

Defense Presentation Plan

General Themes

- World is complex and continuous
- To understand it, and make decisions, we often create categories
 - For example, rather than evaluating student performance using every single test score, we aggregate grades
- Information is lost at the boundaries, but defining categories requires us to be explicit about how we group and interpret the world: the alternative is a lack of objectivity, as categorization will still happen, just without explicit definitions
- These problems exist in infectious diseases: when assessing disease burden and risk we consider what proportions of the population are infected and immunized
 - Takes the continuous phenomena of pathogen load and immune response to place individuals into discrete categories for summarization
 - When deciding how to allocate finite resources for interventions e.g., vaccine doses, we often determine high-priority regions based upon their outbreak status: we have to turn continuous case counts into an “outbreak”, before supplemental immunization activities can be initiated
- In my thesis I discuss where these categorization decisions arise in the analysis of infectious disease data, what we gain from them, and the challenges they present.
 - I do this across 3 phenomena: exposure and infection risk, infection status, and outbreak detection and response

Exposure & Infection Risk

Data4Action

- In Spring 2020, COVID-19 pandemic forced the closure of university campuses, including PSU
- The Fall 2020 semester reintroduced students, creating two distinct possible exposure groups: students and the surrounding community residents
 - Due to the novelty, there were not well-characterized groups directly associated with risk of infection, but preliminary information of transmission and assumptions based on prior respiratory infectious disease transmission suggested that contact rates and spatial proximity would be primary drivers of infection
 - Expected that contact rates would increase resulting from students return, and because both cohorts were immunologically naive, transmission between two cohorts would be persistent and overwhelm any differences in behavior of the two groups
- The Data4Action project was set up to analyze this question: what are the differences between the two well-defined exposure groups?
- Did this by collecting serological samples from community residents during the summer, before the student return, followed by serological samples of the students in the latter half of the Fall semester, and community residents, again, during the winter break
- The expectation was that the community residents would exhibit a large increase in infection in the second sampling wave (higher than observed in neighboring counties/state)
- Instead, we observed that the two communities experienced very different transmission dynamics:
 - Community members saw an increase from c. 3% to 7% (in-line with state and national estimates)
 - Approximately 30% of students sampled were infected
- Clear indication that there was little persistent transmission between two geographically coincident exposure groups.

LCA

- Large differences in seroprevalence led to questions about drivers of infection
 - In the absence of vaccines, possibly a result of behavioral differences e.g., adherence to social distancing and mask-wearing?
- Very possible that similar differences existed within each cohort
- Unlike in the seroprevalence study, no clear way to characterize the cohorts based upon social or demographic data
- Instead, behavioral survey data was collected for the majority of participants in the study, asking questions about intention to follow public health measures and guidelines
- How can we link behavioral data to infection outcomes?
 - Important to recognize that behavioral responses are likely related to one-another. If you routinely wear a mask, for example, its possible that you are more likely to follow social-distancing precautions, as well as other guidelines.
- As none of these measures exist in a silo, we wanted a way to use the behavioral data to build a “risk profile” for the students based upon their intent to follow public health measures
- Used a technique called lca to create a model of 3 different risk profiles that existed within the student population: those that intended to always adhere to PHMs; never; and those who mostly masked and stayed home when ill, but didn’t distance or avoid crowds
- Using these emergent categories, we could now test whether behavior was indeed associated with infection: it was!
 - Those who always adhered to PHM had the lowest seroprevalence (25.4%), never the highest (37.7%), and the middle the middle (32.2%)
 - These patterns held when accounting for factors in the prior analysis that were shown to be associated with COVID-19 infection (e.g., contact with known positive)
- Now we have the groups, what can we do with them?
- Natural next question is what would happen if we were able to implement a public health messaging campaign that was perfectly effective at increasing adherence to PHMs
- LCA showed that only 62% of the students could have risk reduced, so unlikely that all disease could be eliminated, but because disease transmission isn’t linear, it’s possible that we can decrease transmission by more than 62%
- Parameterized a disease transmission model to explore this
- Depending on the level of between-group mixing, we saw between 76-93% reduction in transmission
- No inherent way to link continuous risk profile to infection outcome given data that is traditionally available or readily collectable, and by defining categories we were able to provide a mechanism to explore these questions

