# Agent-Based and Aggregate Modeling: Tradeoffs & Limitations

Nathaniel Osgood

November 13, 2012

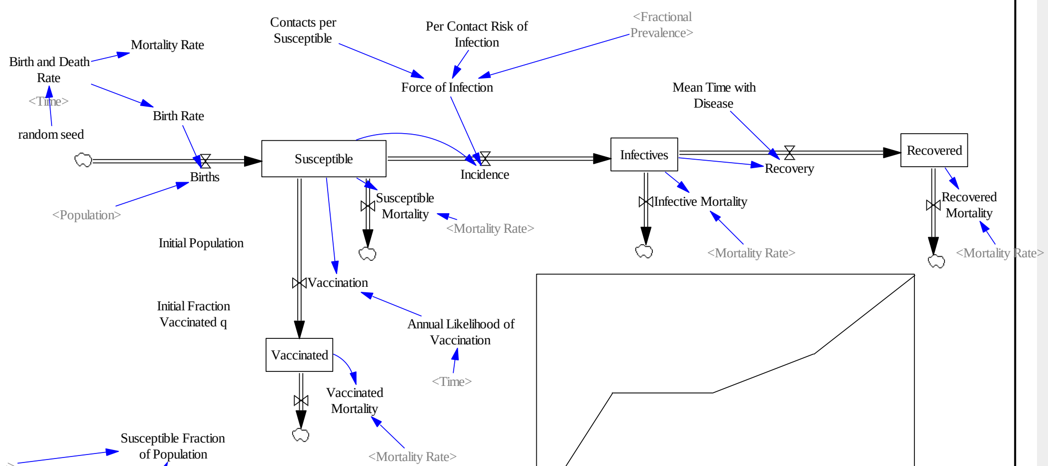
# Distinctions

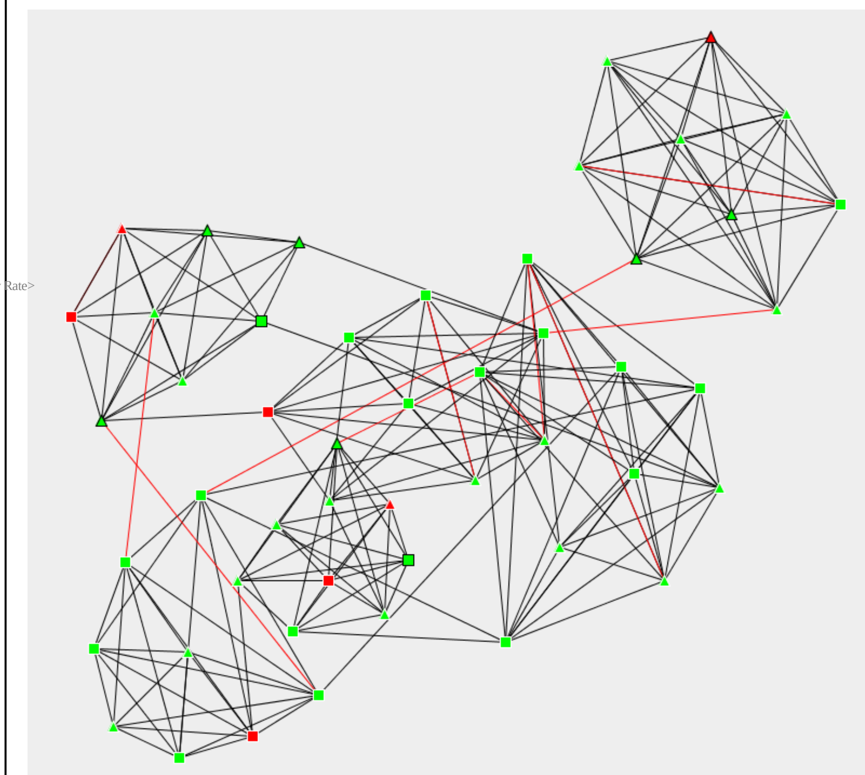
- Inherent, e.g.
  - Qualitative vs. Quantitative
  - Static vs. Dynamic
  - Stochastic vs. Deterministic
  - Capacity to understand single scenario vs. range of scenarios
  - Magnitude of computational resources required
    - Interactive or not
  - Under vs. over-determined calibration
  - Ability to calibrate to/make behaviour depend on individual history
- Important software skills mediation
  - Required level of software development sophistication

# Dynamic Models for Health

- Classic: Aggregate Models
  - Differential equations
  - Population classified into 2 or more state variables according to attributes
  - $|\text{State Variables}|, |\text{Parameters}| \ll |\text{Population}|$
- Recent: Individual-Based Models
  - Governing equations approach varies
  - Each individual evolves
  - $|\text{State Variables}|, |\text{Parameters}| \propto |\text{Population}|$

# Contrasting Model Granularity



# Granularity Selection: Problem Specific

- Selection of granularity is a function of question that are asking – not of the "*true nature of the system*"
  - *Modeling for learning/qualitative insight (requires "caricature model") vs.*
  - *Modeling to quantitatively predict (requires detailed characterization)*
- Quanta of most obvious system components may not align with needs for insight
  - May gain benefits from higher-level representation
    - Many high-level qualitative behaviors of complex systems can be explained with very simple models
    - Often gain greater insight from simpler model: C.f. Gas laws vs. lattice gas model
  - May wish to seek lower level model
    - Small infection spread model : Characterization at level of immune response rather than monolithic person

# Myth of Individual-Based Models as “Modeling from the Bottom Up”

- A single person is a natural locus of description
  - Presents for care
  - Lives
  - Dies
  - Coupled internal systems
- But the world has no natural “bottom”
  - It is frequently desirable to include within a person a great deal of “within the skin” detail
- The issues of model depth & breath are just as pressing in individual-based models as in aggregate modeling

# Contrasting Benefits

## Aggregate Models

- Frequently, easier
  - Construction
  - Calibration
  - Parameterization
  - Formal analysis (Control theoretic & Eigenspace techniques)
  - Understanding
- Performance
  - Lower baseline cost
  - Population size invariance
- Less pronounced stochastics
  - Less frequent need for Monte Carlo ensembles
- Quicker construction, runtime $\Rightarrow$More time for understanding, refinement

## Individual-Based Models

- Better fidelity to many dynamics
- Stronger support for highly targeted policy planning
- Ability to calibrate to & validate off of longitudinal data
- Greater heterogeneity flexibility
- Better for examining finer-grained consequences
  - e.g. transfer effects w/i pop.
  - Network spread
- Simpler description of some causal mechanisms

# Key Needs Motivating Individual-Based Modeling

- Need to calibrate against information on **agent history**
- Need to capture **progression** of agents **along multiple pathways** (e.g. co-morbidities)
- Wish to characterize **learning by and/or memory** of agents based on experience, or **strong history dependence** in agents
- Need to capture distinct **localized perception** among agents
- Seeking to intervene at points in, change behavior on, explain phenomena over or explain dynamics **across networks**
- Seek **distinct interventions for** many **heterogenous categories**
- Need to **capture impact** of intervention across **many categories**
- When it is much simpler to **describe behavior at indiv. level**
- Seek **flexibility** in exploring different **heterogeneity dimensions**
- **Needs of stakeholders** to engage with individual-based models
- Want to describe behaviour at **multiple scales**
- We care about **stochastics/uncertainty** caused by indiv variabilit

