

Spatiotemporal Cluster Detection for COVID-19 Outbreak Surveillance: A Descriptive Analysis Study

Rachel Martonik, Caitlin Oleson, Ellyn Marder

Submitted to: JMIR Public Health and Surveillance
on: June 14, 2023

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript..... 5

Supplementary Files..... 21

 Figures 22

 Figure 23

 Figure 1..... 24

Spatiotemporal Cluster Detection for COVID-19 Outbreak Surveillance: A Descriptive Analysis Study

Rachel Martonik¹ MSc; Caitlin Oleson² MSc; Ellyn Marder² MPH, PhD

¹Deloitte Arlington US

²Washington State Department of Health Olympia US

Corresponding Author:

Rachel Martonik MSc

Deloitte

1919 North Lynn Street

Arlington

US

Abstract

Background: During the peak of the winter 2020-21 surge, the number of weekly reported COVID-19 outbreaks in Washington State was 231 and the majority of these outbreaks were in high-priority settings. Local health jurisdictions (LHJs), which were primarily responsible for case and outbreak investigations, were overly burdened. Systematic cluster detection using real-time surveillance data could reduce this burden.

Objective: To improve outbreak detection, the Washington State Department of Health initiated a systematic statewide cluster detection model to identify timely and actionable COVID-19 clusters for investigation and resource prioritization. This report details the implementation of the model using SaTScan, along with an assessment of the tool's effectiveness.

Methods: Six LHJs participated in a pilot before statewide implementation in August 2021. Clusters during July 17–December 17, 2021 were analyzed by LHJ population size and incidence. Clusters were matched to reported outbreaks and compared by setting

Results: A weekly, LHJ-specific retrospective space-time permutation model identified 2874 new clusters. The median cluster size was 15 cases and the median number of clusters was 4. Nearly 60% of clusters were timely (ending within one week before the analysis). There were 2874 reported outbreaks during this same time period; 363 (12.8%) matched to ?1 cluster. The most frequent settings among reported and matched outbreaks were schools and youth programs (28.7%, 29.8%), workplaces (21.5%, 15.4%), and long-term care facilities (18.8%, 27.3%). Settings with the highest percentage matching were community settings (22.2%) and congregate housing (20.8%). Approximately one-third (32.8%) of matched outbreaks had all cases linked after the cluster was identified.

Conclusions: Our goal was to routinely and systematically identify timely and actionable COVID-19 clusters throughout the state. Regardless of population or incidence, the model identified reasonably sized, timely clusters statewide, successfully meeting the goals. Among some high priority settings subject to public health interventions throughout the pandemic, such as schools and community settings, the model identified clusters which were matched to reported outbreaks. In workplaces, another high priority setting, results suggest the SaTScan model might be able to identify outbreaks sooner than existing outbreak detection methods.

(JMIR Preprints 14/06/2023:49871)

DOI: <https://doi.org/10.2196/preprints.49871>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to the public.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <http://www.jmir.org/>, I will be able to make my full manuscript PDF available to the public.



Original Manuscript

Spatiotemporal Cluster Detection for COVID-19 Outbreak Surveillance: A Descriptive Analysis Study

Rachel Martonik, MS¹, Caitlin Oleson, MSc², Ellyn P. Marder, DrPH, MPH²

Author affiliation and titles: ¹Deloitte, ²Washington State Department of Health

Rachel Martonik, Specialist Master

Caitlin Oleson, Epidemiologist II

Ellyn P. Marder, Senior Epidemiologist

Preprint
JMIR Publications

Abstract

Background: During the peak of the winter 2020-21 surge, the number of weekly reported COVID-19 outbreaks in Washington State was 231; the majority occurred in high-priority settings such as workplaces, community settings, and schools. The Washington State Department of Health (WA DOH) utilized automated address matching to identify cases and clusters at health care facilities. No other systematic, statewide outbreak detection methods were in place. This was a gap given high volumes of cases which delayed investigations and decreased data completeness, potentially leading to undetected outbreaks. To aid outbreak detection using systematic and automated methods, we initiated statewide cluster detection using SaTScan, implementing a space-time permutation model to identify COVID-19 clusters for local investigation.

Objectives: To improve outbreak detection, WA DOH initiated a systematic cluster detection model to identify timely and actionable COVID-19 clusters for local health jurisdiction (LHJ) investigation and resource prioritization. This report details the implementation of the model with an assessment of the tool's effectiveness.

Methods: Six LHJs participated in a pilot to test various model parameters including analysis type, geographic aggregation, maximum cluster radius, and data lag. Parameters were determined through heuristic criteria to detect clusters early when they are smaller so that interventions are more feasible. This study reviews all clusters detected after statewide implementation during July 17–December 17, 2021. The clusters were analyzed by LHJ population and disease incidence. Clusters were compared with reported outbreaks.

Results: A weekly, LHJ-specific retrospective space-time permutation model identified 2874 new clusters during the study period. While the weekly analysis included case data from the prior three weeks, nearly 60% of all clusters identified were timely—having occurred within one week of the analysis and early enough for intervention to prevent further transmission. There were 2874 reported outbreaks during this same period. Of those, 363 (12.8%) matched to at least one SaTScan cluster. The most frequent settings among reported and matched outbreaks were schools and youth programs (28.7%, 29.8%), workplaces (21.5%, 15.4%), and long-term care facilities (18.8%, 27.3%). Settings with the highest percentage of clusters that matched outbreaks were community settings (22.2%) and congregate housing (20.8%). Approximately one-third (32.8%) of matched outbreaks were identified by the model before cases were associated to the outbreak event in our surveillance system.

Conclusions: Our goal was to routinely and systematically identify timely and actionable COVID-19 clusters throughout the state. Regardless of population or incidence, the model identified reasonably sized, timely clusters statewide, successfully meeting the goals. Among some high priority settings subject to public health interventions throughout the pandemic, such as schools and community settings, the model identified clusters which were matched to reported outbreaks. In workplaces, another high priority setting, results suggest the model might be able to identify outbreaks sooner than existing outbreak detection methods.

Introduction

By March 2021, 340 323 COVID-19 cases were reported in Washington State. During the peak of the winter 2020–21 surge, the weekly incidence rate was 284.9 per 100 000 population and the number of reported outbreaks per week was 231 [1]. It is important to identify and investigate outbreaks of COVID-19 to reduce the spread of disease and protect vulnerable populations, such as long-term care facility residents [2]. Outbreaks in these settings are a high priority because residents have a greater risk of severe outcomes [3]. However, case and outbreak investigations can be resource-intensive [4]; thus, it is helpful to utilize automated and systematic methods as much as possible. To identify outbreaks in long-term care facilities, the Washington State Department of Health (DOH) routinely identified cases occurring at known health care facility addresses through an automated process; results are shared with local health jurisdictions (LHJs) [5].