Discretization of Infection & Outbreak Status

- Switching gears, we can think about discretization of infection and outbreak status
- When we think about resource allocation, we often think about outbreaks
 - It makes sense that regions experiencing large outbreaks need additional resources to mitigate transmission and minimize burden as much as possible
- However, how do we define and alert for outbreaks?
 - In practice, for a lot of disease systems, particularly where burden is high and capacity is limited, the need to move quickly is imperative - there isn’t time to build long, computationally expensive, models to estimate parameters like the *real-time* reproduction number
 - Instead, thresholds can be used to define outbreaks: if the number of cases exceeds a threshold, then an outbreak is declared

- We can therefore see we are left with a classification process, outbreak and non-outbreak, with associated opportunities for understanding and challenges with discretization errors
 - If the threshold is comparatively low, there will be instances where we exceed momentarily the threshold due to transient imported cases, but doesn't result in an exponential growth in cases
 - Conversely, if the alert threshold is too high, we might miss some smaller outbreaks
- But importantly, all this is predicated on accurate case counts
- We know that individual diagnostic tests are not perfect themselves: they have to translate pathogen load, for example, into an infection status
 - This is often talked about as the sensitivity and specificity of the test
 - The % of positive individuals who will return a positive result, and the % of negative individuals who will return a negative result
- We can see both of these problems can be evaluated in similar terms
- The important question, however, is how do inaccuracies in the diagnostic test result affect the ability to detect outbreaks?
 - If diagnostic test accuracy isn't ultimately that important, it opens up the opportunity to develop and expand the use of less accurate tests that don't have the same constraints as traditional diagnostics used for surveillance purposes: fewer lab resources required as samples don't need to be sent to expensive (central) processing facilities that have associated delays in case reporting and have expensive storage requirements

Outbreak Detection

- To examine, I created simulations of outbreaks resulting from co-circulating pathogens, using measles and rubella to provide context to the model
 - Simulating the backdrop to our target disease is essential to examine the scenarios which might break the detection ability of a less accurate diagnostic
 - There is a concept called the PPV, which tells us what we're actually interested in with a test - the probability of a positive result given that the individual is actually infected
 - Unlike the sensitivity, it doesn't condition on a positive person being tested
 - We can imagine that if 100 people come to a clinic, 99 with measles and 1 with rubella, a lack of specificity in the diagnostic that would produce a false positive is relatively inconsequential for our estimates of disease prevalence
 - If the conditions are reversed and only one of those individuals has measles and 99 of them have rubella, suddenly an inaccurate test would produce a meaningful amount of false positive results that is much larger than the number of true positive test results.
 - Used an SEIR model to simulate 100 years of measles and rubella infections, and generated a time series of test positive individuals given a perfect or imperfect diagnostic test and a specified testing rate
 - Defined a true outbreak as a period when the number of cases exceeds 5 measles cases per day, for at least 30 consecutive days, and more than 500 cases must be observed in a continuous 'hump' of measles cases
 - To detect an outbreak, categorize test positive time series into block where the number of observed cases exceeds the threshold, for example, more than 2 test positives.
- Now we have a blocks of outbreak and alert periods, we can compare then to evaluate how good our outbreak detection is
 - Conceptually, if there is a lot of overlap, then our system is good
 - We can evaluate the system in terms of its sensitivity and its PPV: how many of the true outbreaks so it detect, and given an alert, what the chance that it actually relates to an outbreak

- Both are competing ideals so, much like a diagnostic test, we can calculate an alert accuracy metric as the average of the two to try and reduce the chance of alerts being overly reactive or overly specific
- Finally, calculate this accuracy over multiple alert thresholds for each test find the best alert conditions given a test, testing rate, and amount of background noise, because, as discussed, these factors all interact with each other
- Although described noise as simulating rubella outbreaks, I also simulated noise drawn from a Poisson distribution to demonstrate the effects of noise *structure* as well as magnitude, on the conclusions
- What we see is that as we increase the testing rate, the ‘optimal’ threshold increases, regardless of the test
- Now, we can compare accuracies:
 - Start with low Poisson noise and walk through plateauing accuracy with increasing testing
 - Explain $\Lambda(1)$ as equal noise
 - Walk through increasing Poisson noise: magnitude doesn’t make much difference
 - Add all dynamical noise:
 - Pretty quickly the performance of imperfect tests drops off as false positives overwhelm the system, which can’t be accounted for with changes to testing rates

