

# Introduction of Infectious Disease Modelling

# Objectives

---

By the end of this session on **Intro to ID Modelling** you should learn:

- How does infectious disease modelling fit into the field of epidemiology
- What are the different goals of ID mechanistic models
- What are the different types of ID models
- What types of ID data are collected
- How are ID data and ID models related
- What does typical measles and rubella data look like

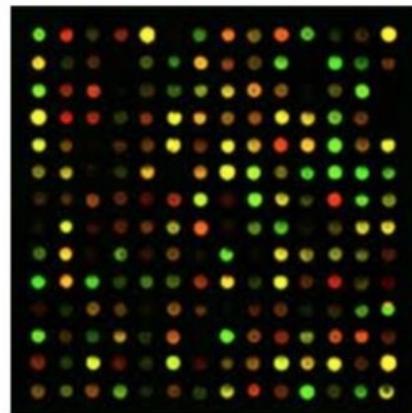
# Contextualizing Infectious Disease Modelling

# Many ways to study infectious diseases

Medicine



Genomics



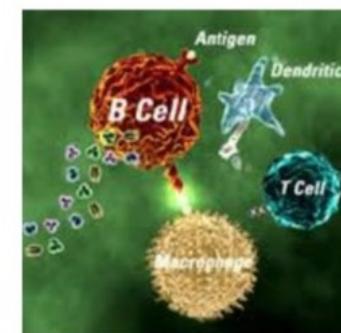
Microbiology



Some disciplines focus on infectious diseases from the:

- (1) individual level
- (2) within-host scale
- (3) microbe perspective

Immunology



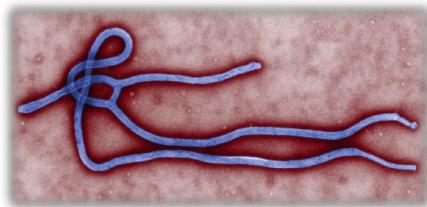
Vaccines & Drugs



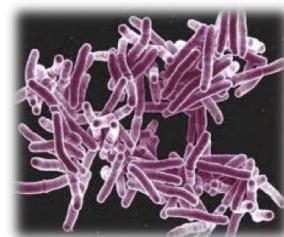
# Disease vs Infectious Disease

**Disease** - A deviation from the normal physiological status of an organism that negatively affects its survival or reproduction

**Infectious Disease** - A disease in one organism (the host) that is caused by another organism (pathogen or parasite) which has entered the host's body



Ebola Virus



Tuberculosis Bacteria



HIV

Pathogen: Microorganism that causes disease  
(virus, bacteria, parasite)

# Agents of Infectious Diseases

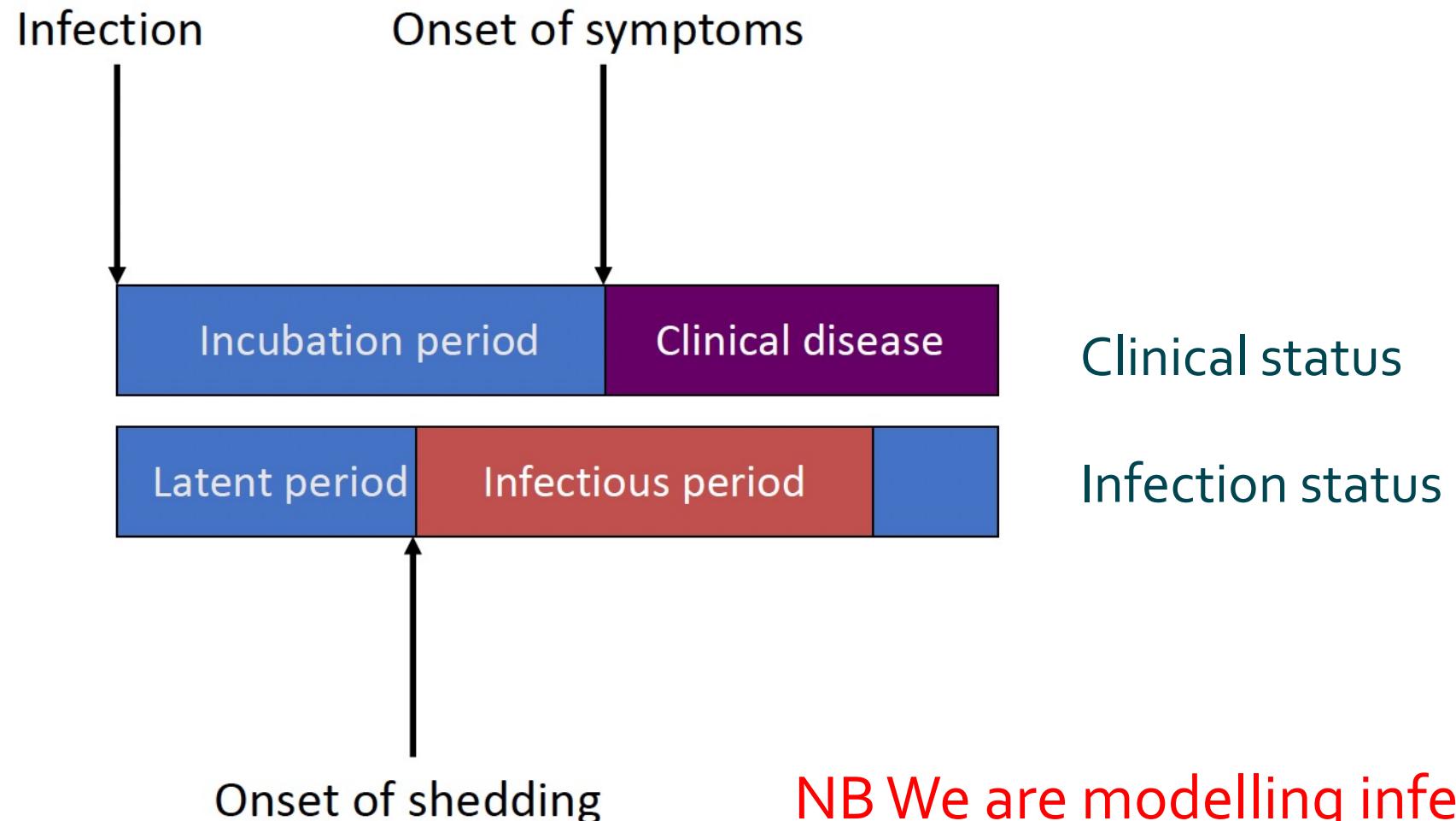
Infectious disease is a disease caused by a pathogen

**Pathogens** (agents) – organisms that are capable of producing diseases.

- **Viruses** (Examples: HIV -> AIDS, influenza -> flu, Measles morbillivirus -> measles)
- **Bacteria** (Examples: Vibrio cholerae -> Cholera, Yersinia pestis -> plague, Mycobacterium tuberculosis -> Tuberculosis)
- **Fungi** (Examples: Aspergillus -> Aspergillosis, tinea -> Athlete's foot)
- **Protozoa** (Examples: Plasmodium falciparum -> Malaria, Trypanosoma cruzi -> Chagas Disease)
- **Helminths** (Examples: Schistosoma mansoni -> Schistosomiasis, Hookworm -> hookworm infection)

