

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

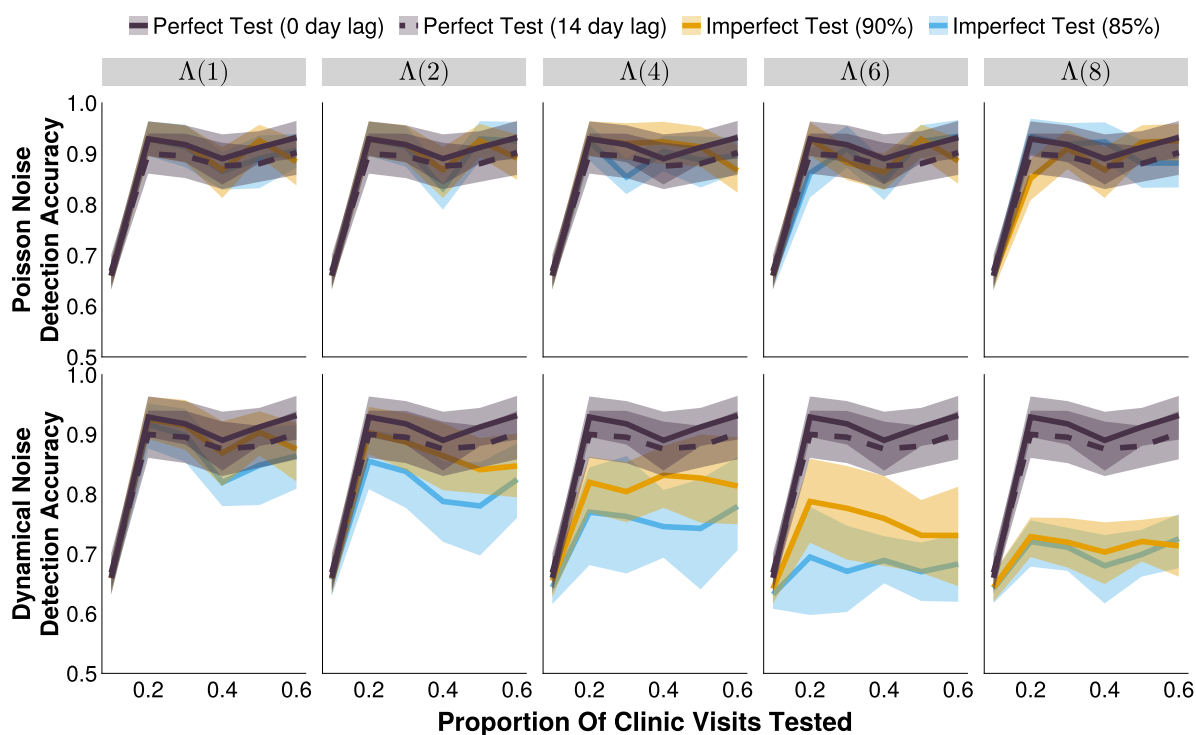


Figure 1:

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

Main Sections

- About my research
- D4A
- LCA
- Outbreak Detection
- CSD

About My Research

Overview

- Focus on heterogeneity and uncertainty from discretization
 - Define two terms
- Heterogeneity & discretization can exist on multiple scales
 - Going to think about 2 levels for this talk:
 - Individual
 - Population

Individual

- IgG OD distribution
 - Imagine we are developing a test for a novel pathogen, and we have some sera from pre-pandemic times, and from individuals we are sure were recently infected
 - We want a test to be able to discriminate between two infection statuses, but overlap between the distributions means there will be false positives/negative.
 - Will need to make a judgement call about which is better/more important (foreshadow outbreak detection projects)

Population

- Much like individual, heterogeneity and discretization exist on a population level
- Infection time series often need to be broken up into outbreak/non-outbreak times for resource allocation

How does it present in epi models?

- Classical SIR model introduces categories - without them you couldn't model infections
- Often taken further to introduce subcategories for particular groups and model interactions e.g., age classes and traditional POLYMOD mixing matrix

Does it matter for epi models?

