Estimation

Goals:

- The basic recipe for estimation
- The method you use should be tailored to the data and to the use

Basic Recipe of Estimation

1. Make an observation of the world
2. Build a model that can replicate that observation
3. Define a measure of distance between observation and model
4. Search over many (all?) parameters to find the ones that minimize that distance

Basic Recipe of Estimation

1. Make an observation of the world
    1. Time series
    2. Proportion immune
2. Build a model that can replicate that observation
    - Can be dynamic or static
4. Define a measure of distance between observation and model
    - For a given set of parameters, how far apart are observation and the model?
5. Search over many (all?) parameters to find the ones that minimize that distance
    - Analytically, brute force, or algorithmically

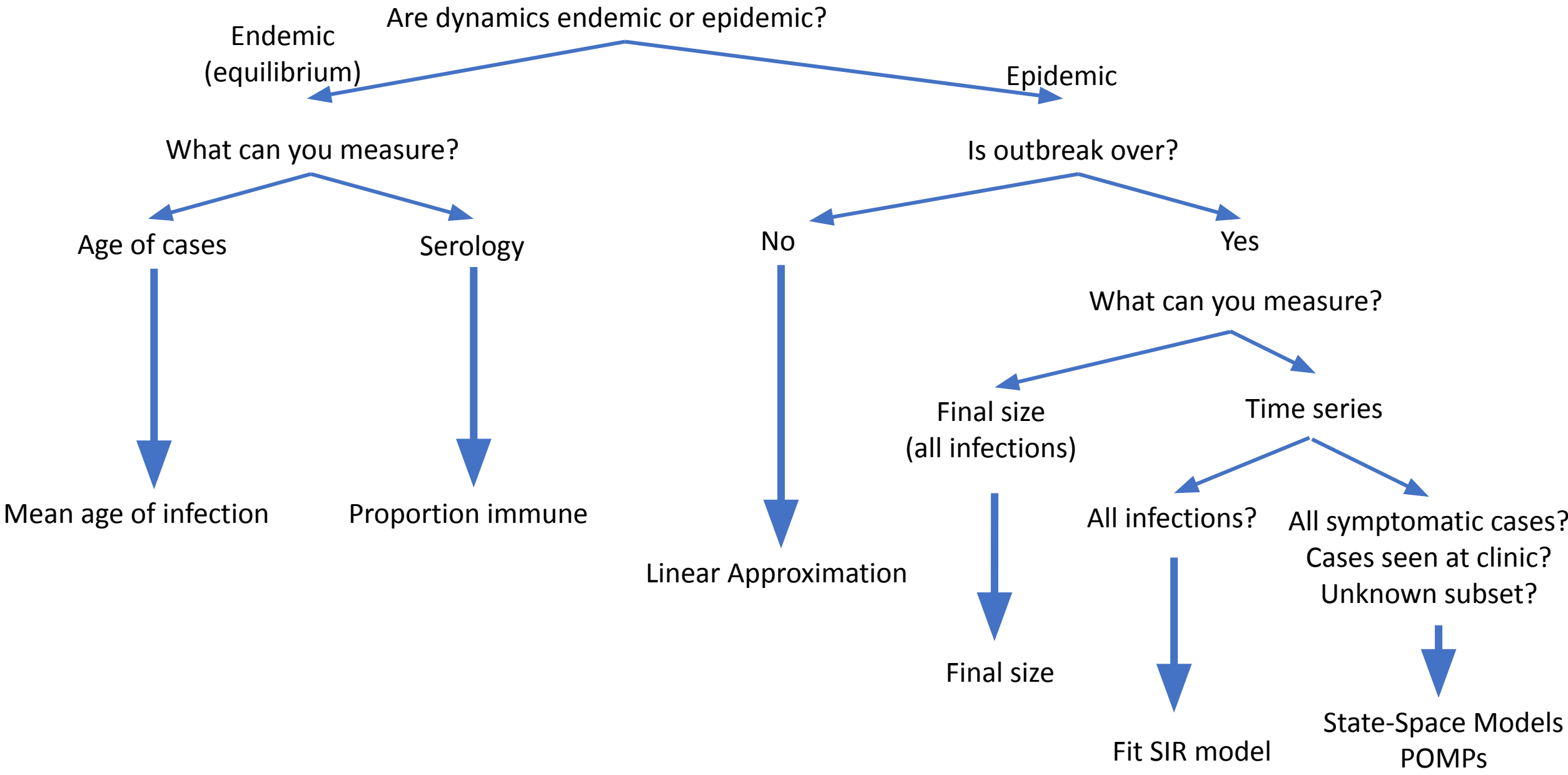
Basic Recipe of Estimation

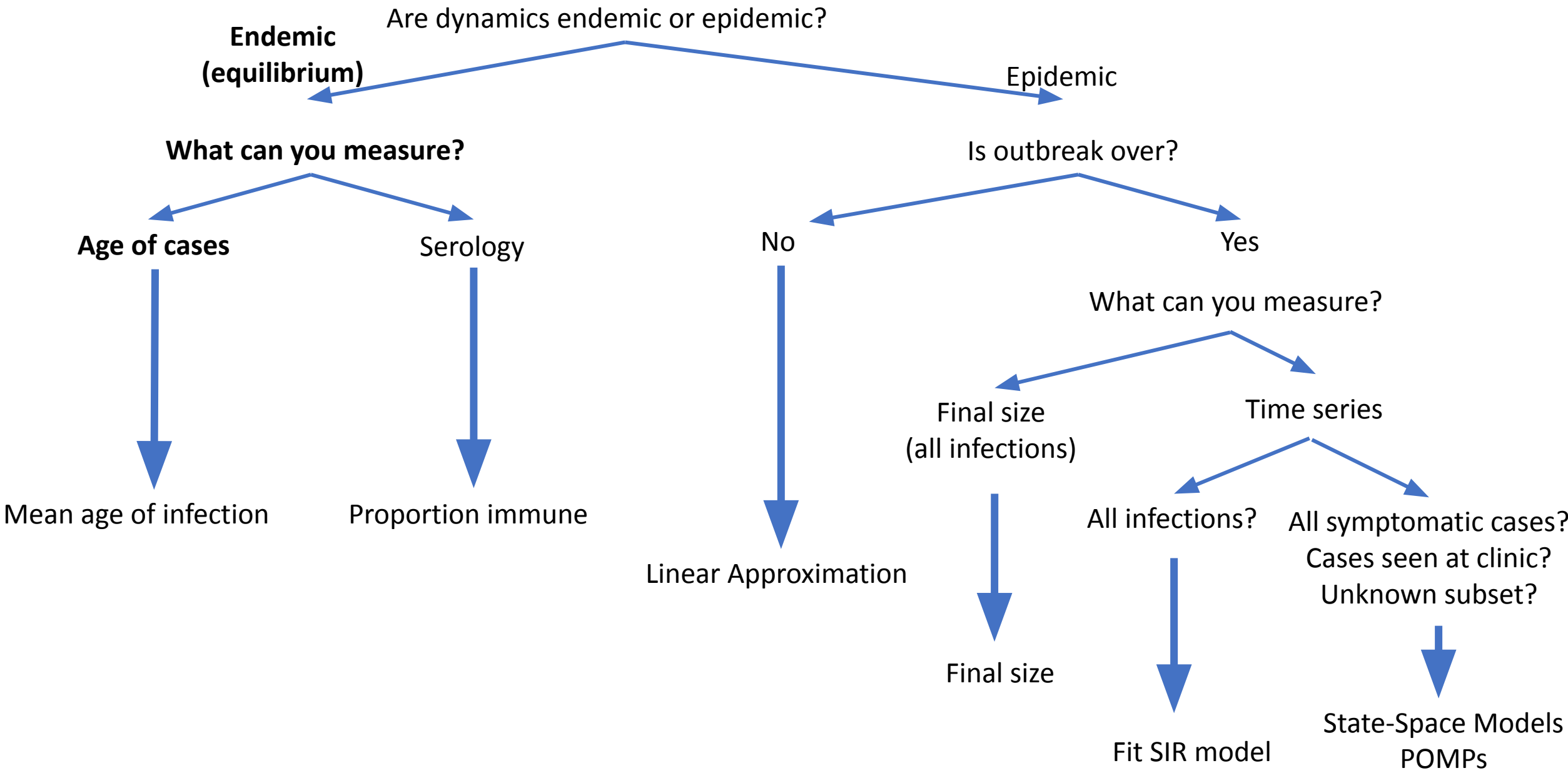
1. Make an observation of the world

What you can observe depends on where you are:

1. Are the dynamics at equilibrium?
2. What can you measure?
3. Is the outbreak over?
    - Is it still growing exponentially?