# Key Needs Motivating Aggregate-Based Modeling

- Need to **execute quickly** (e.g. for user interaction)
- Understand/describe system behaviour across **all possible values for parameters**
  - Seeking to mathematically **analyze the model** (e.g. to determine location or stability of equilibria) for insight
  - To determine shape of **all possible trajectories**
- Want to **use mathematical tools** (e.g. control theory )to identify high-leverage parameters, optimal policies
- Need to **extensively calibrate to much historic data**
- **Desire of stakeholders** to work at higher level
- Behavior for different subgroups **differs only in degree**
- No recourse to **software engineering knowledge**
- **Lack of detailed knowledge** of network structure/ individual-level behaviour/Individual-level data

# Individual Descriptions are Sometimes Simpler

- Understanding of *individual* behavior sometimes exceeds that of collective behavior
  - Response to locally visible incentives
  - Company's response to competition
  - Young person's response to peer pressure
  - Individual's response to scarcity of good
- Sometimes it is very difficult to derive *a priori* the aggregate dynamics resulting from individual behavior
- Individual model can be simpler, more transparent

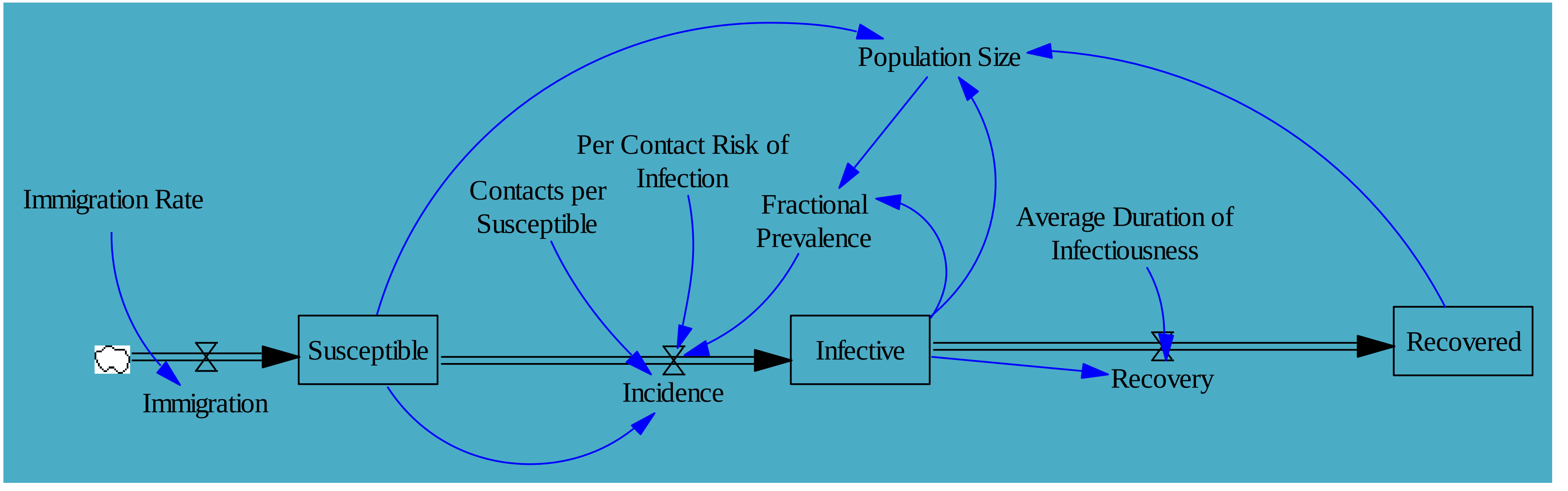
# Aggregate Descriptions are Sometimes Simpler

- Aggregate descriptions frequently allow us to abstract away from myriad lower-level hypotheses
  - May afford us an easier mode of description without the need to explicitly posit involved lower-level hypotheses
  - Can be readily formulated from partial data & applied globally
- Consider
  - Using a mixing matrix computed from partial mixing data
  - Formulating population-wide
    - Hypothesized contact networks
    - Mobility patterns driving contact

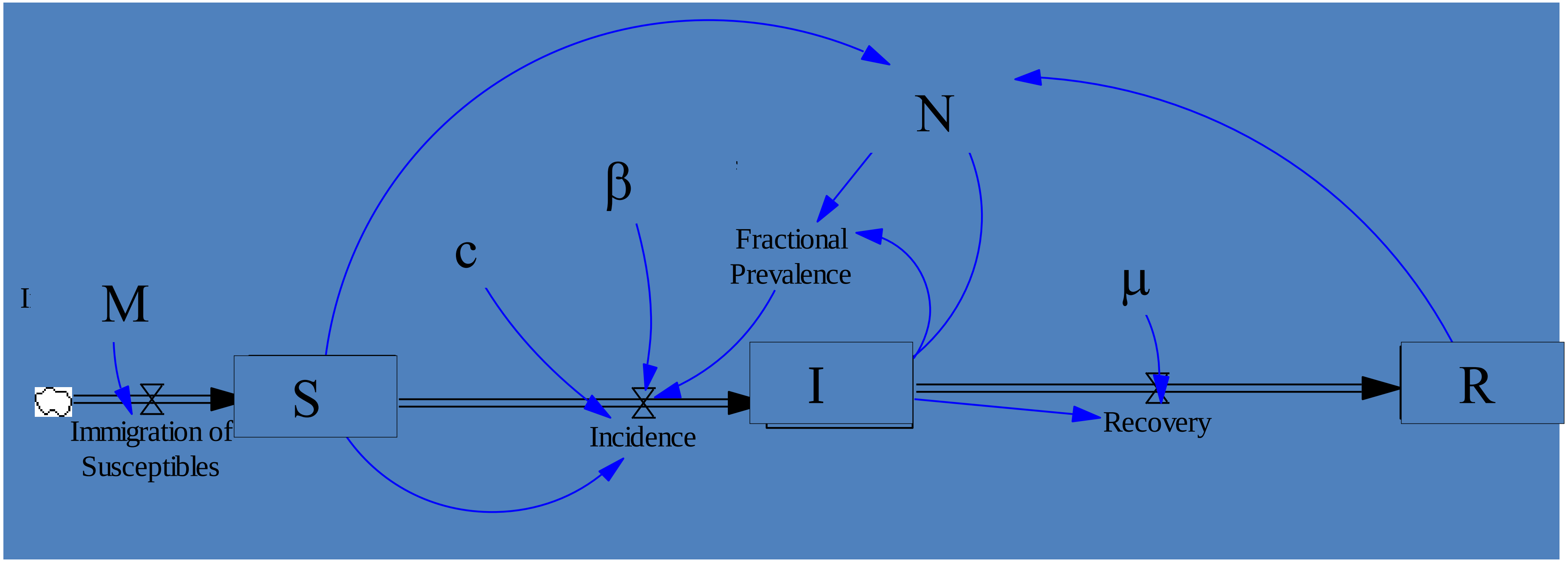
# Some Uses of Formal Approaches

- Explaining observed behavior patterns
- Identifying possible behavior modes over a wide variety of possible scenarios (e.g. via eigenspace & phase plane analysis)
- Identifying how behavior depends on parameters (stability, location of equilibria)
- Creating "self-correcting" models (via control theory)
  - Individual-based models are typically not identifiable
- Formal calibration methods