In Washington State, there are 35 LHJs, which are the administrative public health authorities that manage and oversee health services at the county or regional level, including case and outbreak investigations for reportable diseases such as COVID-19. Aside from facility address matching, other systematic methods for cluster detection were rare. Reported outbreaks likely undercounted the true number of outbreaks due to varying detection methods and data disruptions. When the volume of cases dramatically increased during short periods of time, health care systems, including hospitals, laboratories, and public health, experienced delays which impacted the completeness and timeliness of reporting and processing case information. Furthermore, limited testing capacity during surges resulted in underreporting of mild or asymptomatic cases, further contributing to incomplete data [6–10]. To aid outbreak detection using systematic and automated methods, DOH initiated statewide cluster detection using SaTScan, implementing a space-time permutation model to identify timely and actionable COVID-19 clusters for local investigation.

Infectious disease space-time surveillance allows health officials to target resources and interventions at specific areas of emerging disease [11, 12]. Scan statistics are a public health surveillance method using data models to identify and evaluate emerging clusters of cases in a temporal, spatial, or space-time setting [13]. SaTScan is a free software widely used for spatial-temporal analysis [14]. In public health, SaTScan has been used for many reportable conditions [15], including foodborne illnesses [16–18], Lyme disease [19], HIV [20], and opioid overdoses [21, 22]. Space-time cluster detection of COVID-19 cases was implemented throughout the pandemic in many jurisdictions [11, 23–27]. SaTScan is a flexible and adaptable tool, allowing users to specify a variety of parameters (e.g., analysis type, probability model type, geographic and temporal data aggregation, study period length, spatial and temporal cluster sizes) [27]. Through this customization, the user can tailor the analysis to their specific datasets and objectives. Daily prospective Poisson models [11, 23, 24, 28] were frequently used for COVID-19; however, models varied based on data availability, the geographical area of study, and specific study goals [12, 29]. This report details our efforts to implement a customized systematic approach for COVID-19 cluster detection in Washington State using SaTScan along with an assessment of the tool's effectiveness.

Methods

Data Sources and Model Development

Case Data

During this period, laboratories in Washington State were required to report all SARS-CoV-2 test results to public health. Results included specimen collection date and patient demographics. COVID-19 cases were defined according to national surveillance [30]. Cases were geocoded based on residential address using DOH's Geocoder Web Service. Cases were excluded if the patient

resided out of state, did not have a positive molecular or antigen test during the model study period (which is defined as the time period included in each analysis), or if their address could not be geocoded.

Model Development

For the pilot, we chose a state-level retrospective space-time permutation model. The space-time permutation model was used because it does not require population testing rates which were not reliably available at the time. Case-patients were aggregated on specimen collection date, as a proxy for infection date, and census tract of their residential address, which was the most complete address available. We selected a maximum cluster duration of 7 days based on the median incubation period of COVID-19 [31]. Given this maximum cluster duration, we selected 21 days for the study period, which is the minimum recommended (i.e., three times the maximum cluster duration). We utilized a day-of-week adjustment given the variability in reporting on weekends. We tested three maximum cluster radii of 5km, 10km, and 20km to determine which provided the best yield across urban and rural settings. We used a 3-day lag to account for data reporting delays.

Six LHJs (Benton-Franklin Health District, Clark County Public Health, Kitsap Public Health District, Tacoma–Pierce County Health Department, Spokane Regional Health District, Whatcom County Health Department) participated in the pilot and provided feedback on the initial model. LHJs reported that the clusters identified from the initial model were larger (in both size and radius) than was actionable. Based on this feedback, we tested the following parameters to identify smaller, more actionable clusters: for analysis type, statewide vs. LHI-specific and weekly retrospective vs. daily prospective; for geographic aggregation, census tract vs. census block group; for maximum cluster radius, 5 km vs. 10 km vs. 20 km; for data lag, 1-day vs. 2-day vs. 3-day. To compare the retrospective and prospective models, we ran the daily model for 14 days and the weekly model on days 7 and 14. To determine the most appropriate lag, we analyzed data over several weeks using each lag and compared the identified clusters. We used heuristic criteria, prioritizing smaller (size and radius) and timely clusters, to finalize the parameters and features. Timely clusters were defined as having occurred within one week of the analysis—early enough for intervention to prevent further transmission. The final model was an LHI-specific, retrospective space-time permutation with case-patients aggregated on specimen collection date and census block group of their residential address. The maximum cluster duration was 7 days, study period 21 days, and maximum cluster radius of 20km. We utilized a day-of week adjustment and 2-day lag.

SaTScan Cluster Data

The SaTScan cluster dataset included all clusters detected since the inaugural weekly analysis in August through December 2021, representing cases with specimen collection dates during July 17–December 17, 2021. This period was selected for its relatively stable incidence and predominance of Delta variant. Variables included were LHI, cluster start and end date, radius, number of cases, and p-value. Statistical significance was defined as $\alpha < 0.1$ to account for multiple testing. Because of the weekly retrospective design with the 21-day study period, the same or similar cluster could be identified in subsequent weeks. To deduplicate clusters included in this comparison, we compared clusters week-to-week using the Jaccard Similarity (JS) Index [32]. Clusters with a Jaccard Similarity (JS) Index $\leq 12.0\%$ were categorized as new; all other clusters were considered ongoing.

Reported Outbreak Data

Outbreak detection and reporting varies statewide and by outbreak setting. Known COVID-19 outbreaks are captured in our state surveillance system, where they are linked to cases. Because COVID-19 outbreak definitions and reporting requirements differ by setting and have changed over time, we applied a standardized definition for this analysis. Outbreaks were defined as ≥ 2 cases that

were associated with an outbreak event in the state surveillance system and had specimen collection dates within 14 days of each other. Start and end dates were based on the first and last specimen collection date. We included outbreaks with start and end dates during July 17–December 17, 2021 and collapsed the original 64 outbreak settings into 11 broad categories: school and youth programs, workplace, retail, congregate housing, food and beverage establishments, community, colleges, travel, long term care facilities, outpatient and inpatient settings. Outbreaks linked to military settings as indicated by LHJs were excluded.

