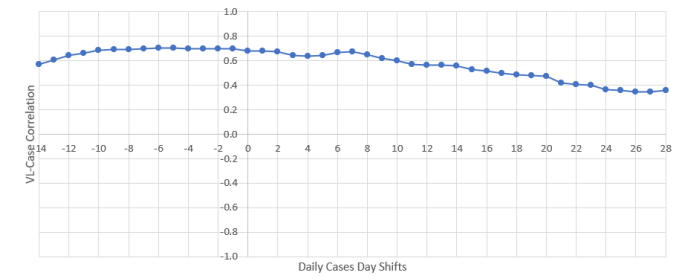


Site 22 – CB*

Site Weekly Viral Load and
Weekly Cases over Time - Site 22

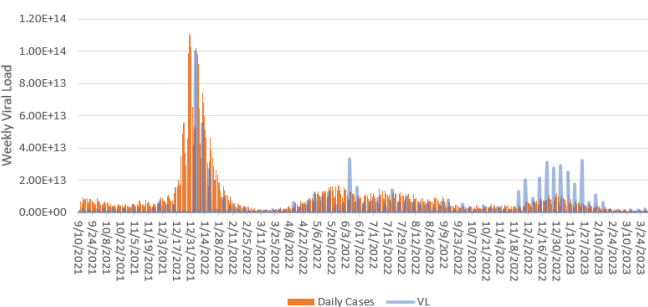


SMP Site-Specific Weekly Viral Loads and
Weekly Case Counts Correlation Coefficient
by Day Shift -- Site 22

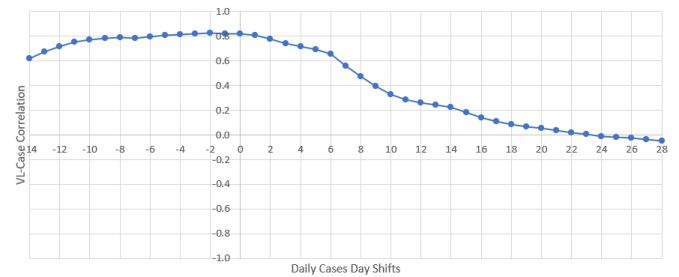


Site 23 – UOSA

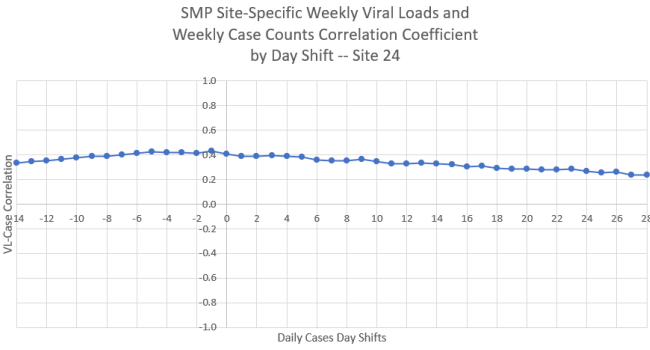
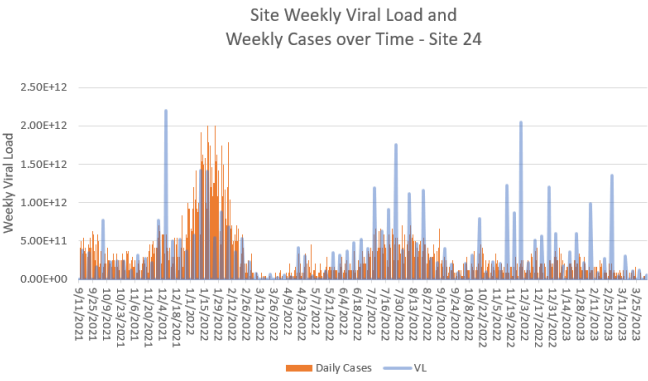
Site Weekly Viral Load and
Weekly Cases over Time - Site 23



SMP Site-Specific Weekly Viral Loads and
Weekly Case Counts Correlation Coefficient
by Day Shift -- Site 23



Site 24 – Wolf

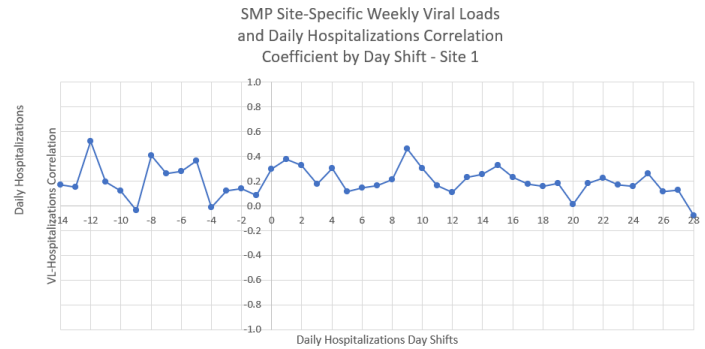
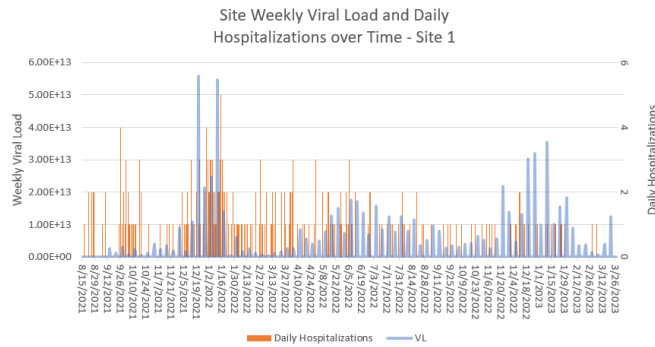


Appendix – 9.5 By Site VL & Daily Hospitalizations

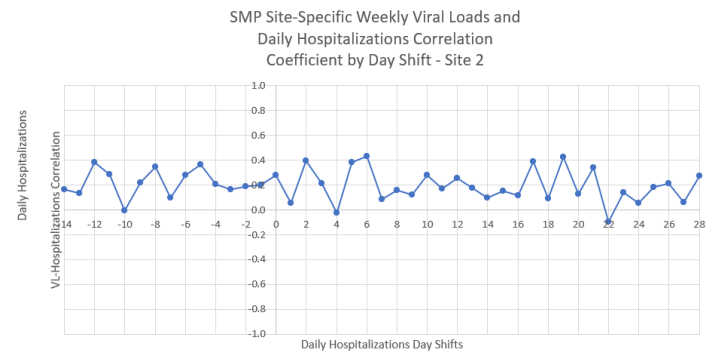
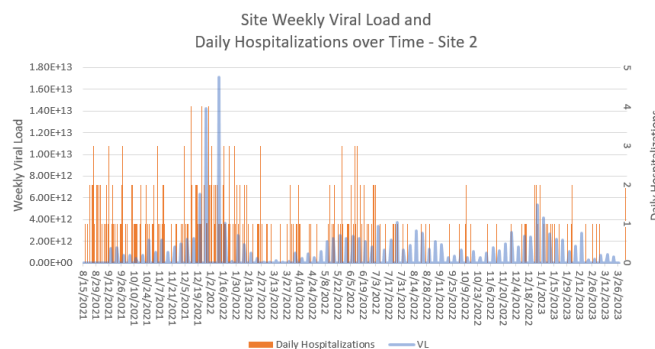
Some sites withdrew from SMP early (*).

Note: secondary axes are intentionally not labeled, to protect case anonymity.

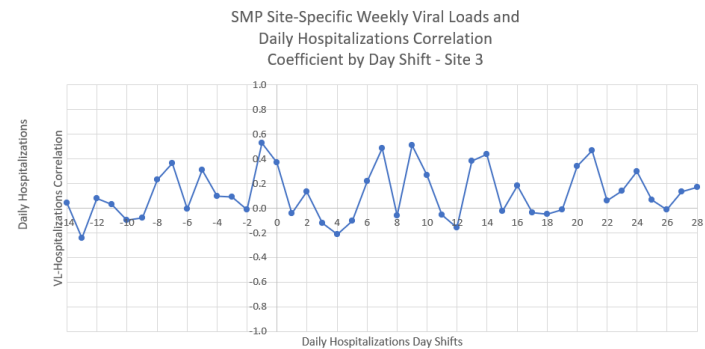
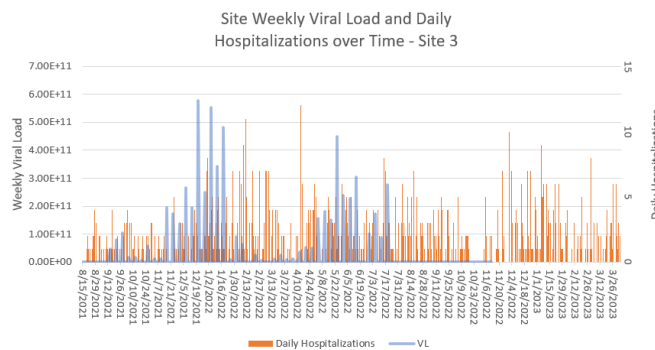
Site 1 – AlexRenew