# Example Aggregate Model Structure



# Mathematical Notation

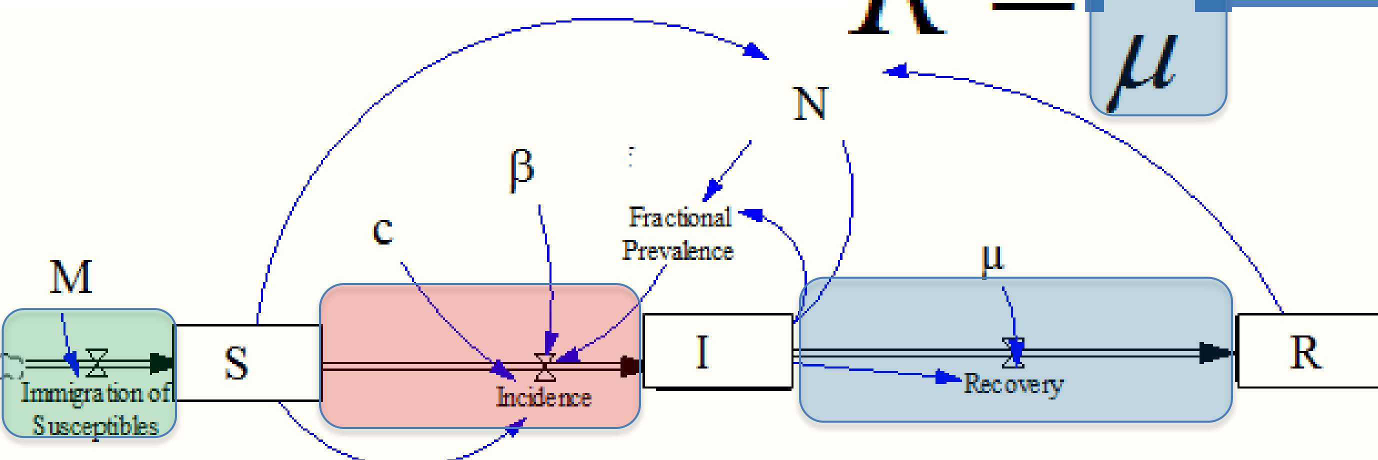


# Underlying (Ordinary) Differential Equations

$$\dot{S} = M - c\left(\frac{I}{N}\right)\beta S$$

$$\dot{I} = c\left(\frac{I}{N}\right)\beta S - \frac{I}{\mu}$$

$$\dot{R} = \frac{I}{\mu}$$



# Model Mathematical Analysis

System Linearization (Jacobian)

$$\begin{bmatrix} -\beta\,\mathit{Infectives} - \delta & -\beta\,S - \delta \\ \beta\,\mathit{Infectives} & \beta\,S - \dfrac{1}{\mu + \dfrac{\tau\,\mathit{Infectives}}{h}} + \dfrac{\mathit{Infectives}\,\tau}{\left(\mu + \dfrac{\tau\,\mathit{Infectives}}{h}\right)^2 h} \end{bmatrix}$$

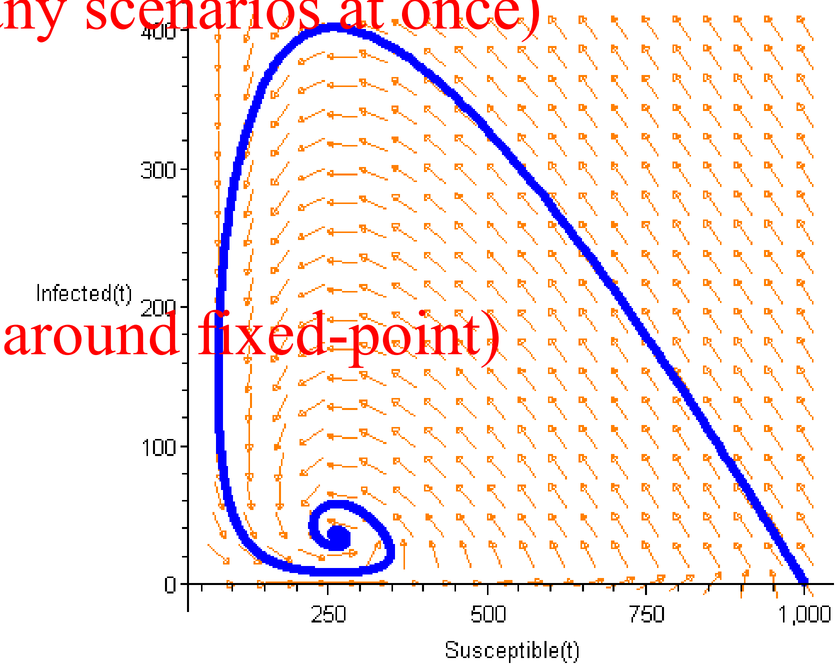
Fixed-Point Criteria

$$\dot{S} = -c\left(\frac{I}{N}\right)\hat{\beta}S + R\delta = 0$$

$$\dot{I} = c\left(\frac{I}{N}\right)\hat{\beta}S - \frac{I}{\mu + \tau\dfrac{I}{h}} = 0$$

$$\dot{R} = \frac{I}{\mu + \tau\dfrac{I}{h}} - R\delta = 0$$

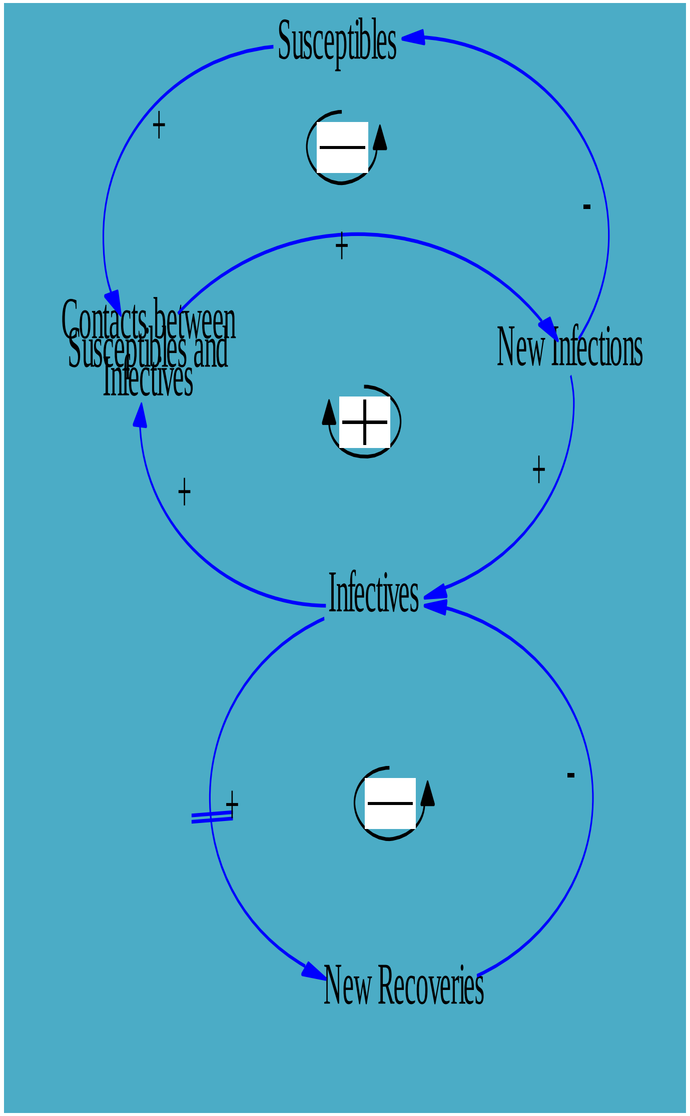
State space diagram (reasoning about many scenarios at once)