Original data collection was conducted as part of WA DOH COVID-19 surveillance. As such, in accordance with the Office for Human Research Protections guidelines, institutional review board (IRB) approval was not required [33]. As part of public health surveillance activities authorized by a public health authority, the secondary analysis of those data described in this study was excluded from the definition of research provided by the revised Common Rule [34] and did not receive approval or an exemption from an IRB. The researchers were granted access to the data as part of their official duties within the WA DOH.

Analysis

Descriptive Analysis

We compared SaTScan clusters among LHJs categorized by population size. LHJs with a population $\geq 400\,000$ persons were classified as large; 90000–399 999 as medium; 25000–89999 as small; and ≤ 24999 as rural. Statistical significance was determined using chi-square and Wilcoxon tests, using $\alpha < .05$. To assess correlation among study period incidence (SPI) and cluster characteristics, we calculated correlation coefficients. SPI was defined as the average number of cases per 100 000 population during the SaTScan 21-day study period. Cluster size was defined as number of cases per cluster and timeliness as clusters ending within one week before the analysis date.

Outbreak-Cluster Comparison

To assess the ability of the model to detect true outbreaks, we matched SaTScan clusters, including statistically significant and non-significant clusters, with reported outbreaks. We validated a random sample of outbreak-cluster pairs with ≥ 1 case in common to refine the matching criteria. Because the mean size of outbreaks and clusters vastly differed, we selected two matching criteria. First, $\geq 20\%$ of outbreak cases had to be identified in a cluster and, second, $\geq 6\%$ of cluster cases had to be identified in an outbreak. We calculated summary statistics of the matched pairs. Matched and unmatched outbreaks were compared to determine the types and characteristics of outbreaks identified as SaTScan clusters. Statistical significance was assessed using chi-square tests and $\alpha < .05$. We used R (R Core Team, 2021) to run SaTScan (Version 9.6) and for all other analyses.

Results

Model Development

During pilot testing, the statewide model identified clusters that were large in radius and size, which did not meet the goal of actionable clusters. Independent models for each LHM with the same parameters produced smaller clusters, which did meet that goal. Therefore, our final design included LHM-specific models that were combined to provide statewide results.

Next, we compared daily prospective and weekly retrospective models. There were 12 significant clusters identified by the daily prospective model; 10 (83%) of which were also identified by the weekly retrospective model. While the daily prospective used more recent data, the model required more resources to produce and review. Given overlap between the model output and resource constraints, we selected the weekly retrospective analysis for the final model.

In comparing 1-, 2-, and 3-day lag periods, we found about half (47/90, 52%) identified similar clusters regardless of lag. The remainder identified similar clusters across the 2- and 3-day lag. We chose a 2-day lag to maximize timeliness.

The initial model, which aggregated cases to census tracts, produced clusters that were too large geographically for public health action. Using the smaller unit of census block group was favorable based on heuristic criteria.

Lastly, we tested various maximum cluster radii. The 5 km model identified the smallest clusters (median radius: 3.3 km, range: 0.0–5.0 km), followed by the 10 km model (median radius: 5.2 km, range: 0.0–10.0 km), and then the 20 km model (median radius: 6.1 km, range: 0.0–19.5 km). The 20 km radius was determined to capture both smaller clusters in urban areas and larger clusters in rural areas, and thus was selected for the final model.

Descriptive Analysis

Cluster Results

During July 17–December 17, 2021, there were 341 505 COVID-19 cases reported among Washington State residents. Of these, 316 642 (92.7%) had a geocoded residential address. A median of 41217 cases were included in each weekly SaTScan analysis (range: 24996–61069); a total of 4659 clusters were identified (weekly median: 236; range: 197–275). Clusters included 98172 unique cases (28.6% of reported COVID-19 cases); most (2874/4659, 61.7%) were new clusters. The remaining analyses are limited to new clusters.

Among the 2874 new clusters, 887 (30.9%) occurred in large LHJs, 890 (31.0%) in medium LHJs, 734 (25.5%) in small LHJs, and 363 (12.6%) in rural LHJs (Table 1). Overall, the median cluster size was 15 cases (range: 2–1045) and significantly differed across population groups (large $P < .001$; medium $P < .001$; small $P < .001$; rural $P < .001$). Rural LHJs had the smallest median cluster size (4 cases, range: 2–28) and large LHJs the largest (27 cases, range: 5–1045). The median number of weekly clusters per LHJ also differed significantly: rural LHJs had the least (2, range: 1–6; $P < .001$), followed by small (3, range: 1–9; $P < .001$), medium (5, range: 1–12; $P < .001$), and large LHJs (9, range: 2–21; $P < .001$). Overall, the median cluster radius was 2.4 km. The radius significantly differed for medium (2.5 km; $P = 0.045$), small (3.8 km; $P < .001$), and rural LHJs (0.0 km; $P < .001$). Of the 2874 clusters, 486 (16.9%) were statistically significant in the weekly analysis. Rural LHJs had a significantly smaller percentage of significant clusters (37/363, 10.2%; $P < .001$). There was no significant difference in duration (overall: 4 days, range 1–7) or timeliness (overall: 486/2874, 58.2%) by LHJ population group.

Table 1. Characteristics of new COVID-19 clusters identified by the SaTScan model, by local health jurisdiction (LHJ) population group, Washington State, July 17–December 17, 2021

	LHJ Population Group ^a												Total (n=2874)	
	Large (n=887)			Medium (n=890)			Small (n=734)			Rural (n=363)				
	Median	(Range)	<i>P</i> value	Median	(Range)	<i>P</i> value	Median	(Range)	<i>P</i> value	Median	(Range)	<i>P</i> value	Median	(Range)
Cluster size ^{b,c}	27	(5–1045)	<.001	17	(2–362)	<.001	9	(2–113)	<.001	4	(2–28)	<.001	15	(2–1045)
Weekly clusters per LHJ ^c	9	(2–21)	<.001	5	(1–12)	<.001	3	(1–9)	<.001	2	(1–6)	<.001	4	(1–21)
Radius (km) ^c	2.2	(0.0–19.9)	.30	2.5	(0.0–20.0)	.045	3.8	(0.0–20.0)	<.001	0.0	(0.0–19.8)	<.001	2.4	(0.0–20.0)
Duration (days) ^c	4	(1–7)	<.001	4	(1–7)	.11	4	(1–7)	<.001	4	(1–7)	.002	4	(1–7)
	<i>n</i>	(%)	<i>P</i> value	<i>n</i>	(%)	<i>P</i> value	<i>n</i>	(%)	<i>P</i> value	<i>n</i>	(%)	<i>P</i> value	<i>n</i>	(%)
Significant clusters ^{d,e}	162	(18.3)	.30	171	(19.2)	.47	116	(15.8)	.07	37	(10.2)	.002	486	(16.9)

Small clusters ^{d,f}	490	(55.2)	.27	525	(59.0)	.81	452	(61.6)	.29	207	(57.0)	.83	1674	(58.2)
-------------------------------	-----	--------	-----	-----	--------	-----	-----	--------	-----	-----	--------	-----	------	--------

Large: County population ≥400,000 persons; medium: 90,000–399,999 persons; small: 25,000–89,999 persons; and rural: ≤24,999 persons.

a: Cases per cluster.

b: Statistical significance tested with unpaired two-sided Wilcoxon test.

c: Statistical significance tested with chi-square test.

d: Cluster's p-value is <.01.

e: Cluster's end date was within one week of analysis date.