What you can observe depends on where you are:




Are dynamics endemic or epidemic?
Endemic (equilibrium)
Epidemic
What can you measure?
Age of cases
Serology
Mean age of infection
Proportion immune
Is outbreak over?
No
Yes
Linear Approximation
What can you measure?
Final size (all infections)
Time series
Final size
All infections?
All symptomatic cases?
Cases seen at clinic?
Unknown subset?
Fit SIR model
State-Space Models
POMPs

Mean Age of Infection

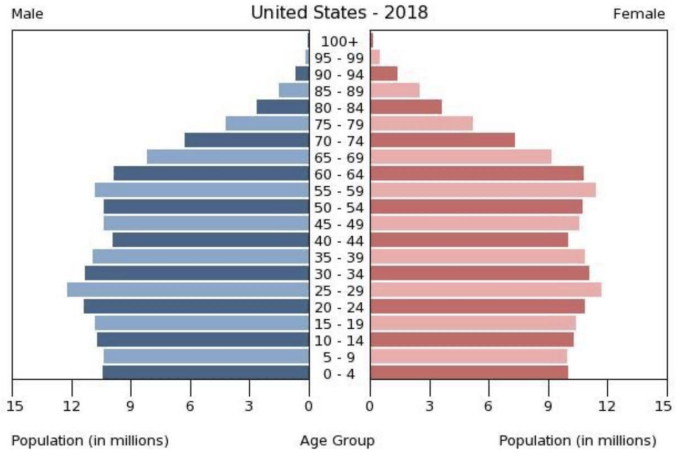
$$A = \frac{L}{R_0 - 1}$$

$$R_0 = \frac{L}{A} + 1$$

- A is the mean age of infection
- L is the life expectancy at birth

Basic Recipe of Estimation

1. Make an observation of the world
2. Build a model that can replicate that observation
   - Validity of the estimate depends on the match between the model and reality
   - $R_0 = {}^{L}/_{A}$ only holds if the age distribution of the population is exponentially distributed (e.g. constant death rate at all ages)
   - In many populations, mortality is low in the young and high in the elderly. This leads to a population age distribution that is more rectangular



Mean Age of Infection

$$A = \frac{L}{R_0}$$

$$R_0 = \frac{L}{A}$$

- A is the mean age of infection
- L is the life expectancy at birth

Not a big difference, but a bias that is generated by choosing the wrong model

Basic Recipe of Estimation

1. Make an observation of the world
2. Build a model that can replicate that observation
3. Define a measure of distance between observation and model
4. Search over many (all?) parameters to find the ones that minimize that distance

This part is trivial for mean age of infection IF the population has homogeneous mixing. Based on what we did earlier, how would you estimate R_0 for a population with age-specific mixing or force of infection?

Basic Recipe of Estimation

1. Make an observation of the world
2. Build a model that can replicate that observation
3. Define a measure of distance between observation and model
4. Search over many (all?) parameters to find the ones that minimize that distance

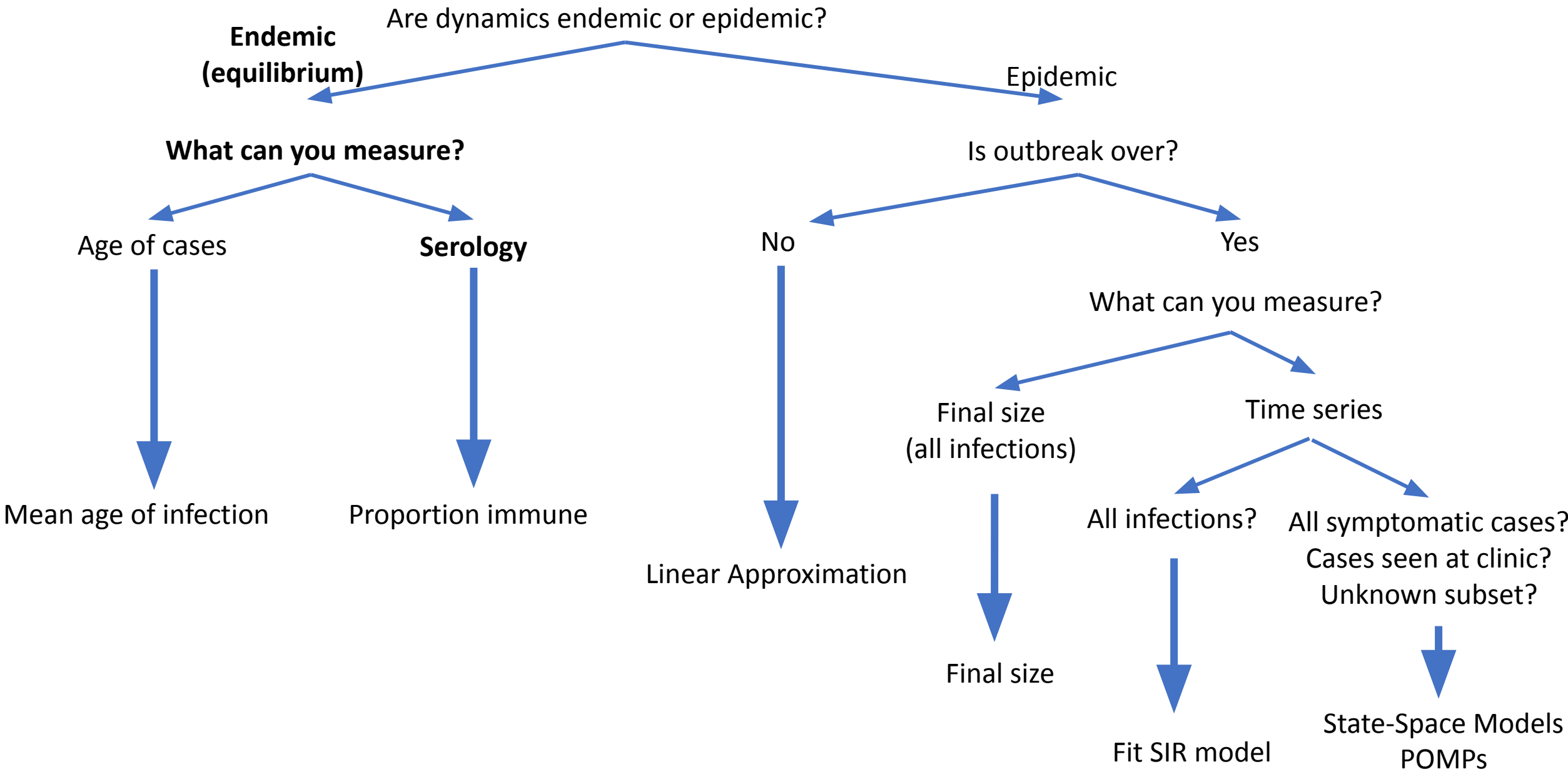
    1. Build SIR model with mixing matrix and demography that reflects the population of interest
    2. For a given R_0, simulate model to equilibrium and evaluate mean age of infection
    3. Do this for all R_0 and estimate is that which is closest to observed mean age

Basic Recipe of Estimation

1. Make an observation of the world
2. Build a model that can replicate that observation
3. Define a measure of distance between observation and model
4. Search over many (all?) parameters to find the ones that minimize that distance

1. Build SIR model with mixing matrix and demography that reflects the population of interest
2. For a given R_0, simulate model to equilibrium and evaluate mean age of infection
3. Do this for all R_0 and estimate is that which is closest to observed mean age

What if you have the whole age distribution of reported cases?


Are dynamics endemic or epidemic?
Endemic (equilibrium)
Epidemic
What can you measure?
Age of cases
Serology
Mean age of infection
Proportion immune
Is outbreak over?
No
Yes
Linear Approximation
What can you measure?
Final size (all infections)
Time series
Final size
All infections?
All symptomatic cases? Cases seen at clinic? Unknown subset?
Fit SIR model
State-Space Models POMPs

Proportion Immune

- For the standard SIR model, the equilibrium proportion immune is

$$\text{P(immune)} = 1 - \frac{1}{R_0}$$

The proportion immune can be estimated using a serological survey.
This is less likely to be biased by access to care.

Proportion Immune

- For the standard SIR model, the equilibrium proportion immune is

$$\text{P(immune)} = 1 - \frac{1}{R_0}$$

The proportion immune can be estimated using a serological survey. This is less likely to be biased by access to care.

