The Pennsylvania State University

The Graduate School

DISSERTATION TITLE

A Dissertation in

Department of Biology

by

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
Chair of Committee

Dr. Matthew J. Ferrari
Major Field Member & Dissertation Advisor

Dr. Maciej Boni
Major Field Member

Dr. Stephen Berg
Outside Unit & Field Member

Abstract

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Test

ACKNOWLEDGEMENTS

Testing of acknowledgments section

New Lines

Chapter 1 | Introduction

Much of the world is complex, and in seeking to understand it we must make aggregations and assumptions to create groups ***[REF]***. 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 e.g., as grades, is essential to the synthesis and interpretation of information, particularly for action; as it would be impractical to simultaneously evaluate every score a student achieved during school ***[REF]***. Infectious diseases are no different. Rather than tracking measles viral loads in a population, for example, individuals are broadly categorized as susceptible, infected, or removed ***[REF]***. Cases are counted, providing estimates of the number of new and cumulative infected individuals, respectively, at any one time; without these groupings, it would be hard to answer questions about disease spread and burden, and allocate resources like preventative vaccines appropriately ***[REF]***. 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 compounding these categorization decisions.

In the first half of my dissertation (Chapters 2 & 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 to shut down, requiring their students to return to their respective homes ***[REF]***. 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 now 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 in use calculations of seroprevalence (the proportion of a population that have sufficiently high levels of antibodies, indicating past exposure to a pathogen). 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. 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 ***[REF]***, and demonstrate the limits of non-pharmaceutical interventions alone to reduce infections within the student population.

In the second half of my dissertation (Chapters 4 & 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 [1,2]. 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 ***[REF]***. In practice, no threshold will be able to perfectly discriminate between these groups of individuals, leading to classification errors ***[REF]***. 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 ***[REF]***. The specificity is the opposite: the ability to correct detect the *lack* of infection in an uninfected individual, also called the true negative rate ***[REF]***. 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 ***[REF]***. 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, and attempts to separate a time series of test positive cases by suspected outbreak status will face similar issues of sensitivity/specificity/PPV/NPV. 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, $R_{\text{effective}} = 1$. $R_{\text{effective}}$ is the average number of secondary infections each infectious individual is expected to generate before they recover (given the current population size and susceptibility), where 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 policy-makers to account for the compounding layers of uncertainty, finding zones of 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 | The Maximal Expected Benefit of SARS-CoV-2 Interventions Among University Students: A Simulation Study Using Latent Class Analysis

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

[1] Department of Biology, Pennsylvania State University, University Park, PA, USA 16802
[2] Center for Infectious Disease Dynamics, Pennsylvania State University, University Park, PA, USA 16802
[3] Ross & Carole Nese College of Nursing, Pennsylvania State University, University Park, PA, USA 16802
[4] College of Health and Human Development, Pennsylvania State University, University Park, PA, USA 16802
[5] Social Science Research Institute, Pennsylvania State University, University Park, PA, USA 16802
[6] Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA, USA 16802
[7] Department of Veterinary and Biomedical Sciences, Pennsylvania State University, University Park, PA, USA 16802
[8] 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.

* Corresponding author. Callum R.K. Arnold. Address: Department of Biology, Pennsylvania State University, University Park, PA, USA 16802. Email: contact@callumarnold.com.

Background

Within epidemiology, the importance of heterogeneity, whether that host, population, statistical, or environmental, has long been recognized [3–7]. For example, when designing targeted interventions, it is crucial to understand and account for differences that may exist within populations [8–10]. 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 [3,4,8,11–16]. 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 [17–21]. When examining optimal influenza vaccination policy in the United Kingdom, Baguelin et al. [22] 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 [23]. Similarly, epidemiological models have demonstrated the potential for HIV pre-exposure prophylaxis to reduce racial disparities in HIV incidence [24]. Therefore, heterogeneity can be used to inform more complete theories of change, increasing intervention effectiveness [25]

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 [9,11,17,19,22,24,26]. 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 paper 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 [27], universities across the United States began to close their campuses and accommodations, shifting to remote instruction [28,29]. By Fall 2020, academic institutions transitioned to a hybrid working environment (in-person and online), requiring students to return to campuses [30–32]. In a prior paper [33] 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 [34]) on the community infection rates, testing serum for the presence of anti-Spike Receptor Binding Domain (S/RBD) IgG, indicating prior exposure [35]. Despite widespread concern that campus re-openings would lead to substantial increases in surrounding community infections [30,36,37], very little sustained transmission was observed between the two geographically coincident populations [33].

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 [38], 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 [38]. 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 [33], but in brief, students were eligible for the student cohort if they were: ≥ 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 [39] 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 [40]. 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 [41]. Further details in the Supplement of the previous paper [33].

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 [42] in the R programming language, version 4.3.3 (2024-02-29) [43]. 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 [44,45]. We examined the latent class models using Bayesian Information Criterion, which is a commonly recommended as part of LCA model evaluation [46,47], 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 [47,48]. The following variables were determined a priori to be potential risk factors for exposure [33]: 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 [49], as described in the supplement of the previous paper [33].

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 (ϕ), and a scaling factor for the whole transmission matrix (ρ). The off-diagonal values are equal to a within-group value (diagonal) multiplied by a scaling factor (ϕ). This scaling factor can either multiply the within-group beta value of the source group (e.g., $\beta_{HL} = \phi \cdot \beta_{LL}$; Eq. 1A), or the recipient group (e.g., $\beta_{LH} = \phi \cdot \beta_{LL}$; Eq. 1B), each with a different interpretation.

$$\rho \begin{pmatrix} \beta_{HH} & \beta_{HM} & \beta_{HL} \\ \beta_{MH} & \beta_{HM} & \beta_{ML} \\ \beta_{LH} & \beta_{HM} & \beta_{LL} \end{pmatrix} \to \rho \begin{pmatrix} \beta_{HH} & \phi\beta_{MM} & \boldsymbol{\phi\beta_{LL}} \\ \phi\beta_{HH} & \beta_{MM} & \boldsymbol{\phi\beta_{LL}} \\ \phi\beta_{HH} & \phi\beta_{MM} & \boldsymbol{\beta_{LL}} \end{pmatrix} \text{ mixing structure } \boldsymbol{A}$$

$$\to \rho \begin{pmatrix} \beta_{HH} & \phi\beta_{HH} & \phi\beta_{HH} \\ \phi\beta_{MM} & \beta_{MM} & \beta_{MM} \\ \boldsymbol{\phi\beta_{LL}} & \boldsymbol{\phi\beta_{LL}} & \boldsymbol{\beta_{LL}} \end{pmatrix} \text{ mixing structure } \boldsymbol{B} \tag{1}$$

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 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 prevents infection, rather than protecting against onwards transmission. A range of between-group scaling values (ϕ) were simulated to perform sensitivity analysis for the degree of assortativity. Results are only shown for matrix structure $\boldsymbol{A}$, but alternative assumptions about between-group mixing can be found in the supplement (Supplemental Figures 1-4). 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 [50].

Results

Demographics

Full details can be found in the prior paper [33], 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 1). Two 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 (43.4% and 53.1% respectively), and avoiding face-touching with unwashed hands (43.5%).

The four- and the three-class LCA models had the lowest BIC respectively (Table 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 Tables 1 & 2).

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 [33]. The mean and median BIC and AIC indicated similar predictive ability of the three- and four-class LCA models (Table 3). Given these factors, the three-class model was selected for use in simulation for parsimony, requiring fewer assumptions and parameters to fit.

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 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 5).

Compartmental Model

The ABC distance distributions indicated that near-homogeneous levels of between-group mixing better fit the data (Figure 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 2).

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 [51–54]. 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 Figures 1-4).

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 [38], 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" [46], whereby groups are assigned and then specific causal meaning is given to each cluster, affecting subsequent analyses and interpretation of results. In this paper, 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.

Additional Information

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 report.

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 and Dr. Matthew J. Ferrari (Department of Biology, Pennsylvania State University) conducted the data analysis.

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.

Author Contributions

Conceptualization: CA, MJF

Data curation: CA, MJF

Formal analysis: CA, MJF

Funding acquisition: MJF

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

Methodology: CA, NB, MJF

Project administration: MJF

Software: CA, MJF

Supervision: MJF

Validation: CA, MJF

Visualization: CA, MJF

Writing - original draft: CA

Writing - review and editing: all authors.