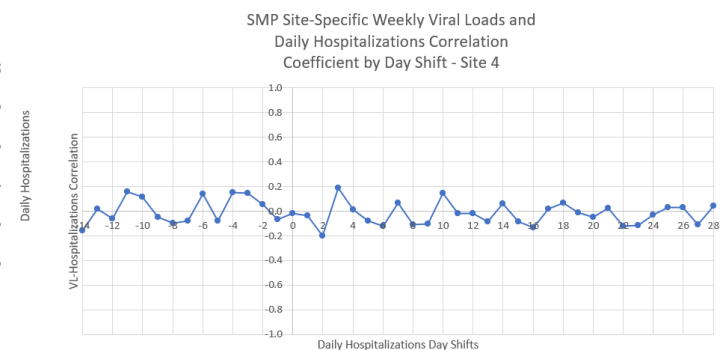
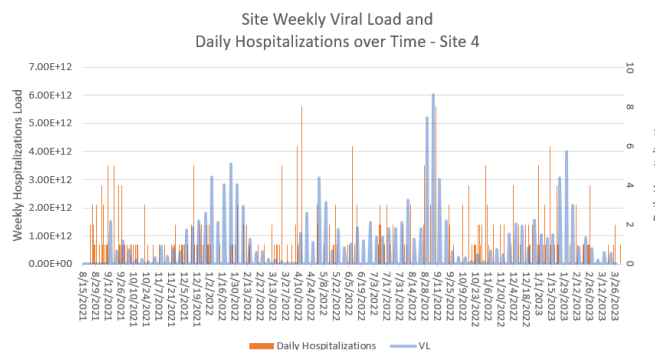
Site 2 – Aquia



Site 3 – Basham



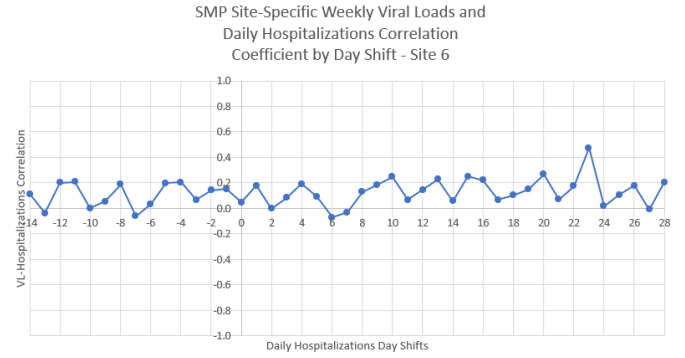
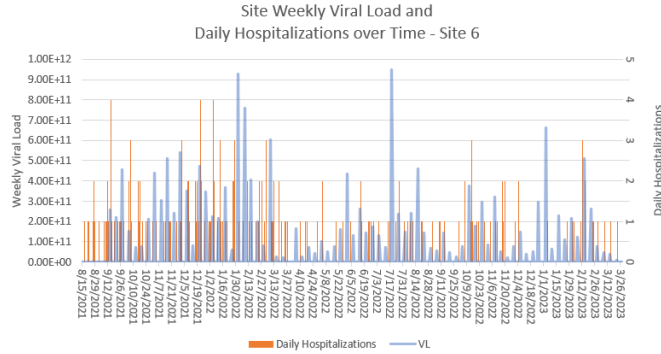
Site 4 – Blacksburg



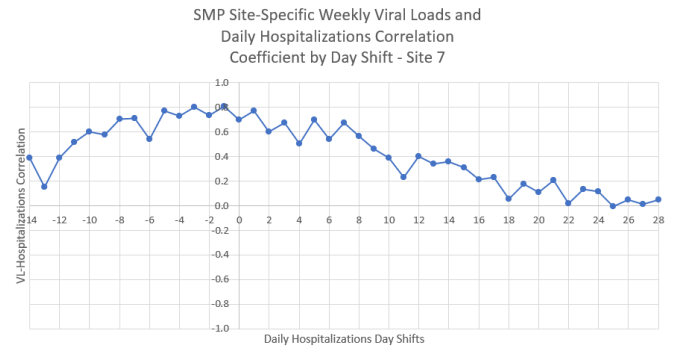
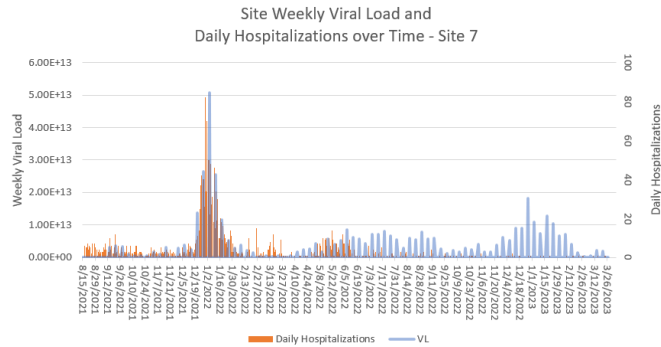
Site 5 – Danville

[sewershed unavailable]

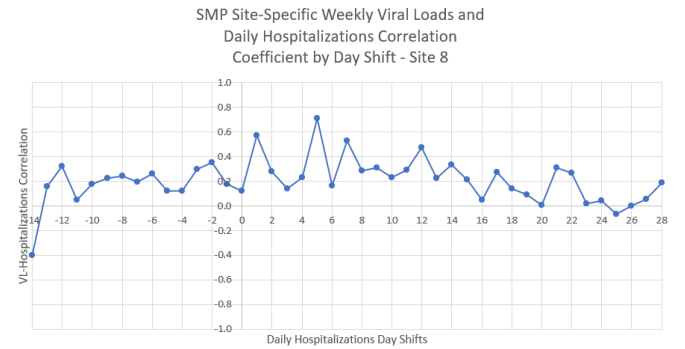
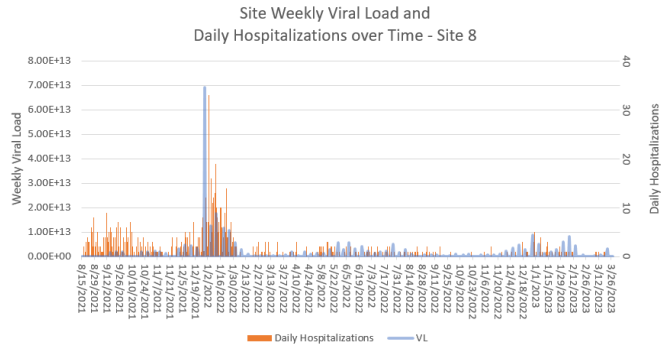
Site 6 – Coeburn



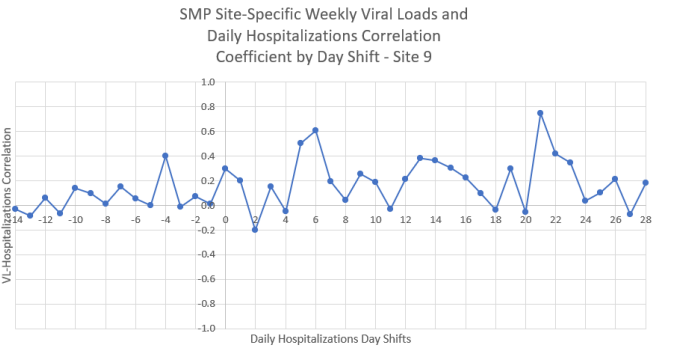
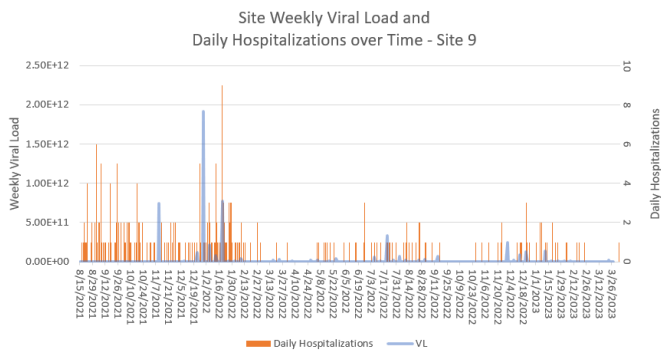
Site 7 – HLM



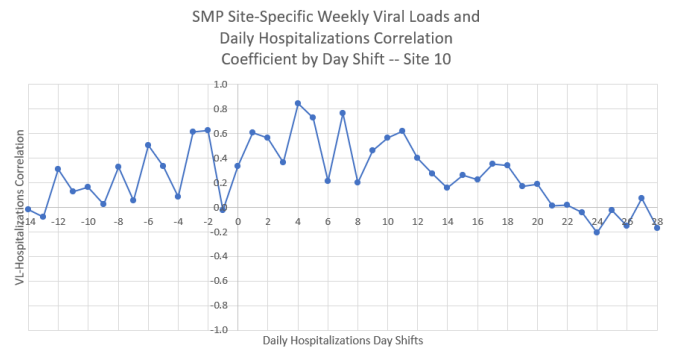
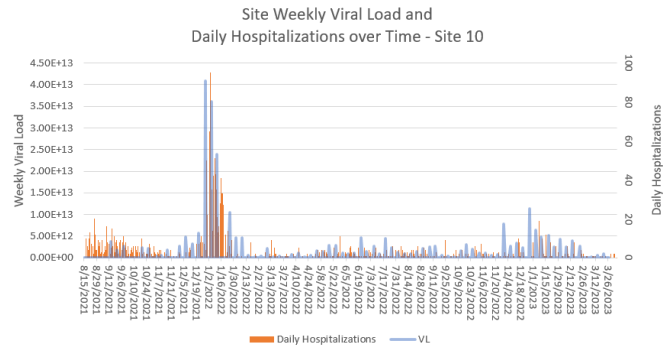
Site 8 – NA