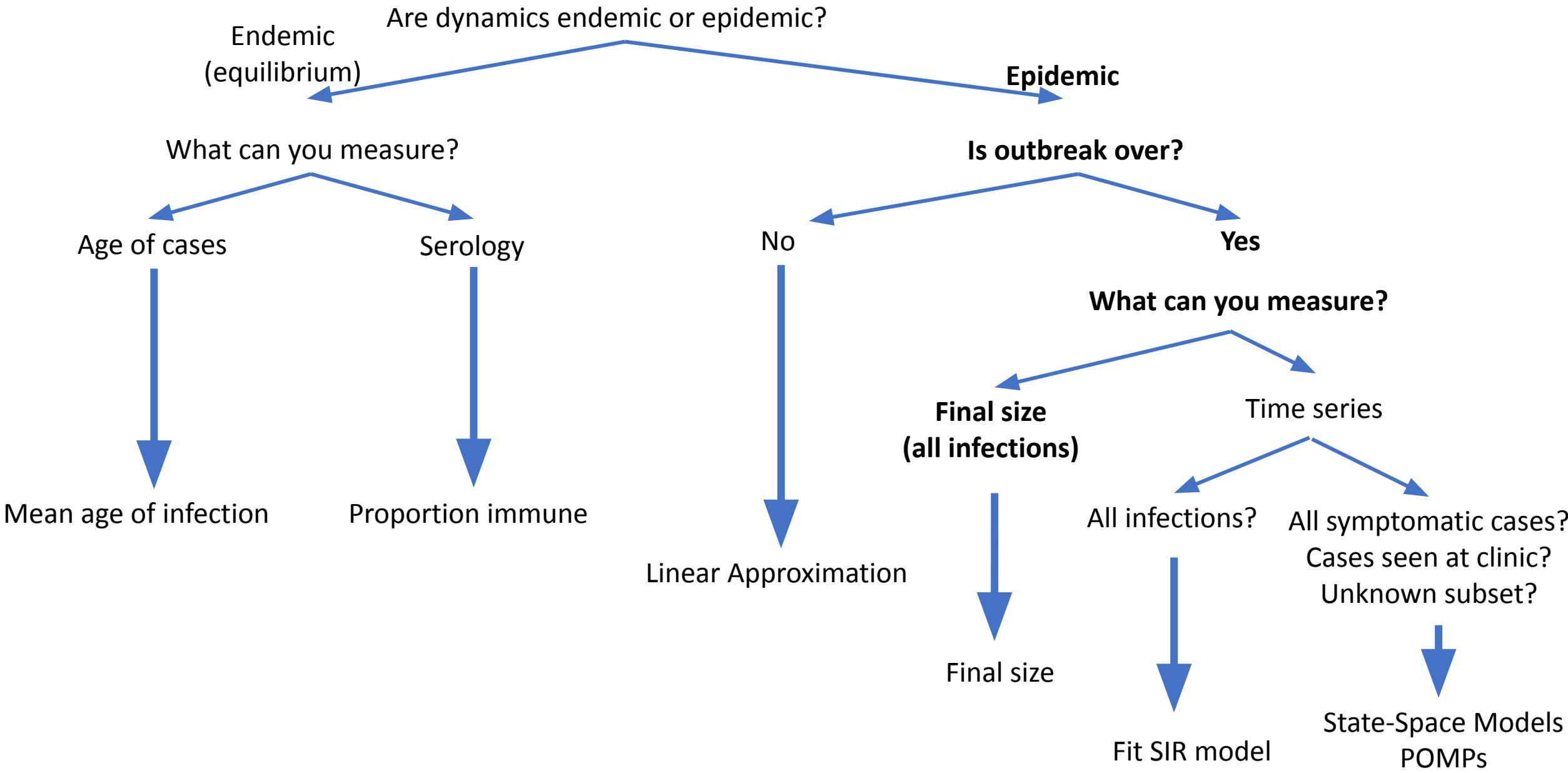
Proportion Immune

- For the standard SIR model, the equilibrium proportion immune is

$$\text{P(immune)} = 1 - \frac{1}{R_0}$$

The proportion immune can be estimated using a serological survey

As with mean age of infection – this result holds for the simple case, but the equilibrium proportion immune at equilibrium (or the age-specific seroprevalence curve) can be simulated for any specific assumptions about model structure, age-specific mixing, demographics


Are dynamics endemic or epidemic?
Endemic (equilibrium)
Epidemic
What can you measure?
Age of cases
Serology
Mean age of infection
Proportion immune
Is outbreak over?
No
Yes
Linear Approximation
What can you measure?
Final size (all infections)
Time series
Final size
All infections?
All symptomatic cases?
Cases seen at clinic?
Unknown subset?
Fit SIR model
State-Space Models
POMPs

Epidemic Final Size

$$R_\infty = 1 - e^{-R_0 R_\infty}$$

Use Newton's method for estimation !

- Citation: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3506030/

Epidemic Final Size

$$R_\infty = 1 - e^{-R_0 R_\infty}$$

- A useful result but challenging for estimation because we rarely can see ALL infections.

- Citation: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3506030/

Epidemic Final Size

$$R_\infty = 1 - e^{-R_0 R_\infty}$$

- A useful result but challenging for estimation because we rarely can see ALL infections. We're more likely to see all symptomatic cases … or all symptomatic cases that had access to clinical care and diagnostics.
- Requires an assumption that dynamics don't change during outbreak

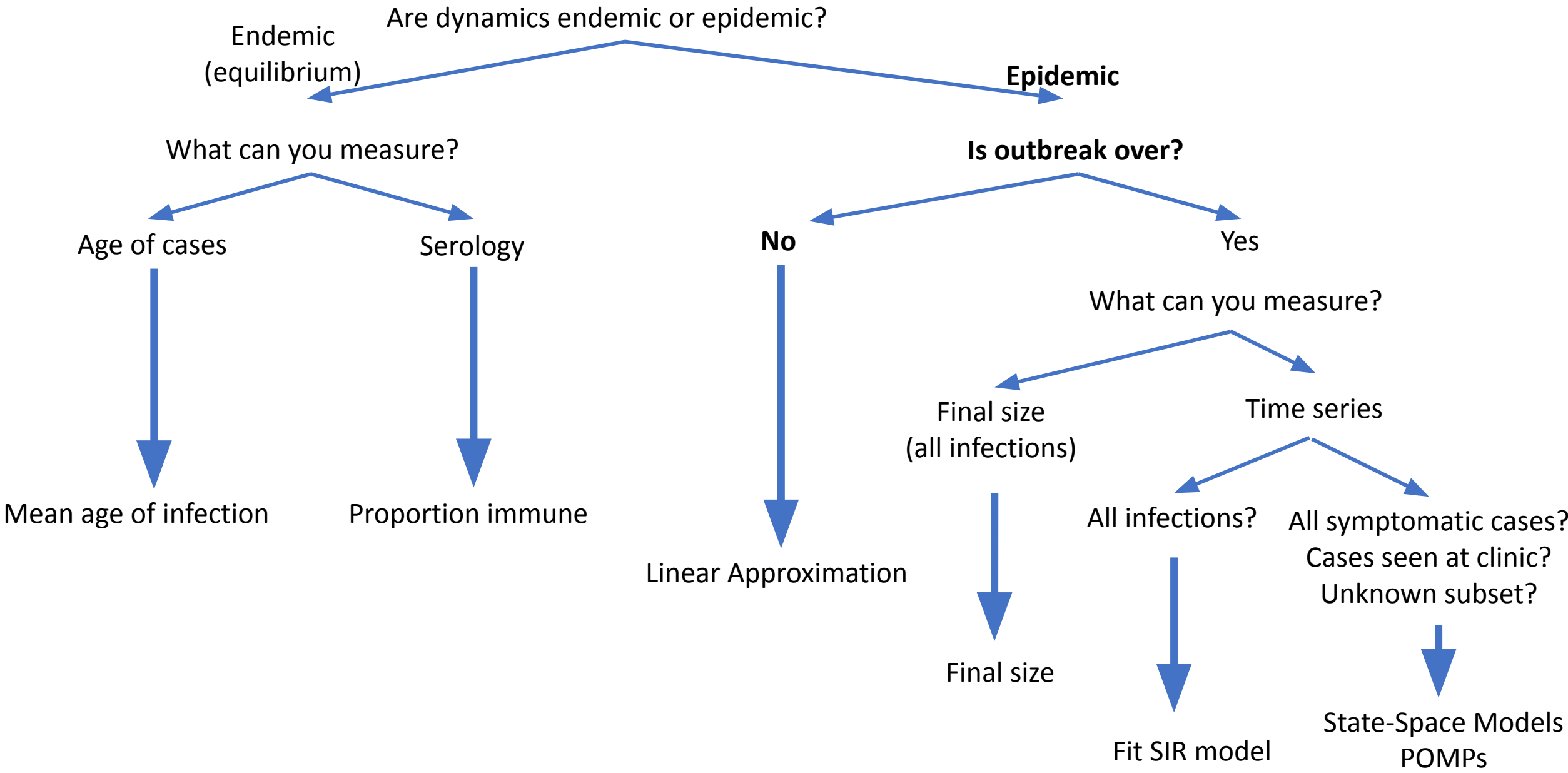
- Citation: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3506030/

Epidemic Final Size

$$R_\infty = 1 - e^{-R_0 R_\infty}$$

- A useful result but challenging for estimation because we rarely can see ALL infections. We're more likely to see all symptomatic cases … or all symptomatic cases that had access to clinical care and diagnostics.
- Requires an assumption that dynamics don't change during outbreak
- More commonly applied AFTER R_0 is estimated by another method to predict final size*
- Citation: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3506030/

* What assumption does this require?


Are dynamics endemic or epidemic?
Endemic
(equilibrium)
Epidemic
What can you measure?
Age of cases
Serology
Mean age of infection
Proportion immune
Is outbreak over?
No
Linear Approximation
Yes
What can you measure?
Final size
(all infections)
Final size
Time series
All infections?
Fit SIR model
All symptomatic cases?
Cases seen at clinic?
Unknown subset?
State-Space Models
POMPs

Fitting Time Series

- In the initial phase of an outbreak, the epidemic grows exponentially

Fitting Time Series

- In the initial phase of an outbreak, the epidemic grows exponentially

  How does herd immunity change this?

Fitting Time Series

- In the initial phase of an outbreak, the epidemic grows exponentially

Initial Geometric Growth on time scale of infectious period

$$I_1 = I_0 * R_0$$
$$I_2 = I_1 * R_0$$
$$I_2 = I_0 * R_0^2$$
$$\vdots$$
$$I_T = I_0 R_0^T$$

$$\log(I_T) = \log(I_0) + T\log(R_0)$$

Fitting Time Series

- In the initial phase of an outbreak, the epidemic grows exponentially

Initial Geometric Growth on time scale of infectious period

$$I_1 = I_0 * R_0$$
$$I_2 = I_1 * R_0$$
$$I_2 = I_0 * R_0^2$$
$$\vdots$$
$$I_T = I_0 R_0^T$$

$$\log(I_T) = \log(I_0) + T\log(R_0)$$

Exponential time scale may not be convenient for observation. What is generation time for monkeypox?

Fitting Time Series

- In the initial phase of an outbreak, the epidemic grows exponentially

Initial Geometric Growth on time scale of infectious period

$$I_1 = I_0 * R_0$$
$$I_2 = I_1 * R_0$$
$$I_2 = I_0 * R_0^2$$
$$\vdots$$
$$I_T = I_0 R_0^T$$

$$\log(I_T) = \log(I_0) + T\log(R_0)$$

Exponential time scale may not be convenient for observation. What is generation time for monkeypox?

