

The Pennsylvania State University

The Graduate School

**THE ROLE OF CATEGORIES IN SHAPING OUR UNDERSTANDING  
OF INFECTIOUS DISEASE OUTBREAKS**

A Dissertation in

Department of Biology

by

Callum R. K. Arnold

© 2024 Callum R. K. Arnold

Submitted in Partial Fulfilment  
of the Requirements  
for the Degree of

Doctor of Philosophy

December 2024

The dissertation of Callum R. K. Arnold was reviewed and approved by the following:

Dr. Katriona Shea  
Professor of Biology  
Alumni Professor in the Biological Sciences  
Chair of Committee

Dr. Matthew J. Ferrari  
Professor of Biology  
Center for Infectious Disease Dynamics Director  
Major Field Member & Dissertation Advisor

Dr. Maciej Boni  
Professor of Biology, Temple University  
Major Field Member

Dr. Stephen Berg  
Assistant Professor of Statistics  
Outside Unit & Field Member

## **Abstract**

Infectious diseases have impacted civilisations throughout history, shaping human interactions and influencing the growth of populations. Despite advances in the control of disease, they still pose a substantial threat to human life. To minimize disease burden it is essential to gain a greater understanding of the underlying dynamics of disease within populations, and efficient control efforts require explicit decisions and the targeted application of resources, both of which are aided by clearly defined categories. I consider how the classification of 3 axes of infectious disease data (risk of exposure, infection status, and accumulation of cases in populations), provides opportunities and challenges to these aims. I illustrate this in the context of two outbreak settings: COVID-19, and measles. Due to the novelty of COVID-19, disease spread rapidly but infection risk groups were not well characterized. For measles, however, drivers of disease transmission are well documented, but the episodic nature of cases combined with imperfect surveillance systems make outbreak detection and response imprecise. In Chapter 2 & Chapter 3, I use empirical data to characterize the emergent risk classes. For Chapter 4 and Chapter 5, I use mathematical models to simulate infections and quantify the outbreak detection and predictive ability of surveillance systems.

The COVID-19 pandemic threw the world into a state of uncertainty, shutting down university campuses, towns, and cities across the United States and wider world. With the return of students to the Pennsylvania State University in the Fall of 2020, two exposure groups emerged: students and the surrounding community members. Contrary to the assumptions at the time that the spatial proximity would overwhelm other drivers of disease transmission, resulting in similar exposure rates between the cohorts, the students were infected at a far higher rate than the county residents (Chapter 2). To evaluate the potential for within-cohort variation in exposure rates, I used clustering techniques on behavioral survey data to uncover latent groups of within the student body, with respect to risk behavior (Chapter 3). Doing so provided a

mechanism to examine the effects and limits of targeted interventions to improve adherence to public health measures. This demonstrated that only 62% of the students could have their risk of exposure reduced, but due to the non-linear nature of outbreaks, a fully effective intervention could reduce total infection rates by up to 93%.

In Chapter 4, I simulate co-circulating pathogens (measles and rubella) to investigate the role of diagnostic uncertainty in case identification and generation of alerts in an outbreak detection system. When the non-target disease does not experience large peaks and troughs in incidence, imperfect diagnostic tests can perform as well in detecting outbreaks as perfect diagnostic tests, particularly if a perfect test is associated with a test result delay. However, the performance of an outbreak detection system based upon incidence estimates from an imperfect diagnostic test degrades substantially with increasing magnitudes of background dynamical noise, like from uncontrolled rubella.

Designing *proactive* rather than *reactive* outbreak detection systems requires the calculation of summary statistics that are correlated with emergence of an outbreak. In Chapter 5, I show that anticipating outbreaks with imperfect diagnostic tests is as context-dependant as reactive systems. When there are low levels of background noise, imperfect diagnostics can provide advanced warning of outbreak emergence. However, when the incidence of measles is small relative to the incidence of rubella, false positive test results dominate the time series used to compute the summary statistics, rendering them unable to discriminate between emergent and non-emergent time periods.

While discretizations are associated with a loss of information, the process of creating categories can often provide necessary context to understanding infectious disease data. Categorization can allow for decision-making in the face of overwhelming volumes of information. Creating dynamically-relevant, explicit, groups is the key to tackling these problems in a transparent and objective fashion.

# Table of Contents

List of Figures.....	ix
List of Tables.....	xiv
Acknowledgements.....	xvii
Chapter 1   Introduction .....	1
Chapter 2   A Longitudinal Study of the Impact of University Student Return to Campus on the SARS-CoV-2 Seroprevalence Among the Community Members .....	7
Abstract .....	7
Background .....	7
Methods .....	7
Results .....	7
Conclusions .....	8
Background .....	9
Methods .....	10
Design, Setting, and Participants .....	10
Outcomes .....	12
Statistical Methods .....	12
Results .....	14
Demographics .....	14
Prior Positive Results and Seroprevalence .....	17
Variables Associated with IgG Positivity .....	20
Discussion .....	22
Limitations and Strengths .....	25
Acknowledgements .....	27
Funding .....	27
Author Contributions .....	27
Conflicts of Interest and Financial Disclosures .....	27
Data Access, Responsibility, and Analysis .....	27
Collaborators .....	28
Data Availability .....	28
Notes .....	28
Chapter 3   The Maximal Expected Benefit of SARS-CoV-2 Interventions Among University Students: A Simulation Study Using Latent Class Analysis .....	29
Abstract .....	29
Background .....	30
Methods .....	32
Design, Setting, and Participants .....	32

Outcomes .....	33
Statistical Methods .....	33
Results .....	36
Demographics .....	36
LCA Fitting .....	36
Compartmental Model .....	41
Discussion .....	43
Limitations and Strengths .....	45
Acknowledgements .....	47
Funding .....	47
Author Contributions .....	47
Conflicts of Interest and Financial Disclosures .....	47
Data Access, Responsibility, and Analysis .....	47
Collaborators .....	48
Data Availability .....	48
 Chapter 4   Individual and Population Level Uncertainty Interact to Determine Performance of Outbreak Surveillance Systems .....	49
Abstract .....	49
Background .....	49
Methods .....	49
Results .....	50
Conclusions .....	50
Background .....	51
Methods .....	54
Model Structure .....	54
Defining Outbreaks .....	56
Triggering Alerts .....	58
Results .....	60
Discussion .....	66
Limitations and Strengths .....	68
Acknowledgements .....	71
Funding .....	71
Author Contributions .....	71
Conflicts of Interest and Financial Disclosures .....	71
Data Access, Responsibility, and Analysis .....	71
Data Availability .....	71
 Chapter 5   Diagnostic Uncertainty Limits the Potential of Early Warning Signals to Identify Epidemic Emergence .....	72
Abstract .....	72

Background .....	73
Materials & Methods .....	75
Model Structure .....	75
Computing & Evaluating EWS .....	78
Results .....	81
Correlation with Emergence .....	81
Predictive Ability .....	84
Discussion .....	86
Acknowledgements .....	91
Funding .....	91
Author Contributions .....	91
Conflicts of Interest and Financial Disclosures .....	91
Data Access, Responsibility, and Analysis .....	91
Data Availability .....	91
 Chapter 6   Synthesis .....	92
Discretization of Risk Groups .....	92
Discretization of Infection Status .....	96
Conclusion .....	102
 Appendix A   Supplementary Material for Chapter 2 .....	103
Laboratory Methods: .....	103
Production and Purification of SARS-CoV-2 Receptor-binding domain (S/RBD)...	
103	
Estimation of IgG antibodies against SARS-CoV-2 .....	103
Statistical Methods .....	104
Treatment of Missing Data .....	104
Alternative Estimate of True Prevalence .....	105
Comparison of Community Member Infections by Similarity To Student Cohort.	
106	
Tables .....	106
Figures .....	109
 Appendix B   Supplementary Material for Chapter 3 .....	111
LCA Model Fitting .....	111
Matrix Structure Sensitivity Analysis .....	112
PHMs Confer Protection .....	112
Eq 1C (Identical Off-Diagonal Values) .....	115

Appendix C   Supplementary Material for Chapter 4 .....	117
Tables .....	117
Figures .....	119
Appendix D   Supplementary Material for Chapter 5 .....	121
Tables .....	121
Figures .....	123
Bibliography .....	132
Section 5   Vita   Callum Arnold .....	151
Education .....	151
Awards .....	151
Select Publications .....	151

## LIST OF FIGURES

Figure 2.1: Map of Centre County, Pennsylvania, USA. Blue indicates the 5 townships and 1 borough that comprise the Centre Region. Red indicates the location of The Pennsylvania State University (PSU), University Park (UP) Campus. Inset illustrates the proportion of the county population in each region; PSU indicates the estimated student population that returned to campus for the Fall 2020 term. ....	17
Figure 2.2: Raw seroprevalence (circles) with 95% binomial confidence intervals for the community residents at the first visit at the start of the Fall 2020 term (light blue), returning students at the end of the fall 2020 term (red), and community residents at either the first or the second visit after student departure (dark blue). ....	19
Figure 2.3: Estimated true prevalence (circles, with 95% confidence intervals) among participants at each sampling interval corrected for estimated assay sensitivity as a function of the assumed assay specificity (x-axis). Light blue indicates community residents at the first visit at the start of the Fall 2020 term, red indicates returning students at the end of the Fall 2020 term, and dark blue indicates community residents at the second visit after student departure. ....	20
Figure 3.1: Distribution of the distance from the ABC fits, with the minimum and maximum distances illustrated by the whiskers, and the median distance by the point. Between-group mixing of 1.0 equates to between-group mixing as likely as within-group mixing .....	42
Figure 3.2: A) The reduction in final infection size across a range of intervention effectiveness (1.0 is a fully effective intervention), accounting for a range of assortativity. Between-group mixing of 1.0 equates to between-group mixing as likely as within-group mixing; B) The relative distribution of group sizes at three levels of intervention effectiveness (0.0, 0.5, 1.0) ....	43
Figure 4.1: A schematic of the outbreak definition and alert detection system. A) Measles incidence time series. B) Noise incidence time series. C) Observed time series of test positive cases according to a given testing scenario. The orange bands present in all 3 panels represent regions of the measles time series that meet the outbreak definition criteria. In panel C, the dark blue bands represent regions of the test positive time series that breach the alert threshold (the horizontal dashed line), and constitute an alert. ...	58
Figure 4.2: The accuracy of outbreak detection systems under different testing rates and noise structures, at their respective optimal alert thresholds. The shaded bands illustrate the 80% central interval, and the solid/dashed lines represent the mean estimate. Imperfect tests have the same values for sensitivity and specificity. Solid lines represent	

tests with 0-day turnaround times, and dashed lines represent tests with result delays.  $\Lambda(4)$  indicates the mean noise incidence is 4 times higher than the mean measles incidence, for example. ..... 62

Figure 4.3: The detection delay of outbreak detection systems under different testing rates and noise structures, at their respective optimal alert thresholds. The shaded bands illustrate the 80% central interval, and the solid/dashed lines represent the mean estimate. Imperfect tests have the same values for sensitivity and specificity. Solid lines represent tests with 0-day turnaround times, and dashed lines represent tests with result delays.  $\Lambda(4)$  indicates the mean noise incidence is 4 times higher than the mean measles incidence. ..... 63

Figure 4.4: The proportion of the time series in alert of outbreak detection systems under different testing rates and noise structures, at their respective optimal alert thresholds. The shaded bands illustrate the 80% central interval, and the solid/dashed lines represent the mean estimate. Imperfect tests have the same values for sensitivity and specificity. Solid lines represent tests with 0-day turnaround times, and dashed lines represent tests with result delays.  $\Lambda(4)$  indicates the mean noise incidence is 4 times higher than the mean measles incidence. ..... 65

Figure 4.5: The number of unavoidable cases of an outbreak detection systems under different testing rates and noise structures, at their respective optimal alert thresholds. The shaded bands illustrate the 80% central interval, and the solid/dashed lines represent the mean estimate. Imperfect tests have the same values for sensitivity and specificity. Solid lines represent tests with 0-day turnaround times, and dashed lines represent tests with result delays.  $\Lambda(4)$  indicates the mean noise incidence is 4 times higher than the mean measles incidence. ..... 66

Figure 5.1: The change in alert accuracy for the most correlated EWS metrics under increasing diagnostic uncertainty, and low and high levels of Poisson or dynamical noise. Low noise refers to simulations where the average incidence of noise is equal to the average incidence of measles. High noise refers to simulations where the average incidence of noise is equal to 7 times the average incidence of measles. The tests sensitivity equals the test specificity for all diagnostic tests. ..... 85

Figure 5.2: Survival curves for the autocovariance EWS metric computed on emergent and null simulations, with a perfect test, and an imperfect test that is 90% sensitive and specific. The histogram depicts the times when the tipping point is reached ( $R_E = 1$ ) under the emergent simulation, right-truncating the curves. The trajectory of the solid lines are identical in each facet, as the perfect test is unaffected by noise cases. The histogram is identical between facets as it represents the timing of the tipping points,

and all testing scenarios use the same underling measles simulations that terminate at the tipping point .....	86
Supplemental Figure A.1: Imputation method comparison among returning students..... 109	
Supplemental Figure A.2: Sensitivity analysis of true prevalence amongst returning student and community subgroups, using pooled estimate of IgG test sensitivity against self-reported prior positive test .....	110
Supplemental Figure B.1: PHMs confer protection to the practitioner. Distribution of the distance from the ABC fits, with the minimum and maximum distances illustrated by the whiskers, and the median distance by the point. Between-group mixing of 1.0 equates to between-group mixing as likely as within-group mixing .....	113
Supplemental Figure B.2: PHMs confer protection to the practitioner. A) The reduction in final infection size across a range of intervention effectiveness (1.0 is a fully effective intervention), accounting for a range of assortativity. Between-group mixing of 1.0 equates to between-group mixing as likely as within-group mixing; B) The relative distribution of group sizes at three levels of intervention effectiveness (0.0, 0.5, 1.0) .	114
Supplemental Figure B.3: Identical off-diagonal values. Distribution of the distance from the ABC fits, with the minimum and maximum distances illustrated by the whiskers, and the median distance by the point. Between-group mixing of 1.0 equates to between-group mixing as likely as within-group mixing .....	115
Supplemental Figure B.4: Identical off-diagonal values. A) The reduction in final infection size across a range of intervention effectiveness (1.0 is a fully effective intervention), accounting for a range of assortativity. Between-group mixing of 1.0 equates to between-group mixing as likely as within-group mixing; B) The relative distribution of group sizes at three levels of intervention effectiveness (0.0, 0.5, 1.0), .	116
Supplemental Figure C.1: The number of alerts of outbreak detection systems under different testing rates and noise structures, at their respective optimal alert thresholds. The shaded bands illustrate the 80% central interval, and the solid/dashed lines represent the mean estimate. Imperfect tests have the same values for sensitivity and specificity. Solid lines represent tests with 0-day turnaround times, and dashed lines represent tests with result delays. $\Lambda(4)$ indicates the mean noise incidence is 4 times higher than the mean measles incidence, for example. ....	119
Supplemental Figure C.2: The duration of alerts of outbreak detection systems under different testing rates and noise structures, at their respective optimal alert thresholds. The shaded bands illustrate the 80% central interval, and the solid/dashed lines	

represent the mean estimate. Imperfect tests have the same values for sensitivity and specificity. Solid lines represent tests with 0-day turnaround times, and dashed lines represent tests with result delays. $\Lambda(4)$ indicates the mean noise incidence is 4 times higher than the mean measles incidence, for example. ....	120
Supplemental Figure D.1: The change in alert accuracy for the least correlated EWS metrics under increasing diagnostic uncertainty, and low and high levels of Poisson or dynamical noise. Low noise refers to simulations where the average incidence of noise is equal to the average incidence of measles. High noise refers to simulations where the average incidence of noise is equal to 7 times the average incidence of measles. The test sensitivity equals the test specificity for all diagnostic tests. ....	123
Supplemental Figure D.2: The strength of the correlation ( $ AUC - 0.5 $ ) for each EWS metric with emergence, at low levels of Poisson noise, for diagnostic tests of varying accuracy, and was computed after the completion of the burn-in period. The test sensitivity equals the test specificity for all diagnostic tests. ....	124
Supplemental Figure D.3: The strength of the correlation ( $ AUC - 0.5 $ ) for each EWS metric with emergence, at high levels of Poisson noise, for diagnostic tests of varying accuracy, and was computed after the completion of the burn-in period. The test sensitivity equals the test specificity for all diagnostic tests. ....	125
Supplemental Figure D.4: The strength of the correlation ( $ AUC - 0.5 $ ) for each EWS metric with emergence, at low levels of dynamical noise, for diagnostic tests of varying accuracy, and was computed after the completion of the burn-in period. The test sensitivity equals the test specificity for all diagnostic tests. ....	126
Supplemental Figure D.5: The strength of the correlation ( $ AUC - 0.5 $ ) for each EWS metric with emergence, at high levels of dynamical noise, for diagnostic tests of varying accuracy, and was computed after the completion of the burn-in period. The test sensitivity equals the test specificity for all diagnostic tests. ....	127
Supplemental Figure D.6: The maximal alert accuracy achieved by each EWS metric under low levels of Poisson noise. Q) refers to the long-running quantile threshold to return a flag, and C) the number of consecutive flags to trigger an alert, that in combination produce the maximal accuracy. S) refers to the resulting specificity of the alert system. The test sensitivity equals the test specificity for all diagnostic tests. ....	128
Supplemental Figure D.7: The maximal alert accuracy achieved by each EWS metric under high levels of Poisson noise. Q) refers to the long-running quantile threshold to return a flag, and C) the number of consecutive flags to trigger an alert, that in combination produce the maximal accuracy. S) refers to the resulting specificity of the alert system. The test sensitivity equals the test specificity for all diagnostic tests. ....	129

Supplemental Figure D.8: The maximal alert accuracy achieved by each EWS metric under low levels of dynamical noise. Q) refers to the long-running quantile threshold to return a flag, and C) the number of consecutive flags to trigger an alert, that in combination produce the maximal accuracy. S) refers to the resulting specificity of the alert system. The test sensitivity equals the test specificity for all diagnostic tests. .... 130

Supplemental Figure D.9: The maximal alert accuracy achieved by each EWS metric under high levels of dynamical noise. Q) refers to the long-running quantile threshold to return a flag, and C) the number of consecutive flags to trigger an alert, that in combination produce the maximal accuracy. S) refers to the resulting specificity of the alert system. The test sensitivity equals the test specificity for all diagnostic tests. .... 131

## LIST OF TABLES

Table 2.1: Demographic characteristics of study participants. Non-D4A participants are all participants in the initial anonymous survey from which Data4Action participants were drawn. D4A participants are divided into subsets for which antibody assays were conducted (N=1313) and those for which assays were not conducted (N=218). ....	15
Table 2.2: Demographic characteristics of the returning student participants .....	16
Table 2.3: IgG ELISA results as a function of self-reported prior SARS-CoV-2 diagnostic test outcome among returning student cohort participants .....	18
Table 2.4: Crude and adjusted odds ratios (OR; aOR) of risk factors among returning PSU UP student cohort .....	21
Table 3.1: Participants' intention to always or not always follow 8 public health measures .....	37
Table 3.2: Log likelihood, AIC, and BIC of two to seven class LCA model fits .....	38
Table 3.3: Mean and median AIC and BIC of multiply-imputed logistic regressions for two to seven class LCA models against IgG serostatus .....	38
Table 3.4: Class-conditional item response probabilities shown in the main body of the table for a three-class LCA model, with footers indicating the size of the respective classes, and the class-specific seroprevalence .....	40
Table 3.5: Adjusted odds ratio (aOR) for risk factors of infection among the returning PSU UP student cohort .....	41
Table 4.1: Compartmental model parameters .....	55
Table 4.2: Optimal outbreak alert thresholds for imperfect and perfect diagnostic tests, under dynamical and Poisson noise structures where the average daily noise incidence is 8 times the average daily measles incidence $\Lambda(8)$ . The test sensitivity equals the test specificity for all diagnostic tests. ....	61
Table 5.1: Compartmental model parameters .....	77
Table 5.2: Numerical computations for EWS metrics, where $\delta = 1$ time step, $b = 52$ weeks .....	79

Table 5.3: The ranking and mean value of Kendall's Tau computed on emergent time series, and the $ AUC - 0.5 $ for each metric. The values are computed on the full time series, and the subset from after the completion of the burn-in period, with a perfect test .....	82
Table 5.4: $ AUC - 0.5 $ for EWS metrics, ranked in descending order of magnitude, computed on the subset of the emergent time series after the burn-in period, for a perfect test and an imperfect diagnostic test with 90% sensitivity and 90% specificity, under high and low Poisson and dynamical noise systems .....	83
Supplemental Table A.1: Wave 1 seroprevalence among community cohort members that are similar/not similar in age ( $\leq 30$ ) and household income ( $\leq 50k$ USD p.a.) to returning students. ....	106
Supplemental Table A.2: Wave 2 seroprevalence among community cohort members that are similar/not similar in age ( $\leq 30$ ) and household income ( $\leq 50k$ USD p.a.) to returning students. ....	107
Supplemental Table A.3: Wave 2 cumulative seroprevalence among community cohort members that are similar/not similar in age ( $\leq 30$ ) and household income ( $\leq 50k$ USD p.a.) to returning students. ....	107
Supplemental Table A.4: Propensity of following public health measures in returning students and community members with PSU ELISA results; subset of community members that received the "Health Messaging" survey. P-value refers to Chi-square test with Yates' continuity correction of proportions in the predictor level by cohort. ....	107
Supplemental Table A.5: Raw prevalence in each subgroup by adherence to PH public health .....	108
Supplemental Table B.1: Class-conditional item response probabilities shown in the main body of the table for a four-class LCA model, with footers indicating the size of the respective classes, and the class-specific seroprevalence .....	111
Supplemental Table C.1: Mean outbreak detection accuracy for imperfect and perfect diagnostic tests, at their respective optimal alert thresholds, under dynamical and Poisson noise structures where the average daily noise incidence is 8 times the average daily measles incidence $\Lambda(8)$ . The test sensitivity equals the test specificity for all diagnostic tests. ....	117
Supplemental Table C.2: Mean outbreak detection delays (days) for imperfect and perfect diagnostic tests, at their respective optimal alert thresholds, under dynamical and Poisson noise structures where the average daily noise incidence is 8 times the	

average daily measles incidence $\Lambda(8)$ . The test sensitivity equals the test specificity for all diagnostic tests. ....	118
Supplemental Table C.3: Mean unavoidable cases per annum (scaled to Ghana's 2022 population) for imperfect and perfect diagnostic tests, at their respective optimal alert thresholds, under dynamical and Poisson noise structures where the average daily noise incidence is 8 times the average daily measles incidence $\Lambda(8)$ . The test sensitivity equals the test specificity for all diagnostic tests. ....	118
Supplemental Table D.1: The ranking and mean value of Kendall's Tau computed on the subset of the emergent time series after the burn-in period, for a perfect test and an imperfect test with sensitivity and specificity equal to 90%, under high and low Poisson and dynamical noise systems .....	121
Supplemental Table D.2: The ranking and alert accuracy of the EWS-based alert system computed on the subset of the emergent time series after the burn-in period, for a perfect test and an imperfect test with sensitivity and specificity equal to 90%, under high and low Poisson and dynamical noise systems .....	122

## ACKNOWLEDGEMENTS

To begin, I would like to thank my committee, Kat Shea, Maciej Boni, Stephen Berg, and Matt Ferrari, for their support, patience, and guidance throughout my PhD. I am so very grateful to have had the opportunity to work and discuss science with you all; my work is immeasurably better as a result. In particular, I would like to pay tribute to Matt. Not only have you helped guide me through all academic puzzles an aspiring researcher could ask for, facilitated numerous collaboration opportunities, and been tireless in helping me become a better scientist, but you are one of the few that pair it with a deep-seated kindness and humanity. I can truly say that the last four years would not have been as productive or joyous without your care and understanding; always providing time and support when I needed it.

I would like to thank the many excellent scientists I had the good fortune of collaborating with. Catherine Herzog, Vivek Kapur, Andrew Read, Rachel Smith, Sreenidhi Srinivasan, and the rest of the Data4Action group for their assistance and comments throughout the first two chapters of this dissertation. Birgit Nikolay, Thomas Roederer, Etienne Gignoux, Catherine Eisenhauer, and Klaudia Porten at Médecins Sans Frontières for providing mentorship during a wonderful few weeks in Paris, and the many months that followed. Amy Winter, Bill Moss, Bryan Patenaude, and Alex Kong for irreplaceable feedback in the development of the outbreak detection project.

I am also incredibly grateful to the whole of the Ferrari lab: Deepit Bhatia, Ariel Greiner, Brian Lambert, Tiffany Leung, Candace Todd, Kai Zou, and honorary member, Emily Howerton. Thank you for acting as a sounding board and providing constructive criticism; my work is notably better for your engagement and strength to listen to me repeat the same presentation multiple meetings in a row.

To the friends throughout my time: John, Suzana, Alexa, Kate, Des, Deepit, Emily, Chiara, Alex, Jared, Ryan, Scott, Betty, Dave, Jen. Thank you for joining me on various outdoor excursions, or just grabbing a coffee or something to eat, and helping make both State College and Toronto wonderful places to live for the past 4 years.

And last, but certainly not least, I would like to thank my family for their unwavering support. I wouldn't be here without the love of my parents, Cindy and Chris, my sister, Madeleine, and my aunt, Nikki. As with everything, your care rendered any challenge manageable, providing necessary perspective when I had a tendency to get in my own head. My greatest champions, who all helped nurture a curiosity and interest in science, my grand-parents: Fiona, Richard, Jean, and Harry. My in-laws, Bruno and Christine, could not have been bigger cheerleaders from before the PhD was even a possibility. Words cannot describe my appreciation for the generosity offered during every step of the last 8 years. And finally, I would like to thank my wife, Caroline. You are truly my better half. Deciding to attend that one, and only, early-morning triathlon training session was by far the best decision I ever made. I feel so privileged to have been your companion through the mundane and exciting, witness your grace and strength during trying times, and I can't see what adventures lie ahead for us.

The work detailed in this dissertation was generously supported by: the Office of the Provost and the Clinical Translational Science Institute, Huck Life Sciences Institute, and the Social Science Research Institutes at the Pennsylvania State University; the National Center for Advancing Translational Sciences, National Institutes of Health; Gavi, the Vaccine Alliance; the Bill & Melinda Gates Foundation; the National Science Foundation; and the University Graduate Fellowship at the Pennsylvania State University. Findings and conclusions do not necessarily reflect the view of the funding agencies.

What a useful thing a pocket-map is!" I remarked.

"That's another thing we've learned from your Nation," said Mein Herr, "map-making. But we've carried it much further than you. What do you consider the largest map that would be really useful?"

"About six inches to the mile."

"Only six inches!" exclaimed Mein Herr. "We very soon got to six yards to the mile. Then we tried a hundred yards to the mile. And then came the grandest idea of all ! We actually made a map of the country, on the scale of a mile to the mile!"

"Have you used it much?" I enquired. "It has never been spread out, yet," said Mein Herr: "the farmers objected: they said it would cover the whole country, and shut out the sunlight ! So we now use the country itself, as its own map, and I assure you it does nearly as well."

— *Sylvie and Bruno Concluded, Chapter XI*, Lewis Carroll

# Chapter 1 | Introduction

Much of the world is complex, and in seeking to understand it we must often make simplifying aggregations and assumptions, often to create groups [1–3]. When quantifying a student’s academic achievements, or measuring distances, for example, decisions are made to create discrete categories from the underlying continuous data. Categorization is essential to the synthesis and interpretation of information, particularly for action; grades are utilized as it would be impractical to simultaneously evaluate every score a student achieved during school. Infectious diseases are no different. Rather than tracking pathogen loads in a population, for example, individuals are typically categorized as either susceptible, exposed/latent, infected, or removed [4,5]. Cases are counted, providing estimates of the number of new and cumulative infected individuals at any one time; labeling individuals as infected is integral to calculation of disease burden [6], and the equitable and efficient allocation of vaccinations [7]. This approach, however, leaves us with some questions; namely, how many groups are appropriate, how should the breakpoints between groups be defined, and are there meaningful differences between the groups that allow for inferences about the system in question? At every scale in an infectious disease system, from variability of infectivity within an individual’s infection cycle, to defining outbreaks in a population from the accumulation of infections, these questions must be addressed. In this dissertation I explore how variability in continuous measures can be discretized, and the interactions that arise from the compounding uncertainty of these categorization decisions.

The use of discretization is prevalent throughout the history of infectious disease biology and epidemiology; to produce more accurate mathematical models of disease transmission, detailed maps of contact rates have been generated that categorize a

population into age groups and log the relative frequency of the interactions within and between these groups [8–10]. Extending this work, networks of individuals and animals have been created to provide a more granular picture of interactions that can exist within an infectious disease system (human or animal), characterizing nodes, vertices, and edges to represent individuals, locations, movement patterns, and connection [11,12]. Furthermore, there is variation that exists between the infectious properties of a population. The (basic) reproduction number is defined as the expected number of secondary cases generated from one infectious individual (in a completely susceptible population) [4]. While often used to describe the transmissibility of a disease within a population, it is a discretization of the underlying, continuous, infectiousness associated with each individual. As a consequence of variability in the pathogen load in an individual, the number of their contacts (and the susceptibility of those contacts), and the length of their infectious period, there is a distribution of secondary infection numbers that exists, and the reproduction number is the mean of this distribution. Individuals at the tail of the distribution can be categorized as super-spreaders and super-shedders [13], providing additional insight when compiling summary information about the spread of disease. Descriptions of heterogeneity also aid in the design of targeted interventions. By identifying the most at-risk groups in a population, limited resources can be most appropriately allocated [14]. For example, accounting for injection drug use status can help target interventions to reduce human immunodeficiency virus (HIV) and Hepatitis C Virus incidence [15].

Throughout my dissertation I specifically address the discretization of 3 categories of continuous phenomena: risk classes, infection status, and outbreaks. Chapter 2 & Chapter 3 address the definition of exposure and risk classes in the context of the COVID-19 pandemic and the implications for intervention effectiveness. In Chapter 4 & Chapter 5 I examine the impact of uncertainty in the binary classification of an

individual's infection status via diagnostics on the determination of outbreak status, and consequentially, outbreak response and prediction.

In the first half of my dissertation (Chapter 2 & Chapter 3 ), I explore how differences in infection rates between geographically co-incident groups can be evaluated in the context of the categorization process. In the spring of 2020, the COVID-19 pandemic resulted in many university campuses across the US shutting down, requiring their students to return to their respective homes [16]. When students were re-introduced to the Pennsylvania State University campus during the start of the Fall 2020 semester, two spatially entwined but demographically and behaviorally disparate groups were defined: returning students and the surrounding community members. Through this grouping, it is possible to characterize the burden of SARS-CoV-2 infection (the underlying virus that causes the disease COVID-19). Without discrete categories, there is no denominator for use in calculations of seroprevalence (the proportion of a population that have sufficiently high levels of antibodies, indicating past exposure to a pathogen); a key metric in the evaluation of disease burden. In Chapter 2 I show that substantial, unexpected, differences in infection rates can be observed between the student and community populations, highlighting that opportunities exist for infection mitigation efforts to minimize spread between spatially-linked subgroups of a population. To examine differences in COVID-19 infections that may exist in the student body, it was, once again, imperative to define groups to compare. However, with no clear differences in traditional demographic measures that could be used to categorize individuals, such as age, I use Latent Class Analysis (LCA) to define these group from behavioral survey data [17,18]. The process of discovering categories with unsupervised clustering methods provides a mechanism to quantify the variation in risk perception and behavior, that cannot be directly measured. In Chapter 3, I map the association between these emergent risk groups with infection rates from serological data to parameterize a mechanistic model of infection,

and demonstrate the limits of solely using non-pharmaceutical interventions to reduce infections within the student population.

In the second half of my dissertation (Chapter 4 & Chapter 5), I examine the necessity and implications of categorizations for action in regions with persistent and emerging infection dynamics. Infectious disease surveillance has 3 primary objectives: to observe and quantify the burden of disease, monitor trends in prevalence, and detect and inform response to outbreaks [19,20]. In pursuit of these goals, numerous continuous values must be discretized. Firstly, cases must be counted, which requires a set of criteria to convert the underlying infection dynamics within an individual into a binary status: infected or not. This criteria often comes in the form of a diagnostic test, like an enzyme linked immunosorbent assay (ELISA). ELISAs measure the presence and quantity of antibodies in a biological sample that are produced by a person's immune system in response to pathogen exposure, and attempts to discriminate between two hypothetical infection/exposure states [21]. In practice, no threshold will be able to perfectly discriminate between these groups of individuals, leading to classification errors [22]. The sensitivity of a test refers to its ability to correctly detect the presence of infection when an infectious individual is tested, also called the true positive rate [23–25]. The specificity is the opposite: the ability to correctly detect the *lack* of infection in an uninfected individual, also called the true negative rate [23–25]. An important third characteristic of diagnostic tests that arises from the discretization of a continuous measure is the positive predictive value (PPV) of a test. The PPV is the probability that a positive test result actually reflects a positive individual [23,24]. Unlike the sensitivity, it is not preconditioned on the assumption that the individual tested is truly positive. The complement to the PPV is the negative predictive value (NPV); the probability that a negative test result accurately reflects reality. When counting for infectious disease surveillance, decisions are made on the basis of these imperfect categorizations.

In my 4th chapter I explore how fallible diagnostic tests interact with non-target background infections (that change the PPV of test results), producing different time series that are used to detect outbreaks. Additionally, the very notion of an outbreak is itself a categorization of a continuous phenomenon, separating a time series of test positive cases by suspected outbreak status, and faces similar issues of sensitivity/specificity/PPV/NPV [26–28]. My work demonstrates how uncertainty that arises at each step of the outbreak detection process must be accounted for, highlighting contexts where different combinations of diagnostic tests and outbreak classification criteria can produce equivalent outbreak detection accuracies.

In the final chapter, I address how these discontinuity errors affect efforts to build *proactive* rather than *reactive* outbreak alert systems. In contrast to traditional outbreak detection systems that require the observation of test positive cases to trigger an alert i.e., respond to the detection of an ongoing outbreak, proactive alert systems have been developed to predict the risk and potential of future outbreaks. Instead of categorizing incidence to define a prediction target, proactive alert systems calculate summary statistics of test positive time series to predict the approach to the *tipping point* of infectious diseases,  $R_E = 1$ .  $R_E$  (the effective reproduction number) is the expected number of secondary infections that would be generated by each infectious individual before they recover (given the current population size and susceptibility). Values greater than or equal to 1 indicate transmission would be self-sustaining if a population is seeded with infection(s). Predicting the approach to this tipping point would provide advance warning of potential outbreaks, allowing proactive decisions to be made. I show that when imperfect diagnostic tests are utilized to create the underlying summary statistics, much like reactive outbreak detection systems, the alert performance is heavily influenced by the shape and magnitude of the non-target background infections. Addressing the context explicitly when designing a reactive or proactive outbreak surveillance system allows public health officials and policy-makers

to account for the compounding layers of uncertainty, finding zones of alert accuracy equivalence where particular objectives can be given greater prioritization e.g., speed of response vs. the number of false alerts.

When evaluated in its entirety, my dissertation provides a clear and principled approach to evaluating the effects of categorizing continuous infectious disease data. I demonstrate that through acknowledging the imperfect nature of discretization, it is possible to identify meaningfully different clusters of individuals and outcomes that can inform our understanding of the populations most at risk of infection, and how outbreak surveillance systems can be designed to best address context-specific priorities.

# **Chapter 2 | A Longitudinal Study of the Impact of University Student Return to Campus on the SARS-CoV-2 Seroprevalence Among the Community Members**

Callum R.K. Arnold<sup>1,2</sup>, Sreenidhi Srinivasan<sup>2,7</sup>, Sophie Rodriguez<sup>7</sup>, Natalie Ryzdak<sup>8</sup>, Catherine M. Herzog<sup>2,7</sup>, Abhinay Gontu<sup>8</sup>, Nita Bharti<sup>1,2</sup>, Meg Small<sup>4,6</sup>, Connie J. Rogers<sup>5</sup>, Margeaux M. Schade<sup>6</sup>, Suresh V. Kuchipudi<sup>2,8</sup>, Vivek Kapur<sup>2,7,9</sup>, Andrew F. Read<sup>1,2,7</sup>, Matthew J. Ferrari<sup>1,2</sup>

<sup>1</sup> Department of Biology, Pennsylvania State University, University Park, PA, USA 16802

<sup>2</sup> Center for Infectious Disease Dynamics, Pennsylvania State University, University Park, PA, USA 16802

<sup>3</sup> Ross & Carole Nese College of Nursing, Pennsylvania State University, University Park, PA, USA 16802

<sup>4</sup> College of Health and Human Development, Pennsylvania State University, University Park, PA, USA 16802

<sup>5</sup> Department of Nutritional Sciences, Pennsylvania State University, University Park, PA, USA 16802

<sup>6</sup> Social Science Research Institute, Pennsylvania State University, University Park, PA, USA 16802

<sup>7</sup> Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA, USA 16802

<sup>8</sup> Department of Veterinary and Biomedical Sciences, Pennsylvania State University, University Park, PA, USA 16802

<sup>9</sup> Department of Animal Science, Pennsylvania State University, University Park, PA, USA 16802

## **Abstract**

### **Background**

Returning university students represent large-scale, transient demographic shifts and a potential source of transmission to adjacent communities during the COVID-19 pandemic.

### **Methods**

In this prospective longitudinal cohort study, we tested for IgG antibodies against SARS-CoV-2 in a non-random cohort of residents living in Centre County prior to the Fall 2020 term at the Pennsylvania State University and following the conclusion of the Fall 2020 term. We also report the seroprevalence in a non-random cohort of students collected at the end of the Fall 2020 term.

### **Results**

Of 1313 community participants, 42 (3.2%) were positive for SARS-CoV-2 IgG antibodies at their first visit between 07 August and 02 October 2020. Of 684 student participants who returned to campus for fall instruction, 208 (30.4%) were positive for SARS-CoV-2 antibodies between 26 October and 21 December. 96 (7.3%) community participants returned a positive IgG antibody result by 19 February. Only contact with known SARS-CoV-2-positive individuals and attendance at small gatherings (20-50 individuals) were

significant predictors of detecting IgG antibodies among returning students (aOR, 95% CI: 3.1, 2.07-4.64; 1.52, 1.03-2.24; respectively).

### **Conclusions**

Despite high seroprevalence observed within the student population, seroprevalence in a longitudinal cohort of community residents was low and stable from before student arrival for the Fall 2020 term to after student departure. The study implies that heterogeneity in SARS-CoV-2 transmission can occur in geographically coincident populations.

**Key words:** Latent Class Analysis; SIR Model; Approximate Bayesian Computation; Behavioral Survey; IgG Serosurvey.

## Background

Demographic shifts, high population densities, and population mobility are known to impact the spread of infectious diseases [29–33]. While this has been well characterized at large scales [34–36], it has proved more challenging to demonstrate at smaller geographic scales [37–39]. The return of college and university students to in-person and hybrid (in-person and online) instruction in the Fall 2020 term during the COVID-19 pandemic represented a massive demographic shift in many communities in the United States (US); specifically, increased total population and proportion living in high density living facilities, with a concomitant increase in person-to-person interactions [40]. This shift had the potential to increase SARS-CoV-2 transmission in returning students and to surrounding communities, particularly for non-urban campuses where incidence lagged larger population centers [41]. Modeling analyses conducted prior to student return raised concerns that university re-opening would result in significant SARS-CoV-2 transmission in both the returning student and community resident populations [42,43].

During the Fall 2020 term, many universities in the US experienced high rates of COVID-19 cases among students [44], with a 56% increase in incidence among counties home to large colleges or universities relative to matched counties without such institutions [40]. While there is strong evidence of high incidence rates associated with a return to campus at US colleges and universities [40], the increase in risk in surrounding communities, and transmission rate from campuses to communities, have been less well characterized. The observed increases in COVID-19 cases in these communities cannot be explicitly attributed to campus origin, absent detailed contact tracing.

This investigation reports the initial results of a longitudinal serosurvey of community residents in Centre County, Pennsylvania, USA, which is home to The Pennsylvania State University (PSU), University Park (UP) campus. The return of approximately 35,000 students to the UP campus in August 2020 represented a nearly 20% increase in the county population [45]. During the Fall 2020 term, more than 4,500 cases of SARS-CoV-2 infections were detected among the student population [46]. Between 7 August and 2 October 2020 (before and just after student return), we enrolled a cohort of community residents and tested serum for the presence of anti-Spike Receptor Binding Domain (S/RBD) IgG, which would indicate prior SARS-CoV-2 exposure [47]. This was repeated in the same cohort during December 2020 (post-departure of students), and we present seroprevalence for both sampling waves. Additionally, returning students were enrolled in a longitudinal cohort, and IgG seroprevalence results are presented from the first wave of sampling (between October and November 2020, prior to the end of the term). The hypothesis tested was that following the return of the students for the Fall 2020 term, the community and student cohorts would experience similarly elevated seroprevalence levels relative to the initial community seroprevalence.