Acknowledgements

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

Figures

Figure 1

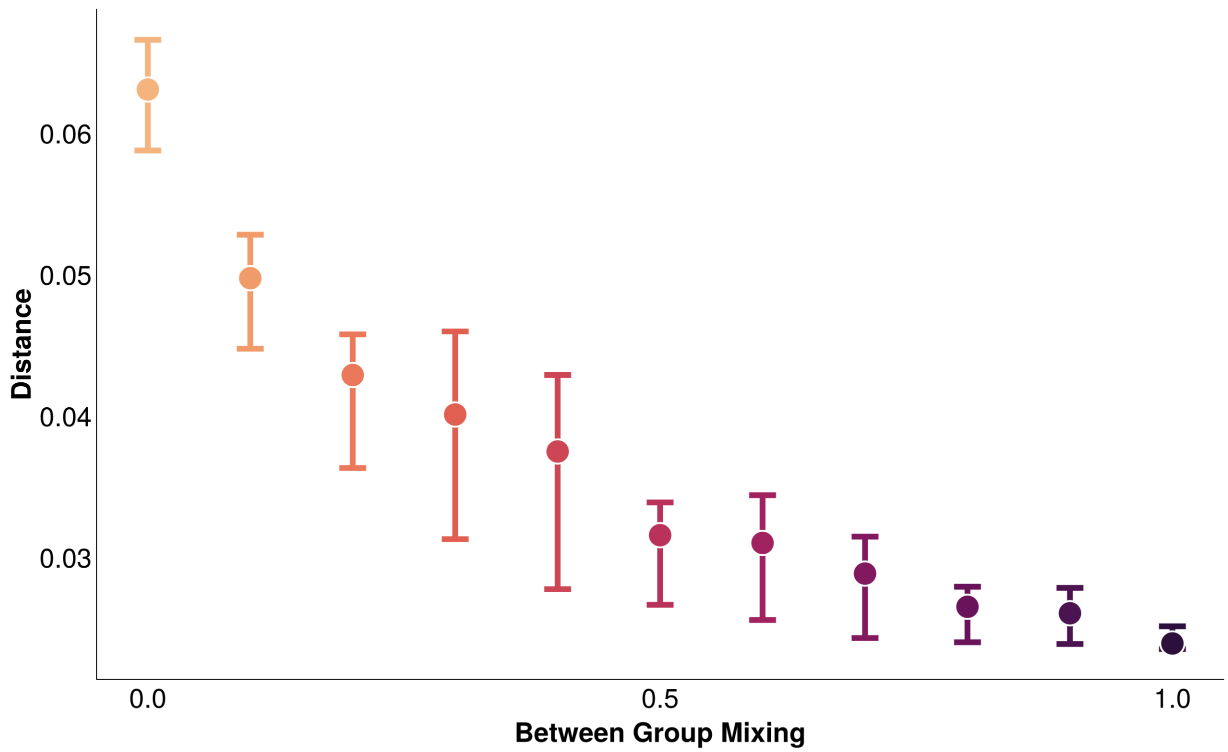

Figure 1: Distribution of the distance from the ABC fits, with the minimum and maximum distances illustrated by the whiskers, and the median distance by the point. Between-group mixing of 1.0 equates to between-group mixing as likely as within-group mixing

Figure 2

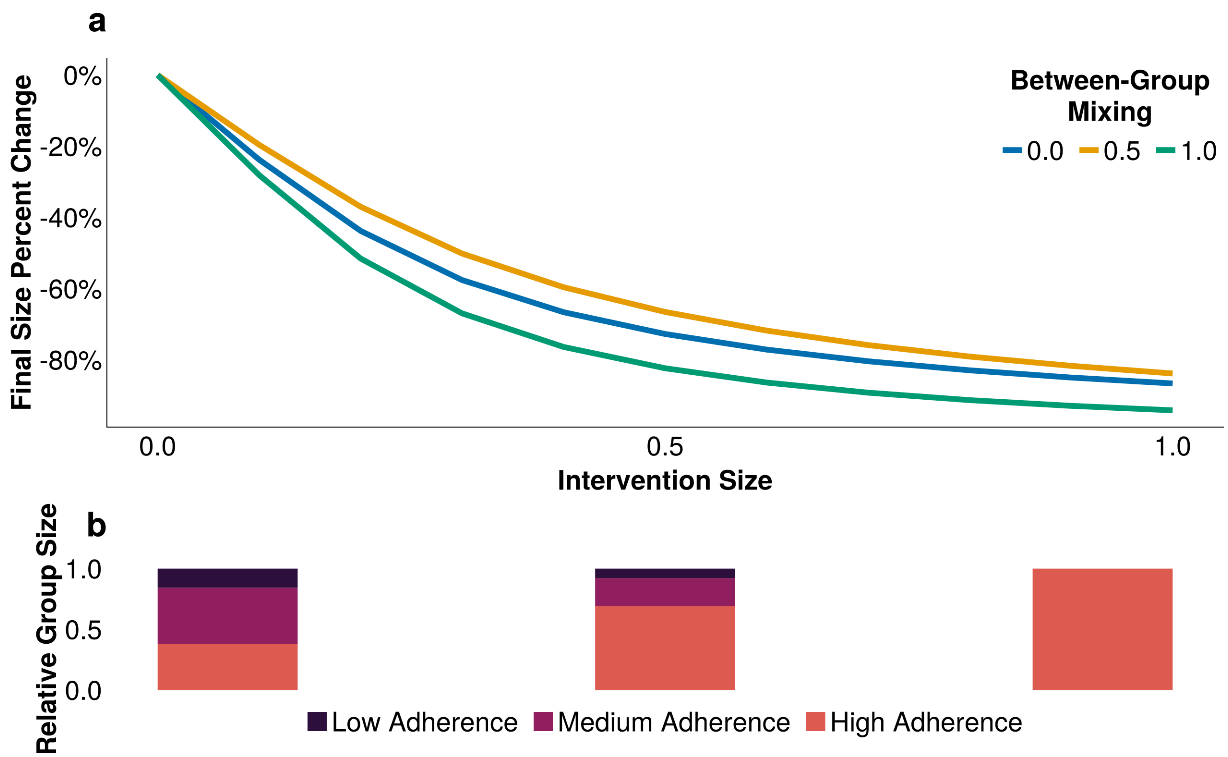

Figure 2: A) The reduction in final infection size across a range of intervention effectiveness (1.0 is a fully effective intervention), accounting for a range of assortativity. Between-group mixing of 1.0 equates to between-group mixing as likely as within-group mixing; B) The relative distribution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

Tables

Table 1

Intention to always:	Always	Not Always
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%)

Table 1: Participants' intention to always or not always follow 8 public health measures

Table 2

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

Table 2: Log likelihood, AIC, and BIC of two to seven class LCA model fits

Table 3

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

Table 3: Mean and median AIC and BIC of multiply-imputed logistic regressions for two to seven class LCA models against IgG serostatus

Table 4

Measure **Intention to Always:**	**Low Adherence**	**Medium Adherence**	**High Adherence**
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
Group Size	15.75%	46.21%	38.04%
Seroprevalence	37.7%	32.2%	25.4%

Table 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

Table 5

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)

Table 5: Adjusted odds ratio (aOR) for risk factors of infection among the returning PSU UP student cohort

Chapter 3 | Individual and Population Level Uncertainty Interact to Determine Performance of Outbreak Detection

Callum R.K. Arnold[1,2], Alex C. Kong[3], Amy K. Winter[4], William J. Moss[3,5], Bryan N. Patenaude[3], Matthew J. Ferrari[1,2]

[1] Department of Biology, Pennsylvania State University, University Park, PA, USA 16802
[2] Center for Infectious Disease Dynamics, Pennsylvania State University, University Park, PA, USA 16802
[3] Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA 21205
[4] Department of Epidemiology, College of Public Health, University of Georgia, Athens, GA, USA 30602
[5] 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 an enzyme-linked immunosorbent assay (ELISA). 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 ELISA-like diagnostic tests (c. 93%), given moderately high testing rates. With larger delays (14 days) between the ELISA test administration and result date, RDTs could outperform the ELISA. Similar numbers of unavoidable cases and outbreak alert delays could be achieved between the test types. With dynamical noise, however, the accuracy of ELISA scenarios was far superior to those achieved with RDTs (c. 93% vs. 73%). For dynamical noise, RDT-based scenarios typically favored more sensitive alert threshold than ELISA-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, RDTs can perform as well as ELISA 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

At the heart of an outbreak detection system is a surveillance program, often utilizing individual diagnostic tests as required components of case detection and epidemiological investigations before an outbreak can be declared [1,55–58]. For diseases with non-specific symptoms, accurate measurement tools are often necessary to confidently and correctly ascribe changes in symptom prevalence within a population to a particular disease, and therefore detect outbreaks of specific pathogens. As a result, it has been commonplace for surveillance systems to be developed around high-accuracy tests, such as Polymerase Chain Reaction (PCR) tests and immunoglobulin (Ig) tests, when financially and logistically feasible [59–64]. Depending on the disease in question, either sensitivity (the ability to correctly detect a true positive individual) or specificity (the ability to correctly discount a true negative) will be prioritized, as they often are at odds with each other [65–67]. This balance is commonly defined within the Target Product Profile (TPP) of a test [68], which is a set of minimum characteristics that should be met for production and widespread use, helping to guide research and development. For example, in the wake of the 2013 Ebola outbreak in Guinea a TPP was developed that listed the minimum acceptable values of sensitivity and specificity as 95% and 99%, respectively [69]. Recognizing that Ebola is not the major cause of fever and other non-specific symptoms in the region, it is arguably more important to prioritize the specificity of the disease; however, the authors note that the disease severity requires a high level of sensitivity, as the consequences of a missed case are dire at an individual and population level [69].

Much like the accuracy of an individual test, outbreak detection systems face the same issue regarding the prioritization of sensitive or specific alerts [70–72]. For many disease systems, particularly in resource constrained environments where the burden of infectious diseases is typically highest [73,74], 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 [57,71]. 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 [75–80]. 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 resources for future response capabilities. Just as the balance of sensitivity and specificity of a test for an individual must be carefully evaluated, so must the balance at the outbreak level.

The concept of using incidence-based alert triggers to define the discrete event of an "outbreak" with characteristics analogous to individual tests has been well documented in the case of meningitis, measles, and malaria [57,72,79,81–84]. 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 [85]. Despite concerns of their lower diagnostic accuracy slowing their adoption until recently [86], the reduced cold-chain requirements [87] and faster speed of result provided by RDTs has been show to outweigh the cost of false positive/negative results in some settings [85,88–90].