Exponential growth on arbitrary time scale

$$I_t = I_t e^{(R_0-1)(\gamma+\mu)t}$$

$$\ln(Y_t) = \ln(Y_0) + (R_0 - 1)(\gamma + \mu)t$$

γ is the recovery rate ($1/\gamma$ is mean duration of infection)
μ is non-disease mortality rate, which can be ignored if dynamics are fast enough

Fitting Time Series

- In the initial phase of an outbreak, the epidemic grows exponentially

Initial Geometric Growth on time scale of infectious period

$$I_1 = I_0 * R_0$$
$$I_2 = I_1 * R_0$$
$$I_2 = I_0 * R_0^2$$
$$\vdots$$
$$I_T = I_0 R_0^T$$

$$\log(I_T) = \log(I_0) + T\log(R_0)$$

Exponential time scale may not be convenient for observation. What is generation time for monkeypox?

Exponential growth on arbitrary time scale

$$I_t = I_t e^{(R_0-1)(\gamma+\mu)t}$$

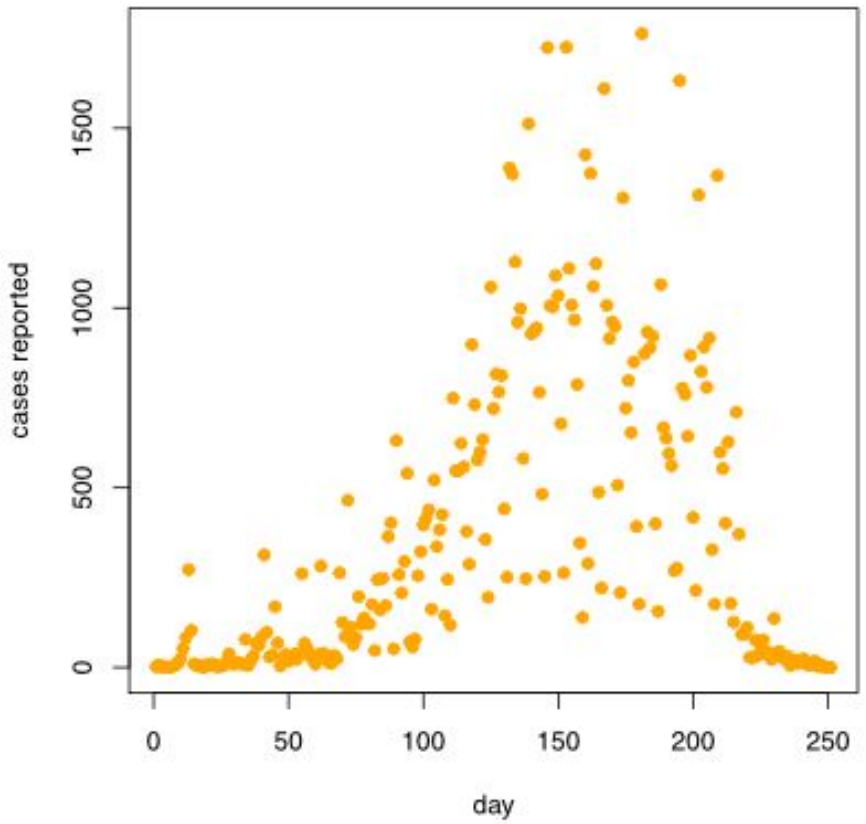
$$\ln(Y_t) = \ln(Y_0) + (R_0 - 1)(\gamma + \mu)t$$

γ is the recovery rate ($1/\gamma$ is mean duration of infection)
μ is non-disease mortality rate, which can be ignored if dynamics are fast enough
Y_t is the number infected by day t

Fitting Time Series

- In the initial phase of an outbreak, the epidemic grows exponentially

2010 measles outbreak in Malawi



Exponential growth on arbitrary time scale

$$I_t = I_t e^{(R_0-1)(\gamma+\mu)t}$$

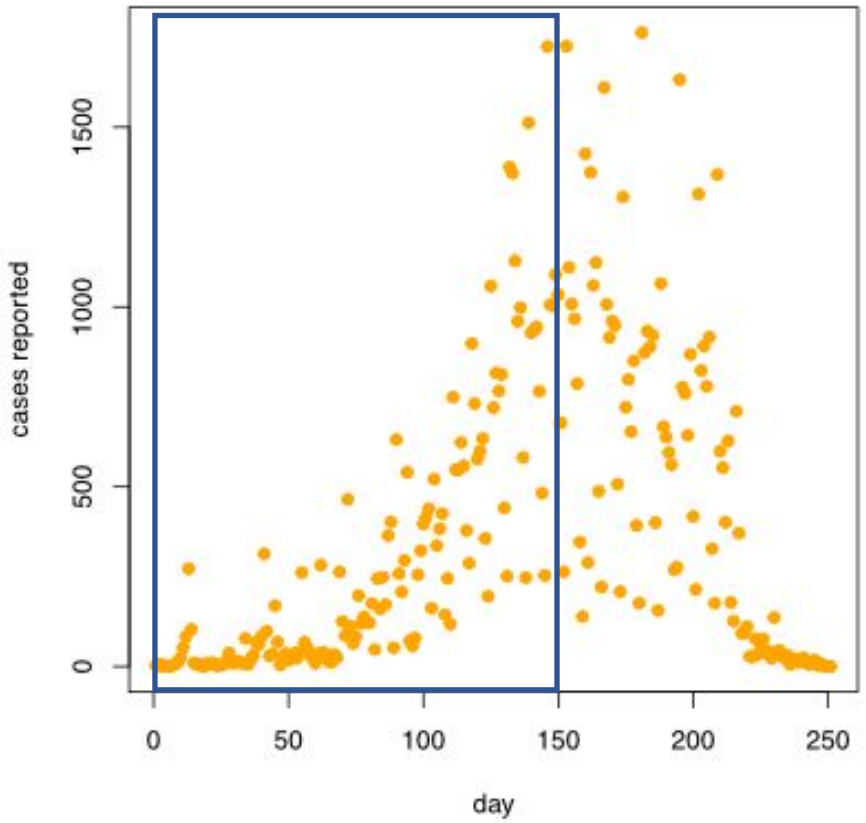
$$\ln(Y_t) = \ln(Y_0) + (R_0 - 1)(\gamma + \mu)t$$

γ is the recovery rate ($1/\gamma$ is mean duration of infection)
μ is non-disease mortality rate, which can be ignored if dynamics are fast enough
Y_t is the number infected by day t

Fitting Time Series

- In the initial phase of an outbreak, the epidemic grows exponentially

2010 measles outbreak in Malawi



Exponential growth on arbitrary time scale

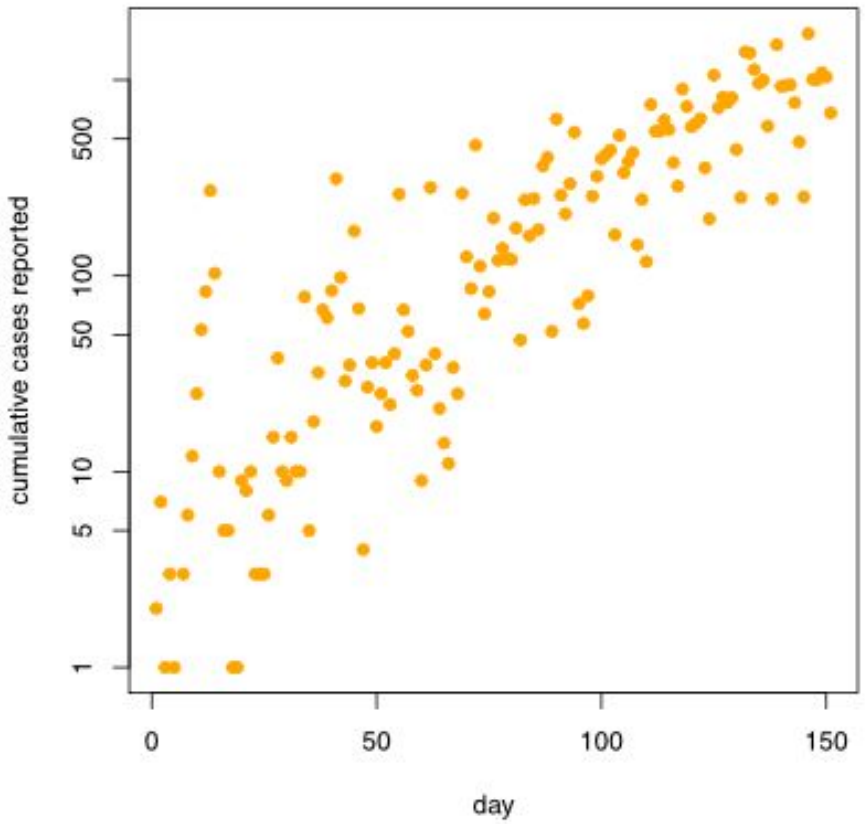
$$I_t = I_t e^{(R_0-1)(\gamma+\mu)t}$$

$$\ln(Y_t) = \ln(Y_0) + (R_0 - 1)(\gamma + \mu)t$$

γ is the recovery rate ($1/\gamma$ is mean duration of infection)
μ is non-disease mortality rate, which can be ignored if dynamics are fast enough
Y_t is the number infected by day t