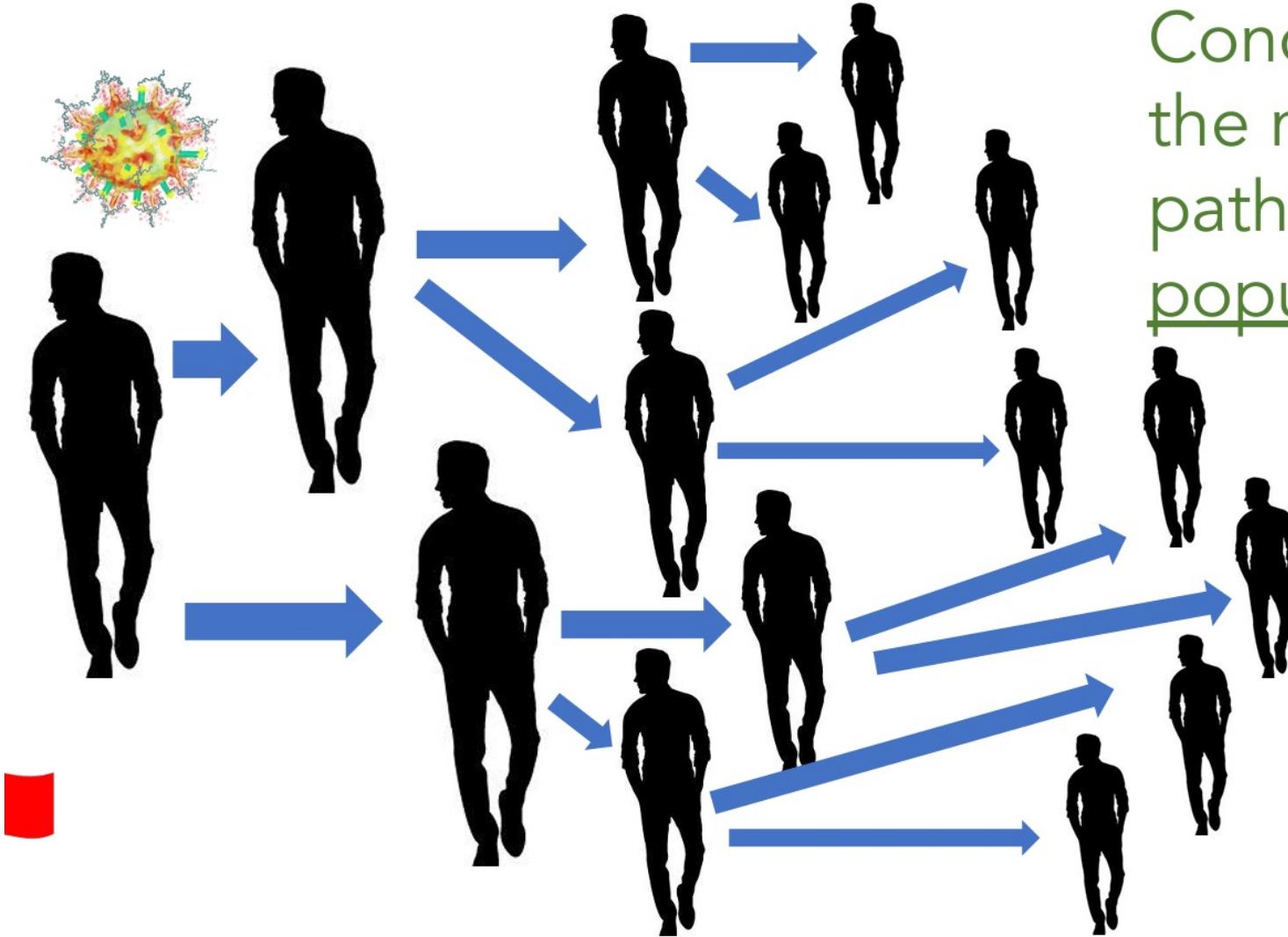
parasites

# From infection to disease



NB We are modelling infections, although we actually observe clinical cases

# Epidemiology Focus on Population Level

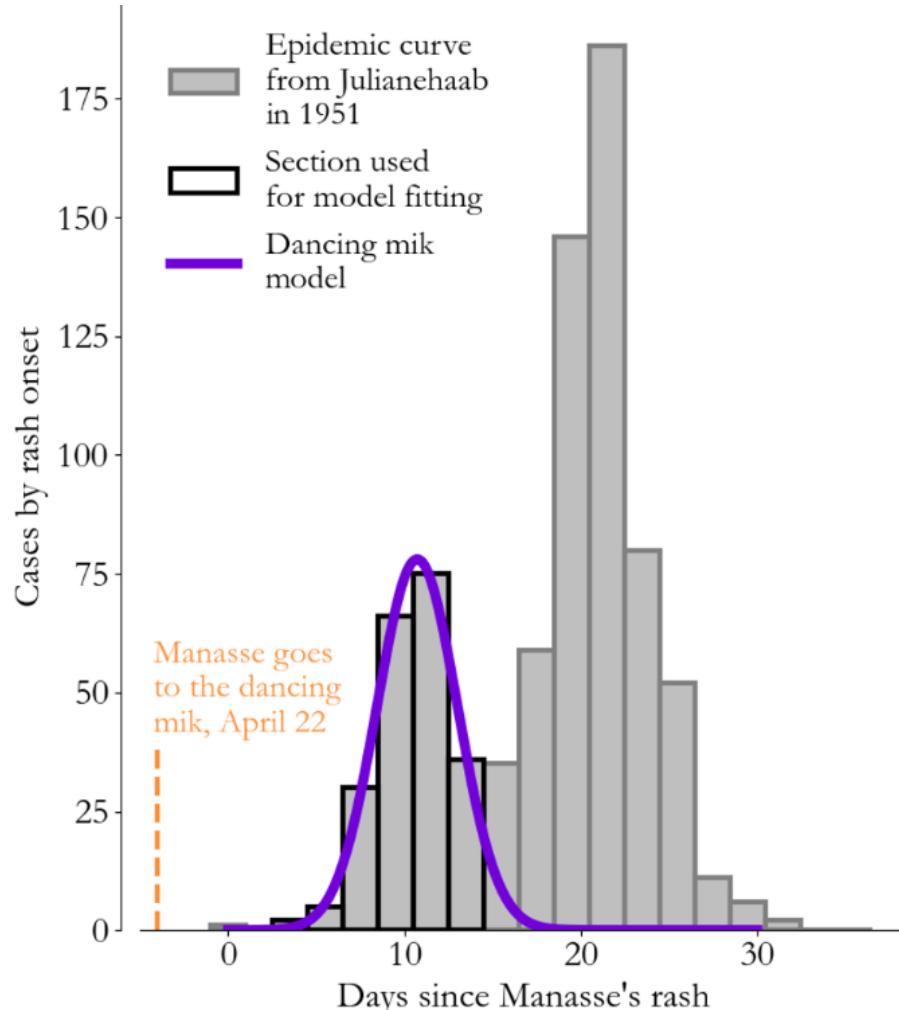


Concept: We can track the movement of pathogens throughout populations

A transmission chain is the set of infection events that occur as a pathogen moves through a population.

# The use of time series data

## Measles Epidemic from Greenland



### Epi Outbreak investigation:

- What pathogen is causing the illness?
- Is it a novel pathogen?
- Who is infected?

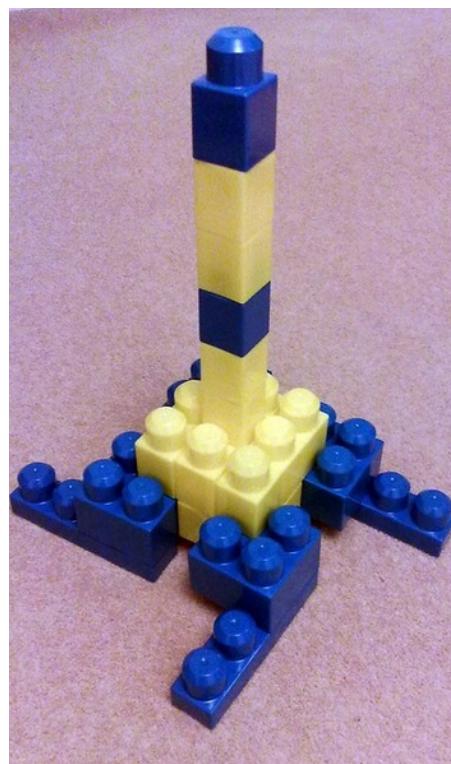
### Biological questions:

- What allows a pathogen to enter the population?
- What does the growth rate tell us?
- Why does the epidemic turn over?
- Why are there two peaks?