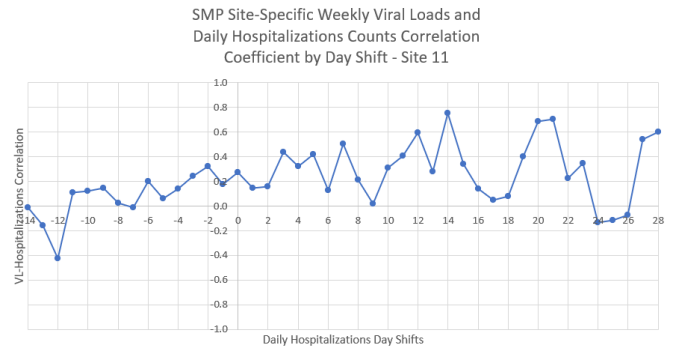
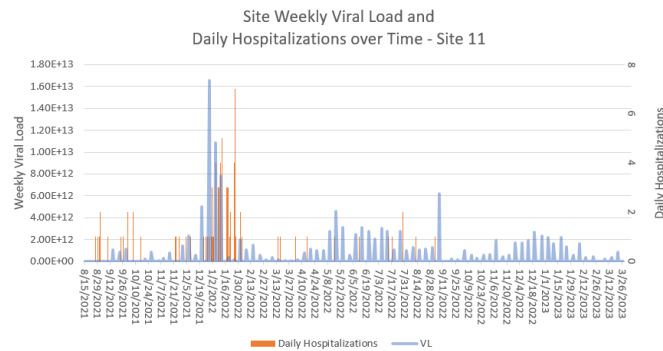
Site 9 – ON



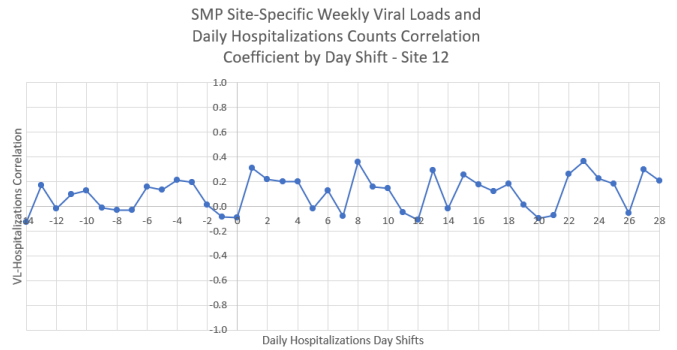
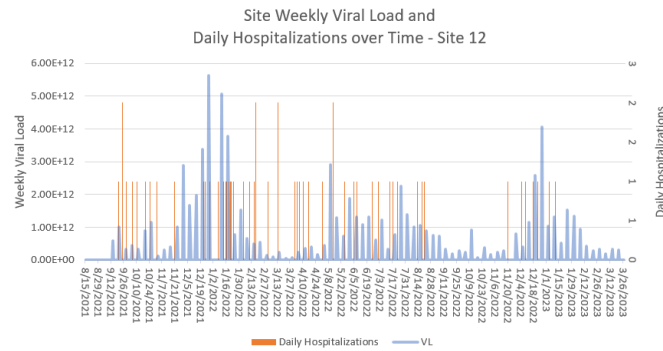
Site 10 – VIP



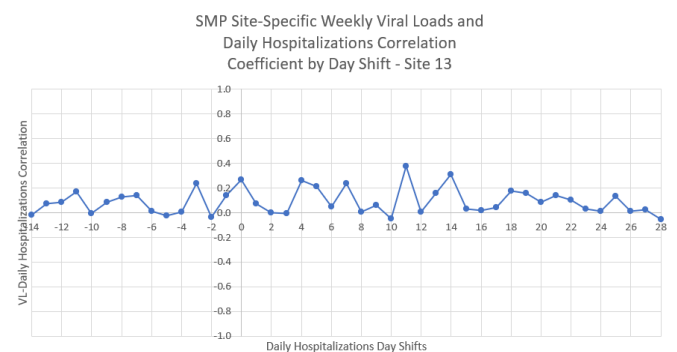
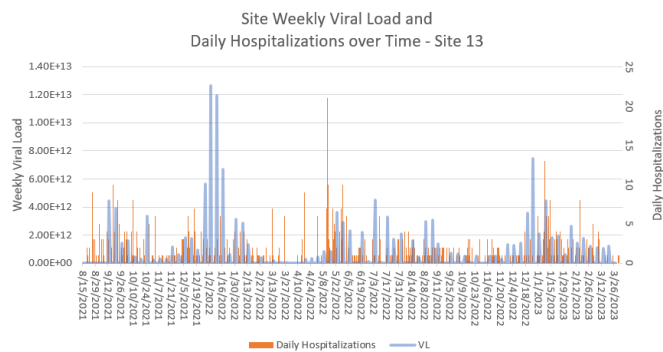
Site 11 – WB



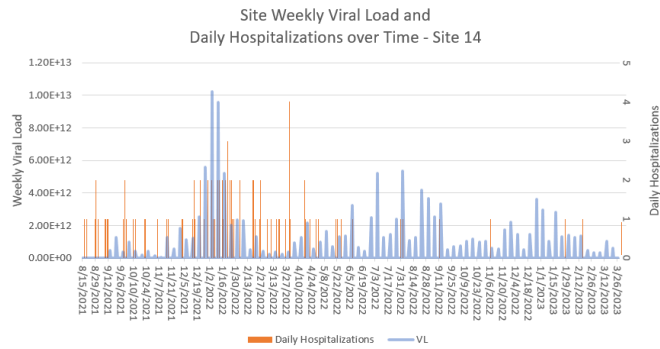
Site 12 – LFR



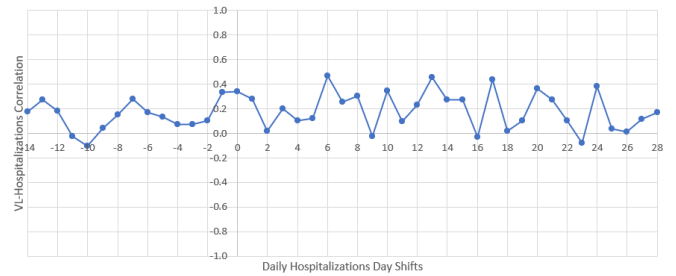
Site 13 – LB



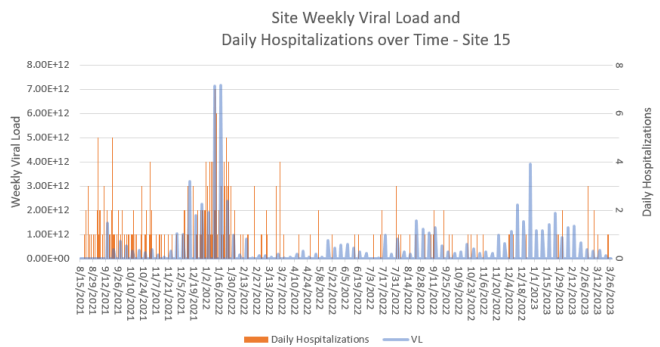
Site 14 – Moores



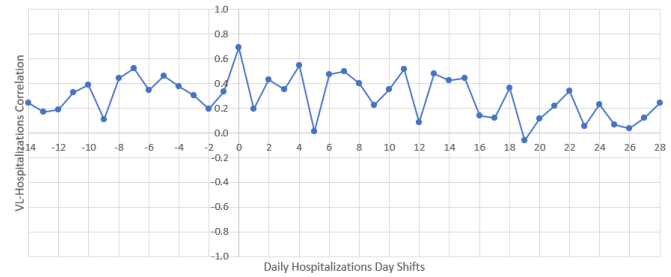
SMP Site-Specific Weekly Viral Loads and Daily Hospitalizations Correlation Coefficient by Day Shift - Site 14



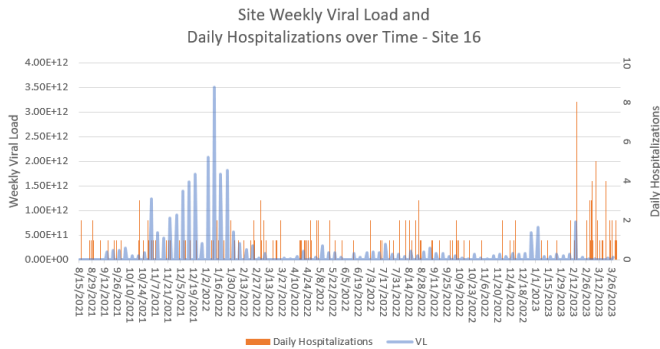
Site 15 – NorthRiver



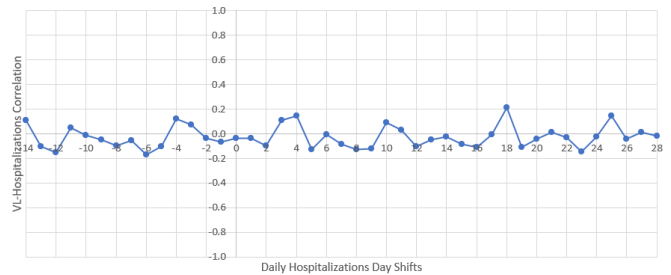
SMP Site-Specific Weekly Viral Loads and Daily Hospitalizations Correlation Coefficient by Day Shift - Site 15



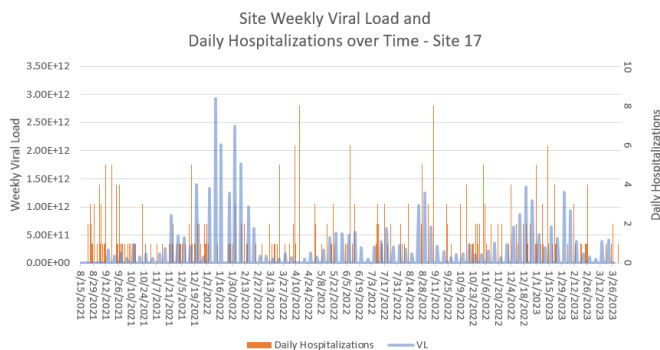
Site 16 – Parkins



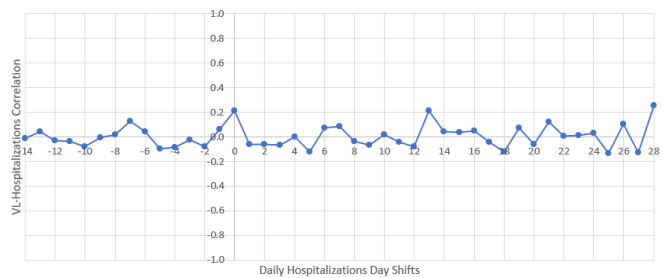
SMP Site-Specific Weekly Viral Loads and Daily Hospitalizations Correlation Coefficient by Day Shift - Site 16