In this paper we examine how the use of imperfect diagnostic tests affects the performance of outbreak detection for measles outbreaks in the context of a febrile rash surveillance system that includes both measles and non-measles cases. Because measles symptoms are non-specific, it is important to account for non-measles sources of febrile rash e.g., rubella, parvovirus, varicella, etc., producing the potential for false positive results in the context of imperfect tests. Currently, measles outbreaks are declared on the basis of either suspected measles cases (i.e., an individual with fever and maculopapular rash [91]) alone, cases confirmed by enzyme immunoassay / enzyme-linked immunosorbent assay, from here-on in referred to as ELISA, to detect the presence of measles-specific IgM antibodies, or a combination of the two. Countries at or near elimination status are encouraged to primarily use PCR or ELISA diagnostic tests for the confirmation of suspected cases [91]. Each of these detection systems have its flaws. Although clinical case definition is very fast and requires minimal resources, it is highly sensitive, and in the face of high “background noise” from non-measles sources of febrile rash, can lead to low positive predictive value (PPV), i.e., the probability that an alert accurately reflects the outbreak status [92]. And while ELISA confirmation is the standard diagnostic test for measles surveillance and has higher specificity [91], the training and facility requirements generally mean that samples must be transported from the point of care to a separate laboratory, which incurs both costs and delays [57,64,87]. In resource-poor settings these delays may be days to weeks [87]. A rapid diagnostic test may meet the WHO’s definition and requirements for using in a surveillance setting, if not for individual patient care [93]. Recent developments show encouraging signs in the field, as well as in theory, providing a compromise between diagnostic accuracy and timeliness in most [85,87,94], though not all [95], settings.

By examining the combination of alert threshold and individual test characteristic in a modeling study that explicitly incorporates dynamical background noise, we aim to illustrate the need to develop a TPP for the whole detection system, not just one component. To evaluate the alert system performance, we develop a set of outbreak definition criteria and surveillance metrics, drawing inspiration from acceptance sampling, ecological surveillance systems, and epidemiological surveillance system guidelines and reviews [96–101]. Using these metrics we overcome issues encountered by early warning systems that rely solely on dynamical values such as $R_{\text{effective}}$ in defining outbreaks [102–106], for example, characterizing the end of an epidemic period is important in a time series where multiple outbreaks will occur.

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 [107]. We utilized binomial draws to ensure compartment sizes remained positive valued [108]. We assumed that the transmission rate (β_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 [59,109]. 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) [110]. 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 [111,112]. 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 [113]. The full table of parameters can be found in Table 6. All simulations and analysis was completed in Julia version 1.10.5 [50], with all code stored at https://github.com/arnold-c/OutbreakDetection.

Parameters	Measles	Dynamical noise
R0	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	

Table 6: Compartmental model parameters

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 6) [114,115]. The seasonality for the rubella noise was simulated to be in-phase with measles, anti-phase with measles (peak timing 6 months later), or non-seasonal. Only dynamical in-phase noise and Poisson-only noise are presented in the main text; the anti-phase and non-seasonal dynamical noise scenarios are presented in the supplement.

For each noise structure, we simulated five magnitudes of noise (Λ), representing the average daily noise incidence. Λ 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 Λ (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 [102,104], 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) [99,106,116,117].

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 great 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 RDT equivalent 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 [118]

- An RDT equivalent with 90% sensitivity and specificity, and 0-day lag in result return [87]
- 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 [119], 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 3 a). We then add non-measles noise (Figure 3 b) and test according to the testing scenario, which yields 5 time series of test-positive cases (Figure 3 c): one time series of all clinically compatible cases and 4 reflecting the testing scenarios.

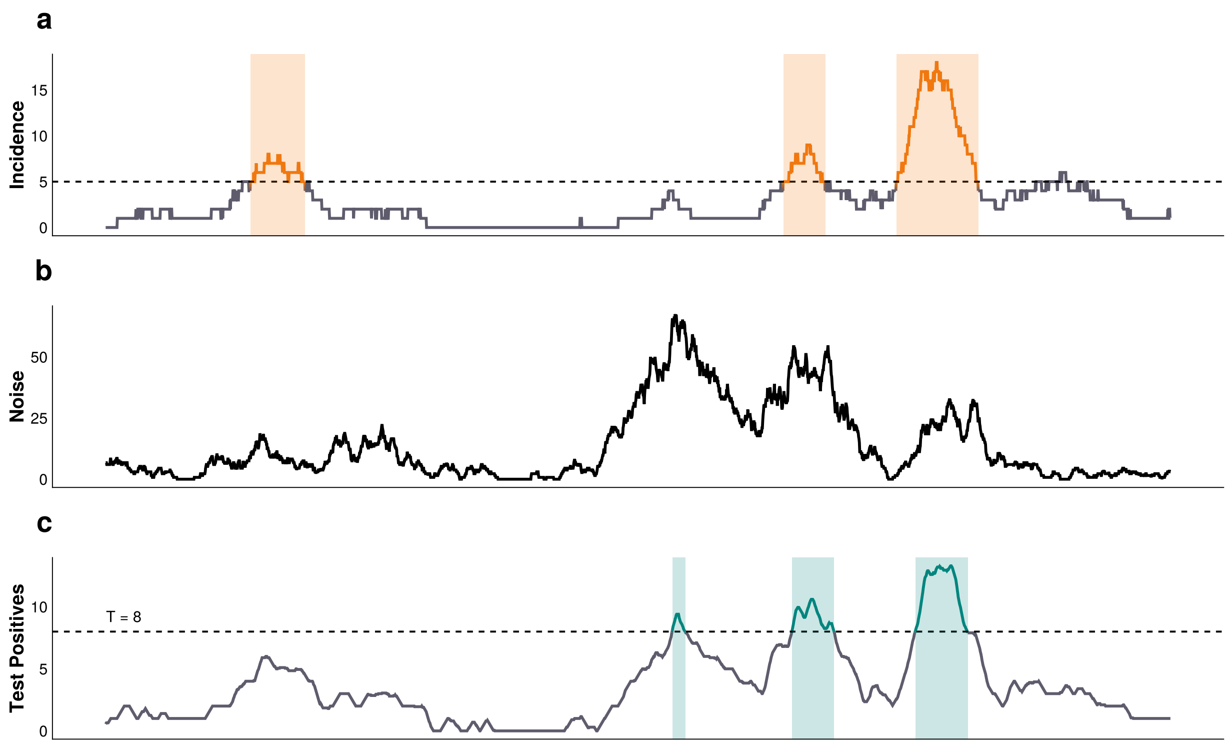

Figure 3: A schematic of the outbreak definition and alert detection system. A) Measles incidence time series. B) Noise time series. C) Observed time series of test positive cases according to a given testing scenario. In panel A, the orange bands represent regions of the measles time series that meet the outbreak definition criteria. In panel C, the green 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 alert and outbreak periods (Figure 3 b and c). 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 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 3 b and c). 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 after a correct alert is first triggered 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 imperfect and delays in 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_O, 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 7). When the average noise incidence was 8 times higher than the average measles incidence ($\Lambda(8)$), the optimal threshold ranged between 1 and 7 test-positive cases per day. Not surprisingly, the biggest driver of this difference was the testing rate; as a large fraction of suspected cases are tested, the optimal threshold increases monotonically for all test and noise types (Table 7).

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 cases, to $\approx$ 90% accuracy at $\geq$ 20% testing, for all test types (Figure 4). For dynamical SEIR noise, the ELISA-like perfect tests perform identically to the Poisson noise case at all magnitudes (Figure 4). For RDT-like tests, which have lower individual sensitivity and specificity, the maximal attainable accuracy is lower than the ELISA-like tests for all testing rates (P) at noise magnitude $\geq \Lambda(2)$ (Figure 4). 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).

	Test Characteristic		**Testing Rate**					
Noise Type	Test Type	Test Lag	10%	20%	30%	40%	50%	60%
Dynamical noise: in-phase	RDT Equivalent (85.0%)	0	1	2	3	4	4	5
Dynamical noise: in-phase	RDT Equivalent (90.0%)	0	1	2	3	4	4	5
Poisson noise	RDT Equivalent (85.0%)	0	1	3	4	5	6	7
Poisson noise	RDT Equivalent (90.0%)	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

Table 7: Optimal threshold for RDT-like, ELISA-like, and perfect tests, under dynamical and Poisson-like noise structures where the average daily noise incidence is 8 times the average daily measles incidence

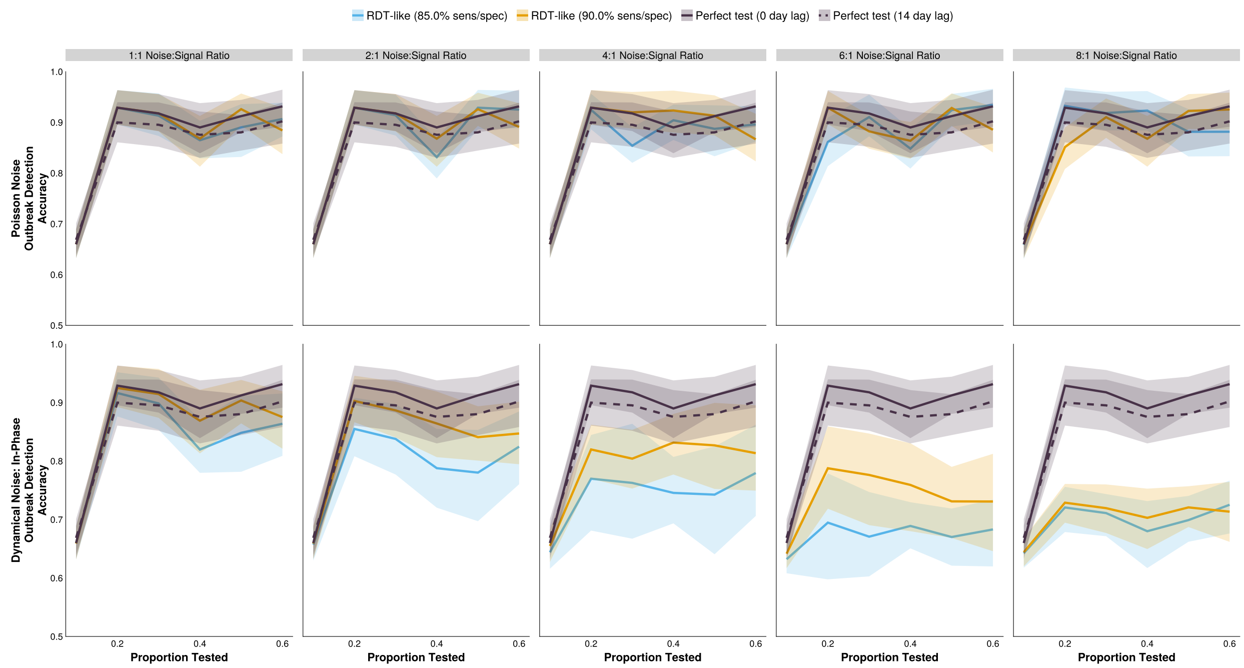

Figure 4: The accuracy of outbreak detection systems under different testing rates and noise structures. The shaded bands illustrate the 80% central interval, and the solid/dashed lines represent the mean estimate. Solid lines represent tests with 0-day turnaround times, and dashed lines represent tests with result delays.