- Yes:
 - During COVID-19 we saw phenomenon called overdispersion - some individuals were drivers of outbreaks (colloquially called superspreaders)
 - Some estimates indicate c. 10% infections caused 80% onward transmission (see references in <https://www.pnas.org/doi/10.1073/pnas.2016623118>; <https://www.nature.com/articles/s41591-020-1092-0>)
- If you don't account for heterogeneity, hard to create models with accurate forecasts or realistic projections
- When designing targeted interventions, important to know where to allocate resources to get the largest impact and do the most good

What do we need to know?

- Number of groups
- Size of groups
- How groups differ with respect to an outcome
 - Risk of onwards transmission

Presentation Outline

- This talk is going to look at each of these two scales, first the individual, and then transition to population

D4A & LCA

Background

- COVID-19 novel pathogen & forced campus closure in March 2020
- 35k returning students for Fall semester
 - Accounts for 20% increase in county pop
- Two well defined (discretized) groups, with different behaviors and (assumed) risk of infection and onwards transmission (heterogeneity)
- Conventional wisdom would indicate return would increase students' contact rates and lead to large increases in students exposure rates
 - Subsequent high risk of increasing community transmission

What Was Done

- D4A study set up to examine effect
 - Paired behavioral and serological surveys for both community members and students
- Sampled community during summer (without students), then students in Fall semester, and community again during Winter break

Initial Results

- 684 students and 1313 community residents participated in the longitudinal cohort study
- 30.4% seroprevalence among students, 3.2% & 7.3% among community residents
- Despite clear expectations of significant between-group transmission, not observed
- Heterogeneity in behavior?

What about the students?

- Expect that there were differences within the student population as well
- Unlike prior analysis, no clear way to separate our students into groups that can be modeled
- Believe that we can use behavioral data to 'discover' latent groups within student body
 - Behavior should be related to infection outcomes

Discovering latent groups

- LCA is a method to define latent groups according to the behavioral survey responses of individuals, and assign a probability of group membership to each individual based on their responses
- Based on the results, saw that an LCA model with 3 classes was best fit to the data

Are the classes associated with outcome?

- Yes, with classes that did the least exhibiting the highest seroprevalence, and with OR > 1 when compared to the class that were the most adherent

Incorporating classes into a model

- Now we have classes that map behavior to infection rates as expected, we estimated transmission parameters to allow exploration of interventions
- Explain fitting process (only did structure **A**)

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

1

Fitting with Assortativity

- Plot of ABC distance by assortativity level

What can we achieve through interventions?

- Modeled interventions by moving individuals from the low adherence groups to the high adherence groups
- Fully effective intervention can reduce transmission by 76-93%, depending on degree of assortativity

LCA takeaways

- Using an interdisciplinary approach we were able to model heterogeneity in transmission risk in a manner that matched the empirical data
 - Not possible to do with traditional demographic-based approach for contact matrices
- Doing this showed that only 62% could have risk reduced, and expected benefit of behavioral-based interventions was limited in capacity to prevent an outbreak (although targeted interventions could have an outsized effect relative to scope)

Integrating Individual Population Level Heterogeneity

- Shifting focus to population level & incorporation of individual level uncertainty
- Going to talk about outbreak detection, using measles as test case

Outbreak Detection

Measles epidemiology

- As becomes rarer, outbreaks make up a more significant portion of the annual incidence (Graham et al)
- Variability in year-on-year incidence increases
- Outbreaks are particularly important to detect as measles elimination approaches
 - Equally, if countries have lapses in their vaccination coverage and go the wrong way

Outbreak detection SOP

- ELISA & CCD
- Alerts based on thresholds

RDTs offer new opportunity for outbreak surveillance

- Tension between patient-centric and public-focussed approaches to outbreak detection
 - ELISAs are excellent for the patient, but costly and slow
 - RDTs are less accurate, so may give the patient the wrong information

- For measles, this is less of a concern as the treatment for febrile rash doesn't depend on an accurate diagnosis of measles vs rubella
- Despite concerns that lower accuracy of the test, has been used effectively in malaria and COVID-19 surveillance systems
- Want to evaluate if RDTs can be incorporated into a measles surveillance system effectively, and the interaction between individual-level uncertainty and at the outbreak detection level

