

## Synthesis

There is a growing body of literature that focusses on uncertainty in disease transmission; from incorporating viral dynamics into mechanistic models of disease [1,2], to accounting for reporting and testing uncertainty in estimates of the *real-time* reproduction number ( $R_t$ ) [3]. In this dissertation I demonstrate how the classification of continuous infectious disease variables is both essential to inferences about disease dynamics and their underlying systems, and results in compounding uncertainty that limits our predictive and detection abilities for outbreaks. This work has already shown relevance and impact, helping inform the target product profile (TPP) of a potential future rapid diagnostic test (RDT) for measles in outbreak surveillance settings [4].

## Discretization of Risk Groups

In the first half of my dissertation I explore mechanisms by which populations can be classified to understand the transmission of COVID-19 within and between Pennsylvania State University students and community members of its surrounding county (Center County). Clear definitions of risk and transmission groups provide a natural mechanism to explore the heterogeneity in infection that may exist in a population, as it is conceivable that the student and community cohorts differ with respect to many drivers of infection e.g., demography, contact rates, perceptions of infection risk, willingness to take preventative actions. However, despite expectations that the high spatial proximity of these two well-defined groups in immunologically naive populations would overwhelm differences in other drivers of infection and result in similar exposure rates, substantial variation in outcomes was observed (Chapter 2). This, supported by evidence that the Center County community experienced lower per-capita incidence rates than its 5 surrounding counties [5], implies that intervention efforts by the University were able to minimize the risk of onward transmission from the student population. If between-group transmission did occur, it was likely transient in nature. Within the study body, the only factors associated with infection outcome were the recent contact with a known COVID-19 positive individual, and the attendance of gatherings.

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

If we are to design effective interventions, it is essential to first characterize the limits of what can realistically be achieved. For example, it is well known that supplemental immunization activities

aimed to minimize the effects of an outbreak, reactively or prospectively, are partially limited by the vaccination coverage that can be achieved: those who are able to access the care are more likely to also be individuals who can attend routine immunizations [Portnoy REF]. Incorporating access and dose redundancy data can provide more realistic estimates on the effectiveness of future efforts. Similarly, incorporating behavioral information into epidemiological models can shed light on the potential effectiveness of vaccination campaigns. Defining latent groups of individuals by vaccine-seeking or hesitancy behaviors, it would be possible to calculate an expected uptake, augmenting estimates provided through access data. This could allow for a more realistic estimate of intervention effectiveness at reducing disease burden, characterizing a limit for coverage that incorporates previously observed aggregate information (the number of redundant doses provided), and individual intention behavior; there will be situations where individuals will choose not to be vaccinated, despite access expanding to incorporate them. It is possible that vaccine-intentions differ between groups with and without access, so efforts improve access alone may overestimate the protection benefits gained. Modeling these latent sentiments offers a mechanism to account for these discrepancies. Not only does this approach potentially provide more accurate projections of the limits of vaccination coverage, but may also improve disease burden estimates by describing *who* can be immunized; overlap between access and healthcare-seeking behavior with disease vulnerability will highlight gaps that need to be overcome.

This approach can also shed light on effectiveness of non-pharmaceutical interventions in heterogeneous populations, where traditional measures derived from demographic data, such as age-based mixing matrices, may over-simplify the dynamics. Characterizing complex network structures has been a point of focus in the study and prevention of sexually transmitted infections (STIs) [REF]. However, the process is often laborious, disease-specific, and subject to change [REF]. In such systems heterogeneity exists across multiple facets, which may not be adequately captured by standard discretizations. The incorporation of different data sources and categorization approaches may serve as a beneficial middle ground, allowing for a more accurate depiction of the true heterogeneity in disease risk than age classes, while also requiring less hands-on data collection than mapping contact networks; behavioral surveys can be implemented online, and pose potentially less sensitive questions that may otherwise act as a barrier to participation. This is particularly important when designing interventions for novel pathogens, where the most at-risk groups, and therefore most beneficial to target, might not translate to previously observed pathogens and outbreaks and may change over time, and urgent actions are necessary to minimize the impact [REF].