Fitting Time Series

- In the initial phase of an outbreak, the epidemic grows exponentially



Exponential growth on arbitrary time scale

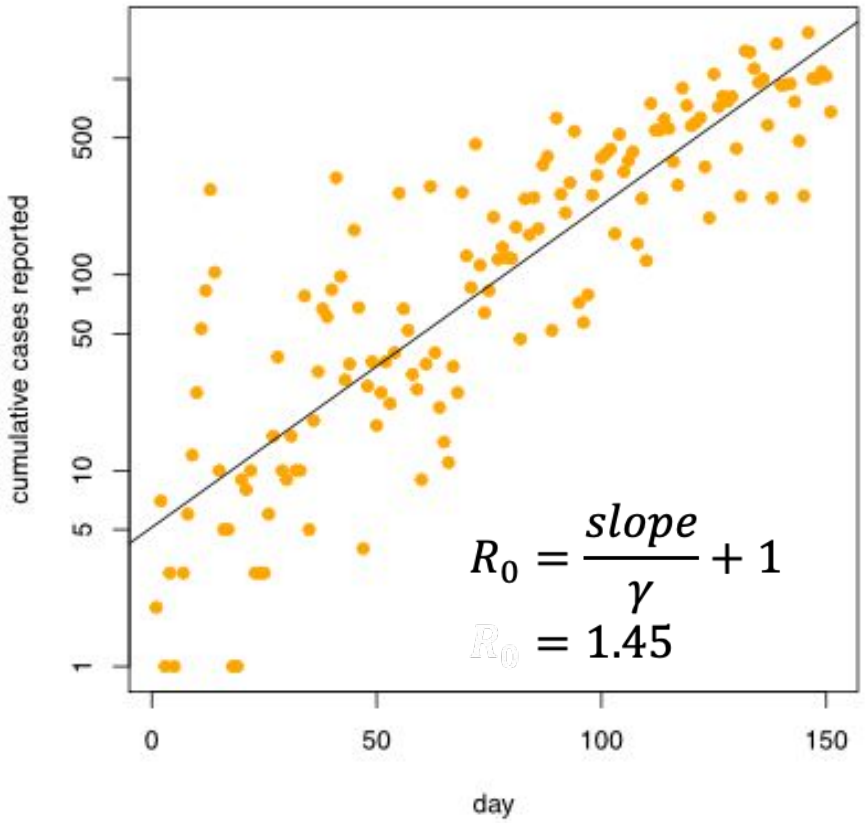
$$I_t = I_t e^{(R_0-1)(\gamma+\mu)t}$$

$$\ln(Y_t) = \ln(Y_0) + (R_0 - 1)(\gamma + \mu)t$$

γ is the recovery rate ($1/\gamma$ is mean duration of infection)
μ is non-disease mortality rate, which can be ignored if dynamics are fast enough
Y_t is the number infected by day t

Fitting Time Series

- In the initial phase of an outbreak, the epidemic grows exponentially



Exponential growth on arbitrary time scale

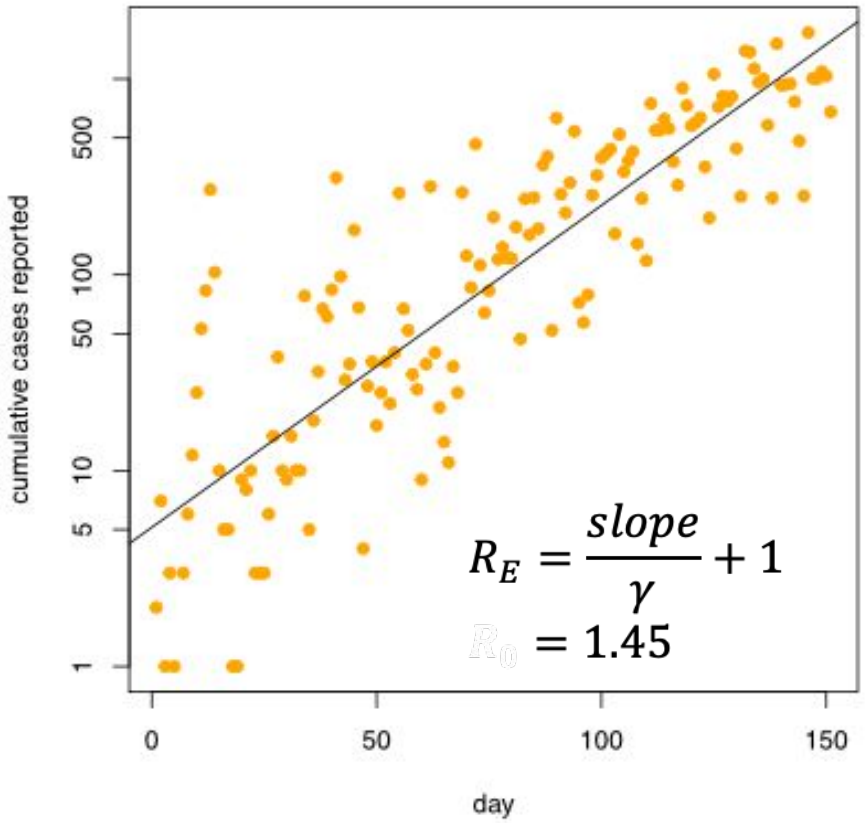
$$I_t = I_t e^{(R_0-1)(\gamma+\mu)t}$$

$$\ln(Y_t) = \ln(Y_0) + (R_0 - 1)(\gamma + \mu)t$$

γ is the recovery rate ($1/\gamma$ is mean duration of infection)
μ is non-disease mortality rate, which can be ignored if dynamics are fast enough
Y_t is the number infected by day t

Fitting Time Series

- In the initial phase of an outbreak, the epidemic grows exponentially



Exponential growth on arbitrary time scale

$$I_t = I_t e^{(R_0-1)(\gamma+\mu)t}$$

$$\ln(Y_t) = \ln(Y_0) + (R_0 - 1)(\gamma + \mu)t$$

γ is the recovery rate ($1/\gamma$ is mean duration of infection)
μ is non-disease mortality rate, which can be ignored if dynamics are fast enough
Y_t is the number infected by day t