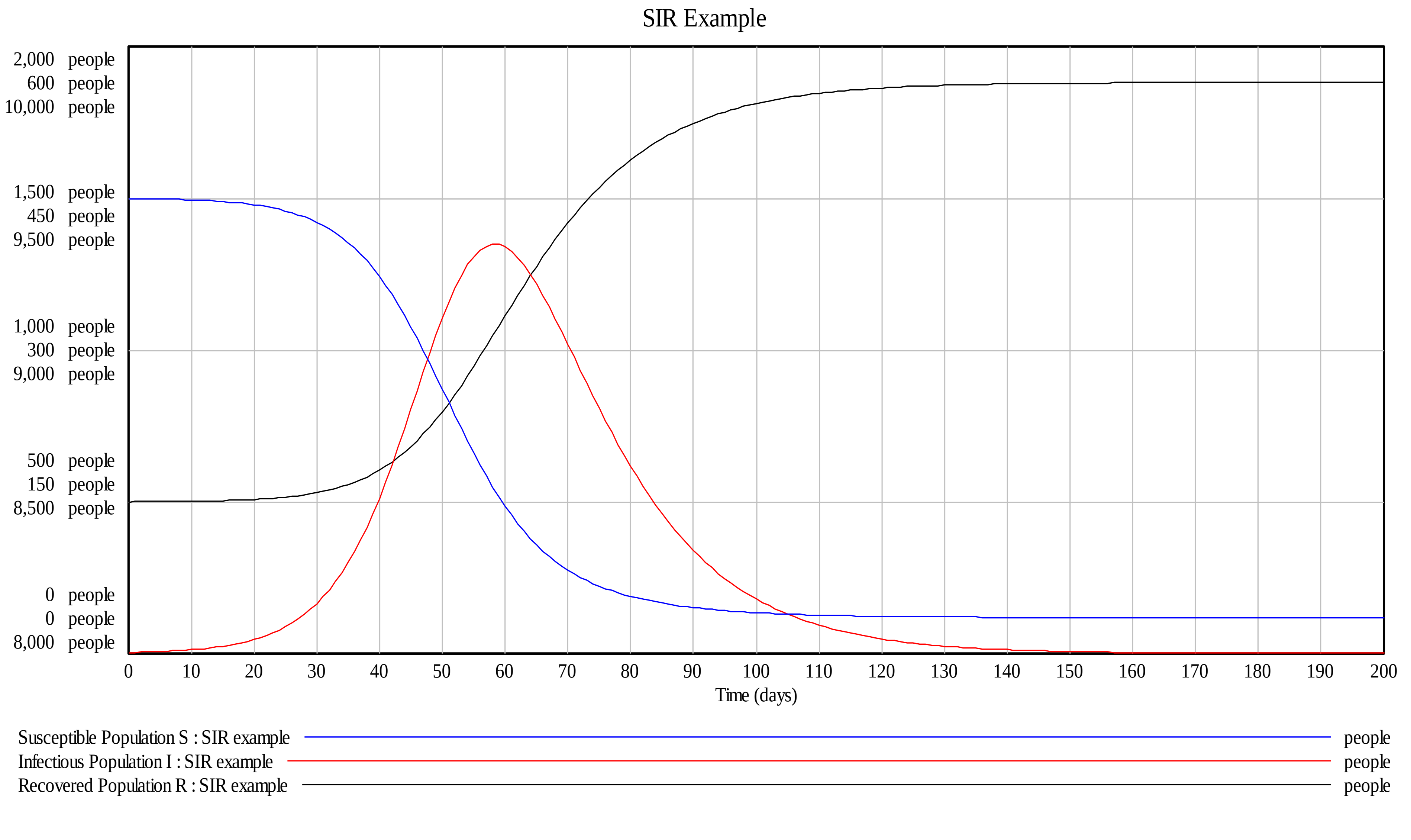

Eigenvalues (e.g. for stability analysis around fixed-point)

$$\frac{1}{2}\beta S - \frac{1}{2}\frac{1}{\mu + \dfrac{\tau\,\mathit{Infectives}}{h}} + \frac{1}{2}\frac{\mathit{Infectives}\,\tau}{\left(\mu + \dfrac{\tau\,\mathit{Infectives}}{h}\right)^2 h} - \frac{1}{2}\beta\,\mathit{Infectives} - \frac{1}{2}\delta$$
$$+ \frac{1}{2}\left(\left(\beta S - \frac{1}{\mu + \dfrac{\tau\,\mathit{Infectives}}{h}} + \frac{\mathit{Infectives}\,\tau}{\left(\mu + \dfrac{\tau\,\mathit{Infectives}}{h}\right)^2 h}\right)^2 - 2\left(\beta S - \frac{1}{\mu + \dfrac{\tau\,\mathit{Infectives}}{h}}\right.\right.$$
$$\left.\left. + \frac{\mathit{Infectives}\,\tau}{\left(\mu + \dfrac{\tau\,\mathit{Infectives}}{h}\right)^2 h}\right)(-\beta\,\mathit{Infectives} - \delta) + (-\beta\,\mathit{Infectives} - \delta)^2 + 4\beta\,\mathit{Infectives}\,(-\beta S - \delta)\right)^{\frac{1}{2}}$$