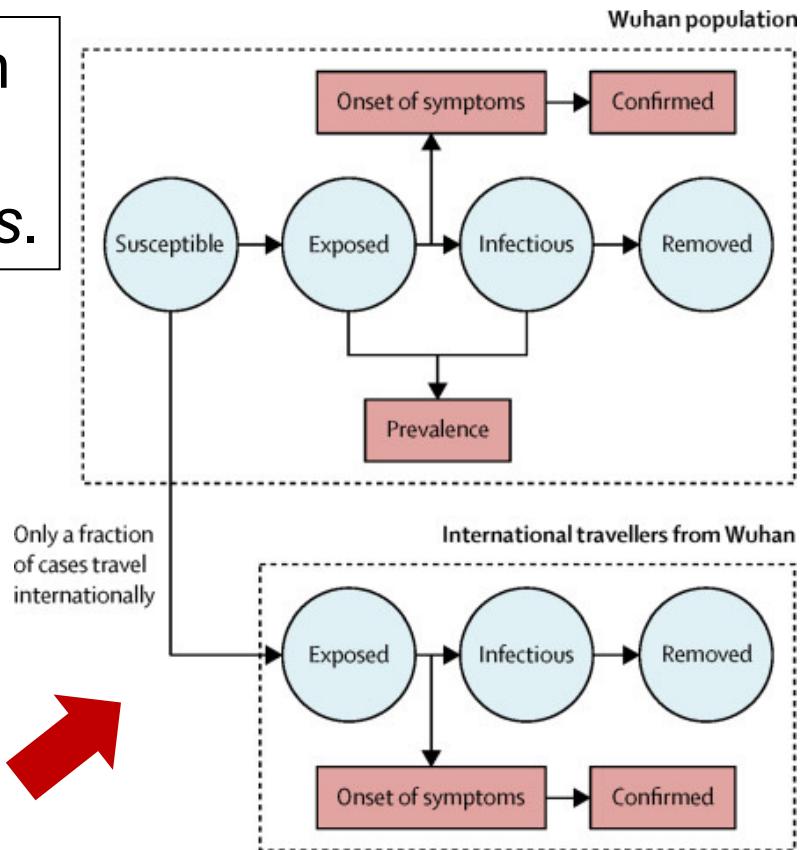
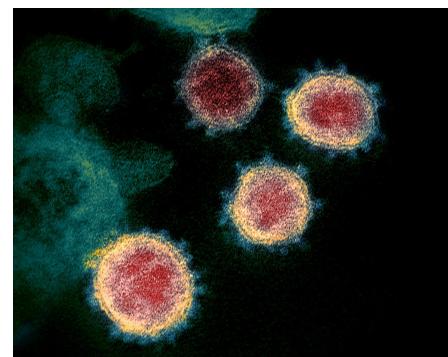
### Intervention questions:

- How to prevent spread?
- When is best to implement control?
- Drugs, vaccines, or other control measures?

# What is a model?

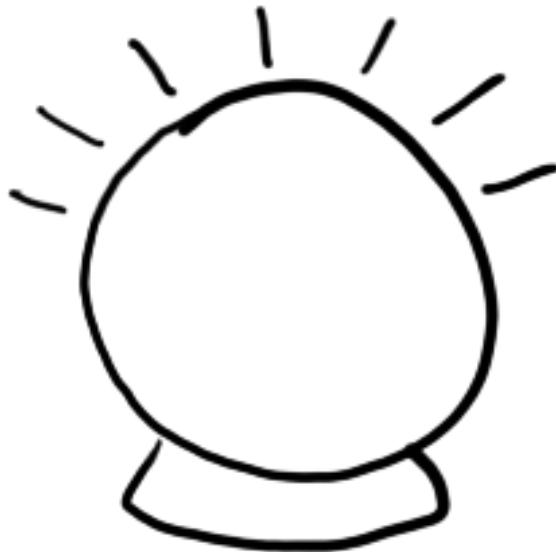


A model is a simplified representation  
of a more complex object/process,  
*designed to address specific questions.*



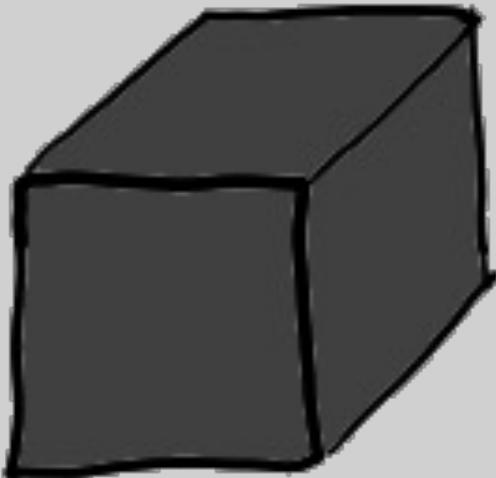
# What model is not ...

A model is **not** a crystal ball.



(it can't create information out of nothing.)

A model is **not** a black box.



(it is only useful when you understand the assumptions and use it to guide thinking, not replace it.)

A model is **not** a perfect representation.  
(the map is not the territory)

“Only *six inches!*” exclaimed Mein Herr. “We very soon got to *six yards* to the mile. Then we tried a *hundred* yards to the mile. And then came the grandest idea of all! We actually made a map of the country, on the scale of *a mile to the mile!*”

“Have you used it much?” I enquired. “It has never been spread out, yet,” said Mein Herr: “the farmers objected: they said it would cover the whole country, and shut out the sunlight! So we now use the country itself, as its own map, and I assure you it does nearly as well. Now let me ask you *another* question. What is the smallest *world* you would care to inhabit?”

Lewis Carroll. (1893) *Sylvie and Bruno Concluded*.

(it is only useful when it simplifies reality.)

# Infectious Disease Models are not new

**Daniel Bernoulli's 1766 analysis of smallpox** might be the first published model of an infectious disease, although not a transmission model. <https://doi.org/10.1002/rmv.443>



## The first counterfactual

- Bernoulli worked out solutions for a system of differential equations to determine life expectancy at birth with inoculation and without inoculation to smallpox

## Results

- Inoculating everyone at birth increased overall life expectancy by about 3 years
- Effective as long as the probability of dying from smallpox right after inoculation is less than ~ 11%, which Bernoulli estimated this risk to be ~1%



# Where does modelling fit in epidemiology?

Classical Epidemiology	Mechanistic Epidemiology
Data-Centric	Process-Centric
Public Health	Disease Ecology
Risk Factors	Infectious Disease Dynamics
Biostatistics	Mathematic Modelling

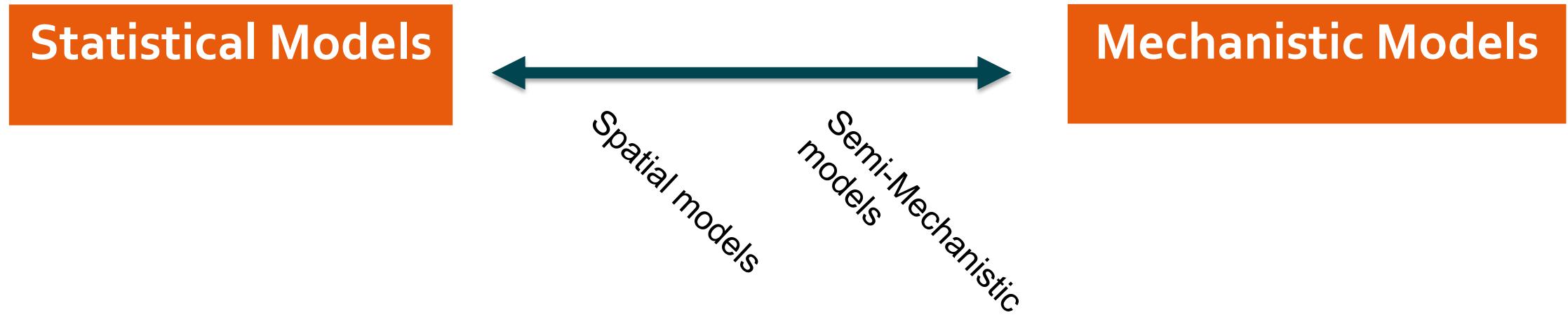
# Where does modelling fit in epidemiology?

(Bio)Statistical Epidemiology (data-centric)	Mechanistic Epidemiology (process-centric)
Account for bias and random error to find correlations that may imply causality	Systems Approach: Explicitly model multiple mechanisms to understand their interactions
Often the first step to assessing relationships	Links observed relationship at different scales
Assume independence of individuals (as some scale)	Explicitly focus on dependence of individuals

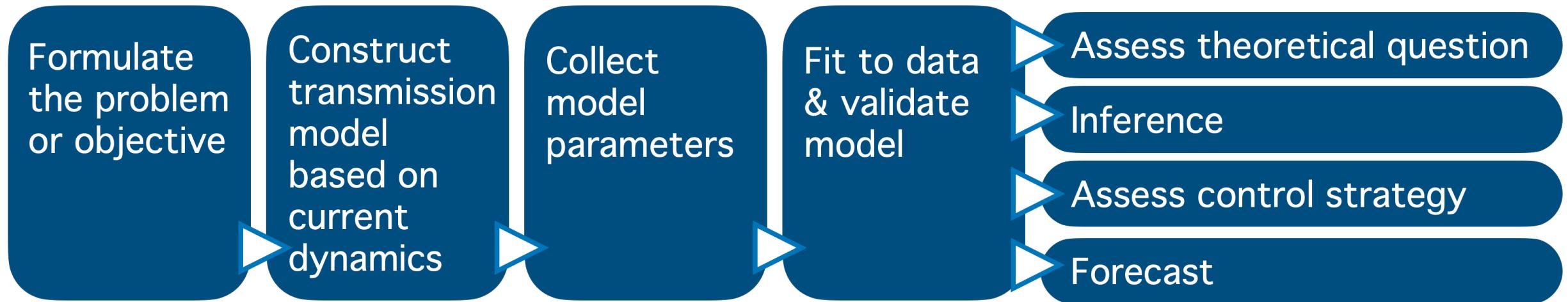
# Where does modelling fit in epidemiology?