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 an ELISA-like 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 test. This always leads to an increase in the median delay from outbreak start to alert, relative to an ELISA-like test with no result delays, as well as RDT-like tests (Figure 5).

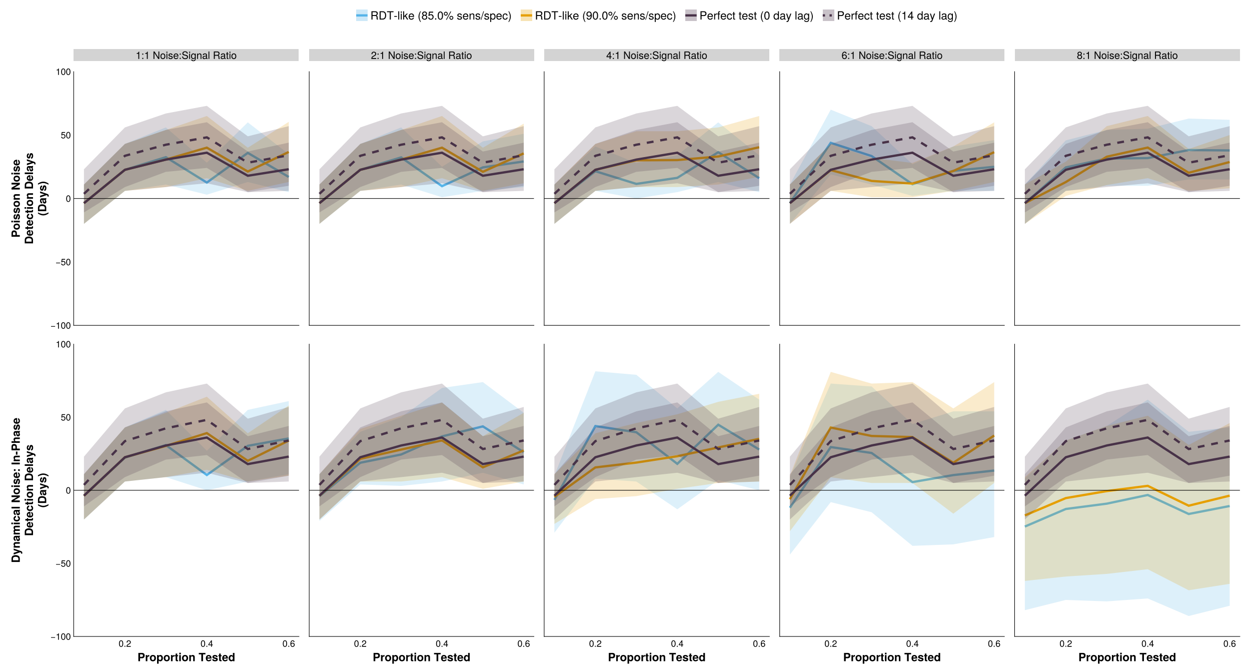

Figure 5: The detection delay of outbreak detection systems under different testing rates and noise structures. The shaded bands illustrate the 80% central interval, and the solid/dashed lines represent the mean estimate. Solid lines represent tests with 0-day turnaround times, and dashed lines represent tests with result delays.

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 ELISA and RDTs, with the separation of outcomes occurring later than observed for surveillance accuracy ($\Lambda(8)$ vs $\Lambda(2)$ — Figure 5 and Figure 4 — respectively). With large amounts of dynamical noise ($\Lambda(8)$), the mean detection delay of the 90% and 85% RDTs 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 6, Supplemental Figures 2 and 3). Long detection delays manifest as large numbers of unavoidable cases (i.e., cases that occur between the outbreak start and its detection) (Figure 7). 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.

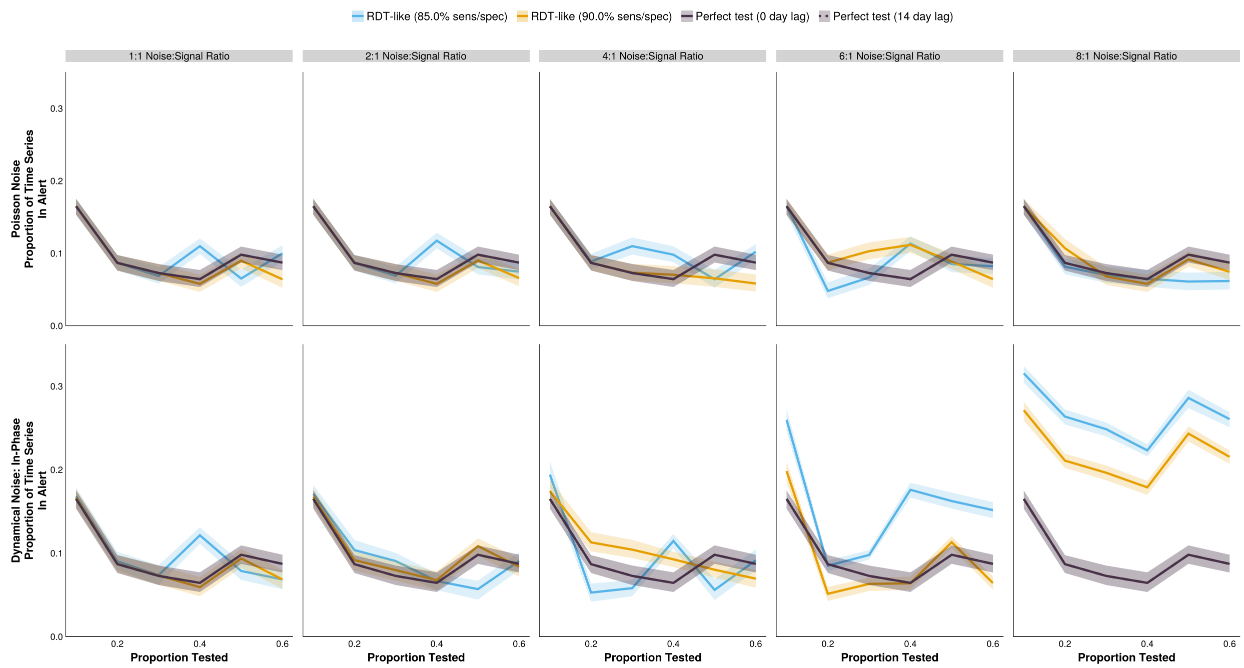

Figure 6: The difference between the proportion of the time series in alert for outbreak detection systems under different testing rates and noise structures. The shaded bands illustrate the 80% central interval, and the solid/dashed lines represent the mean estimate. Solid lines represent tests with 0-day turnaround times, and dashed lines represent tests with result delays.

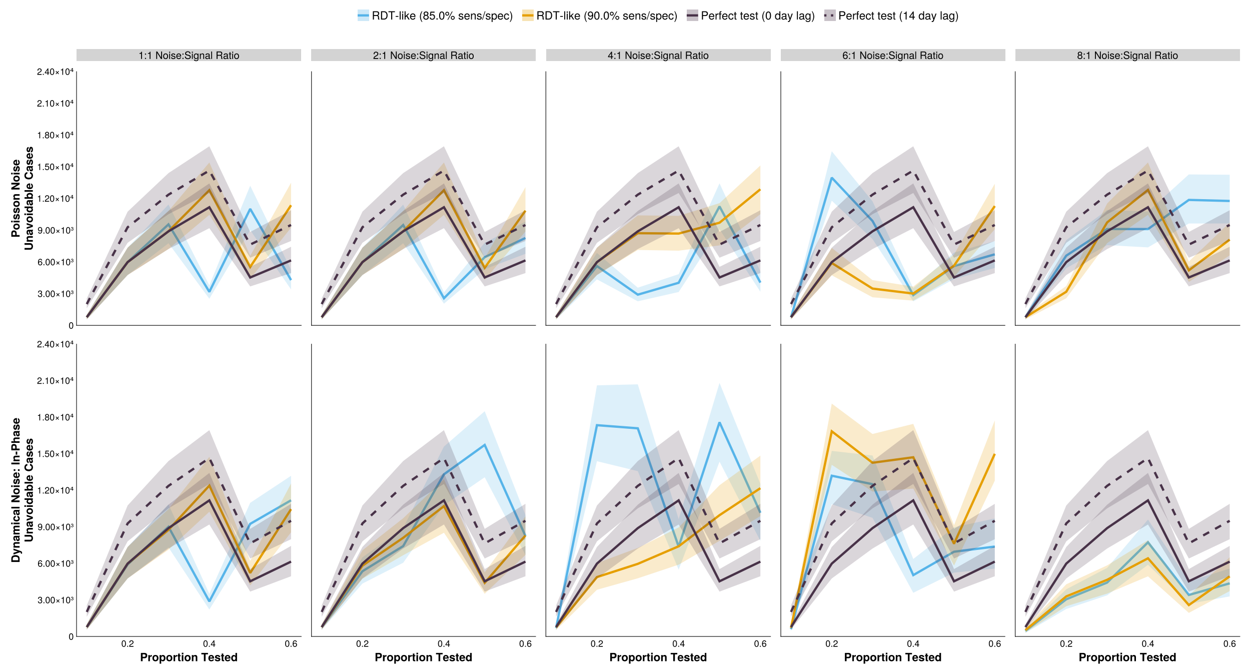

Figure 7: The number of unavoidable cases of outbreak detection systems under different testing rates and noise structures. The shaded bands illustrate the 80% central interval, and the solid/dashed lines represent the mean estimate. Solid lines represent tests with 0-day turnaround times, and dashed lines represent tests with result delays.