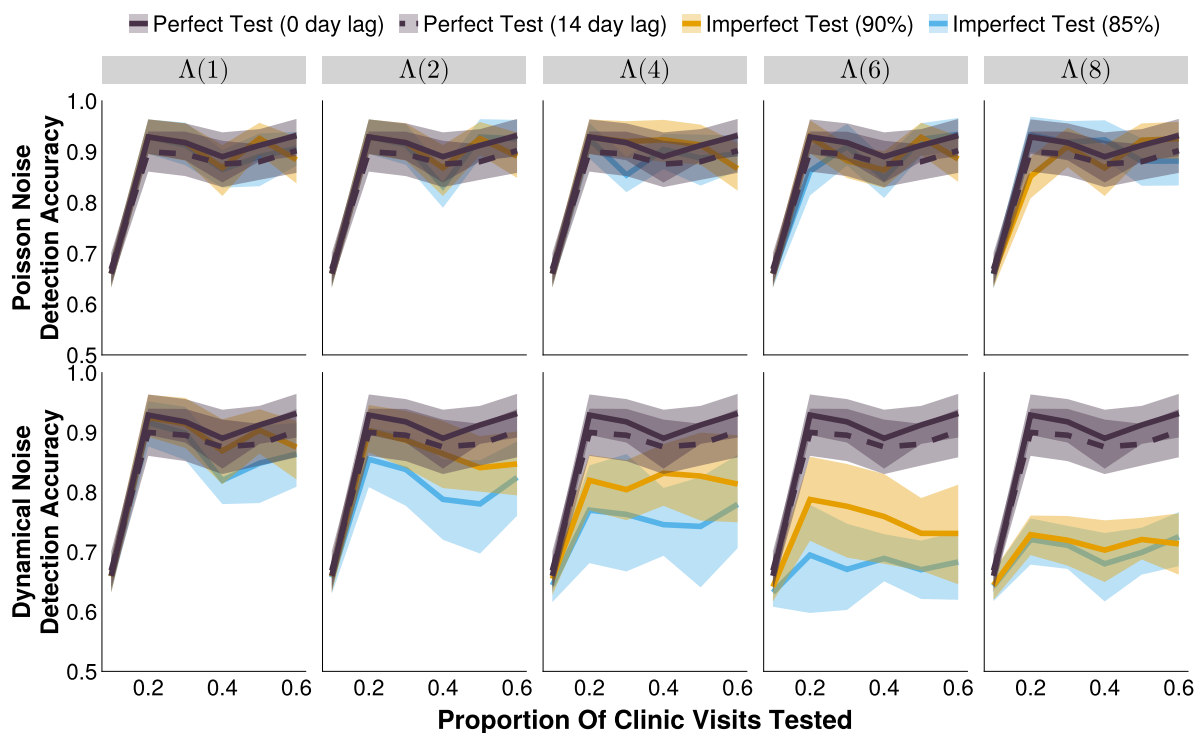


Figure 1:

EWS

- Until now, I’ve just talked about reactive outbreak detection
 - It would be great to be able to take a *proactive* approach and instigate action before infections can get to an uncontrollable level
- There is a field of work in complex systems called critical slowing down
- The general idea is that we can calculate summary statistics, called early warning signals, that provide some advance warning of a *tipping point*.
- Previous work has demonstrated that these EWS do work in infectious disease systems to predict the onset of outbreaks