Statistical Models	Mechanistic Models
Is HIV positively associated with the risk of TB infection?	Based on increased TB due to HIV, how much should we expect increase in TB to increase given HIV prevalence?
Are insecticide treated bednets (ITN) or indoor residual spraying (IRS) more effective for controlling malaria?	How do we expect the spatial distribution of malaria incidence to change after implementing ITB or IRS?
What are risk factors for dying from measles infection?	What is the impact of vaccination on the age profile of measles infection and deaths?

# Where does modelling fit in epidemiology?



# Steps in Developing a Model



# Model Utility and Types

- Choice of model depends crucially on focal question and available data
- Models are a tools and typically several types of models can be deployed for any given disease system
- Models are used principally for understanding nature or making projections under various scenarios

Formulate  
the problem  
or objective

# Model Utility

- We always develop models based on a specific goal or objective
- When setting model goals with public health practitioners, it is **critical** to set and communicate realistic expectations about what the model can and cannot do

# Four Goals for Mechanistic ID Models

Formulate  
the problem  
or objective

## Theoretical Modelling

*What type of emergence behavior is produced by disease systems with different properties?*

- What-if type questions
- Highly abstract
- Explore consequences of hypothetical mechanisms

# Theoretical Model Example

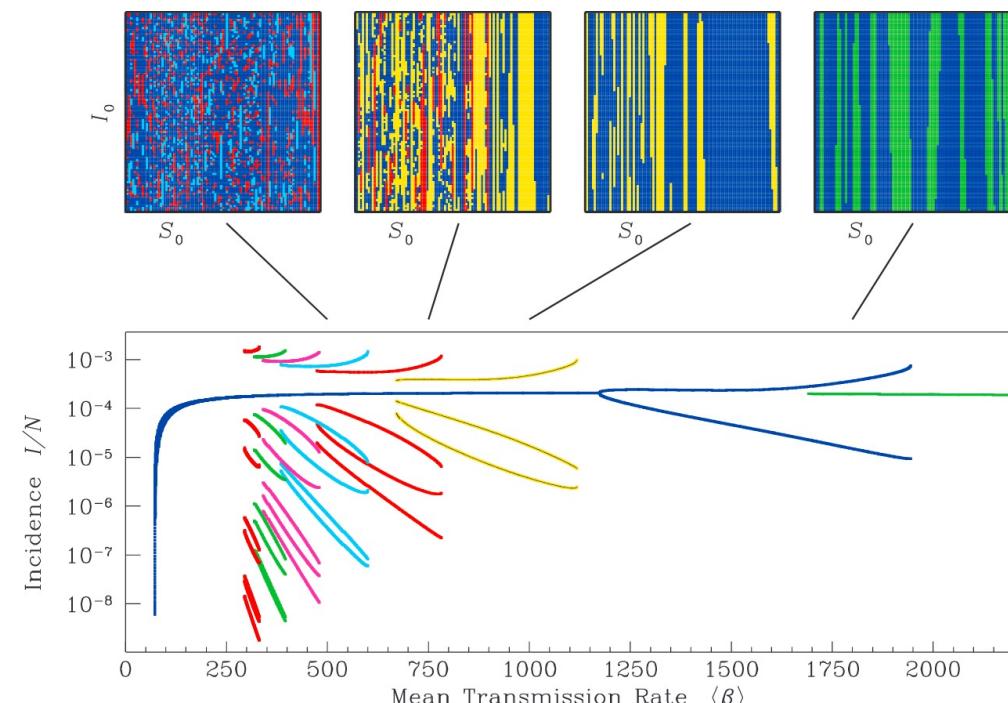
Earn et al. investigated the causes of transitions in measles epidemic patterns from regular (i.e., annual or biennial) to irregular outbreaks. The assumption under-investigation was the dynamical effect of changing birth rates and vaccination rates (i.e., changes in transmission) on incidence given term-time forcing. As a result, the authors displayed a bifurcation diagram looking across multiple transmission rates on incidence patterns and found more stochastic dynamics at lower transmission rates. [[10.1126/science.287.5453.667](https://doi.org/10.1126/science.287.5453.667)]

REPORTS

## A Simple Model for Complex Dynamical Transitions in Epidemics

David J. D. Earn,<sup>1,2\*</sup> Pejman Rohani,<sup>2</sup> Benjamin M. Bolker,<sup>3</sup>  
Bryan T. Grenfell<sup>2</sup>

Dramatic changes in patterns of epidemics have been observed throughout this century. For childhood infectious diseases such as measles, the major transitions are between regular cycles and irregular, possibly chaotic epidemics, and from regionally synchronized oscillations to complex, spatially incoherent epidemics. A simple model can explain both kinds of transitions as the consequences of changes in birth and vaccination rates. Measles is a natural ecological system that exhibits different dynamical transitions at different times and places, yet all of these transitions can be predicted as bifurcations of a single nonlinear model.



# Four Goals for Mechanistic ID Models

Formulate  
the problem  
or objective

## Theoretical Modelling

*What type of emergence behavior is produced by disease systems with different properties?*

- What-if type questions
- Highly abstract
- Explore consequences of hypothetical mechanisms

## Inference Modelling

*What is the true nature of the disease processes that are producing the observed health metrics?*

- Understand specific mechanisms of transmission
- Quantify value of specific parameters
- Account for epistemic and sampling process

# Inference Model Example

*Ecological Monographs*, 72(2), 2002, pp. 185–202  
© 2002 by the Ecological Society of America

## DYNAMICS OF MEASLES EPIDEMICS: SCALING NOISE, DETERMINISM, AND PREDICTABILITY WITH THE TSIR MODEL

May 2002

BRYAN T. GRENFELL,<sup>1,4</sup> OTTAR N. BJØRNSTAD,<sup>2</sup> AND BÁRBEL F. FINKENSTÄDT<sup>1,3</sup>

<sup>1</sup>Department of Zoology, University of Cambridge, Cambridge CB2 3EJ UK

<sup>2</sup>Departments of Entomology and Biology, 501 ASI Bldg., Pennsylvania State University,  
University Park, Pennsylvania 16802 USA

<sup>3</sup>Department of Statistics, University of Warwick, Coventry CV4 7AL UK

## SCALING OF MEASLES DYNAMICS

187

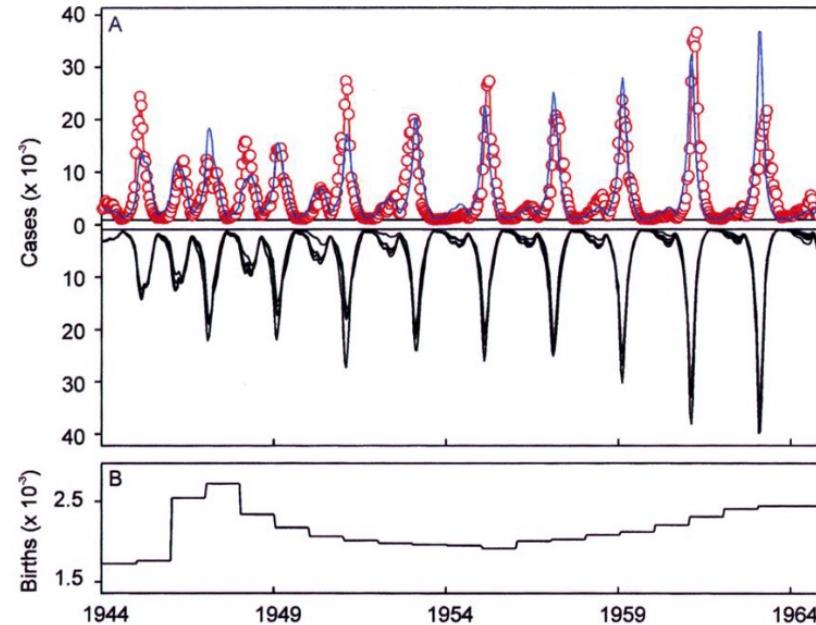


FIG. 1. (A) Measles incidence in 2-wk periods (in hundreds) in London from 1944 to 1965. The circles and the red line represent observed incidence (corrected for underreporting). The blue line represents the deterministic prediction from the TSIR model (using the susceptible and infected density in the first 2-wk period of 1944 as initial conditions). The black lines (and inverted scale) represent five stochastic realizations of the TSIR model. (B) The biweekly number of births (in hundreds) in London. The numbers are averaged within each year. The post-World War II baby boom in the late 1940s is associated with a period of annual cycles in measles incidence.