## Methods

### Design, Setting, and Participants

This human subjects research was conducted with PSU Institutional Review Board approval and in accordance with the Declaration of Helsinki. The study uses a longitudinal cohort design, with two separate cohorts: community residents and returning students. We report on measures from the first two clinic visits for the community resident cohort and the first clinic visit for the returning student cohort.

To assist with recruitment into studies under the Data4Action (D4A) Centre County COVID Cohort Study umbrella, a REDCap survey was distributed to residents of Centre County where respondents could indicate interest in future study participation and provide demographic data. Returning students received a similar survey and were also recruited through cold-emails and word-of-mouth.

Individuals were eligible for participation in the community resident cohort if they were:  $\geq$  18 years old, residing in Centre County at the time of recruitment (June through September 2020); expecting to reside in Centre County until June 2021; fluent in English; and capable of providing their own consent. PSU students who remained in Centre County through spring and summer university closure were eligible for inclusion in the community resident cohort as they experienced similar geographic COVID-19 risks as community residents. Participants were eligible for inclusion in the returning student cohort if they were:  $\geq$  18 years old; fluent in English; capable of providing their own consent; residing in Centre County at the time of recruitment (October 2020); officially enrolled as PSU UP students for the Fall 2020 term; and intended to be living in Centre County through April 2021. In both cohorts, individuals were invited to participate in the survey-only portion of the study if they were: lactating, pregnant, or intended to become pregnant in the next 12 months; unable to wear a mask for the clinic visit; demonstrated acute COVID-19 symptoms within the previous 14 days; or reported a health condition that made them uncomfortable with participating in the clinic visit. Informed consent was obtained for all participants.

Upon enrollment, returning students were supplied with a REDCap survey to examine socio-behavioral phenomena, such as attendance at gatherings and adherence to non-pharmaceutical interventions, in addition to information pertaining to their travel history and contact with individuals who were known or suspected of being positive for SARS-CoV-2. Community residents received similar surveys with questions

relating to potential SARS-CoV-2 household exposures. All eligible participants were scheduled for a clinical visit at each time interval where blood samples were collected.

## Outcomes

The primary outcome was the presence of S/RBD IgG antibodies, measured using an indirect isotype-specific (IgG) screening ELISA developed at PSU [48]. An optical density (absorbance at 450 nm) higher than six standard deviations above the mean of 100 pre-SARS-CoV-2 samples collected in November 2019, determined a threshold value of 0.169 for a positive result. Comparison against virus neutralization assays and RT-PCR returned sensitivities of 98% and 90%, and specificities of 96% and 100%, respectively [49]. Further details in the supplement (Appendix A). The presence of anti-SARS-CoV-2 antibodies has been documented in prior seroprevalence studies as a method of quantifying cumulative exposure [50–52].

## Statistical Methods

Community resident and returning student cohorts' seroprevalence are presented with binomial 95% confidence intervals. We estimated each subgroup's true prevalence, accounting for imperfect sensitivity and specificity of the IgG assay, using the prevalence package in R [53]. We calculated a 95% binomial confidence interval for test sensitivity of the IgG assay for detecting prior self-reported positive tests in the returning student cohort (students had high access to testing from a common University provider) with a uniform prior distribution between these limits. Prevalence estimates were then calculated across all possible values of specificity between 0.85 and 0.99. Estimates were not corrected for demographics as participants were not enrolled using a probability-based sample. We assessed demographic characteristics of the tested

participants relative to all study participants to illustrate potential selection biases (Table 2.1).

Missing values were deemed “Missing At Random” and imputed, as described in the supplement (Appendix A). We estimated the adjusted odds ratios (aOR) of IgG positivity in the student subgroup using multivariable logistic regression implemented with the mice and finalfit packages [54,55], two-sided Chi-squared tests for raw odds ratios (OR), and Welch Two Sample t-test for continuous distributions, and present 95% confidence intervals. We considered the following variables *a priori* to be potential risk factors as they increase contact with individuals outside of a participants’ household [56–59]: close proximity (6 feet or less) to an individual who tested positive for SARS-CoV-2; close proximity to an individual showing key COVID-19 symptoms (fever, cough, shortness of breath); attendance at a small gathering (20-50 people) in the past 3 months; attendance at a medium gathering (51-1000 people) in the past 3 months; lives in University housing; ate in a restaurant in the past 7 days; ate in a dining hall in the past 7 days; only ate in their room/apartment in the past 7 days; travelled in the 3 months prior to returning to campus; and travelled since returning to campus for the Fall term.

We estimated the aOR of IgG positivity at either time point in the returning community subgroup, with the following risk factors determined *a priori* to the study’s inception: being a PSU employee; and the amount of contact with PSU students when “Stay at home” orders are not in place (self-reported on a scale of 1-10). BIC and AIC were used to evaluate the contribution of the variables to the model.

All statistical analyses were conducted using R version 4.2.1 (2022-06-23) [60], with a pipeline created using the targets package [61].

## Results

### Demographics

A total of 9299 community residents were identified through an initial REDCap survey that collected eligibility, demographic, and contact information. 1531 were eligible, indicated willingness to participate, and were enrolled. 1462 completed a first clinic visit between 07 August and 02 October 2020, and 1313 of those completed a second clinic visit between 30 November and 19 February 2020 and for whom both visit 1 and visit 2 samples were analyzed. 1410 returning students were recruited using volunteer sampling and 725 enrolled; of these, 684 completed clinic visits for serum collection between 26 October and 21 December 2020.

Among participants with serum samples: the median age of community residents was 47 years (IQR: 36-58), with 86.5% between the ages 18-65 years, and for the returning students the median age was 20 years (IQR: 19-21), with 99.7% between the ages 18-65 years; 66.9% of the community residents identified as female and 34.6% as male; 81.9% of the community residents identified as white, as did 81.9% of the students. Similar proportions were seen in those enrolled without samples, and among the initial REDCap survey respondents (Table 2.1; Table 2.2). Although all county residents were eligible for participation, 74.9% of community resident participants were from the 5 townships (College, Ferguson, Harris, Half Moon, Patton) and 1 borough (State College) that form the “Centre Region” and account for ≈ 59% of Centre County’s population [45] (Figure 2.1). The median household income group in the community residents providing samples was \$100,000 to \$149,999 USD (IQR: \$50,000 to \$74,999; \$150,000 to \$199,999). The median household income in the county is \$60,403 [45]. 47.4% of the county is female, 87.9% white, and 70.3% are between the ages of 18-65 years old [45]. The study cohort is moderately older and more affluent (in part because of the

exclusion of returning students), and disproportionately female compared to the general Centre County population.

Table 2.1: Demographic characteristics of study participants. Non-D4A participants are all participants in the initial anonymous survey from which Data4Action participants were drawn. D4A participants are divided into subsets for which antibody assays were conducted (N=1313) and those for which assays were not conducted (N=218).

		D4A Participant		
		Assay Subset (N = 1313)	Non-Assay Subset (N = 218)	Non-Participant
Age (years)	Median [IQR]	47.0 [36.0, 58.0]	42.0 [34.0, 60.0]	49.0 [37.0, 60.0]
	Median [Min, Max]	47.0 [19.0, 99.0]	42.0 [18.0, 91.0]	49.0 [18.0, 861]
Race	White	1220 (92.9%)	194 (89.0%)	6206 (79.9%)
	Aggregated Category*	12 (0.9%)	2 (0.9%)	256 (3.3%)
	Listed more than one race or ethnicity	6 (0.5%)	0 (0.0%)	18 (0.2%)
	Missing	5 (5.7%)	22 (10.1%)	1288 (16.6%)
Gender	Female	879 (66.9%)	113 (51.8%)	0 (0.0%)
	Male	424 (32.3%)	54 (24.8%)	0 (0.0%)
	Non-binary/ Transgender/Self-described	10 (0.8%)	1 (0.5%)	0 (0.0%)
	Prefer not to answer	0 (0.0%)	0 (0.0%)	0 (0.0%)
	Missing	0 (0.0%)	50 (22.9%)	7768 (100%)
Household Income (USD)	\$200,000 and over	137 (10.4%)	21 (9.6%)	681 (8.8%)
	\$150,000 to \$199,999	186 (14.2%)	24 (11.0%)	764 (9.8%)
	\$100,000 to \$149,999	348 (26.5%)	54 (24.8%)	1502 (19.3%)
	\$75,000 to \$99,999	179 (13.6%)	31 (14.2%)	1093 (14.1%)
	\$50,000 to \$74,999	175 (13.3%)	27 (12.4%)	957 (12.3%)
	\$25,000 to \$49,999	142 (10.8%)	22 (10.1%)	747 (9.6%)
	Under \$25,000	43 (3.3%)	13 (6.0%)	256 (3.3%)
	Prefer not to answer	102 (7.8%)	26 (11.9%)	799 (10.3%)
	Missing	1 (0.1%)	0 (0.0%)	969 (12.5%)

Table 2.2: Demographic characteristics of the returning student participants

		<b>Assay Subset (N = 684)</b>	<b>Non-Assay Subset (N = 41)</b>
Age (years)	Median [IQR]	20.0 [19.0, 21.0]	20.0 [20.0, 21.0]
	Median [Min, Max]	20.0 [18.0, 67.0]	20.0 [18.0, 32.0]
	Missing	1 (0.1%)	18 (43.9%)
Race	White	560 (81.9%)	27 (65.9%)
	Aggregated Category*	86 (12.6%)	5 (12.2%)
	Listed more than one race	32 (4.7%)	2 (4.9%)
	Missing	6 (0.9%)	7 (17.1%)
Gender	Female	441 (64.5%)	19 (46.3%)
	Male	237 (34.6%)	22 (53.7%)
	Genderqueer/nonconforming/transgender/different identity	5 (0.7%)	0 (0.0%)
	Missing	1 (0.1%)	0 (0.0%)
University Housing	Not Uni housing	501 (73.2%)	27 (65.9%)
	Uni housing	181 (26.5%)	8 (19.5%)
	Missing	2 (0.3%)	6 (14.6%)

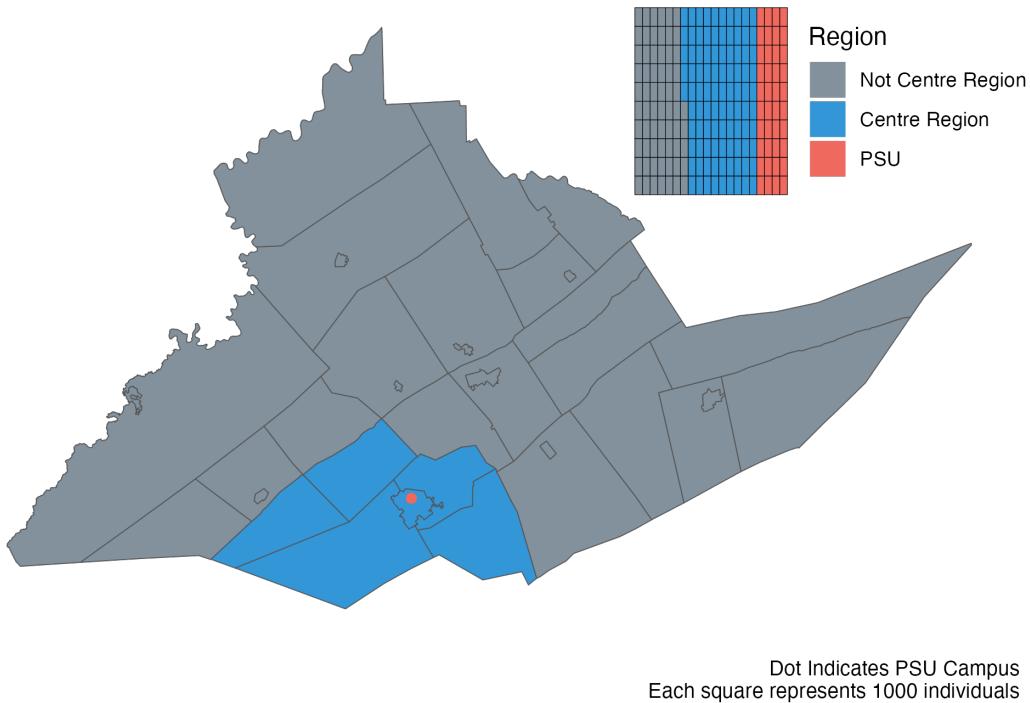


Figure 2.1: Map of Centre County, Pennsylvania, USA. Blue indicates the 5 townships and 1 borough that comprise the Centre Region. Red indicates the location of The Pennsylvania State University (PSU), University Park (UP) Campus. Inset illustrates the proportion of the county population in each region; PSU indicates the estimated student population that returned to campus for the Fall 2020 term.

### Prior Positive Results and Seroprevalence

Of the returning student participants, 673 (92.8%) had at least one test prior to enrollment in the study; of these, 107 (15.9%) self-reported a positive result (Table 2.3). Of these, 100 (93.5%) indicated that this test result occurred after their return to campus (median: 25 September; IQR: 10 September, 07 October). Of the 684 returning students with an ELISA result, 95 of the 102 (93.1%) with a self-reported prior positive test result were positive for SARS-CoV-2 IgG antibodies. Of the 582 returning students with ELISA results who did not report a positive SARS-CoV-2 test, 113 (16.5%) were positive for SARS-CoV-2 IgG antibodies. Of the total 648 returning students with ELISA results, 208

(30.41%) were positive for SARS-CoV-2 IgG antibodies (Figure 2.2). Among the community resident participants, 42 of 1313 (3.2%) were positive for SARS-CoV-2 antibodies at their first visit (Figure 2.2). Between their first and second visit, 54 participants converted from negative to positive and 19 converted from positive to negative; 96 (7.3%) were positive for SARS-CoV-2 IgG antibodies at either visit. There were no differences by age or the number of days separating visit samples, between those that seroconverted and seroreverted ( $p = 0.91$ ;  $p = 0.91$ , respectively). The Wave 1 quantitative OD values of those who seroreverted ( $n = 19$ ) were significantly lower than individuals who remained positive from waves 1 to 2 ( $n = 23$ ) (Welch's t-test,  $p = 0.001$ ; mean of 0.32 vs 0.63). Community residents who were of similar age and household income as the returning students (age  $\leq 30$  years and income  $\leq 50$ k USD) did not have significantly different seroprevalence than community residents age  $\geq 30$  years or with income  $\geq 50$ k USD (Supplemental Table A.1, Supplemental Table A.2, Supplemental Table A.3).

Table 2.3: IgG ELISA results as a function of self-reported prior SARS-CoV-2 diagnostic test outcome among returning student cohort participants

ELISA Result	Prior Test			
	Prior Positive (N = 107)	No Prior Positive (N = 550)	Awaiting Results (N = 16)	No Prior Test
Positive	95 (88.8%)	102 (18.5%)	3 (18.8%)	8 (15.4%)
Negative	7 (6.5%)	419 (76.2%)	13 (81.3%)	37 (71.2%)
Missing	5 (4.7%)	29 (5.3%)	0 (0.0%)	7 (13.5%)

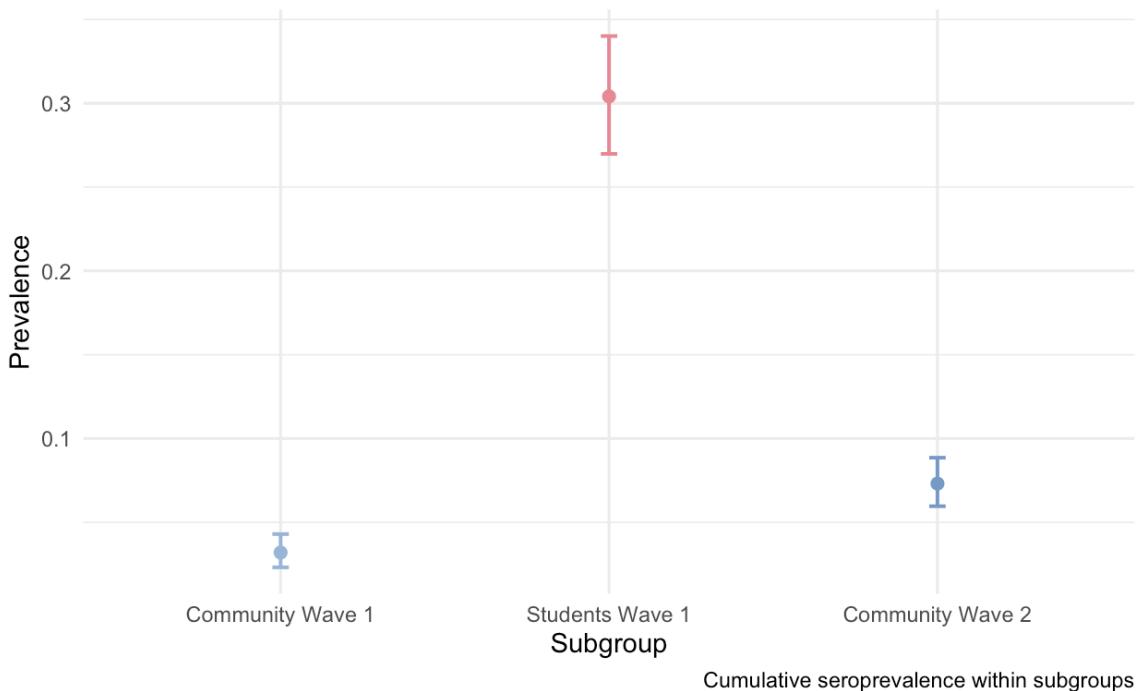


Figure 2.2: Raw seroprevalence (circles) with 95% binomial confidence intervals for the community residents at the first visit at the start of the Fall 2020 term (light blue), returning students at the end of the fall 2020 term (red), and community residents at either the first or the second visit after student departure (dark blue).

Of returning students with a self-reported prior positive SARS-CoV-2 test, 93.1% (95% CI: 86.4-97.2%) had positive IgG antibodies; this was used as an estimate of sensitivity of the IgG assay for detecting previously detectable infection (see Appendix A for an alternative calculation of sensitivity that includes community resident responses). For all values of specificity below 0.95, the 95% credible intervals for the prevalence in the community residents overlapped for the pre- and post-term time points, and neither overlapped with the returning student subgroup (Figure 2.3).

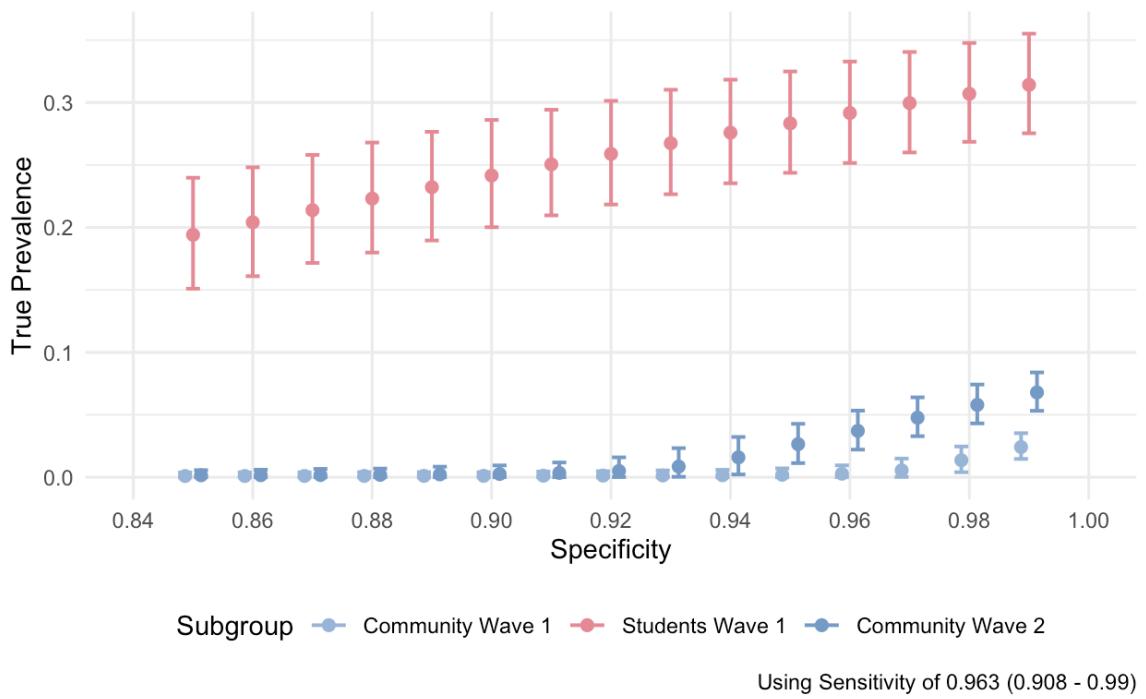


Figure 2.3: Estimated true prevalence (circles, with 95% confidence intervals) among participants at each sampling interval corrected for estimated assay sensitivity as a function of the assumed assay specificity (x-axis). Light blue indicates community residents at the first visit at the start of the Fall 2020 term, red indicates returning students at the end of the Fall 2020 term, and dark blue indicates community residents at the second visit after student departure.

### Variables Associated with IgG Positivity

Among the returning students, only close proximity to a known SARS-CoV-2 positive individual (aOR: 3.09, 2.07-4.62) and attending small gatherings in the past 3 months (aOR: 1.52, 1.03-2.23) were significantly associated with a positive ELISA classification in the multivariable model (Table 2.4). Attending medium gatherings (51-1000 people) (OR: 1.78, 1.17-2.69), and close proximity to an individual showing key COVID-19 symptoms (OR: 1.67, 1.18-2.36) were also associated with the IgG positivity in crude calculations of association. Among the community cohort, the amount of student contact was not associated with cumulative IgG positivity. However, PSU employees experienced reduced odds of positivity (OR: 0.56, 0.35-0.90). Neither AIC or BIC were

improved by the addition of student contact as a variable over employment status only, or using student contact as the only variable.

Table 2.4: Crude and adjusted odds ratios (OR; aOR) of risk factors among returning PSU UP student cohort

Risk Factor	Response	Negative	Positive	OR (univariable)	aOR (multiple imputation)
Close proximity to known COVID-19 Positive Individual	No	277 (58.3%)	61 (29.5%)	-	-
	Yes	198 (41.7%)	146 (70.5%)	3.35 (2.37-4.78, p < 0.001)	3.10 (2.07-4.64, p < 0.001)
Close proximity to individual showing COVID-19 symptoms	No	346 (73.0%)	128 (61.8%)	-	-
	Yes	128 (27.0%)	79 (38.2%)	1.67 (1.18-2.36, p = 0.004)	0.87 (0.58-1.30, p = 0.494)
Travelled in the 3 months prior to campus arrival	No	209 (45.4%)	82 (40.8%)	-	-
	Yes	251 (54.6%)	119 (59.2%)	1.21 (0.86-1.69, p = 0.269)	1.05 (0.73-1.53, p = 0.785)
Travelled since campus arrival	No	183 (38.5%)	82 (39.6%)	-	-
	Yes	292 (61.5%)	125 (60.4%)	0.96 (0.68-1.34, p = 0.789)	0.85 (0.59-1.23, p = 0.394)
Attended a gathering of 20-50 people since arrival for the Fall Semester	No	280 (59.1%)	82 (39.6%)	-	-
	Yes	194 (40.9%)	125 (60.4%)	2.20 (1.58-3.08, p < 0.001)	1.52 (1.03-2.24, p = 0.034)
Attended a gathering of 51-1000 people since arrival for the Fall Semester	No	396 (85.3%)	154 (76.6%)	-	-
	Yes	68 (14.7%)	47 (23.4%)	1.78 (1.17-2.69, p = 0.007)	1.32 (0.83-2.10, p = 0.238)
Ate in a dining hall in the past 7 days	No	394 (83.1%)	163 (79.1%)	-	-
	Yes	80 (16.9%)	43 (20.9%)	1.30 (0.85-1.96, p = 0.214)	1.30 (0.74-2.28, p = 0.356)
Ate in a restaurant in the past 7 days	No	250 (52.5%)	96 (46.8%)	-	-
	Yes	226 (47.5%)	109 (53.2%)	1.26 (0.91-1.75, p = 0.173)	1.12 (0.78-1.61, p = 0.539)
Only ate in their room in the past 7 days	No	158 (33.2%)	76 (36.9%)	-	-
	Yes	318 (66.8%)	130 (63.1%)	0.85 (0.61-1.20, p = 0.350)	0.91 (0.61-1.34, p = 0.625)
Lives in University housing	No	349 (73.5%)	152 (73.4%)	-	-
	Yes	126 (26.5%)	55 (26.6%)	1.00 (0.69-1.45, p = 0.991)	0.89 (0.54-1.45, p = 0.630)

Both the returning students and community residents self-reported high masking compliance; 87.0% and 76.1%, respectively, reported always wearing mask or cloth face covering when in public (Supplemental Table A.4, Supplemental Table A.5). Less than one third of both groups (29.1% and 30.0%, respectively) self-reported always maintaining 6-feet of distance from others in public. Less than half (43.0%) of returning students indicated that they always avoided groups of 25 or greater, in contrast with 65.8% of community residents.

## Discussion

The return of students to in-person instruction on the PSU UP campus was associated with a large increase in COVID-19 incidence in the county, evidenced by over 4,500 student cases at PSU [46]. In a sample of 684 returning students, 30.4% were positive for SARS-CoV-2 antibodies. Out of approximately 35,000 students who returned to campus, this implies that the detected cases may account for  $\approx 40\%$  of all infections among PSU UP students. Despite this high overall incidence of SARS-CoV-2 infection in the county during the Fall 2020 term, the studied cohort of community residents (who disproportionately identified as female and lived in close proximity to campus) saw only a modest increase in the prevalence of SARS-CoV-2 IgG antibodies (3.2 to 7.3%) between September and December 2020; consistent with a nation-wide estimate of seroprevalence for the summer of 2020 [52]. The true prevalence of prior SARS-CoV-2 infection in the cohorts depends on the assumed sensitivity and specificity. However, for most realistic values of sensitivity and specificity, there was little evidence of a significant increase among the community resident sample. Within the community cohort, 19 individuals seroreverted. Given the high specificity of the ELISA, the probability of observing 19 or greater false positives is  $< 0.0001$ , so it is possible that this reflects waning immunity. We note that these 19 individuals had lower OD values

in wave 1 than those that remained positive from wave 1 to wave 2, which is consistent with waning from an initially low antibody titer.

While in-person student instruction has been associated with an increase in per capita COVID-19 incidence [40], these results suggest that outbreaks in the returning student and the community resident cohorts we studied were asynchronous, implying limited between-cohort transmission. A recent analysis of age-specific movement and transmission patterns in the US suggested that individuals between the ages of 20-34 disproportionately contributed to spread of SARS-CoV-2 [62]. Despite close geographic proximity to a college-aged population, transmission in our community resident sample appears distinctly lagged; suggestive of the potential for health behaviors to prevent infection.

Within the student group, presence of SARS-CoV-2 antibodies was significantly associated with close proximity to known SARS-CoV-2-positive individuals and attendance of small events. No other risk factors were correlated with an increase in IgG test positivity, aligning with other research [52]. It is not possible to discern how much the likelihood of contact with a SARS-CoV-2 positive individual is due to the high campus prevalence versus individual behaviors. Considered independently, eating in dining halls within the past 7 days was weakly associated with testing positive for SARS-CoV-2 antibodies, and participation in medium-sized events (51-1000 individuals) and close proximity to a symptomatic individual were significantly associated with testing positive for SARS-CoV-2 antibodies, which is consistent with patterns observed elsewhere [58,59]. Within the community group, being a PSU employee was significantly associated with lower odds of IgG test positivity. There were no significant differences in the age distributions of by employment status. Bharti *et al.* [63] identified lower per-capita incidence in Centre County residents relative to the 5 surrounding counties, as well as a greater movement restriction and less time spent outside the home. Whilst this paper only examined Centre County residents, it is plausible that PSU

employees were more able to work remotely and similarly reduced their movement and non-household contacts, relative to non-PSU employees. The low number of positive community cases meant that it was not possible to identify other associations with IgG positivity.

Though the participants reflect a convenience sample, the large differences in SARS-CoV-2 seroprevalence suggest that the cohorts did not experience a synchronous, well-mixed epidemic despite their close geographic proximity College campuses have been observed to have high COVID-19 attack rates, and counties containing colleges and universities have been observed to have significantly higher COVID-19 incidence than demographically matched counties without such institutions [40]. While college and university operations may present a significant exposure risk, this analysis suggests the possibility that local-scale heterogeneity in mixing may allow for asynchronous transmission dynamics despite close geographic proximity. Thus, the disproportionately high incidence in the student population, which comprises less than one quarter of the county population, may bias assessment of risk in the non-student population. Risk assessment in spatial units (e.g., counties) that have strong population sub-structuring should consider these heterogeneities and their consequences to inform policy.

While SARS-CoV-2 transmission between the student and community resident populations is likely to have occurred (perhaps multiple times), the large difference in seroprevalence between the student and resident participants after the Fall term are consistent with either rare or non-persistent transmission events between the students and residents, or both. This suggests that it is possible to minimize risks brought about by sub-populations with high SARS-CoV-2 incidence using behavioral interventions. This observation may have implications for outbreak management in other high risk, highly mobile populations (e.g., displaced populations, seasonal workers, military deployment). However, we note that this was achieved in the context of disproportionate investment in prevention education, testing, contact tracing, and

infrastructure for isolation and quarantine by PSU in the high-prevalence sub-population (students).

With respect to the health behaviors measured, both students and community residents reported high masking rates ( $> 75\%$ ) and low distancing rates in public ( $< 30\%$ ). However, students had significantly higher masking and gathering rates than community residents, thus a next step is to identify factors that may explain these differences. Minimizing risk, however, may come at significant social, psychological, educational, economic, and societal costs [64]. Thus, operational planning for both institutions of higher education and their resident communities should consider both the risk of SARS-CoV-2 transmission and the costs of mitigation efforts.

## **Limitations and Strengths**

Neither the resident nor the student participants were selected using a probability-based sample. Thus, these participants may not be representative of the populations. Those who chose to participate in this study may have been more cognizant and compliant with public health mitigation measures. Specifically, the resident participants disproportionately lived in the townships immediately surrounding the UP campus, where extensive health messaging [65] and preventative campaigns were enacted, and they have a higher median income than the residents of Centre County overall.

Serotype analysis was not performed, so it may be possible that each sampling time-point reflects the dynamics of different (previous) Variants of Concern (VOCs). However, most samples were provided before VOCs were identified within the United States; Alpha (B.1.1.7) was first identified in Colorado on December 29, 2020, halfway through community wave 2, and Beta (B.1.351) was first identified in South Carolina on January 28, 2021, a few days before the completion of community wave 2 sampling [66].

To our knowledge, this is one of the first studies to explicitly examine the effects of a large and transient student population on the SARS-CoV-2 prevalence of a geographically proximate community population using a longitudinal cohort design. Other studies have observed this influence using a cross-sectional or matched case-control design, but here we present the results of a time-ordered study with large cohort sizes.

## Acknowledgements

### Funding

This work was supported by funding from the Office of the Provost and the Clinical and Translational Science Institute, Huck Life Sciences Institute, and Social Science Research Institutes at the Pennsylvania State University. The project described was supported by the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant UL1 TR002014. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The funding sources had no role in the collection, analysis, interpretation, or writing of the work.

### Author Contributions

*Conceptualization:* MJF, NB, MS, AR, VK

*Data curation:* MJF, CA

*Formal analysis:* CA, MJF, CMH

*Funding acquisition:* MJF, AR

*Investigation:* SS, SR, NR, AG, MMS, CJR

*Methodology:* CA, MJF

*Project administration:* MJF, MMS

*Software:* CA, MJF, CMH

*Supervision:* MJF, VK, SK

*Validation:* CA, MJF, CMH, SS

*Visualization:* CA, MJF

*Writing - original draft:* CA, SS, CMH, VK, MJF

*Writing - review and editing:* all authors.

### Conflicts of Interest and Financial Disclosures

The authors declare no conflicts of interest.

### Data Access, Responsibility, and Analysis

Callum Arnold and Dr. Matthew J. Ferrari had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Callum Arnold, Dr. Matthew J. Ferrari (Department of Biology, Pennsylvania State University), and Dr. Catherine M. Herzog (Huck Institutes of the Life Sciences, Pennsylvania State University) conducted the data analysis.

## Collaborators

1. Florian Krammer, Mount Sinai, USA for generously providing the transfection plasmid pCAGGS-RBD
2. Scott E. Lindner, Allen M. Minns, Randall Rossi produced and purified RBD
3. The D4A Research Group: Dee Bagshaw, Clinical & Translational Science Institute, Cyndi Flanagan, Clinical Research Center and the Clinical & Translational Science Institute, Thomas Gates, Social Science Research Institute, Margeaux Gray, Dept. of Biobehavioral Health, Stephanie Lanza, Dept. of Biobehavioral Health and Prevention Research Center, James Marden, Dept. of Biology and Huck Institutes of the Life Sciences, Susan McHale, Dept. of Human Development and Family Studies and the Social Science Research Institute, Glenda Palmer, Social Science Research Institute, Rachel Smith, Dept. of Communication Arts and Sciences and Huck Institutes of the Life Sciences, and Charima Young, Penn State Office of Government and Community Relations.
4. The authors thank the following for their assistance in the lab: Liz D. Cambron, Elizabeth M. Schwartz, Devin F. Morrison, Julia Fecko, Brian Dawson, Sean Gullette, Sara Neering, Mark Signs, Nigel Deighton, Janhayi Damani, Mario Novelo, Diego Hernandez, Ester Oh, Chauncy Hinshaw, B. Joanne Power, James McGee, Riëtte van Biljon, Andrew Stephenson, Alexis Pino, Nick Heller, Rose Ni, Eleanor Jenkins, Julia Yu, Mackenzie Doyle, Alana Stracuzzi, Brielle Bellow, Abriana Cain, Jaime Farrell, Megan Kostek, Amelia Zazzera, Sara Ann Malinchak, Alex Small, Sam DeMatte, Elizabeth Morrow, Ty Somberger, Haylea Debolt, Kyle Albert, Corey Price, Nazmiye Celik

## Data Availability

The datasets generated during and/or analyzed during the current study are not publicly available due to containing personally identifiable information but are available from the corresponding author on reasonable request.

## Notes

This work has been published in Nature Scientific Reports and is available online at the following link: <https://www.nature.com/articles/s41598-022-12499-5>

Arnold, C.R.K., Srinivasan, S., Rodriguez, S. et al. A longitudinal study of the impact of university student return to campus on the SARS-CoV-2 seroprevalence among the community members. Sci Rep 12, 8586 (2022). <https://doi.org/10.1038/s41598-022-12499-5>

# **Chapter 3 | The Maximal Expected Benefit of SARS-CoV-2 Interventions Among University Students: A Simulation Study Using Latent Class Analysis**

Callum R.K. Arnold<sup>1,2</sup>, Nita Bharti<sup>1,2</sup>, Cara Exten<sup>3</sup>, Meg Small<sup>4,5</sup>, Sreenidhi Srinivasan<sup>2,6</sup>,  
Suresh V. Kuchipudi<sup>2,7</sup>, Vivek Kapur<sup>2,6,8</sup>, Matthew J. Ferrari<sup>1,2</sup>

<sup>1</sup> Department of Biology, Pennsylvania State University, University Park, PA, USA 16802

<sup>2</sup> Center for Infectious Disease Dynamics, Pennsylvania State University, University Park, PA, USA 16802

<sup>3</sup> Ross & Carole Nese College of Nursing, Pennsylvania State University, University Park, PA, USA 16802

<sup>4</sup> College of Health and Human Development, Pennsylvania State University, University Park, PA, USA 16802

<sup>5</sup> Social Science Research Institute, Pennsylvania State University, University Park, PA, USA 16802

<sup>6</sup> Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA, USA 16802

<sup>7</sup> Department of Veterinary and Biomedical Sciences, Pennsylvania State University, University Park, PA, USA 16802

<sup>8</sup> Department of Animal Science, Pennsylvania State University, University Park, PA, USA 16802

## **Abstract**

Non-pharmaceutical public health measures (PHMs) were central to pre-vaccination efforts to reduce Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) exposure risk; heterogeneity in adherence placed bounds on their potential effectiveness, and correlation in their adoption makes assessing the impact attributable to an individual PHM difficult. During the Fall 2020 semester, we used a longitudinal cohort design in a university student population to conduct a behavioral survey of intention to adhere to PHMs, paired with an IgG serosurvey to quantify SARS-CoV-2 exposure at the end of the semester. Using Latent Class Analysis on behavioral survey responses, we identified three distinct groups among the 673 students with IgG samples: 256 (38.04%) students were in the most adherent group, intending to follow all guidelines, 306 (46.21%) in the moderately-adherent group, and 111 (15.75%) in the least-adherent group, rarely intending to follow any measure, with adherence negatively correlated with seropositivity of 25.4%, 32.2% and 37.7%, respectively. Moving all individuals in an SIR model into the most adherent group resulted in a 76-93% reduction in seroprevalence, dependent on assumed assortativity. The potential impact of increasing PHM adherence was limited by the substantial exposure risk in the large proportion of students already following all PHMs.

**Key words:** Latent Class Analysis; SIR Model; Approximate Bayesian Computation; Behavioral Survey; IgG Serosurvey.

## Background

Within epidemiology, the importance of heterogeneity, whether that host, population, statistical, or environmental, has long been recognized [13,67–70]. For example, when designing targeted interventions, it is crucial to understand and account for differences that may exist within populations [71–73]. These differences can present in a variety of forms: heterogeneity in susceptibility, transmission, response to guidance, and treatment effects etc.; all of which affect the dynamics of an infectious disease [67,68,71,74–79]. While heterogeneity may exist on a continuous spectrum, it can be difficult to incorporate into analysis and interpretation, so individuals are often placed in discrete groups according to a characteristic that aims to represent the true differences [8–10,80,81]. When examining optimal influenza vaccination policy in the United Kingdom, Baguelin et al. [82] classified individuals within one of seven age groups. Explicitly accounting for, and grouping, individuals by whether they inject drugs can help target interventions to reduce human immunodeficiency virus (HIV) and Hepatitis C Virus incidence [15]. Similarly, epidemiological models have demonstrated the potential for HIV pre-exposure prophylaxis to reduce racial disparities in HIV incidence [83]. Therefore, heterogeneity can be used to inform more complete theories of change, increasing intervention effectiveness [14]

When discretizing a population for the purposes of inclusion within a mechanistic model, three properties need to be defined: 1) the number of groups, 2) the size of the groups, and 3) the differences between the groups. Typically, as seen in the examples above, demographic data is used e.g., age, sex, race, ethnicity, socio-economic status, etc., often in conjunction with the contact patterns and rates [8,10,72,74,82–84]. There are several reasons for this: the data is widely available, and therefore can be applied almost universally; it is easily understandable; and there are clear demarcations of the groups, addressing properties 1) and 2). However, epidemiological models often aim to

assess the effects of heterogeneity with respect to infection, e.g., “how does an individual’s risk tolerance affect their risk of infection for influenza?”. When addressing questions such as these, demographic data does not necessarily provide a direct link between the discretization method and the heterogeneous nature of the exposure and outcome, particularly if behavioral mechanisms are a potential driver. Instead, it relies on assumptions and proxy measures e.g., an individual’s age approximates their contact rates, which in turn approximates their risk of transmission. This chapter demonstrates an alternative approach to discretizing populations for use within mechanistic models, highlighting the benefits of an interdisciplinary approach to characterize heterogeneity in a manner more closely related to the risk of infection.

In early 2020, shortly after the World Health Organization (WHO) declared the SARS-CoV-2 outbreak a public health emergency of international concern [85], universities across the United States began to close their campuses and accommodations, shifting to remote instruction [86,87]. By Fall 2020, academic institutions transitioned to a hybrid working environment (in-person and online), requiring students to return to campuses [44,88,89]. In the Chapter 2 and accompanying paper [90] we documented the results of a large prospective serosurvey conducted in State College, home to The Pennsylvania State University (PSU) University Park (UP) campus. We examined the effect of 35,000 returning students (representing a nearly 20% increase in the county population [45]) on the community infection rates, testing serum for the presence of anti-Spike Receptor Binding Domain (S/RBD) immunoglobulin G (IgG), indicating prior exposure [47]. Despite widespread concern that campus re-openings would lead to substantial increases in surrounding community infections [88,91,92], very little sustained transmission was observed between the two geographically coincident populations [90].