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), RDTs can perform as well, if not better than ELISA tests at moderate to high testing rates, and at a fraction of the cost [87]. 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 ELISA's 93%. 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 effects of imperfect diagnostic tests that produce higher rates of false positives and negatives than ELISA-like diagnostics.

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 5), unavoidable cases (Figure 7), and proportion of the time series in alert status (Figure 6).

Surveillance is counting for action [2]. 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 RDTs 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, is has been well documented that the performance of an individual test is highly sensitive to its timing within a person's infection cycle [59,89,90,120,121], 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.

Funding

- *Something about GAVI/Gates*

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 report.

Acknowledgements

Author Contributions

Conceptualization: CA, MJF

Data curation: MJF, CA

Formal analysis: CA, MJF

Funding acquisition: MJF, WM, AW

Investigation: CA, MJF

Methodology: CA, MJF

Project administration: MJF

Software: CA

Supervision: MJF, WM, AW, BP

Validation: CA, MJF

Visualization: CA

Writing - original draft: CA, 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 and Dr. Matthew J. Ferrari (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 4 | Synthesis

Summary

Appendix A | Supplementary Material for Chapter 3

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

[1] Department of Biology, Pennsylvania State University, University Park, PA, USA 16802
[2] Center for Infectious Disease Dynamics, Pennsylvania State University, University Park, PA, USA 16802
[3] Ross & Carole Nese College of Nursing, Pennsylvania State University, University Park, PA, USA 16802
[4] College of Health and Human Development, Pennsylvania State University, University Park, PA, USA 16802
[5] Social Science Research Institute, Pennsylvania State University, University Park, PA, USA 16802
[6] Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA, USA 16802
[7] Department of Veterinary and Biomedical Sciences, Pennsylvania State University, University Park, PA, USA 16802
[8] Department of Animal Science, Pennsylvania State University, University Park, PA, USA 16802

† Corresponding author. Callum R.K. Arnold. Address: Department of Biology, Pennsylvania State University, University Park, PA, USA 16802. Email: contact@callumarnold.com.

Results

LCA Model Fitting

Measure Intention to Always:	Low Adherence	Low-Medium Adherence	Medium-High Adherence	High Adherence
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
Group Size	13.82%	30.91%	16.49%	38.78%
Seroprevalence	35.50%	31.20%	36.00%	25.70%

Supplemental Table 8: 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

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 (Supplemental Eq 1A), rather than conferring protection for the practitioner (Supplemental Eq 1B). Another alternative uses a single scaled value of β_{LL}, representing all between-group interactions experiencing the same risk of transmission that is a fraction of the transmission observed between Low Adherence individuals (Supplemental Eq 1C).

$$\rho\begin{pmatrix}\beta_{HH} & \beta_{HM} & \beta_{HL}\\ \beta_{MH} & \beta_{HM} & \beta_{ML}\\ \beta_{LH} & \beta_{HM} & \beta_{LL}\end{pmatrix} \to \rho\begin{pmatrix}\beta_{HH} & \phi\beta_{MM} & \boldsymbol{\phi\beta_{LL}}\\ \phi\beta_{HH} & \beta_{MM} & \boldsymbol{\phi\beta_{LL}}\\ \phi\beta_{HH} & \phi\beta_{MM} & \boldsymbol{\beta_{LL}}\end{pmatrix} \text{ mixing structure } \boldsymbol{A}$$

$$\to \rho\begin{pmatrix}\beta_{HH} & \phi\beta_{HH} & \phi\beta_{HH}\\ \phi\beta_{MM} & \beta_{MM} & \beta_{MM}\\ \boldsymbol{\phi\beta_{LL}} & \boldsymbol{\phi\beta_{LL}} & \boldsymbol{\beta_{LL}}\end{pmatrix} \text{ mixing structure } \boldsymbol{B} \qquad (2)$$

$$\to \rho\begin{pmatrix}\beta_{HH} & \boldsymbol{\phi\beta_{LL}} & \boldsymbol{\phi\beta_{LL}}\\ \boldsymbol{\phi\beta_{LL}} & \beta_{MM} & \boldsymbol{\phi\beta_{LL}}\\ \boldsymbol{\phi\beta_{LL}} & \boldsymbol{\phi\beta_{LL}} & \beta_{LL}\end{pmatrix} \text{ mixing structure } \boldsymbol{C}$$

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.

Eq 1B (PHMs Confer Protection)

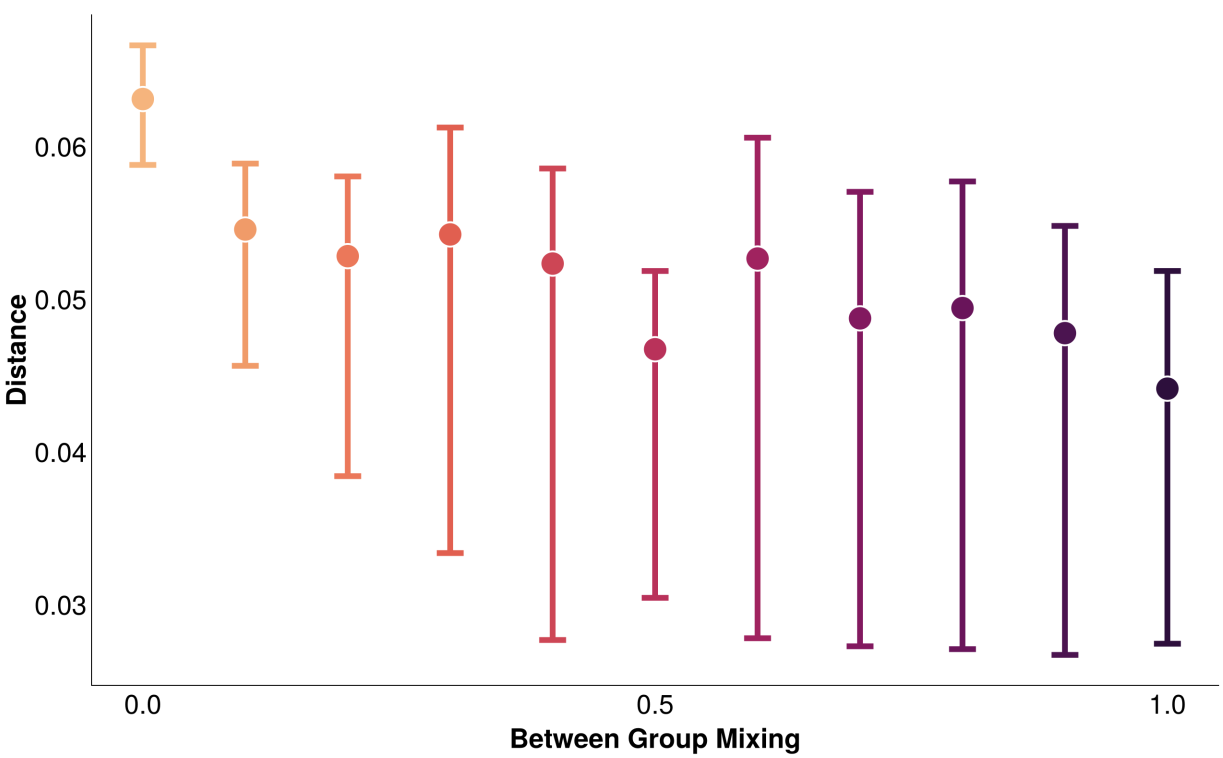

Figure 8: PHMs confer protection to the practitioner. Distribution of the distance from the ABC fits, with the minimum and maximum distances illustrated by the whiskers, and the median distance by the point. Between-group mixing of 1.0 equates to between-group mixing as likely as within-group mixing