# Four Goals for Mechanistic ID Models

Formulate  
the problem  
or objective

## Theoretical Modelling

*What type of emergence behavior is produced by disease systems with different properties?*

- What-if type questions
- Highly abstract
- Explore consequences of hypothetical mechanisms

## Inference Modelling

*What is the true nature of the disease processes that are producing the observed health metrics?*

- Understand specific mechanisms of transmission
- Quantify value of specific parameters
- Account for epistemic and sampling process

## Strategic Modelling

*How will an epidemic unfold and different control strategies work under various conditions?*

- Conditional predictions of what could happen under specific scenarios
- *Focus on contrast's between scenarios*

# Strategic Modelling Example



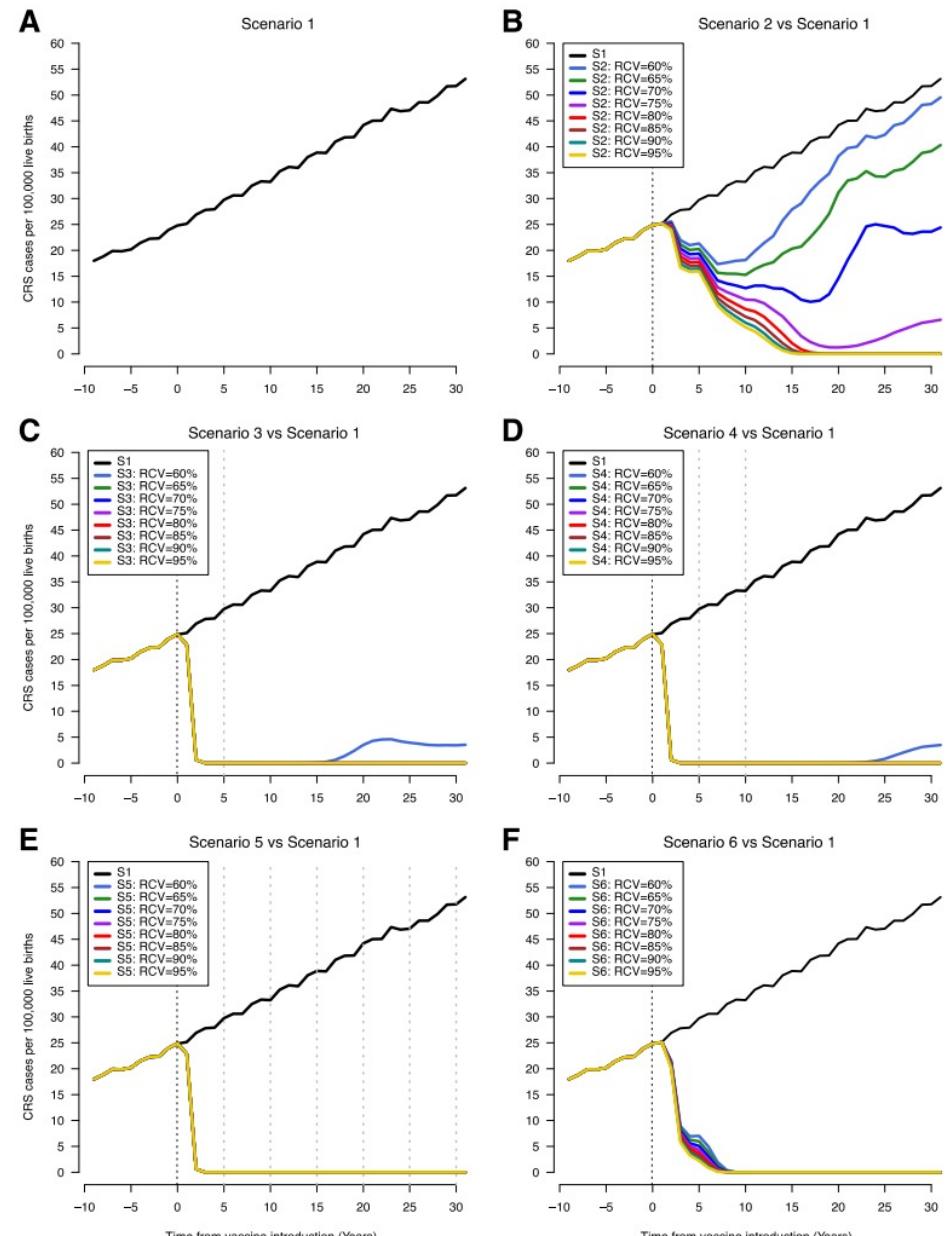
Article

## Rubella Vaccine Introduction in the South African Public Vaccination Schedule: Mathematical Modelling for Decision Making

Nkengafac Villyen Motaze <sup>1,2,\*</sup>, Ijeoma Edoka <sup>3</sup>, Charles S. Wiysonge <sup>2,4,5</sup>,

**Table 1.** Possible scenarios for rubella-containing vaccine (RCV) introduction in South Africa.

Scenario	Routine Vaccination in Expanded Program on Immunization (EPI)	Target Age Group for Routine Vaccination	Target Age Group for Initial Mass Campaign	Follow-Up Mass Campaigns	
				Target Age Group	Timing
1			No RCV in EPI		
2	RCV introduction	1 year	No initial campaign	No follow-up campaign	N/A
3	RCV introduction	1 year	1 to 14 years	No follow-up campaign	N/A
4	RCV introduction	1 year	1 to 14 years	1 to 4 years	One follow-up campaign 5 years after initial campaign
5	RCV introduction	1 year	1 to 14 years	1 to 4 years	Six follow-up campaigns every 5 years after initial campaign for 30 years
6	RCV introduction	1 year and 9 years	No initial campaign	No follow-up campaign	N/A



**Figure 2.** Time series of congenital rubella syndrome (CRS) incidence (CRS cases per 100,000 live births) showing scenario 1 (A) and comparing scenario 1 with scenarios 2–6 (B–F). The vertical black dotted

# Four Goals for Mechanistic ID Models

Formulate  
the problem  
or objective

## Theoretical Modelling

*What type of emergence behavior is produced by disease systems with different properties?*

- What-if type questions
- Highly abstract
- Explore consequences of hypothetical mechanisms

## Inference Modelling

*What is the true nature of the disease processes that are producing the observed health metrics?*

- Understand specific mechanisms of transmission
- Quantify value of specific parameters
- Account for epistemic and sampling process

## Strategic Modelling

*How will an epidemic unfold and different control strategies work under various conditions?*