- I'm interested in how they work in the face of diagnostic uncertainty: do the patterns still hold when we are less confident in our test results?
- To tackle this problem, we must once again define categories to be used in the evaluation of such an alert system
- Unlike converting cases into an outbreak status, instead we are trying to predict whether the state of the observational process is approaching a *tipping point* or not
- For infectious diseases this *tipping point* is the effective reproduction number R_E - To tackle this problem, we must once again define categories to be used in the evaluation of such an alert system
 - R_E is the average number of secondary infections caused by a single infectious individual in a population
 - When $R_E \geq 1$, seeded infections are self sustaining and can spread uncontrollably (effectively a necessary precursor to outbreaks)
- Using a similar simulation structure as the *reactive* outbreak detection project, we simulate test positive time series for various test and noise structures & magnitude combinations
- Show emergent time series
 - Step through Reff, then test positives, then variance EWS
 - Show Tau - prior work has mostly been identifying that EWS are correlated with emergence
 - To say this definitively we need to also simulate null time series where there isn't emergence
 - Reveal null comparison to highlight still a correlation with imperfect test
- Much like reactive outbreak detection, we need criteria to trigger action
 - Define this based on properties of the EWS computed
- Calculate an alert as exceeding a quantile of the long-running distribution, producing a 'flag', and may require multiple consecutive 'flags' to produce an 'alert'
- Because we simulate both emergent and null time series, we can calculate the alert accuracy
 - The alert sensitivity is the % of emergent time series that produce an alert
 - The alert specificity is the % of null time series that *do not* produce an alert
 - Averaging these two produce an accuracy
- Similar to the *reactive* detection project, we calculate the optimal alert conditions given the EWS being computed, the test, noise structure and amount of background noise, because, as discussed, these factors all interact with each others.
- Walk through autocovariance accuracy figure:
 - With low Poisson noise, decreasing test accuracy has little effect on the alert accuracy
 - Increasing Poisson noise similarly has little effect
 - However, increasing amounts of dynamical noise do: imperfect tests cannot reliably detect emergence (show both noise magnitudes)
 - Then show other 3 metrics: some are more robust to dynamical noise (mean) with imperfect diagnostics

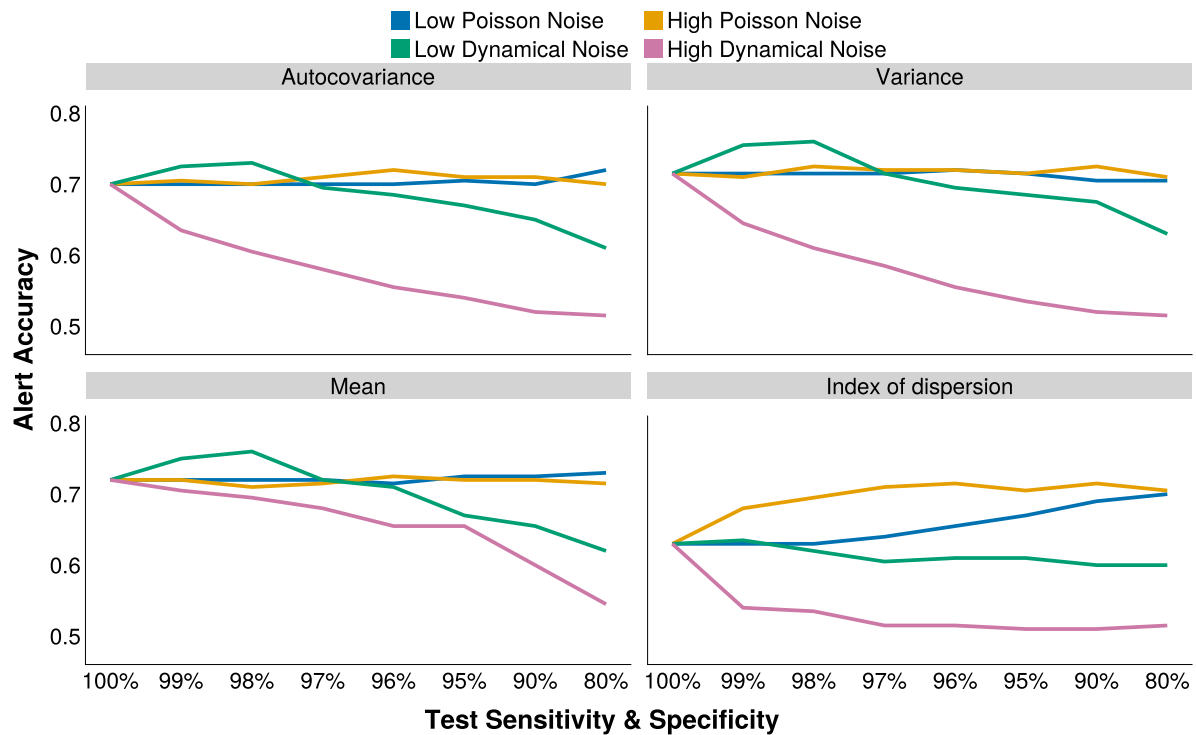


Figure 2:

- Also important to examine the speed of warnings in the system
 - Do this by visualizing when alerts occur relative to the tipping point
- Walk through autocovariance survival plot
 - Histogram shows the timing of the tipping points
 - For the emergent time series with a perfect test, we start with 100% of simulations that haven't alerted
 - Each step down denotes a time series that has alerted
 - Levels off when no more alerts happen before the tipping point (false negatives)
 - Can produce same curve for the null time series
 - The difference at the end signifies the accuracy: we want a larger gap between them as indicates the emergent typically alert before the tipping point, but the nulls do not
 - For autocovariance under low Poisson noise and a perfect test, the steeper decline indicates faster alerts than the null time series
 - Layering in imperfect tests we can see a slightly more sensitive system as both curves decline faster and end lower than with a perfect test
- Extend for all noise structure and magnitude combinations
 - The histograms and solid curves are identical because they use the same simulations and make no mistakes in identifying measles from noise

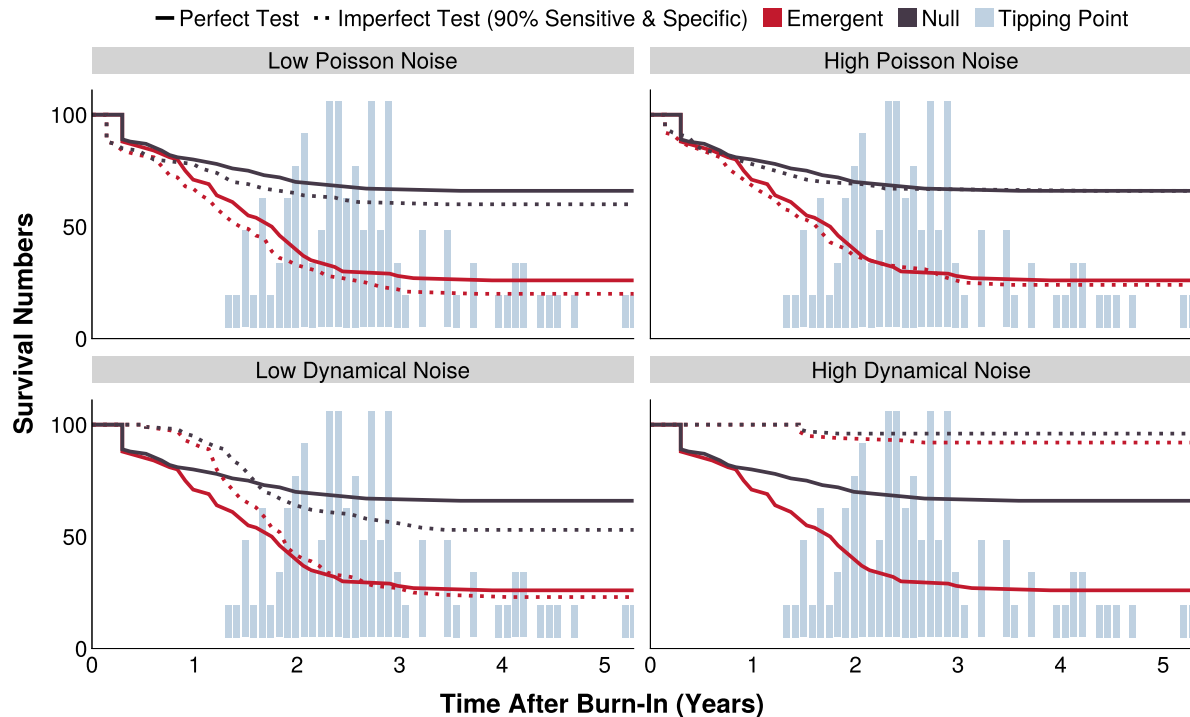


Figure 3:

- Much like the reactive system, a proactive alert system can be predictive with imperfect tests, so long as the noise is unstructured or does not vastly outnumber the magnitude of the target disease

Wrap-Up

- First 2 projects about categorizing explanatory variables
- Last 2 projects about categorizing explanatory & outcome variables
- Discretization is necessary evil
 - To understand systems we need to have things to compare
 - For action, we have to have thresholds and cut-off points to motivate activities
- However, comes with a loss of information
 - Fundamental limits on predictability as
 - categories may not map directly of explanatory variables or outcomes
- My work has in part been to identify scope of predictability of a system, even if it's not very predictive
 - We need to know the limits of what can be done to appropriately plan interventions