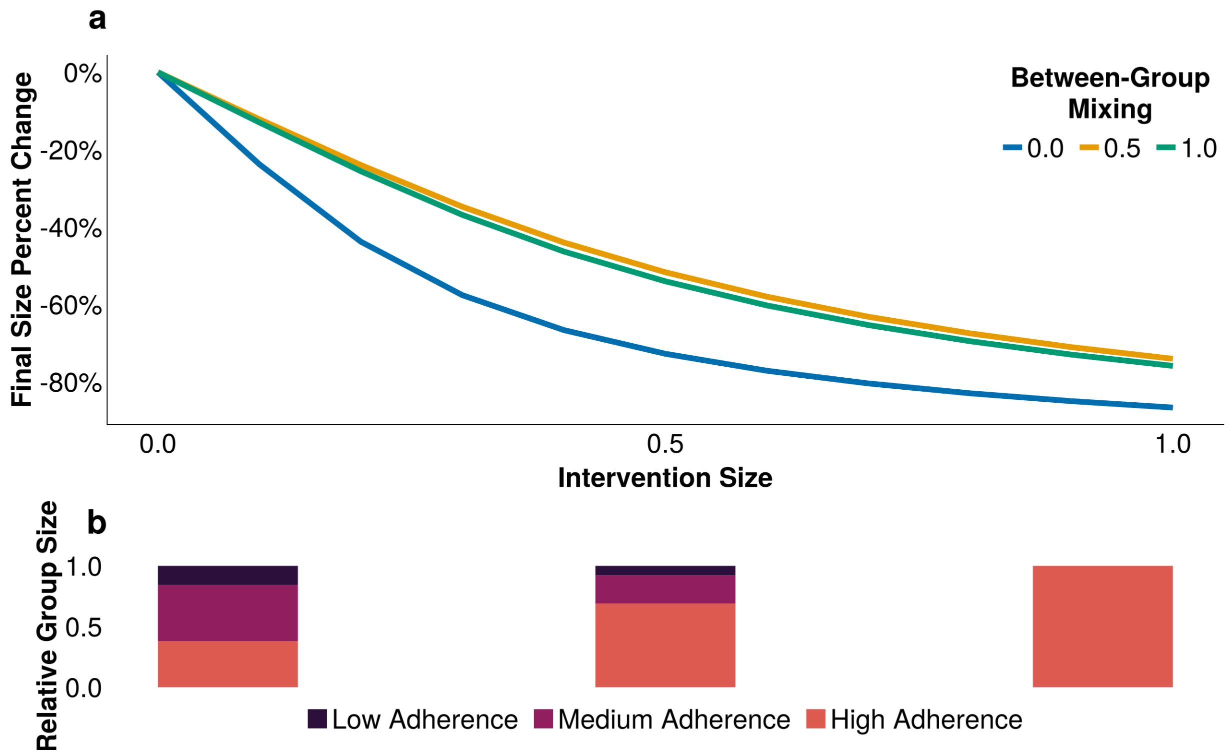

Figure 9: PHMs confer protection to the practitioner. A) The reduction in final infection size across a range of intervention effectiveness (1.0 is a fully effective intervention), accounting for a range of assortativity. Between-group mixing of 1.0 equates to between-group mixing as likely as within-group 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)

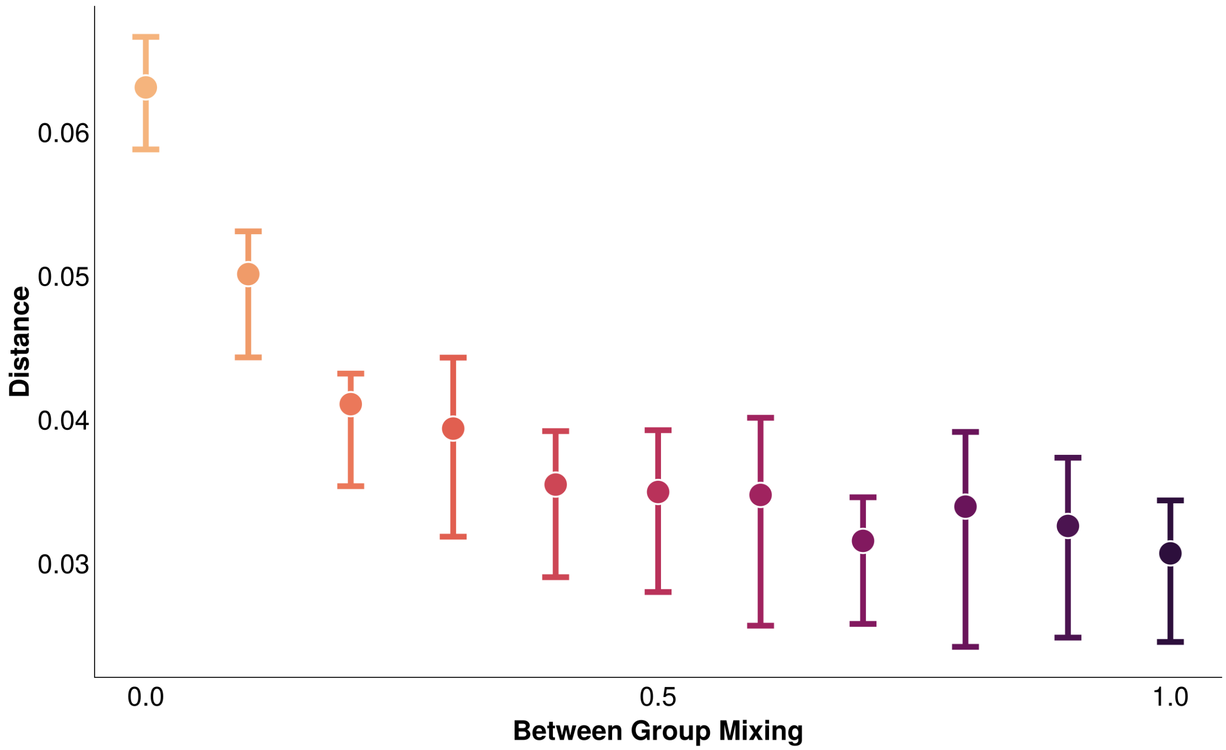

Figure 10: Identical off-diagonal values. Distribution of the distance from the ABC fits, with the minimum and maximum distances illustrated by the whiskers, and the median distance by the point. Between-group mixing of 1.0 equates to between-group mixing as likely as within-group mixing

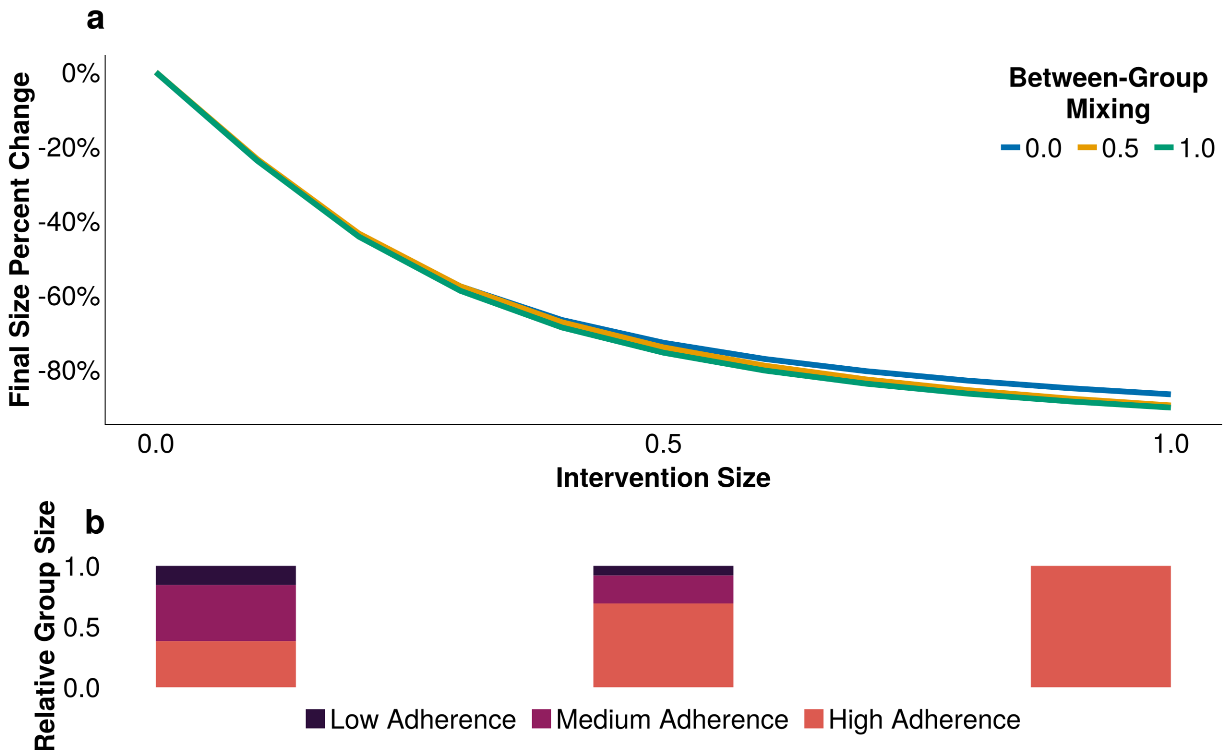

Figure 11: Identical off-diagonal values. A) The reduction in final infection size across a range of intervention effectiveness (1.0 is a fully effective intervention), accounting for a range of assortativity. Between-group mixing of 1.0 equates to between-group mixing as likely as within-group mixing; B) The relative distribution of group sizes at three levels of intervention effectiveness (0.0, 0.5, 1.0)

Appendix B | Supplementary Material for Chapter 4

Callum R.K. Arnold[1,2,‡], Alex C. Kong[3], Amy K. Winter[4], William J. Moss[3,5], Bryan N. Patenaude[3], Matthew J. Ferrari[1,2]

[1] Department of Biology, Pennsylvania State University, University Park, PA, USA 16802
[2] Center for Infectious Disease Dynamics, Pennsylvania State University, University Park, PA, USA 16802
[3] Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA 21205
[4] Department of Epidemiology, College of Public Health, University of Georgia, Athens, GA, USA 30602
[5] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA 21205

‡ Corresponding author. Callum R.K. Arnold. Address: Department of Biology, Pennsylvania State University, University Park, PA, USA 16802. Email: contact@callumarnold.com.

Results

Tables

	Test Characteristic		**Testing Rate**					
Noise Type	Test Type	Test Lag	10%	20%	30%	40%	50%	60%
Dynamical noise: in-phase	RDT Equivalent (85.0%)	0	0.64	0.72	0.71	0.68	0.7	0.73
Dynamical noise: in-phase	RDT Equivalent (90.0%)	0	0.64	0.73	0.72	0.71	0.72	0.71
Poisson noise	RDT Equivalent (85.0%)	0	0.66	0.93	0.92	0.93	0.88	0.88
Poisson noise	RDT Equivalent (90.0%)	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

Table 9: Mean outbreak detection accuracy of each testing scenario at their specific optimal thresholds, when the average noise incidence is 8 times higher than the average measles incidence. A) the noise structure is dynamical, and the seasonality is in-phase with the measles incidence. B) the noise structure is Poisson only.