Given the high infection rate observed among the student body (30.4% seroprevalence), coupled with the substantial heterogeneity in infection rates between

the two populations, we hypothesized that there may be further variation in exposure within the student body, resulting from behavioral heterogeneity. Despite extensive messaging campaigns conducted by the University [65], it is unlikely that all students equally adhered to public health guidance regarding SARS-CoV-2 transmission prevention. We use students' responses to the behavioral survey to determine and classify individuals based on their intention to adhere to public health measures (PHMs). We then show that these latent classes are correlated with SARS-CoV-2 seroprevalence. Finally, we parameterize a mechanistic model of disease transmission within and between these groups, and explore the impact of public health guidance campaigns, such as those conducted at PSU [65]. We show that interventions designed to increase student compliance with PHMs would likely reduce overall transmission, but the relatively high initial compliance limits the scope for improvement via PHM adherence alone.

## Methods

### Design, Setting, and Participants

This research was conducted with PSU Institutional Review Board approval and in accordance with the Declaration of Helsinki, and informed consent was obtained for all participants. The student population has been described in detail previously [90], but in brief, students were eligible for the student cohort if they were:  $\geq 18$  years old; fluent in English; capable of providing their own consent; residing in Centre County at the time of recruitment (October 2020) with the intention to stay through April 2021; and officially enrolled as PSU UP students for the Fall 2020 term. Upon enrollment, students completed a behavioral survey in REDCap [93] to assess adherence and attitudes towards public health guidance, such as attendance at gatherings, travel patterns, and

non-pharmaceutical interventions. Shortly after, they were scheduled for a clinic visit where blood samples were collected. Students were recruited via word-of-mouth and cold-emails.

## Outcomes

The primary outcome was the presence of S/RBD IgG antibodies, measured using an indirect isotype-specific (IgG) screening ELISA developed at PSU [48]. An optical density (absorbance at 450 nm) higher than six standard deviations above the mean of 100 pre-SARS-CoV-2 samples collected in November 2019, determined a threshold value of 0.169 for a positive result. Comparison against virus neutralization assays and RT-PCR returned sensitivities of 98% and 90%, and specificities of 96% and 100%, respectively [49]. Further details in the supplement of the previous chapter (Appendix A).

## Statistical Methods

To identify behavioral risk classes, we fit a range of latent class analysis (LCA) models (two to seven class models) to the student's behavioral survey responses, using the poLCA package [94] in the R programming language, version 4.3.3 (2024-02-29) [60]. We considered their answers regarding the frequency with which they intended to engage in the following behaviors to be *a priori* indicators of behavioral risk tolerance: wash hands with soap and water for at least 20s; wear a mask in public; avoid touching their face with unwashed hands; cover cough and sneeze; stay home when ill; seek medical attention when experiencing symptoms and call in advance; stay at least 6 feet (about 2 arms lengths) from other people when outside of their home; and, stay out of crowded places and avoid mass gatherings of more than 25 people. The behavioral survey collected responses on the Likert scale of: Never, Rarely, Sometimes, Most of the

time, and Always. For all PHMs, Always and Most of the time accounted for > 80% of responses (with the exception of intention to stay out of crowded places and avoid mass gatherings, where Always and Most of the time accounted for 78.8% of responses). To reduce the parameter space of the LCA and minimize overfitting, the behavioral responses were recoded as Always and Not Always. Measures of SARS-CoV-2 exposure e.g., IgG status, were not included in the LCA model fitting, as they reflect the outcome of interest. We focused on responses regarding intention to follow behaviors because this information can be feasibly collected during a public health campaign for a novel or emerging outbreak; it has also been shown that intentions are well-correlated with actual behaviors for coronavirus disease 2019 (COVID-19) public health guidelines, as well as actions that have short-term benefits [95,96]. We examined the latent class models using Bayesian Information Criterion, which is a commonly recommended as part of LCA model evaluation [17,18], to select the model that represented the best balance between parsimony and maximal likelihood fit.

Using the best-fit LCA model, we performed multivariate logistic regression of modal class assignment against IgG seropositivity to assess the association between the latent classes and infection. This “three-step” approach is recommended over the “one-step” LCA model fit that includes the outcome of interest as a covariate in the LCA model [18,97]. The following variables were determined a priori to be potential risk factors for exposure [90]: close proximity (6 feet or less) to an individual who tested positive for SARS-CoV-2; close proximity to an individual showing key COVID-19 symptoms (fever, cough, shortness of breath); lives in University housing; ate in a restaurant in the past 7 days; ate in a dining hall in the past 7 days; only ate in their room/apartment in the past 7 days; travelled in the 3 months prior to returning to campus; and travelled since returning to campus for the Fall term. Variables relating to attending gatherings were not included in the logistic regression due to overlap with intention variables of the initial LCA fit. Missing variables were deemed “Missing At

Random” and imputed using the mice package [55], as described in the supplement of the previous chapter (Appendix A).

We parameterized a deterministic compartmental Susceptible-Infected-Recovered (SIR) model using approximate Bayesian computation (ABC) against the seroprevalence within each latent class. The recovery rate was set to 8 days. Diagonal values of the transmission matrix were constrained such that  $\beta_{HH} \leq \beta_{MM} \leq \beta_{LL}$  ( $H$  represents high-adherence to public health guidelines, and  $M$  and  $L$  represent medium- and low-adherence, respectively), with the following parameters fit: the transmission matrix diagonals, a scaling factor for the off-diagonal values ( $\phi$ ), and a scaling factor for the whole transmission matrix ( $\rho$ ). The off-diagonal values are equal to a within-group value (diagonal) multiplied by a scaling factor ( $\phi$ ). This scaling factor can either multiply the within-group beta value of the source group (e.g.,  $\beta_{HL} = \phi \cdot \beta_{LL}$ ; Equation 3.1), or the recipient group (e.g.,  $\beta_{LH} = \phi \cdot \beta_{LL}$ ; Supplemental Equation B.1), each with a different interpretation.

$$\rho \begin{pmatrix} \beta_{HH} & \beta_{HM} & \beta_{HL} \\ \beta_{MH} & \beta_{MM} & \beta_{ML} \\ \beta_{LH} & \beta_{HM} & \beta_{LL} \end{pmatrix} \rightarrow \rho \begin{pmatrix} \beta_{HH} & \phi\beta_{MM} & \phi\beta_{LL} \\ \phi\beta_{HH} & \beta_{MM} & \phi\beta_{LL} \\ \phi\beta_{HH} & \phi\beta_{MM} & \beta_{LL} \end{pmatrix} \text{ mixing structure} \quad 3.1$$

The former assumes that between-group transmission is dominated by the transmissibility of the source individuals, implying that adherence to the PHMs primarily prevents onwards transmission, rather than protecting the recipient against infection. The latter assumes that between-group transmission is dominated by the susceptibility of the recipient individuals, implying that adherence to the PHMs primarily protects the recipient from infection, rather than preventing onwards transmission. A range of between-group scaling values ( $\phi$ ) were simulated to perform sensitivity analysis for the degree of assortativity. Results are only shown for the matrix structure in Equation 3.1, but results for simulations that imply PHMs prevent

infections through protecting the recipient can be found in the supplement (Supplemental Equation B.1, Supplemental Figure B.1, Supplemental Figure B.2). We also present the results of simulations where all between-group mixing is assumed to be equal (Supplemental Equation B.2, Supplemental Figure B.3, Supplemental Figure B.4). To examine the effect of an intervention to increase PHM adherence, we redistributed a proportion of low- and medium adherence individuals to the high adherence latent class, i.e., a fully effective intervention is equivalent to a single-group SIR model of high adherent individuals. Model fitting and simulation was conducted using the Julia programming language, version 1.10.5 [98].

## Results

### Demographics

Full details can be found in the Chapter 2 and accompanying paper [90], but briefly: 1410 returning students were recruited, 725 were enrolled, and 684 students completed clinic visits for serum collection between 26 October and 21 December 2020. Of these, 673 students also completed the behavioral survey between 23 October and 8 December 2020. The median age of the participants was 20 years (IQR: 19-21), 64.5% identified as female and 34.6% as male, and 81.9% identified as white. A large proportion (30.4%) were positive for IgG antibodies, and 93.5% (100) of the 107 students with a prior positive test reported testing positive only after their return to campus.

### LCA Fitting

Of the 673 participants, most students intended to always mask (81.0%), always cover their coughs/sneezes (81.9%), and always stay home when ill (78.2%) (Table 3.1). Three

of the least common intentions were social distancing by maintaining a distance of at least 6 feet from others outside of their home, avoiding crowded places and mass gatherings > 25 people, and avoiding face-touching with unwashed hands (43.4% and 53.1%, 43.5%, respectively).

Table 3.1: Participants' intention to always or not always follow 8 public health measures

<b>Intention to always:</b>	<b>Always</b>	<b>Not Always</b>
Avoid face-touching with unwashed hands	293 (43.54%)	380 (56.46%)
Cover cough and sneeze	551 (81.87%)	122 (18.13%)
Seek medical attention when have symptoms and call in advance	480 (71.32%)	193 (28.68%)
Stay at least 6 feet (about 2 arms lengths) from other people when outside of home.	292 (43.39%)	381 (56.61%)
Stay home when ill	526 (78.16%)	147 (21.84%)
Stay out of crowded places and avoid mass gatherings > 25 people	357 (53.05%)	316 (46.95%)
Tested for COVID-19 twice or more	544 (80.83%)	129 (19.17%)
Wash hands often with soap and water for at least 20 seconds.	434 (64.49%)	239 (35.51%)
Wear a face cover (mask) in public	545 (80.98%)	128 (19.02%)

The four- and the three-class LCA models had the lowest BIC respectively (Table 3.2). Examining the four-class model, there was minimal difference in the classification of individuals, relative to the three-class model. In the four-class model, the middle class (of the three-class model) was split into two groups with qualitatively similar class-conditional item response probabilities i.e., conditional on class membership, the probability of responding "Always" to a given question, except for hand washing and avoiding face-touching with unwashed hands (Supplemental Table B.1).

Table 3.2: Log likelihood, AIC, and BIC of two to seven class LCA model fits

Classes	Log Likelihood	Akaike Information Criterion	Bayesian Information Criterion
2	-2895.40	5828.81	5914.53
3	-2715.67	5489.35	5620.19
4	-2673.50	5425.00	5600.96
5	-2658.46	5414.93	5636.00
6	-2647.01	5412.03	5678.22
7	-2636.05	5410.10	5721.41

We fit a logistic regression model to predict binary IgG serostatus that included inferred class membership, in addition to other predictor variables we previously identified in [90]. The mean and median BIC and AIC indicated similar predictive ability of the three- and four-class LCA models (Table 3.3). Given these factors, the three-class model was selected for use in simulation for parsimony, requiring fewer assumptions and parameters to fit.

Table 3.3: Mean and median AIC and BIC of multiply-imputed logistic regressions for two to seven class LCA models against IgG serostatus

Classes	AIC (Mean)	BIC (Mean)	AIC (Median)	BIC (Median)
2	794.33	839.44	794.18	839.29
3	794.29	843.92	794.23	843.86
4	797.52	851.66	797.50	851.64
5	799.69	858.34	799.70	858.35
6	796.91	860.08	796.84	860.00
7	794.68	862.36	794.67	862.35

In the three-class model, approximately 15.75% of individuals were members of the group that rarely intended to always follow the PHMs, 38.04% intended to always follow all guidelines, and the remaining 46.21% mostly intended to mask, test, and manage symptoms, but not distance or avoid crowds (Table 3.4). We have labelled the three classes as “Low-”, “High-” and “Medium-Adherence” groups, respectively, for ease

of interpretation. Examining the class-conditional item response probabilities, the Medium Adherence class had a probability of 0.88 of always wearing a mask in public, but a probability of only 0.19 of social distancing when outside of their homes, for example. Calculating the class-specific seroprevalence, the Low Adherence group had the highest infection rates (37.7%, 95% Binomial CI: 28.5-47.7%), the medium adherence the next highest (32.2%, 95% Binomial CI: 27.0-37.7%), and the most adherent group experienced the lowest infection rates (25.4%, 95% Binomial CI: 20.2-31.1%). Incorporating latent class membership into the imputed GLM model described in our previous paper (30) retained the relationship between adherence and infection. Relative to the least adherent group, the Medium Adherence group experienced a non-significant reduction in infection risk (aOR, 95% CI: 0.73, 0.45-1.18), and the most adherent group a significant reduction (aOR, 95% CI: 0.59, 0.36-0.98) (Table 3.5).

Table 3.4: Class-conditional item response probabilities shown in the main body of the table for a three-class LCA model, with footers indicating the size of the respective classes, and the class-specific seroprevalence

<b>Measure Intention to Always:</b>	<b>Low Adherence</b>	<b>Medium Adherence</b>	<b>High Adherence</b>
Wash my hands often with soap and water for at least 20 seconds.	0.04	0.57	0.96
Wear a face cover (mask) in public	0.13	0.88	0.99
Avoid face-touching with unwashed hands	0.00	0.21	0.86
Cover cough and sneeze	0.22	0.86	1.00
Stay home when ill	0.07	0.83	1.00
Seek medical attention when have symptoms and call in advance	0.03	0.70	0.98
Stay at least 6 feet (about 2 arms lengths) from other people when outside of my home.	0.00	0.19	0.87
Stay out of crowded places and avoid mass gatherings > 25 people	0.03	0.39	0.88
Tested for COVID-19 twice or more	0.76	0.82	0.81
<b>Group Size</b>	15.75%	46.21%	38.04%
<b>Seroprevalence</b>	37.7%	32.2%	25.4%

Table 3.5: Adjusted odds ratio (aOR) for risk factors of infection among the returning PSU UP student cohort

Covariate (response) / reference levels	aOR (multiple imputation)
Close proximity to known COVID-19 positive individual (yes) / no	3.41 (2.29-5.08, p<0.001)
Close proximity to individual showing COVID-19 symptoms (yes) / no	0.86 (0.58-1.29, p=0.474)
Lives in University housing (yes) / no	0.90 (0.55-1.47, p=0.685)
Latent Class (medium adherence) / low adherence	0.73 (0.45-1.18, p=0.203)
Latent Class (high adherence) / low adherence	0.59 (0.36-0.98, p=0.043)
Travelled in the 3 months prior to campus arrival (yes) / no	1.12 (0.76-1.63, p=0.57)
Travelled since campus arrival (yes) / no	0.87 (0.6-1.25, p=0.447)
Ate in a dining hall in the past 7 days (yes) / no	1.32 (0.76-2.29, p=0.332)
Ate in a restaurant in the past 7 days (yes) / no	1.14 (0.8-1.64, p=0.465)
Only ate in their room in the past 7 days (yes) / no	0.87 (0.59-1.29, p=0.499)

### Compartmental Model

The ABC distance distributions indicated that near-homogeneous levels of between-group mixing better fit the data (Figure 3.1). After model parameterization, we examined the effect of increasing adherence to public health guidance. Moving all individuals into the High Adherence class resulted in a 76-93% reduction in final size; when moderate between-group mixing is simulated, a fully effective intervention results in approximately 80% reduction in final seroprevalence, and when between-group mixing is as likely as within-group mixing, a 93% reduction is observed (Figure 3.2).

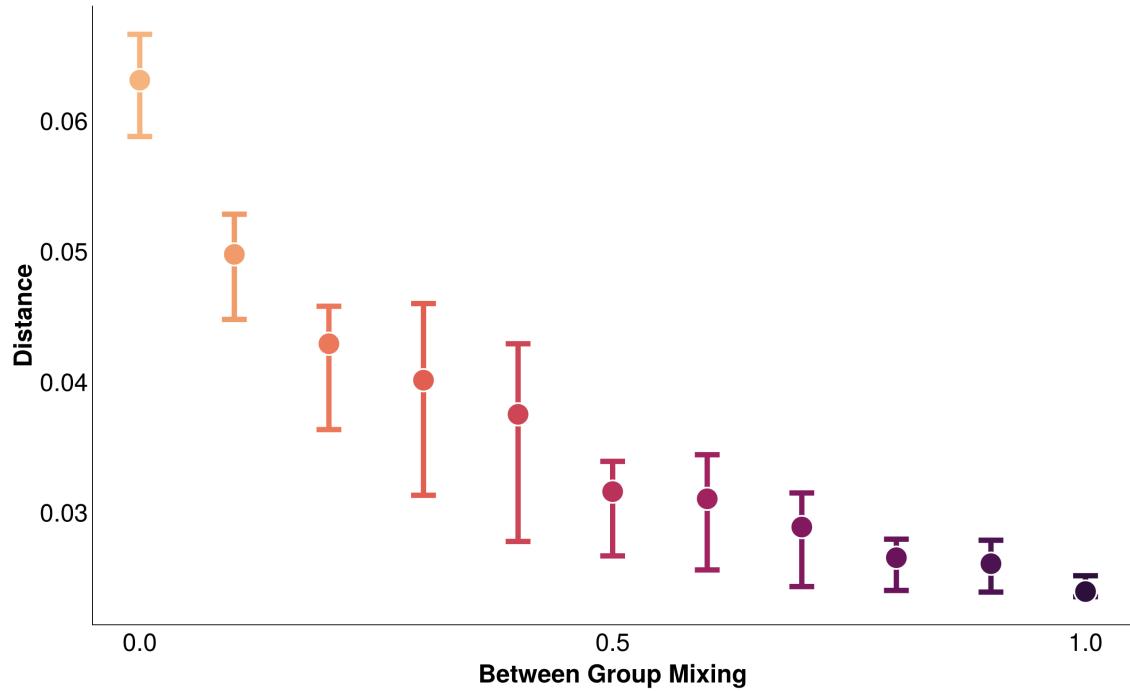


Figure 3.1: Distribution of the distance from the ABC fits, with the minimum and maximum distances illustrated by the whiskers, and the median distance by the point. Between-group mixing of 1.0 equates to between-group mixing as likely as within-group mixing

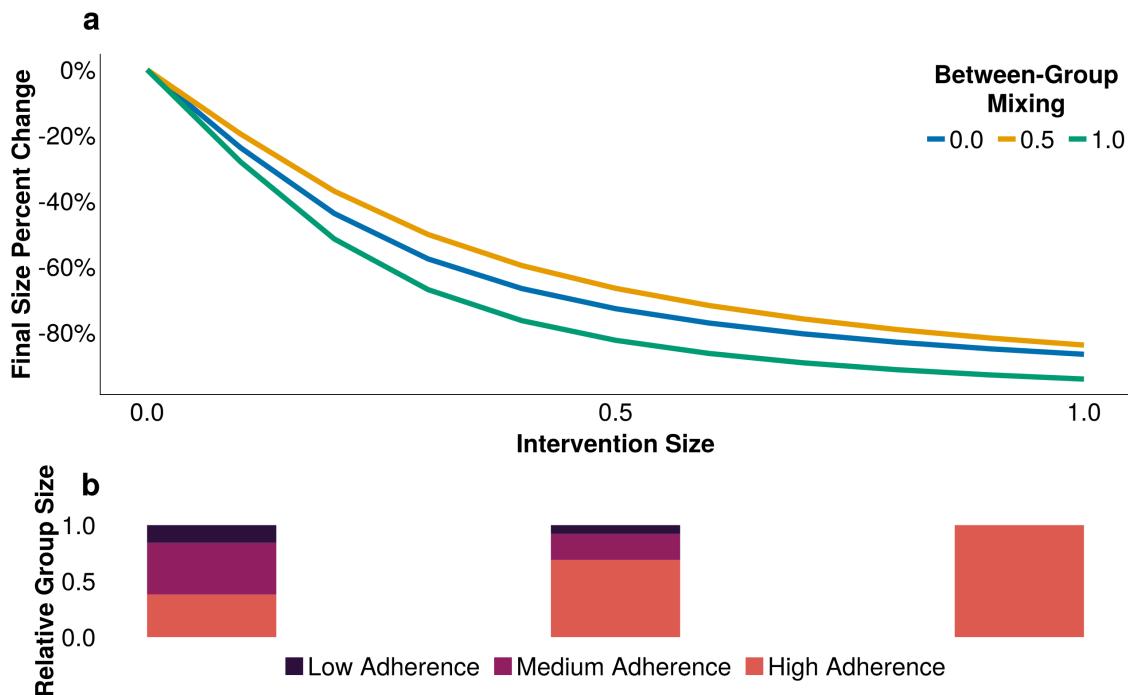


Figure 3.2: A) The reduction in final infection size across a range of intervention effectiveness (1.0 is a fully effective intervention), accounting for a range of assortativity. Between-group mixing of 1.0 equates to between-group mixing as likely as within-group mixing; B) The relative distribution of group sizes at three levels of intervention effectiveness (0.0, 0.5, 1.0)

## Discussion

In this interdisciplinary analysis, we collected behavioral data from surveys and integrated it with serosurveillance results. This approach allowed us to use LCA to categorize a population's transmission potential with measures related to risk tolerance and behavior. The LCA model was fit without inclusion of infection status data, but class membership was correlated with IgG seroprevalence. The classes that were the most adherent to PHMs experienced the lowest infection rates, and the least adherent exhibited the highest seroprevalence.

Although a four-class LCA model was a marginally better fit for the data, there were not substantial differences in class assignment relative to the three-class LCA model. The three-class model was selected for use in simulation for parsimony, requiring fewer assumptions and parameters to fit. Upon parametrizing the compartmental model, smaller ABC distance values were observed for moderate to high levels of between-group mixing, implying some degree of assortativity in our population, though the exact nature cannot be determined from our data. Examining the three classes, 38% of individuals already intended to always follow all PHMs. As a result, only 62% of the study population could have their risk reduced with respect to the PHMs surveyed. Further, the infection rates observed in the High Adherence group indicates that even a perfectly effective intervention aimed at increasing adherence to non-pharmaceutical PHMs (i.e., after the intervention, all individuals always followed every measure) would not eliminate transmission in a population, an observation that aligns with prior COVID-19 research [99–102]. The extent to which the infection in the High Adherence group is a result of mixing with lower adherence classes cannot be explicitly described, but the sensitivity analysis allows for an exploration of the effect and ABC fits suggest near-homogeneous mixing occurred. Varying the structure of the transmission matrix yielded very similar quantitative and qualitative results (Supplemental Figure B.1, Supplemental Figure B.2, Supplemental Figure B.3, Supplemental Figure B.4).

Examining the impact of increasing adherence to PHMs (modeled as increasing the proportion of the population in the High Adherence class), a fully effective intervention saw between a 76–93% reduction in the final size of the simulation outbreak. The small but appreciable dependence of the reduction’s magnitude on the degree of between-group mixing can be explained as such: with higher levels of between-group mixing, the initial SIR parameterization results in lower transmission parameters for the High-High adherence interactions, as more infections in the High Adherence group originate

from interactions with Low and Medium Adherence individuals. Increasing adherence, therefore, results in a greater reduction of the overall transmission rate than in simulations with less assortativity.

## **Limitations and Strengths**

The student population was recruited using convenience sampling, and therefore may not be representative of the wider population. Those participating may have been more cognizant and willing to follow public health guidelines. Similarly, because of the University's extensive messaging campaigns and efforts to increase access to non-pharmaceutical measures [65], such as lateral flow and polymerase-chain reaction diagnostic tests, the students likely had higher adherence rates than would be observed in other populations. However, these limitations are not inherent to the modeling approach laid out, and efforts to minimize them would likely result in stronger associations and conclusions due to larger differences in the latent behavioral classes and resulting group infection rates.

It is well known that classification methods, like LCA, can lead to the “naming fallacy” [17], whereby groups are assigned and then specific causal meaning is given to each cluster, affecting subsequent analyses and interpretation of results. In this chapter, this effect is reduced by virtue of the analysis plan being pre-determined, and the relationship with the outcome showing a positive association with the classes in the mechanistically plausible direction (i.e., increasing adherence to PHMs results in reduced infection rates). Our decision to conduct the simulation analysis with the three-class model was, in part, to avoid the potential bias that would arise from naming or assigning an order to the two intermediate risk groups.

Despite these limitations, this work presents a novel application of a multidisciplinary technique, outlining how alternate data sources can guide future

model parameterization and be incorporated into traditional epidemiological analysis, particularly within demographically homogeneous populations where there is expected or observed heterogeneity in transmission dynamics. This is particularly important in the design of interventions that aim to target individual behaviors, allowing the categorization of populations into dynamically-relevant risk groups and aiding in the efficient use of resources through targeted actions. Future research should consider including perceived agency and efficacy for PHM adherence.

## Acknowledgements

### Funding

This work was supported by funding from the Office of the Provost and the Clinical and Translational Science Institute, Huck Life Sciences Institute, and Social Science Research Institutes at the Pennsylvania State University. The project described was supported by the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant UL1 TR002014. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The funding sources had no role in the collection, analysis, interpretation, or writing of the work.

### Author Contributions

*Conceptualization:* CA, MJF

*Data curation:* CA, MJF

*Formal analysis:* CA

*Funding acquisition:* MJF

*Investigation:* NB, CE, MS, SS, SK, VS

*Methodology:* CA, NB, MJF

*Project administration:* MJF

*Software:* CA

*Supervision:* MJF

*Validation:* CA

*Visualization:* CA

*Writing - original draft:* CA

*Writing - review and editing:* all authors.

### Conflicts of Interest and Financial Disclosures

The authors declare no conflicts of interest.

### Data Access, Responsibility, and Analysis

Callum Arnold and Dr. Matthew J. Ferrari had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Callum Arnold (Department of Biology, Pennsylvania State University) conducted the data analysis.

## Collaborators

1. Florian Krammer, Mount Sinai, USA for generously providing the transfection plasmid pCAGGS-RBD
2. Scott E. Lindner, Allen M. Minns, Randall Rossi produced and purified RBD
3. The D4A Research Group: Dee Bagshaw, Clinical & Translational Science Institute, Cyndi Flanagan, Clinical Research Center and the Clinical & Translational Science Institute; Thomas Gates, Social Science Research Institute; Margeaux Gray, Dept. of Biobehavioral Health; Stephanie Lanza, Dept. of Biobehavioral Health and Prevention Research Center; James Marden, Dept. of Biology and Huck Institutes of the Life Sciences; Susan McHale, Dept. of Human Development and Family Studies and the Social Science Research Institute; Glenda Palmer, Social Science Research Institute; Connie J. Rogers, Dept. of Nutritional Sciences; Rachel Smith, Dept. of Communication Arts and Sciences and Huck Institutes of the Life Sciences; and Charima Young, Penn State Office of Government and Community Relations.
4. The authors thank the following for their assistance in the lab: Sophie Rodriguez, Natalie Rydzak, Liz D. Cambron, Elizabeth M. Schwartz, Devin F. Morrison, Julia Fecko, Brian Dawson, Sean Gullette, Sara Neering, Mark Signs, Nigel Deighton, Janhayi Damani, Mario Novelo, Diego Hernandez, Ester Oh, Chauncy Hinshaw, B. Joanne Power, James McGee, Riëtte van Biljon, Andrew Stephenson, Alexis Pino, Nick Heller, Rose Ni, Eleanor Jenkins, Julia Yu, Mackenzie Doyle, Alana Stracuzzi, Brielle Bellow, Abriana Cain, Jaime Farrell, Megan Kostek, Amelia Zazzera, Sara Ann Malinchak, Alex Small, Sam DeMatte, Elizabeth Morrow, Ty Somberger, Haylea Debolt, Kyle Albert, Corey Price, Nazmiye Celik

## Data Availability

The datasets generated during and/or analyzed during the current study are not publicly available as they contain personally identifiable information, but are available from the corresponding author on reasonable request.

# **Chapter 4 | Individual and Population Level Uncertainty Interact to Determine Performance of Outbreak Surveillance Systems**

Callum R.K. Arnold<sup>1,2</sup>, Alex C. Kong<sup>3</sup>, Amy K. Winter<sup>4,5</sup>, William J. Moss<sup>3,6</sup>, Bryan N. Patenaude<sup>3</sup>, Matthew J. Ferrari<sup>1,2</sup>

<sup>1</sup> Department of Biology, Pennsylvania State University, University Park, PA, USA 16802

<sup>2</sup> Center for Infectious Disease Dynamics, Pennsylvania State University, University Park, PA, USA 16802

<sup>3</sup> Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA 21205

<sup>4</sup> Department of Epidemiology and Biostatistics, College of Public Health, University of Georgia, Athens, GA, USA 30602

<sup>5</sup> Center for the Ecology of Infectious Disease, University of Georgia, Athens, GA, UGA 30602

<sup>6</sup> Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA 21205

## **Abstract**

### **Background**

Infectious disease surveillance and outbreak detection systems often utilize diagnostic testing to validate case identification. The metrics of sensitivity, specificity, and positive predictive value are commonly discussed when evaluating the performance of diagnostic tests, and to a lesser degree, the performance of outbreak detection systems. However, the interaction of the two levels' (the test and the alert system) metrics, is typically overlooked. Here, we describe how equivalent regions of detection accuracy can exist over a range of diagnostic test characteristics, examining the sensitivity to background noise structure and magnitude.

### **Methods**

We generated a stochastic SEIR model with importation to simulate true measles and non-measles sources of febrile rash (noise) daily incidence. We generated time series of febrile rash (i.e., measles clinical case definition) by summing the daily incidence of measles and either independent Poisson noise or non-measles dynamical noise (consistent with rubella virus). For each time series we assumed a fraction of all cases were seen at a healthcare clinic, and a subset of those were diagnostically confirmed using a test with sensitivity and specificity consistent with either a rapid diagnostic test (RDT) or perfect diagnostic test. From the resulting time series of test positive cases, we define an outbreak alert as the exceedance of a threshold by the 7-day rolling average of observed (test positive) cases. For each threshold level, we calculated percentages of alerts that were aligned with an outbreak (analogous to the positive predictive value), the percentage of outbreaks detected (analogous to the sensitivity), and combined these two measures into an accuracy metric for outbreak detection. We selected the optimal

threshold as the value that maximizes accuracy. We show how the optimal threshold and resulting accuracy depend on the diagnostic test, testing rate, and the type and magnitude of the non-measles noise.

## Results

The optimal threshold for each test increased monotonically as the percentage of clinic visits who were tested increased. With Poisson-only noise, similar outbreak detection accuracies could be achieved with imperfect RDT-like tests as with perfect diagnostic tests (c. 93%), given moderately high testing rates. With larger delays (14 days) between the perfect test administration and result date, RDTs could outperform the perfect tests. Similar numbers of unavoidable cases and outbreak alert delays could be achieved between the test types. With dynamical noise, however, the accuracy of perfect test scenarios was far superior to those achieved with RDTs (c. 93% vs. 73%). For dynamical noise, RDT-based scenarios typically favored a more sensitive alert threshold than perfect test-based scenarios (at a given testing rate), observed with lower numbers of unavoidable cases and detection delays.

## Conclusions

The performance of an outbreak detection system is highly sensitive to the structure and the magnitude of background noise. Under the assumption that the noise is relatively static over time, imperfect diagnostic tests can perform as well as perfect tests in a surveillance system. However, when the noise is temporally correlated, as from a separate SEIR process, imperfect tests cannot overcome their accuracy limitations through higher testing rates.

**Key words:** Rapid-Diagnostic Tests; ELISA; Infectious Disease Surveillance; Outbreak Detection.

## Background

Diagnostics are medical devices and techniques used to detect the presence of a specific pathogen in a host [103]. This may include *in vivo* measures, such as x-ray imagery, or *in vitro* tests to directly quantify the presence of the pathogen itself, e.g. polymerase chain reaction (PCR), or the host's response to the pathogen e.g., enzyme immunoassay/enzyme-linked immunosorbent assay (EIA/ELISA) [21,103,104]. Any given diagnostic will vary in its ability to correctly identify the presence of the pathogen, which is described by its sensitivity and specificity. The sensitivity of a diagnostic is the ability to correctly identify a positive result, conditional on a positive individual being tested i.e., a true positive result [23–25]. The specificity is the opposite: the ability to correctly determine a true negative result, conditional on a negative individual being tested [23–25]. Due to the translation of quantitative measures e.g., immunoglobulin M (IgM) antibody titers, into a binary outcomes (positive/negative), the sensitivity and specificity of a diagnostic are often at odds with one another. For example, using a low optical density value to define the threshold for detection for an ELISA will produce a diagnostic that is highly sensitive, as it only requires a small host response to the pathogen and many resulting antibody titers will exceed this value. However, this may lead to low specificity due to an increase in spurious false positive results from non-infected individuals. To account for these differences, the target product profile (TPP) of a diagnostic provides a minimum set of characteristics that should be met, helping to guide the development and use [105].

The choice to prioritize sensitivity or specificity will be pathogen and context specific. When the cost of a false negative result is disproportionately high relative to a false positive, such as for Ebola [106], highly specific tests are required. This balance will, however, vary as the prevalence of infection in a population varies. Higher presence of infection in a population will increase the positive predictive value (PPV) of

the test i.e., the probability that a positive test reflects a positive individual, that unlike the sensitivity of the test, is not conditioned upon the infection status of the tested individual [23,24]. Regions of high disease burden may therefore prioritize test sensitivity, in contrast to a lower burden location's preference for test specificity and PPV, all else being equal.

At the heart of an outbreak detection system is a surveillance program that enumerates the baseline rate of case incidence and defines an outbreak as a time period with anomalously high incidence relative to that baseline [19,107–109]. As many disease symptoms reflect generic host responses to infection e.g., febrile rash, and infection with a given pathogen can give rise to a wide range of disease symptoms and severity across individuals, accurate methods of case identification are required. Given the imperfect nature of diagnostic classification, any result for an individual is uncertain. Accumulating multiple individual test results to produce population-level counts will propagate this uncertainty, and may result in over- or under-counts due to a preponderance of the diagnostic to produce false positive and negative individual test results, respectively. This process becomes increasingly important when the prevalence of the surveillance program's target disease is low relative to the presence of other sources of clinically-compatible cases; the PPV of an individual diagnostic decreases, increasing the number of false positives, making it harder to identify true anomalies in disease incidence. As a result, it has been commonplace for surveillance systems to be developed around high-accuracy tests, such as PCR and ELISA tests, when financially and logically feasible [110–115].

Outbreak detection systems, like diagnostic tests, must prioritize the sensitivity or specificity of an alerts [116–118]. For many disease systems, particularly in resource constrained environments where the burden of infectious diseases is typically highest [119,120], cases are counted and if a pre-determined threshold is breached, be that weekly, monthly, or some combination of the two, an alert is triggered that may launch

a further investigation and/or a response [117,121]. In effect, this discretizes a distinctly continuous phenomenon (observed cases) into a binary measure (outbreak or no outbreak) for decision making purposes. For reactive management approaches, such as vaccination campaigns and non-pharmaceutical based interventions that are designed to reduce transmission or limit and suppress outbreaks, early action has the potential to avert the most cases [122–127]. While this framing would point towards a sensitive (i.e., early alert) surveillance system being optimal, each action comes with both direct and indirect financial and opportunity costs stemming from unnecessary activities that limit resource availability for future responses. Much like the need to carefully evaluate the balance of an individual diagnostic test's sensitivity and specificity, it is essential to consider at the outbreak level.

The concept of using incidence-based alert triggers to detect the discrete event of an outbreak with characteristics analogous to individual tests has been well documented in the case of meningitis, measles, and malaria [26,27,118,121,126,128,129]. However, an overlooked, yet critical, aspect of an outbreak detection system is the interplay between the individual test and outbreak alert characteristics. With their success within malaria surveillance systems, and particularly since the COVID-19 pandemic, rapid diagnostic tests (RDTs) have garnered wider acceptance, and their potential for use in other disease systems has been gaining interest [130]. Despite concerns of their lower diagnostic accuracy slowing their adoption until recently [131], the reduced cold-chain requirements [132], reduced training and laboratory requirements and costs [115,121,132], and faster speed of result provided by RDTs has been shown to outweigh the cost of false positive/negative results in some settings [130,133–135].

In this chapter we examine how the use of imperfect diagnostic tests affects the performance of outbreak detection in the context of measles where RDTs are being developed with promising results [130,132,136,137] (though not exclusively [138]). We

evaluate the scenarios under which equivalence in outbreak detection can be achieved, where altering testing rates can offset the reduction in diagnostic discrimination of imperfect tests relative to perfect tests, and meaningful improvements can be attained with respect to specific metrics e.g., speed of response. By examining the combination of the alert threshold and individual test characteristics in a modeling study that explicitly incorporates dynamical background noise, we illustrate the need to develop TPPs for surveillance programs as a whole.

## Methods

### Model Structure

We constructed a stochastic compartmental non-age structured Susceptible-Exposed-Infected-Recovered (SEIR) model of measles, and simulated using a modified Tau-leaping algorithm with a time step of 1 day [139]. We utilized binomial draws to ensure compartment sizes remained positive valued [140]. We assumed that the transmission rate ( $\beta_t$ ) is sinusoidal with a period of one year and 20% seasonal amplitude.  $R_0$  was set to 16, with a latent period of 10 days and infectious period of 8 days [110,141]. The population was initialized with 500,000 individuals with Ghana-like birth and vaccination rates, and the final results were scaled up to the approximate 2022 population size of Ghana (33 million) [142]. Ghana was chosen to reflect a setting with a high-performing measles vaccination program that has not yet achieved elimination status (c. 80% coverage for two doses of measles-containing vaccine), and must remain vigilant to outbreaks [143,144]. We assumed commuter-style imports at each time step to avoid extinction; the number of imports each day were drawn from a Poisson distribution with mean proportional to the size of the population and  $R_0$  [145]. The full table of parameters can be found in Table 4.1. All simulations and analysis was

completed in Julia version 1.10.5 [98], with all code stored at <https://github.com/arnoldc/OutbreakDetection>.

Table 4.1: Compartmental model parameters

Parameters	Measles	Dynamical noise
$R_0$	16	5
Latent period (s)	10 days	7 days
Infectious period (g)	8 days	14 days
Seasonal amplitude	0.2	0.2
Vaccination rate at birth (r)	80%	(5-85)%
Birth/death rate (m)	27 per 1000 per annum	
Importation rate		$\frac{1.06*\mu*R_0}{\sqrt{N}}$
Population size (N)	500,000, scaled to 33M	
Initial proportion susceptible	0.05	
Initial proportion exposed	0.0	
Initial proportion infected	0.0	
Initial proportion recovered	0.95	

To examine the sensitivity of the detection system to background noise, we generated a time series of symptomatic febrile rash by combining the measles incidence time series with a noise time series. The noise time series was modeled as either Poisson-only noise, to represent the incidence of non-specific febrile rash due to any of a number of possible etiologies, or dynamical noise modeled as a rubella SEIR process. For Poisson-only noise, the time series of non-measles febrile rash cases each day was constructed by independent draws from a Poisson distribution. For dynamical noise, we generated time series of cases from an SEIR model that matched the measles model in structure, but had  $R_0 = 5$ , mean latent period of 7 days, and mean infectious period of 14 days. We also added additional Poisson noise with mean equal to 15% of the average daily rubella incidence to account for non-rubella sources of febrile rash (Table 4.1) [146,147]. The seasonality for the rubella noise was simulated to be in-phase with measles.

For each noise structure, we simulated five magnitudes of noise ( $\Lambda$ ), representing the average daily noise incidence.  $\Lambda$  was calculated as a multiple ( $c$ ) of the average daily measles incidence ( $\langle \Delta I_M \rangle$ ):  $\Lambda = c \cdot \langle \Delta I_M \rangle$  where  $c \in \{1, 2, 4, 6, 8\}$ . Noise magnitudes will be denoted as  $\Lambda(c)$  for the rest of the manuscript e.g.,  $\Lambda(8)$  to denote scenarios where the average noise incidence is 8 times that of the average measles incidence. For the Poisson-noise scenarios, independent draws from a Poisson distribution with mean  $c \cdot \langle \Delta I_M \rangle$  were simulated to produce the noise time series i.e.,  $\Lambda(c) = \text{Pois}(c \cdot \langle \Delta I_M \rangle)$ . For the dynamical noise scenarios, the rubella vaccination rate at birth was set to 85.38%, 73.83%, 50.88%, 27.89%, or 4.92% to produce equivalent values of  $\Lambda$  (to within 2 decimal places):  $\Lambda(c) = \langle \Delta I_R \rangle + \text{Pois}(0.15 \cdot \langle \Delta I_R \rangle)$ . We simulated 100 time series of 100 years for each scenario, before summarizing the distributions of outbreak detection methods.

## Defining Outbreaks

It is common to use expert review to define outbreaks when examining empirical data, but this is not feasible in a modeling study where tens of thousands of years are being simulated. Previous simulation studies define an outbreak as a period where  $R_t > 1$  with the aim of detecting an outbreak during the grow period [148,149], or use a threshold of  $> 2$  standard deviations (s.d.) over the mean seasonal incidence observed in empirical data (or from a ‘burn-in’ period of the simulation) [150–153].

Here we simulate time series of 100 years and we define a measles outbreak as a region of the time series that meets the following three criteria:

- The daily measles incidence must be greater than, or equal to, 5 cases
- The daily measles incidence must remain above 5 cases for greater than, or equal to, 30 consecutive days

- The total measles incidence must be greater than, or equal to, 500 cases within the bounds of the outbreak

Only events meeting all 3 criteria are classified as outbreaks. The incidence of non-measles febrile rash (i.e., noise) does not affect the outbreak status of a region but may affect the alert status triggered by the testing protocol.