Fitting Time Series

- In the initial phase of an outbreak, the epidemic grows exponentially

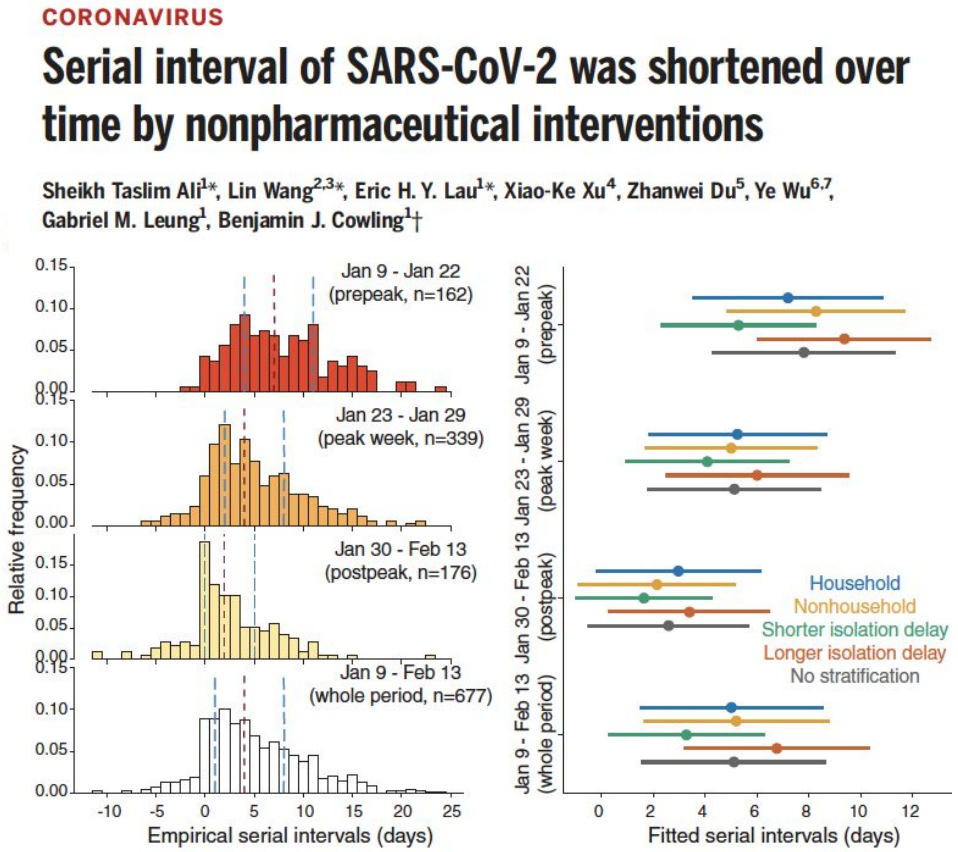


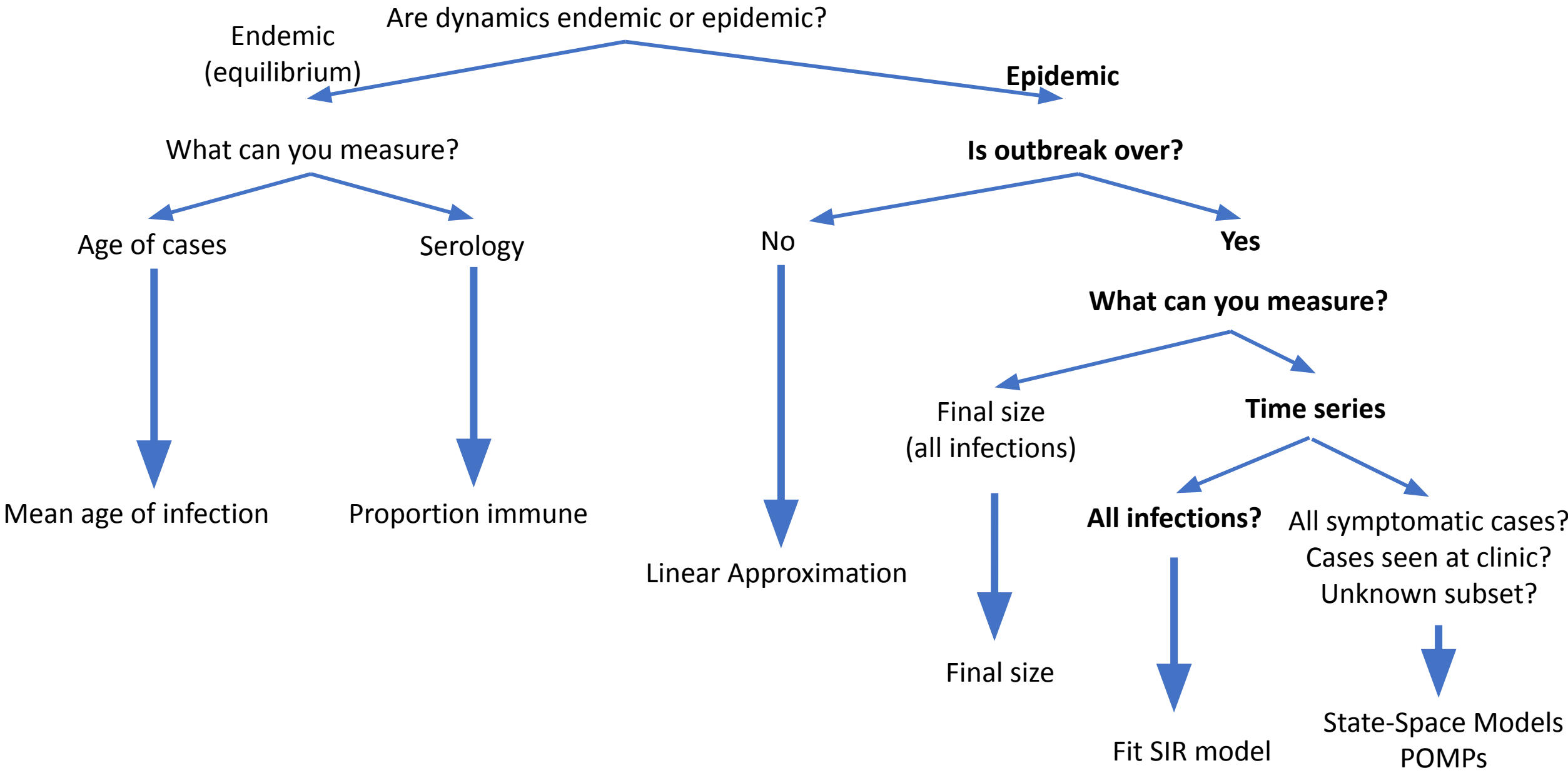
Fig. 1. Serial intervals of SARS-CoV-2 substantially shortened over time in mainland China. (A) Empirical

Exponential growth on arbitrary time scale

$$I_t = I_t e^{(R_0-1)(\gamma+\mu)t}$$

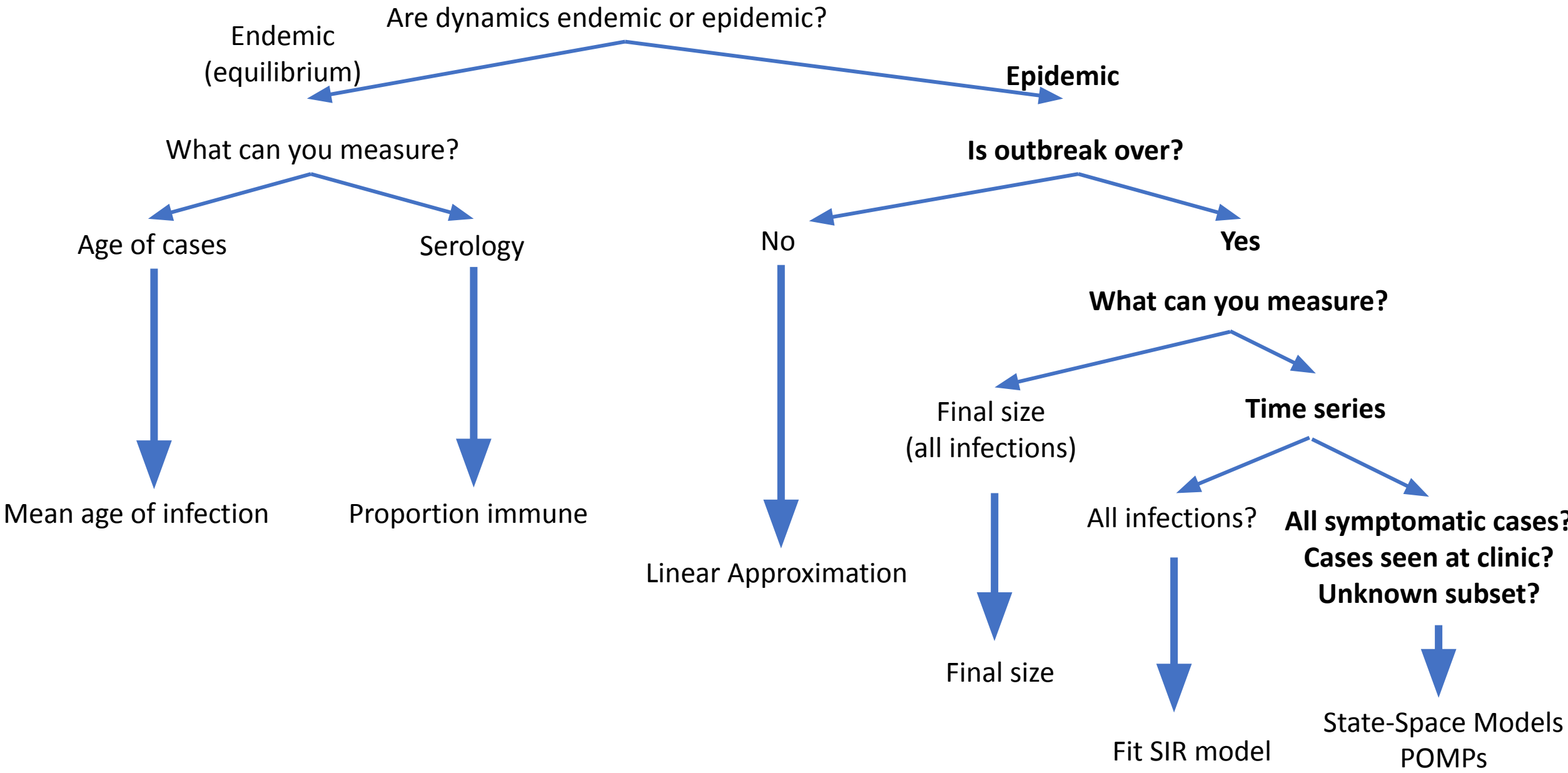
$$\ln(Y_t) = \ln(Y_0) + (R_0 - 1)(\gamma + \mu)t$$

γ is the recovery rate ($1/\gamma$ is mean duration of infection)
μ is non-disease mortality rate, which can be ignored if dynamics are fast enough
Y_t is the number infected by day t


Are dynamics endemic or epidemic?
Endemic (equilibrium)
Epidemic
What can you measure?
Age of cases
Serology
Mean age of infection
Proportion immune
Is outbreak over?
No
Yes
Linear Approximation
What can you measure?
Final size (all infections)
Time series
Final size
All infections?
All symptomatic cases?
Cases seen at clinic?
Unknown subset?
Fit SIR model
State-Space Models
POMPs

See R Worksheet

- What should be the measure of distance?
  - Trajectory matching?
  - Likelihood?
  - Something else?


Are dynamics endemic or epidemic?
Endemic (equilibrium)
Epidemic
What can you measure?
Age of cases
Serology
Mean age of infection
Proportion immune
Is outbreak over?
No
Yes
Linear Approximation
What can you measure?
Final size (all infections)
Time series
Final size
All infections?
All symptomatic cases?
Cases seen at clinic?
Unknown subset?
Fit SIR model
State-Space Models
POMPs

Fitting Real Time Series: Measurement Error

- Frequently we can only see a part of the time series, or the time series is obscured:
  - Under-reporting
  - Diagnostic uncertainty

Globally Reported: 147,000 measles cases
Globally Estimated: 7.5 million measles cases

Morbidity and Mortality Weekly Report (*MMWR*)

CDC

Progress Toward Regional Measles Elimination — Worldwide, 2000–2020

Weekly / November 12, 2021 / 70(45);1563–1569

Meredith G. Dixon, MD[1]; Matt Ferrari, PhD[2]; Sebastien Antoni, MPH[3]; Xi Li, MD[1]; Allison Portnoy, ScD[4]; Brian Lambert[2]; Sarah Hauryski[2]; Cynthia Hatcher, MPH[1]; Yoann Nedelec, MPH[3]; Minal Patel, MD[1,3]; James P. Alexander Jr., MD[1]; Claudia Steulet[3]; Marta Gacic-Dobo, MSc[3]; Paul A. Rota, PhD[5]; Mick N. Mulders, PhD[3]; Anindya S. Bose, MD[3]; Alexander Rosewell, PhD[3]; Katrina Kretsinger, MD[1]; Natasha S. Crowcroft, MD[3] (View author affiliations)

Study: US COVID cases, deaths far higher than reported

Filed Under: COVID-19
Mary Van Beusekom | News Writer | CIDRAP News | Jan 05, 2021 Share Tweet LinkedIn Email Print & PDF