	Test Characteristic		**Testing Rate**					
Noise Type	Test Type	Test Lag	10%	20%	30%	40%	50%	60%
Dynamical noise: in-phase	RDT Equivalent (85.0%)	0	452	3053	4424	7728	3406	4361
Dynamical noise: in-phase	RDT Equivalent (90.0%)	0	515	3289	4650	6417	2578	4933
Poisson noise	RDT Equivalent (85.0%)	0	766	6592	9111	9107	11865	11765
Poisson noise	RDT Equivalent (90.0%)	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

Table 10: Mean unavoidable cases per annum of each testing scenario at their specific optimal thresholds, scaled up to Ghana's 2022 population, when the average noise incidence is 8 times higher than the average measles incidence. A) the noise structure is dynamical, and the seasonality is in-phase with the measles incidence. B) the noise structure is Poisson only.

	Test Characteristic		**Testing Rate**					
Noise Type	Test Type	Test Lag	10%	20%	30%	40%	50%	60%
Dynamical noise: in-phase	RDT Equivalent (85.0%)	0	-24.82	-12.79	-9.15	-3.21	-16.22	-10.77
Dynamical noise: in-phase	RDT Equivalent (90.0%)	0	-17.21	-5.34	-0.55	3.03	-10.55	-3.66
Poisson noise	RDT Equivalent (85.0%)	0	-3.75	24.49	31.45	31.85	38.17	37.87
Poisson noise	RDT Equivalent (90.0%)	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

Table 11: Mean outbreak alert delay (days) of each testing scenario at their specific optimal thresholds, when the average noise incidence is 8 times higher than the average measles incidence. A) the noise structure is dynamical, and the seasonality is in-phase with the measles incidence. B) the noise structure is Poisson only.

Figures

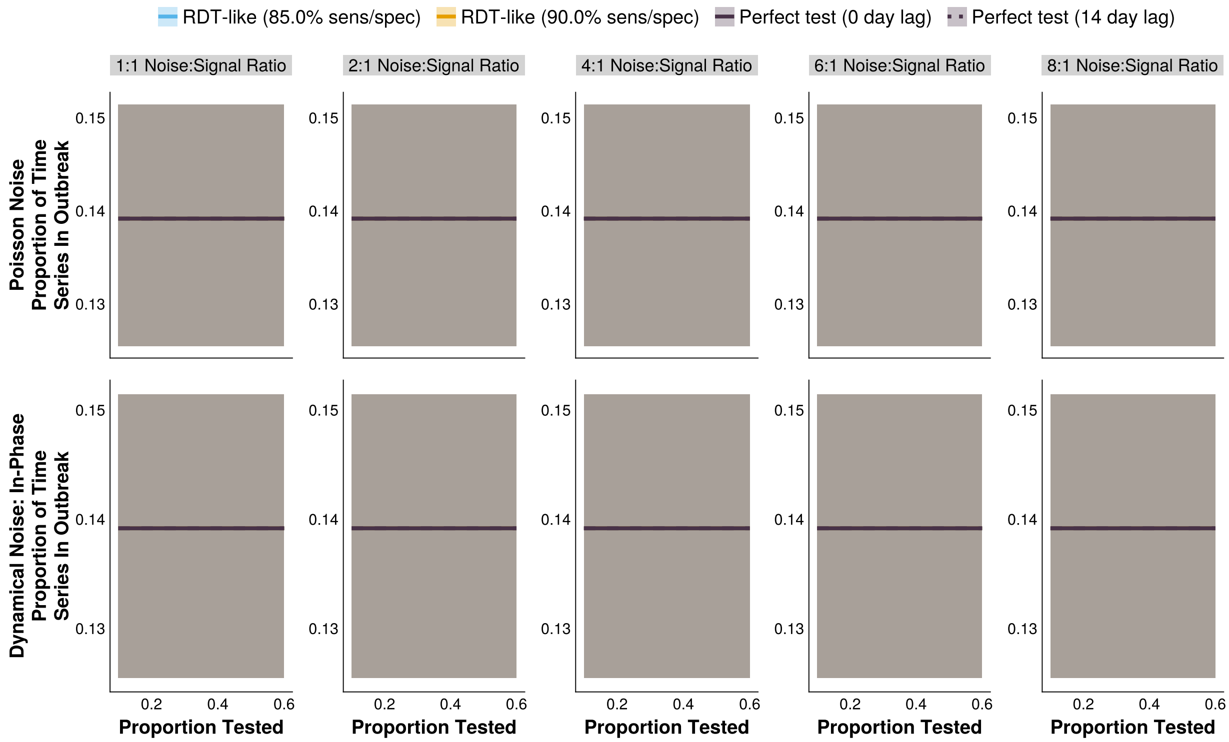

Figure 12: The difference between the proportion of the time series in outbreak for outbreak detection systems under different testing rates and noise structures. The shaded bands illustrate the 80% central interval, and the solid/dashed lines represent the mean estimate. Solid lines represent tests with 0-day turnaround times, and dashed lines represent tests with result delays.

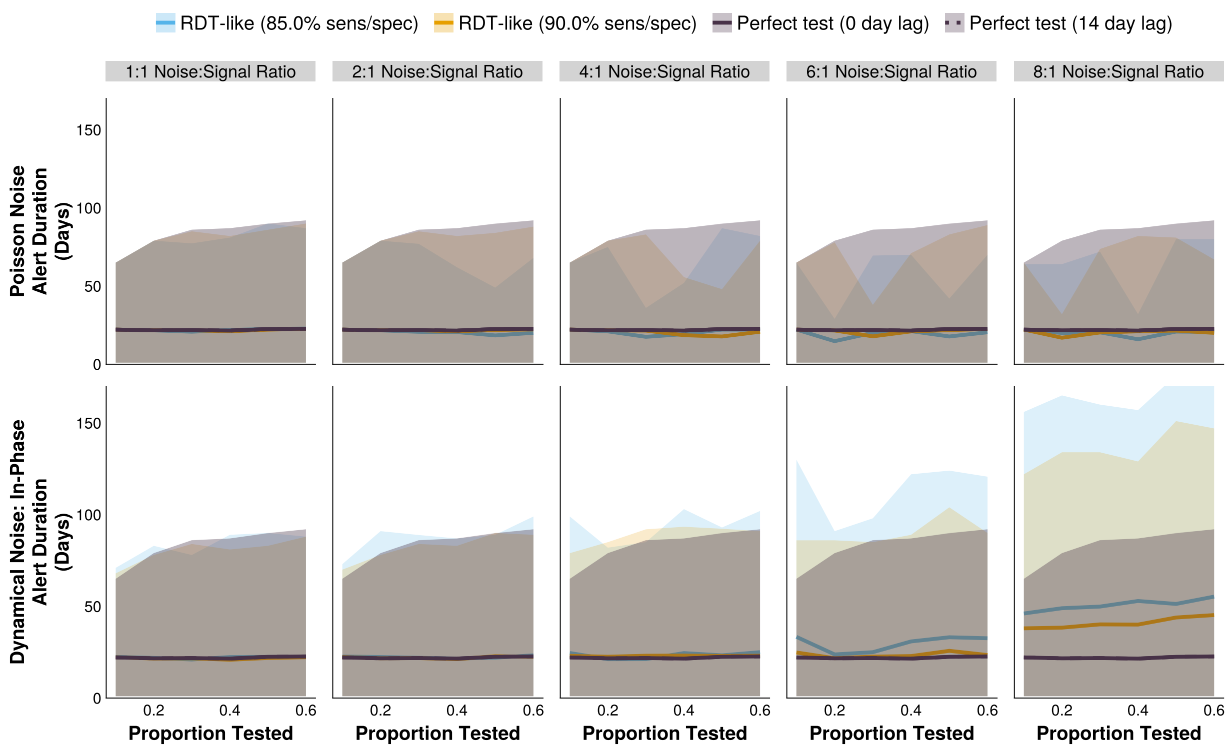

Figure 13: The difference between the alert durations for outbreak detection systems under different testing rates and noise structures. The shaded bands illustrate the 80% central interval, and the solid/dashed lines represent the mean estimate. Solid lines represent tests with 0-day turnaround times, and dashed lines represent tests with result delays.