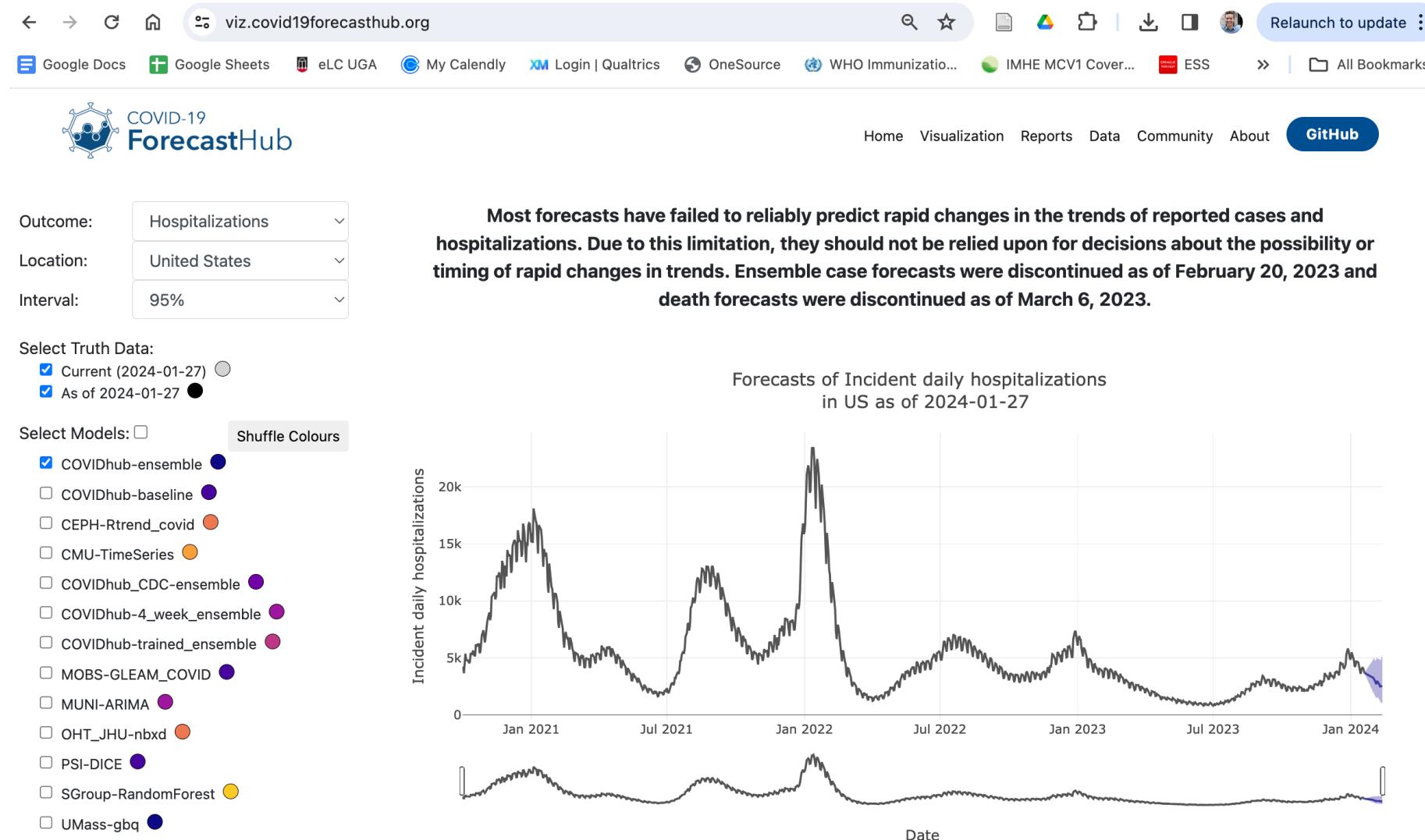
- Conditional predictions of what could happen under specific scenarios
- *Focus on contrast's between scenarios*

## Forecast Modelling

*How will an epidemic unfold in the coming weeks or months?*

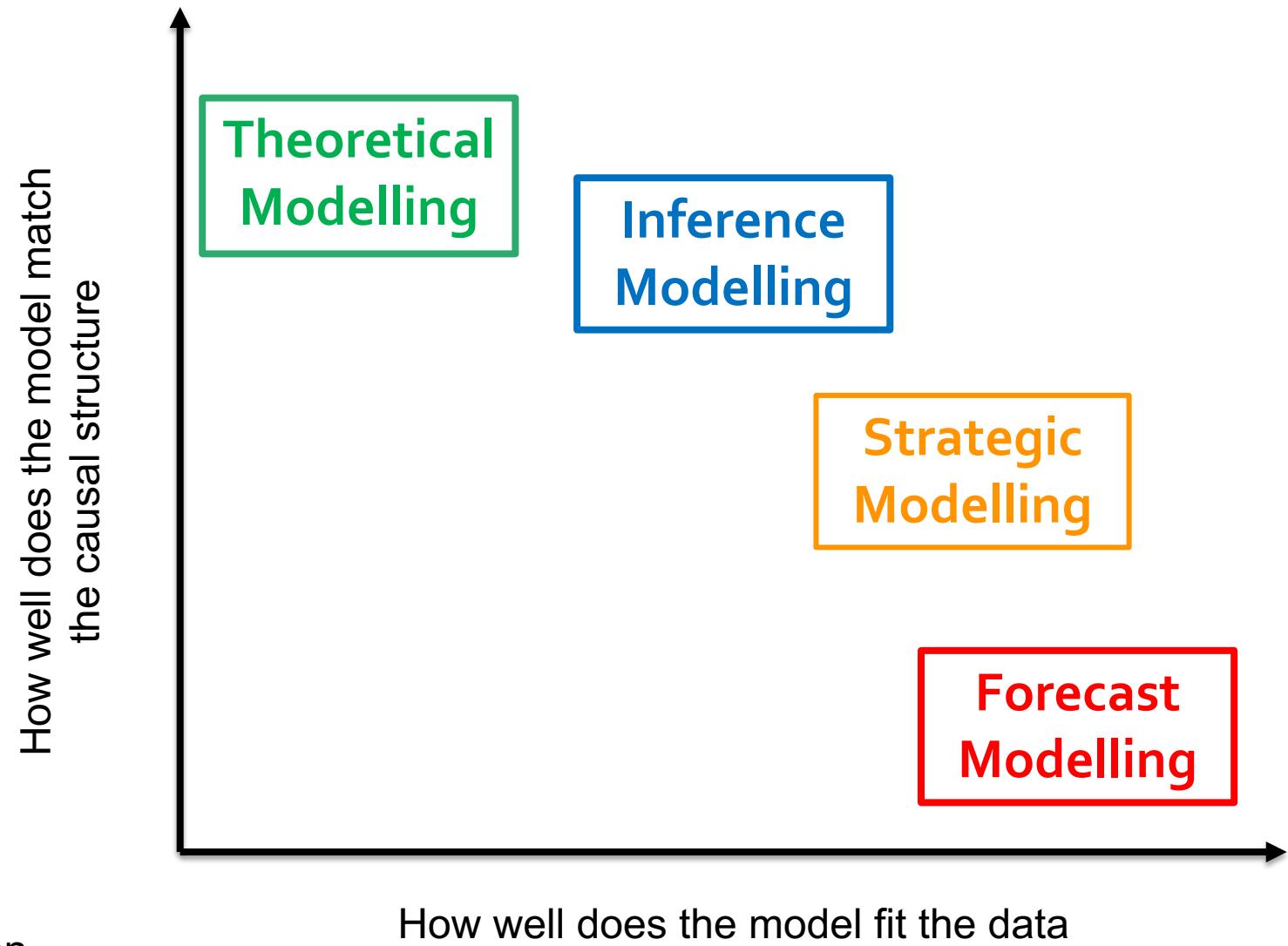
- Unconditional prediction of what will happen
- Choice of specific metrics