The SPI was significantly positively correlated with the weekly number of cluster cases among large ($R=.82$; $P < .001$), medium ($R = 0.76$; $P = .01$), and small ($R = 0.73$; $P = .01$) LHJs and with cluster size among large ($R = 0.73$; $P = .01$) and medium LHJs ($R = 0.82$; $P < .001$) (Figure 1). There were few non-significant moderate correlations between SPI and median cluster size, number of weekly clusters, and percentage of reported cases associated with clusters. There was no correlation between timeliness of clusters and SPI (large $R=-0.19$; medium $R=-0.22$; small $R=-0.12$; and rural $R=-0.42$). [insert Figure 1.]

Outbreak-Cluster Comparison

There were 2874 reported outbreaks during July 17–December 17, 2021, and 363 (12.8%) matched to ≥1 SaTScan cluster (Table 2). Among reported outbreaks, the most common settings were schools and youth programs (825/2874, 28.7%), workplaces (617/2874, 21.5%), and long-term care facilities (541/2874, 18.8%). Among the 363 matched outbreaks, the most common settings were also schools and youth programs (108/363, 29.8%), long-term care facilities (99/363, 27.3%), and workplaces (56/363, 15.4%), as well as congregate housing (44/363, 12.1%). Of the 2874 reported outbreaks, the settings with the largest percentage of reported outbreaks which matched to clusters were community settings (16/72, 22.2%), congregate housing (44/212, 20.8%), long-term care facilities (99/541, 18.3%) and school and youth programs (108/825, 13.1%). Settings with <10% of matched outbreaks were workplaces (56/617, 9.1%), food and beverage establishments (10/113, 8.8%), inpatient (6/77, 7.8%), outpatient (8/104, 7.7%), colleges (1/16, 6.3%), and retail settings (14/287, 4.9%).

The median duration among all reported outbreaks was 8 days (range: 1–151) and 10 days (range: 1–151) among matched outbreaks. Among all reported outbreaks, the median size was 4 cases (range: 2–256) and among matched outbreaks, it was 6 cases (range: 2–232).

Outbreaks in community settings (16/72, 22.2%; $P = .01$), congregate housing (44/212, 20.8%; $P < .001$), and long-term care facilities (99/541, 18.3%; $P < .001$) were more likely to match to clusters than other settings, while outbreaks in workplaces (56/617, 9.1%; $P = .003$) and retail settings (14/287, 4.9%; $P < .001$) were less likely to match. For one-third (119/363, 32.8%) of matched outbreaks, all cases were associated to the outbreak event in the surveillance system after the SaTScan cluster was identified. This resulted in “report dates” later than the “analysis date.” Of the 363 matched outbreaks, those in workplaces (24/56, 42.9%; $P < .001$) and schools and youth program settings (42/108, 38.9%; $P < .001$) were more likely to have report dates later than analysis dates, while outbreaks in community settings (4/16, 25.0%; $P < .001$), long-term care facilities (20/99, 20.2%; $P < .001$), and congregate housing (8/44, 18.2%; $P < .001$) were less likely (Table 2). Table 2. Reported outbreaks compared with matched^a outbreaks by setting type, Washington State, July 17–December 17, 2021.

	Reported outbreaks		Matched outbreaks					
	n	(%)	n	Column percentage (%)	Row percentage (%)	χ^2 P value	Report date after analysis date ^b n/N (%)	χ^2 P value
Non-health care settings								
Schools and youth programs	825	(28.7)	108	(29.8)	(13.1)	.63	42/108	(38.9) <.001
Workplace ^c	617	(21.5)	56	(15.4)	(9.1)	.003	24/56	(42.9) <.001
Retail settings ^d	287	(10.0)	14	(3.9)	(4.9)	<.001	7/14	(50.0) .16
Congregate housing ^e	212	(7.4)	44	(12.1)	(20.8)	<.001	8/44	(18.2) <.001
Food and beverage establishments ^f	113	(3.9)	10	(2.8)	(8.8)	.22	6/10	(60.0) .62
Community ^g	72	(2.5)	16	(4.4)	(22.2)	.01	4/16	(25.0) <.001
Colleges	16	(0.6)	1	(0.3)	(6.3)	.44	1/1	(100.0) -
Travel	10	(0.3)	1	(0.3)	(10.0)	.80	1/1	(100.0) -

Health care settings									
Long-term care facilities ^b	541	(18.8)	99	(27.3)	(18.3)	<.001	20/99	(20.2)	<.001
Outpatient	104	(3.6)	8	(2.2)	(7.7)	.12	5/8	(62.5)	.77
Inpatient	77	(2.7)	6	(1.7)	(7.8)	.20	1/6	(16.7)	.24
Total	2,874	(100.0)	363	(100.0)	(12.8)		119/363	(32.8)	

^aMatched is defined as $\geq 20\%$ of outbreak cases identified in a SaTScan cluster and $\geq 6\%$ of SaTScan cluster cases identified in an outbreak.

^bAll cases were associated to the outbreak event in the surveillance system after the SaTScan cluster was identified.

^cWorkplace manufacturing (food n=24; non-food n=162), agricultural employee provided housing (n=71), construction (n=49), professional services/office-based (n=43), utilities (n=38), agencies, facilities, etc (n=37), and other settings (n=193).

^dRetail services includes grocery (n=144), retail (n=129), personal care and service (hair, nails) (n=13), and other (n=1).

^eCongregate housing includes shelter/homeless services (n=126), correctional settings (n=59), congregate housing (n=10), juvenile justice settings (n=6), and other (n=11).