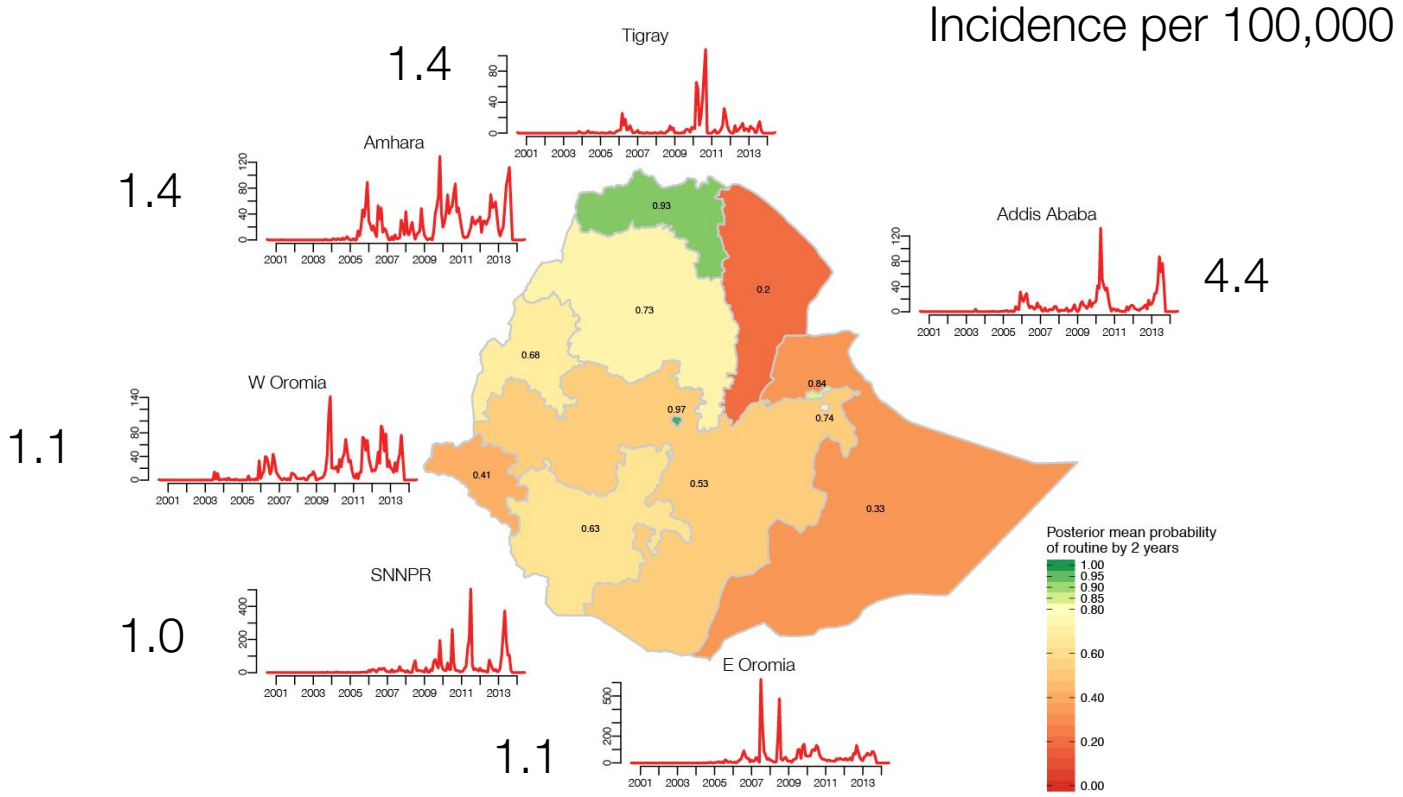
An estimated 14.3% of the US population had antibodies against COVID-19 by mid-November 2020, suggesting that that the virus has infected vastly more people than reported—but still not enough to come close to the proportion needed for herd immunity, according to a study published today in *JAMA Network Open*.

Government of Alberta, Chris Schwarz / Flickr cc

In the cross-sectional study, researchers from study sponsors Pfizer and Merck analyzed data from random community seroprevalence surveys and five such regional and national Centers for Disease Control and Prevention (CDC) surveys to estimate infection underreporting multipliers. Seroprevalence surveys reveal the proportion of a population that has antibodies against a certain disease, such as COVID-19.

After adjusting for underreporting using validated multipliers, the analysis revealed an estimated median 46,910,006 infections with SARS-CoV-2, the virus

Measles Incidence in Ethiopia

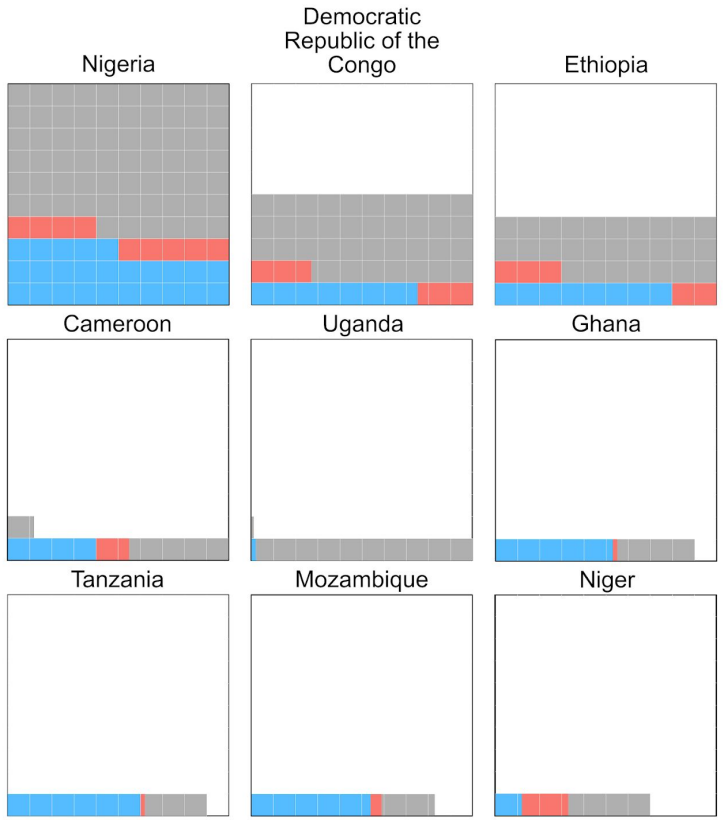


Fitting Real Time Series: Measurement Error

- Frequently we can only see a part of the time series, or the time series is obscured:
  - Under-reporting
  - Diagnostic uncertainty

Many infections cause similar syndromes (a collection of clinical symptoms):

Upper respiratory infections	-> influenza, COVID
Fever + rash	-> measles
Acute flaccid paralysis	-> polio
Acute diarrhea	-> cholera
Acute fever	-> malaria