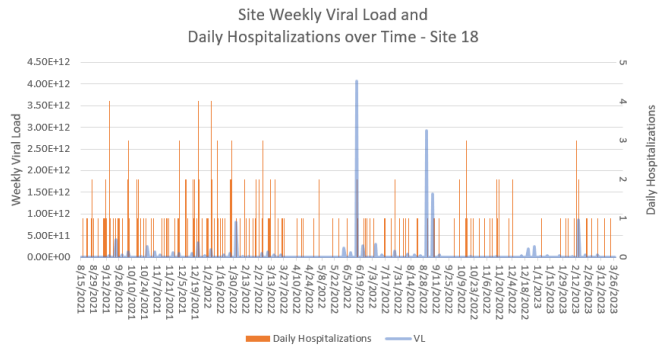
Site 17 – Peppers



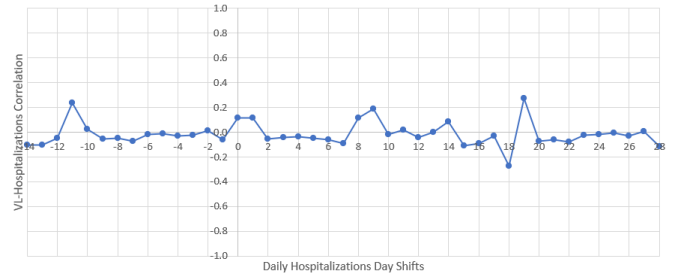
SMP Site-Specific Weekly Viral Loads and Daily Hospitalizations Correlation Coefficient by Day Shift - Site 17



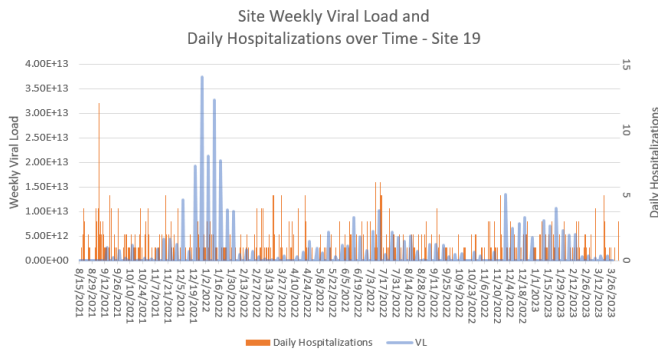
Site 18 – Pound



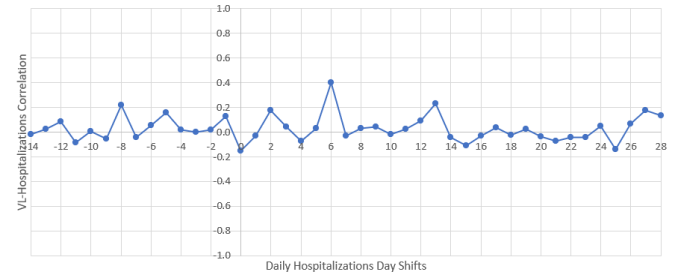
SMP Site-Specific Weekly Viral Loads and Daily Hospitalizations Correlation Coefficient by Day Shift - Site 18



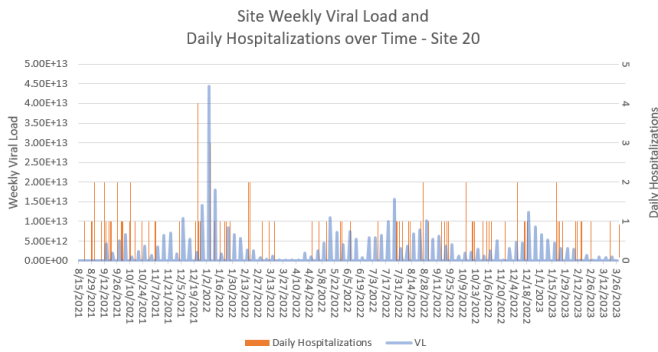
Site 19 – Richmond



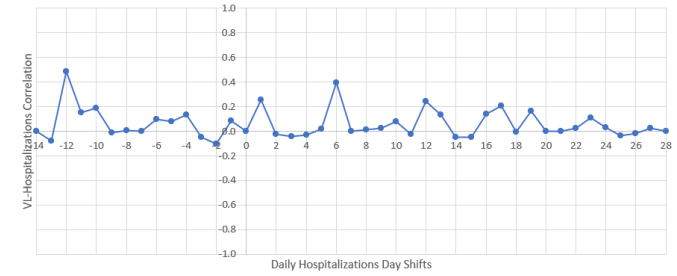
SMP Site-Specific Weekly Viral Loads and Daily Hospitalizations Correlation Coefficient by Day Shift - Site 19



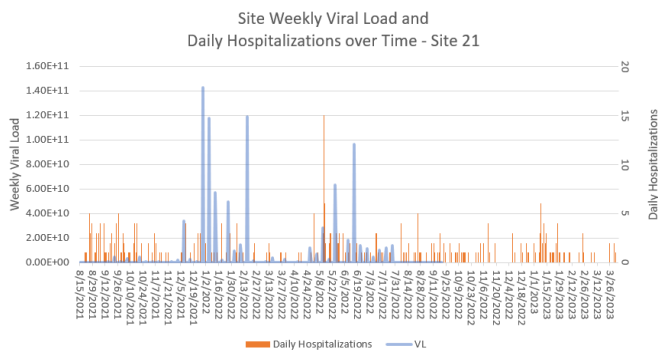
Site 20 – Roanoke



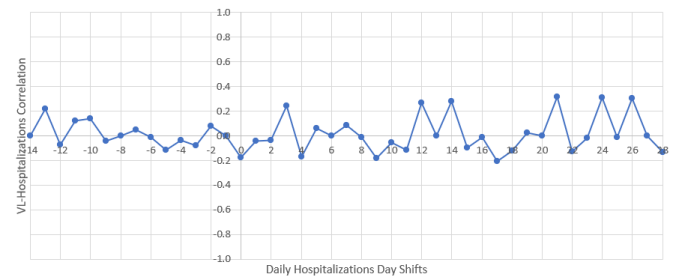
SMP Site-Specific Weekly Viral Loads and Daily Hospitalizations Correlation Coefficient by Day Shift - Site 20



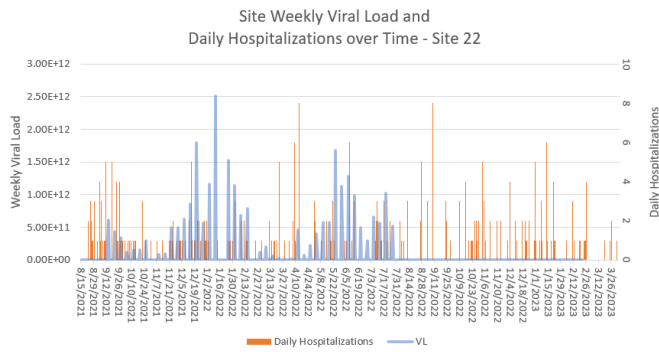
Site 21 – Rutledge



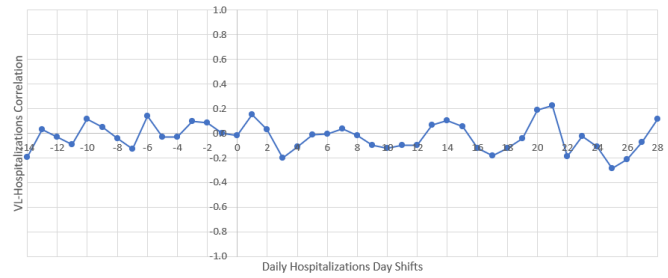
SMP Site-Specific Weekly Viral Loads and Daily Hospitalizations Correlation Coefficient by Day Shift - Site 21



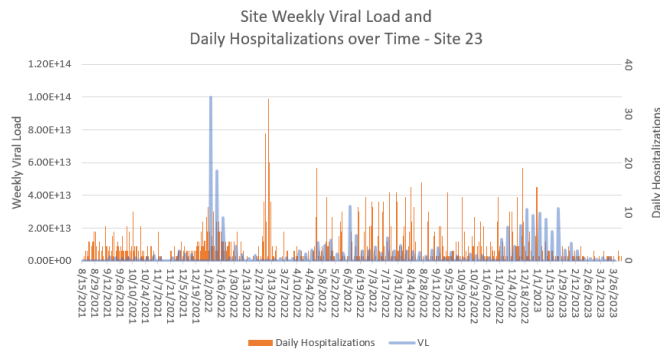
Site 22 – CB



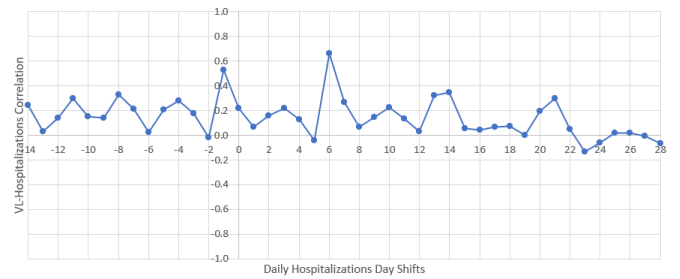
SMP Site-Specific Weekly Viral Loads and Daily Hospitalizations Correlation Coefficient by Day Shift - Site 22