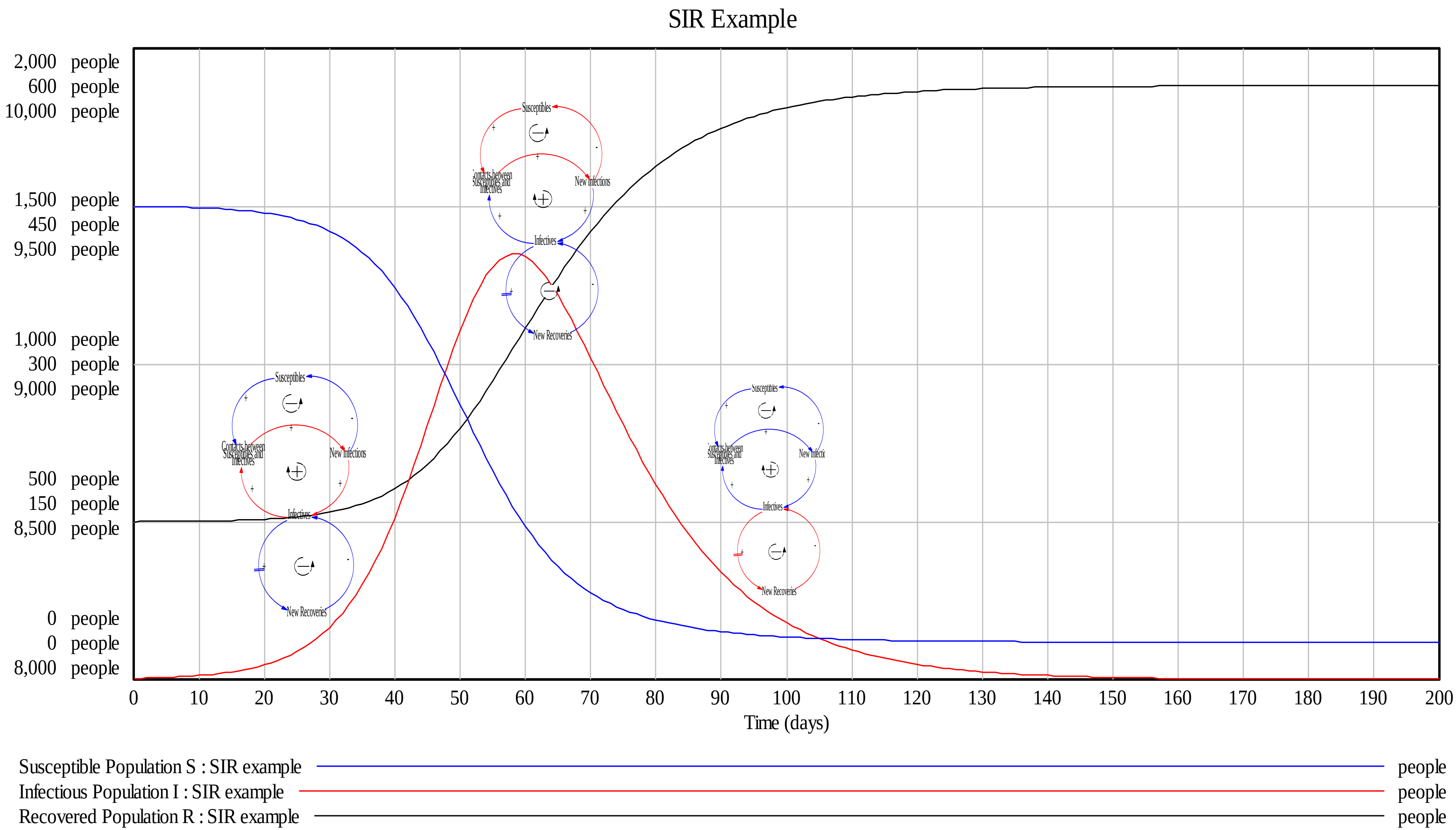
# Feedbacks Driving Infectious Disease Dynamics



# Example Dynamics of SIR Model (No Births or Deaths)



# Shifting Feedback Dominance



# Fidelity to Dynamics

- Adequate characterization of system's causal processes may require fine-grain representation
  - Rich heterogeneity
  - Learning and adaptation
  - Response to local incentives
  - Memoryful processes
  - Behavior over persistent networks
- Aggregate behavior is *not* necessarily the same as |Population|*(Behavior of "average" individual)
- May be able to calibrate an aggregate model to results of individual-level model *post hoc*

# Example of Concern: History Information

- Heterogeneity with respect to individual history can be highly important for future health
  - Whether vaccinated
  - *in utero* exposure
  - Degree of glycemic control over the past decade
  - Exposure to adiposity
  - Previous exposure to a pathogen
- In some areas of health, we have access to longitudinal data that provides information on individual historical trajectories.

# Capturing History Information

- Individual based model
  - Both discrete & continuous history information can be readily captured
    - Categorical/discrete: State (in statechart) or variable
    - Continuous: Variable
  - Readily able to capture records of trajectories
- Aggregate model
  - Categorical/discrete: Limited discrete history information can be captured by disaggregating stocks
    - Curse of dimensionality provides tight limits on # of aspects of history can be recorded
  - Continuous: Almost always infeasible
  - Very complex to provide distributions of trajectories (via convolution of potentially changing PSFs of stocks)

# Longitudinal Fidelity: Aggregate Models

- An aggregate model provides an ongoing series of *cross-sectional* descriptions of system state
  - In Calibration & validation, we can do rich comparison of these cross-sectional descriptions against available point or time-series data
  - Because the model does not track individuals, we generally cannot explicitly extract model longitudinal trajectories from the model for comparison with empirical giving longitudinal trajectories

# Longitudinal Fidelity: Individual-Based Models

- An individual-based model provides easily accessible *cross-sectional* and longitudinal descrip. of system state
  - The system state at a particular moment in time is cross-sectional
  - By following & recording the trajectories of particular individuals, we can obtain longitudinal description
- In Calibration & validation, we can do rich comparison of both longitudinal and cross-sectional descriptions against available point or time-series data
  - It is in principle possible to have a model that accords with cross-sectional data, but which is at odds longitudinally

# Comparisons of Model & History that are Difficult in an Aggregate Model

- Proportions of people with certain history characteristics (e.g. fraction of women who develop T2DM who have had 2 or more bouts of gestational diabetes, those with a certain duration of time separating TB infection sand active TB)
  - Can be very valuable for calibration
  - This is critical for assessing model accord with observed effect size (Relative Risk/Odds ratio)
- Model vs. historic trajectories (e.g. for timing of some transitions) for people with certain history characteristics

# Example of Additional Information from Longitudinal Data

- Consider trying to distinguish pairs of situations
- e.g.: Smoking
  - Situation 1: One set of people quit & stay quit as former smokers, another set remain as current smokers
  - Situation 2: The entire set of people cycle through situations where they quit, relapse & repeat
- These two situations have very different health consequences
- We'd probably choose vary different sets of interventions for these two situations
- Similar examples are easy to imagine for obesity, STIs, TB, glycemic control & diabetes, etc.

# Trajectories Summary

- If any of the following are true….
  - You have significant longitudinal information you'd strongly like the model to match
  - You have good reason to think that trajectory history has important consequences for health
  - You seek to examine the effect of policies that make use of information on individual history (e.g. # previous treatments)
- Then you should strongly consider building a model that captures this history information
  - By disaggregating stocks, you can capture limited *discrete* history information in an aggregate model (e.g. whether a person was exposed *in utero*, Time Since Quit for FS, whether a woman has had a history of gestational diabetes)
  - There is significantly greater flexibility for collecting continuous or discrete history information for guiding individual dynamics & for calibration/validation comparison to historic longitudinal data

# Importance of Heterogeneity

- Heterogeneity often significantly impacts policy effectiveness
  - Policies preferentially affect certain subgroups
    - Infection may be maintained within certain subgroups even though would tend to go extinct with random mixing in the entire population
  - Policies alter balance of heterogeneity in population
    - Shifts in the underlying heterogeneity can change aggregate population statistics
  - Given a non-linear relationship, inaccurate to use the mean as a proxy for whole distribution
- **Assessing policy effectiveness often requires representing heterogeneity**
- ***Flexibility* in representing heterogeneity is hard to achieve in aggregate (coarse-grained) models**

# Impacts of Heterogeneity on Policy Effectiveness

- Value of breast cancer detection (Park & Lees)
- Impact of airbags on deaths (Shepherd&Zeckhauser)
- Value of hernia operations (Neuhauser)
- Impact of cardiovascular disease interventions (Chiang)
- Controlling blood pressure (Shepherd&Zeckhauser)
- Effectiveness of mobile cardiac care unit (Shepherd&Zeckhauser)
- Value of breast cancer treatment (Fox)
- Taeuber paradox (Keyfitz)

# Heterogeneity & Equity Considerations

- Failure to disaggregate (to represent heterogeneity) can impose implicit value judgements! e.g.
  - Treating situation as net zero cost if favouring group A while disadvantaging group B