^fFood and beverage establishments include food service/restaurants (n=101), bar/nightclub (n=11), and other (n=1).

^gCommunity includes those in places of worship (n=27), hospitality (n=16), large gatherings (n=13), private events (n=11), and other (n=16).

^hLong-term care facilities include assisted living (n=260), nursing home (n=158), adult family home (n=88), enhanced services facility/ICF/IID (n=18), senior living (n=14), and other (n=3).

A reported outbreak could match to more than one cluster and vice versa. The 363 outbreaks matched to 349 SaTScan clusters, which resulted in 384 unique outbreak-cluster pairs. Matched outbreaks had a median size of 6 cases and duration of 10 days. Matched clusters had a median size of 17 cases and duration of 5 days. The median number of linked cases (those common to the outbreak and cluster in a matched pair) was 3 (range: 1–58). This represents a median of 40.0% of outbreak cases that were part of the cluster and 13.0% of cluster cases that were part of the outbreak. The median JS was 10.0% (Table 3). For an illustrative example, an 8-case workplace outbreak matched a 29-case cluster by 8 linked cases, with 100% of outbreak cases in the cluster and 28% of cluster cases in the outbreak, resulting in JS=28%.

Table 3. Characteristics of COVID-19 reported outbreaks and detected clusters that matched, Washington State, July 17–December 17, 2021

	Matched outbreaks (n=363)		Matched clusters (n=349)	
	Median	(Range)	Median	(Range)
Number of cases	6	(2–232)	17	(4–147)
Duration, days	10	(1–151)	5	(1–7)
	Matched outbreak-cluster pairs (n=384)			
	Median	(Range)		
Number of linked cases	3	(1–58)		
Percentage of outbreak cases linked	40	(20–100)		
Percentage of cluster cases linked	13.0	(6–100)		
Jaccard similarity (%)	10.0	(5–100)		

^aMatched is defined as $\geq 20\%$ of outbreak cases identified in a SaTScan cluster and $\geq 6\%$ of cluster cases identified in an outbreak

Discussion

During a time when LHJs were heavily burdened with case and outbreak investigations, DOH initiated systematic statewide cluster detection for COVID-19 using SaTScan. SaTScan parameters can be adjusted based on study objectives, data availability, and disease and population characteristics [27, 35]. Washington State required a model tailored to its overtaxed information systems and decentralized public health system, which serves almost 8 million residents in 39 diverse counties. We collaborated with LHJs to implement a space-time permutation model that maximized real-time surveillance data and identified opportunities for public health intervention.

We assessed a daily prospective model, which would have provided the most real-time surveillance; however, it did not produce substantially different results than a weekly retrospective model.

Considering time and resource constraints faced by LHJs, we determined the weekly retrospective model was the most appropriate. Despite having parameters unique from other studies [11, 23, 24], the model worked well in Washington and identified clusters early enough for intervention. It is important to identify outbreaks as quickly as possible (when there are fewer cases) to prevent further transmission and illness. While outbreak size may vary due to population and disease transmission levels, our analysis showed that nearly 60% of clusters were timely regardless of population or incidence. By identifying clusters promptly, LHJs can conduct targeted investigations, implement containment measures quickly, and prioritize resource distribution to areas with the greatest need.

The number of weekly clusters and cluster size were both correlated with population: more and larger clusters were identified in highly populated urban areas which generally have larger health departments with greater epidemiologic capacity [36]. These data support the use of a systematic statewide approach as it benefits LHJs with varying populations and resource capacity, ensuring that all LHJs, regardless of resources, can effectively manage outbreaks by leveraging a standardized, statewide surveillance system.

Space-time scan statistics have successfully identified respiratory disease outbreaks [13, 24, 49], including COVID-19 [37]. To determine if the model could identify true outbreaks, we compared clusters with reported outbreaks. Generally, we did not expect a high degree of alignment between reported outbreaks and SaTScan clusters, given SaTScan was intended to fill gaps in current outbreak detection methods. Given the use of residential address as the geographic input for the model, outbreaks in congregate living settings (e.g., long-term care facilities, corrections) had a high cluster matching rate as expected. Outbreaks that occur further from the home, such as in workplaces, would likely be harder for the model to detect.

During the COVID-19 pandemic, several settings have been a high priority and focus for public health interventions, including workplaces, community settings, and schools. Workplaces, the second most common setting among reported outbreaks in Washington State, play a central role in COVID-19 community transmission [38-40], and have been subject to safety prevention requirements over the course of the pandemic [41] to reduce the spread of the disease. As expected, workplace outbreaks matched fewer clusters than all reported outbreaks. Incorporating workplace address may have helped, however occupational and industry data for COVID-19 are largely incomplete and often not standardized, which may create barriers to systematically identify workplace outbreaks [39, 42]. Efforts to collect and standardize these data and improvement of occupational health surveillance may address this [38].

Community settings have also been a high priority for public health interventions. In fall 2021, when SaTScan was first implemented in Washington State, many efforts were made to implement safety precautions, including limiting capacity at events and gatherings [43], masking [44], and vaccination requirements at large events [45]. Less than 3% of all reported outbreaks were among community settings, yet 22.2% of these outbreaks were linked to clusters, suggesting SaTScan might be a useful method for systematically detecting these types of outbreaks.

The model performed moderately well for schools and youth programs, another high priority and focus of public health interventions. For the 2021–22 school year in Washington State, public health officials collaborated on guidance to allow students to return safely to in-person instruction [46]. Given their residential proximity, we expected a high cluster match rate for educational settings. Schools and youth programs were the most common setting among reported outbreaks as well as matched outbreaks. However, only 13.1% of school outbreaks matched to a cluster compared to 12.8% overall, arguably leaving room for improvement. The model could be further tailored to capture outbreaks in educational settings by incorporating school district-level models or including age as an input variable.

Lastly, about a third of outbreaks that matched to clusters had a report date in the surveillance system that was later than when the cluster was identified by the model. This suggested that the model might be able to identify outbreaks earlier than existing outbreak detection methods. Detecting outbreaks

sooner allows quicker implementation of containment measures, reducing the spread of the disease and minimizing the impact on communities. By automating the detection process, local public health workers can spend more time on the outbreak intervention measures and less on the initial outbreak detection and investigation work.