# Forecast Model Example



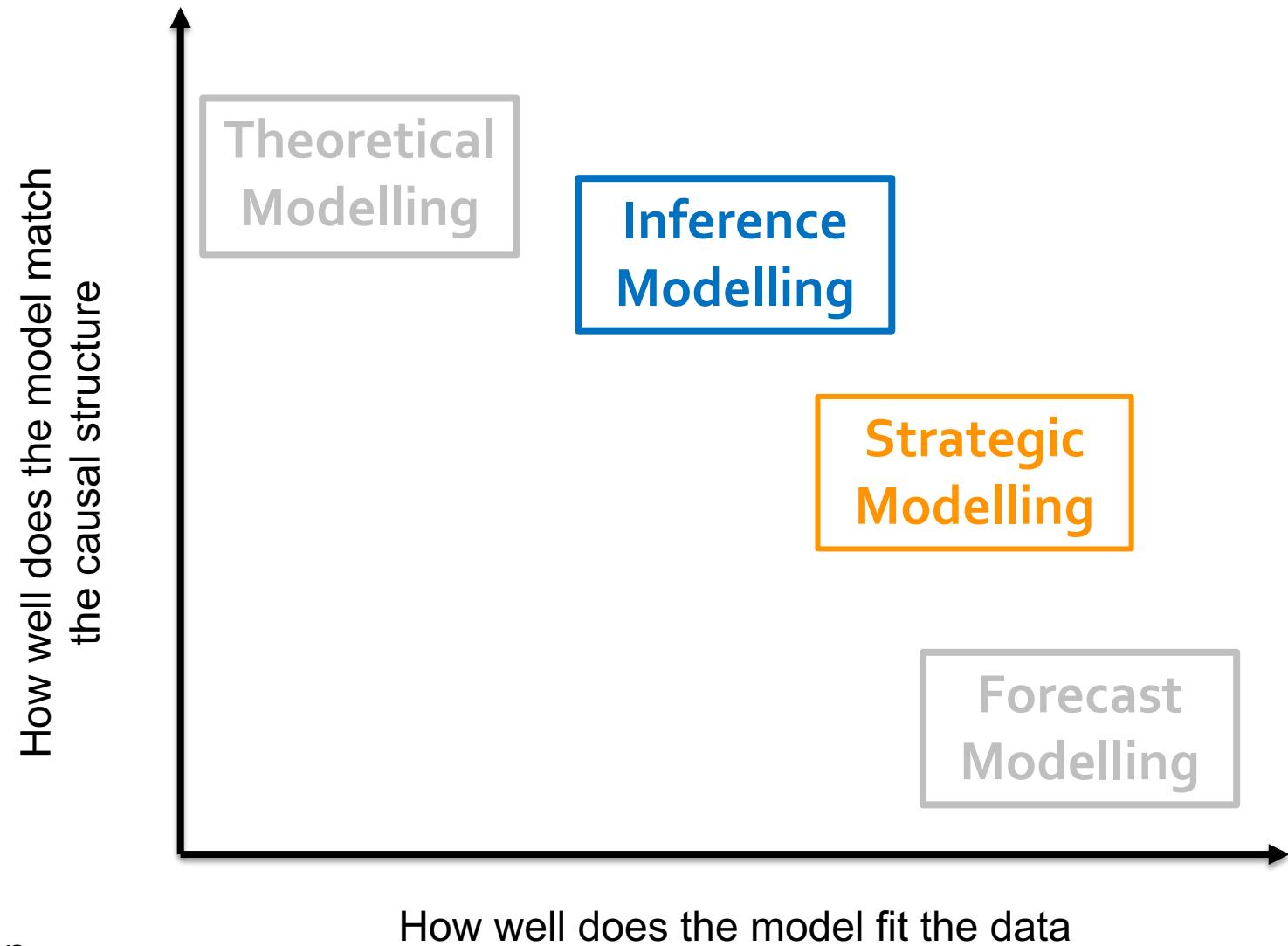
# Four Goals for Mechanistic ID Models

Formulate  
the problem  
or objective



# Four Goals for Mechanistic ID Models

Formulate  
the problem  
or objective



# Reporting Standards Based on Model Goals

## Theoretical Modelling

- Clear description of mechanistic mapping b/w model and phenomena
- Clear description of assumptions and model structure
- Evaluation of how non-target model components influence results

## Inference Modelling

- Clear description of how models are linked to data
- Description of sources of uncertainty
- Description of how threats to inference addressed (e.g., confounding)

## Strategic Modelling

- Clear description of how interventions are mechanistically modeled
- Definition and justification of scenarios
- Clear definition of what outcomes will be used to make contrasts b/w scenarios

## Forecast Modelling (Pollet et al 2021)

- Describe forecasting targets
- Define time horizon
- Describe how forecasts were validated

Construct transmission model based on current dynamics

# Scale of modelling

INTERFACE

[rsif.royalsocietypublishing.org](https://rsif.royalsocietypublishing.org)

Research



Cite this article: Klepac P, Megiddo I, Grenfell BT, Laxminarayan R. 2016 Self-enforcing regional vaccination agreements. *J. R. Soc. Interface* 13: 20150907. <http://dx.doi.org/10.1098/rsif.2015.0907>

Received: 22 October 2015

Accepted: 23 December 2015

## Self-enforcing regional vaccination agreements

Petra Klepac<sup>1</sup>, Itamar Megiddo<sup>2</sup>, Bryan T. Grenfell<sup>3,4,5</sup> and Ramanan Laxminarayan<sup>2,6</sup>

<sup>1</sup>Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, UK  
<sup>2</sup>Center for Disease Dynamics, Economics and Policy, Washington, DC 20036, USA  
<sup>3</sup>Ecology and Evolutionary Biology, and <sup>4</sup>Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, NJ 08544, USA  
<sup>5</sup>Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, USA  
<sup>6</sup>Public Health Foundation of India, New Delhi 110070, India

IM, 0000-0001-8391-6660

In a highly interconnected world, immunizing infections are a transboundary problem, and their control and elimination require international cooperation and coordination. In the absence of a global or regional body that can impose a universal vaccination strategy, each individual country sets its own strategy. Mobility of populations across borders can promote

Social, political, economic, health systems

Populations (“between host models”)

Individuals, physiological systems

Cells, genes, proteins (“within host models”)

## Estimating Progression Rates for Human Papillomavirus Infection From Epidemiological Data

Mark Jit, PhD, Nigel Gay, MSc, Kate Soldan, PhD, Yoon Hong Choi, PhD, William John Edmunds, PhD

A Markov model was constructed in order to estimate type-specific rates of cervical lesion progression and regression in women with high-risk human papillomavirus (HPV). The model was fitted to age- and type-specific data regarding the HPV DNA and cytological status of women undergoing cervical screening in a recent screening trial, as well as cervical cancer incidence. It incorporates different assumptions about the way lesions regress, the accuracy of cytological screening, the specificity of HPV DNA testing, and the age-specific prevalence of HPV infection. Combinations of assumptions generate 162 scenarios for squamous cell carcinomas and 54 scenarios for adenocarcinomas. Simulating an unscreened cohort of women infected with high-risk HPV indicates that the probability of an infection continuing to persist and to develop into invasive cancer depends on the length of time it has already persisted. The scenarios and parameter sets that produce the best fit to available epidemiological data provide a basis for modeling the natural history of HPV infection and disease. Key words: cervical intraepithelial neoplasms; mathematical model; Markov process; papillomavirus infections; uncertainty. (*Med Decis Making* 2010;30:84–98)

## The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study

Kiesha Prem\*, Yang Liu\*, Timothy W Russell, Adam J Kucharski, Rosalind M Eggo, Nicholas Davies, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group<sup>1</sup>, Mark Jit, Petra Klepac

### Summary

**Background** In December, 2019, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a novel coronavirus, emerged in Wuhan, China. Since then, the city of Wuhan has taken unprecedented measures in response to the outbreak, including extended school and workplace closures. We aimed to estimate the effects of physical distancing measures on the progression of the COVID-19 epidemic, hoping to provide some insights for the rest of the world.



*Lancet Public Health* 2020; 5:e264–70  
Published Online March 25, 2020  
[https://doi.org/10.1016/S2468-7667\(20\)30073-6](https://doi.org/10.1016/S2468-7667(20)30073-6)

## ARTICLES

<https://doi.org/10.1038/s41559-018-0786-x>

*nature ecology & evolution*

## Within-host dynamics shape antibiotic resistance in commensal bacteria

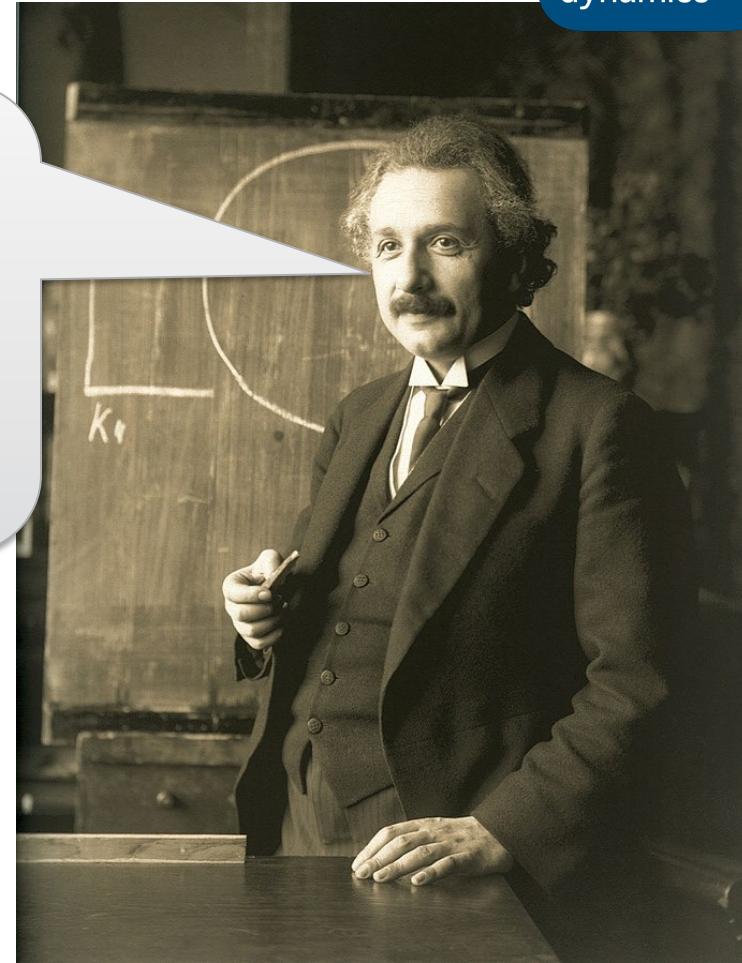
Nicholas G. Davies<sup>1,2\*</sup>, Stefan Flasche<sup>1,2</sup>, Mark Jit<sup>1,2,3</sup> and Katherine E. Atkins<sup>1,2,4</sup>