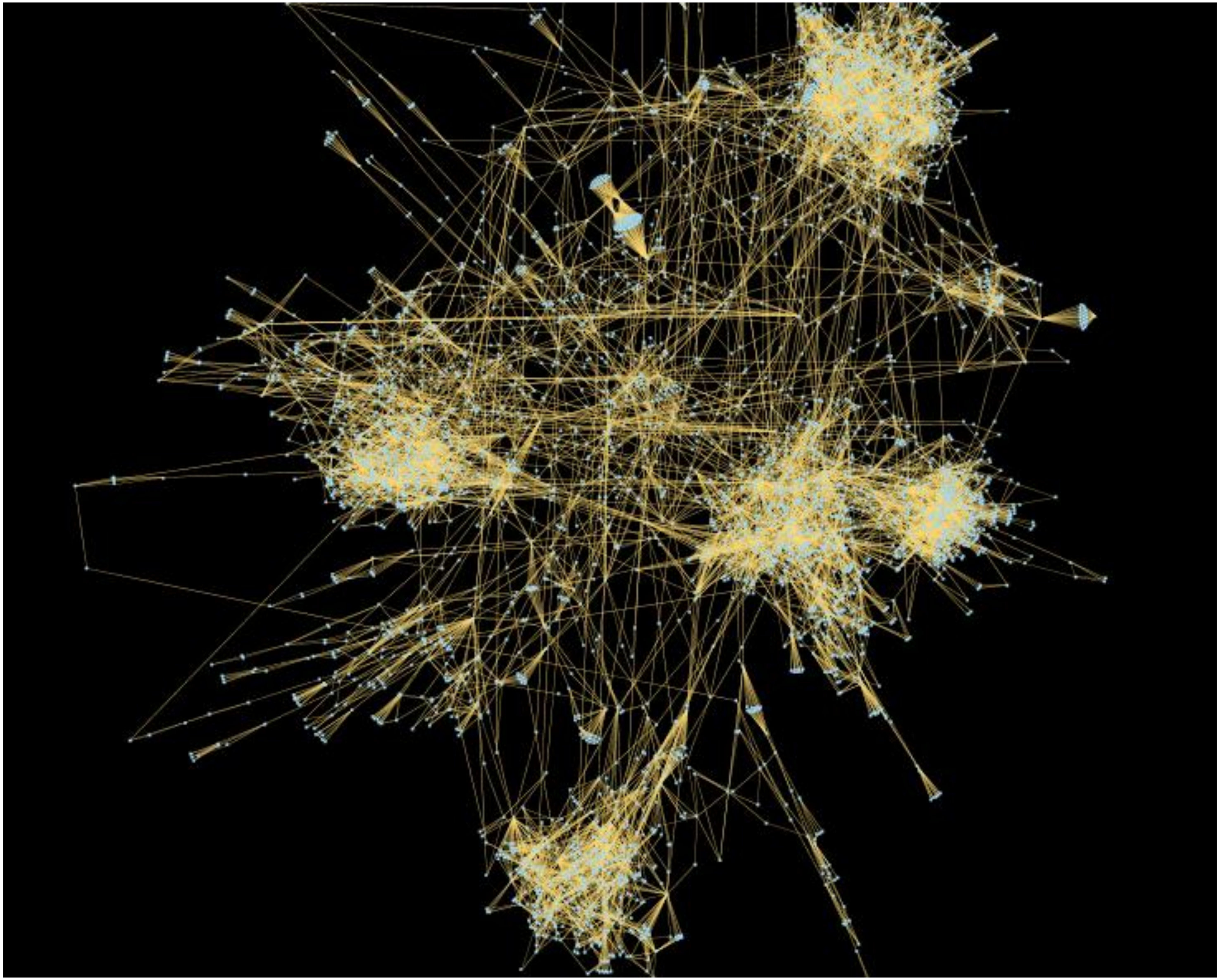
# Slides With Elements Adapted from External Source

Redacted from Public PDF for Copyright Reasons

# Importance of Core Groups

- Someone with high # of partners is both
  - More likely to be infected by one of the partners
    - Connect to lots of partners
    - More likely than the average individual to be connected with another high-contact person (in turn more likely to be connected)
  - Likely to pass on the infection more susceptible persons
- Often high-contact individuals connect in networks
- We may see very different infection rates in high contact-rate individuals
  - Core groups may be the key factor sustaining the infection
- Via targeted interventions on high contact people, may be able to achieve great "bang for the buck"
- ***Because of all of these considerations, we often seek to explicitly represent & reason about interventions targeting these individuals & their networks***