The findings in this report are subject to several limitations. First, the model does not account for irregular geographics, such as major waterways. Second, residential address was used to determine outbreak detection as this was the most complete and available. However, residential address may not be the most representative of where exposure occurs and may limit the identification of clusters where transmissions happened away from home. Third, reported outbreaks are identified through many indicators and sources which vary by LHJ, contain known biases including lags or gaps during surges [6-9], and likely underestimate actual outbreaks [47-48]. Fourth, because the final model was run independently for each LHJ, cross-jurisdictional clusters were not identified and thus not matched with similar outbreaks. Lastly, as expected, relatively few SaTScan clusters matched outbreaks, and without further investigation to identify epidemiologic or genomic linkages, we cannot determine if unmatched clusters are true outbreaks.

Conclusion

Our goal was to implement a statewide systematic cluster detection process to identify COVID-19 clusters. DOH successfully developed and implemented a SaTScan space-time permutation model that met these goals within the state's unique structure and systems. Regardless of LHJ population, the model identified reasonably sized, timely clusters for investigation and resource prioritization. Cluster size increased with incidence but likely remain actionable even for smaller LHJs. Among reported outbreaks that matched clusters, the model performed well in congregate living settings, schools, and events, which have been high priority settings throughout the COVID-19 pandemic. There is opportunity to tailor the model to further improve cluster detection in some settings, such as workplaces and schools. Evidence suggests running a weekly model might identify some outbreaks sooner than existing outbreak detection methods, particularly in workplace outbreaks where data collection remains a challenge. In summary, our SaTScan model was able to identify timely, actionable clusters, especially in high priority settings, which can serve to support outbreak detection to reduce further COVID-19 transmission.

Acknowledgments

The authors thank Benton-Franklin Health District, Clark County Public Health, Kitsap Public Health District, Tacoma-Pierce County Health Department, Spokane Regional Health District, and Whatcom County Health Department for their participation in the pilot. We also acknowledge the contributions of the Non-healthcare Congregate Settings Program and the Healthcare-Associated Infections Epidemiology Program at the Washington State Department of Health. The authors received no financial support for the research, authorship, and/or publication of this article.

Data Availability

The datasets generated and/or analyzed during the current study are not publicly available due to the presence of personally identifiable information (PII). For privacy reasons, these data will not be made available upon request.

Conflicts of Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. This publication contains general information only and Deloitte is not, by means of this publication, rendering accounting, business, financial, investment, legal, tax, or other professional advice or services. This publication is not a substitute for such professional advice or

services, nor should it be used as a basis for any decision or action that may affect your business. Before making any decision or taking any action that may affect your business, you should consult a qualified professional advisor. Deloitte shall not be responsible for any loss sustained by any person who relies on this publication.

Abbreviations

WA DOH: Washington State Department of Health

LHJ: Local health jurisdiction



References

1. COVID-19 case and outbreak data. Unpublished raw data [database]. Tumwater, WA: Washington Disease Reporting System; 2023.
2. Pray IW, Kocharian A, Mason J, Westergaard R, Meiman J. Trends in Outbreak-Associated Cases of COVID-19 — Wisconsin, March–November 2020. *MMWR Morb Mortal Wkly Rep* 2021;70:114–117. DOI: <http://dx.doi.org/10.15585/mmwr.mm7004a2>
3. Telford CT, Onwubiko U, Holland DP, et al. Preventing COVID-19 Outbreaks in Long-Term Care Facilities Through Preemptive Testing of Residents and Staff Members - Fulton County, Georgia, March–May 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69(37):1296-1299. Published 2020 Sep 18. doi:10.15585/mmwr.mm6937a4
4. Bonney T, Grant MP. Local health department engagement with workplaces during the COVID-19 pandemic-Examining barriers of and facilitators to outbreak investigation and mitigation. *Front Public Health*. 2023;11:1116872. Published 2023 Mar 17. doi:10.3389/fpubh.2023.1116872
5. Wu, C., Tasslimi, A., et al. (2024). Enhanced COVID-19 surveillance methods for early detection in high-risk settings: Utilizing fuzzy address matching and exact phone number matching to improve case identification in congregate care and correctional facilities. Abstract presented at the Council of State and Territorial Epidemiologists (CSTE) Annual Conference, 2024.
6. Pagel C, Yates CA. Tackling the pandemic with (biased) data. *Science*. 2021;374(6566):403-404. doi:10.1126/science.abi6602
7. Britton T, Scalia Tomba G. Estimation in emerging epidemics: biases and remedies. *J R Soc Interface*. 2019;16(150):20180670. doi:10.1098/rsif.2018.0670
8. Overton CE, Stage HB, Ahmad S, et al. Using statistics and mathematical modelling to understand infectious disease outbreaks: COVID-19 as an example. *Infect Dis Model*. 2022;5:409-441. doi:<https://doi.org/10.1016/j.idm.2020.06.008>
9. Boore A, Herman KM, Perez AS, et al. Surveillance for Foodborne Disease Outbreaks --- United States, 2007. *MMWR*. 2010;59(31):973-979.
10. Badker R, Miller K, Pardee C, et al. Challenges in reported COVID-19 data: best practices and recommendations for future epidemics. *BMJ Glob Health*. 2021;6(5):e005542. doi:10.1136/bmjgh-2021-005542
11. Hohl A, Delmelle EM, Desjardins MR, Lan Y. Daily surveillance of COVID-19 using the prospective space-time scan statistic in the United States. *Spat Spatiotemporal Epidemiol*. 2020;34:100354. doi:10.1016/j.sste.2020.100354
12. Leal-Neto OB, Santos FAS, Lee JY, Albuquerque JO, Souza WV. Prioritizing COVID-19 tests based on participatory surveillance and spatial scanning. *Int J Med Inform*. 2020;143:104263. doi:10.1016/j.ijmedinf.2020.104263
13. Kulldorff M, Heffernan R, Hartman J, Assunção R, Mostashari F (2005) A Space–Time Permutation Scan Statistic for Disease Outbreak Detection. *PLoS Med* 2(3): e59. <https://doi.org/10.1371/journal.pmed.0020059>
14. SaTScan Software for the spatial, temporal, and space-time scan statistics. Accessed September 20, 2022. <https://www.satscan.org/>
15. Greene SK, Peterson ER, Kapell D, et al. Daily Reportable Disease Spatiotemporal Cluster Detection, New York City, New York, USA, 2014–2015. *Emerg Infect Dis*. 2016;22(10):1808-1812. doi:10.3201/eid2210.160097.
16. Hughes GJ, Gorton R. An evaluation of SaTScan for the prospective detection of space-time *Campylobacter* clusters in the North East of England. *Epidemiol Infect*. 2013;141(11):2354-2364. doi:10.1017/S0950268812003135