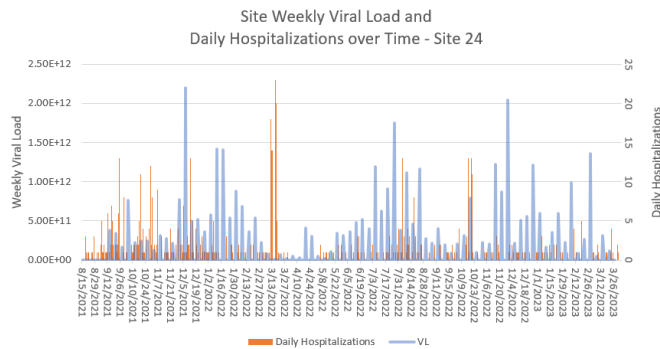
Site 23 – UOSA



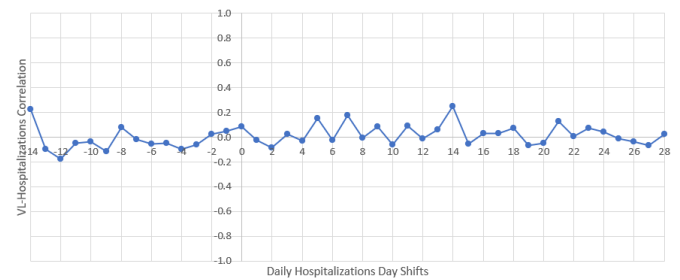
SMP Site-Specific Weekly Viral Loads and Daily Hospitalizations Correlation Coefficient by Day Shift - Site 23



Site 24 – Wolf



SMP Site-Specific Weekly Viral Loads and Daily Hospitalizations Correlation Coefficient by Day Shift - Site 24

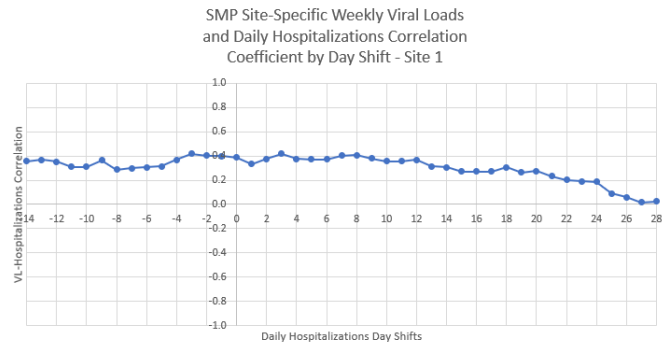
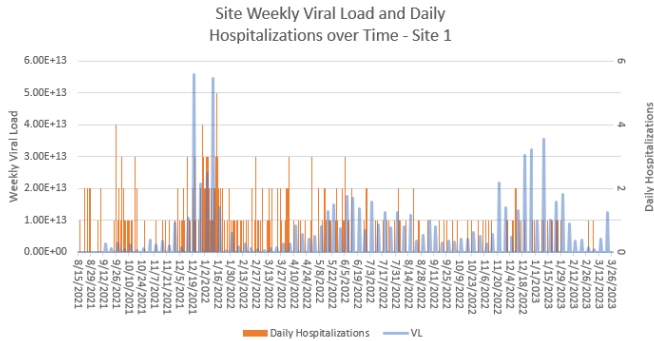


Appendix – 9.6 By Site VL & Weekly Hospitalizations

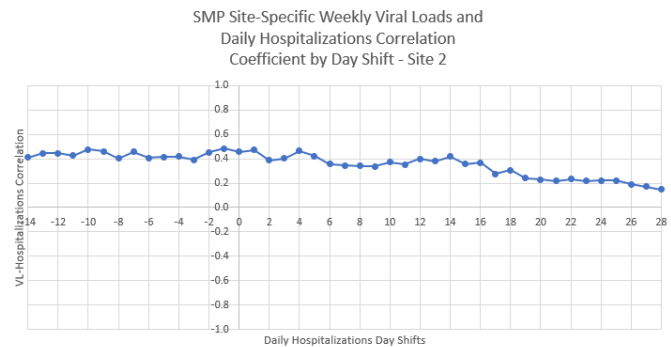
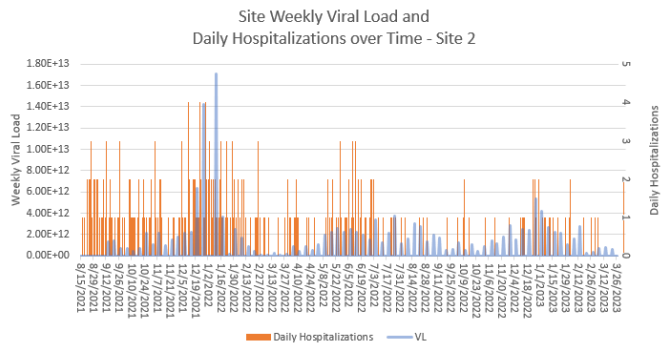
Some sites withdrew from SMP early (*).

Note: secondary axes are intentionally not labeled, to protect case anonymity.

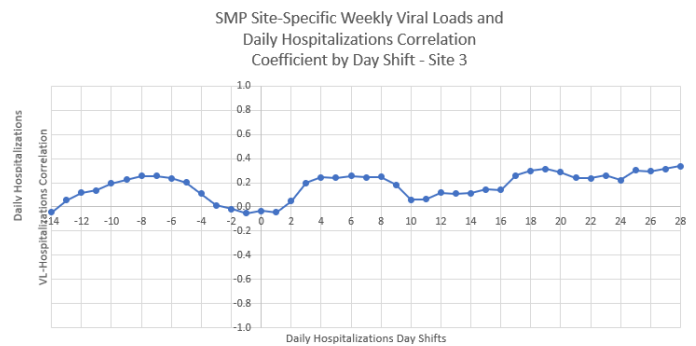
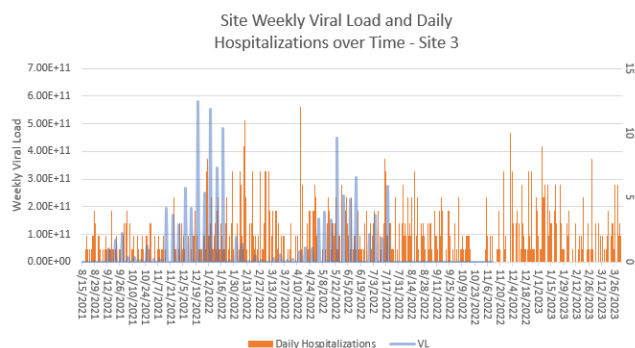
Site 1 – AlexRenew



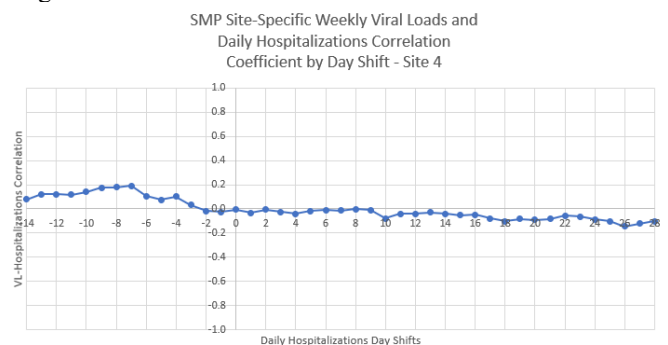
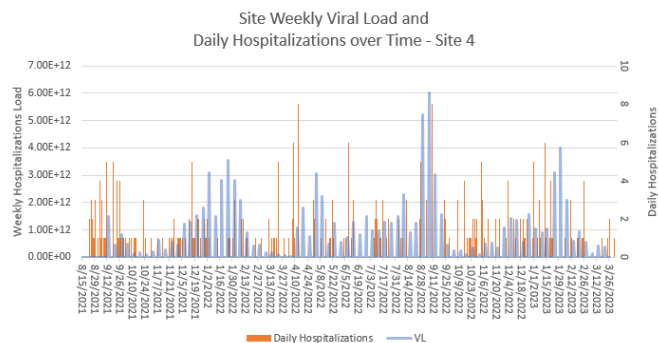
Site 2 – Aquia



Site 3 – Basham



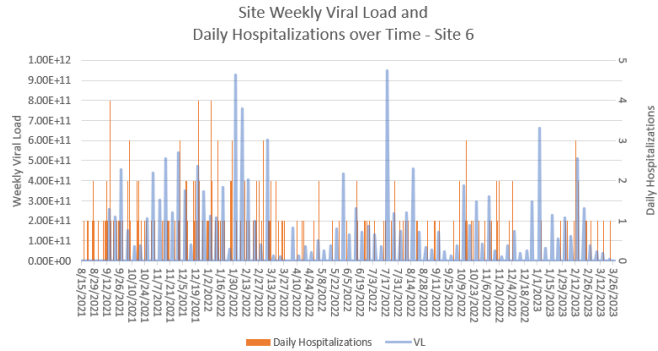
Site 4 – Blacksburg



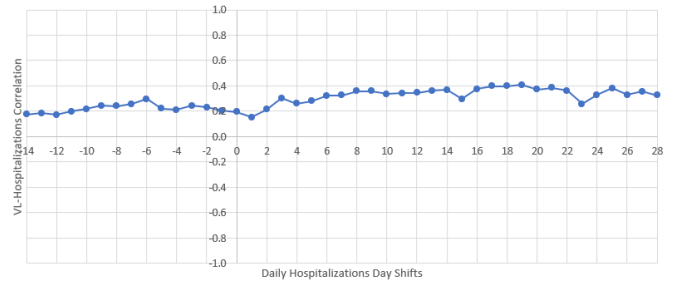
Site 5 – Danville

[sewershed unavailable]