Cases of fever and rash that match the measles clinical case definition

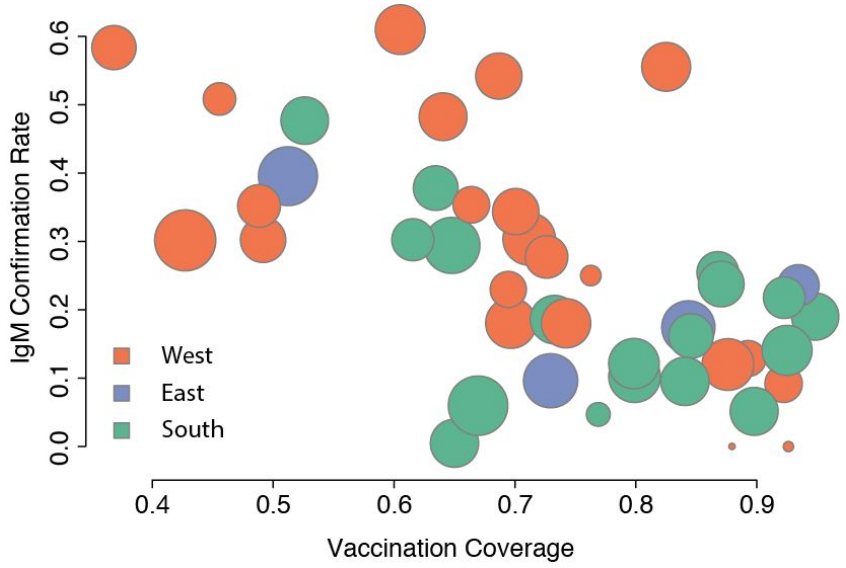
Untested, clinical cases

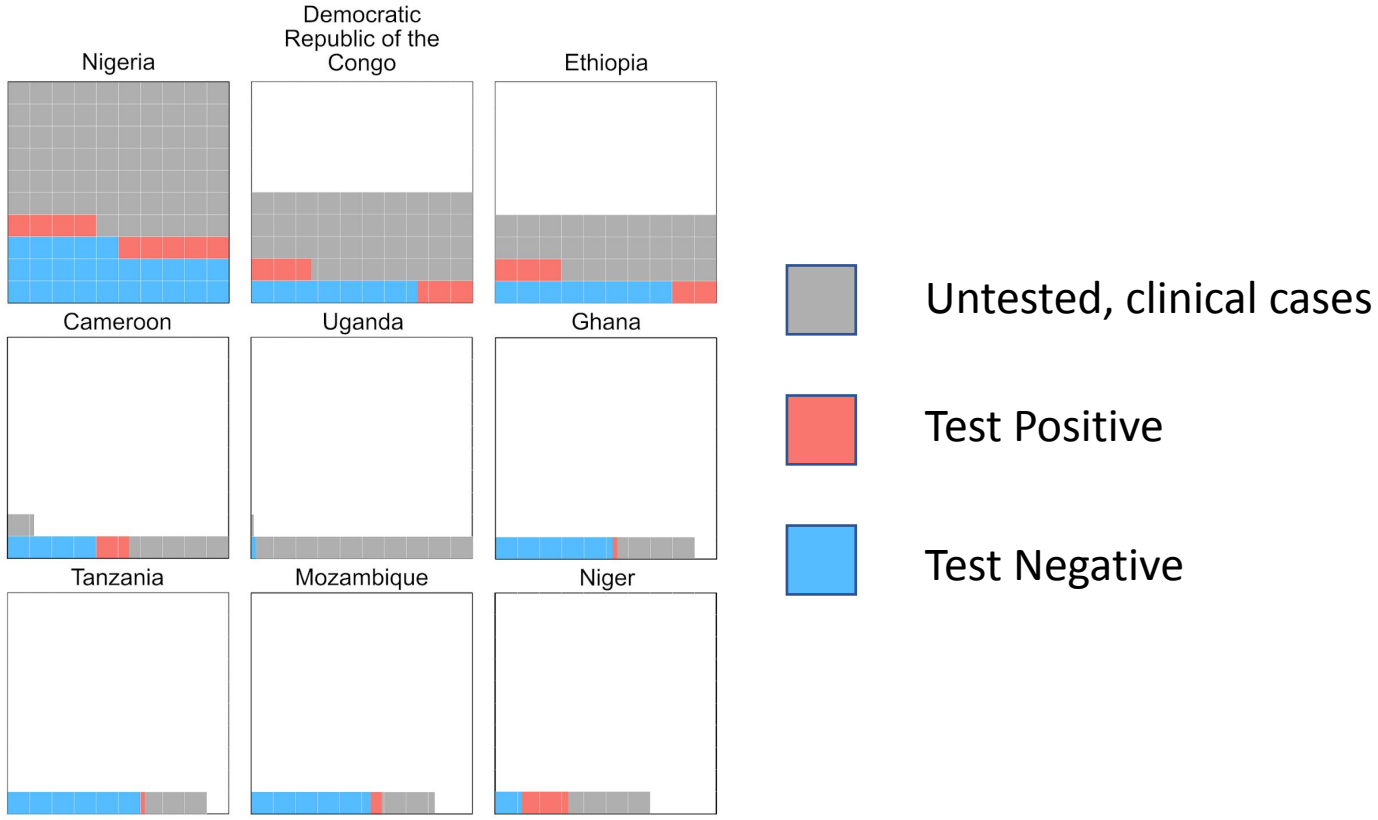
Test Positive

Test Negative

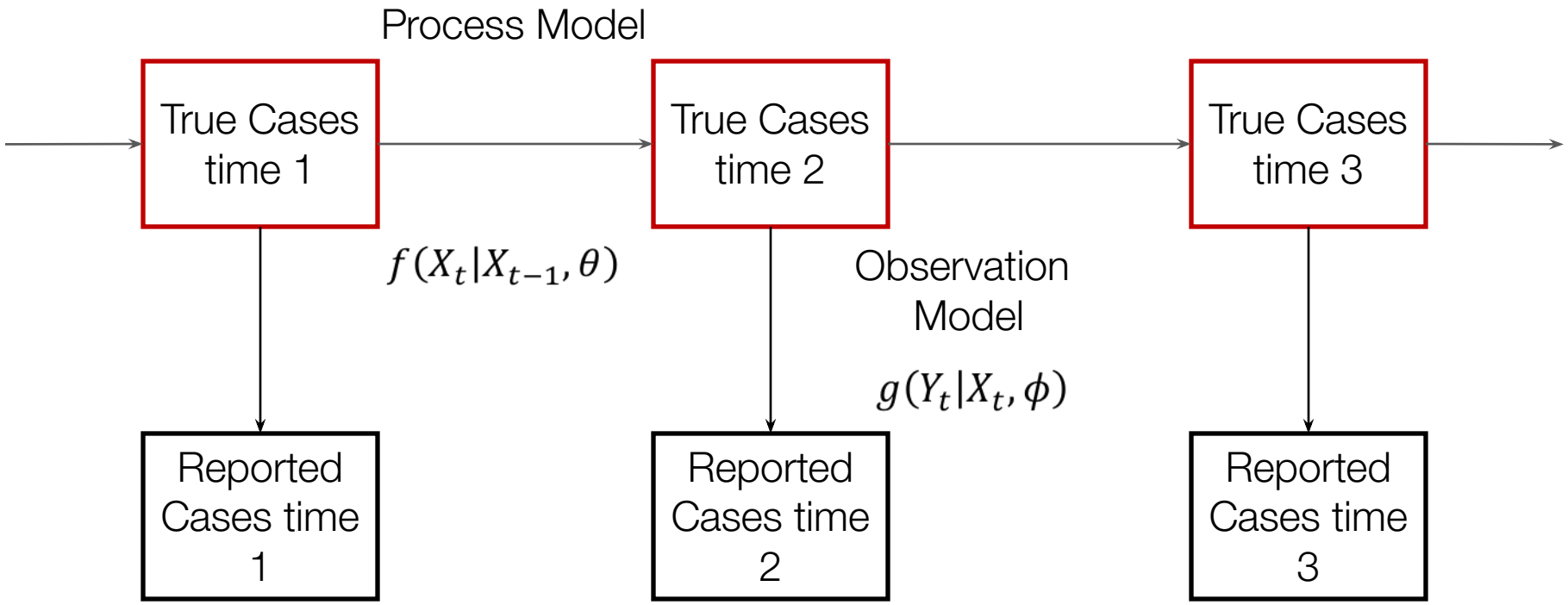
Fitting Real Time Series: Measurement Error

- Frequently we can only see a part of the time series, or the time series is obscured:
  - Under-reporting
  - Diagnostic uncertainty





We need TWO models

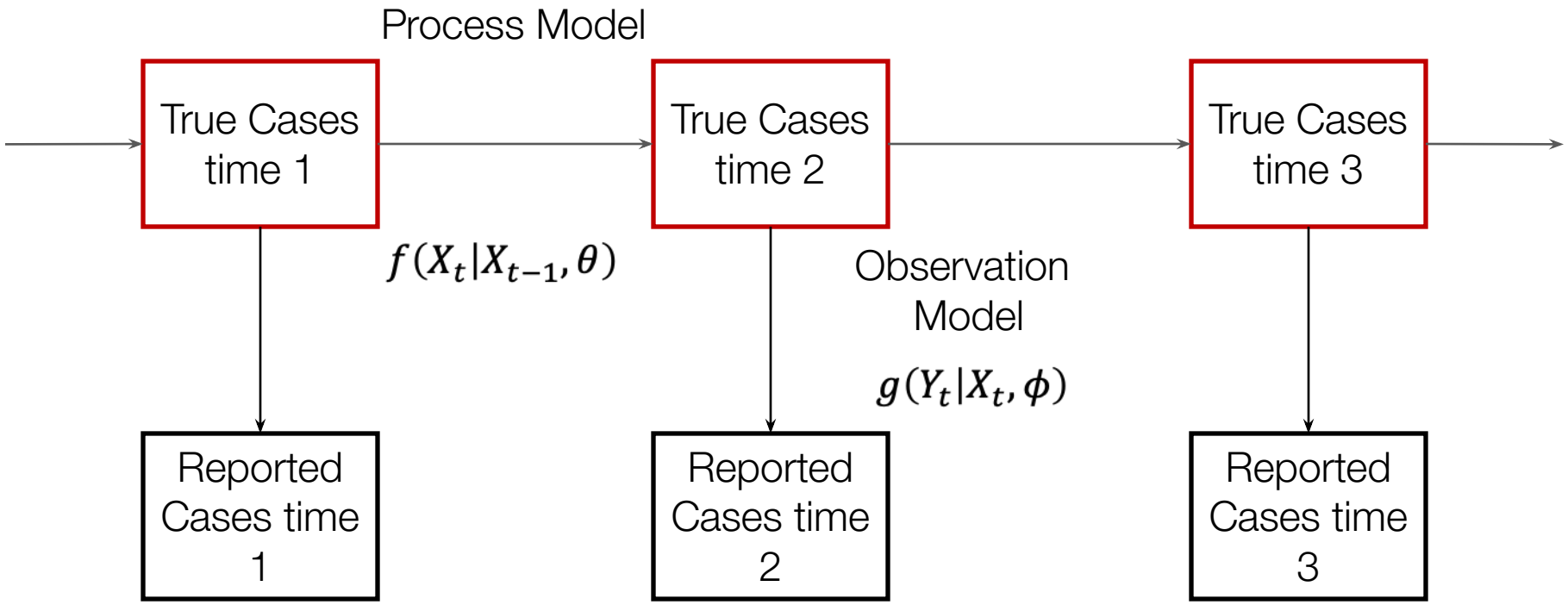


$f(Y_t|Y_{t-1},\theta,\phi)$

- can be stated as a function of these two models, unobserved states are latent variables
- long history in engineering, more recently in population dynamics

We need TWO models

Here we might have the additional goal of estimating the true states; i.e. the true burden of disease among those who were not measured.



$f(Y_t|Y_{t-1},\theta,\phi)$

- can be stated as a function of these two models, unobserved states are latent variables
- long history in engineering, more recently in population dynamics

Basic Recipe of Estimation

1. Make an observation of the world
2. Build a model that can replicate that observation
3. Define a measure of distance between observation and model
4. Search over many (all?) parameters to find the ones that minimize that distance

The basic recipe offers a systematic approach to parameter estimation. If you can simulate, you can estimate … even if it isn’t very efficient