17. Edens C, Alden NB, Danila RN, et al. Multistate analysis of prospective Legionnaires' disease cluster detection using SaTScan, 2011–2015. Okell LC, ed. PLoS One. 2019;14(5):e0217632. doi:10.1371/journal.pone.0217632
18. Latash J, Greene SK, Stavinsky F, et al. Salmonellosis Outbreak Detected by Automated Spatiotemporal Analysis — New York City, May–June 2019. MMWR. 2020;69(26):815-819.
19. Kugeler KJ, Farley GM, Forrester JD, Mead PS. Geographic Distribution and Expansion of Human Lyme Disease, United States. Emerg Infect Dis. 2015;21(8):1455-1457. doi:10.3201/eid2108.141878
20. Hixon BA, Omer SB, del Rio C, Frew PM. Spatial Clustering of HIV Prevalence in Atlanta, Georgia and Population Characteristics Associated with Case Concentrations. J Urban Health. 2011;88(1):129-141. doi:10.1007/s11524-010-9510-0
21. Chen J, Chang HG, Hammer M, D'Anna N, Gelberg K. Using Scan Statistic to Detect Heroin Overdose Clusters with Hospital Emergency Room Visit Data. Online J Public Health Inform. 2018;10(1). doi:10.5210/ojphi.v10i1.8344
22. Marks C, Carrasco-Escobar G, Carrasco-Hernández R, et al. Methodological approaches for the prediction of opioid use-related epidemics in the United States: a narrative review and cross-disciplinary call to action. Transl Res. 2021;234:88-113. doi:10.1016/j.trsl.2021.03.018
23. Greene, SK, Peterson ER, Balan D, et al. Detecting COVID-19 clusters at high spatiotemporal resolution, New York City, New York, USA, June–July 2020. Emerg Infect Dis. 2021;27(5):1500-1504. doi:10.3201/eid2705.203583.
24. Desjardins MR, Hohl A, Delmelle EM. Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: Detecting and evaluating emerging clusters. Appl Geogr. 2020;118:102202. doi:10.1016/j.apgeog.2020.102202
25. Alves HJ de P, Fernandes FA, Lima KP de, Batista BD de O, Fernandes TJ. Incidence and lethality of COVID-19 clusters in Brazil via circular scan method. Rev Bras Biom. 2021;39(4):556-570. doi:10.28951/rbb.v39i4.539
26. AlQadi H, Bani-Yaghoub M, Balakumar S, Wu S, Francisco A. Assessment of retrospective COVID-19 spatial clusters with respect to demographic factors: Case study of Kansas City, Missouri, united states. IJERPH. 2021;18(21):11496. doi:10.3390/ijerph182111496
27. Levin-Rector, A, Kulldorff, M, et al. Prospective Spatiotemporal Cluster Detection using SaTScan: A Tutorial for Designing and Finetuning a System to Detect Reportable Communicable Disease Outbreaks. JMIR Public Health and Surveillance. 2023 July 7. DOI: <https://doi.org/10.2196/preprints.50653>
28. Martines MR, Ferreira RV, Toppa RH, Assunção LM, Desjardins MR, Delmelle EM. Detecting space–time clusters of COVID-19 in Brazil: mortality, inequality, socioeconomic vulnerability, and the relative risk of the disease in Brazilian municipalities. J Geogr Syst. 2021;23(1):7-36. doi:10.1007/s10109-020-00344-0
29. Aturinde A, Mansourian A. Space—Time Surveillance of COVID-19 Seasonal Clusters: A Case of Sweden. ISPRS International Journal of Geo-Information. 2022; 11(5):307. <https://doi.org/10.3390/ijgi11050307>
30. Update to the standardized surveillance case definition and national notification for 2019 novel coronavirus disease (COVID-19). Published August 2021. Accessed September 20, 2022. https://cdn.ymaws.com/www.cste.org/resource/resmgr/ps1/ps2021/21-ID-01_COVID-19_updated_Au.pdf
31. Xu X, Wu Y, et al. Assessing changes in incubation period, serial interval, and generation time of SARS-CoV-2 variants of concern: a systematic review and meta-analysis. BMC Med. 2023 Sep 29;21(1):374. doi: 10.1186/s12916-023-03070-8. PMID: 37775772; PMCID: PMC10541713.
32. Glen, Stephanie. Jaccard index / similarity coefficient. Statistics How To. Published December 3, 2016. Accessed July 13, 2022. <https://www.statisticshowto.com/jaccard-index/>