Each day, 60% of the measles and non-measles febrile rash cases visit the clinic for treatment, and a percentage (P) of these clinic visits are tested; all clinic visits are deemed to be suspected measles cases because they meet the clinical case definition. The percentage of clinic visits (P) that are tested is varied between 10% and 60%, in 10% increments. Each “testing scenario” combines a testing rate (P) with one of the following tests:

- An imperfect test with 85% sensitivity and specificity, and 0-day lag in result return. That is, 85% of true measles cases will be correctly labelled as positive, and 15% of non-measles febrile rash individuals that are tested will be incorrectly labelled as positive for measles. This acts as a lower bound of acceptability for a hypothetical measles RDT [136]
- An imperfect test with 90% sensitivity and specificity, and 0-day lag in result return [132]
- A perfect test with 100% sensitivity and specificity, and a 0-day test result delay. This is more accurate than is observed for current ELISA tests [22], but it used to evaluate the theoretical best-case scenario
- A perfect test with 100% sensitivity and specificity, and a 14-day test result delay

For each time series of true measles cases, we define outbreaks as the range of time that meets the definition above (Figure 4.1a). We then add non-measles noise (Figure 4.1b) and test according to the testing scenario, which yields 5 time series of test positive

cases (Figure 4.1c): one time series of all clinically compatible cases and 4 reflecting the testing scenarios.

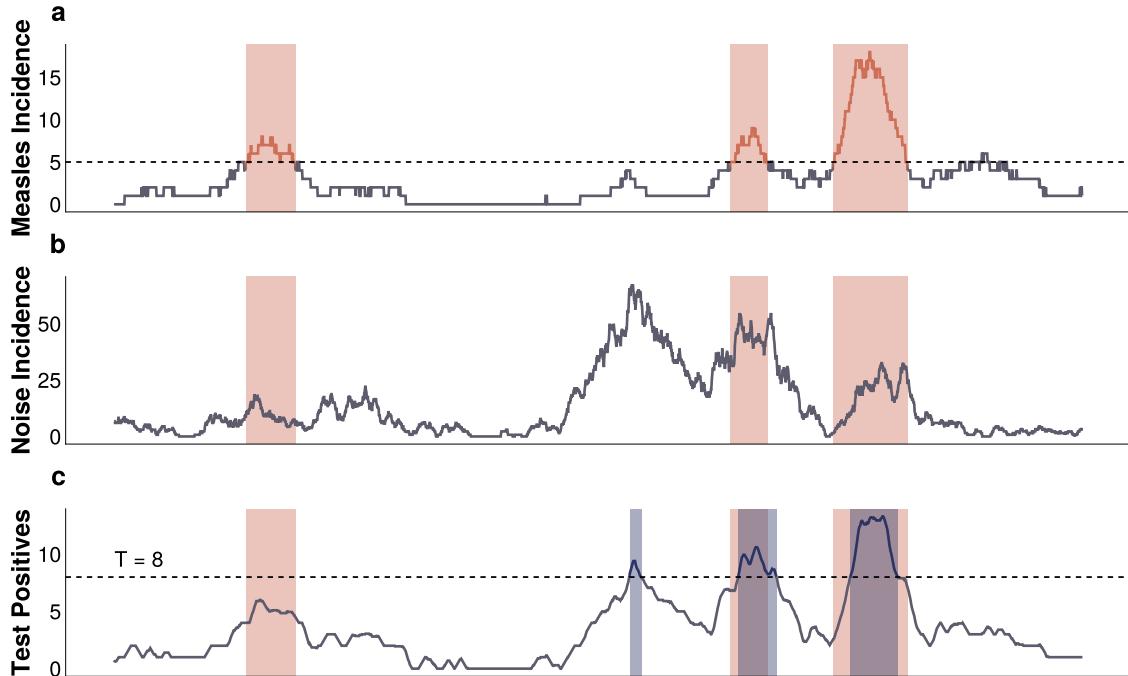


Figure 4.1: A schematic of the outbreak definition and alert detection system. A) Measles incidence time series. B) Noise incidence time series. C) Observed time series of test positive cases according to a given testing scenario. The orange bands present in all 3 panels represent regions of the measles time series that meet the outbreak definition criteria. In panel C, the dark blue bands represent regions of the test positive time series that breach the alert threshold (the horizontal dashed line), and constitute an alert.

## Triggering Alerts

We define an “alert” as any consecutive string of 1 or more days where the 7-day moving average of the test positive cases is greater than, or equal to, a pre-specified alert threshold,  $T$ . For each time series of test positive cases, we calculate the percentage of alerts that are “correct”, defined as any overlap of 1 or more days between the outbreak and alert periods (Figure 4.1c). This is analogous to the PPV of the alert system, and will be referred to as such for the rest of the manuscript. Note that it is

possible to have multiple alerts within a single outbreak if the 7-day moving average of test positive cases drops below, and then re-crosses, the threshold,  $T$ , and we count each as correct. For all outbreaks in the measles time series, we calculate the percentage that contain at least 1 alert within the outbreak's start and end dates (Figure 4.1c). We refer to this as the sensitivity of the alert system. We also calculate the detection delay as the time from the start of an outbreak to the start of its first alert. If the alert period starts before the outbreak and continues past the start date of the outbreak, this would be considered a correct alert with a negative delay i.e., an early warning triggered by false positive test results. Finally, for each time series we calculate the number of unavoidable and avoidable outbreak cases. Unavoidable cases are those that occur before a correct alert, or those that occur in an undetected outbreak. Avoidable cases are defined as those that occur within an outbreak on and after the day of the first correct alert i.e., cases that could theoretically be prevented with a perfectly effective and timely response. Not all cases defined as avoidable would be in practice (due to delays in, and the imperfect nature of, responses); the specifics of operation response are beyond the scope of this work.

We define the accuracy of the surveillance system for a given time series as the mean of the system's PPV and sensitivity. To examine the interaction of the test with the surveillance system's characteristics (i.e., testing rate, noise structure and magnitude), we varied the alert threshold,  $T$ , between 1 and 15 cases per day. Each of the 100 simulations per scenario produces an accuracy, and we identified the optimal alert threshold,  $T_0$ , as the value that produced the highest median accuracy for a given scenario. We then compare testing scenarios at their respective optimal alert threshold. This allows for conclusions to be made about the surveillance system as a whole, rather than just single components.

## Results

The threshold that maximized surveillance accuracy depends on diagnostic test characteristics, the testing rate, and the structure of the non-measles noise (Table 4.2). When the average noise incidence was 8 times higher than the average measles incidence ( $\Lambda(8)$ ), the optimal threshold ( $T_O$ ) ranged between 1 and 7 test positive cases per day. Not surprisingly, the biggest driver of this difference was the testing rate; as a larger fraction of suspected cases are tested, the optimal threshold increases monotonically for all test and noise types (Table 4.2).

The maximal attainable surveillance accuracy at the optimal threshold depends strongly on the structure and magnitude of the background noise. For Poisson noise, at all magnitudes, the maximum surveillance accuracy increases rapidly from 65% at 10% testing of suspected clinic cases, to  $\approx 90\%$  accuracy at  $\geq 20\%$  testing, for all test types (Figure 4.2). For dynamical SEIR noise, the perfect tests perform identically to the Poisson noise case at all magnitudes (Figure 4.2). For imperfect diagnostic tests, which have lower individual sensitivity and specificity, the maximal attainable accuracy is lower than the perfect tests for all testing rates ( $P$ ) at noise magnitude  $\geq \Lambda(2)$  (Figure 4.2). Notably, the surveillance accuracy declines with increasing noise and, at all noise levels, is not improved with higher testing rates as the signal becomes increasingly dominated by false positive test results (Figure 4.2).

Table 4.2: Optimal outbreak alert thresholds for imperfect and perfect diagnostic tests, under dynamical and Poisson noise structures where the average daily noise incidence is 8 times the average daily measles incidence  $\Lambda(8)$ . The test sensitivity equals the test specificity for all diagnostic tests.

Noise Type	Test Characteristic		Testing Rate					
	Test Type	Test Lag	10%	20%	30%	40%	50%	60%
Dynamical noise	Imperfect Test (85%)	0	1	2	3	4	4	5
Dynamical noise	Imperfect Test (90%)	0	1	2	3	4	4	5
Poisson noise	Imperfect Test (85%)	0	1	3	4	5	6	7
Poisson noise	Imperfect Test (90%)	0	1	2	4	5	5	6
All noise structures	Perfect Test	0	1	2	3	4	4	5
All noise structures	Perfect Test	14	1	2	3	4	4	5

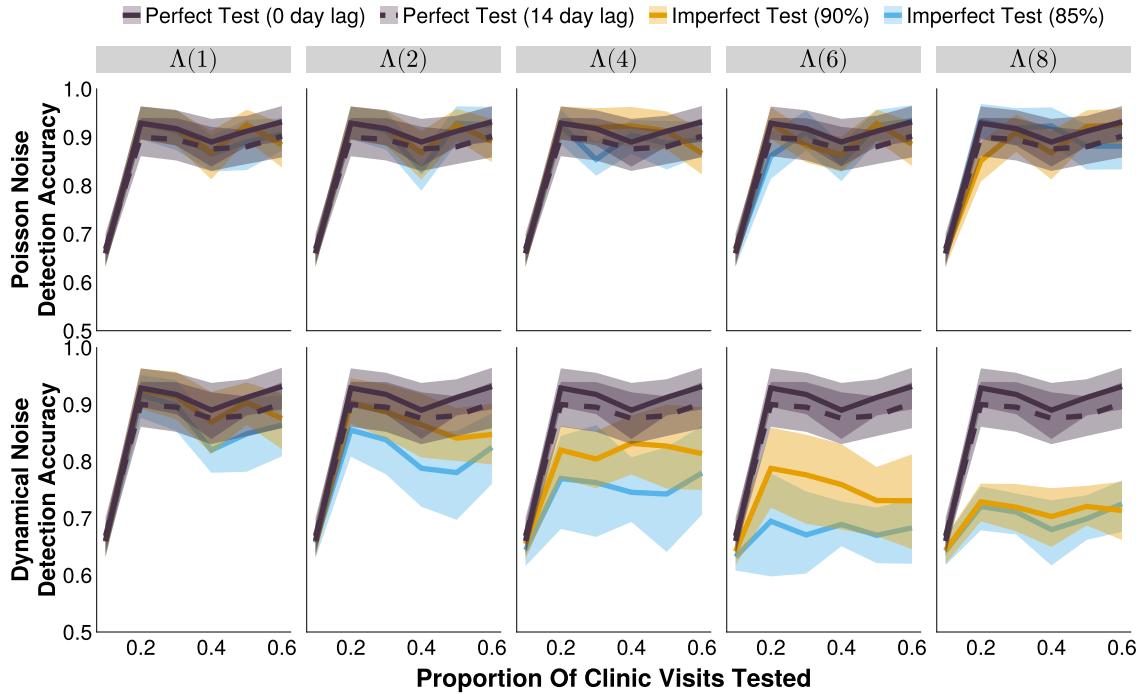


Figure 4.2: The accuracy of outbreak detection systems under different testing rates and noise structures, at their respective optimal alert thresholds. The shaded bands illustrate the 80% central interval, and the solid/dashed lines represent the mean estimate. Imperfect tests have the same values for sensitivity and specificity. Solid lines represent tests with 0-day turnaround times, and dashed lines represent tests with result delays.  $\Lambda(4)$  indicates the mean noise incidence is 4 times higher than the mean measles incidence, for example.

Introducing a lag in test result reporting necessarily decreases surveillance accuracy because an alert can only begin once the test results are in-hand, which increases the chance that an outbreak will end before results can be translated to an alert. For the conditions simulated here, introducing a 14-day lag in test reporting for a perfect test reduces the surveillance accuracy by  $\approx 3\%$ . For all simulated scenarios, this is consistent with, or higher than, the accuracy achievable with an RDT-like imperfect test. This always leads to an increase in the median delay from outbreak start to alert, relative to a perfect test with no result delays, as well as imperfect tests (Figure 4.3).

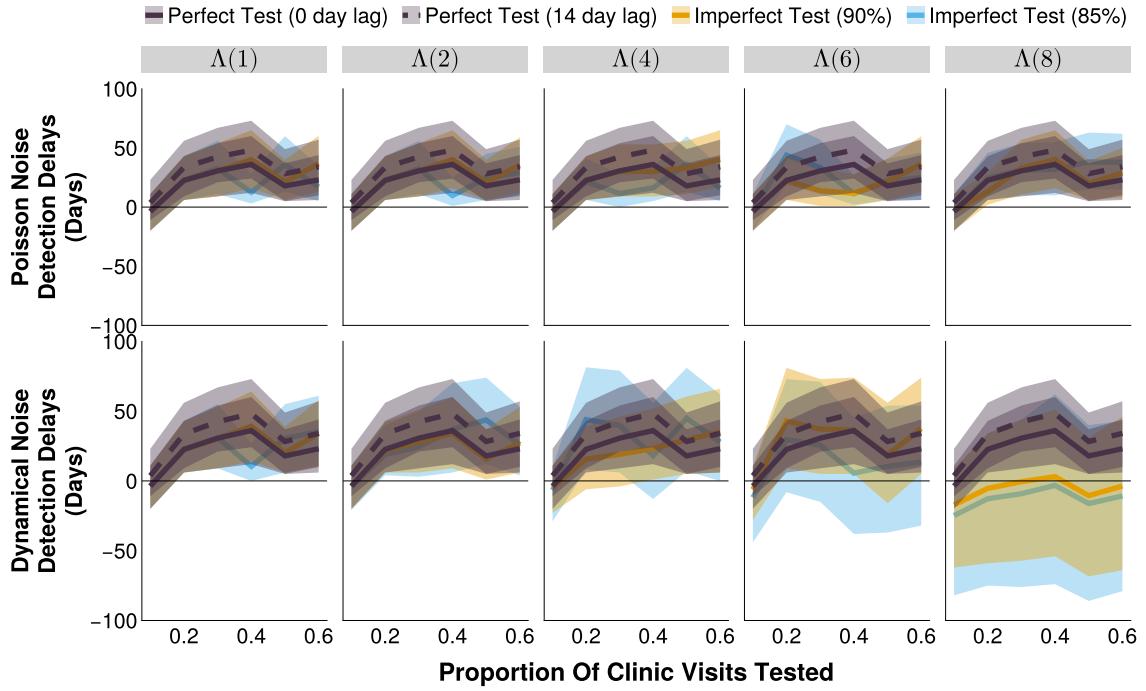


Figure 4.3: The detection delay of outbreak detection systems under different testing rates and noise structures, at their respective optimal alert thresholds. The shaded bands illustrate the 80% central interval, and the solid/dashed lines represent the mean estimate. Imperfect tests have the same values for sensitivity and specificity. Solid lines represent tests with 0-day turnaround times, and dashed lines represent tests with result delays.  $\Lambda(4)$  indicates the mean noise incidence is 4 times higher than the mean measles incidence.

It is notable that surveillance metrics do not change monotonically with an increase in testing rate, and this holds regardless of the type of test. This effect is exaggerated for some metrics (detection delays, proportion of time in alert, and number of unavoidable cases) than others (accuracy). In general, the increase in accuracy with higher testing rates is accompanied with longer testing delays. This reflects the change from highly sensitive systems with low thresholds to more specific systems with higher thresholds at higher testing rates. For Poisson noise, similar detection delays are observed for all test and noise magnitudes, with most of the variation attributable to the change in the testing rate (means of -3.7 to 36.1 days). Under dynamical noise, there are clearer differences in the performance of perfect and imperfect diagnostic tests, with the separation of outcomes occurring later than observed for surveillance accuracy ( $\Lambda(8)$  vs

$\Lambda(2)$  – Figure 4.3 and Figure 4.2 – respectively). With large amounts of dynamical noise ( $\Lambda(8)$ ), the mean detection delay of the 90% and 85% imperfect tests range from –17.5 days to 3.2 days, and from –25.2 days to –3.4 days, respectively. Negative delays indicate that alerts are being triggered before the start of the outbreak and is correlated with the proportion of the time series that is under alert, with larger negative delays associated with more and/or longer alert periods (Figure 4.4, Supplemental Figure C.1, Supplemental Figure C.2). Long detection delays manifest as large numbers of unavoidable cases (i.e., cases that occur between the outbreak start and its detection) (Figure 4.5). Given the exponential trajectory of infections in the initial phase of an outbreak, the pattern of unavoidable cases follows the same shape as for detection delays, but more exaggerated.

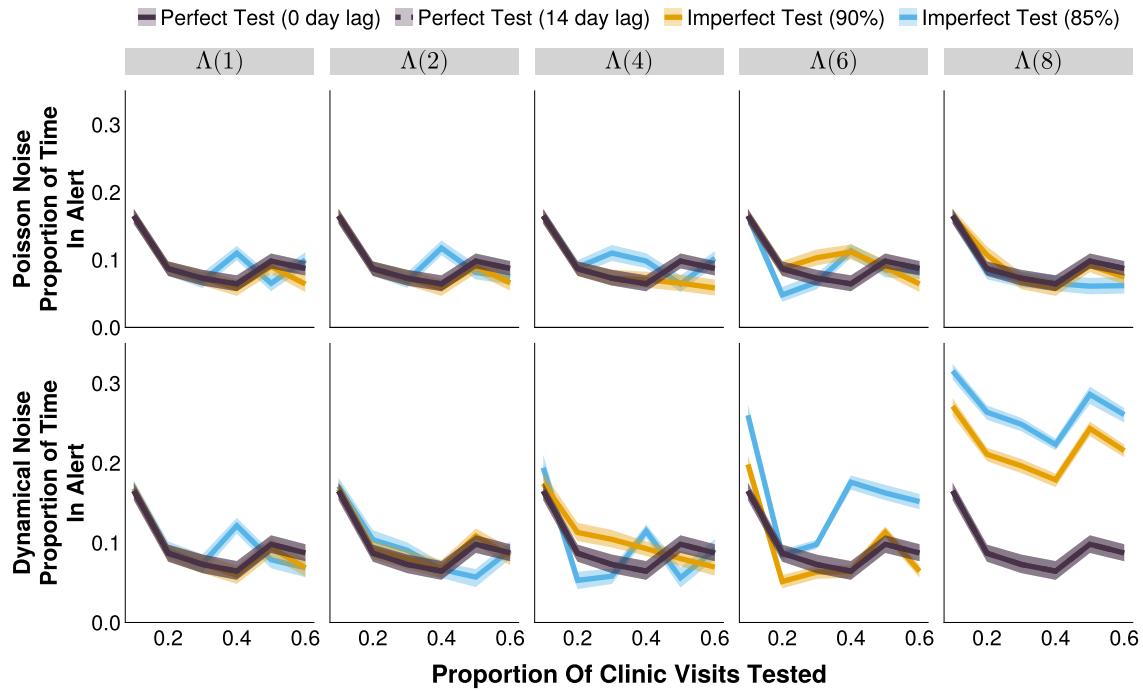


Figure 4.4: The proportion of the time series in alert of outbreak detection systems under different testing rates and noise structures, at their respective optimal alert thresholds. The shaded bands illustrate the 80% central interval, and the solid/dashed lines represent the mean estimate. Imperfect tests have the same values for sensitivity and specificity. Solid lines represent tests with 0-day turnaround times, and dashed lines represent tests with result delays.  $\Lambda(4)$  indicates the mean noise incidence is 4 times higher than the mean measles incidence.

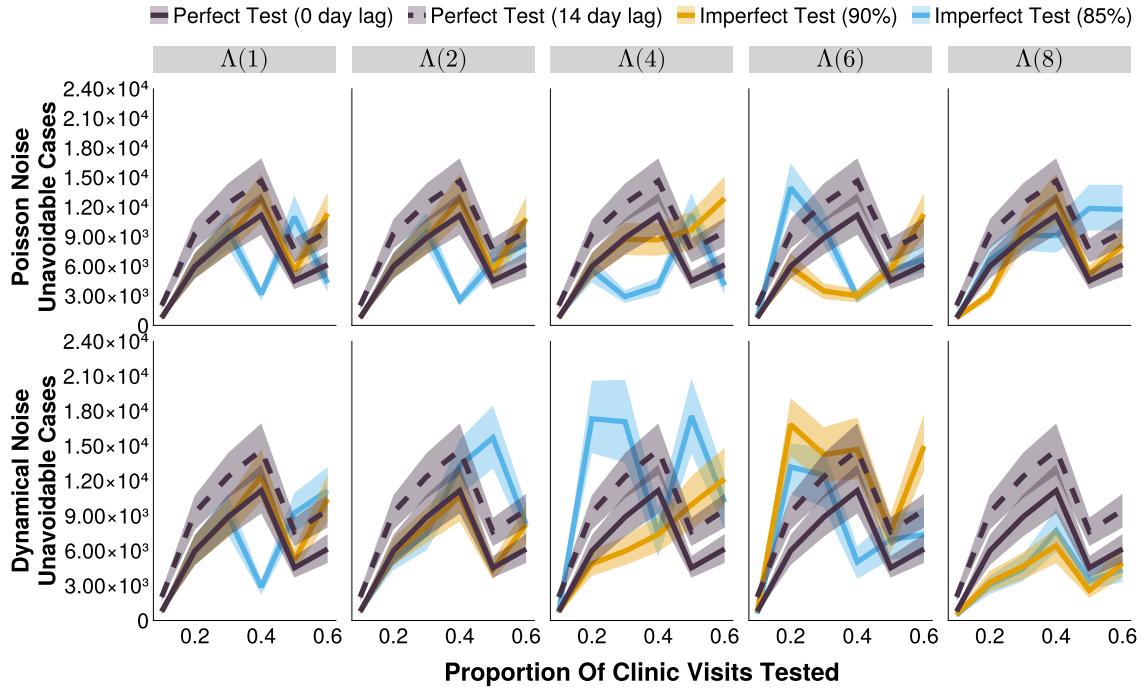


Figure 4.5: The number of unavoidable cases of an outbreak detection systems under different testing rates and noise structures, at their respective optimal alert thresholds.

The shaded bands illustrate the 80% central interval, and the solid/dashed lines represent the mean estimate. Imperfect tests have the same values for sensitivity and specificity. Solid lines represent tests with 0-day turnaround times, and dashed lines represent tests with result delays.  $\Lambda(4)$  indicates the mean noise incidence is 4 times higher than the mean measles incidence.

## Discussion

The performance of an outbreak detection system is highly sensitive to the structure and level of background noise in the simulation. Despite the mean daily noise incidence set to equivalent values between the dynamical and Poisson-only simulations, drastically different results are observed.

Under the assumption that non-measles febrile rash is relatively static in time (Poisson noise scenarios), imperfect (RDT-like) diagnostics can perform as well, if not better than perfect (ELISA-like) tests at moderate to high testing rates, and at a fraction

of the cost [132]. However, if it is expected that the noise is dynamic, imperfect tests cannot overcome their accuracy limitations through higher testing rates, saturating at c. 74% accuracy, relative to 93% achieved with perfect tests. This discrepancy occurs because, despite the same average incidence of noise in each (comparable) scenario, the relative proportion of measles to noise on any one day varies throughout the dynamical noise time series, exacerbating the increase in false positive and negative test results as the diagnostic's sensitivity and specificity declines.

For all noise structures and diagnostic tests, increasing testing rate was not accompanied by a monotonic change in the associated metrics. The reason behind this unintuitive result stems from the use of integer-valued alert thresholds. For a given diagnostic test, increasing the testing rate will result in an increase in the number of observed (test positive) cases. This, however, may not translate to an integer increase in the moving average of test positive results, which is used to trigger an alert. Even with a perfect test, the alert system must discriminate between endemic/imported cases and epidemic cases. As such, the threshold may stay the same as the optimal value selected for the previous testing rate, providing an overly sensitive system that will be triggered more frequently by endemic cases. Or, it can increase, resulting in a system with a higher PPV per alert, but lower surveillance sensitivity. Both options may translate to a lower surveillance accuracy than observed when fewer individuals are tested. But more importantly, this can result in contiguous testing rates selecting for system sensitivity vs PPV differently, translating to discontinuous changes in the outbreak delays (Figure 4.3), unavoidable cases (Figure 4.5), and proportion of the time series in alert status (Figure 4.4).

Surveillance is counting for action [20]. What actions are taken depend upon the constraints imposed, and the values held, within a particular surveillance context. This analysis is therefore not a complete optimization, which would require explicit decisions to be made about the preference for increased speed at the cost of higher false

alert rates and lower PPV (and visa versa). These will be country-specific decisions, and they may change throughout time; for example, favoring RDTs when there are low levels of background infections, and ELISAs during large (suspected) rubella outbreaks. These trade-offs must be explicitly acknowledged when designing surveillance systems, and we present a framework to account for the deep interconnectedness of individual and population-level uncertainties that arise from necessary categorizations.

## **Limitations and Strengths**

To our knowledge, this is one of the first simulation studies to examine the relationship between individual test characteristics and the wider surveillance program. By explicitly modeling the interaction between the two, we make a case that surveillance systems should take a holistic approach; prematurely constraining one component can lead to drastically different, and suboptimal, results. Additionally, by defining outbreak bounds concretely we have been able to calculate metrics of outbreak detection performance that draw parallels to those used when evaluating individual diagnostic tests. This provides an intuitive understanding and simple implementation of this method in resource-constrained environments, something that may not be possible with many outbreak detection and early warning system simulations in the literature. An evaluation of all outbreak detection algorithms is beyond the scope of this work, but a more computationally expensive approach based on nowcasting incidence may help overcome the shortcomings of imperfect diagnostics in high-noise scenarios.

For computational simplicity, this paper did not include demography in the model structure. And while a simulation-based approach allows for complete determination of true infection status i.e., measles vs non-measles febrile rash cases, and therefore an accurate accounting of the outbreak and alert bounds, these simulations do not specifically represent any real-world setting. The evaluation of empirical data does

provide this opportunity, but at the cost of knowing the true infection status of individuals, confounding of multiple variables, limiting analysis to only those who are observed (i.e., not those in the community who do not visit a healthcare center), and removing the possibility to explore the sensitivity of the results to parameters of interest to a surveillance program e.g., testing rate, and the test itself.

Additionally, it has been well documented that the performance of an individual test is highly sensitive to its timing within a person's infection cycle [110,134,135,154,155], so it is possible that different conclusions would be drawn if temporal information about the test administration was included in the simulation.

Finally, the optimal threshold for a testing scenario is affected by the use of integer-values; smaller steps could be chosen to potentially minimize discontinuities. Similarly, the optimal threshold depends heavily on the costs ascribed to incorrect actions, be that failing to detect an outbreak or incorrectly mounting a response for an outbreak that doesn't exist. In the simulations we have weighted them equally, but it is likely that they should not be deemed equivalent; missing an outbreak may result in many thousands of cases, whereas an unnecessary alert would generally launch an initial low-cost investigation for full determination of the outbreak status. This is particularly important in countries with vast heterogeneity in transmission: different weightings should be applied to higher vs. lower priority/risk regions to account for discrepancies in consequences of incorrect decisions.

Given these limitations, the explicit values (i.e., optimal thresholds, accuracies etc.) should be interpreted with caution, and the exact results observed in the real-world will likely be highly dependent on unseen factors, such as the proportion of measles and non-measles sources of febrile rash that seek healthcare. However, the general patterns should hold, and more importantly, the analysis framework provides a consistent and

holistic approach to evaluating the trade-off between individual level tests and the alert system enacted to detect outbreaks.

## Acknowledgements

### Funding

This work was supported by Gavi, the Vaccine Alliance, through Contract No. 266639 PO4500012795, A01. The funding sources had no role in the collection, analysis, interpretation, or writing of the work.

### Author Contributions

*Conceptualization:* all authors

*Data curation:* CA, MJF

*Formal analysis:* CA

*Funding acquisition:* AW, BP, MJF, WM

*Investigation:* CA

*Methodology:* CA, MJF

*Project administration:* MJF

*Software:* CA

*Supervision:* AW, BP, MJF, WM

*Validation:* CA

*Visualization:* CA

*Writing - original draft:* CA

*Writing - review and editing:* all authors

### Conflicts of Interest and Financial Disclosures

The authors declare no conflicts of interest.

### Data Access, Responsibility, and Analysis

Callum Arnold and Dr. Matthew J. Ferrari had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Callum Arnold (Department of Biology, Pennsylvania State University) conducted the data analysis.

### Data Availability

All code and data for the simulations can be found at <https://github.com/arnold-c/OutbreakDetection>

# Chapter 5 | Diagnostic Uncertainty Limits the Potential of Early Warning Signals to Identify Epidemic Emergence

Callum R.K. Arnold<sup>1,2</sup>, Matthew J. Ferrari<sup>1,2</sup>

<sup>1</sup> Department of Biology, Pennsylvania State University, University Park, PA, USA 16802

<sup>2</sup> Center for Infectious Disease Dynamics, Pennsylvania State University, University Park, PA, USA 16802

## Abstract

Methods to detect the emergence of infectious diseases, and approach to the “critical transition”  $R_E = 1$ , have the potential to avert substantial disease burden by facilitating pre-emptive actions like vaccination campaigns. Early warning signals (EWS), summary statistics of infection case time series, show promise in providing such advanced warnings. As EWS are computed on test positive case data, the accuracy of this underlying data is integral to their predictive ability, but will vary with changes in the diagnostic test accuracy and the incidence of the target disease relative to clinically-compatible background noise. We simulated emergent and null time series as the sum of an SEIR-generated measles time series, and background noise generated by either independent draws from a Poisson distribution, or an SEIR simulation with rubella-like parameters. We demonstrate that proactive outbreak detection with EWS metrics is resilient to decreasing diagnostic accuracy, so long as background infections remain proportionally low. Under situations with large, episodic, noise, imperfect diagnostic tests cannot appropriately discriminate between emergent and null periods. Not all EWS metrics performed equally; we find that the mean was the least affected by changes to the noise structure and magnitude, given a moderately accurate diagnostic test ( $\geq 95\%$  sensitive and specific), and the autocovariance and variance were the most predictive when the noise incidence did not exhibit large temporal variations. In these situations, diagnostic test accuracy should not be a precursor to the implementation of an EWS metric-based alert system.

**Key words:** Diagnostic Test Uncertainty; Infectious Disease Surveillance; Outbreak Detection; Early Warning Signals; Critical Slowing Down.

## Background

Despite sustained advances over decades, infectious diseases still pose a substantial threat to human life, causing an estimated 33.8 B infections, and 57.0 M deaths, per annum in 2019 (rising to an estimated 67.9 M deaths in 2021, as a result of the COVID-19 pandemic) [156]. For many diseases, effective and affordable vaccines have played a substantial role in reducing this burden, averting 154 million deaths since the introduction of the Expanded Programme on Immunization in 1974 [157]. As burden decreases with increasing control, dynamics may shift from predictable annual incidence to increasingly variable and episodic dynamics [158]. Many populations that have achieved apparent control, suffer from large-scale resurgent outbreaks due to the build up of susceptibles in the absence of persistent transmission [127,159,160]. While rapid detection and response has the potential to minimize the impact of these outbreaks [124,125,161,162], early warning systems that can trigger pre-emptive action prior to outbreaks would be the ideal.

Infectious disease surveillance systems are crucial for detecting outbreaks [20,163], and could be leveraged to anticipate the risk of outbreaks [164–166]. Outbreak detection and response systems are reactive in nature; cases are collated, counted, and if a pre-determined threshold is met or breached, an action is undertaken (e.g., preliminary investigation, or reactive vaccination campaign) [117,121]. However, due to the exponential trajectory of incidence in the early stages of an outbreak, the reactive nature necessarily results in excess infections that cannot be prevented [124,125,161]. To limit the burden of disease, ideally, epidemiologists could utilize the output of a surveillance system (e.g., the trend in cases of a pathogen) to predict the risk of a future outbreak, triggering a *proactive* action, such as a preventative vaccination campaign.

The risk of an outbreak can be quantified in terms of the effective reproduction number,  $R_E$ , defined as the expected number of secondary cases due to each infectious individual [4].  $R_E = 1$  represents a “critical transition”, below which epidemics should not spread, and above which outbreaks should propagate. There has been growing interest, in many fields, to identify and develop early warning signals (EWS) that are predictive of the approach to such critical transitions in dynamical systems [167–171]. The appeal of an alert system based upon EWS metrics is that they are model-free, only requiring the calculation of summary statistics of a time series. Prior work has demonstrated that for infectious disease systems, computing EWS metrics on the progression of population susceptibility may be most predictive [166], but collecting this information is often intractable, and utilizing either the incidence or prevalence data has provided similarly useful predictions [164,166,172]. If an EWS is predictive, critical slowing down theory suggests that the EWS values will change as a transition is approached, such as an increase in the variance. Prior work has demonstrated that EWS metrics are theoretically correlated with a critical transition for infectious disease systems, under emergent and extinction conditions [164–166,172–174].

While identifying EWS that are correlated with a transition is an important first step, systems to preempt outbreaks also require a discrete decision threshold to trigger preventive action (e.g., vaccination) [175,176]. To address this, various threshold-based and statistical learning based approaches have been developed [170,177–180]. For these, a distribution of the EWS metric is quantified during a non-outbreak regime and a decision threshold is triggered when the EWS metrics at time  $t$  exceeds some quantile or derived statistic of this distribution; often the mean plus 2 times the standard deviation. Prior work has shown that a single exceedance is often too sensitive and requiring multiple consecutive flags to trigger an alert improves the accuracy in a ‘noisy’ system by reducing the false positive rate [177,181,182].

Until now, the relatively nascent topic of EWS for outbreak detection has only explored imperfect surveillance in the setting of under-reporting and temporal aggregation of case data [164,183]. Our goal is to characterize the performance of EWS metrics for outbreak detection in a surveillance system with diagnostic uncertainty due to co-circulating pathogens and imperfect diagnostic tests, i.e., non-target disease that may be misdiagnosed as the target disease. For diseases with non-specific symptoms, e.g., measles and rubella that often co-circulate and have similar clinical presentation [121,144], an imperfect diagnostic test will result in false positive and negative cases. In this chapter we show the conditions under which diagnostic uncertainty overwhelms the time series used to calculate EWS summary statistics, limiting the ability to predict epidemic transitions.

## Materials & Methods

### Model Structure

We modeled the dynamics of a target pathogen (measles), for which we want to detect outbreaks, with a stochastic compartmental non-age structured Susceptible-Exposed-Infected-Recovered (SEIR) model. The SEIR model was simulated using a Tau-leaping algorithm with a time step of 1 day, with binomial draws so that no jump resulted in negative compartment sizes [139,140]. We assumed no seasonality in the transmission rate ( $\beta_t$ ), and set the latent and infectious periods equal to 10 days and 8 days, respectively, and an  $R_0$  equal to 16, approximating measles parameters values [110,141]. Demographic parameters (birth and death rates) broadly reflecting those observed in Ghana were selected to evaluate the performance of EWS metrics in a setting where high, yet sub-elimination, vaccination coverage is observed, requiring ongoing vigilance [143,144]. An initial population of 500,000 individuals was simulated, with commuter-

style imports drawn from a Poisson distribution with mean proportional to the size of the population and  $R_0$ , to maintain a level of endemicity [145].

To evaluate the predictive ability of EWS metrics in environments with background disease that could produce false positive test results if tested with an imperfect diagnostic, we generated a time series of “suspected measles” by summing the measles and background noise time series. The noise time series is modeled as either: independent draws of a Poisson distribution, with mean equal to a multiple ( $c$ ) of the daily average measles incidence, where  $c \in \{1, 7\}$ ; or from an SEIR time series with rubella-like parameters with additional noise drawn from a Poisson distribution with mean equal to 15% of the daily average of the rubella incidence time series, to account for non-rubella sources of clinically-compatible febrile rash e.g., parvovirus (Table 5.1) [146,147]. Under dynamical (SEIR-generated) noise simulations, the vaccination rate at birth was selected to produce equivalent magnitudes of daily average noise incidence as observed in the Poisson-like noise simulations (10.20% and 87.34%). Throughout the rest of the manuscript, these will be referred to as low and high Poisson/dynamical noise scenarios, accordingly. Each day, all clinically-compatible febrile rash cases (that is, both the measles and noise time series) were tested using one of the following diagnostic tests, producing a time series of test positive cases.

- A perfect test with 100% sensitivity and specificity. This was chosen to reflect the best-case scenario that the imperfect diagnostic-based alert scenarios could be compared against.
- An imperfect diagnostic with sensitivity and specificity equal to either 99%, 98%, 97%, 96%, 95%, 90%, or 80%.

Table 5.1: Compartmental model parameters

Parameters	Measles - Emergent	Measles - Null	Dynamical noise
R <sub>0</sub>	16		5
Latent period (s)	10 days		7 days
Infectious period (g)	8 days		14 days
Vaccination rate at birth during burn-in period ( $r_i$ )	Unif (92.69%, 100%)		10.20%, 83.74%
Vaccination rate at birth after burn-in period ( $r_e$ )	Unif (60%, 80%)	Unif (92.69%, 100%)	10.20%, 83.74%
Birth/death rate (m)	27 per 1000 per annum		
Importation rate	$\frac{1.06 * \mu * R_0}{\sqrt{N}}$		
Population size (N)	500,000		
Initial proportion susceptible	0.05		
Initial proportion exposed	0.0		
Initial proportion infected	0.0		
Initial proportion recovered	0.95		

To evaluate the performance of the EWS metrics at predicting the approach to the critical transition ( $R_E = 1$ ) from below, we simulated “emergent” scenarios where  $R_E$  increases until 1, and “null” scenarios where  $R_E$  is below 1. For both emergent and null scenarios, we generated 100 time series. All measles simulation incorporated a 5-year burn-in period to produce sufficient data for calculation of the EWS metrics upon aggregation, as well as to produce greater variation in the trajectory of  $R_E$ . For each time series, the vaccination rate at birth during the burn-in period was sampled from a Uniform distribution between 92.69% and 100% coverage. These bounds were selected to ensure the maximum value of  $R_E$  that could be reached within 10 years (twice the

length of the burn-in period) was 0.9. We simulated emergent scenarios by lowering the vaccination rate at birth after completion of the burn-in period, allowing the proportion of the population that is susceptible to grow. For each emergent time series, the vaccination rate at birth was independently drawn from a Uniform distribution between 60% and 80% coverage, allowing the rate of growth in  $R_E$ , and therefore the time of the critical transition, to vary in each emergent time series. For each null time series, the vaccination rate at birth was set to the coverage sampled during the burn-in period, ensuring  $R_E$  would not cross the critical transition within the scope of the simulation, though it may grow slowly. Each of the 100 emergent and 100 null time series are paired during the pre-processing steps i.e., up until the completion of the burn-in period, paired emergent and null simulations share the same vaccination rate at birth, and they are both truncated to identical lengths (the time step when  $R_E = 1$  in that pair's emergent simulation).

All simulations and analysis was completed in Julia version 1.10.5 [98], with all code stored at <https://github.com/arnold-c/CSDNoise>.

## Computing & Evaluating EWS

Each set of null and emergent time series are aggregated by month and numerical estimates of the EWS metrics were then calculated on the aggregated time series, de-trended using backwards-facing moving averages with bandwidth  $b = 52$  weeks. For example, the EWS metric, the mean, is given by the expectation in Equation 5.1 where:  $X_s$  represents the aggregated incidence at time point (month)  $s$ , and  $\delta = 1$  time step (in the simulation results presented, 1 month). At the beginning of the time series when  $t < b$ ,  $b$  is set equal to  $t$ .

$$\hat{\mu}_t = \sum_{s=t-b\delta}^{s=t} \frac{X_s}{b} \quad 5.1$$

In this chapter we evaluate the performance of the following EWS metrics: the mean, variance, coefficient of variation, index of dispersion, skewness, kurtosis, autocovariance, and autocorrelation at lag-1, which have previously been shown to be correlated or predictive of disease emergence [164,166,175,175,183]. The full list of numerical formulas for each EWS metric can be found in Table 5.2.

