Building out a research-level model

What all does this actually entail?

Brainstorming example

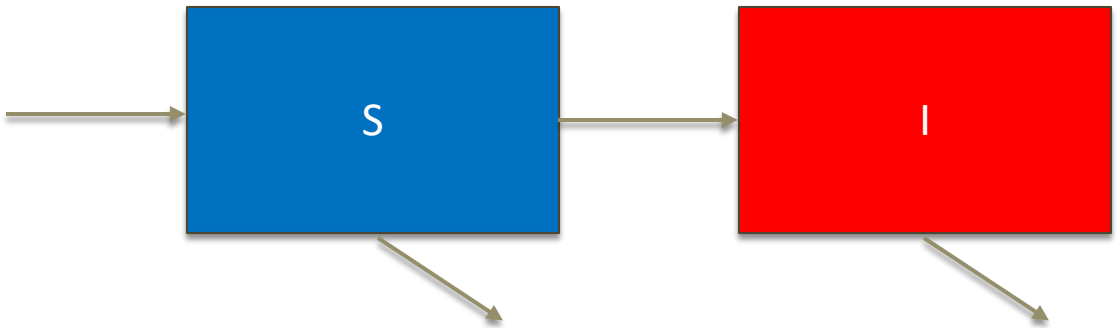
We will talk through an example involving…

HIV among men who have sex with men ☺

In base EpiModel, imagine you have built out a model with:

- time step = 1 day
- a population with no attributes other than infection status
- three network models on this population – main, casual, one-time
- homogeneous act rate for main, another for casual
- basic vital dynamics and epi for an SI model:
  - arrival rate
  - homogeneous transmission probability
  - departure rate for susceptibles
  - departure rate for infecteds

Brainstorming example



Adding heterogeneity

Let's first build out one attribute. What shall we chose?

Age

- What attribute(s) will we add?
- What module(s) will we add?
- What will the module(s) comprise?
- What existing model elements will be revised?
- What network statistics or model parameters do we need?
- What data do we need?

Adding age

- What attribute(s) will we add?
  - age
  - but in what unit?
  - and what will we assign as starting values?
- What modules will we add?
  - aging
- What will the module(s) comprise?
  - each node ages by one day per time step

Adding age

- What existing model elements will be revised?
- What network statistics or model parameters do we need?
- What data do we need?
  - Initialization – need to assign everyone age (what will that look like?)
  - Arrivals - need to assign an age (what will that look like?)
  - Departures - can now also depend on age (what will that look like?)

Adding age

- What existing model elements will be revised?
- What network statistics or model parameters do we need?
- What data do we need?

  - Networks:
    - Main effect of age – how parameterize? What data?
    - Mixing on age – how parametrize? What data?
    - Dissolution by age? – how parametrize? What data?

Adding age

- What existing model elements will be revised?
- What network statistics or model parameters do we need?
- What data do we need?
  - Act rate?
    - probably yes, but what are the details?
  - Transmission probability given SI act?

Adding testing

- Now let's add in HIV testing. We'll brainstorm:
  - A bare-minimum approach
  - A more realistic approach

Adding testing: bare minimum

- What attribute(s) will we add?
  - Diagnosis status
    - What values does it take?
    - How do we initialize it?
- What module(s) will we add?
  - Testing

Adding testing: bare minimum

- What will the module(s) comprise?
  - Each node has some probability of testing for HIV each time step
    - (Or, rather, each node who hasn't already tested HIV+)
    - Bare minimum: everyone has the exact same probability, all the time
  - We flip a virtual coin to determine who tests
  - For those that do, their current infection status becomes their diagnosis status

- What existing model elements will be revised?
  - Bare minimum: initialization and arrivals
  - (But if that's it, then what's the point?)

Adding testing: bare minimum

- What network statistics or model parameters do we need?
  - Daily average probability that an MSM who has not previously tested positive for HIV gets an HIV test
- What data do we need?
  - Survey data about # of tests in last X months/years?
  - Clinical data about # of HIV tests in some setting?
    - (But do we know the population denominator?)

Adding testing: more realistic

- What attribute(s) will we add?
  - Diagnosis status
  - Time since last test?
- What module(s) will we add?
  - Testing

Adding testing: more realistic

- What will the module(s) comprise?
  - Each node who hasn't already tested for HIV+ has some probability of testing for HIV each time step
  - To be more realistic, what should this probability depend on?
    - Time since last test?
    - Age of node?
    - Recent sexual behavior?
  - We flip a virtual coin to determine who tests
  - For those that do, we determine their test result. Based on?
    - Current infection status
    - Time since infection
    - Window period of test
    - Sensitivity and specificity of test more generally

Adding testing: more realistic

- What existing model elements will be revised?
  - Initialize and arrivals
  - Items to disaggregate by dx status:
    - Network:
      - Overall probability of forming a main or casual partnership (nodefactor)
      - Overall probability of breaking a main or casual partnership (nodefactor)
      - Overall probability of a one-time contact (nodefactor)
      - Partner selection for each type (serosorting; nodematch)
    - Other
      - Act rate

Adding testing: more realistic

- What net statistics /model parameters do we need? And what data do we need?
  - Perhaps need to start thinking about ***why*** different people test
    - On some reasonably regular schedule?
    - As the result of a potential exposure?
    - Spontaneously when presented by an easy opportunity?
    - (In practice – as part of getting on and being on PrEP – but we don't have PrEP in our model yet)
    - Then each is governed by various parameters
  - Perhaps we need to consider a new attribute – tester type
  - Does everybody only one type?