33. Office for Human Research Protections. OHRP Guidelines on COVID-19. U.S. Department of Health and Human Services. Published April 8, 2020. Accessed July 20, 2024. <https://www.hhs.gov/ohrp/regulations-and-policy/guidance/ohrp-guidance-on-covid-19/index.html>
34. U.S. Department of Health and Human Services. Draft Guidance on Activities Deemed Not to Be Research: Public Health Surveillance. U.S. Department of Health and Human Services. <https://www.hhs.gov/ohrp/regulations-and-policy/requests-for-comments/draft-guidance-activities-deemed-not-be-research-public-health-surveillance/index.html>. Accessed May 18, 2024.
35. SaTScan User Guide for Version 10.1. Published July, 2022. https://www.satscan.org/cgi-bin/satscan/register.pl/SaTScan_Users_Guide.pdf?todo=process_userguide_download
36. O’Keefe KA, Shafir SC, Shoaf KI. Local Health Department Epidemiologic Capacity: A Stratified Cross-Sectional Assessment Describing the Quantity, Education, Training, and Perceived Competencies of Epidemiologic Staff. *Front Public Health*. 2013;1. doi:10.3389/fpubh.2013.00064
37. Xu F, Beard K. A comparison of prospective space-time scan statistics and spatiotemporal event sequence based clustering for COVID-19 surveillance. Odoi A, ed. *PLoS One*. 2021;16(6):e0252990. doi:10.1371/journal.pone.0252990
38. Baker MG. Occupational Health Surveillance as a Tool for COVID-19 Prevention. *Am J Public Health*. 2021;111(6):999-1001. doi:10.2105/AJPH.2021.306269
39. Baker MG, Peckham TK, Seixas NS. Estimating the burden of United States workers exposed to infection or disease: A key factor in containing risk of COVID-19 infection. *PLoS One*. 2020;15(4):e0232452. Published 2020 Apr 28. doi:10.1371/journal.pone.0232452
40. Luckhaupt SE, Horter L, Groenewold MR, et al. COVID-19 Outbreaks Linked to Workplaces, 23 US Jurisdictions, August-October 2021. *Public Health Rep*. 2023;138(2):333-340. doi:10.1177/00333549221138294
41. L&I Requirements and Guidance for Preventing COVID-19. Labor and Industries Washington State. Updated September 14, 2022. Accessed August 28, 2022. <https://www.lni.wa.gov/agency/outreach/coronavirus/requirements-and-guidance-for-preventing-covid-19>
42. Pray IW, Grajewski B, Morris C, et al. Measuring work-related risk of COVID-19: comparison of COVID-19 incidence by occupation and industry - Wisconsin, September 2020-May 2021 [published online ahead of print, 2022 Aug 4]. *Clin Infect Dis*. 2022;ciac586. doi:10.1093/cid/ciac586
43. Brown CM, Vostok J, Johnson H, et al. Outbreak of SARS-CoV-2 Infections, Including COVID-19 Vaccine Breakthrough Infections, Associated with Large Public Gatherings - Barnstable County, Massachusetts, July 2021. *MMWR*. 2021;70(31):1059-1062. Published 2021 Aug 6. doi:10.15585/mmwr.mm7031e2
44. Inslee J. Washington Ready. Vol 20-25.27.; 2021. Accessed September 27, 2022. https://www.governor.wa.gov/sites/default/files/proclamations/proc_20-25.17.pdf
45. Inslee J. Large Event COVID-19 Vaccine Verification. Vol 21-16.; 2021. https://www.governor.wa.gov/sites/default/files/proclamations/21-16%20-%20COVID-19%20-%20Large%20Events%20%28tmp%29.pdf?utm_medium=email&utm_source=govdelivery
46. COVID-19 testing for K-12 schools. Learn to Return. Accessed August 29, 2022. <https://learntoreturn.org/>
47. COVID-19 Outbreaks in Washington State K-12 Schools. Updated September 30, 2022. Accessed October 12, 2022.
48. Bonwitt, J, Deya, R, Currie, D, et al. COVID-19 Surveillance and Investigations in Workplaces — Seattle & King County, Washington, June 15–November 15, 2020. *MMWR*.

2021;70(25):. Published 2021 June 24. doi: <http://dx.doi.org/10.15585/mmwr.mm7025a3>

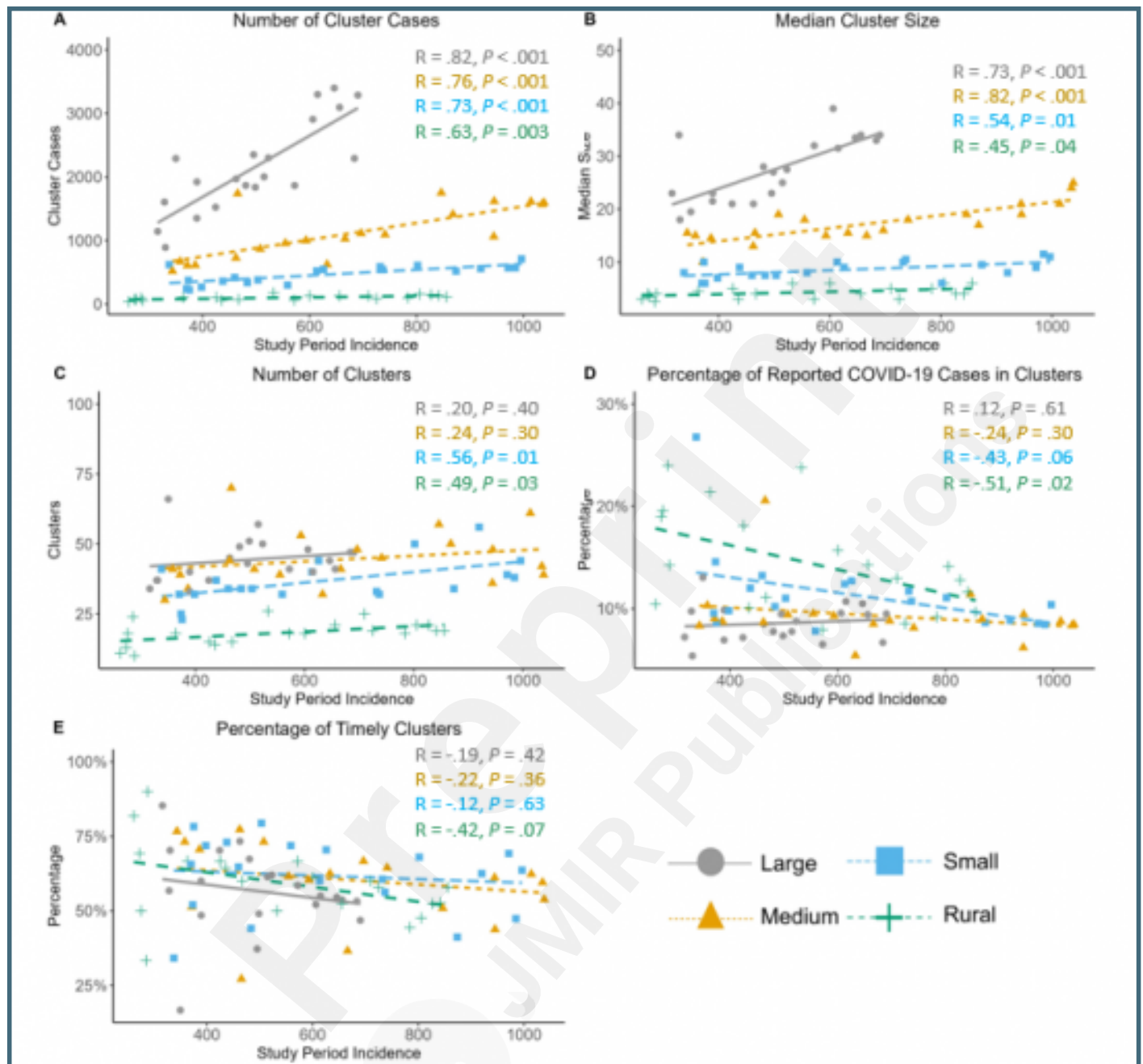
49. van den Wijngaard CC, van Asten L, van Pelt W, et al. Syndromic surveillance for local outbreaks of lower-respiratory infections: would it work?. PLoS One. 2010;5(4):e10406. Published 2010 Apr 29. doi:10.1371/journal.pone.0010406



Supplementary Files

Figures

Untitled.



Study period incidence by cluster characteristics stratified by LHJ population group.