Table 5.2: Numerical computations for EWS metrics, where  $\delta = 1$  time step,  $b = 52$  weeks

EWS Metric	Formula
Mean ( $\hat{\mu}_t$ )	$\sum_{s=t-b\delta}^{s=t} \frac{X_s}{b}$
Variance ( $\hat{\sigma}_t^2$ )	$\sum_{s=t-b\delta}^{s=t} \frac{(X_s - \hat{\mu}_s)^2}{b}$
Coefficient of Variation ( $\widehat{CV}_t$ )	$\frac{\hat{\sigma}_t}{\hat{\mu}_t}$
Index of Dispersion ( $\widehat{IoD}_t$ )	$\frac{\hat{\sigma}_t^2}{\hat{\mu}_t}$
Skewness ( $\widehat{Skew}_t$ )	$\frac{1}{\hat{\sigma}_t^3} \sum_{s=t-b\delta}^{s=t} \frac{(X_s - \hat{\mu}_s)^3}{b}$
Kurtosis ( $\widehat{Kurt}_t$ )	$\frac{1}{\hat{\sigma}_t^4} \sum_{s=t-b\delta}^{s=t} \frac{(X_s - \hat{\mu}_s)^4}{b}$
Autocovariance ( $\widehat{ACov}_t$ )	$\sum_{s=t-b\delta}^{s=t} \frac{(X_s - \hat{\mu}_s)(X_{s-\delta} - \hat{\mu}_{s-\delta})}{b}$
Autocorrelation lag-1 ( $\widehat{AC-1}_t$ )	$\frac{\widehat{ACov}_t}{\hat{\sigma}_t \hat{\sigma}_{t-\delta}}$

Once the EWS metrics have been computed, the correlation within emergent time series is computed using Kendall's Tau-B, signifying if an EWS metric consistently increases (or decreases) in magnitude throughout the time series [184,185]. Kendall's Tau is computed on two lengths of time series: from the beginning of the simulation

until the critical transition is met, and from the completion of the burn-in period until the critical transition. To evaluate the strength of the correlation, we use the area under the receiver operator curve (AUC), as described in prior papers [164,172,175]. Briefly, the calculation of the AUC compares whether the distributions of Kendall's Tau differ substantially between emergent and null simulations for a given alert scenario and EWS metric. AUC is calculated using the rank order of the EWS metrics for both emergent and null time series using Equation 5.2 [186] where  $r_{\text{null}}$  equals the sum of ranks for the null time series, and  $n_{\text{null}}$  and  $n_{\text{emergent}}$  refer to the number of null and emergent simulations, respectively. An AUC of 0.5 indicates the EWS is similarly correlated with both emergent and null time series, offering no benefit; values  $> 0.5$  indicate a positive correlation with emergent time series, and  $< 0.5$  indicates the EWS metric is negatively correlated with the emergent simulations. AUC values are transformed as  $|\text{AUC} - 0.5|$  to highlight the strength of the correlation with emergence, with values close to 0 exhibiting poor performance, and a value of 0.5 indicating perfect correlation [164].

$$\text{AUC} = \frac{r_{\text{null}} - n_{\text{null}}(n_{\text{null}} + 1)/2}{n_{\text{emergent}}n_{\text{null}}} \quad 5.2$$

The primary mode of evaluation for the EWS metrics relies on computing and triggering an alert based upon a set of alert conditions. The alert scenario is defined as the combination of diagnostic test, noise structure and magnitude, and EWS metric. The combination of the alert scenario and EWS alert hyperparameters (the quantile threshold,  $Q$ , value of the long-running metric distribution that must be exceeded to create a flag, and the number of consecutive flags,  $C$ , required to trigger an alert),

produce distinct counts of emergent and null time series that result in an alert. For example, a simulation may require that at two consecutive time points ( $C = 2$ ), the corresponding values of the EWS are larger than 95% of previously observed EWS values ( $Q = 0.95$ ). The sensitivity of the system is defined as the proportion of the emergent simulations that result in an alert, and the specificity is the proportion of the null simulations that do not result in an alert. Taking the mean of the sensitivity and specificity produces the accuracy of the system. For each alert scenario, a grid search over the EWS hyperparameters ( $Q \in [0.5, 0.99]$ ,  $C \in [2, 30]$ ) is performed to identify the set of EWS hyperparameters that maximizes alert accuracy for a given alert scenario. If multiple hyperparameter combinations produce identical alert system accuracies, the combination with the highest alert specificity is selected. After the optimal EWS hyperparameters have been selected, the accuracy of each EWS metric are compared across alert scenarios, at their respective maximal values. Finally, the speed and timing of detection relative to the critical transition is evaluated using Kaplan-Meier survival estimates [187]. Alert accuracy was only evaluated on EWS calculated after the completion of the burn-in period.

## Results

### Correlation with Emergence

The strength and direction of the raw correlation ( $\text{Tau}$ ) between EWS metrics and the approach to the critical transition in emergent time series is strongly dependent upon the length of the time series evaluated;  $\text{Tau}$  is higher when calculated after the burn-in period for the top 5 ranked metrics (Table 5.3). Normalizing the correlation in the emergent time series against the correlation observed in null simulations yields comparable results when calculated from the full time series and only after the burn-in

(Table 5.3). Consistent with previous studies, the autocovariance, variance, mean, and index of dispersion show the strongest correlations with emergence ( $|AUC - 0.5| = 0.20, 0.20, 0.18, 0.13$ , evaluated after the burn-in period, respectively) [164,183].

Table 5.3: The ranking and mean value of Kendall's Tau computed on emergent time series, and the  $|AUC - 0.5|$  for each metric. The values are computed on the full time series, and the subset from after the completion of the burn-in period, with a perfect test

Rank	Tau		$ AUC - 0.5 $	
	Full Time Series	After Burn-In Period	Full Time Series	After Burn-In Period
1	Index of dispersion (0.26)	Variance (0.62)	Mean (0.19)	Autocovariance (0.20)
2	Autocovariance (0.25)	Index of dispersion (0.58)	Autocovariance (0.17)	Variance (0.20)
3	Variance (0.25)	Autocovariance (0.58)	Variance (0.17)	Mean (0.18)
4	Mean (0.22)	Autocorrelation (0.38)	Index of dispersion (0.14)	Index of dispersion (0.13)
5	Autocorrelation (0.17)	Mean (0.38)	Coefficient of variation (0.12)	Autocorrelation (0.12)
6	Coefficient of variation (-0.01)	Coefficient of variation (0.15)	Autocorrelation (0.10)	Coefficient of variation (0.11)
7	Kurtosis (-0.04)	Skewness (0.06)	Skewness (0.10)	Skewness (0.10)
8	Skewness (-0.08)	Kurtosis (-0.02)	Kurtosis (0.02)	Kurtosis (0.03)

With an imperfect diagnostic test, the correlation with emergence was more influenced by the noise structure (Poisson vs. dynamical) than the noise magnitude (Table 5.4). For an imperfect test with 90% sensitivity and specificity, the correlation between all EWS metrics and emergence was relatively unaffected by the magnitude of Poisson noise. The top four metrics with a perfect diagnostic test (autocovariance, variance, mean, and index of dispersion) maintained their positions as the most strongly correlated metrics.

For simulations with rubella-like SEIR dynamical noise, the correlation of all metrics was lower at low dynamical noise compared to low Poisson noise (Table 5.4). With low levels of dynamical noise, the autocovariance, variance, and mean remained the most correlated with emergence ( $|AUC - 0.5| = 0.16, 0.14$ , and  $0.13$ , respectively). At high dynamical noise, these correlations disappeared, with all EWS metrics exhibiting  $|AUC - 0.5| \leq 0.05$ .

A full characterization of the strength of association between each metric and emergence, across all diagnostic tests and noise structures, can be seen in supplement (Supplemental Figure D.2, Supplemental Figure D.3, Supplemental Figure D.4, Supplemental Figure D.5).

Table 5.4:  $|AUC - 0.5|$  for EWS metrics, ranked in descending order of magnitude, computed on the subset of the emergent time series after the burn-in period, for a perfect test and an imperfect diagnostic test with 90% sensitivity and 90% specificity, under high and low Poisson and dynamical noise systems

Rank	Perfect Test	90% Sensitive & Specific Imperfect Test			
		Poisson Noise		Dynamical Noise	
	All Noise	Low	High	Low	High
1	Autocovariance (0.20)	Autocovariance (0.23)	Autocovariance (0.22)	Autocovariance (0.16)	Mean (0.05)
2	Variance (0.20)	Variance (0.21)	Mean (0.20)	Variance (0.14)	Variance (0.04)
3	Mean (0.18)	Mean (0.20)	Variance (0.18)	Mean (0.13)	Autocovariance (0.03)
4	Index of dispersion (0.13)	Index of dispersion (0.17)	Index of dispersion (0.18)	Index of dispersion (0.09)	Coefficient of variation (0.02)
5	Autocorrelation (0.12)	Autocorrelation (0.17)	Coefficient of variation (0.17)	Autocorrelation (0.07)	Skewness (0.01)
6	Coefficient of variation (0.11)	Coefficient of variation (0.10)	Autocorrelation (0.16)	Skewness (0.06)	Autocorrelation (0.01)
7	Skewness (0.10)	Skewness (0.08)	Skewness (0.10)	Coefficient of variation (0.05)	Kurtosis (0.01)
8	Kurtosis (0.03)	Kurtosis (0.05)	Kurtosis (0.07)	Kurtosis (0.01)	Index of dispersion (0.00)

## Predictive Ability

Each alert scenario (the combination of diagnostic test, noise structure and magnitude, and EWS metric) produced its optimal accuracy at a different combination of EWS hyperparameters (the quantile threshold of the long-running metric distribution to be exceeded to return a flag, and the number of consecutive flags required to trigger an alert) (Supplemental Figure D.6, Supplemental Figure D.7, Supplemental Figure D.8, Supplemental Figure D.9). At their respective maximal accuracies, the relative ranking of the EWS metrics computed with a perfect diagnostic test remained consistent to the ranking based upon  $|AUC - 0.5|$ : mean (accuracy = 0.72), variance (0.72), autocovariance (0.70), index of dispersion (0.63), autocorrelation (0.62), skewness (0.60), kurtosis (0.58), and coefficient of variation (0.50) (Supplemental Table D.2).

When EWS metrics were computed on time series generated from imperfect diagnostic tests, each metric's accuracy generally remained constant, with a few notable exceptions (Figure 5.1, Supplemental Figure D.1). For the 4 most correlated metrics (autocovariance, variance, mean, and index of dispersion), the accuracy achieved with imperfect diagnostic tests was comparable for low and high Poisson noise, for all diagnostic test accuracies (Figure 5.1). The accuracy of outbreak detection using index of dispersion increased with decreasing diagnostic test sensitivity and specificity for low and high levels of Poisson noise (Figure 5.1, Supplemental Figure D.6, Supplemental Figure D.7). For low dynamical noise, accuracy increased slightly for diagnostic test sensitivity and specificity greater than 97% and then declined (Figure 5.1). For high dynamical noise, accuracy declined monotonically with decreasing test sensitivity and specificity (Figure 5.1, Supplemental Figure D.9). Results for the 4 least well correlated EWS metrics are presented in the supplement (Supplemental Figure D.1).

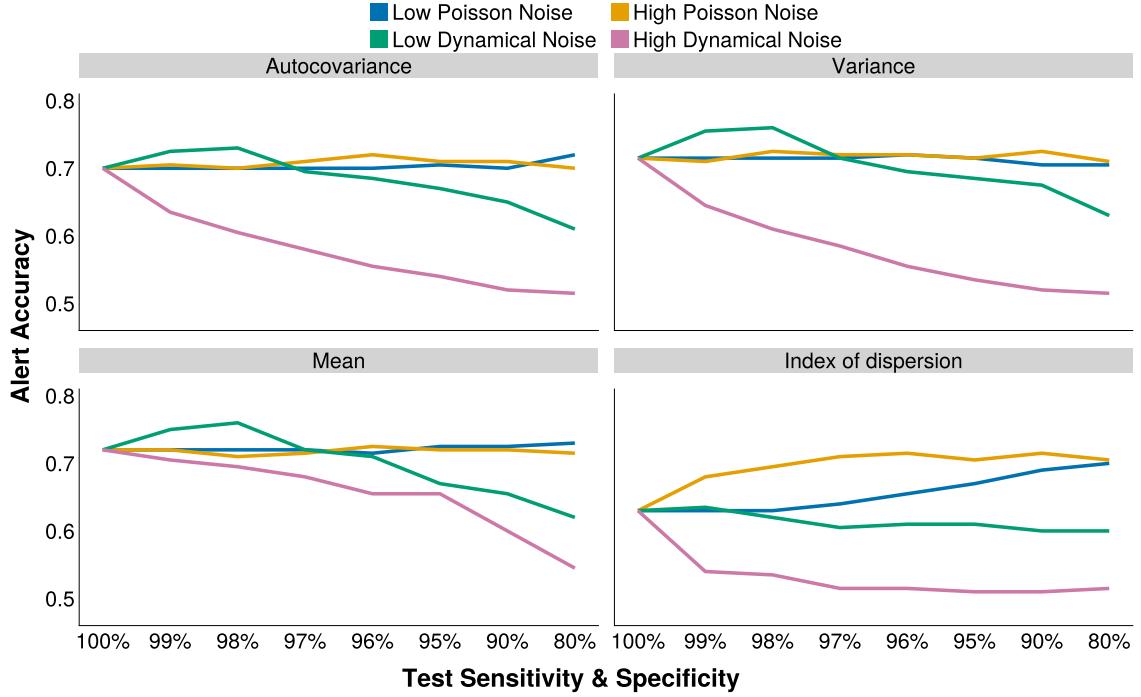


Figure 5.1: The change in alert accuracy for the most correlated EWS metrics under increasing diagnostic uncertainty, and low and high levels of Poisson or dynamical noise. Low noise refers to simulations where the average incidence of noise is equal to the average incidence of measles. High noise refers to simulations where the average incidence of noise is equal to 7 times the average incidence of measles. The tests sensitivity equals the test specificity for all diagnostic tests.

Outbreak detection produced false positives under the null simulations for all EWS metrics, except for the coefficient of variation computed on time series resulting from perfect tests, which also failed to alert in emergent simulations. Here we illustrate the comparison of timing of alerts for the autocovariance metric for the null and emergent simulations (Figure 5.2). Outbreak detection using the autocovariance metric resulted in comparable timing of alerts for perfect and imperfect tests under low and high Poisson noise (Figure 5.2). For low dynamical noise, the imperfect test resulted in a similar number of true positives under the emergent scenario, but tended to trigger those alerts later than with a perfect test. Notably, an imperfect test resulted in more false positives under the null scenario and tended to trigger those alerts later. With high dynamical

noise, an imperfect test failed to produce many alerts under either the null or emergent scenarios (Figure 5.2).

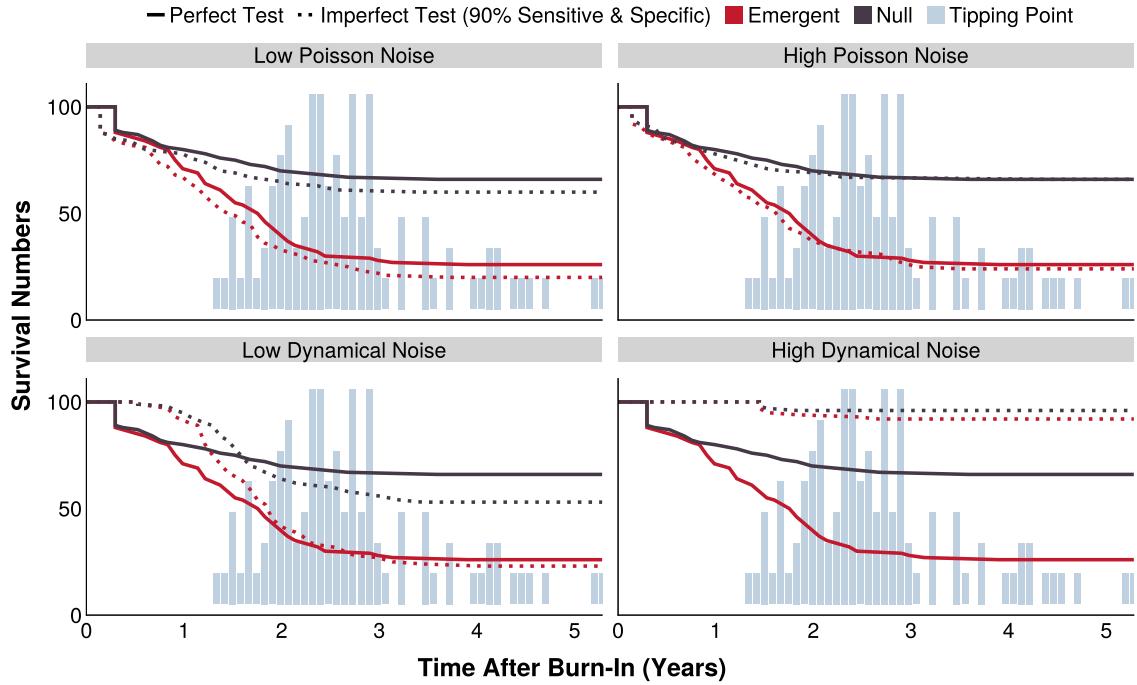


Figure 5.2: Survival curves for the autocovariance EWS metric computed on emergent and null simulations, with a perfect test, and an imperfect test that is 90% sensitive and specific. The histogram depicts the times when the tipping point is reached ( $R_E = 1$ ) under the emergent simulation, right-truncating the curves. The trajectory of the solid lines are identical in each facet, as the perfect test is unaffected by noise cases. The histogram is identical between facets as it represents the timing of the tipping points, and all testing scenarios use the same underlying measles simulations that terminate at the tipping point

## Discussion

Outbreak detection using EWS metrics is robust to diagnostic uncertainty depending on the structure and magnitude of the noise due to non-target infections in the surveillance time series. Under Poisson noise, outbreak detection using a time series of test positive cases resulted in similar performance to a corresponding time series with a

perfect diagnostic, regardless of the incidence of the non-target infections. However, when the background noise due to non-target infections in the time series of suspected cases is consistent with a dynamical SEIR-type process (e.g. tends to produce cycles or periods of consistent outbreaks), the accuracy of outbreak detection declines with decreasing diagnostic test sensitivity and specificity and with increasing relative incidence of the non-target infections. Thus, the performance of outbreak warning systems using EWS depends on both the properties of the individual diagnostics used and structure and magnitude of non-target disease incidence, which may vary with the local context.

This analysis was motivated by the case of anticipating the progression to  $R_E > 1$  for measles in the context of other sources of febrile rash; e.g., rubella, parvovirus, and arboviruses such as dengue fever and chikungunya [147,188–190]. For much of the WHO’s African Region, the co-circulation of measles and rubella is common, although there are stark differences in the relative proportion of incidence by country [144]. In Guinea Bissau, for example, the estimated incidence rates of rubella are approximately 9 times that of measles, in Botswana they are similar, and in the Democratic Republic of Congo measles incidence is estimated to be 20 times higher [144]. Imperfect diagnostic tests will not provide equal value to each of these locations. The Democratic Republic of Congo would be a good candidate for the integration of less accurate diagnostic if it allowed for improvements to other aspects of the disease surveillance system e.g., increases in the testing rates and case-finding activities due to lower financial and logistical costs [132]. However, when large rubella outbreaks can produce meaningful peaks in test positive cases resulting from the use of imperfect diagnostics, such as in Guinea Bissau, the EWS metrics struggle to discriminate between emergent and null periods, reducing their utility.

When evaluating the ability for the EWS metrics to accurately discriminate between emergent and null simulations, it is important to contextualize the results with

the system's relative speed and specificity. Alert systems necessarily make compromises in their design: improvements to speed generally come at the cost of increased numbers of false alerts [118,129]. Depending on the context, it may be desirable to place a greater weight in preference/penalty for one of these axes; in scenarios where the expected cost to launch a preliminary investigation is low relative to the unaverted DALYs resulting from incorrect inaction in an overly specific system, higher false alert rates may be acceptable. This analysis provides a framework to explicitly explore these trade-offs, including through the comparison of survival curves. A larger separation at the end of the time series between the emergent and null simulation lines indicates higher alert accuracy, as there is a greater difference in the true positive and false positive rates. Faster and more sustained declines indicate a (relatively) more sensitive alert system with more advanced warning of emergence.

Under the simulation constraints placed here, generally, the use of imperfect diagnostic tests does not increase the speed of the warning for the EWS metrics that are predictive of emergence. This is likely a consequence of imperfect diagnostic tests producing more false positive cases, which, without appropriate penalization, would otherwise lead to high false alert rates under the same EWS hyperparameters. Adjusting the relative weighting of alert sensitivity and specificity used to compute the alert accuracy would allow for an exploration of alternative scenarios. Additionally, multiple hyperparameter combinations can produce identical alert accuracy i.e., some combinations will favor the speed of alert at the expense of its specificity, and *vice versa*. In situations where identical accuracies are achieved, we present the results associated with the most specific alert system i.e., reduce the number of false positive alerts. An equivalent set of analyses could be computed to favor the hyperparameter combinations that produced the fastest and most frequent alerts.

For EWS metrics to reflect the underlying dynamics of critical slowing down, careful detrending of the data is required [169,191,192]. Our analysis utilizes a

backward-facing uniform moving average to detrend the data: this was chosen as it can be easily intuited and implemented in a surveillance system. However, it has previously been stated that insufficient detrending may lead to spurious patterns that do not arise from a system’s dynamical response [169], with some EWS more sensitive than others, and that the associations may vary with the bandwidth size selected [191,192]. While it may be preferred to detrend using the mean over multiple realizations [191], this is clearly not possible in a real-world situation. Future work could explore the effects of different detrending methods (e.g., Gaussian weighting moving average, smaller and/or larger bandwidths) on the effectiveness of EWS metrics in systems with diagnostic uncertainty [191,192]. Similarly, prior work has demonstrated the benefits of constructing composite metrics, for example, calculating a composite metric as the sum of the standardized differences for each of the individual metrics, before defining a threshold for the distribution of the composite [170]. However, there are other techniques that could be applied to produce a composite statistic [177,178], each requiring different decisions as to the appropriate weightings to be assigned to the underlying single metrics. For this reason, we only present the results from the individual metrics to illustrate the effects of diagnostic uncertainty, but future work should aim to extend the approach detailed to composite EWS metrics.

Despite being relatively well-established in areas of study such as ecology, ecosystem collapse, and climate science [167,169–171,180,193,194], the exploration and development of EWS for infectious disease systems is in its relative infancy. Until recently, a large proportion of the prior work in the area has been to establish the existence of these metrics that theoretically could be used in such a system [166,173,174]. While this is a crucial first step, for use in a proactive outbreak alert system, EWS metrics must be able to provide advance warning of the approach to the tipping point  $R_E = 1$ . Correlations alone are not sufficient to indicate when and what actions must be taken. To address this, there is a growing body of work that seeks to

evaluate the use of various threshold and risk-based approaches within infectious disease systems [175,177,178,183]. Our work expands upon these efforts, characterizing the limits of predictability for EWS metrics in systems with diagnostic uncertainty and background noise.

## Acknowledgements

### Funding

This work was supported by the Bill & Melinda Gates Foundation through grant No. INV-016091\_2020. This project was also supported by the National Science Foundation, by grant: NSF-NIH-NIFA Ecology and Evolution of Infectious Disease award DEB 1911962. The funding sources had no role in the collection, analysis, interpretation, or writing of the work.

### Author Contributions

*Conceptualization:* CA, MJF

*Data curation:* CA, MJF

*Formal analysis:* CA

*Funding acquisition:* MJF

*Investigation:* CA

*Methodology:* CA, MJF

*Writing - original draft:* CA

*Writing - review and editing:* all authors.

### Conflicts of Interest and Financial Disclosures

The authors declare no conflicts of interest.

### Data Access, Responsibility, and Analysis

Callum Arnold and Dr. Matthew J. Ferrari had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Callum Arnold (Department of Biology, Pennsylvania State University) conducted the data analysis.

### Data Availability

All code and data for the simulations can be found at <https://github.com/arnold-c/CSDNoise>

# Chapter 6 | Synthesis

There is a growing body of literature that focusses on uncertainty in disease transmission; from incorporating viral dynamics into mechanistic models of disease [134,135], to accounting for reporting and testing uncertainty in estimates of the *real-time* reproduction number ( $R_t$ ) [195]. In this dissertation I demonstrate how the classification of continuous infectious disease variables is both essential to inferences about disease dynamics and their underlying systems, and results in compounding uncertainty that limits our detection and predictive abilities for outbreaks. This work has already shown relevance and impact, helping inform the target product profile (TPP) of a potential future rapid diagnostic test (RDT) for measles in outbreak surveillance settings [136].

## Discretization of Risk Groups

In the first half of my dissertation I explore mechanisms by which populations can be classified to understand the transmission of COVID-19 within and between Pennsylvania State University students and community members of its surrounding county (Center County). Clear definitions of risk and transmission groups provide a natural mechanism to explore the heterogeneity in infection that may exist in a population. The differences with respect to many drivers of infection e.g., demography, contact rates, perceptions of infection risk, willingness to take preventative actions, between university students and county residents imply differences in risk. However, the high spatial proximity of these two well-defined groups that were both immunologically naive to an emerging pathogen might have overwhelmed differences in other drivers of infection, and resulted in similar exposure rates. We show that,

contrary to the assumptions at the time, there was substantial variation in exposure in these geographically coincident populations (Chapter 2). This, supported by evidence that the Center County community experienced lower per-capita incidence rates than its 5 surrounding counties [63], implies that intervention efforts by the University were able to minimize the risk of onward transmission from the student population. If between-group transmission did occur, it was likely transient in nature. Within the study body, the only factors associated with infection outcome were the recent contact with a known COVID-19 positive individual, and the attendance of gatherings.

In light of these findings, we hypothesized that, in the absence of pharmaceutical interventions, seroprevalence differences were driven by heterogeneity in behavior. We also theorized that similar differences in infection rates may exist *within* each cohort, with each group being inhomogeneous behaviorally. However, unlike in Chapter 2 where clearly-defined exposure groups were pre-existent, no clear demarcations within the cohorts existed. To make inferences about the heterogeneity in transmission, I sought to categorize the student body with respect to a latent (unobservable) variable: risk behavior. Given behavioral survey data of intentions to adhere to non-pharmaceutical public health measures (PHMs), I clustered the students using Latent Class Analysis (LCA). Doing so returned a probability of class assignment for each individual, along with the propensity for each class (to intend) to follow or not follow each of the individual PHMs. Calculating the observed seroprevalence of these strata demonstrated that the behavioral survey data were able to define meaningfully different groups with respect to both the student behaviors and their infection outcomes. Through discretizing the population, it became possible to evaluate the potential effectiveness of interventions aimed at increasing adherence to PHMs. Defining risk behavior groups provided group-specific seroprevalence rates that could be used to parameterize a mathematical model of transmission. Doing so placed realistic bounds on the expected benefit of an intervention, which did not rely upon *a priori*

assumptions as to the intervention's effectiveness in reducing transmission. As a large proportion of the population were already in the most adherent group that always intended to follow public health guidance, interventions targeted at them would serve no direct effect: their risk of transmission was dictated by interactions with less-adherent individuals.

If we are to design effective interventions, it is essential to first characterize the limits of what can realistically be achieved. For example, supplemental immunization activities (SIAs) aimed to minimize the effects of an outbreak, reactively or prospectively, are partially limited by the vaccination coverage that can be achieved. While SIAs can achieve modest improvements over routine immunization coverage, individuals who are able to access the care provided by SIAs are more likely to also be individuals who can attend routine immunizations, minimizing the potential impact of expanded efforts through traditional mechanisms [196–198]. Incorporating access and dose redundancy data can provide more realistic estimates on the effectiveness of future efforts. Similarly, incorporating behavioral information into epidemiological models can shed light on the potential effectiveness of vaccination campaigns. Defining latent groups of individuals by vaccine-seeking or hesitancy behaviors, it would be possible to calculate an expected uptake, augmenting estimates provided through data on accessibility. This could allow for a more realistic estimate of intervention effectiveness at reducing disease burden, characterizing a limit for coverage that incorporates previously observed aggregate information (the number of redundant doses provided), and individual intention behavior. For example, this may identify situations where individuals will choose not to be vaccinated, despite access expanding to incorporate them. It is possible that vaccine-intentions differ between groups with and without access, so efforts improve access alone may overestimate the protection benefits gained. Modeling these latent sentiments offers a mechanism to account for these discrepancies. Not only does this approach potentially provide more accurate

projections of the limits of vaccination coverage, it may also improve disease burden projections by describing *who* can be immunized; overlap between access and healthcare-seeking behavior with disease vulnerability will highlight gaps that need to be overcome.

This approach can also shed light on effectiveness of non-pharmaceutical interventions in heterogeneous populations, where traditional measures derived from demographic data, such as age-based mixing matrices, may over-simplify the dynamics. Characterizing complex network structures has been a point of focus in the study and prevention of sexually transmitted infections (STIs) [11,199,200]. However, the process if often laborious, requires the computation of difficult-to-calculate network measures like partner concurrency, and is subject to change [201–203]. In such systems heterogeneity exists across multiple facets, which may not be adequately captured by standard discretization. The incorporation of different data sources and categorization approaches may serve as a beneficial middle ground, allowing for a more accurate depiction of the true heterogeneity in disease risk than provided by age classes and demographic data, while also requiring less hands-on data collection than mapping contact networks; behavioral surveys can be implemented online, and pose potentially less sensitive questions than those necessary to build a sexual network, which may reduce a barrier to participation [204]. This is particularly important when designing interventions for novel pathogens, where information is limited, prior observations may not hold and are likely to change over time, and urgent actions are necessary to minimize the impact [205,206].

## Discretization of Infection Status

It is imperative to evaluate the effects of discretizing not only exposure classes, but also the outcomes of an infectious disease system. In public health surveillance programs, incidence of disease occurrence must be tabulated before the need for, and level of, action can be decided upon. Uncertainty in the observational process can occur through two mechanisms: uncertainty in who and how many individuals are counted (observed and tested), and uncertainty in the underlying test results. Incorporating uncertainty in the reporting of cases has been of increasing interest in the parameterization of mathematical models, particularly for the estimation of  $R_t$  [134,195,207,208] and disease burden [157,209,210]. However, the accuracy of the diagnostic used to make this determination affects the incidence, necessarily turning a quantitative input (pathogen load/host response) into a binary value, with associated classification errors. Previous inclusions of diagnostic uncertainty solely address under-reporting as only one pathogen is simulated, removing the opportunity for false-positive test results [134,135,195,207,208]. Here I address the potential impact of false positives by simulating a two-pathogen system.

To account for diagnostic uncertainty, in Chapter 4 and Chapter 5 I simulated both a target pathogen and background noise, using representative parameters of measles and rubella, respectively. Through variations to the rubella vaccination coverage, I was able to evaluate the degradation of outbreak detection performance with tests of decreasing sensitivity and specificity. When the background noise did not exhibit large peaks and troughs i.e., was drawn from a Poisson distribution or when there was sufficiently high vaccination coverage in the rubella dynamical noise simulations, both imperfect and perfect diagnostic tests could adequately discriminate between outbreak and non-outbreak periods. However, this was not the case at higher levels of dynamical noise ( $\approx 6$  times, or greater, than the average incidence of measles). In these situations

the reduced diagnostic accuracy of imperfect tests could not be alleviated through increasing the testing rate. Across the WHO's African Region, there is typically far more circulating measles than rubella; in 2023, the average incidence of measles was 60.3 cases per 1M population, in contrast to 5.2 rubella cases per 1M population [144]. In general, this would point to the widespread viability of imperfect diagnostic tests within infectious disease surveillance systems for the purposes of outbreak detection. However, there is great variability in the incidence rates by country. At the one end, Burkina Faso experienced measles incidence of 69.7 measles cases per 1M population, relative to 0.4 rubella cases per 1M; testing 80% of reported cases, there were 4.2 times as many IgM positive and epidemiologically-linked measles cases as other clinically-compatible infections (i.e. rubella or otherwise) [144]. As a result, imperfect diagnostics would likely be able to accurately discriminate between outbreaks of measles from changes in the incidence of other sources of febrile rash. In contrast, Zimbabwe reported rubella incidence that was approximately 7.3 times higher than measles incidence, presenting a location where the increased uncertainty of imperfect diagnostics would result in poor outbreak detection [144]. Lastly, Eritrea experienced 5.0 measles cases, and 1.3 rubella cases, per 1M population [144]. While this may appear to be a good candidate for the use of imperfect tests, the incidence rate for clinically compatible cases that were not measles or rubella was 21.3 cases per 1M population [144]. As a result, before imperfect diagnostic tests could be implemented with confidence, a careful evaluation of the dynamics of the non-rubella background noise cases would be required; if it demonstrated large episodic outbreaks, its scale in relation to the expected measles incidence would necessitate the use of high-accuracy diagnostic tests.

In Chapter 5, I build off the work in the previous chapter to explore the effects of the outlined diagnostic uncertainty on our ability to predict risk of future outbreaks. Traditional outbreak detection that utilizes the exceedance of an incidence threshold is

necessarily reactive in nature. Under ideal circumstances, it would be possible to measure trends in summary statistics derived from infection data to infer the emergence of outbreaks before they occur. This would allow for proactive actions that could avert the most cases, reducing the morbidity and mortality resulting from a pathogen. Prior work has demonstrated the viability of early warning signals (EWS), albeit only accounting for errors stemming from testing rates, not diagnostic uncertainty [164,165,175,183]. I demonstrate that, similar to reactive outbreak detection, predictive systems can be designed around the use of imperfect diagnostics e.g., RDTs, for case identification, so long as the magnitude of dynamical noise is low relative to the incidence of the target pathogen. Not all EWS metrics performed well, but, aligning with the literature, the mean, variance, autocovariance, and index of dispersion were able to discriminate between emergent and non-emergent time series in these situations. Additionally, the evaluation of EWS performance required that the alert in emergent simulations occur before the tipping point  $R_E = 1$  to be considered successful. As this tipping point essentially acts as a necessary precursor to an outbreak (though exceptions can occur due to the stochastic nature of infectious disease transmission), each warning would be provided with sufficient time for action to potentially avert an outbreak.

The primary focus of this section of work (Chapter 4 & Chapter 5) has been to motivate new approaches to the design of surveillance systems at large, not any one specific implementation. These two chapters illustrate the complexity in evaluating the performance of reactive and proactive outbreak detection systems. The system's accuracy is dependent on the interaction between multiple axes: the proportion of individuals tested, the diagnostic test characteristic, the shape and magnitude of the background noise, the and the criteria required to trigger an alert. Each of which can be adjusted, and without consideration of the knock-on effects, can easily lead to a suboptimal path being followed. We found that similar accuracy could be achieved for a

range of conditions by balancing traits of each of the levers of the surveillance system. By relaxing prior constraints on the outbreak detection system, such as the number of test positives (Chapter 5) or the maximum number of consecutive data points (Chapter 5) required to trigger an outbreak alert, for example, it is possible for imperfect diagnostic tests to obtain outbreak detection performance that mirrors performance under perfect tests. In doing so, this opens the opportunity for the development and use of cheaper, less labor intensive, and logistically simpler diagnostics, such as (rapid) lateral flow diagnostic assays [130–132]. Further, we identified conditions where accuracy was fundamentally limited; even with a perfect test, EWS occasionally trigger under null simulations due to endemic and imported cases. These boundaries can only be illustrated through comprehensive exploration across all axes of the surveillance system.

This approach provides public health personnel and policy makers the ability to consider all options and “optimize” the surveillance system to best suit their needs. Trade-offs and the balance of priorities must be evaluated on a context-specific basis. There has long been a tension in designing disease surveillance programs: the needs of an individual may be at odds with those of the wider population. During the early stages of the COVID-19 pandemic, when vaccines were first being developed and their potential effectiveness unknown, there were discussions around who should be prioritized during initial roll-outs: the elderly and those with known co-morbidities, who would most likely receive the largest direct benefit from immunization; or younger individuals with larger numbers of contacts, whose vaccination would most likely result in the largest reduction in incidence, with an indirect benefit to vulnerable individuals [7]. For infectious disease outbreak surveillance, a similar conflict exists. Systems are often built upon routine, passive, surveillance that uses health facility visits for case identification [109,211]. As a result, reducing the accuracy of the diagnostic test used may be associated with a corresponding reduction in cost and technical requirements,

allowing for a greater proportion of the population to be tested. This, in turn, may improve outbreak detection, providing an indirect effect to any specific individual, but at the expense of a more accurate diagnosis and care provided to the patient seeking treatment at a healthcare facility. In a region with high disease burden and large period outbreaks, this may be an acceptable trade-off; in other locations, unacceptable.

In addition to ethics-based decisions, countries must decide how to balance the relative costs and benefits of more sensitive versus specific alert systems. My analysis does not represent a true optimization; the partial observation of the system necessitates decisions and actions be made on the basis of incomplete information. In particular, the true outbreak status (and  $R_E$ ) are never known, and can only be inferred. As a result, incorrect inferences will be made and the need to prioritize the speed of outbreak detection and response against the false positive rate of alerts will depend upon the costs associated with each, which changes by region, and over time. In Chapter 4 & Chapter 5 I provide equal weight to the associated alert speed and specificity metrics utilized in the system's evaluation. In locations that experience large, devastating, outbreaks, where response mobilization may be heavily delayed, a greater premium could be placed on sensitive alert systems, if only to launch an active preliminary investigation. Furthermore, the evaluation of each alert threshold, be that incidence-based alert triggers for reactive surveillance programs, or EWS-based approaches for proactive systems, was conducted at the 'optimal' set of hyperparameters. This approach requires mapping the performance of each test across a wide range of possible parameter values and combinations. While feasible in a simulation scenario, it poses a challenge for empirical use. Careful selection of 'training data' would be required to emulate this approach in a real-world setting to ensure representative time series are utilized, producing appropriate parameter values.

Unlike in simulations, outbreak preparation and response scenarios often impose additional constraints, such as the financial resources available. Explicitly incorporating

the effects of these constraints may limit the space of attainable alert performance, potentially disproportionately for particular diagnostic tests or background noise magnitudes. The work in Chapter 5, in particular, highlighted that not all outbreak detection scenarios are equally robust: the ‘mean’ EWS metric provided slightly suboptimal alert performance under most scenarios, relative to the autocovariance and variance, but was more resilient to higher levels of dynamical noise given the use of moderately accurate RDTs. The analysis approach detailed here provides a mechanism to compare these constraints to find zones of acceptable performance.

In the future, efforts should be made to formally integrate the design and implementation of outbreak surveillance and early warning systems into partially observed Markov decision process (POMPD) models. Doing so would provide a more complete characterization of the uncertainty that develops and propagates throughout the numerous observational and decision processes. My work has demonstrated that even in scenarios with perfect diagnostic tests, detecting and predicting outbreaks relative to baseline infection dynamics is exceedingly difficult. Generally, a binary threshold is used to ascribe infectious status to individuals, despite prior work demonstrating that infectiousness is not a binary state and the timing of diagnostic test within an infection cycle affects its accuracy [134,135,154]. Incorporating this aspect of the observational process into models would generate realistic biases in data generation, and provide a meaningful extension to the analysis presented.

Further, with the rise in machine-learning and big-data driven approaches to outbreak detection, a more thorough exploration of the uncertainty resulting from discretizing the prediction target is required. In surveillance systems, we never know the true state of  $R_E$ , or how to categorize a time series into outbreak and non-outbreak periods, both of which are the target for predictive algorithms. It is conceivable that biases in the observational process arising from false negative, and importantly, false

positive test results, impacts the ability of  $R_t$  estimates to approximate  $R_E$ . Exploration of these issues should be the target of future work.

## Conclusion

Without categorization, understanding disease systems and decision-making can become intractably complex. To paraphrase “Del rigor en la ciencia” by Jorge Luis Borges; “a map of everything is a map of nothing”. Through discretization we can describe trends in disease burden, discover emergent risk groups, and plan targeted actions to most efficiently use limited resources. But the choices we make to define breakpoints and strata introduce challenges that must be addressed, particularly at the boundaries. With careful, intentional, investigation, it is possible to balance the trade-offs that arise from the uncertainty. Throughout my dissertation I characterize the benefits and pitfalls of discretizing continuous phenomena, and present novel approaches to integrate across scales. This topic presents many exciting avenues for future research, with those that address findings from all levels of the infection observation process offering the most opportunity to minimize burden and save lives.

# **Appendix A | Supplementary Material for Chapter 2**

## **Laboratory Methods:**

### **Production and Purification of SARS-CoV-2 Receptor-binding domain (S/RBD)**

Transfections of plasmid pSL1510 (pCAGGS-RBD from Florian Krammer, Mount Sinai, USA) was performed using the Expi293 Expression System from ThermoFisher. Cells were cultured per manufacturer's instructions (37°C, 8% CO<sub>2</sub>, in shaker flasks at 120-130 rpm), and the supernatant was harvested by simple centrifugation on the third day for downstream processing. Cell viability and concentration were monitored throughout to ensure that the culture remained in log phase growth. The detailed protocol is deposited in protocols.io [48]. Briefly, culture supernatant was incubated with pre-equilibrated Ni-NTA resin in 1X PBS at 4°C for 1 h on a nutator, after which a gravity column was used to elute the protein.