# Example of Network Clustering



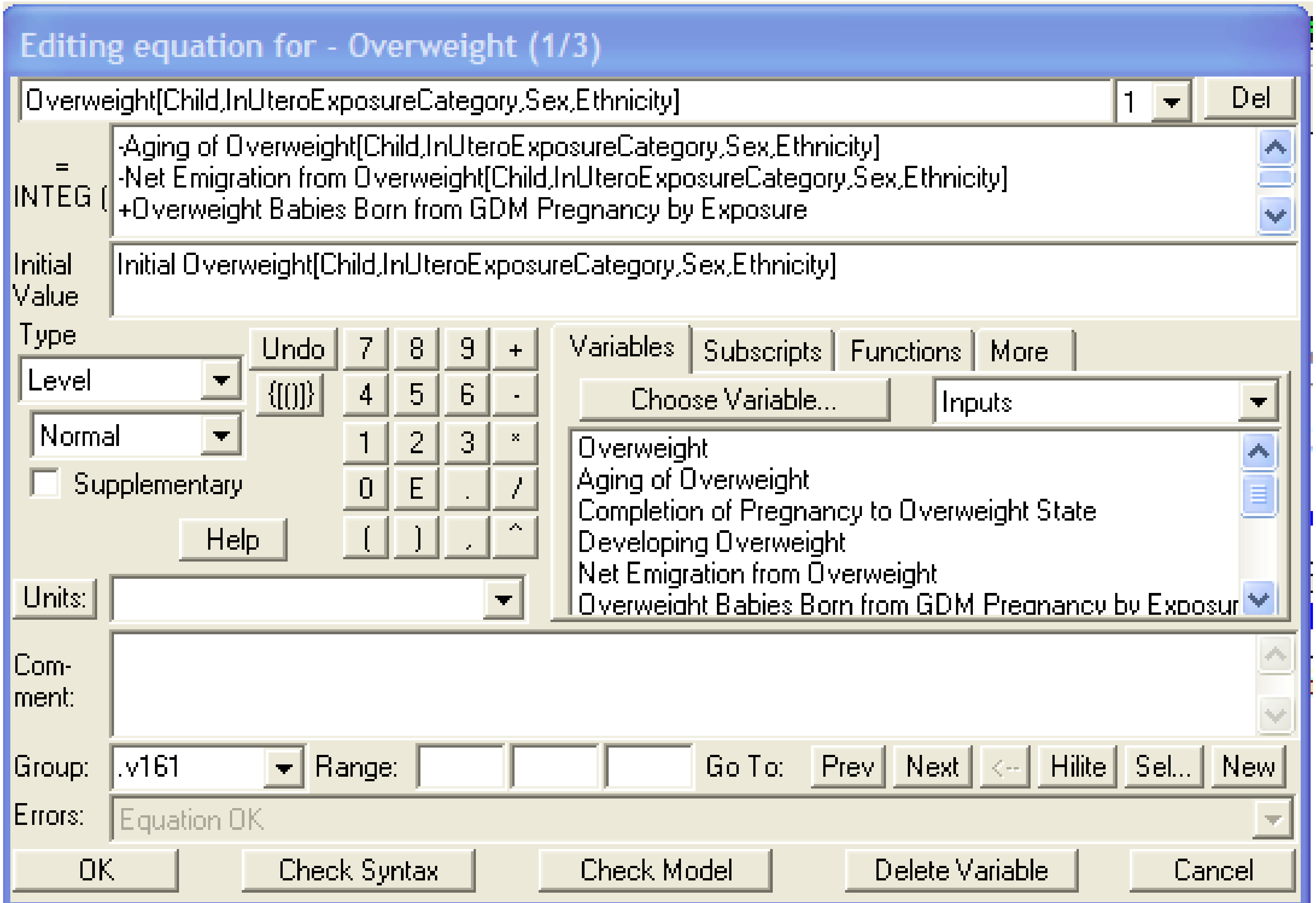
# Frequent Heterogeneity Concerns

- No clear boundaries at which to divide people up into discrete categories
- Many dimensions of heterogeneity simultaneously
  - Capturing state with respect n factors requires n dimensions of heterogeneity!
- Need to consider progression along many dimensions simultaneously

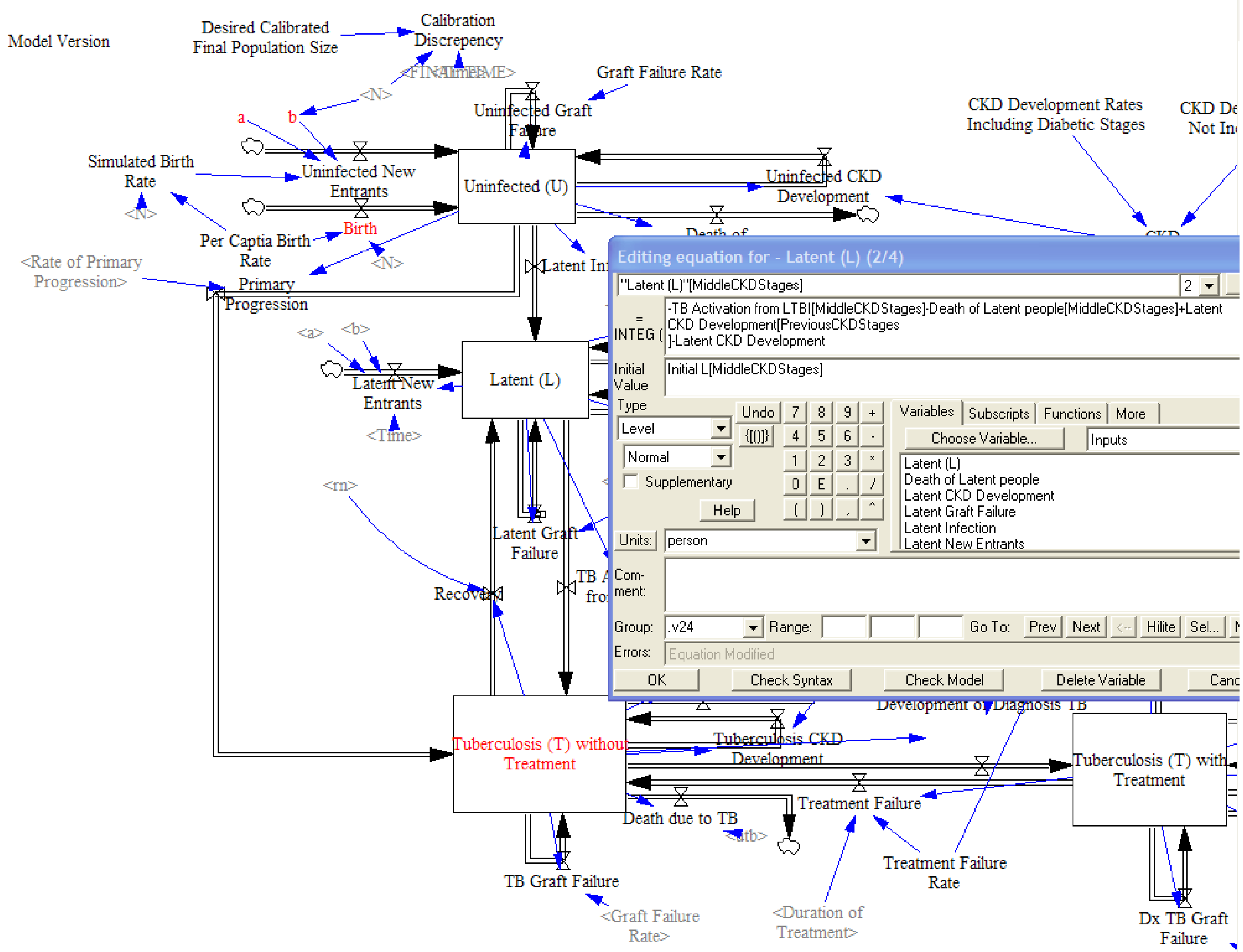
# Challenges for Aggregate Model Formulation: Heterogeneity

- Two aggregate means for representing heterogeneity are limited:
  - Attribute-based disaggregation (e.g. via subscripts)
    - Need *n* dimensions to capture individual state with respect to *n* factors of heterogeneity
    - **Poor (geometric) scaling to large # dimensions**
    - Global structural, equation changes required to incorporate new heterogeneity dimensions
    - Awkwardness in stratifying
  - Co-flows
    - Efficient and precise but highly specialized