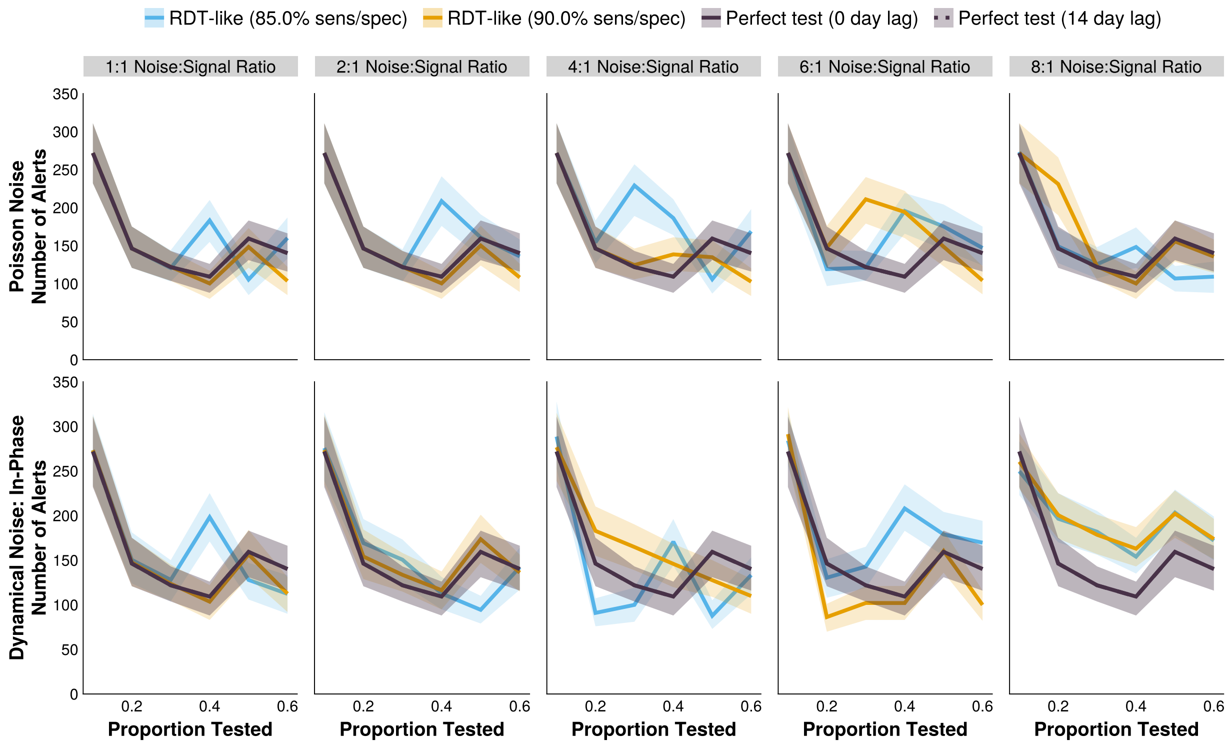

Figure 14: The difference between the number of alerts under different testing rates and noise structures. The shaded bands illustrate the 80% central interval, and the solid/dashed lines represent the mean estimate. Solid lines represent tests with 0-day turnaround times, and dashed lines represent tests with result delays.

Section 2.1.2 | Bibliography

[1] 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

[2] World Health Organization. Surveillance in Emergencies 2024

[3] 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

[4] Nold A. Heterogeneity in Disease-Transmission Modeling. Mathematical Biosciences 1980;52:227–40. https://doi.org/10.1016/0025-5564(80)90069-3

[5] 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

[6] 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

[7] 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

[8] 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

[9] 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

[10] 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

[11] 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

[12] New Results 2023

[13] 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

[14] 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

[15] 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

[16] 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

[17] 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

[18] 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

[19] 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

[20] Hay J, Routledge I, Takahashi S. Serodynamics: A Review of Methods for Epidemiological Inference Using Serological Data 2024. https://doi.org/10.31219/osf.io/kqdsn

[21] 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

[22] 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

[23] 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

[24] 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

[25] 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

[26] 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

[27] 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.

[28] Leadership, School & District Management 2020

[29] Collegian J C | T D. TIMELINE | Penn State's Journey through the Coronavirus Pandemic 2021

[30] 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

[31] 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.

[32] Times T N Y. Tracking the Coronavirus at U.S. Colleges and Universities - The New York Times 2020

[33] 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

[34] United States Census Bureau. U.S. Census Bureau QuickFacts: Centre County, Pennsylvania 2019

[35] 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

[36] 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

[37] 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

[38] Pennsylvania State University. Mask Up or Pack Up 2021

[39] 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

[40] 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

[41] 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

[42] Articles 2011

[43] R Core Team. R: A Language and Environment for Statistical Computing 2021

[44] 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

[45] 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

[46] 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

[47] 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

[48] 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

[49] van Buuren S, Groothuis-Oudshoorn K. {mice}: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software 2011;45:1–67

[50] 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

[51] 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

[52] 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

[53] Brauner J M, Mindermann S, Sharma M, Johnston D, Salvatier J, Gavenčiak T, et al. Inferring the Effectiveness of Government Interventions against COVID-19. Science 2021;371:eabd9338. https://doi.org/10.1126/science.abd9338

[54] 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

[55] 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

[56] PAHO. An Integrated Approach to Communicable Disease Surveillance. 2000

[57] World Health Organization. Measles Outbreak Guide. Geneva, Switzerland: World Health Organization; 2022

[58] Cragg L. Outbreak Response. Applied Communicable Disease Control 2018:134–51

[59] Gastanaduy P A, Redd S B, Clemmons N S, Lee A D, Hickman C J, Rota P A, et al. Measles. Manual for the Surveillance of Vaccine-Preventable Diseases 2019

[60] Commissioner O of the. Coronavirus (COVID-19) Update: FDA Informs Public About Possible Accuracy Concerns with Abbott ID NOW Point-of-Care Test 2020

[61] 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

[62] 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

[63] World Health Organization. Cholera 2023

[64] 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

[65] 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

[66] Shreffler J, Huecker M R. Diagnostic Testing Accuracy: Sensitivity, Specificity, Predictive Values and Likelihood Ratios. StatPearls 2024

[67] 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

[68] World Health Organization. Target Product Profiles n.d.

[69] 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

[70] German R R. Sensitivity and Predictive Value Positive Measurements for Public Health Surveillance Systems. Epidemiology 2000;11:720–7

[71] World Health Organization. Operational Thresholds. Meningitis Outbreak Response in Sub-Saharan Africa: WHO Guideline 2014

[72] 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

[73] 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

[74] Roser M, Ritchie H, Spooner F. Burden of Disease. Our World in Data 2023

[75] 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

[76] Tao Y, Shea K, Ferrari M. Logistical Constraints Lead to an Intermediate Optimum in Outbreak Response Vaccination n.d.:20

[77] 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

[78] 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

[79] World Health Organization. Confirming, Investigating and Managing an Outbreak. Response to Measles Outbreaks in Measles Mortality Reduction Settings: Immunization, Vaccines and Biologicals 2009;3

[80] 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

[81] 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

[82] 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

[83] 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

[84] 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

[85] 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

[86] 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

[87] 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

[88] 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

[89] 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

[90] Middleton C, Larremore D B. Modeling the Transmission Mitigation Impact of Testing for Infectious Diseases 2023

[91] Essential Programme on Immunization (EPI), Immunization, Vaccines and Biologicals (IVB). Measles. Vaccine Preventable Diseases Surveillance Standards 2018:30

[92] Hutchins S S, Amler R, Maes E F, Grabowsky M, Bromberg K, Glasglow V, et al. Evaluation of the Measles Clinical Case Definition. The Journal of Infectious Diseases 2004;189:S153–9. https://doi.org/10.1086/379652

[93] xvii, 357 p. In:. The Selection and Use of Essential in Vitro Diagnostics: Report of the Third Meeting of the WHO Strategic Advisory Group of Experts on In Vitro Diagnostics, 2020 (Including the Third WHO Model List of Essential in Vitro Diagnostics), Geneva: World Health Organization; 2021

[94] 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

[95] 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

[96] Yemshanov D, Haight R G, Liu N, Chen C, MacQuarrie C J K, Ryall K, et al. Acceptance Sampling for Cost-Effective Surveillance of Emerald Ash Borer in Urban Environments. Forestry: An International Journal of Forest Research 2020;93:280–96. https://doi.org/10.1093/forestry/cpz028

[97] Christensen J, Gardner I A. Herd-Level Interpretation of Test Results for Epidemiologic Studies of Animal Diseases. Preventive Veterinary Medicine 2000;45:83–106. https://doi.org/10.1016/S0167-5877(00)00118-5

[98] Murato Y, Hayama Y, Shimizu Y, Sawai K, Yamamoto T. Evaluation of Sampling Methods for Effective Detection of Infected Pig Farms during a Disease Outbreak. PLOS ONE 2020;15:e241177. https://doi.org/10.1371/journal.pone.0241177

[99] Stern L, Lightfoot D. Automated Outbreak Detection: A Quantitative Retrospective Analysis. Epidemiology and Infection 1999;122:103–10

[100] Calba C, Goutard F L, Hoinville L, Hendrikx P, Lindberg A, Saegerman C, et al. Surveillance Systems Evaluation: A Systematic Review of the Existing Approaches. BMC Public Health 2015;15:448. https://doi.org/10.1186/s12889-015-1791-5

[101] Guidelines Working Group. Updated Guidelines for Evaluating Public Health Surveillance Systems. Atlanta, Georgia: 2001

[102] 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

[103] Proverbio D, Kemp F, Magni S, Goņcalves J. Performance of Early Warning Signals for Disease Re-Emergence: A Case Study on COVID-19 Data. PLOS Computational Biology 2022;18:e1009958. https://doi.org/10.1371/journal.pcbi.1009958

[104] Stolerman 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

[105] 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

[106] 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

[107] 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

[108] 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

[109] Guerra F M, Bolotin S, Lim G, Heffernan J, Deeks S L, Li Y, et al. The Basic Reproduction Number (R0) of Measles: A Systematic Review. The Lancet Infectious Diseases 2017;17:e420–8. https://doi.org/10.1016/S1473-3099(17)30307-9

[110] World Bank. Ghana n.d.

[111] World Health Organization. Measles Vaccination Coverage 2024

[112] 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

[113] Keeling M J, Rohani P. Modeling Infectious Diseases in Humans and Animals. Princeton: Princeton University Press; 2008

[114] 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

[115] Morales M, Lanzieri T, Reef S. Rubella. CDC Yellow Book 2024: Health Information for International Travel 2023

[116] 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

[117] 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

[118] FIND. Target Product Profile for Surveillance Tests for Measles and Rubella. Geneva, Switzerland: 2024

[119] Hiebert J, Zubach V, Charlton C L, Fenton J, Tipples 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

[120] 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

[121] 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

Section C | Vita

Callum arnold