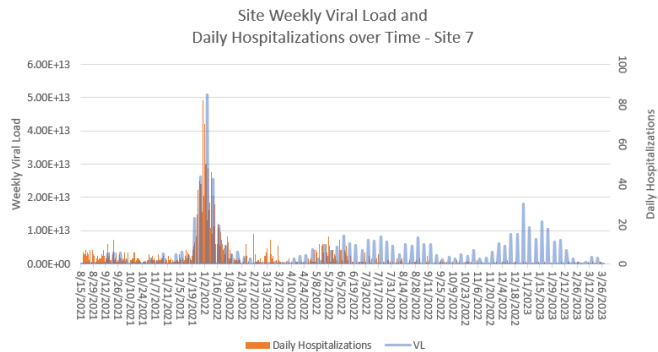
Site 6 – Coeburn



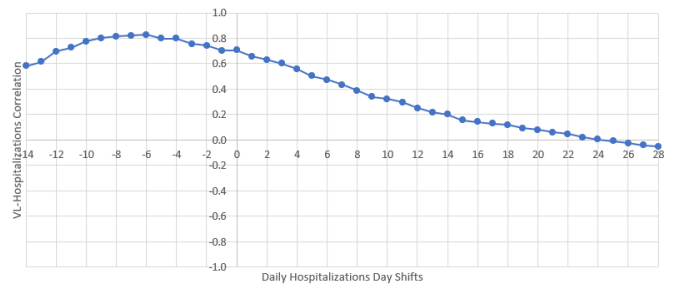
SMP Site-Specific Weekly Viral Loads and Daily Hospitalizations Correlation Coefficient by Day Shift - Site 6



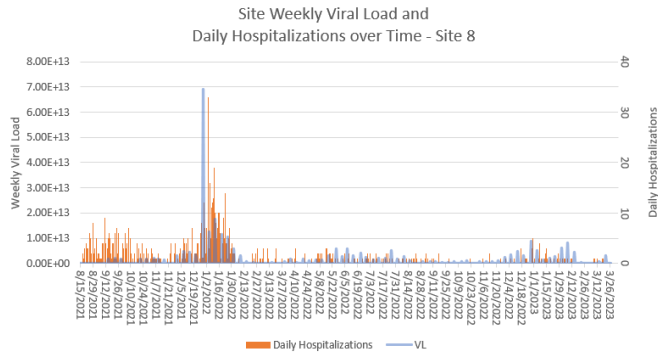
Site 7 – HLM



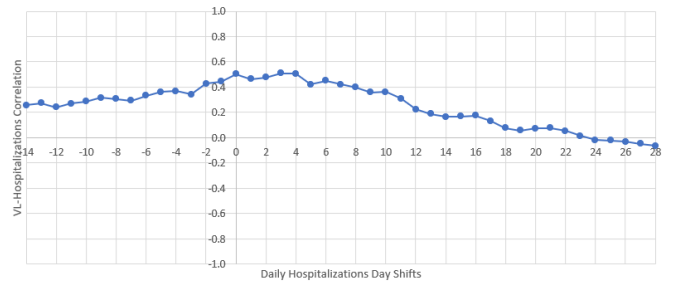
SMP Site-Specific Weekly Viral Loads and Daily Hospitalizations Correlation Coefficient by Day Shift - Site 7



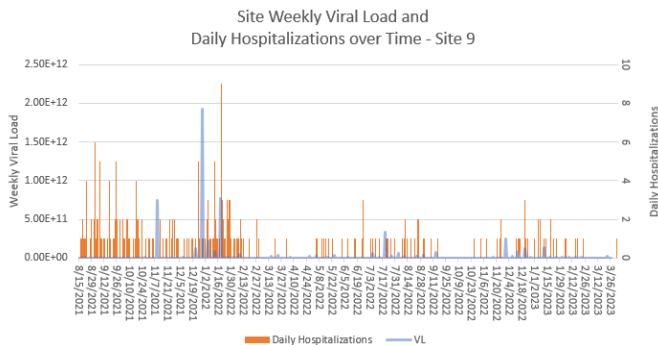
Site 8 – NA



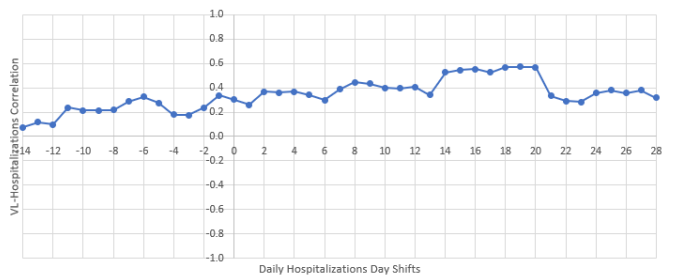
SMP Site-Specific Weekly Viral Loads and Daily Hospitalizations Correlation Coefficient by Day Shift - Site 8



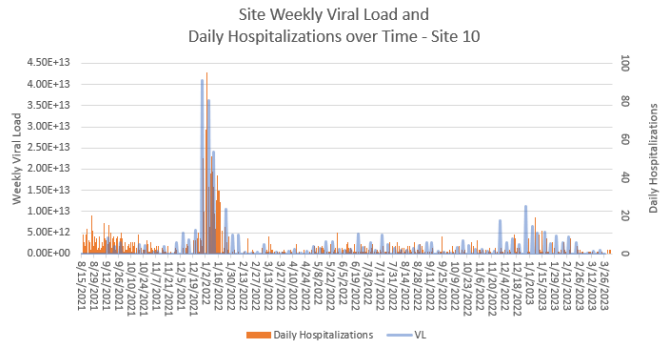
Site 9 – ON



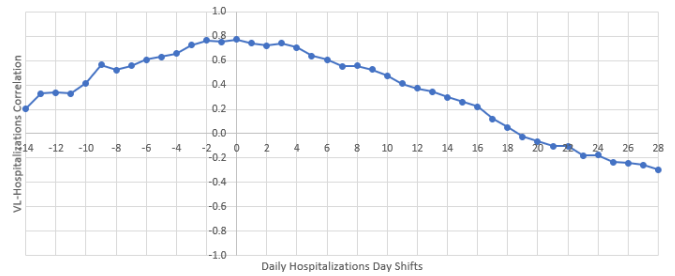
SMP Site-Specific Weekly Viral Loads and Daily Hospitalizations Correlation Coefficient by Day Shift - Site 9



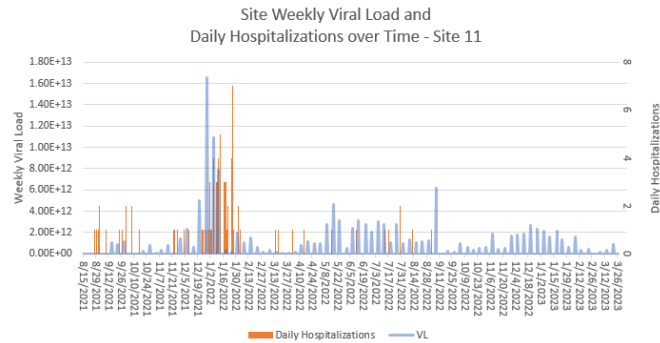
Site 10 – VIP



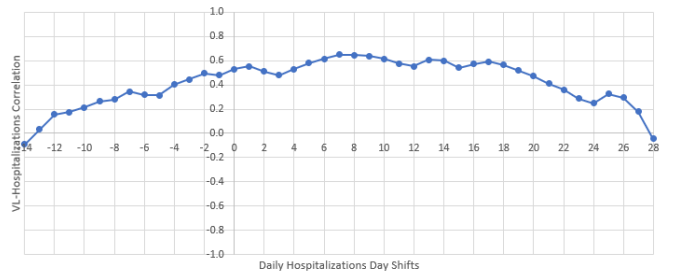
SMP Site-Specific Weekly Viral Loads and Daily Hospitalizations Correlation Coefficient by Day Shift – Site 10



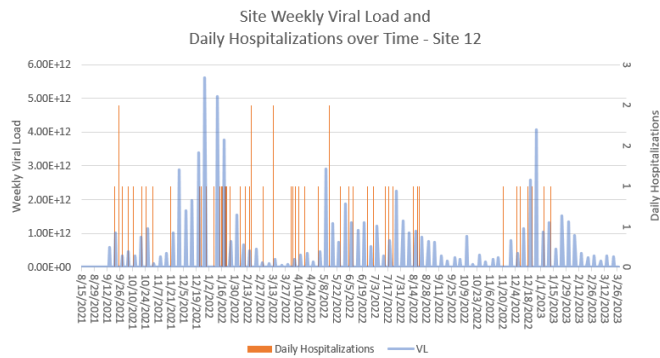
Site 11 – WB



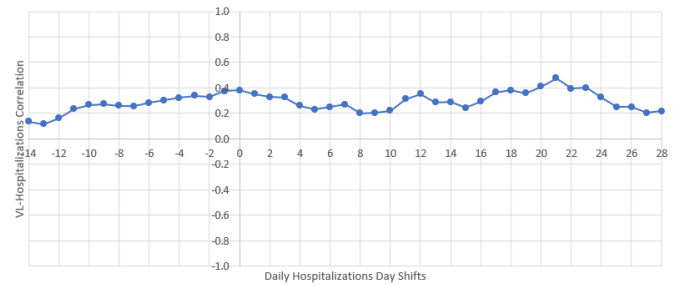
SMP Site-Specific Weekly Viral Loads and Daily Hospitalizations Counts Correlation Coefficient by Day Shift - Site 11