Model structure - compartmental model for measles

Model structure - dynamical noise

Model structure - Poisson noise

How testing works

How outbreaks are defined

How alerts are triggered & characterized

Schematic of accuracy calculation

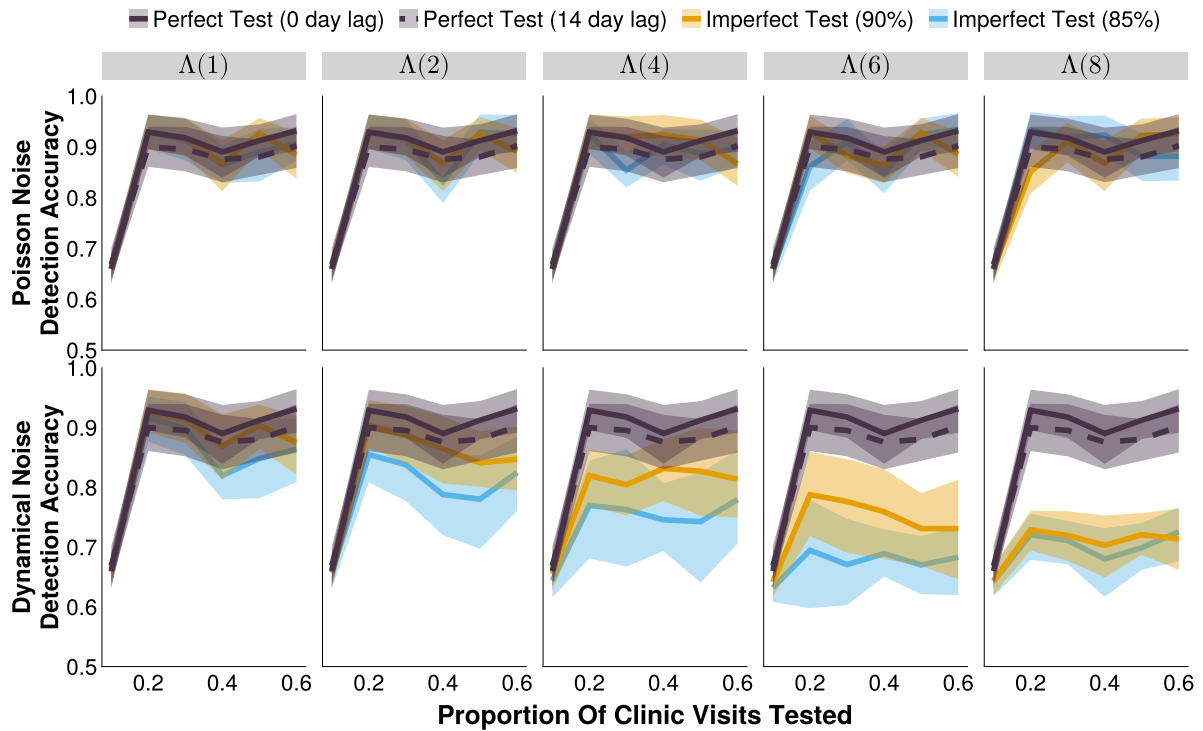
Calculation of optimal threshold

Example of one set of optimal thresholds

- Show table for Poisson noise
- Explain how thresholds increase with increased testing rate
- Slight differences between tests
- Can take this and calculate for each type and magnitude of noise, but instead going to transform this into some plots

Accuracy plot

- For Poisson noise, RDTs can perform equivalently
- For dynamical, break down around 4:1 noise:measles ratio
 - Can't test your way out of the problem