### **Estimation of IgG antibodies against SARS-CoV-2**

An in-house indirect isotype-specific (IgG) ELISA against SARS-CoV-2 receptor-binding domain (S/RBD) was developed [48]. Commercially purchased human monoclonal antibody reactive to spike regions of SARS-CoV-1 and SARS-CoV-2 were used as positive controls in the assay (Two isotypes of CR3022, IgG1: Ab01680-10.0; Absolute Antibody, USA). The cut-off for this IgG ELISA was determined as an optical density (absorbance at 450 nm) higher than six standard deviations above the mean of the tested pre-COVID-19 serum samples (n=100). Briefly, serum was separated from the blood collected from study participants and inactivated at 56°C for 30 minutes. Microtiter plates were coated with purified recombinant S/RBD. Negative serum control

was included on each microtiter plate. 1:50 dilutions of serum were added, incubated for 1 hour, washed, incubated with goat anti-human IgG (Fc specific) (A0170, Sigma-Aldrich, USA), and washed. 3,3',5,5'-Tetramethylbenzidine dihydrochloride (TMB) was used as the ELISA substrate (T3405, Sigma-Aldrich, USA) was added, the plates were developed until the top dilution reached the saturation pointes, and the reaction was stopped with H<sub>2</sub>SO<sub>4</sub>. Plates were read at an absorbance of 450 nm.

To further evaluate the performance of this assay, we used a total of 92 convalescent plasma samples (RT-PCR positive individuals). PCR data was used here as the comparator method (gold standard) to establish clinical truth for all samples, showing a 90% sensitivity, 100% specificity, 100% positive predictive value (PPV), and 92% negative predictive value (NPV). Similarly, comparing outcomes from 200 virus neutralization assays showed a 98% sensitivity, 96% specificity, 98% PPV, and 98% NPV [49].

## Statistical Methods

### Treatment of Missing Data

In the subset of individuals in the returning student subgroup that had ELISA results, there are few missing values for the model variables, with the exception of “working as a service professional” (421/684). As a result of high missingness, service professional was removed as a predictor in the model. Exploration of the missing values in the remaining predictor variables demonstrate no bias by outcome, confirmed using Chi-squared tests of missingness in predictors by outcome level. Little’s test of Missing Completely At Random (MCAR) indicated that the data was not MCAR ( $p = 0.0728$ ) [212], and three imputation methods (MICE, k-Nearest Neighbour with 5 neighbours,

and Bagged Tree) [55,213,214] were used to compare model fits (Supplemental Supplemental Figure A.1). Most missing values occurred across all variables, and there was no observable pattern among the majority of variables: there was some evidence that missingness in “travel in the 3 months prior to return” was associated with “travelling since campus return” response, and that missingness in “eaten in a restaurant in the past 7 days” was associated with “IgG classification”. As such, the predictor variables were deemed to be ‘Missing At Random’, and MICE was used to impute missing values.

### **Alternative Estimate of True Prevalence**

In the main text we present estimates of the true prevalence in the returning student and community resident cohorts that corrects for the sensitivity of the assay. We estimated sensitivity based on the returning student samples only because the student population had high access to RT-PCR diagnostic tests. Here we present an alternative analysis using an estimate of sensitivity including the community residents. 9 community residents self-reported a positive COVID-19 diagnosis by a medical professional prior to the first visit; an additional 19 community residents reported a positive COVID-19 diagnosis between the first and second visit. Of these, 17 were positive for IgG antibodies. Pooled with the student results, this results in a sensitivity estimate of 0.89 (95% CI: 0.82-0.94). This implies a lower sensitivity in the community resident participants, though the number of observations is low. Supplemental Supplemental Figure A.2 shows the estimated true prevalence assuming a uniform prior on the interval (0.82, 0.94) on sensitivity in the community resident population and a uniform prior on the interval (0.85, 0.99) on specificity. For all values of specificity greater than 0.85, there is no change in the qualitative result that the 95% confidence intervals for prevalence in the community residents overlap for both visits for

specificity values less than 0.95, and are distinctly different to the prevalence within the returning student subgroup.

### **Comparison of Community Member Infections by Similarity To Student Cohort**

Given the spread of community resident ages and household incomes, we examined the seroprevalence among community members of a similar age and household income as the students (age  $\leq 30$ y and household income  $\leq 50$ k USD) and compared the seroprevalence against the rest of the community cohort (age  $\geq 30$ y or household income  $\geq 50$ k USD). If risk of infection was correlated with age and income status, rather than student status, we would expect to see higher seroprevalence in this subset of community residents. There were no differences in Wave 1 or Wave 2 seroprevalence, or Wave 2 cumulative seroprevalence ( $p = 0.142$ ,  $p = 1$ ,  $p = 0.691$ , respectively) (Supplemental Supplemental Table A.1, Supplemental Table A.2, Supplemental Table A.3).

## **Tables**

Supplemental Table A.1: Wave 1 seroprevalence among community cohort members that are similar/not similar in age ( $\leq 30$ ) and household income ( $\leq 50$ k USD p.a.) to returning students.

Wave 1 Assay Result	Not Similar to Students (N=1209)	Similar to Students (N=104)
Negative	1173 (97.0%)	98 (94.2%)
Positive	36 (3.0%)	6 (5.8%)

Supplemental Table A.2: Wave 2 seroprevalence among community cohort members that are similar/not similar in age ( $\leq 30$ ) and household income ( $\leq 50k$  USD p.a.) to returning students.

Wave 2 Assay Result	Not Similar to Students (N=1209)	Similar to Students (N=104)
Negative	1138 (94.1%)	98 (94.2%)
Positive	71 (5.9%)	6 (5.8%)

Supplemental Table A.3: Wave 2 cumulative seroprevalence among community cohort members that are similar/not similar in age ( $\leq 30$ ) and household income ( $\leq 50k$  USD p.a.) to returning students.

Wave 2 Assay Result	Not Similar to Students (N=1209)	Similar to Students (N=104)
Negative	1122 (92.8%)	95 (91.3%)
Positive	87 (7.2%)	9 (8.7%)

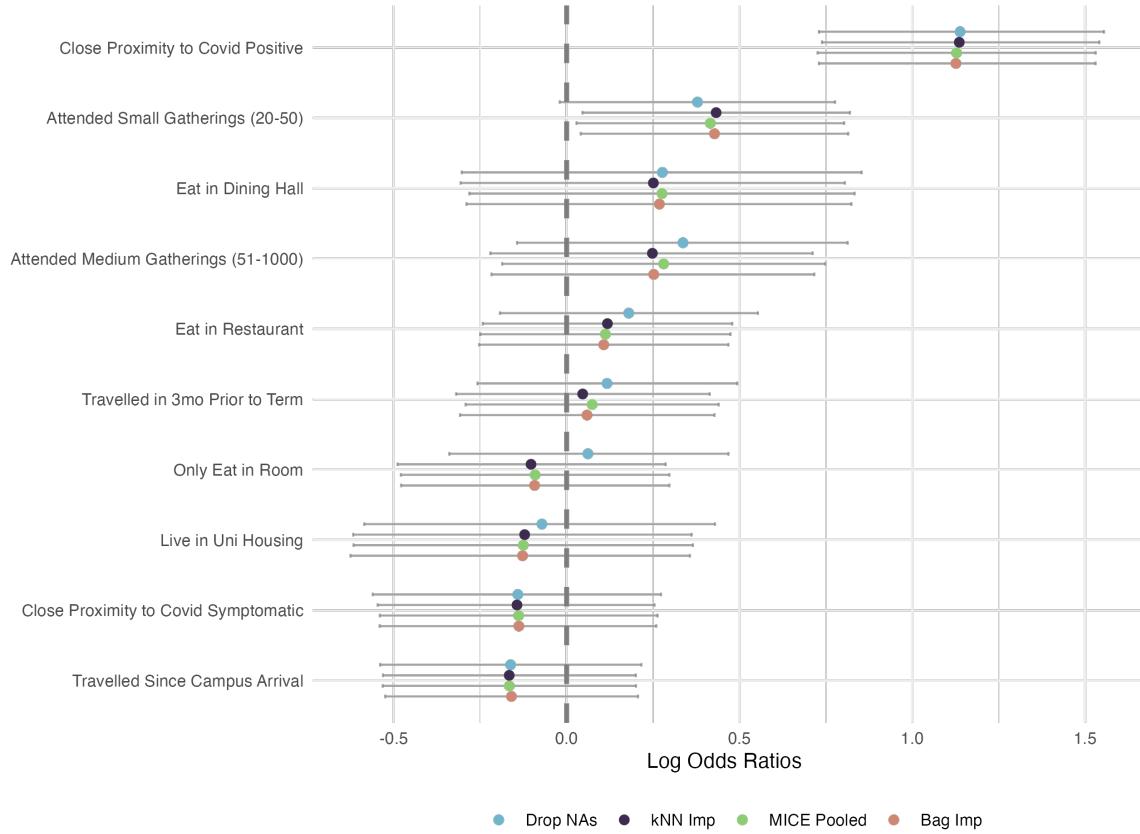
Supplemental Table A.4: Propensity of following public health measures in returning students and community members with PSU ELISA results; subset of community members that received the “Health Messaging” survey. P-value refers to Chi-square test with Yates’ continuity correction of proportions in the predictor level by cohort.

PH Measure		Community - Health Messaging	Returning Students	p
Total N (%)		835 (55.0%)	684 (45.0%)	
Mask Wearing	Always	633 (76.1%)	593 (87.0%)	< 0.001
	Not Always	199 (23.9%)	89 (13.0%)	
Distancing in Public	Always	249 (30.0%)	198 (29.1%)	0.749
	Not Always	582 (70.0%)	483 (70.9%)	
Avoiding Crowds of $\geq 25$ People	Always	549 (65.8%)	293 (43.0%)	< 0.001
	Not Always	285 (34.2%)	389 (57.0%)	

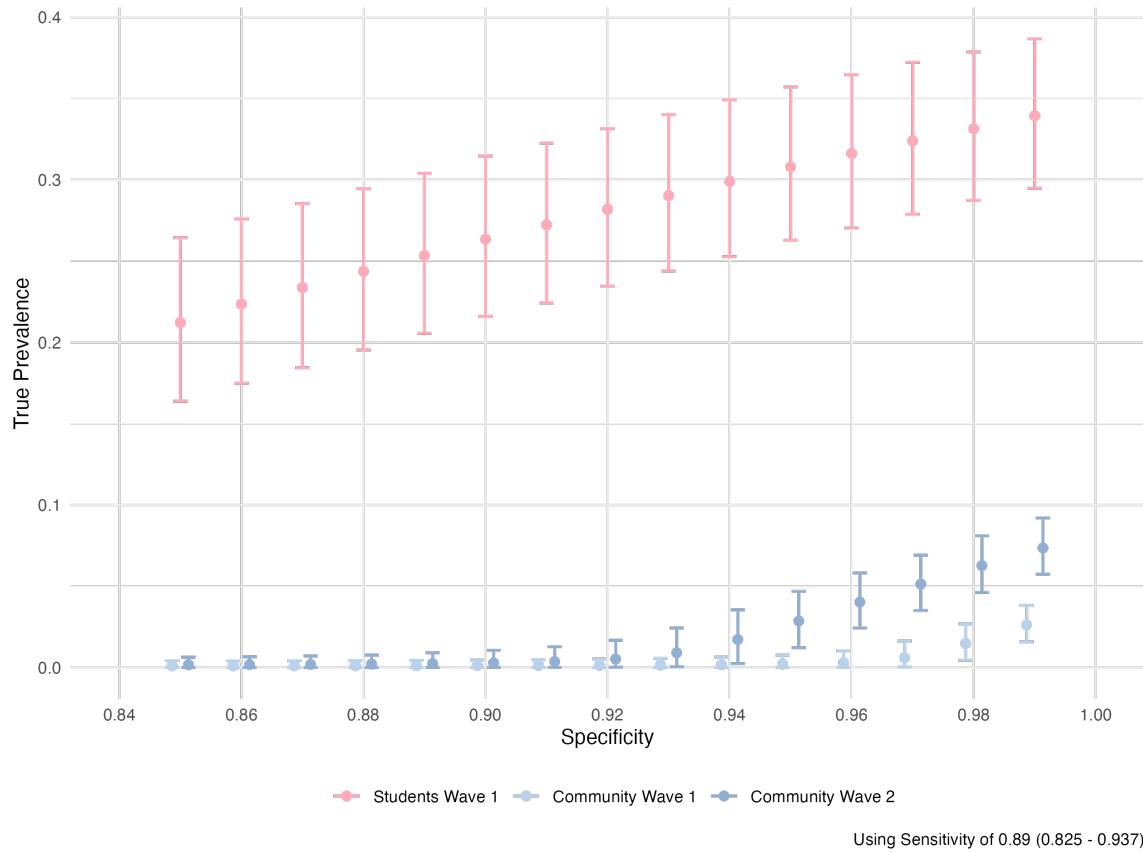
Supplemental Table A.5: Raw prevalence in each subgroup by adherence to PH public health

PH Measure		Community - Health Messaging		Returning Students	
Total N (%)	Adherence	Negative (N=804)	Positive (N=31)	Negative (N=476)	Positive (N=208)
Mask Wearing	Always	610 (75.9%)	23 (74.2%)	410 (86.1%)	183 (88.0%)
	Not Always	191 (23.8%)	8 (25.8%)	65 (13.7%)	24 (11.5%)
	Missing	3 (0.4%)	0 (0%)	1 (0.2%)	1 (0.5%)
Distancing in Public	Always	242 (30.1%)	7 (22.6%)	150 (31.5%)	48 (23.1%)
	Not Always	558 (69.4%)	24 (77.4%)	324 (68.1%)	159 (76.4%)
	Missing	4 (0.5%)	0 (0%)	2 (0.4%)	1 (0.5%)
Avoiding Crowds of $\geq$ 25 People	Always	530 (65.9%)	19 (61.3%)	219 (46.0%)	74 (35.6%)
	Not Always	273 (34.0%)	12 (38.7%)	256 (53.8%)	133 (63.9%)
	Missing	1 (0.1%)	0 (0%)	1 (0.2%)	1 (0.5%)

## Figures



Supplemental Figure A.1: Imputation method comparison among returning students



Supplemental Figure A.2: Sensitivity analysis of true prevalence amongst returning student and community subgroups, using pooled estimate of IgG test sensitivity against self-reported prior positive test

## Appendix B | Supplementary Material for Chapter 3

### LCA Model Fitting

Supplemental Table B.1: Class-conditional item response probabilities shown in the main body of the table for a four-class LCA model, with footers indicating the size of the respective classes, and the class-specific seroprevalence

<b>Measure Intention to Always:</b>	<b>Low Adherence</b>	<b>Low- Medium Adherence</b>	<b>Medium- High Adherence</b>	<b>High Adherence</b>
Wash my hands often with soap and water for at least 20 seconds.	0.04	0.38	0.93	0.95
Wear a face cover (mask) in public	0.11	0.88	0.88	0.99
Avoid face-touching with unwashed hands	0.00	0.00	0.62	0.85
Cover cough and sneeze	0.22	0.77	1.00	1.00
Stay home when ill	0.06	0.82	0.85	0.99
Seek medical attention when have symptoms and call in advance	0.02	0.68	0.75	0.98
Stay at least 6 feet (about 2 arms lengths) from other people when outside of my home.	0.00	0.22	0.10	0.92
Stay out of crowded places and avoid mass gatherings > 25 people	0.02	0.46	0.23	0.92
Tested for COVID-19 twice or more	0.76	0.81	0.84	0.81
<b>Group Size</b>	13.82%	30.91%	16.49%	38.78%
<b>Seroprevalence</b>	35.50%	31.20%	36.00%	25.70%

## Matrix Structure Sensitivity Analysis

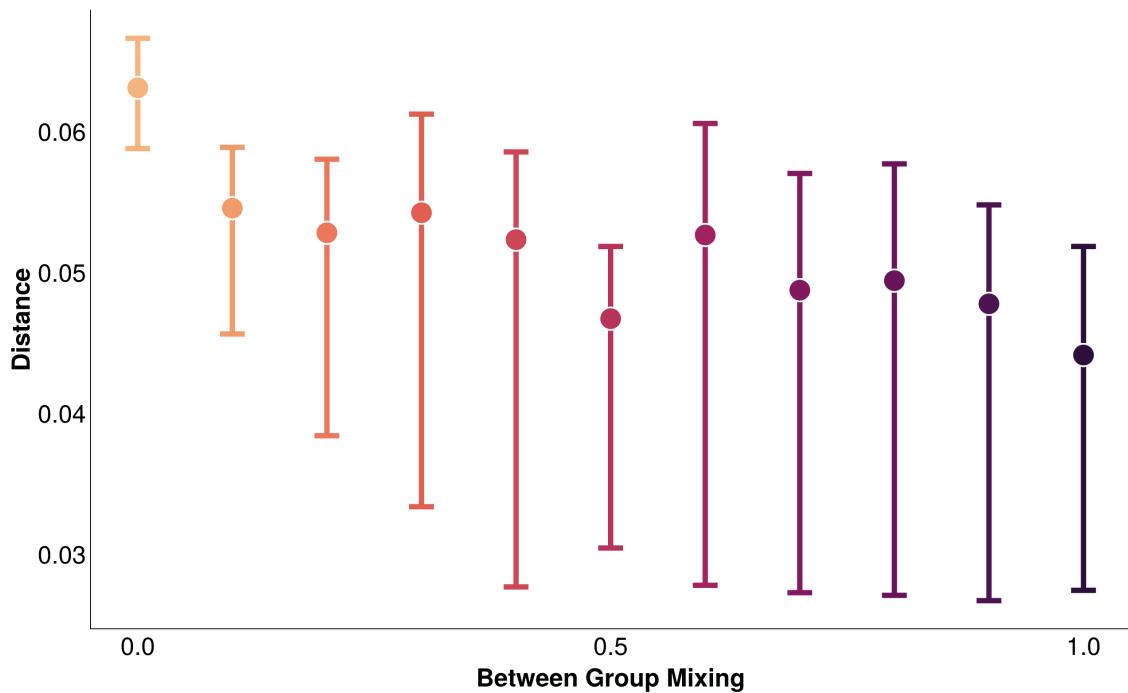
In the main body of the text, we present the results for the three-class model that corresponds to a scenario where public health measures (PHMs) reduce onwards risk of transmission (Equation 3.1) Two alternative were also simulated.

1. A transmission matrix that implies PHMs confer protection for the practitioner (Supplemental Equation B.1).
2. A transmission matrix that uses a single scaled value of  $\beta_{LL}$ , thereby all between-group interactions experience the same risk of transmission. This risk is a fraction of the transmission observed between Low Adherence individuals (Supplemental Equation B.2).

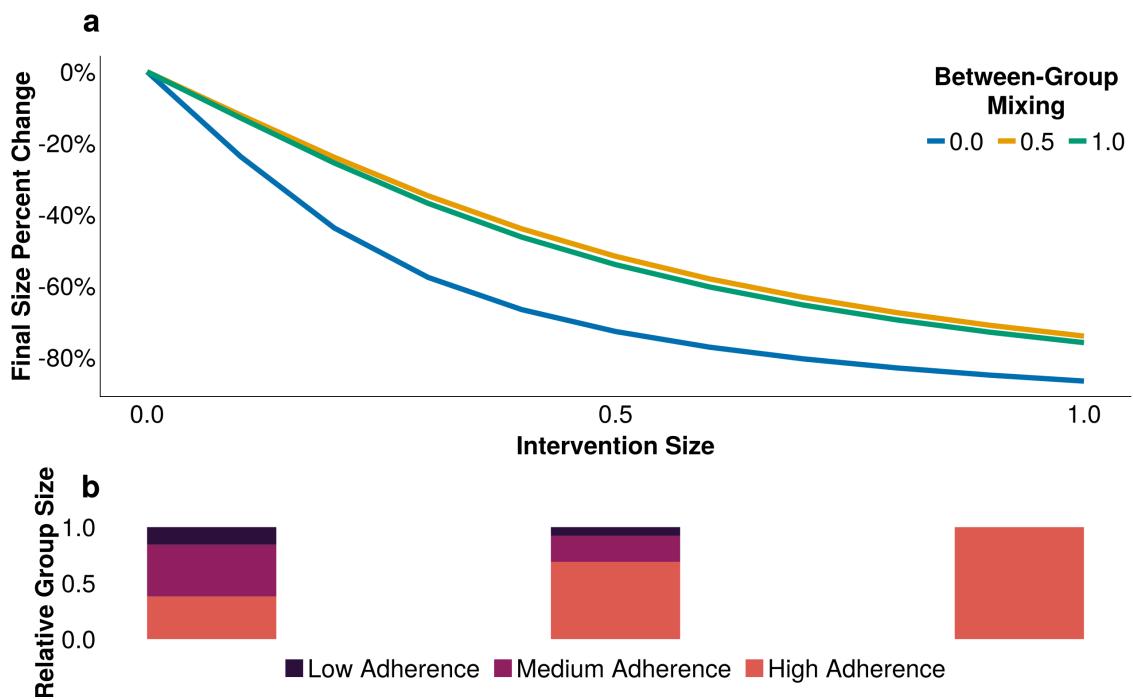
Below are results for alternative scenarios, which show qualitatively similar results to the main body of the text, albeit with a wider distribution in the Approximate Bayesian Computation distance metrics.

### PHMs Confer Protection

$$\rho \begin{pmatrix} \beta_{HH} & \beta_{HM} & \beta_{HL} \\ \beta_{MH} & \beta_{MM} & \beta_{ML} \\ \beta_{LH} & \beta_{HM} & \beta_{LL} \end{pmatrix} \rightarrow \rho \begin{pmatrix} \beta_{HH} & \phi\beta_{HH} & \phi\beta_{HH} \\ \phi\beta_{MM} & \beta_{MM} & \beta_{MM} \\ \color{red}{\phi\beta_{LL}} & \color{red}{\phi\beta_{LL}} & \color{red}{\beta_{LL}} \end{pmatrix} \text{ mixing structure} \quad B.1$$



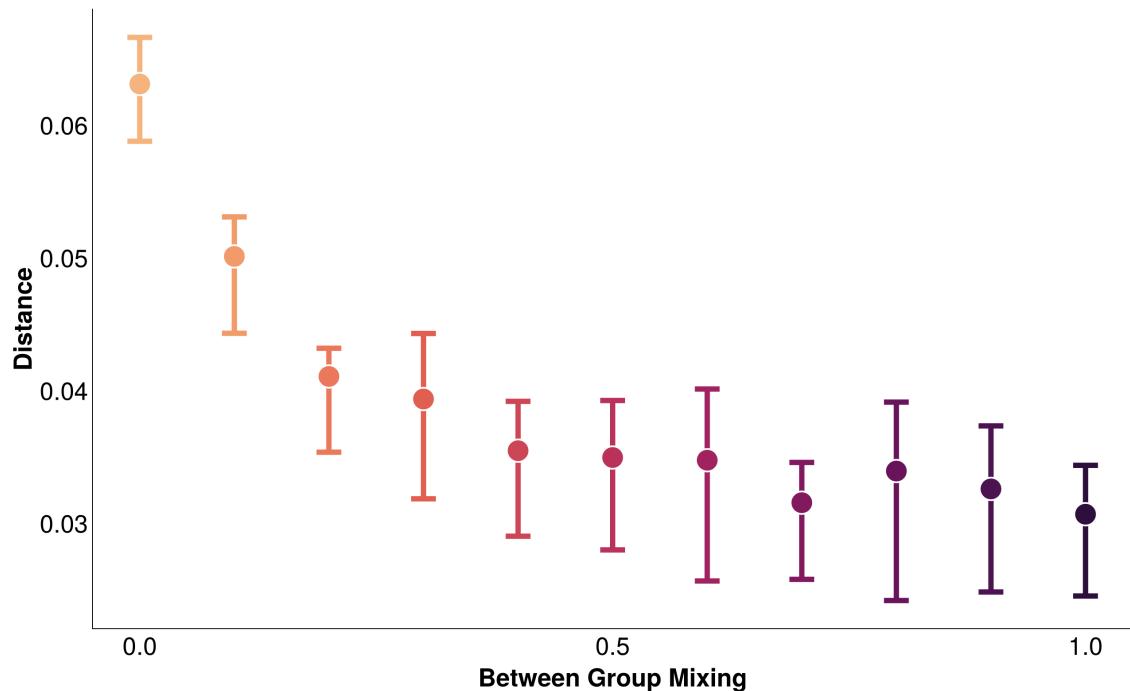
Supplemental Figure B.1: PHMs confer protection to the practitioner. Distribution of the distance from the ABC fits, with the minimum and maximum distances illustrated by the whiskers, and the median distance by the point. Between-group mixing of 1.0 equates to between-group mixing as likely as within-group mixing



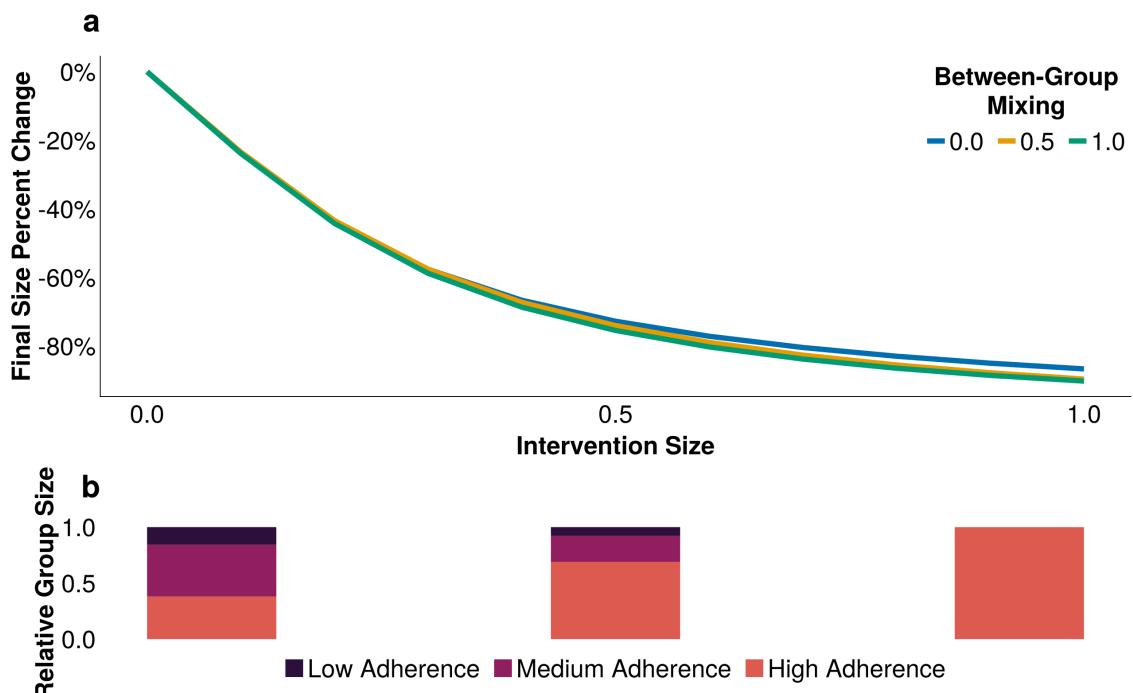
Supplemental Figure B.2: PHMs confer protection to the practitioner. A) The reduction in final infection size across a range of intervention effectiveness (1.0 is a fully effective intervention), accounting for a range of assortativity. Between-group mixing of 1.0 equates to between-group mixing as likely as within-group mixing; B) The relative distribution of group sizes at three levels of intervention effectiveness (0.0, 0.5, 1.0)

**Eq 1C (Identical Off-Diagonal Values)**

$$\rho \begin{pmatrix} \beta_{HH} & \beta_{HM} & \beta_{HL} \\ \beta_{MH} & \beta_{HM} & \beta_{ML} \\ \beta_{LH} & \beta_{HM} & \beta_{LL} \end{pmatrix} \rightarrow \rho \begin{pmatrix} \beta_{HH} & \phi\beta_{LL} & \phi\beta_{LL} \\ \phi\beta_{LL} & \beta_{MM} & \phi\beta_{LL} \\ \phi\beta_{LL} & \phi\beta_{LL} & \beta_{LL} \end{pmatrix} \text{ mixing structure} \quad \text{B.2}$$



Supplemental Figure B.3: Identical off-diagonal values. Distribution of the distance from the ABC fits, with the minimum and maximum distances illustrated by the whiskers, and the median distance by the point. Between-group mixing of 1.0 equates to between-group mixing as likely as within-group mixing



Supplemental Figure B.4: Identical off-diagonal values. A) The reduction in final infection size across a range of intervention effectiveness (1.0 is a fully effective intervention), accounting for a range of assortativity. Between-group mixing of 1.0 equates to between-group mixing as likely as within-group mixing; B) The relative distribution of group sizes at three levels of intervention effectiveness (0.0, 0.5, 1.0),

## Appendix C | Supplementary Material for Chapter 4

### Tables

Supplemental Table C.1: Mean outbreak detection accuracy for imperfect and perfect diagnostic tests, at their respective optimal alert thresholds, under dynamical and Poisson noise structures where the average daily noise incidence is 8 times the average daily measles incidence  $\Lambda(8)$ . The test sensitivity equals the test specificity for all diagnostic tests.

Noise Type	Test Characteristic		Testing Rate					
	Test Type	Test Lag	10%	20%	30%	40%	50%	60%
Dynamical noise	Imperfect Test (85%)	0	0.64	0.72	0.71	0.68	0.7	0.73
Dynamical noise	Imperfect Test (90%)	0	0.64	0.73	0.72	0.71	0.72	0.71
Poisson noise	Imperfect Test (85%)	0	0.66	0.93	0.92	0.93	0.88	0.88
Poisson noise	Imperfect Test (90%)	0	0.66	0.85	0.91	0.87	0.92	0.93
All noise structures	Perfect Test	0	0.66	0.93	0.91	0.89	0.91	0.93
All noise structures	Perfect Test	14	0.67	0.9	0.89	0.88	0.88	0.91

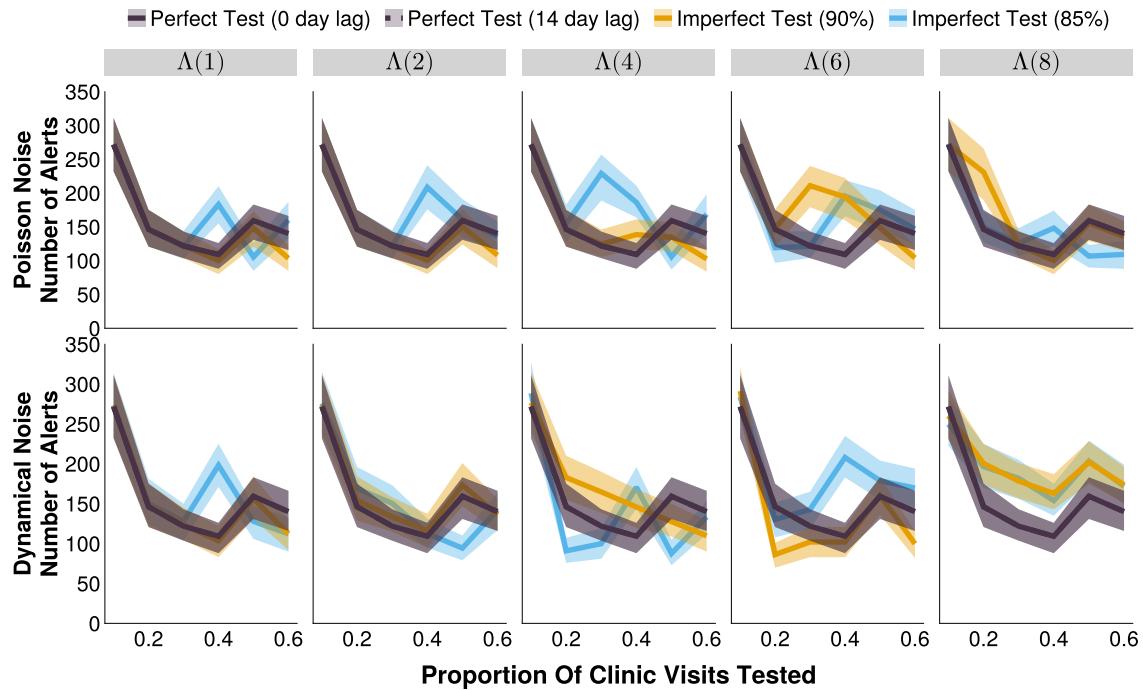
Supplemental Table C.2: Mean outbreak detection delays (days) for imperfect and perfect diagnostic tests, at their respective optimal alert thresholds, under dynamical and Poisson noise structures where the average daily noise incidence is 8 times the average daily measles incidence  $\Lambda(8)$ . The test sensitivity equals the test specificity for all diagnostic tests.

Noise Type	Test Type	Test Lag	Testing Rate					
			10%	20%	30%	40%	50%	60%
Dynamical noise	Imperfect Test (85%)	0	-24.82	-12.79	-9.15	-3.21	-16.22	-10.77
Dynamical noise	Imperfect Test (90%)	0	-17.21	-5.34	-0.55	3.03	-10.55	-3.66
Poisson noise	Imperfect Test (85%)	0	-3.75	24.49	31.45	31.85	38.17	37.87
Poisson noise	Imperfect Test (90%)	0	-3.69	12.94	32.92	40.14	20.26	28.74
All noise structures	Perfect Test	0	-3.69	22.61	30.64	36.08	17.92	23.05
All noise structures	Perfect Test	14	3.69	33.64	42.38	48.18	28.21	34.07

Supplemental Table C.3: Mean unavoidable cases per annum (scaled to Ghana's 2022 population) for imperfect and perfect diagnostic tests, at their respective optimal alert thresholds, under dynamical and Poisson noise structures where the average daily noise incidence is 8 times the average daily measles incidence  $\Lambda(8)$ . The test sensitivity equals the test specificity for all diagnostic tests.

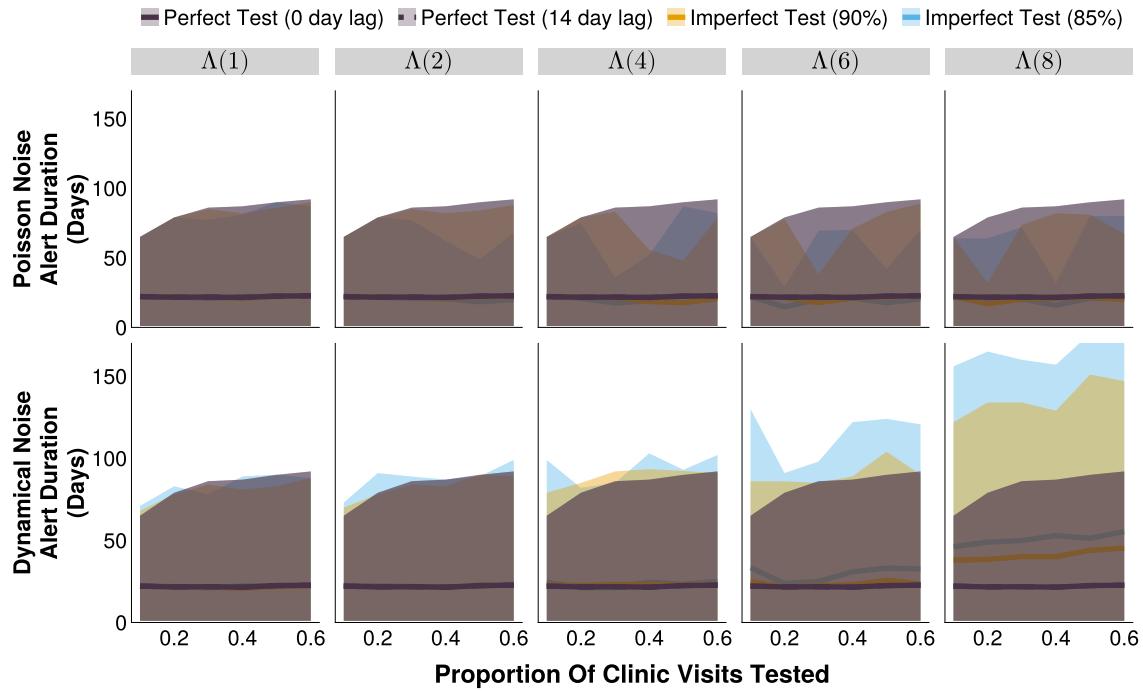
Noise Type	Test Type	Test Lag	Testing Rate					
			10%	20%	30%	40%	50%	60%
Dynamical noise	Imperfect Test (85%)	0	452	3053	4424	7728	3406	4361
Dynamical noise	Imperfect Test (90%)	0	515	3289	4650	6417	2578	4933
Poisson noise	Imperfect Test (85%)	0	766	6592	9111	9107	11865	11765
Poisson noise	Imperfect Test (90%)	0	770	3178	9736	12808	5205	8111
All noise structures	Perfect Test	0	770	5980	8893	11172	4529	6144
All noise structures	Perfect Test	14	2015	9277	12363	14643	7641	9495

## Figures



Supplemental Figure C.1: The number of alerts of outbreak detection systems under different testing rates and noise structures, at their respective optimal alert thresholds.

The shaded bands illustrate the 80% central interval, and the solid/dashed lines represent the mean estimate. Imperfect tests have the same values for sensitivity and specificity. Solid lines represent tests with 0-day turnaround times, and dashed lines represent tests with result delays.  $\Lambda(4)$  indicates the mean noise incidence is 4 times higher than the mean measles incidence, for example.



Supplemental Figure C.2: The duration of alerts of outbreak detection systems under different testing rates and noise structures, at their respective optimal alert thresholds.

The shaded bands illustrate the 80% central interval, and the solid/dashed lines represent the mean estimate. Imperfect tests have the same values for sensitivity and specificity. Solid lines represent tests with 0-day turnaround times, and dashed lines represent tests with result delays.  $\Lambda(4)$  indicates the mean noise incidence is 4 times higher than the mean measles incidence, for example.