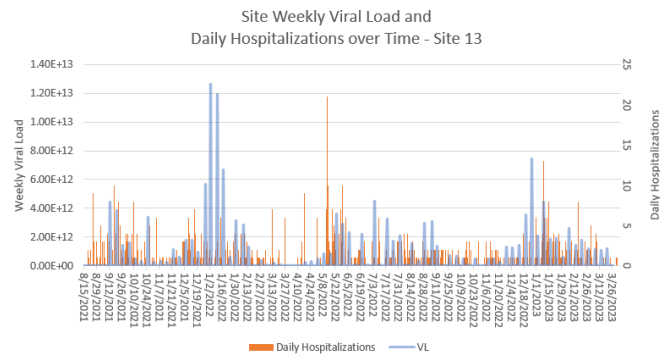
Site 12 – LFR



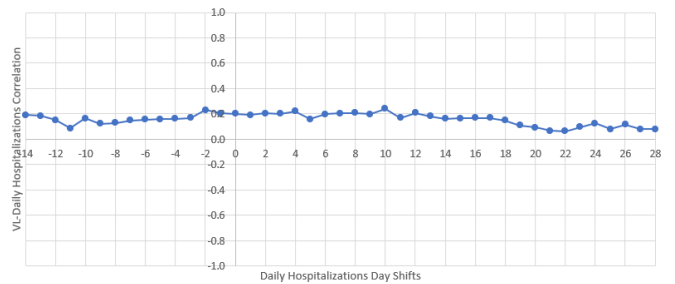
SMP Site-Specific Weekly Viral Loads and Daily Hospitalizations Counts Correlation Coefficient by Day Shift - Site 12



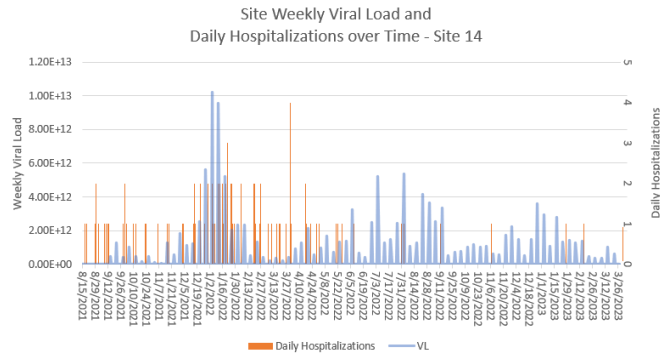
Site 13 – LB



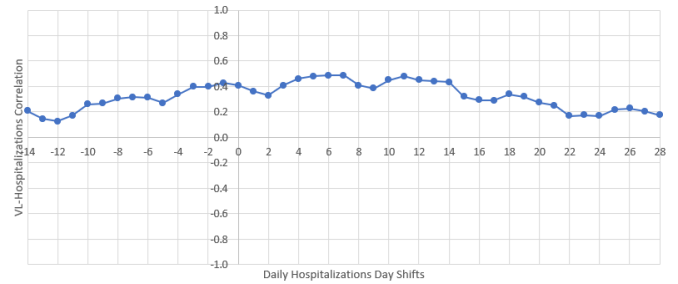
SMP Site-Specific Weekly Viral Loads and Daily Hospitalizations Correlation Coefficient by Day Shift - Site 13



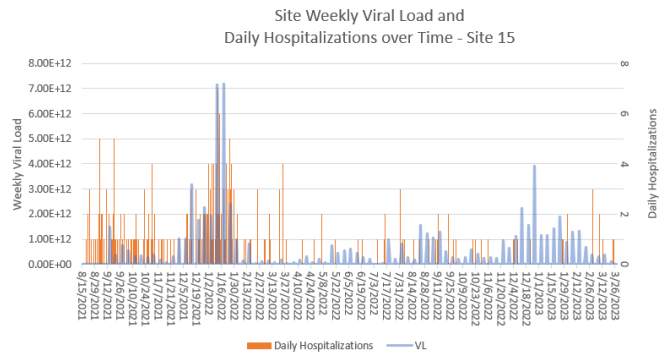
Site 14 – Moores



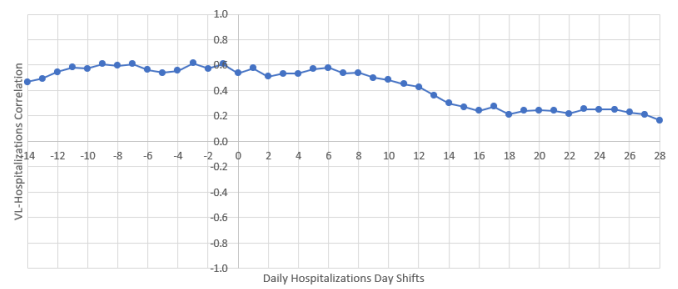
SMP Site-Specific Weekly Viral Loads and Daily Hospitalizations Correlation Coefficient by Day Shift - Site 14



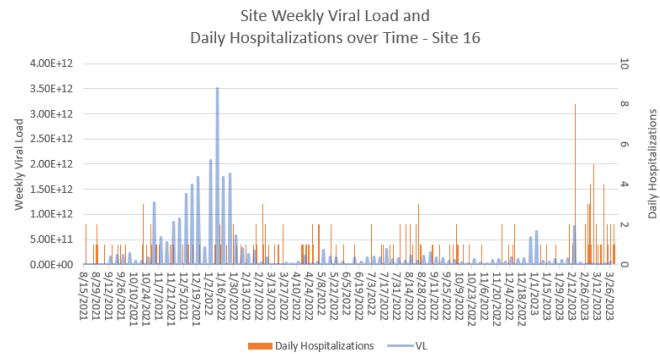
Site 15 – NorthRiver



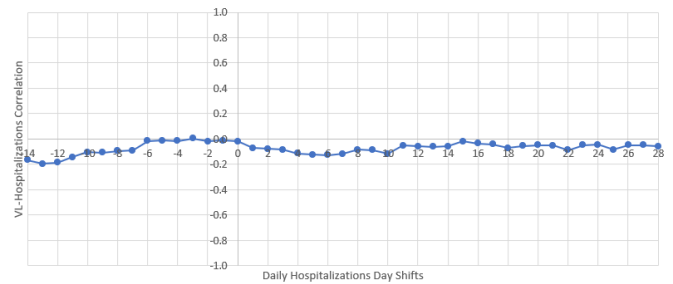
SMP Site-Specific Weekly Viral Loads and Daily Hospitalizations Correlation Coefficient by Day Shift - Site 15



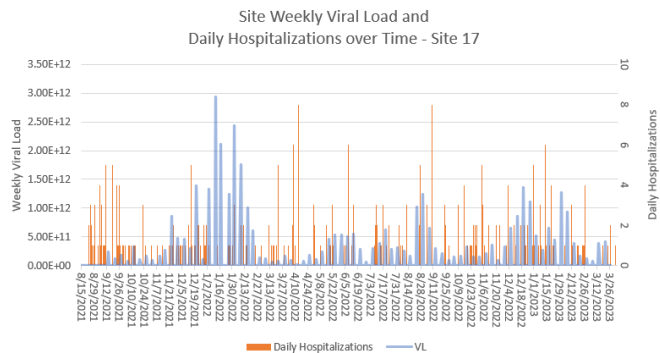
Site 16 – Parkins



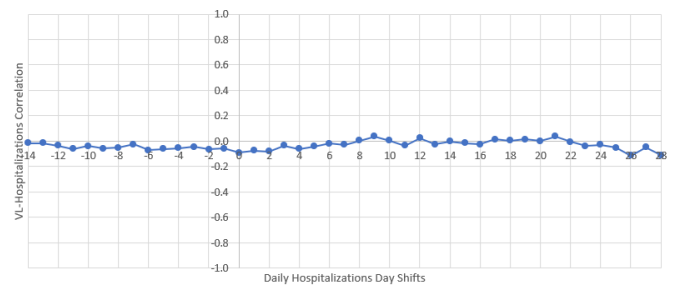
SMP Site-Specific Weekly Viral Loads and Daily Hospitalizations Correlation Coefficient by Day Shift - Site 16



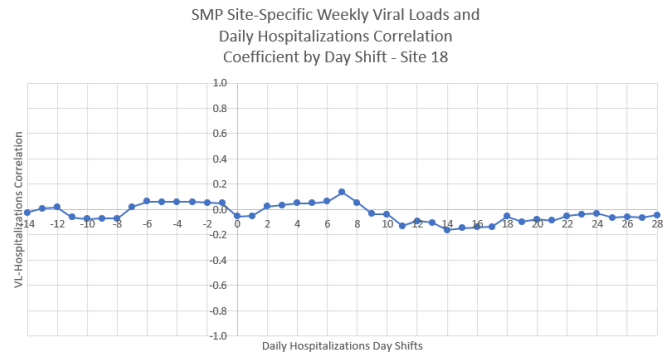
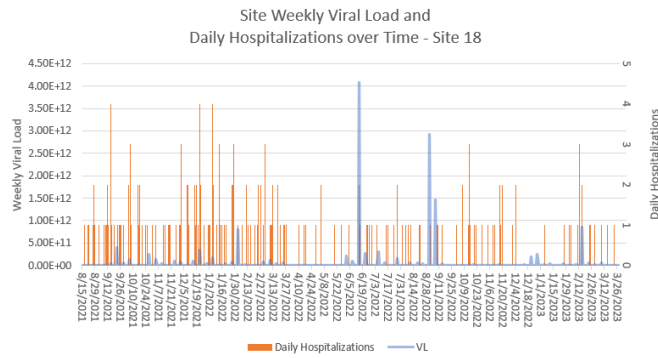
Site 17 – Peppers



