

# Introduction Plan

## Notes

### General Framing

- Continuum, not heterogeneity
- Discretization for decision making purposes
  - Not main focus, though
- Without discretization, we can't draw any conclusions
  - Just mapping all variation all at once
- Start broad and general with examples
  - Age classes -> individual OD -> diagnostic test
  - Have to make decisions about how many classes (1 -> ABM)
    - Represent heterogeneity without losing information or being overly complex
    - LCA paper finds this for behavior
    - OD paper finds this for classification of outbreaks
- By discretizing, we can learn about the underlying processes:
  - LCA:
    - There is no inherent continuous measure of risk that we can measure, but LCA provides mechanism to group behaviors that can approximate risk, together
    - How risk behaviors cluster
    - How different the clusters are
    - What is the resilience of the system / limits of action
  - OD & CSD:
    - What are the limits of possible discrimination
    - How does individual level variability affect discriminatory ability

### Beginning

- If we discretize from continuous:
  - How many?
  - Where's the break point?
- Then we can ask:
  - What's the difference between the groups?
    - May show that discretization doesn't provide us with groups that are meaningful in differences and/or action
- Why do we want to categorize?
  - Learn about:
    - Clustering
    - Scope of differences
    - Discover unknown drivers that may not be present in the data (unsupervised clustering)
      - Data4Action project
      - First look for differences in behavior -> covariates
  - To do stuff:
    - Introduce sens/spec/accuracy/PPV
      - Diagnostic tests (IgG & outbreaks)
      - Talk in general terms

## Chapters

- LCA:
- OD:
  - Know discrete 2/3 groups (outbreak not present/small/large etc)
  - We have noise in the system, so want to understand PPV of an alert
    - Will depend upon the context (e.g. prevalence of outbreaks)
  - Reactive setting
- CSD:
  - Similar to OD, but anticipating risk of outbreak risk

## Principal Points

- Epidemiological triad:
  - Host
  - Agent
  - Environment
- When we think about infectious diseases and their control, we must think about each section of the triad and the interaction between them
- Many things exist on a continuum
  - Heterogeneity is about differences with respect that continuum
- Heterogeneity important for disease transmission, and how we think about disease and outbreaks, across range of scale
  - Each branch of the triad exhibits heterogeneity and the combination will result in an infection risk for an individual and the population
    - Multiple combinations of each can result in similar infection risks
  - Can broadly group the scales in which heterogeneity acts into 2 levels:
    - Individual level
      - Susceptibility to infection
        - Pharmaceutical interventions aim to address this
      - Contact patterns & exposure
      - Behavior
    - Population level
      - What tests are used - sensitivity and specificity are defined on population levels relative to deviations from true negative mean values (IgG/IgM/OD etc.)
      - Public health actions and interventions, e.g., outbreak responses, come at various levels of funding etc
      - Statistical heterogeneity indicates that different studies of the same effect will have different outcomes and need to be summarized in meta-analyses
      - Groups of individuals may respond more similarly to each other than to those in another group e.g., spatial proximity, or contact patterns by age
        - Overdispersion of  $R_0$ 
          - $R_0$  defined at a population level to begin with, and can vary based on any 3 of the epidemiological triad components
- Heterogeneity is impossible to comprehensively model, and typically artificial groups are defined to simplify the modeling process
  - Much like heterogeneity, discretization occurs at both the individual and population scale:
    - Individual level
      - Susceptibility/serostatus of an individual based on a binary diagnostic test
      - Compartmental models are an example as an individual's immunological status exists on a continuum but is split into S I and R compartments, for example.
    - Population level
      - How to define an outbreak
        - Outbreak definitions are incredibly important as governments and public health agencies may only have resources to deploy to a handful of locations per budgetary cycle, so incorrectly describing an outbreak can disastrous outcomes through missed infections (and it may be too late to meaningfully respond to) or opportunity costs from an inappropriately sensitive alert and response

- Models are often made in continuous time, but observational data is often discrete - counts aggregated by each day at the minimum, and commonly weekly, biweekly, or monthly.
- This process is called categorization, or discretization, and the process may result in groups that don't truly align with the underlying heterogeneity.
  - Important to carefully evaluate assumptions being made and think about uncertainty should be accounted for in the analysis
    - Observation
- Early models by Anderson & May assumed mean-field (homogeneous) mixing and no heterogeneity within each compartment
- As computational resources have expanded and become more common-place, it has become possible to increase model granularity: Agent-Based Models are routinely implemented in fields where heterogeneity is known to be incredibly important (STI models are often ABMs or network models)
  - Not necessarily improved predictive accuracy for respiratory infections as need far more data to parameterize the model, and many assumptions about the state of interactions still required

## Intro Summary

- Chapter 2: D4A
  - Defining heterogeneity in disease exposure rates between well-defined geographically coincident populations
  - Despite clear expectations that spatial proximity and moderate-high  $R_0$  disease in immunologically naive population would result in similar exposure rates, outcome heterogeneity is observed
    - Indicates underlying differences (heterogeneity) in behavior, given the absence of pharmaceutical interventions
- Chapter 3: LCA
  - Defining heterogeneity within a demographically homogeneous population to explain how behavior can drive disease transmission, and place limits on the effectiveness of expected interventions
  - Multidisciplinary research provides opportunities to model latent heterogeneous groups in a manner that should be more proportional to transmission dynamics than typical methods reliant on demographic information
- Chapter 4: Outbreak detection
  - Examine the interplay between the heterogeneity that exists in a population (infections) and how the uncertainty in the methods we use to define both infectious individuals and outbreaks compound to affect outbreak detection
  - Partially observed Markov process means we have to make decisions with incomplete information - can never truly optimize our outcomes
- Chapter 5: Early Warning Systems
  - Examine how compounding uncertainty introduced in Chapter 4 impacts the performance of early warning systems with imperfect observational processes
  - Indicates the limits of predictability
- Overarching summary:
  - There is a growing body of literature focused on uncertainty in disease transmission measurements and incorporation into models
  - Thesis demonstrates how heterogeneity can drive infection dynamics & introduce uncertainty at each level the outbreak and observation process, and the importance of acknowledging the interaction between the levels.