# Appendix D | Supplementary Material for Chapter 5

## Tables

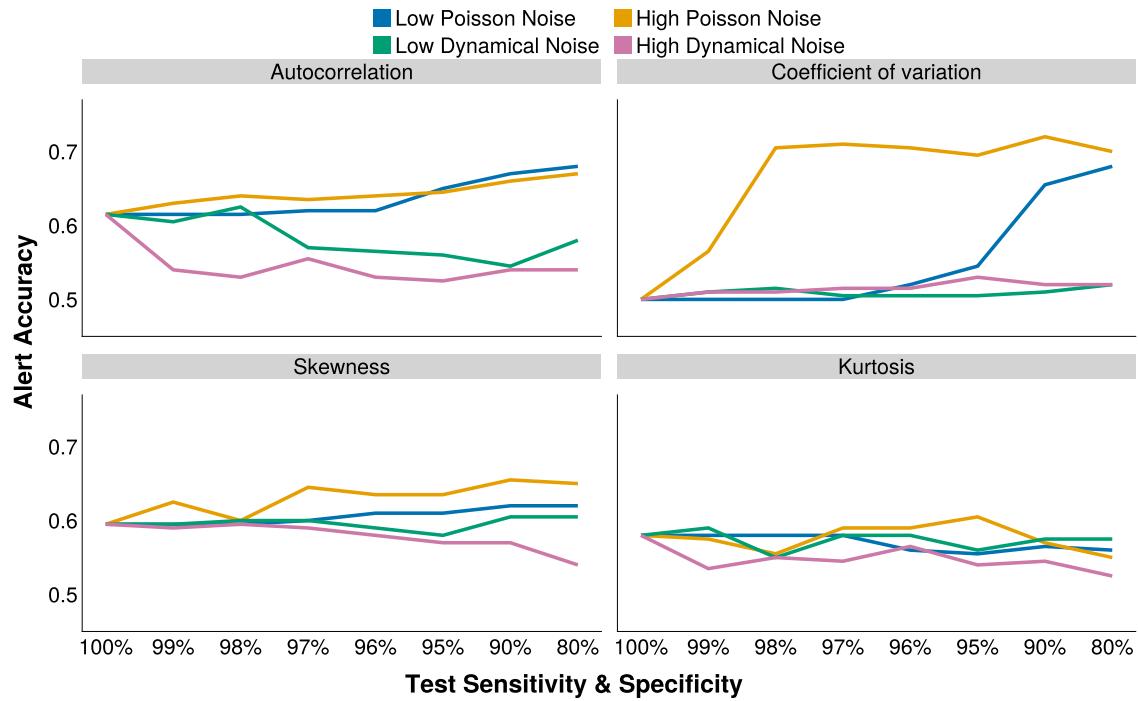
Supplemental Table D.1: The ranking and mean value of Kendall's Tau computed on the subset of the emergent time series after the burn-in period, for a perfect test and an imperfect test with sensitivity and specificity equal to 90%, under high and low Poisson and dynamical noise systems

Rank	Perfect Test	90% Sensitive & Specific Imperfect Test			
		Poisson Noise		Dynamical Noise	
	All Noise	Low	High	Low	High
1	Variance (0.62)	Variance (0.61)	Variance (0.60)	Variance (0.66)	Index of dispersion (0.47)
2	Index of dispersion (0.58)	Index of dispersion (0.59)	Index of dispersion (0.60)	Autocovariance (0.63)	Autocorrelation (0.45)
3	Autocovariance (0.58)	Autocovariance (0.55)	Coefficient of variation (0.59)	Index of dispersion (0.57)	Coefficient of variation (0.45)
4	Autocorrelation (0.38)	Coefficient of variation (0.51)	Autocovariance (0.51)	Mean (0.48)	Autocovariance (0.39)
5	Mean (0.38)	Autocorrelation (0.41)	Mean (0.37)	Autocorrelation (0.42)	Variance (0.38)
6	Coefficient of variation (0.15)	Mean (0.35)	Autocorrelation (0.36)	Coefficient of variation (0.12)	Skewness (0.11)
7	Skewness (0.06)	Skewness (0.14)	Skewness (0.10)	Skewness (-0.05)	Kurtosis (-0.19)
8	Kurtosis (-0.02)	Kurtosis (0.01)	Kurtosis (0.02)	Kurtosis (-0.11)	Mean (-0.21)

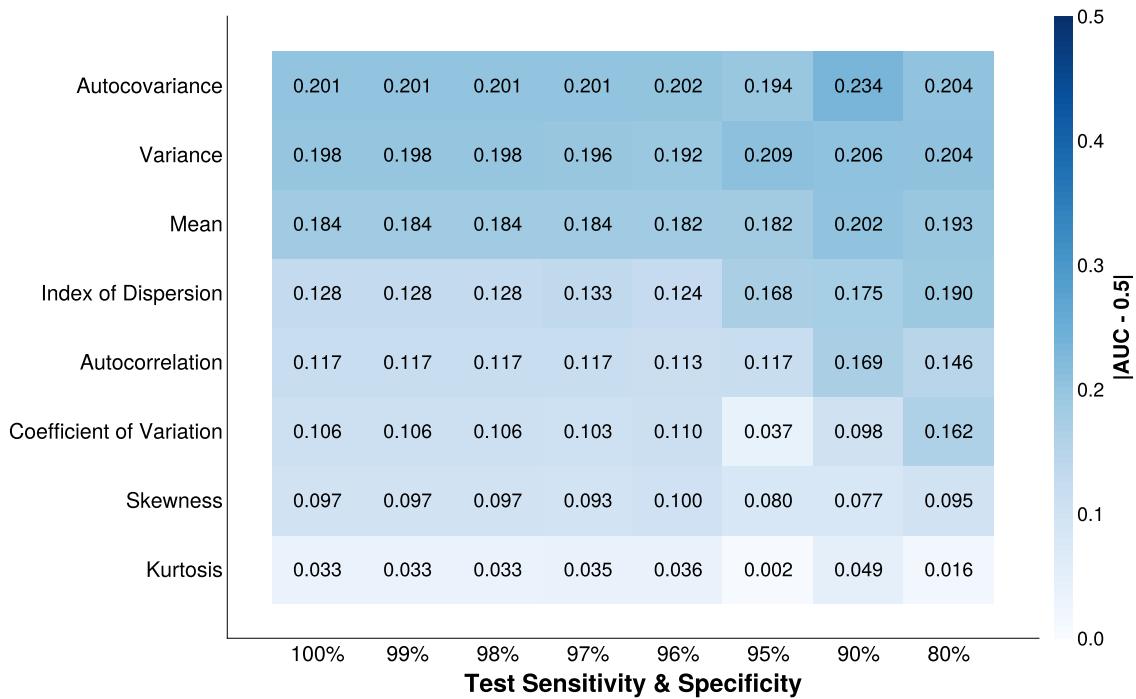
Supplemental Table D.2: The ranking and alert accuracy of the EWS-based alert system computed on the subset of the emergent time series after the burn-in period, for a perfect test and an imperfect test with sensitivity and specificity equal to 90%, under high and low Poisson and dynamical noise systems

Rank	Perfect Test	90% Sensitive & Specific Imperfect Test			
		Poisson Noise		Dynamical Noise	
	All Noise	Low	High	Low	High
1	Mean (0.72)	Mean (0.73)	Variance (0.73)	Variance (0.68)	Mean (0.60)
2	Variance (0.71)	Variance (0.71)	Coefficient of variation (0.72)	Mean (0.66)	Skewness (0.57)
3	Autocovariance (0.70)	Autocovariance (0.70)	Mean (0.72)	Autocovariance (0.65)	Kurtosis (0.55)
4	Index of dispersion (0.63)	Index of dispersion (0.69)	Index of dispersion (0.72)	Skewness (0.60)	Autocorrelation (0.54)
5	Autocorrelation (0.61)	Autocorrelation (0.67)	Autocovariance (0.71)	Index of dispersion (0.60)	Autocovariance (0.52)
6	Skewness (0.59)	Coefficient of variation (0.66)	Autocorrelation (0.66)	Kurtosis (0.57)	Coefficient of variation (0.52)
7	Kurtosis (0.58)	Skewness (0.62)	Skewness (0.66)	Autocorrelation (0.55)	Variance (0.52)
8	Coefficient of variation (0.50)	Kurtosis (0.56)	Kurtosis (0.57)	Coefficient of variation (0.51)	Index of dispersion (0.51)

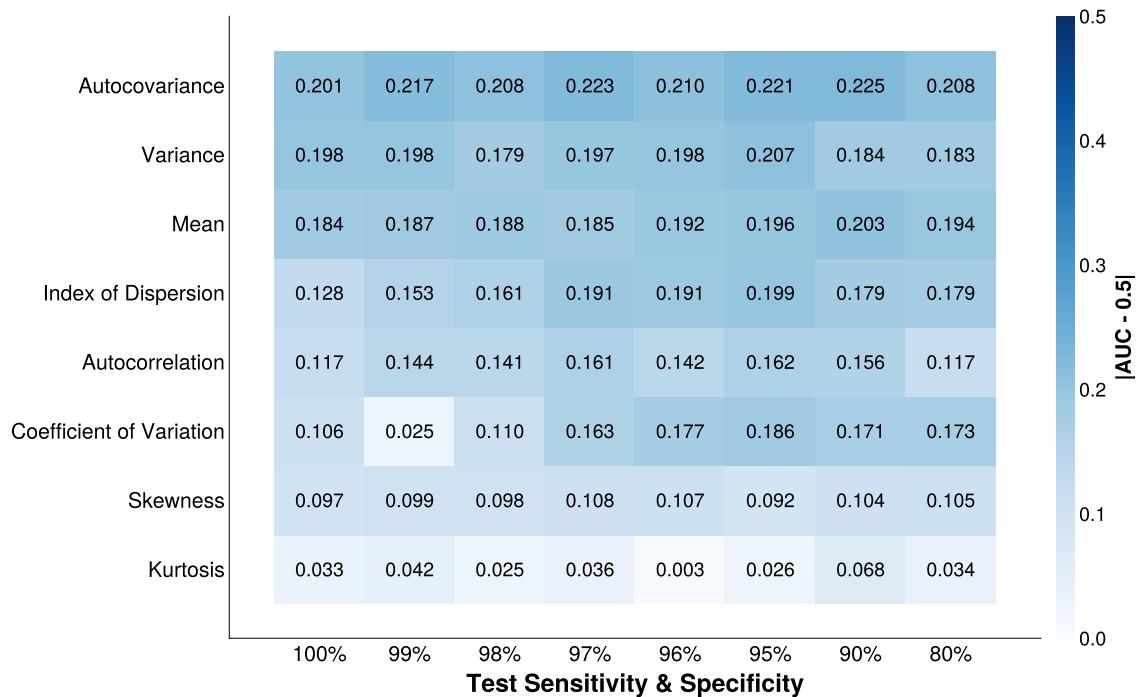
## Figures



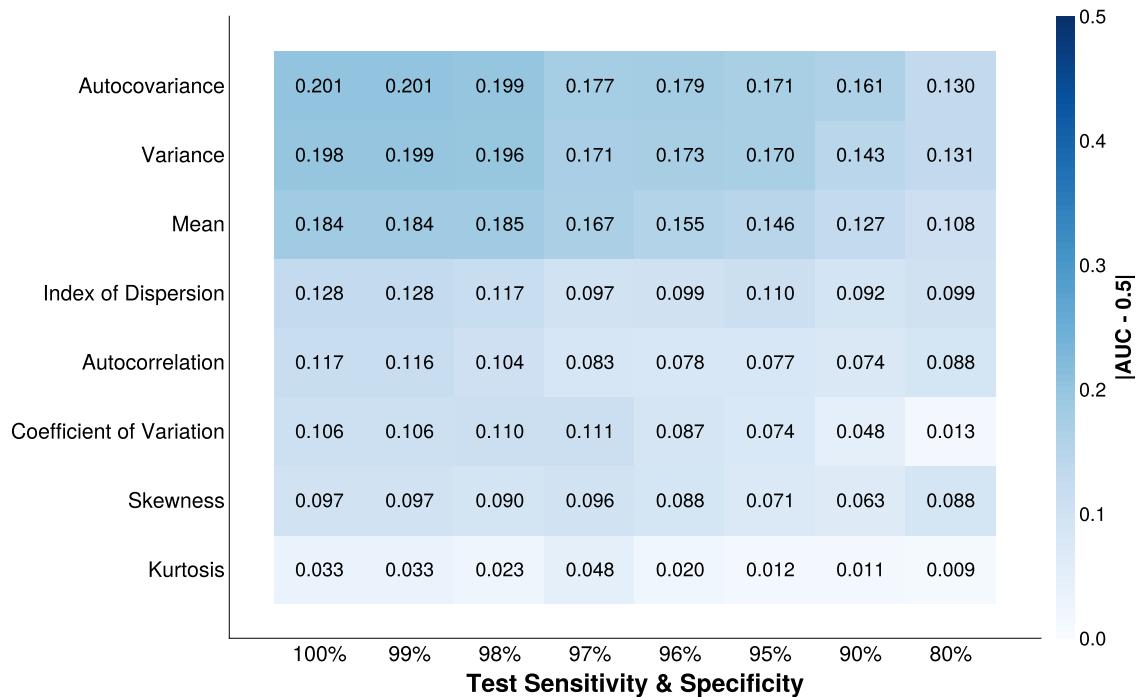
Supplemental Figure D.1: The change in alert accuracy for the least correlated EWS metrics under increasing diagnostic uncertainty, and low and high levels of Poisson or dynamical noise. Low noise refers to simulations where the average incidence of noise is equal to the average incidence of measles. High noise refers to simulations where the average incidence of noise is equal to 7 times the average incidence of measles. The test sensitivity equals the test specificity for all diagnostic tests.



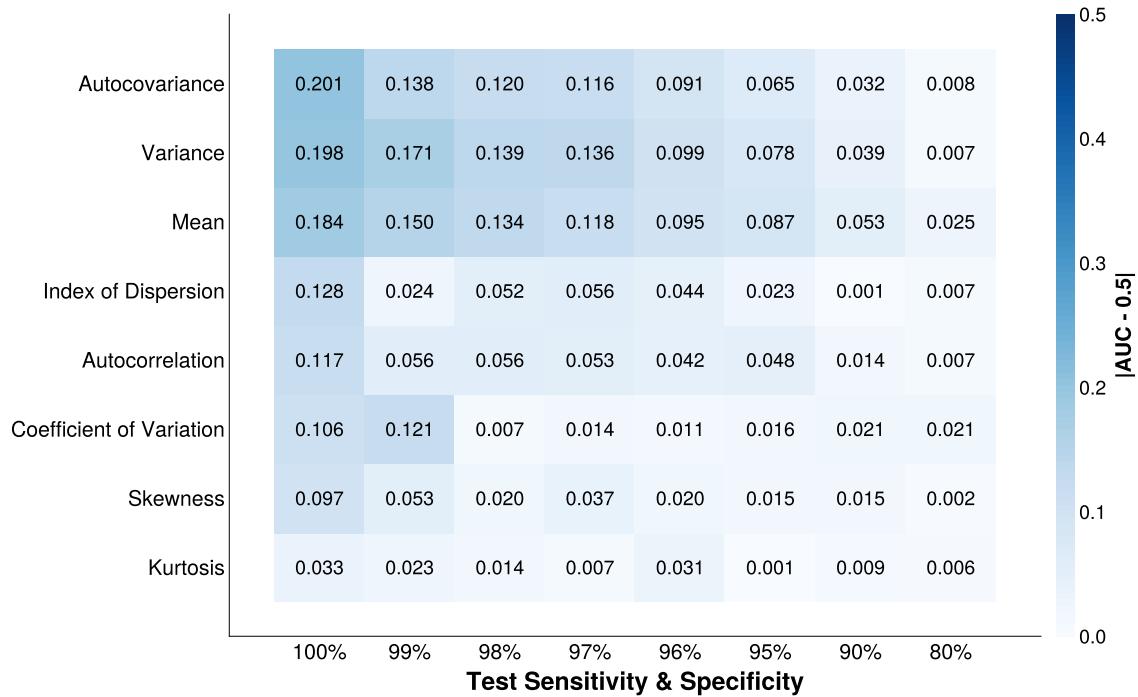
Supplemental Figure D.2: The strength of the correlation ( $|AUC - 0.5|$ ) for each EWS metric with emergence, at low levels of Poisson noise, for diagnostic tests of varying accuracy, and was computed after the completion of the burn-in period. The test sensitivity equals the test specificity for all diagnostic tests.



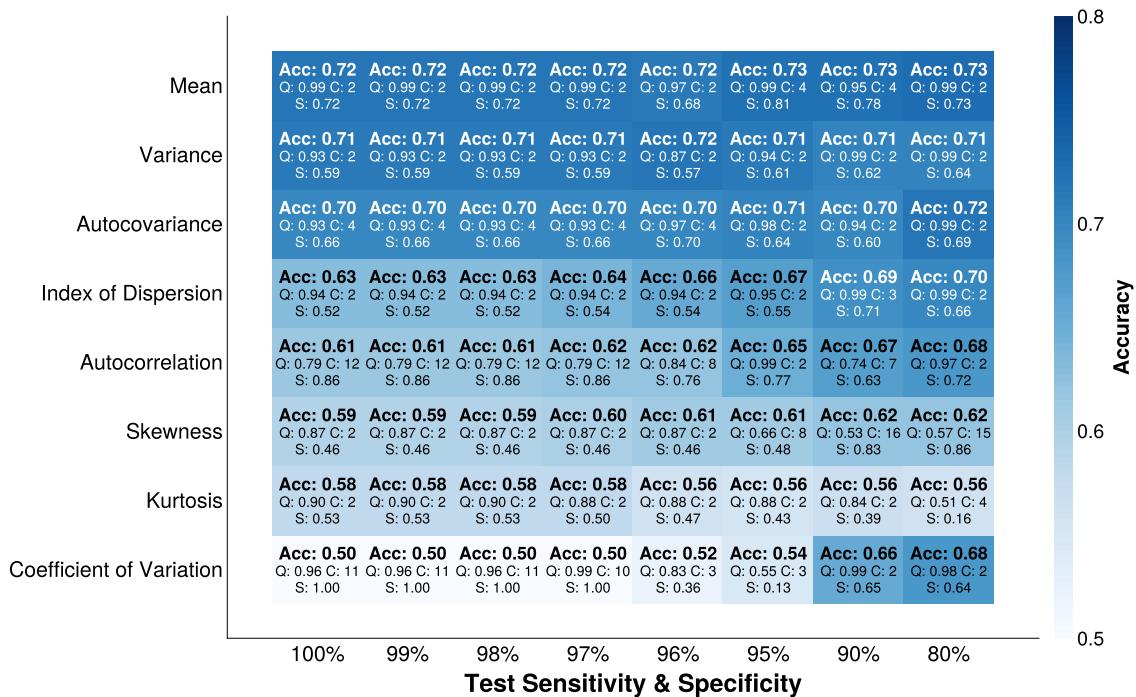
Supplemental Figure D.3: The strength of the correlation ( $|AUC - 0.5|$ ) for each EWS metric with emergence, at high levels of Poisson noise, for diagnostic tests of varying accuracy, and was computed after the completion of the burn-in period. The test sensitivity equals the test specificity for all diagnostic tests.



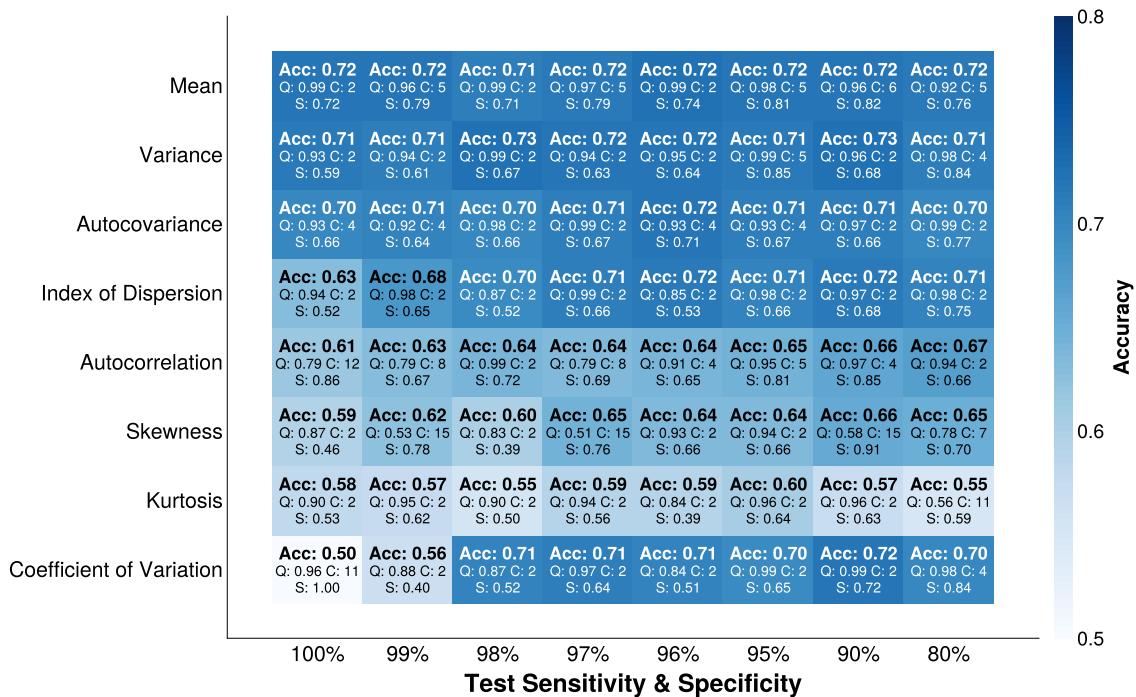
Supplemental Figure D.4: The strength of the correlation ( $|AUC - 0.5|$ ) for each EWS metric with emergence, at low levels of dynamical noise, for diagnostic tests of varying accuracy, and was computed after the completion of the burn-in period. The test sensitivity equals the test specificity for all diagnostic tests.



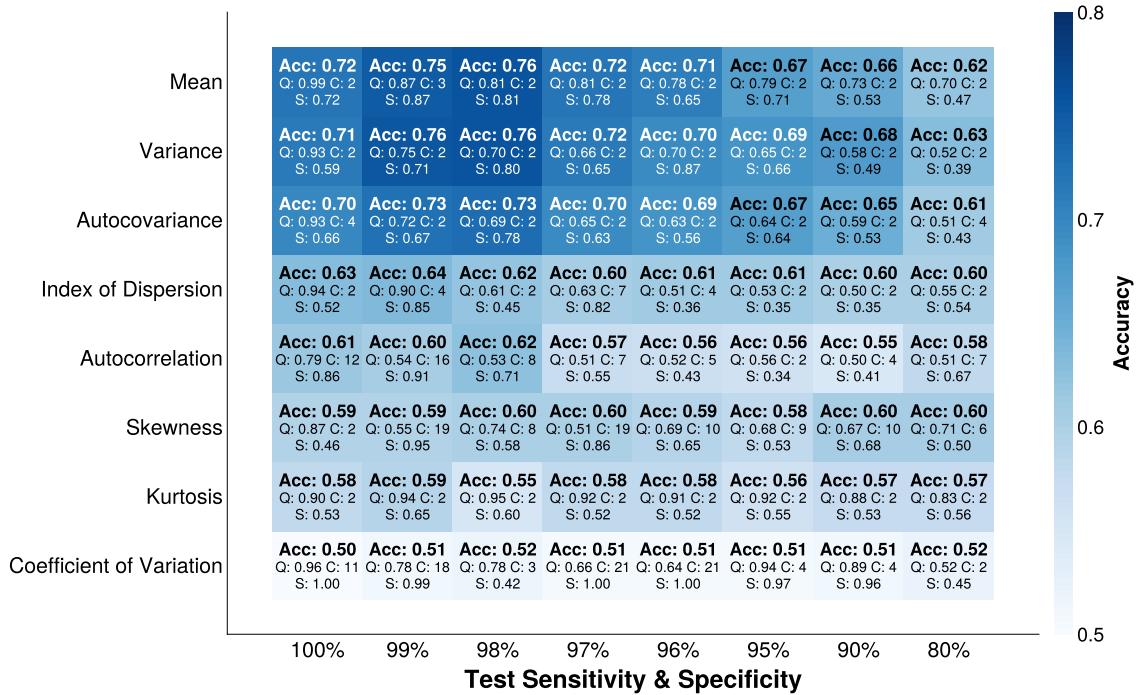
Supplemental Figure D.5: The strength of the correlation ( $|AUC - 0.5|$ ) for each EWS metric with emergence, at high levels of dynamical noise, for diagnostic tests of varying accuracy, and was computed after the completion of the burn-in period. The test sensitivity equals the test specificity for all diagnostic tests.



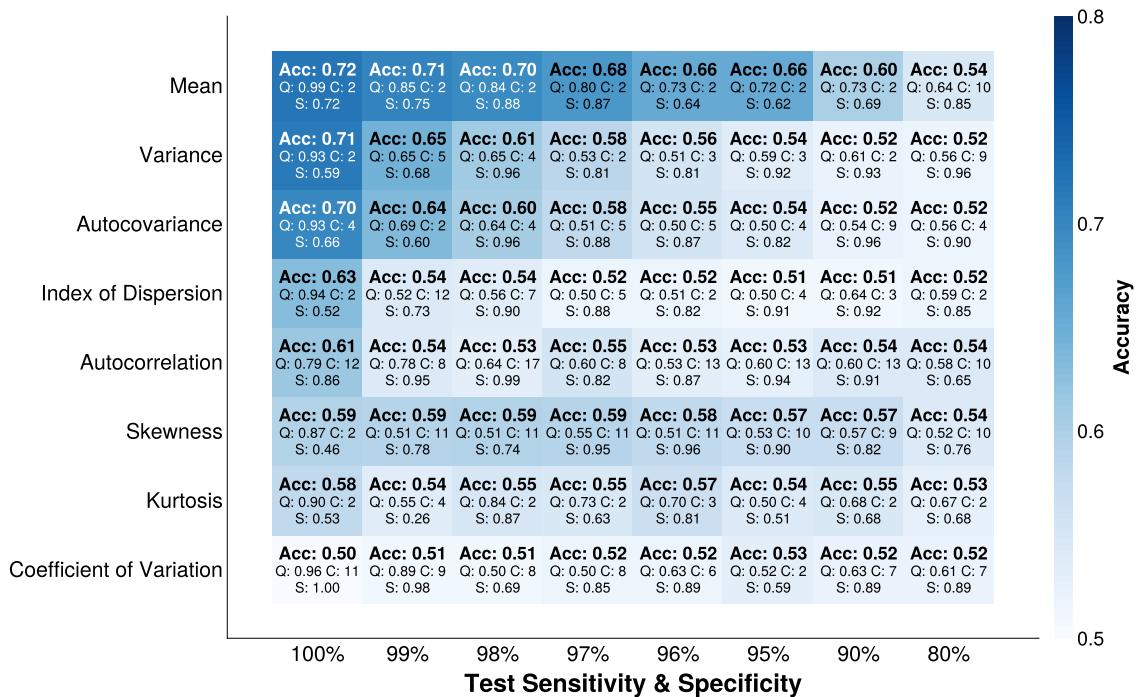
Supplemental Figure D.6: The maximal alert accuracy achieved by each EWS metric under low levels of Poisson noise. Q) refers to the long-running quantile threshold to return a flag, and C) the number of consecutive flags to trigger an alert, that in combination produce the maximal accuracy. S) refers to the resulting specificity of the alert system. The test sensitivity equals the test specificity for all diagnostic tests.



Supplemental Figure D.7: The maximal alert accuracy achieved by each EWS metric under high levels of Poisson noise. Q) refers to the long-running quantile threshold to return a flag, and C) the number of consecutive flags to trigger an alert, that in combination produce the maximal accuracy. S) refers to the resulting specificity of the alert system. The test sensitivity equals the test specificity for all diagnostic tests.



Supplemental Figure D.8: The maximal alert accuracy achieved by each EWS metric under low levels of dynamical noise. Q) refers to the long-running quantile threshold to return a flag, and C) the number of consecutive flags to trigger an alert, that in combination produce the maximal accuracy. S) refers to the resulting specificity of the alert system. The test sensitivity equals the test specificity for all diagnostic tests.



Supplemental Figure D.9: The maximal alert accuracy achieved by each EWS metric under high levels of dynamical noise. Q) refers to the long-running quantile threshold to return a flag, and C) the number of consecutive flags to trigger an alert, that in combination produce the maximal accuracy. S) refers to the resulting specificity of the alert system. The test sensitivity equals the test specificity for all diagnostic tests.

## Bibliography

- [1] Box G E P. Science and Statistics. *Journal of the American Statistical Association* 1976;71:791–9. <https://doi.org/10.2307/2286841>
- [2] Mervis C B, Rosch E. Categorization of Natural Objects. *Annual Review of Psychology* 1981;32:89–115. [https://doi.org/10.1146/annurev-ps.32.020181.000513](https://doi.org/10.1146/annurev.ps.32.020181.000513)
- [3] Rhodes M, Baron A. The Development of Social Categorization. *Annual Review of Developmental Psychology* 2019;1:359. <https://doi.org/10.1146/annurev-devpsych-121318-084824>
- [4] Vynnycky E, White R G. Introduction. An Introduction to Infectious Disease Modelling 2010:1–12
- [5] Kermack W O, McKendrick A G, Walker G T. A Contribution to the Mathematical Theory of Epidemics. *Proceedings of the Royal Society of London Series A, Containing Papers of a Mathematical and Physical Character* 1997;115:700–21. <https://doi.org/10.1098/rspa.1927.0118>
- [6] Hessel F. Burden of Disease. *Encyclopedia of Public Health* 2008:94–6. [https://doi.org/10.1007/978-1-4020-5614-7\\_297](https://doi.org/10.1007/978-1-4020-5614-7_297)
- [7] Bubar K M, Reinholt K, Kissler S M, Lipsitch M, Cobey S, Grad Y H, et al. Model-Informed COVID-19 Vaccine Prioritization Strategies by Age and Serostatus. *Science* 2021. <https://doi.org/10.1126/science.abe6959>
- [8] Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. *PLOS Medicine* 2008;5:e74. <https://doi.org/10.1371/journal.pmed.0050074>
- [9] Klepac P, Kissler S, Gog J. Contagion! The BBC Four Pandemic – The Model behind the Documentary. *Epidemics* 2018;24:49–59. <https://doi.org/10.1016/j.epidem.2018.03.003>
- [10] Davies N G, Klepac P, Liu Y, Prem K, Jit M, Eggo R M. Age-Dependent Effects in the Transmission and Control of COVID-19 Epidemics. *Nature Medicine* 2020;26:1205–11. <https://doi.org/10.1038/s41591-020-0962-9>
- [11] Kretzschmar M, Morris M. Measures of Concurrency in Networks and the Spread of Infectious Disease. *Mathematical Biosciences* 1996;133:165–95. [https://doi.org/10.1016/0025-5564\(95\)00093-3](https://doi.org/10.1016/0025-5564(95)00093-3)
- [12] Herrera-Diestra J L, Tildesley M, Shea K, Ferrari M J. Cattle Transport Network Predicts Endemic and Epidemic Foot-and-Mouth Disease Risk on Farms in Turkey. *PLOS Computational Biology* 2022;18:e1010354. <https://doi.org/10.1371/journal.pcbi.1010354>

- [13] Lloyd-Smith J O, Schreiber S J, Kopp P E, Getz W M. Superspreading and the Effect of Individual Variation on Disease Emergence. *Nature* 2005;438:355–9. <https://doi.org/10.1038/nature04153>
- [14] Bryan C J, Tipton E, Yeager D S. Behavioural Science Is Unlikely to Change the World without a Heterogeneity Revolution. *Nature Human Behaviour* 2021;5:980. <https://doi.org/10.1038/s41562-021-01143-3>
- [15] Levitt A, Mermin J, Jones C M, See I, Butler J C. Infectious Diseases and Injection Drug Use: Public Health Burden and Response. *The Journal of Infectious Diseases* 2020;222:S213–7. <https://doi.org/10.1093/infdis/jiaa432>
- [16] Levenson E, Boyette C, Mack J. Colleges and Universities across the US Are Canceling In-Person Classes Due to Coronavirus 2020
- [17] Weller B E, Bowen N K, Faubert S J. Latent Class Analysis: A Guide to Best Practice. *Journal of Black Psychology* 2020;46:287–311. <https://doi.org/10.1177/0095798420930932>
- [18] Nylund-Gibson K, Choi A Y. Ten Frequently Asked Questions about Latent Class Analysis. *Translational Issues in Psychological Science* 2018;4:440. <https://doi.org/10.1037/tps0000176>
- [19] Murray J, Cohen A L. Infectious Disease Surveillance. *International Encyclopedia of Public Health* 2017:222–9. <https://doi.org/10.1016/B978-0-12-803678-5.00517-8>
- [20] World Health Organization. *Surveillance in Emergencies* 2024
- [21] Alhajj M, Zubair M, Farhana A. Enzyme Linked Immunosorbent Assay. *StatPearls* 2024
- [22] Hiebert J, Zubach V, Charlton C L, Fenton J, Tipple G A, Fonseca K, et al. Evaluation of Diagnostic Accuracy of Eight Commercial Assays for the Detection of Measles Virus-Specific IgM Antibodies. *Journal of Clinical Microbiology* 2021;59:e3161. <https://doi.org/10.1128/JCM.03161-20>
- [23] Westreich D. Diagnostic Testing, Screening, and Surveillance. *Epidemiology by Design: A Causal Approach to the Health Sciences* 2019:0. <https://doi.org/10.1093/oso/9780190665760.003.0005>
- [24] Shreffler J, Huecker M R. Diagnostic Testing Accuracy: Sensitivity, Specificity, Predictive Values and Likelihood Ratios. *StatPearls* 2024
- [25] Parikh R, Mathai A, Parikh S, Chandra Sekhar G, Thomas R. Understanding and Using Sensitivity, Specificity and Predictive Values. *Indian Journal of Ophthalmology* 2008;56:45–50
- [26] Kaninda A-V, Belanger F, Lewis R, Batchassi E, Aplogan A, Yakoua Y, et al. Effectiveness of Incidence Thresholds for Detection and Control of

- Meningococcal Meningitis Epidemics in Northern Togo. International Journal of Epidemiology 2000;29:933–40. <https://doi.org/10.1093/ije/29.5.933>
- [27] Trotter C L, Cibrelus L, Fernandez K, Lingani C, Ronveaux O, Stuart J M. Response Thresholds for Epidemic Meningitis in Sub-Saharan Africa Following the Introduction of MenAfriVac®. Vaccine 2015;33:6212–7. <https://doi.org/10.1016/j.vaccine.2015.09.107>
- [28] Guidelines Working Group. Updated Guidelines for Evaluating Public Health Surveillance Systems. Atlanta, Georgia: 2001
- [29] Bengtsson L, Gaudart J, Lu X, Moore S, Wetter E, Sallah K, et al. Using Mobile Phone Data to Predict the Spatial Spread of Cholera. Scientific Reports 2015;5:8923. <https://doi.org/10.1038/srep08923>
- [30] Wilson M E. Travel and the Emergence of Infectious Diseases. Emerging Infectious Diseases 1995;1:39–46
- [31] Chinazzi M, Davis J T, Ajelli M, Gioannini C, Litvinova M, Merler S, et al. The Effect of Travel Restrictions on the Spread of the 2019 Novel Coronavirus (COVID-19) Outbreak. Science 2020;368:395–400. <https://doi.org/10.1126/science.aba9757>
- [32] Viboud C, Bjørnstad O N, Smith D L, Simonsen L, Miller M A, Grenfell B T. Synchrony, Waves, and Spatial Hierarchies in the Spread of Influenza. Science 2006;312:447–51. <https://doi.org/10.1126/science.1125237>
- [33] Bharti N, Tatem A J, Ferrari M J, Grais R F, Djibo A, Grenfell B T. Explaining Seasonal Fluctuations of Measles in Niger Using Nighttime Lights Imagery. Science (New York, NY) 2011;334:1424–7. <https://doi.org/10.1126/science.1210554>
- [34] Wells C R, Sah P, Moghadas S M, Pandey A, Shoukat A, Wang Y, et al. Impact of International Travel and Border Control Measures on the Global Spread of the Novel 2019 Coronavirus Outbreak. Proceedings of the National Academy of Sciences 2020;117:7504–9. <https://doi.org/10.1073/pnas.2002616117>
- [35] Bogoch I I, Creatore M I, Cetron M S, Brownstein J S, Pesik N, Miniota J, et al. Assessment of the Potential for International Dissemination of Ebola Virus via Commercial Air Travel during the 2014 West African Outbreak. The Lancet 2015;385:29–35. [https://doi.org/10.1016/S0140-6736\(14\)61828-6](https://doi.org/10.1016/S0140-6736(14)61828-6)
- [36] Jia J S, Lu X, Yuan Y, Xu G, Jia J, Christakis N A. Population Flow Drives Spatio-Temporal Distribution of COVID-19 in China. Nature 2020;582:389–94. <https://doi.org/10.1038/s41586-020-2284-y>
- [37] Funk S, Ciglenecki I, Tiffany A, Gignoux E, Camacho A, Eggo R M, et al. The Impact of Control Strategies and Behavioural Changes on the Elimination of Ebola from Lofa County, Liberia. Philosophical Transactions of the Royal Society

- of London Series B, Biological Sciences 2017;372. <https://doi.org/10.1098/rstb.2016.0302>
- [38] Sniadack D H, Moscoso B, Aguilar R, Heath J, Bellini W, Chiu M C. Measles Epidemiology and Outbreak Response Immunization in a Rural Community in Peru. Bulletin of the World Health Organization 1999;77:545–52
- [39] Stoddard S T, Forshey B M, Morrison A C, Paz-Soldan V A, Vazquez-Prokopec G M, Astete H, et al. House-to-House Human Movement Drives Dengue Virus Transmission. Proceedings of the National Academy of Sciences 2012. <https://doi.org/10.1073/pnas.1213349110>
- [40] Leidner A J, Barry V, Bowen V B, Silver R, Musial T, Kang G J, et al. Opening of Large Institutions of Higher Education and County-Level COVID-19 Incidence – United States, July 6–September 17, 2020 Vaughn Barry, PhD1; Virginia B. Bowen, PhD1; Rachel Silver, MPH1; Trieste Musial, MS2; Gloria J. Kang, PhD1; Matthew D. Ritchey, DPT3; Kelly Fletcher, MPH2; Lisa Barrios, DrPH1; Eric Pevzner, PhD. MMWR Morbidity and Mortality Weekly Report 2021;70. <https://doi.org/10.15585/mmwr.mm7001a4>
- [41] Goetz S, Tian Z, Schmidt C, Meadowcroft D, NERCRD, Pennsylvania State University. Rural COVID-19 Cases Lag Urban Areas but Are Growing Much More Rapidly. 2020
- [42] Christensen H, Turner K, Trickey A, Booton R D, Hemani G, Nixon E, et al. COVID-19 Transmission in a University Setting: A Rapid Review of Modelling Studies. Medrxiv 2020:2020. <https://doi.org/10.1101/2020.09.07.20189688>
- [43] Andersen M S, Bento A I, Basu A, Marsicano C, Simon K. College Openings, Mobility, and the Incidence of COVID-19 Cases. Medrxiv 2020:2020. <https://doi.org/10.1101/2020.09.22.20196048>
- [44] Times T N Y. Tracking the Coronavirus at U.S. Colleges and Universities - The New York Times 2020
- [45] United States Census Bureau. U.S. Census Bureau QuickFacts: Centre County, Pennsylvania 2019
- [46] Pennsylvania State University. COVID-19 Dashboard 2021
- [47] Long Q-X, Liu B-Z, Deng H-J, Wu G-C, Deng K, Chen Y-K, et al. Antibody Responses to SARS-CoV-2 in Patients with COVID-19. Nature Medicine 2020;26:845–8. <https://doi.org/10.1038/s41591-020-0897-1>
- [48] Gontu A, Srinivasan S, Nair M S, Lindner S E, Minns A M, Rossi R, et al. Quantitative Estimation of IgM and IgG Antibodies Against SARS-CoV-2. Protocolsio 2020. <https://doi.org/dx.doi.org/10.17504/protocols.io.bivgke3w>

- [49] Gontu A, Srinivasan S, Salazar E, Nair M S, Nissly R H, Greenawalt D, et al. Limited Window for Donation of Convalescent Plasma with High Live-Virus Neutralizing Antibody Titers for COVID-19 Immunotherapy. *Communications Biology* 2021;4:1–9. <https://doi.org/10.1038/s42003-021-01813-y>
- [50] Uyoga S, Adetifa I M O, Karanja H K, Nyagwange J, Tuju J, Wanjiku P, et al. Seroprevalence of Anti-SARS-CoV-2 IgG Antibodies in Kenyan Blood Donors. *Science* 2021;371:79–82. <https://doi.org/10.1126/science.abe1916>
- [51] Stringhini S, Zaballa M-E, Perez-Saez J, Pullen N, Mestral C de, Picazio A, et al. Seroprevalence of Anti-SARS-CoV-2 Antibodies after the Second Pandemic Peak. *The Lancet Infectious Diseases* 2021;0. [https://doi.org/10.1016/S1473-3099\(21\)00054-2](https://doi.org/10.1016/S1473-3099(21)00054-2)
- [52] Kalish H, Klumpp-Thomas C, Hunsberger S, Baus H A, Fay M P, Siripong N, et al. Mapping a Pandemic: SARS-CoV-2 Seropositivity in the United States. *Medrxiv* 2021:2021. <https://doi.org/10.1101/2021.01.27.21250570>
- [53] Devleesschauwer B, Torgerson P, Charlier J, Levecke B, Praet N, Roelandt S, et al. Prevalence: Tools for Prevalence Assessment Studies 2014
- [54] Harrison E, Drake T, Ots R. Finalfit: Quickly Create Elegant Regression Results Tables and Plots When Modelling 2021
- [55] van Buuren S, Groothuis-Oudshoorn K. {mice}: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software* 2011;45:1–67
- [56] Huang Y-T, Tu Y-K, Lai P-C. Estimation of the Secondary Attack Rate of COVID-19 Using Proportional Meta-Analysis of Nationwide Contact Tracing Data in Taiwan. *Journal of Microbiology, Immunology and Infection* 2020. <https://doi.org/10.1016/j.jmii.2020.06.003>
- [57] Cheng H-Y, Jian S-W, Liu D-P, Ng T-C, Huang W-T, Lin H-H, et al. Contact Tracing Assessment of COVID-19 Transmission Dynamics in Taiwan and Risk at Different Exposure Periods Before and After Symptom Onset. *JAMA Internal Medicine* 2020;180:1156–63. <https://doi.org/10.1001/jamainternmed.2020.2020>
- [58] Leclerc Q J, Fuller N M, Knight L E, Funk S, Knight G M. What Settings Have Been Linked to SARS-CoV-2 Transmission Clusters?. *Wellcome Open Research* 2020;5. <https://doi.org/10.12688/wellcomeopenres.15889.2>
- [59] Brooks-Pollock E, Read J M, House T, Medley G F, Keeling M J, Danon L. The Population Attributable Fraction (PAF) of Cases Due to Gatherings and Groups with Relevance to COVID-19 Mitigation Strategies. *Medrxiv* 2020:2020. <https://doi.org/10.1101/2020.03.20.20039537>
- [60] R Core Team. R: A Language and Environment for Statistical Computing 2021