## Discretization of Infection Status

It is imperative to evaluate the effects of discretizing not only exposure classes, but also the outcomes of an infectious disease system. In public health surveillance programs, incidence of disease occurrence must be tabulated before the need for and level of action can be decided upon. Uncertainty in the observational process can occur through two mechanisms: uncertainty in who and how many individuals are counted (observed and tested), and uncertainty in the underlying test results. Incorporating uncertainty in the reporting of cases has been of increasing interest in the parameterization of mathematical models, particularly for the estimation of  $R_t$  [1,3,6,7] and disease burden [REF]. However, the accuracy of the diagnostic used to make this determination affects the incidence, necessarily turning a quantitative input (pathogen load/host response) into a binary value, with associated classification errors. Previous inclusions of diagnostic uncertainty amount to under-reporting as only one pathogen is simulated, removing the opportunity for false-positive test results [1–3,6,7].

To account for diagnostic uncertainty, in Chapters 4 & 5 I simulated both a target pathogen and background noise, using representative parameters of measles and rubella, respectively. Through variations to the rubella vaccination coverage, I was able to evaluate the degradation of outbreak detection performance with tests of decreasingly sensitivity and specificity. When the background noise did not exhibit large peaks and troughs i.e., was drawn from a Poisson distribution or when there was sufficiently high vaccination coverage in the rubella dynamical noise simulations, both imperfect and perfect diagnostic tests could adequately discriminate between outbreak and non-outbreak periods. However, this was not the case at higher levels of dynamical noise ( $\approx 6$  times, or greater, than the average incidence of measles). In these situations the reduced diagnostic accuracy of imperfect tests could not be alleviated through increasing the testing rate. Across the WHO's African Region, there is typically far more circulating measles than rubella; in 2023, the average incidence of measles was 60.3 cases per 1M population, in contrast to 5.2 rubella cases per 1M population [8]. In general, this would point to the widespread viability of imperfect diagnostic tests within infectious disease surveillance systems for the purposes of outbreak detection. However, there is great variability in these incidence rates by country. At the one end, countries like Burkina Faso experienced measles incidence of 69.7 measles cases per 1M population, relative to 0.4 rubella cases per 1M; testing 80% of reported cases, there were 4.2 times as many IgM positive and epidemiologically-linked measles cases as other clinically-compatible infections (i.e. rubella or otherwise) [8]. As a result, imperfect diagnostics would likely be able to accurately discriminate between outbreaks of measles from changes in the incidence of other sources of febrile rash. In contrast, Zimbabwe reported rubella incidence that was approximately 7.3 times higher than measles incidence, presenting a location where the increased uncertainty of imperfect diagnostics would result in poor outbreak detection [8]. Lastly, Eritrea experienced 5.0 measles cases, and 1.3 rubella cases, per 1M population [8]. While this may appear to be a good candidate for the use of imperfect tests, the incidence rate for clinically compatible cases that were not measles or rubella, was 21.3 cases per 1M population [8]. As a result, before imperfect diagnostic tests could be implemented with confidence, a careful evaluation of the dynamics of the non-rubella background noise cases would be required; if it demonstrated large episodic outbreaks, its scale in relation to the expected measles incidence would necessitate the use of high-accuracy diagnostic tests.