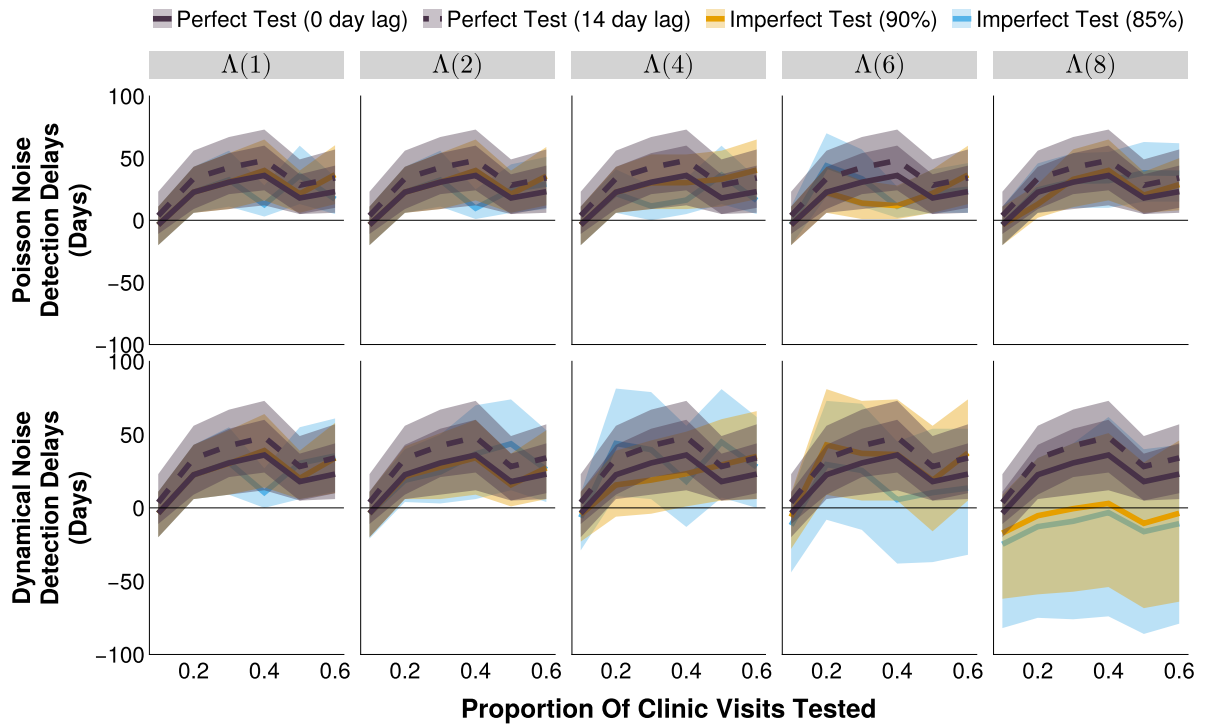


What's causing this difference?

- Dynamical noise has rubella outbreaks that translate into peaks of false positive test results that trigger the alert
- Increasing the threshold would cause genuine measles outbreaks to be missed, lowering the sensitivity of the alert system, more than is made up for by the corresponding specificity of the system

Delays plot

- Corresponding dynamical noise with RDTs show a more sensitive system than with ELISAs (or under Poisson noise)



Proportion in Alert

- Can see same picture in the proportion of the time series in alert



Concluding Thoughts

CSD

Extending to Early Warning Systems

- Rather than just reacting, lots of interest in trying to detect a state of risk for future outbreak
- Prior work on matter used CSD metrics

CSD overview

Potential issues with measures

Want to incorporate learning from OD project

Background with Tycho project

- Show how they are calculated

Simulation

- create simulation experiment where we can discretize time series
 - Reff grows by lowering vaccination rate
- Need null to evaluate
- In CSD don't need outbreak to evaluate, unlike outbreak detection
 - Not conditional on outbreak (reactive vs proactive)
- Plot trajectory of Reff between case and control simulation
- In treatment simulation, trying to use metrics to identify that it will cross threshold, even if outbreak doesn't happen
 - Want systems that don't flag inappropriately in null case
- Explain hyperparameters optimized over
 - distribution builds threshold for flag
 - how many flags required for an alert
 - Show histogram and consecutive flags schematic
- Show accuracy heatmap
 - Just the accuracies
 - Declarative statement about low and high values of accuracy for each type of noise
 - 2 x 2 for noise structure and magnitude
 - The system affects the accuracy
 - Each cell has its own optimal value of percentile and flags
 - Learnt from prior chapter that the system has to be considered
- Within a heatmap
 - The measure that gives the best accuracy depends on the test used
- Then add in speed and specificity
 - Maybe only add in to best 2 metric for each test
 - Within a system, think about what you prefer (speed/spec) so choose a different metric
- Mostly likely just end with table, not survival plot

Summary

- 2 & 3 about explanatory variables
- 4 & 5 about explanatory & outcome variables
- what was learnt within each section
- What was learnt across all chapters

- Discretization is necessary evil
 - Fundamental limits on predictability as
 - categories may not map directly of explanatory variables or outcomes
- Even just identifying scope of predictability of a system, even if it's not very predictive