- [61] Landau W M. The Targets R Package: A Dynamic Make-like Function-Oriented Pipeline Toolkit for Reproducibility and High-Performance Computing. *Journal of Open Source Software* 2021;6:2959. <https://doi.org/10.21105/joss.02959>
- [62] Monod M, Blenkinsop A, Xi X, Hebert D, Bershan S, Tietze S, et al. Age Groups That Sustain Resurging COVID-19 Epidemics in the United States. *Science* 2021:eabe8372. <https://doi.org/10.1126/science.abe8372>
- [63] Bharti N, Lambert B, Exten C, Faust C, Ferrari M, Robinson A. Large University with High COVID-19 Incidence Is Not Associated with Excess Cases in Non-Student Population. *Scientific Reports* 2022;12:3313. <https://doi.org/10.1038/s41598-022-07155-x>
- [64] Brooks S K, Webster R K, Smith L E, Woodland L, Wessely S, Greenberg N, et al. The Psychological Impact of Quarantine and How to Reduce It: Rapid Review of the Evidence. *Lancet (London, England)* 2020;395:912–20. [https://doi.org/10.1016/S0140-6736\(20\)30460-8](https://doi.org/10.1016/S0140-6736(20)30460-8)
- [65] Pennsylvania State University. Mask Up or Pack Up 2021
- [66] Centers for Disease Control and Prevention. CDC Museum COVID-19 Timeline 2022
- [67] Fletcher J. What Is Heterogeneity and Is It Important?. *BMJ : British Medical Journal* 2007;334:94–6. <https://doi.org/10.1136/bmj.39057.406644.68>
- [68] Nold A. Heterogeneity in Disease-Transmission Modeling. *Mathematical Biosciences* 1980;52:227–40. [https://doi.org/10.1016/0025-5564\(80\)90069-3](https://doi.org/10.1016/0025-5564(80)90069-3)
- [69] Trauer J M, Dodd P J, Gomes M G M, Gomez G B, Houben R M G J, McBryde E S, et al. The Importance of Heterogeneity to the Epidemiology of Tuberculosis. *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 2019;69:159–66. <https://doi.org/10.1093/cid/ciy938>
- [70] Zhang Y, Britton T, Zhou X. Monitoring Real-Time Transmission Heterogeneity from Incidence Data. *PLOS Computational Biology* 2022;18:e1010078. <https://doi.org/10.1371/journal.pcbi.1010078>
- [71] Woolhouse M E J, Dye C, Etard J-F, Smith T, Charlwood J D, Garnett G P, et al. Heterogeneities in the Transmission of Infectious Agents: Implications for the Design of Control Programs. *Proceedings of the National Academy of Sciences of the United States of America* 1997;94:338–42
- [72] Wang H, Ghosh A, Ding J, Sarkar R, Gao J. Heterogeneous Interventions Reduce the Spread of COVID-19 in Simulations on Real Mobility Data. *Scientific Reports* 2021;11:7809. <https://doi.org/10.1038/s41598-021-87034-z>

- [73] McDonald S A, Devleesschauwer B, Wallinga J. The Impact of Individual-Level Heterogeneity on Estimated Infectious Disease Burden: A Simulation Study. *Population Health Metrics* 2016;14:47. <https://doi.org/10.1186/s12963-016-0116-y>
- [74] Sevelius J M, Patouhas E, Keatley J G, Johnson M O. Barriers and Facilitators to Engagement and Retention in Care among Transgender Women Living with Human Immunodeficiency Virus. *Annals of Behavioral Medicine : a Publication of the Society of Behavioral Medicine* 2014;47:5–16. <https://doi.org/10.1007/s12160-013-9565-8>
- [75] Tuschhoff B M, Kennedy D A. Detecting and Quantifying Heterogeneity in Susceptibility Using Contact Tracing Data 2023:2023. <https://doi.org/10.1101/2023.10.04.560944>
- [76] Delaney K P. Strategies Adopted by Gay, Bisexual, and Other Men Who Have Sex with Men to Prevent Monkeypox Virus Transmission — United States, August 2022. *MMWR Morbidity and Mortality Weekly Report* 2022;71. <https://doi.org/10.15585/mmwr.mm7135e1>
- [77] Anderson T L, Nande A, Merenstein C, Raynor B, Oommen A, Kelly B J, et al. Quantifying Individual-Level Heterogeneity in Infectiousness and Susceptibility through Household Studies. *Medrxiv* 2022:2022. <https://doi.org/10.1101/2022.12.02.22281853>
- [78] MacDonald K S, Fowke K R, Kimani J, Dunand V A, Nagelkerke N J D, Blake Ball T, et al. Influence of HLA Supertypes on Susceptibility and Resistance to Human Immunodeficiency Virus Type 1 Infection. *The Journal of Infectious Diseases* 2000;181:1581–9. <https://doi.org/10.1086/315472>
- [79] Elie B, Selinger C, Alizon S. The Source of Individual Heterogeneity Shapes Infectious Disease Outbreaks. *Proceedings of the Royal Society B: Biological Sciences* 2022;289:20220232. <https://doi.org/10.1098/rspb.2022.0232>
- [80] Hay J, Routledge I, Takahashi S. Serodynamics: A Review of Methods for Epidemiological Inference Using Serological Data. 2023. <https://doi.org/10.31219/osf.io/kqdsn>
- [81] Yang B, Lessler J, Zhu H, Jiang C Q, Read J M, Hay J A, et al. Life Course Exposures Continually Shape Antibody Profiles and Risk of Seroconversion to Influenza. *PLOS Pathogens* 2020;16:e1008635. <https://doi.org/10.1371/journal.ppat.1008635>
- [82] Baguelin M, Flasche S, Camacho A, Demiris N, Miller E, Edmunds W J. Assessing Optimal Target Populations for Influenza Vaccination Programmes: An Evidence Synthesis and Modelling Study. *PLOS Medicine* 2013;10:e1001527. <https://doi.org/10.1371/journal.pmed.1001527>

- [83] Jenness S M, Maloney K M, Smith D K, Hoover K W, Goodreau S M, Rosenberg E S, et al. Addressing Gaps in HIV Preexposure Prophylaxis Care to Reduce Racial Disparities in HIV Incidence in the United States. *American Journal of Epidemiology* 2019;188:743–52. <https://doi.org/10.1093/aje/kwy230>
- [84] Fox S J, Javan E, Pasco R, Gibson G C, Betke B, Herrera-Diestra J L, et al. Disproportionate Impacts of COVID-19 in a Large US City. *PLOS Computational Biology* 2023;19:e1011149. <https://doi.org/10.1371/journal.pcbi.1011149>
- [85] World Health Organization. Statement on the Second Meeting of the International Health Regulations (2005) Emergency Committee Regarding the Outbreak of Novel Coronavirus (2019-nCoV) n.d.
- [86] Map: Coronavirus and School Closures in 2019-2020. Education Week 2020
- [87] Campos J. TIMELINE | Penn State's Journey through the Coronavirus Pandemic 2021
- [88] Adams G B, Shannon J, Shannon S. Return to University Campuses Associated with 9% Increase in New COVID-19 Case Rate. *Medrxiv* 2020:2020. <https://doi.org/10.1101/2020.10.13.20212183>
- [89] Hadden J. What the Top 25 Colleges and Universities in the US Have Said about Their Plans to Reopen in Fall 2020, from Postponing the Semester to Offering More Remote Coursework n.d.
- [90] Arnold C R K, Srinivasan S, Rodriguez S, Rydzak N, Herzog C M, Gontu A, et al. A Longitudinal Study of the Impact of University Student Return to Campus on the SARS-CoV-2 Seroprevalence among the Community Members. *Scientific Reports* 2022;12:8586. <https://doi.org/10.1038/s41598-022-12499-5>
- [91] Lopman B, Liu C Y, Le Guillou A, Handel A, Lash T L, Isakov A P, et al. A Modeling Study to Inform Screening and Testing Interventions for the Control of SARS-CoV-2 on University Campuses. *Scientific Reports* 2021;11:5900. <https://doi.org/10.1038/s41598-021-85252-z>
- [92] Benneyan J, Gehrke C, Ilies I, Nehls N. Community and Campus COVID-19 Risk Uncertainty Under University Reopening Scenarios: Model-Based Analysis. *JMIR Public Health and Surveillance* 2021;7:e24292. <https://doi.org/10.2196/24292>
- [93] Harris P A, Taylor R, Minor B L, Elliott V, Fernandez M, O'Neal L, et al. The REDCap Consortium: Building an International Community of Software Platform Partners. *Journal of Biomedical Informatics* 2019;95:103208. <https://doi.org/10.1016/j.jbi.2019.103208>
- [94] Linzer D A, Lewis J B. poLCA: An R Package for Polytomous Variable Latent Class Analysis. *Journal of Statistical Software* 2011;42:1–29. <https://doi.org/10.18637/jss.v042.i10>

- [95] Conner M, Wilding S, Norman P. Does Intention Strength Moderate the Intention–Health Behavior Relationship for Covid-19 Protection Behaviors?. *Annals of Behavioral Medicine* 2024;58:92–9. <https://doi.org/10.1093/abm/kaad062>
- [96] McDonald J, McDonald P, Hughes C, Albarracín D. Recalling and Intending to Enact Health Recommendations: Optimal Number of Prescribed Behaviors in Multibehavior Messages. *Clinical Psychological Science : a Journal of the Association for Psychological Science* 2017;5:858–65. <https://doi.org/10.1177/2167702617704453>
- [97] Bolck A, Croon M, Hagenaars J. Estimating Latent Structure Models with Categorical Variables: One-Step Versus Three-Step Estimators. *Political Analysis* 2004;12:3–27. <https://doi.org/10.1093/pan/mph001>
- [98] Bezanson J, Edelman A, Karpinski S, Shah V B. Julia: A Fresh Approach to Numerical Computing. *SIAM Review* 2017;59:65–98. <https://doi.org/10.1137/141000671>
- [99] Flaxman S, Mishra S, Gandy A, Unwin H J T, Mellan T A, Coupland H, et al. Estimating the Effects of Non-Pharmaceutical Interventions on COVID-19 in Europe. *Nature* 2020;584:257–61. <https://doi.org/10.1038/s41586-020-2405-7>
- [100] Banholzer N, Weenen E van, Lison A, Cenedese A, Seeliger A, Kratzwald B, et al. Estimating the Effects of Non-Pharmaceutical Interventions on the Number of New Infections with COVID-19 during the First Epidemic Wave. *PLOS ONE* 2021;16:e252827. <https://doi.org/10.1371/journal.pone.0252827>
- [101] Brauner J M, Mindermann S, Sharma M, Johnston D, Salvatier J, Gavečiak T, et al. Inferring the Effectiveness of Government Interventions against COVID-19. *Science* 2021;371:eabd9338. <https://doi.org/10.1126/science.abd9338>
- [102] Ge Y, Zhang W-B, Wu X, Ruktanonchai C W, Liu H, Wang J, et al. Untangling the Changing Impact of Non-Pharmaceutical Interventions and Vaccination on European COVID-19 Trajectories. *Nature Communications* 2022;13:3106. <https://doi.org/10.1038/s41467-022-30897-1>
- [103] World Health Organization. *Diagnostics* 2024
- [104] Yang S, Rothman R E. PCR-based Diagnostics for Infectious Diseases: Uses, Limitations, and Future Applications in Acute-Care Settings. *The Lancet Infectious Diseases* 2004;4:337. [https://doi.org/10.1016/S1473-3099\(04\)01044-8](https://doi.org/10.1016/S1473-3099(04)01044-8)
- [105] World Health Organization. *Target Product Profiles* 2024
- [106] Chua A C, Cunningham J, Moussy F, Perkins M D, Formenty P. The Case for Improved Diagnostic Tools to Control Ebola Virus Disease in West Africa and

- How to Get There. PLOS Neglected Tropical Diseases 2015;9:e3734. <https://doi.org/10.1371/journal.pntd.0003734>
- [107] Zhou X-N, Bergquist R, Tanner M. Elimination of Tropical Disease through Surveillance and Response. Infectious Diseases of Poverty 2013;2:1. <https://doi.org/10.1186/2049-9957-2-1>
  - [108] PAHO. An Integrated Approach to Communicable Disease Surveillance. 2000
  - [109] Cragg L. Outbreak Response. Applied Communicable Disease Control 2018:134–51
  - [110] Gastanaduy P A, Redd S B, Clemons N S, Lee A D, Hickman C J, Rota P A, et al. Measles. Manual for the Surveillance of Vaccine-Preventable Diseases 2019
  - [111] Commissioner O of the. Coronavirus (COVID-19) Update: FDA Informs Public About Possible Accuracy Concerns with Abbott ID NOW Point-of-Care Test 2020
  - [112] Grassly N C, Pons-Salort M, Parker E P K, White P J, Ferguson N M, Imperial College COVID-19 Response Team. Comparison of Molecular Testing Strategies for COVID-19 Control: A Mathematical Modelling Study. The Lancet Infectious Diseases 2020;20:1381–9. [https://doi.org/10.1016/S1473-3099\(20\)30630-7](https://doi.org/10.1016/S1473-3099(20)30630-7)
  - [113] Ezhilan M, Suresh I, Nesakumar N. SARS-CoV, MERS-CoV and SARS-CoV-2: A Diagnostic Challenge. Measurement 2021;168:108335. <https://doi.org/10.1016/j.measurement.2020.108335>
  - [114] World Health Organization. Cholera 2023
  - [115] Essential Programme on Immunization (EPI), Immunization, Vaccines and Biologicals (IVB). Clinical Specimens for the Laboratory Confirmation and Molecular Epidemiology of Measles, Rubella, and CRS. Manual for the Laboratory-based Surveillance of Measles, Rubella, And Congenital Rubella Syndrome 2018
  - [116] German R R. Sensitivity and Predictive Value Positive Measurements for Public Health Surveillance Systems. Epidemiology 2000;11:720–7
  - [117] World Health Organization. Operational Thresholds. Meningitis Outbreak Response in Sub-Saharan Africa: WHO Guideline 2014
  - [118] Lewis R, Nathan N, Diarra L, Belanger F, Paquet C. Timely Detection of Meningococcal Meningitis Epidemics in Africa. The Lancet 2001;358:287–93. [https://doi.org/10.1016/S0140-6736\(01\)05484-8](https://doi.org/10.1016/S0140-6736(01)05484-8)
  - [119] GBD 2019 Child and Adolescent Communicable Disease Collaborators. The Unfinished Agenda of Communicable Diseases among Children and Adolescents before the COVID-19 Pandemic, 1990–2019: A Systematic Analysis of the Global

- Burden of Disease Study 2019. *The Lancet* 2023;402:313–35. [https://doi.org/10.1016/S0140-6736\(23\)00860-7](https://doi.org/10.1016/S0140-6736(23)00860-7)
- [120] Roser M, Ritchie H, Spooner F. Burden of Disease. Our World in Data 2023
  - [121] World Health Organization. Measles Outbreak Guide. Geneva, Switzerland: World Health Organization; 2022
  - [122] Atkins B D, Jewell C P, Runge M C, Ferrari M J, Shea K, Probert W J M, et al. Anticipating Future Learning Affects Current Control Decisions: A Comparison between Passive and Active Adaptive Management in an Epidemiological Setting. *Journal of Theoretical Biology* 2020;506:110380. <https://doi.org/10.1016/j.jtbi.2020.110380>
  - [123] Tao Y, Shea K, Ferrari M. Logistical Constraints Lead to an Intermediate Optimum in Outbreak Response Vaccination n.d.:20
  - [124] Grais R, Conlan A, Ferrari M, Djibo A, Le Menach A, Bjørnstad O, et al. Time Is of the Essence: Exploring a Measles Outbreak Response Vaccination in Niamey, Niger. *Journal of the Royal Society Interface* 2008;5:67–74. <https://doi.org/10.1098/rsif.2007.1038>
  - [125] Ferrari M J, Fermon F, Nackers F, Llosa A, Magone C, Grais R F. Time Is (Still) of the Essence: Quantifying the Impact of Emergency Meningitis Vaccination Response in Katsina State, Nigeria. *International Health* 2014;6:282–90. <https://doi.org/10.1093/inthealth/ihu062>
  - [126] World Health Organization. Confirming, Investigating and Managing an Outbreak. Response to Measles Outbreaks in Measles Mortality Reduction Settings: Immunization, Vaccines and Biologicals 2009;3
  - [127] Minetti A, Kagoli M, Katsulukuta A, Huerga H, Featherstone A, Chiotcha H, et al. Lessons and Challenges for Measles Control from Unexpected Large Outbreak, Malawi. *Emerging Infectious Diseases* 2013;19:202–9. <https://doi.org/10.3201/eid1902.120301>
  - [128] Cooper L V, Stuart J M, Okot C, Asiedu-Bekoe F, Afreh O K, Fernandez K, et al. Reactive Vaccination as a Control Strategy for Pneumococcal Meningitis Outbreaks in the African Meningitis Belt: Analysis of Outbreak Data from Ghana. *Vaccine* 2019;37:5657–63. <https://doi.org/10.1016/j.vaccine.2017.12.069>
  - [129] Zalwango M G, Zalwango J F, Kadobera D, Bulage L, Nanziri C, Migisha R, et al. Evaluation of Malaria Outbreak Detection Methods, Uganda, 2022. *Malaria Journal* 2024;23:18. <https://doi.org/10.1186/s12936-024-04838-w>
  - [130] Warrener L, Andrews N, Koroma H, Alessandrini I, Haque M, Garcia C C, et al. Evaluation of a Rapid Diagnostic Test for Measles IgM Detection; Accuracy and the Reliability of Visual Reading Using Sera from the Measles Surveillance

- Programme in Brazil, 2015. *Epidemiology & Infection* 2023;151:e151. <https://doi.org/10.1017/S0950268823000845>
- [131] Miller E, Sikes H D. Addressing Barriers to the Development and Adoption of Rapid Diagnostic Tests in Global Health. *Nanobiomedicine* 2015;2:6. <https://doi.org/10.5772/61114>
  - [132] Brown D W, Warrener L, Scobie H M, Donadel M, Waku-Kouomou D, Mulders M N, et al. Rapid Diagnostic Tests to Address Challenges for Global Measles Surveillance. *Current Opinion in Virology* 2020;41:77–84. <https://doi.org/10.1016/j.coviro.2020.05.007>
  - [133] McMorrow M L, Aidoo M, Kachur S P. Malaria Rapid Diagnostic Tests in Elimination Settings—Can They Find the Last Parasite?. *Clinical Microbiology and Infection : the Official Publication of the European Society of Clinical Microbiology and Infectious Diseases* 2011;17:1624–31. <https://doi.org/10.1111/j.1469-0691.2011.03639.x>
  - [134] Larremore D B, Wilder B, Lester E, Shehata S, Burke J M, Hay J A, et al. Test Sensitivity Is Secondary to Frequency and Turnaround Time for COVID-19 Screening. *Science Advances* 2021;7:eabd5393. <https://doi.org/10.1126/sciadv.abd5393>
  - [135] Middleton C, Larremore D B. Modeling the Transmission Mitigation Impact of Testing for Infectious Diseases. *Science Advances* 2024;10:eadk5108. <https://doi.org/10.1126/sciadv.adk5108>
  - [136] FIND. Target Product Profile for Surveillance Tests for Measles and Rubella. Geneva, Switzerland: 2024
  - [137] Shonhai A, Warrener L, Mangwanya D, Slibinskas R, Brown K, Brown D, et al. Investigation of a Measles Outbreak in Zimbabwe, 2010: Potential of a Point of Care Test to Replace Laboratory Confirmation of Suspected Cases. *Epidemiology and Infection* 2015;143:3442–50. <https://doi.org/10.1017/S0950268815000540>
  - [138] Senin A, Noordin N M, Sani J A M, Mahat D, Donadel M, Scobie H M, et al. A Measles IgM Rapid Diagnostic Test to Address Challenges with National Measles Surveillance and Response in Malaysia. *PLOS ONE* 2024;19:e298730. <https://doi.org/10.1371/journal.pone.0298730>
  - [139] Gillespie D T. Approximate Accelerated Stochastic Simulation of Chemically Reacting Systems. *The Journal of Chemical Physics* 2001;115:1716–33. <https://doi.org/10.1063/1.1378322>
  - [140] Chatterjee A, Vlachos D G, Katsoulakis M A. Binomial Distribution Based Tau-Leap Accelerated Stochastic Simulation. *The Journal of Chemical Physics* 2005;122:24112. <https://doi.org/10.1063/1.1833357>

- [141] Guerra F M, Bolotin S, Lim G, Heffernan J, Deeks S L, Li Y, et al. The Basic Reproduction Number ( $R_0$ ) of Measles: A Systematic Review. *The Lancet Infectious Diseases* 2017;17:e420–8. [https://doi.org/10.1016/S1473-3099\(17\)30307-9](https://doi.org/10.1016/S1473-3099(17)30307-9)
- [142] World Bank. Ghana 2024
- [143] World Health Organization. Measles Vaccination Coverage 2024
- [144] Masresha B G, Wiysonge C S, Katsande R, O'Connor P M, Lebo E, Perry R T. Tracking Measles and Rubella Elimination Progress—World Health Organization African Region, 2022–2023. *Vaccines* 2024;12:949. <https://doi.org/10.3390/vaccines12080949>
- [145] Keeling M J, Rohani P. Modeling Infectious Diseases in Humans and Animals. Princeton: Princeton University Press; 2008
- [146] Papadopoulos T, Vynnycky E. Estimates of the Basic Reproduction Number for Rubella Using Seroprevalence Data and Indicator-Based Approaches. *Plos Computational Biology* 2022;18:e1008858. <https://doi.org/10.1371/journal.pcbi.1008858>
- [147] Morales M, Lanzieri T, Reef S. Rubella. *CDC Yellow Book 2024: Health Information for International Travel* 2023
- [148] Jombart T, Ghozzi S, Schumacher D, Taylor T J, Leclerc Q J, Jit M, et al. Real-Time Monitoring of COVID-19 Dynamics Using Automated Trend Fitting and Anomaly Detection. *Philosophical Transactions of the Royal Society B: Biological Sciences* 2021;376:20200266. <https://doi.org/10.1098/rstb.2020.0266>
- [149] Stolerian L M, Clemente L, Poirier C, Parag K V, Majumder A, Masyn S, et al. Using Digital Traces to Build Prospective and Real-Time County-Level Early Warning Systems to Anticipate COVID-19 Outbreaks in the United States. *Science Advances* 2023;9:eabq199. <https://doi.org/10.1126/sciadv.abq0199>
- [150] Stern L, Lightfoot D. Automated Outbreak Detection: A Quantitative Retrospective Analysis. *Epidemiology and Infection* 1999;122:103–10
- [151] Salmon M, Schumacher D, Höhle M. Monitoring Count Time Series in R: Aberration Detection in Public Health Surveillance. *Journal of Statistical Software* 2016;70. <https://doi.org/10.18637/jss.v070.i10>
- [152] Teklehaimanot H D, Schwartz J, Teklehaimanot A, Lipsitch M. Alert Threshold Algorithms and Malaria Epidemic Detection. *Emerging Infectious Diseases* 2004;10:1220–6. <https://doi.org/10.3201/eid1007.030722>

- [153] Leclère B, Buckeridge D L, Boëlle P-Y, Astagneau P, Lepelletier D. Automated Detection of Hospital Outbreaks: A Systematic Review of Methods. *PLOS ONE* 2017;12:e176438. <https://doi.org/10.1371/journal.pone.0176438>
- [154] Kissler S M, Fauver J R, Mack C, Olesen S W, Tai C, Shiue K Y, et al. Viral Dynamics of Acute SARS-CoV-2 Infection and Applications to Diagnostic and Public Health Strategies. *PLOS Biology* 2021;19:e3001333. <https://doi.org/10.1371/journal.pbio.3001333>
- [155] Ratnam S, Tipples G, Head C, Fauvel M, Fearon M, Ward B J. Performance of Indirect Immunoglobulin M (IgM) Serology Tests and IgM Capture Assays for Laboratory Diagnosis of Measles. *Journal of Clinical Microbiology* 2000;38:99–104
- [156] Institute for Health Metrics and Evaluation. GBD Results 2024
- [157] Shattock A J, Johnson H C, Sim S Y, Carter A, Lambach P, Hutubessy R C W, et al. Contribution of Vaccination to Improved Survival and Health: Modelling 50 Years of the Expanded Programme on Immunization. *The Lancet* 2024;403:2307–16. [https://doi.org/10.1016/S0140-6736\(24\)00850-X](https://doi.org/10.1016/S0140-6736(24)00850-X)
- [158] Graham M, Winter A K, Ferrari M, Grenfell B, Moss W J, Azman A S, et al. Measles and the Canonical Path to Elimination. *Science* 2019;364:584–7. <https://doi.org/10.1126/science.aau6299>
- [159] Parums D V. A Review of the Resurgence of Measles, a Vaccine-Preventable Disease, as Current Concerns Contrast with Past Hopes for Measles Elimination. *Medical Science Monitor : International Medical Journal of Experimental and Clinical Research* 2024;30:e944436–1–e944436–10. <https://doi.org/10.12659/MSM.944436>
- [160] Masresha B G. Progress Toward Measles Elimination – African Region, 2017–2021. *MMWR Morbidity and Mortality Weekly Report* 2023;72. <https://doi.org/10.15585/mmwr.mm7236a3>
- [161] Azman A S, Luquero F J, Rodrigues A, Palma P P, Grais R F, Banga C N, et al. Urban Cholera Transmission Hotspots and Their Implications for Reactive Vaccination: Evidence from Bissau City, Guinea Bissau. *Plos Neglected Tropical Diseases* 2012;6:e1901. <https://doi.org/10.1371/journal.pntd.0001901>
- [162] Dergiades T, Milas C, Mossialos E, Panagiotidis T. Effectiveness of Government Policies in Response to the First COVID-19 Outbreak. *PLOS Global Public Health* 2022;2:e242. <https://doi.org/10.1371/journal.pgph.0000242>
- [163] Murray J, Cohen A L. Infectious Disease Surveillance. *International Encyclopedia of Public Health* 2016:222. <https://doi.org/10.1016/B978-0-12-803678-5.00517-8>

- [164] Brett T S, O'Dea E B, Marty É, Miller P B, Park A W, Drake J M, et al. Anticipating Epidemic Transitions with Imperfect Data. *Plos Computational Biology* 2018;14:e1006204. <https://doi.org/10.1371/journal.pcbi.1006204>
- [165] Brett T S, Drake J M, Rohani P. Anticipating the Emergence of Infectious Diseases. *Journal of the Royal Society, Interface* 2017;14:20170115. <https://doi.org/10.1098/rsif.2017.0115>
- [166] Drake J M, Brett T S, Chen S, Epureanu B I, Ferrari M J, Marty É, et al. The Statistics of Epidemic Transitions. *PLOS Computational Biology* 2019;15:e1006917. <https://doi.org/10.1371/journal.pcbi.1006917>
- [167] Scheffer M, Bascompte J, Brock W A, Brovkin V, Carpenter S R, Dakos V, et al. Early-Warning Signals for Critical Transitions. *Nature* 2009;461:53–9. <https://doi.org/10.1038/nature08227>
- [168] Scheffer M. Foreseeing Tipping Points. *Nature* 2010;467:411–2. <https://doi.org/10.1038/467411a>
- [169] Dakos V, Scheffer M, van Nes E H, Brovkin V, Petoukhov V, Held H. Slowing down as an Early Warning Signal for Abrupt Climate Change. *Proceedings of the National Academy of Sciences* 2008;105:14308–12. <https://doi.org/10.1073/pnas.0802430105>
- [170] Drake J M, Griffen B D. Early Warning Signals of Extinction in Deteriorating Environments. *Nature* 2010;467:456–9. <https://doi.org/10.1038/nature09389>
- [171] Boettiger C, Hastings A. Quantifying Limits to Detection of Early Warning for Critical Transitions. *Journal of the Royal Society Interface* 2012. <https://doi.org/10.1098/rsif.2012.0125>
- [172] Southall E, Tildesley M J, Dyson L. Prospects for Detecting Early Warning Signals in Discrete Event Sequence Data: Application to Epidemiological Incidence Data. *PLOS Computational Biology* 2020;16:e1007836. <https://doi.org/10.1371/journal.pcbi.1007836>
- [173] O'Regan S M, Drake J M. Theory of Early Warning Signals of Disease Emergence and Leading Indicators of Elimination. *Theoretical Ecology* 2013;6:333–57. <https://doi.org/10.1007/s12080-013-0185-5>
- [174] Drake J M, Hay S I. Monitoring the Path to the Elimination of Infectious Diseases. *Tropical Medicine and Infectious Disease* 2017;2:20. <https://doi.org/10.3390/tropicalmed2030020>
- [175] Southall E, Brett T S, Tildesley M J, Dyson L. Early Warning Signals of Infectious Disease Transitions: A Review. *Journal of the Royal Society Interface* 2021;18:20210555. <https://doi.org/10.1098/rsif.2021.0555>

- [176] Dablander F, Heesterbeek H, Borsboom D, Drake J M. Overlapping Timescales Obscure Early Warning Signals of the Second COVID-19 Wave. *Proceedings of the Royal Society B: Biological Sciences* 2022;289:20211809. <https://doi.org/10.1098/rspb.2021.1809>
- [177] Southall E, Tildesley M J, Dyson L. How Early Can an Upcoming Critical Transition Be Detected? 2022:2022. <https://doi.org/10.1101/2022.05.27.22275693>
- [178] Brett T S, Rohani P. Dynamical Footprints Enable Detection of Disease Emergence. *PLOS Biology* 2020;18:e3000697. <https://doi.org/10.1371/journal.pbio.3000697>
- [179] Clements C F, Ozgul A. Including Trait-Based Early Warning Signals Helps Predict Population Collapse. *Nature Communications* 2016;7:10984. <https://doi.org/10.1038/ncomms10984>
- [180] O'Brien D A, Clements C F. Early Warning Signal Reliability Varies with COVID-19 Waves. *Biology Letters* 2021;17:20210487. <https://doi.org/10.1098/rsbl.2021.0487>
- [181] Clements C F, Blanchard J L, Nash K L, Hindell M A, Ozgul A. Body Size Shifts and Early Warning Signals Precede the Historic Collapse of Whale Stocks. *Nature Ecology & Evolution* 2017;1:1–6. <https://doi.org/10.1038/s41559-017-0188>
- [182] Clements C F, McCarthy M A, Blanchard J L. Early Warning Signals of Recovery in Complex Systems. *Nature Communications* 2019;10:1681. <https://doi.org/10.1038/s41467-019-09684-y>
- [183] Brett T, Ajelli M, Liu Q-H, Krauland M G, Grefenstette J J, Panhuis W G van, et al. Detecting Critical Slowing down in High-Dimensional Epidemiological Systems. *PLOS Computational Biology* 2020;16:e1007679. <https://doi.org/10.1371/journal.pcbi.1007679>
- [184] Kendall M G. The Treatment of Ties in Ranking Problems. *Biometrika* 1945;33:239–51. <https://doi.org/10.1093/biomet/33.3.239>
- [185] Knight W R. A Computer Method for Calculating Kendall's Tau with Ungrouped Data. *Journal of the American Statistical Association* 1966;61:436–9. <https://doi.org/10.1080/01621459.1966.10480879>
- [186] Flach P A. ROC Analysis. *Encyclopedia of Machine Learning and Data Mining* 2016:1–8. [https://doi.org/10.1007/978-1-4899-7502-7\\_739-1](https://doi.org/10.1007/978-1-4899-7502-7_739-1)
- [187] Clark T G, Bradburn M J, Love S B, Altman D G. Survival Analysis Part I: Basic Concepts and First Analyses. *British Journal of Cancer* 2003;89:232–8. <https://doi.org/10.1038/sj.bjc.6601118>

- [188] Kaida Y, Kanbayashi D, Kurata T, Mori H. Contribution of Parvovirus B19 in Suspected Cases of Measles/Rubella in Osaka, Japan, between 2011 and 2021. *Journal of Medical Virology* 2023;95:e28593. <https://doi.org/10.1002/jmv.28593>
- [189] Pull L, Brichler S, Bouchaud O, Siriez J-Y. Differential Diagnosis of Dengue Fever: Beware of Measles!. *Journal of Travel Medicine* 2012;19:268–71. <https://doi.org/10.1111/j.1708-8305.2012.00628.x>
- [190] Tinto B, Bicaba B, Kagoné T S, Kayiwa J, Rabe I, Merle C S C, et al. Co-Circulation of Two Alphaviruses in Burkina Faso: Chikungunya and O'nyong Nyong Viruses. *PLOS Neglected Tropical Diseases* 2024;18:e11712. <https://doi.org/10.1371/journal.pntd.0011712>
- [191] Gama Dessavre A, Southall E, Tildesley M J, Dyson L. The Problem of Detrending When Analysing Potential Indicators of Disease Elimination. *Journal of Theoretical Biology* 2019;481:183–93. <https://doi.org/10.1016/j.jtbi.2019.04.011>
- [192] Lenton T M, Livina V N, Dakos V, Nes E H V, Scheffer M. Early Warning of Climate Tipping Points from Critical Slowing down: Comparing Methods to Improve Robustness. *Philosophical Transactions Series A, Mathematical, Physical, And Engineering Sciences* 2012;370:1185. <https://doi.org/10.1098/rsta.2011.0304>
- [193] Carpenter S R, Cole J J, Pace M L, Batt R, Brock W A, Cline T, et al. Early Warnings of Regime Shifts: A Whole-Ecosystem Experiment. *Science* 2011;332:1079–82. <https://doi.org/10.1126/science.1203672>
- [194] Dudney J, Suding K N. The Elusive Search for Tipping Points. *Nature Ecology & Evolution* 2020;4:1449–50. <https://doi.org/10.1038/s41559-020-1273-8>
- [195] Pitzer V E, Chitwood M, Havumaki J, Menzies N A, Perniciaro S, Warren J L, et al. The Impact of Changes in Diagnostic Testing Practices on Estimates of COVID-19 Transmission in the United States. *American Journal of Epidemiology* 2021;190:1908–17. <https://doi.org/10.1093/aje/kwab089>
- [196] Portnoy A, Jit M, Helleringer S, Verguet S. Impact of Measles Supplementary Immunization Activities on Reaching Children Missed by Routine Programs. *Vaccine* 2018;36:170–8. <https://doi.org/10.1016/j.vaccine.2017.10.080>
- [197] Portnoy A, Jit M, Helleringer S, Verguet S. Comparative Distributional Impact of Routine Immunization and Supplementary Immunization Activities in Delivery of Measles Vaccine in Low- and Middle-Income Countries. *Value in Health : the Journal of the International Society for Pharmacoeconomics and Outcomes Research* 2020;23:891–7. <https://doi.org/10.1016/j.jval.2020.03.012>
- [198] Lessler J, Metcalf C J E, Grais R F, Luquero F J, Cummings D A T, Grenfell B T. Measuring the Performance of Vaccination Programs Using Cross-Sectional

- Surveys: A Likelihood Framework and Retrospective Analysis. Plos Medicine 2011;8:e1001110. <https://doi.org/10.1371/journal.pmed.1001110>
- [199] Whittles L K, White P J, Didelot X. A Dynamic Power-Law Sexual Network Model of Gonorrhoea Outbreaks. Plos Computational Biology 2019;15:e1006748. <https://doi.org/10.1371/journal.pcbi.1006748>
- [200] Spicknall I H. Modeling the Impact of Sexual Networks in the Transmission of Monkeypox Virus Among Gay, Bisexual, and Other Men Who Have Sex With Men – United States, 2022. MMWR Morbidity and Mortality Weekly Report 2022;71. <https://doi.org/10.15585/mmwr.mm7135e2>
- [201] Uong S, Rosenberg E S, Goodreau S M, Luisi N, Sullivan P, Jenness S M. Assessment of Bias in Estimates of Sexual Network Degree Using Prospective Cohort Data. Epidemiology (Cambridge, Mass) 2020;31:229–37. <https://doi.org/10.1097/EDE.0000000000001151>
- [202] Drumright L N, Gorbach P M, Holmes K K. Do People Really Know Their Sex Partners?: Concurrency, Knowledge of Partner Behavior, and Sexually Transmitted Infections Within Partnerships. Sexually Transmitted Diseases 2004;31:437–42. <https://doi.org/10.1097/01.OLQ.0000129949.30114.37>
- [203] Nelson S J, Manhart L E, Gorbach P M, Martin D H, Stoner B P, Aral S O, et al. Measuring Sex Partner Concurrency: It's What's Missing That Counts. Sexually Transmitted Diseases 2007;34:801–7. <https://doi.org/10.1097/OLQ.0b013e318063c734>
- [204] Ito H, Shigeta K, Yamamoto T, Morita S. Exploring Sexual Contact Networks by Analyzing a Nationwide Commercial-Sex Review Website. PLOS ONE 2022;17:e276981. <https://doi.org/10.1371/journal.pone.0276981>
- [205] Glennon E E, Bruijning M, Lessler J, Miller I F, Rice B L, Thompson R N, et al. Challenges in Modeling the Emergence of Novel Pathogens. Epidemics 2021;37:100516. <https://doi.org/10.1016/j.epidem.2021.100516>
- [206] Lloyd-Smith J O, Funk S, McLean A R, Riley S, Wood J L. Nine Challenges in Modelling the Emergence of Novel Pathogens. Epidemics 2015;10:35–9. <https://doi.org/10.1016/j.epidem.2014.09.002>
- [207] Gostic K M, McGough L, Baskerville E B, Abbott S, Joshi K, Tedijanto C, et al. Practical Considerations for Measuring the Effective Reproductive Number, Rt. PLOS Computational Biology 2020;16:e1008409. <https://doi.org/10.1371/journal.pcbi.1008409>
- [208] Abbott S, Hellewell J, Thompson R N, Sherratt K, Gibbs H P, Bosse N I, et al. Estimating the Time-Varying Reproduction Number of SARS-CoV-2 Using

- National and Subnational Case Counts. Wellcome Open Research 2020;5:112. <https://doi.org/10.12688/wellcomeopenres.16006.2>
- [209] Vos T, Lim S S, Abbafati C, Abbas K M, Abbasi M, Abbasifard M, et al. Global Burden of 369 Diseases and Injuries in 204 Countries and Territories, 1990–2019: A Systematic Analysis for the Global Burden of Disease Study 2019. *The Lancet* 2020;396:1204–22. [https://doi.org/10.1016/S0140-6736\(20\)30925-9](https://doi.org/10.1016/S0140-6736(20)30925-9)
- [210] Winter A K, Lambert B, Klein D, Klepac P, Papadopoulos T, Truelove S, et al. Feasibility of Measles and Rubella Vaccination Programmes for Disease Elimination: A Modelling Study. *The Lancet Global Health* 2022;10:e1412–22. [https://doi.org/10.1016/S2214-109X\(22\)00335-7](https://doi.org/10.1016/S2214-109X(22)00335-7)
- [211] Giesecke J. Routine Surveillance of Infectious Diseases. *Modern Infectious Disease Epidemiology* 2016;137–48
- [212] Little R J A. A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association* 1988;83:1198–202. <https://doi.org/10.1080/01621459.1988.10478722>
- [213] Kuhn M, Wickham H. Tidymodels: A Collection of Packages for Modeling and Machine Learning Using Tidyverse Principles 2020
- [214] Kuhn M, Johnson K. Feature Engineering and Selection: A Practical Approach for Predictive Models. Taylor & Francis; 2019

# Vita | Callum Arnold

## Education

PhD Biology, <i>Pennsylvania State University</i>	December 2024 (Expected)
MSc Global Health Policy, <i>London School of Hygiene &amp; Tropical Medicine</i>	June 2020
MChem Chemistry, <i>Oxford University</i>	June 2015

## Awards

Braddock Award, <i>Pennsylvania State University</i>	2020-2021
University Graduate Fellowship, <i>Pennsylvania State University</i>	2020-2021
Board of Examiner's Academic Achievement Award, <i>London School of Hygiene &amp; Tropical Medicine</i>	June 2020

## Select Publications

1. Shankar, M., Hartner, A., **Arnold, C.**, Gayawan, E., Kang, H., *et al.* (In-Press). How mathematical modelling can inform outbreak response vaccination, *BMC Infectious Diseases*
2. Bolotin, S., Hughes, S. L., Savage, R. D., McLachlan, E., Severini, A., **Arnold, C.**, *et al.* (2023). Maternal varicella antibodies in children aged less than one year: Assessment of antibody decay. *PLOS ONE*, 18(11), e0287765. <https://doi.org/10.1371/journal.pone.0287765>
3. Wright, J., Science, M., Osman, S., **Arnold, C.**, Sumaida, M., *et al.* (2023). Uptake of pertussis immunization in pregnancy and determinants of vaccination in Toronto, Canada. *Vaccine*. <https://doi.org/10.1016/j.vaccine.2023.10.020>
4. **Arnold, C.**, Srinivasan, S., Rodriguez, S., Rydzak, N., Herzog, C. M., *et al.* (2022). A longitudinal study of the impact of university student return to campus on the SARS-CoV-2 seroprevalence among the community members. *Scientific Reports*, 12(1), 8586. <https://doi.org/10.1038/s41598-022-12499-5>
5. Piché-Renaud, P.P., Groves, H., Kitano, T., **Arnold, C.**, Thomas, A., *et al.* (2020). Healthcare workers' perception of a global outbreak of novel coronavirus (COVID-19) and personal protective equipment: Survey of a pediatric tertiary care hospital. *Infection Control & Hospital Epidemiology*, 1–25. <https://doi.org/10.1017/ice.2020.415>
6. Science, M., Savage, R., Severini, A., McLachlan, E., Hughes, S. L., **Arnold, C.**, *et al.* (2019). Measles Antibody Levels in Young Infants. *Pediatrics*, e20190630. <https://doi.org/10.1542/peds.2019-0630>