Adding testing: more realistic

- What net statistics /model parameters do we need?
  - # of partnerships in cross-section by diagnosis status of each member
  - Mean relational duration by diagnosis status of each member
  - # one-time contacts by diagnosis status of each member per time unit
  - Act rate within each type of partnership
  - If you're starting to see some data issues arise, hang on…..

- And note:
  - We've already put age into the model and allowed many of these same things to vary by age – how will the two effects interact?

Adding testing: more realistic

- What data do we need?
  - Survey data about partnerships/contacts by dx status
  - But watch out!
    - Is my partner's/contacts diagnosis status always the same as my understanding of it?
    - Might there be systematic ways these diverge, which vary bu partner type?
    - So maybe we need to think about another module next (disclosure)

Adding testing: more realistic

- What data do we need?
  - Survey data about partnerships/contacts by dx status
  - But watch out!
    - Does a preponderance of HIV-concordant relationships always reflect a process of partnership selection on dx status?
      - No – some of this is endogenous (i.e. p'ships involve transmission)
      - No – other network processes might also impact this
        - One in our model already – age mixing
    - These other phenomena are can be tricky to account for during the estimation process

Adding testing: more realistic

- These other phenomena are can be tricky to account for during the estimation process
  - Age mixing
    - make sure to set HIV prevalence and diagnosis status on the estimation network to reflect their observed levels by age, then have age mixing in the network model
    - Not too hard but requires more data

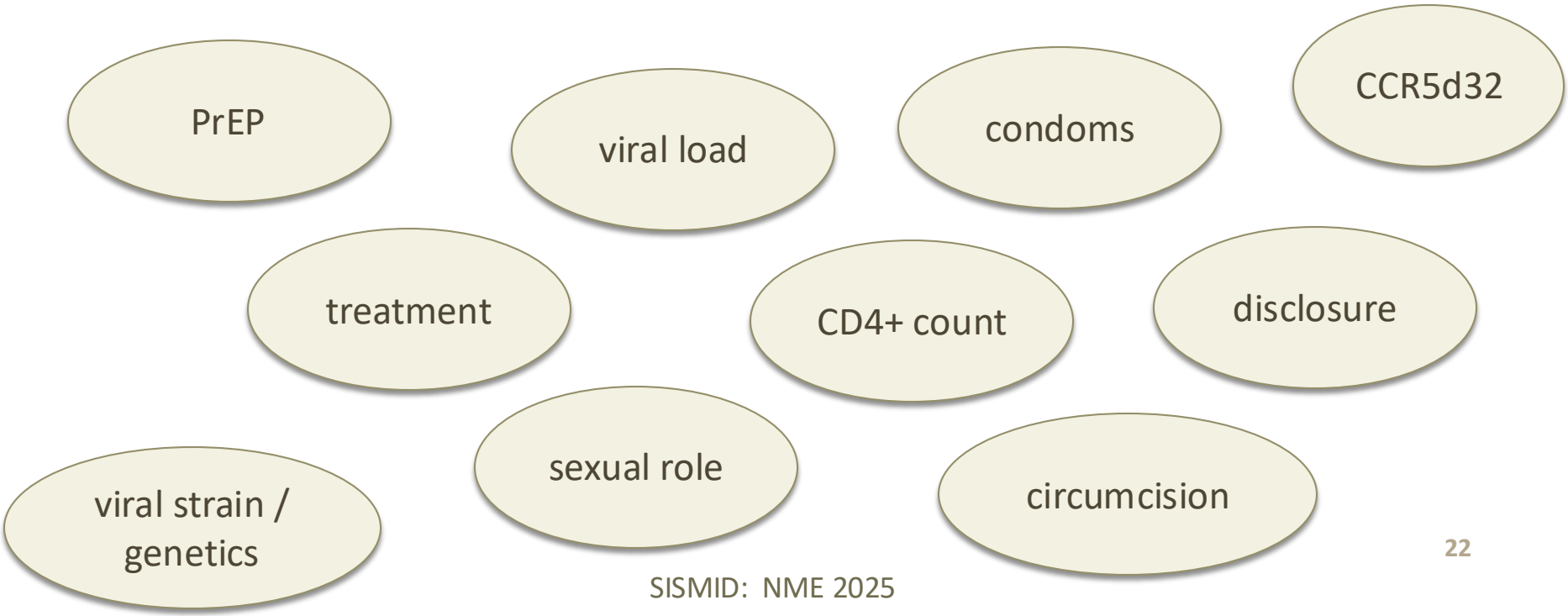
Adding testing: more realistic

- These other phenomena are can be tricky to account for during the estimation process

  - Endogenous transmission:
    - estimate model as is
    - simulate full transmission model with estimated nodematch(dx.status) param – nodematch statistic should end up a bit too high in practice
    - simulate full transmission model with a range of nodematch(dx.status) params below the estimated value and
      - check their nodematch stats
      - Pick the value that comes closest, or try new values if needed
    - This can be automated through a process known as Approxiate Bayesian Computation (ABC), with supporting R packages **EasyABC** and **EpiModelABC**

Adding other phenomena

- That was a lot!
- And we've only just begun.
- What other phenomena might you want to add to a reasonable research-level HIV model?



Adding other phenomena

- Break into groups based on one phenomenon to add
- Brainstorm all of the questions discussed
- Pay particular attention to challenges that arise that seem particularly distinctive