In Chapter 5, I build off the work in the previous chapter to explore the effects of the outlined diagnostic uncertainty on our ability to predict risk of future outbreaks. Traditional outbreak detection that utilizes the exceedance of an incidence threshold is necessarily reactive in nature. Under ideal circumstances, it would be possible to measure trends in summary statistics derived from infection data to infer the emergence of outbreaks before they occur. This would allow proactive actions that could avert the most cases, reducing the morbidity and mortality resulting from a pathogen. Prior work has demonstrated the viability of early warning signals (EWS), albeit only accounting for errors stemming from testing rates, not diagnostic uncertainty [9–12]. I demonstrate that, similar to reactive outbreak detection, predictive systems can be designed around the use of RDTs for case identification, so long as the magnitude of dynamical noise is low relative to the incidence of the target pathogen. Not all EWS metrics performed well, but, aligning with the literature, the mean, variance, autocovariance, and index of dispersion were able to discriminate between emergent and non-emergent time series in these situations. Additionally, the evaluation of EWS performance required the alert in emergent simulations to occur before the tipping point  $R_E = 1$ , which indicates the potential for a future outbreak. As this tipping point essentially acts as a necessary precursor to an outbreak (though exceptions can occur due to the stochastic nature of infectious disease transmission), each warning would be provided with sufficient time for action to potentially avert an outbreak.

The primary focus of this section of work (Chapters 4 & 5) has been to motivate new approaches to the design of surveillance systems at large, not any one specific implementation. These two chapters illustrate the complexity in evaluating the performance of reactive and proactive outbreak detection systems. The system's accuracy is dependent on the interaction between multiple axes: the proportion of individuals tested, the diagnostic test characteristic, the shape and magnitude of the background noise, the and the criteria required to trigger an alert. Each of which can be adjusted, and without consideration of the knock-on effects, can easily lead to a suboptimal path being followed. We found that similar accuracy could be achieved for a range of conditions by balancing traits of each of the levers of the surveillance system. By relaxing prior constraints on the outbreak detection system, such as the number of test positives (Chapter 5) or the maximum number of consecutive data points (Chapter 5) required to trigger an outbreak alert, for example, it is possible for imperfect diagnostic tests to obtain outbreak detection performance that mirrors performance under perfect tests. In doing so, this opens the opportunity for the development and use of cheaper, less labor intensive, and logistically simpler diagnostics, such as (rapid) lateral flow diagnostic assays [13–15]. Further, we identified conditions where accuracy was fundamentally limited; even with a perfect test, EWS occasionally trigger under null simulations due to endemic and imported cases. These boundaries can only be illustrated through comprehensive exploration across all axes of the surveillance system.

“These two chapters illustrate the complexity in evaluating the performance (accuracy) of a detection system because the outcome depends on multiple interacting axes (% tested, test characteristics, threshold, shape and magnitude of noise). We found that similar accuracy could be achieved for a range of setting by trading off characteristics of these axes of the surveillance system. Further,

we identified regions of dynamics where accuracy was fundamentally limited — which can only be illustrated by considering the performance across all these axes.

Within Chapter 5, for example, there are numerous opportunities to balance the speed of detection against the frequency of false alerts. Through the approach laid out in this dissertation, an explicit choice about the weight of speed against false positive rates must be made for the calculation of the accuracy of the alert system. Furthermore, once parameters have been set and simulations run,”

Assuming this approach provides public health personnel and policy makers the ability to consider all options and “optimize” the surveillance system to best suit their needs. Trade-offs and the balance of priorities must be evaluated on a context-specific basis. There has long been a tension in designing disease surveillance programs: the needs of an individual may be at odds with those of the wider population. During the early stages of the COVID-19 pandemic, when vaccines were first being developed and their potential effectiveness unknown, there were discussions around who should be prioritized during initial roll-outs: the elderly and those with known co-morbidities, who would most likely receive the largest direct benefit from immunization; or younger individuals with larger numbers of contacts, whose vaccination would most likely result in the largest reduction in incidence, with an indirect benefit to vulnerable individuals [16]. For infectious disease outbreak surveillance, a similar conflict exists. Systems are often built upon routine, passive, surveillance that uses health facility visits for case identification [17,18]. As a result, reducing the accuracy of the diagnostic test used may be associated with a corresponding reduction in cost and technical requirements, allowing for a greater proportion of the population to be tested. This, in turn, may improve outbreak detection, providing an indirect effect to any specific individual, but at the expense of a more accurate diagnosis and care provided to the patient seeking treatment at a

healthcare facility. In a region with high disease burden and large period outbreaks, this may be an acceptable trade-off; in other locations, unacceptable.