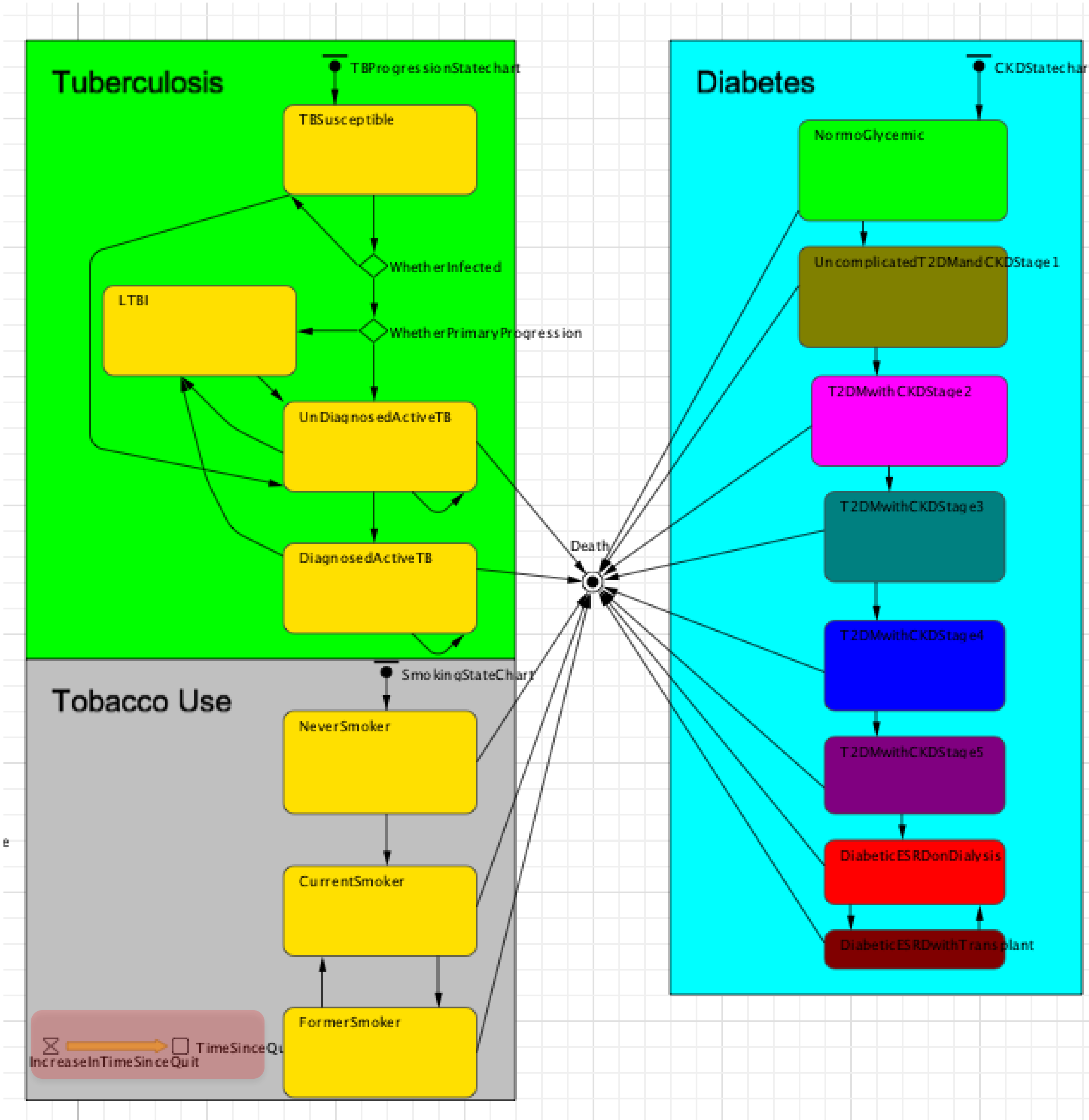
# Fragility of Multi-Dimensional Subscripting



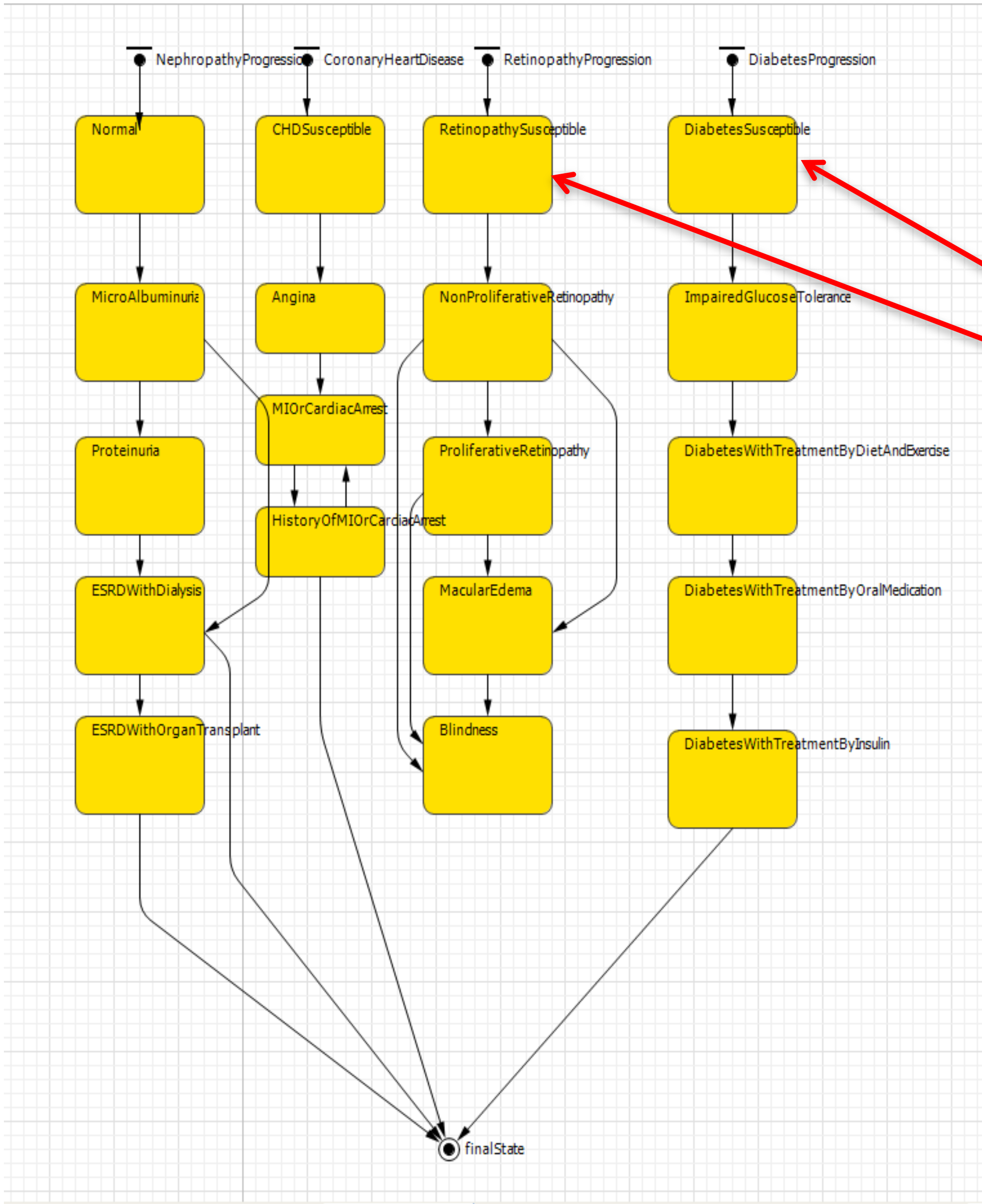
# Combinatorial Subscripting: Multi-Dimensional Progression



# Parallel Transitions



# Parallel State Transition Diagrams




A person is in some particular state with respect to each of these (condition specific) state transition diagrams

This requires representing combinations of possibilities in an aggregate model

# Capturing Heterogeneity in Individual-Based vs. Aggregate Models

- Consider the need to keeping track a new piece of information for each person (with d possible values)
  - E.g. age, sex, ethnicity, education level, strain type, city of residence, etc.
- Aggregate Model: Add a subscript
  - This multiplies the model size (number of state variables into which we divide individuals) by *d*!
- Individual based model: Add field (variable/param)
  - If model already has c fields, this will increase model size by a fraction 1/c.

# Desired: Flexibility in Representing Heterogeneity

- It is desirable to capture heterogeneity in a flexible fashion.
- More judicious exploration of whether to represent heterogeneity
  - Examine whether some observed covariation might simply be due to colinearity
    - Represent added heterogenity dimensions with no causal interaction, see if model covariations matches what is seen in external world
      - e.g. represent age in a TB model, see if rates of LTBI by age in the model match age-specific infection rate observations
  - Try adding in new dimension of heterogeneity & effects, and see if has impact that is both substantive & plausible