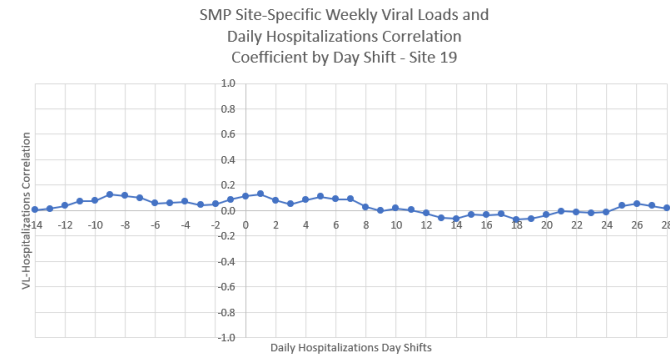
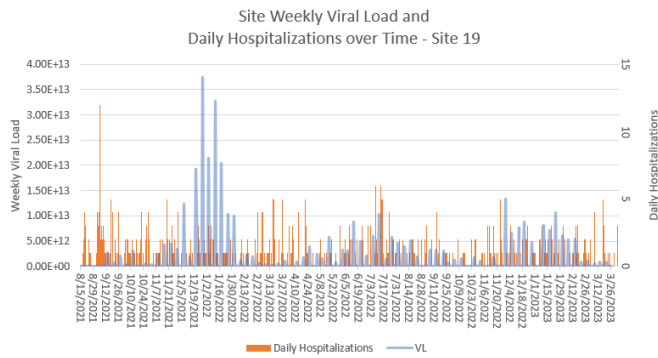
SMP Site-Specific Weekly Viral Loads and Daily Hospitalizations Correlation Coefficient by Day Shift - Site 17



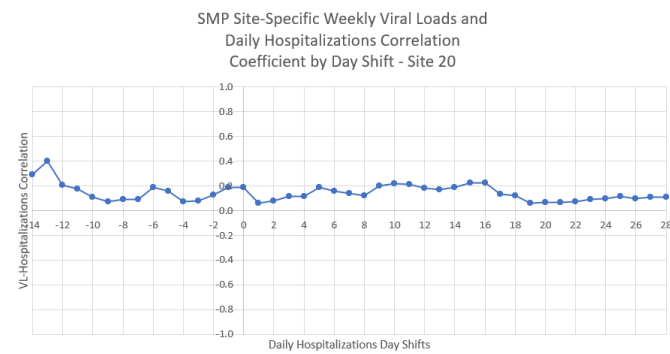
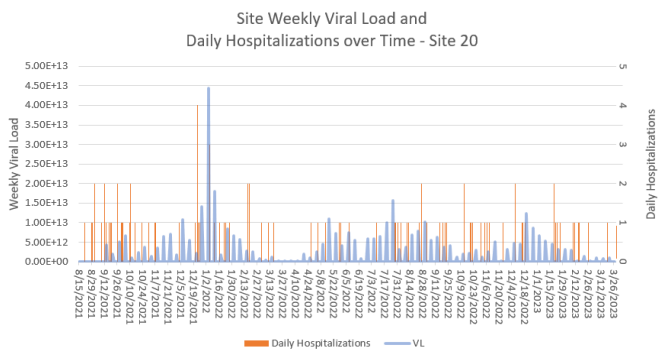
Site 18 – Pound



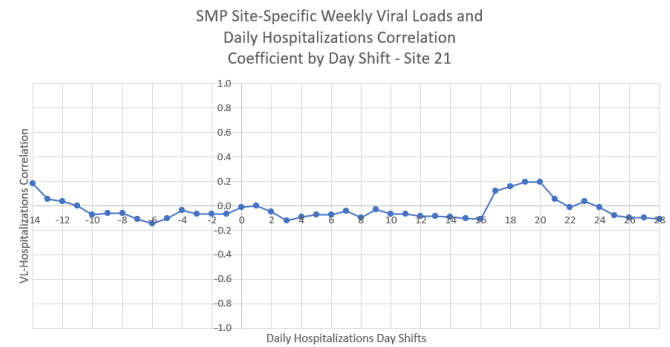
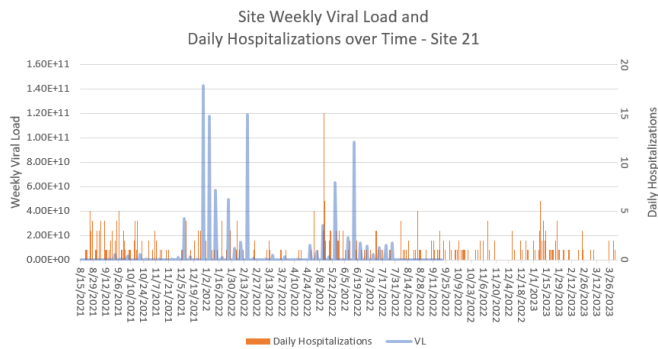
Site 19 – Richmond



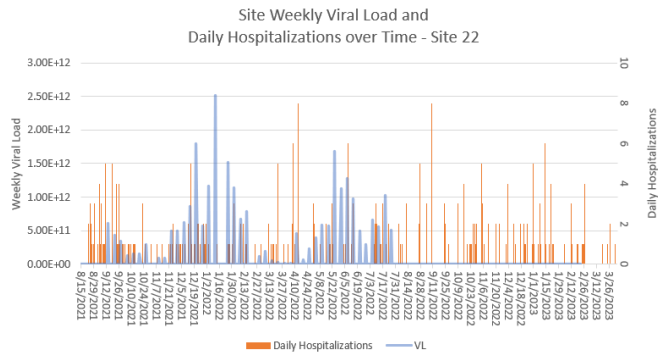
Site 20 – Roanoke



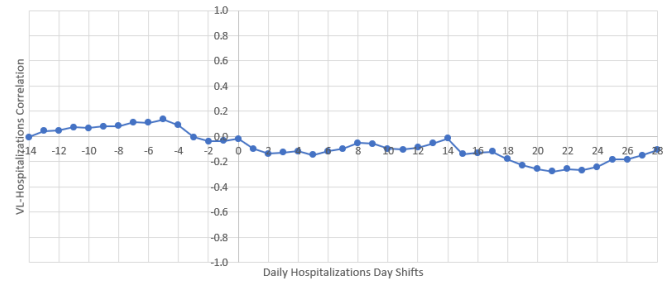
Site 21 – Rutledge



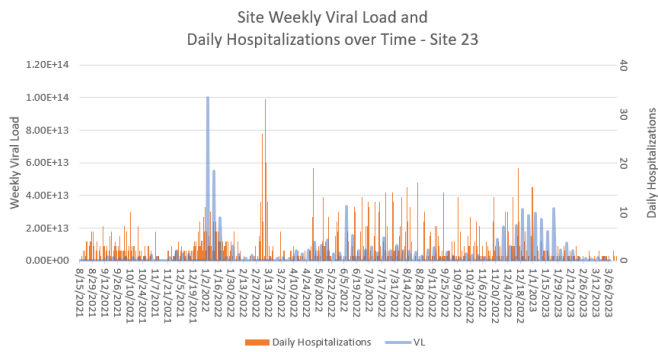
Site 22 – CB



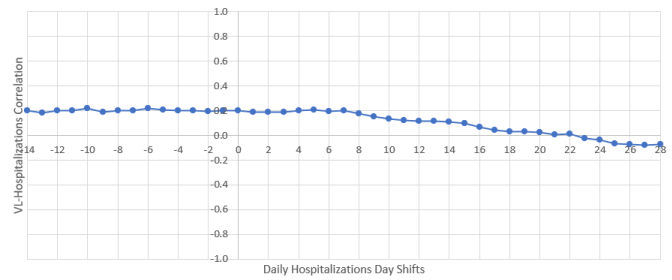
SMP Site-Specific Weekly Viral Loads and Daily Hospitalizations Correlation Coefficient by Day Shift - Site 22



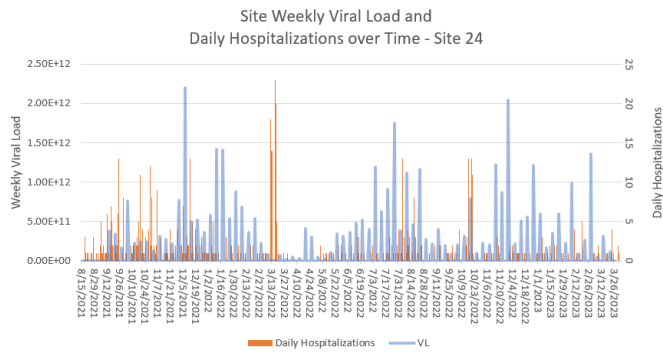
Site 23 – UOSA



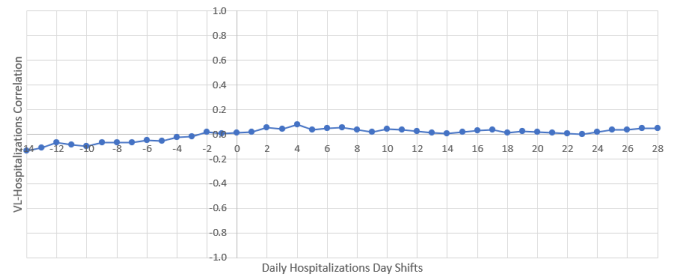
SMP Site-Specific Weekly Viral Loads and Daily Hospitalizations Correlation Coefficient by Day Shift - Site 23



Site 24 – Wolf



SMP Site-Specific Weekly Viral Loads and Daily Hospitalizations Correlation Coefficient by Day Shift - Site 24



[End of Dissertation]