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

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

In the future, efforts should be made to formally integrate the design and implementation of outbreak surveillance and early warning systems into partially observed Markov decision process (POMDP) models. Doing so would provide a more complete characterization of the uncertainty that develops and propagates throughout the numerous observational and decision processes. My work has demonstrated that even in scenarios with perfect diagnostic tests, detecting and predicting outbreaks relative to baseline infection dynamics is exceedingly difficult. Generally, a binary threshold is used to ascribe infectious status to individuals, despite prior work demonstrating that infectiousness is not a binary state and the timing of diagnostic test within an infection cycle affects its accuracy [1,2,19]. Incorporating this aspect of the observational process into models would generate realistic biases in data generation, and provide a meaningful extension to the analysis presented.

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

## Conclusion

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

## Bibliography

- [1] 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>
- [2] Middleton C, Larremore D B. Modeling the Transmission Mitigation Impact of Testing for Infectious Diseases. *Science Advances* 2024;10:eadk5108. <https://doi.org/10.1126/sciadv.adk5108>
- [3] Pitzer V E, Chitwood M, Havumaki J, Menzies N A, Perniciaro S, Warren J L, et al. The Impact of Changes in Diagnostic Testing Practices on Estimates of COVID-19 Transmission in the United States. *American Journal of Epidemiology* 2021;190:1908–17. <https://doi.org/10.1093/aje/kwab089>
- [4] FIND. Target Product Profile for Surveillance Tests for Measles and Rubella. Geneva, Switzerland: 2024
- [5] Bharti N, Lambert B, Exten C, Faust C, Ferrari M, Robinson A. Large University with High COVID-19 Incidence Is Not Associated with Excess Cases in Non-Student Population. *Scientific Reports* 2022;12:3313. <https://doi.org/10.1038/s41598-022-07155-x>
- [6] Gostic K M, McGough L, Baskerville E B, Abbott S, Joshi K, Tedijanto C, et al. Practical Considerations for Measuring the Effective Reproductive Number, Rt. *PLOS Computational Biology* 2020;16:e1008409. <https://doi.org/10.1371/journal.pcbi.1008409>
- [7] Abbott S, Hellewell J, Thompson R N, Sherratt K, Gibbs H P, Bosse N I, et al. Estimating the Time-Varying Reproduction Number of SARS-CoV-2 Using National and Subnational Case Counts. *Wellcome Open Research* 2020;5:112. <https://doi.org/10.12688/wellcomeopenres.16006.2>
- [8] 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>
- [9] 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>
- [10] Southall E, Brett T S, Tildesley M J, Dyson L. Early Warning Signals of Infectious Disease Transitions: A Review. *Journal of the Royal Society Interface* 2021;18:20210555. <https://doi.org/10.1098/rsif.2021.0555>

- [11] Brett T, Ajelli M, Liu Q-H, Krauland M G, Grefenstette J J, Panhuis W G van, et al. Detecting Critical Slowing down in High-Dimensional Epidemiological Systems. *PLOS Computational Biology* 2020;16:e1007679. <https://doi.org/10.1371/journal.pcbi.1007679>
- [12] Brett T S, Drake J M, Rohani P. Anticipating the Emergence of Infectious Diseases. *Journal of the Royal Society, Interface* 2017;14:20170115. <https://doi.org/10.1098/rsif.2017.0115>
- [13] 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>
- [14] 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>
- [15] 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>
- [16] Bubar K M, Reinholt K, Kissler S M, Lipsitch M, Cobey S, Grad Y H, et al. Model-Informed COVID-19 Vaccine Prioritization Strategies by Age and Serostatus. *Science* 2021. <https://doi.org/10.1126/science.abe6959>
- [17] Cragg L. Outbreak Response. *Applied Communicable Disease Control* 2018:134–51
- [18] Giesecke J. Routine Surveillance of Infectious Diseases. *Modern Infectious Disease Epidemiology* 2016:137–48
- [19] 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>