The spread of antibiotic resistance, a major threat to human health, is poorly understood. Simple population-level models of bacterial transmission predict that above a certain rate of antibiotic consumption in a population, resistant bacteria should completely eliminate non-resistant strains, while below this threshold they should be unable to persist at all. This prediction stands at odds with empirical evidence showing that resistant and non-resistant strains coexist stably over a wide range of antibiotic consumption rates. Not knowing what drives this long-term coexistence is a barrier to developing evidence-based strategies for managing the spread of resistance. Here, we argue that competition between resistant and sensitive pathogens within individual hosts gives resistant pathogens a relative fitness benefit when they are rare, promoting coexistence between strains at the population level. To test this hypothesis, we embed mechanistically explicit within-host dynamics in a structurally neutral pathogen transmission model. Doing so allows us to reproduce patterns of resistance observed in the opportunistic pathogens *Escherichia coli* and *Streptococcus pneumoniae* across European countries and to identify factors that may shape resistance evolution in bacteria by modulating the intensity and outcomes of within-host competition.

# What type of model will be most useful?

**“Selecting the correct level of detail is one of the most difficult decisions a modeler faces.** Models that are too simple may lose face validity because they do not incorporate aspects that content experts feel are required, but models that are too complex may be difficult to build, debug, analyze, understand, and communicate.”

*Everything  
should be as  
simple as  
possible, but no  
simpler.*



Einstein 1921 by F Schmutzler, public domain

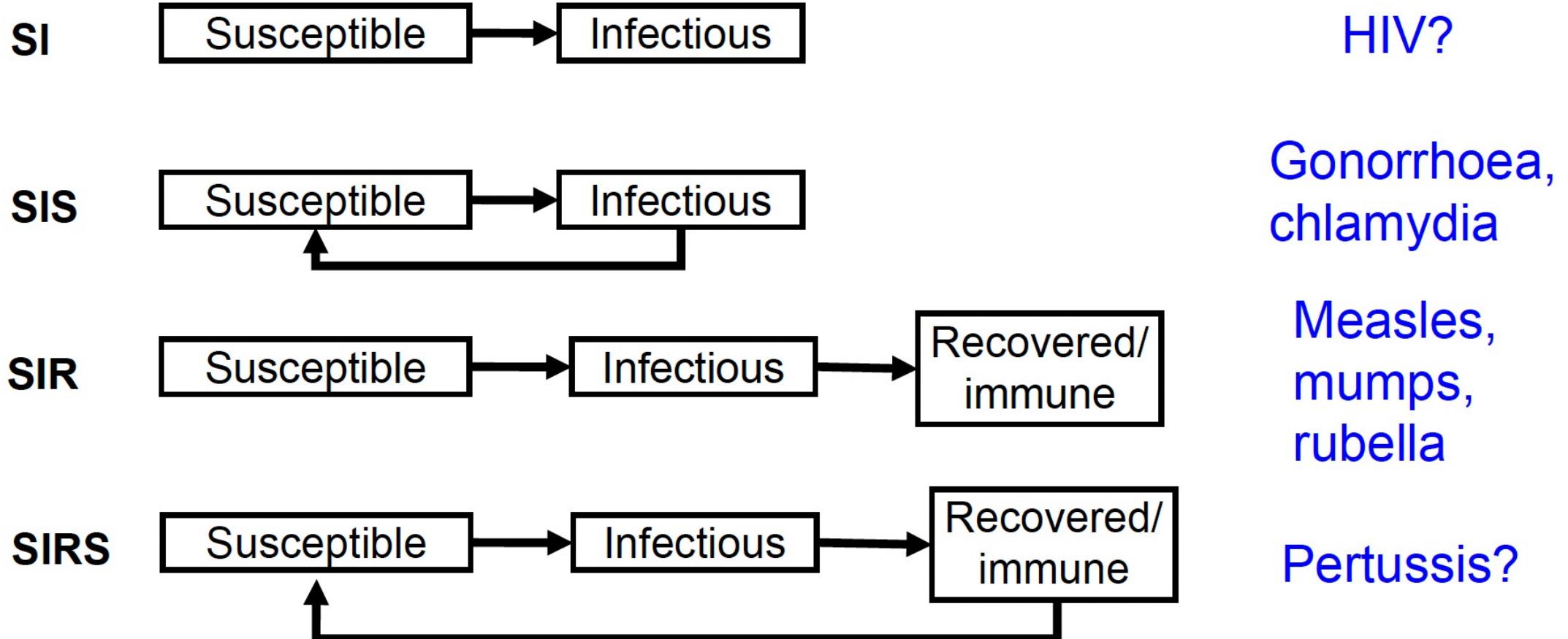
# Model structure

Construct transmission model based on current dynamics

- **Model structure** is usually determined by considering the **relationship between:**
  - **Inputs** relevant to the natural history of disease, clinical pathways, intervention effectiveness etc.
  - **Outputs** most useful to decision makers eg. cases of disease, deaths, hospital admissions, life years gained, QALYs, DALYs.

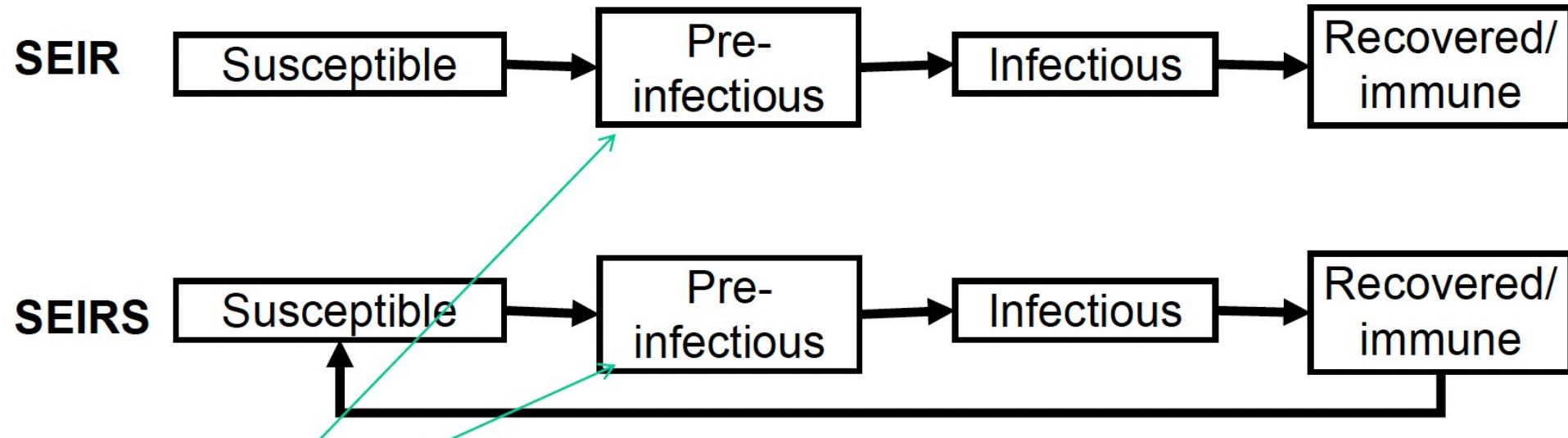
# Model structure

Construct transmission model based on current dynamics



Construct transmission model based on current dynamics

# Model Structure – Common Mistake

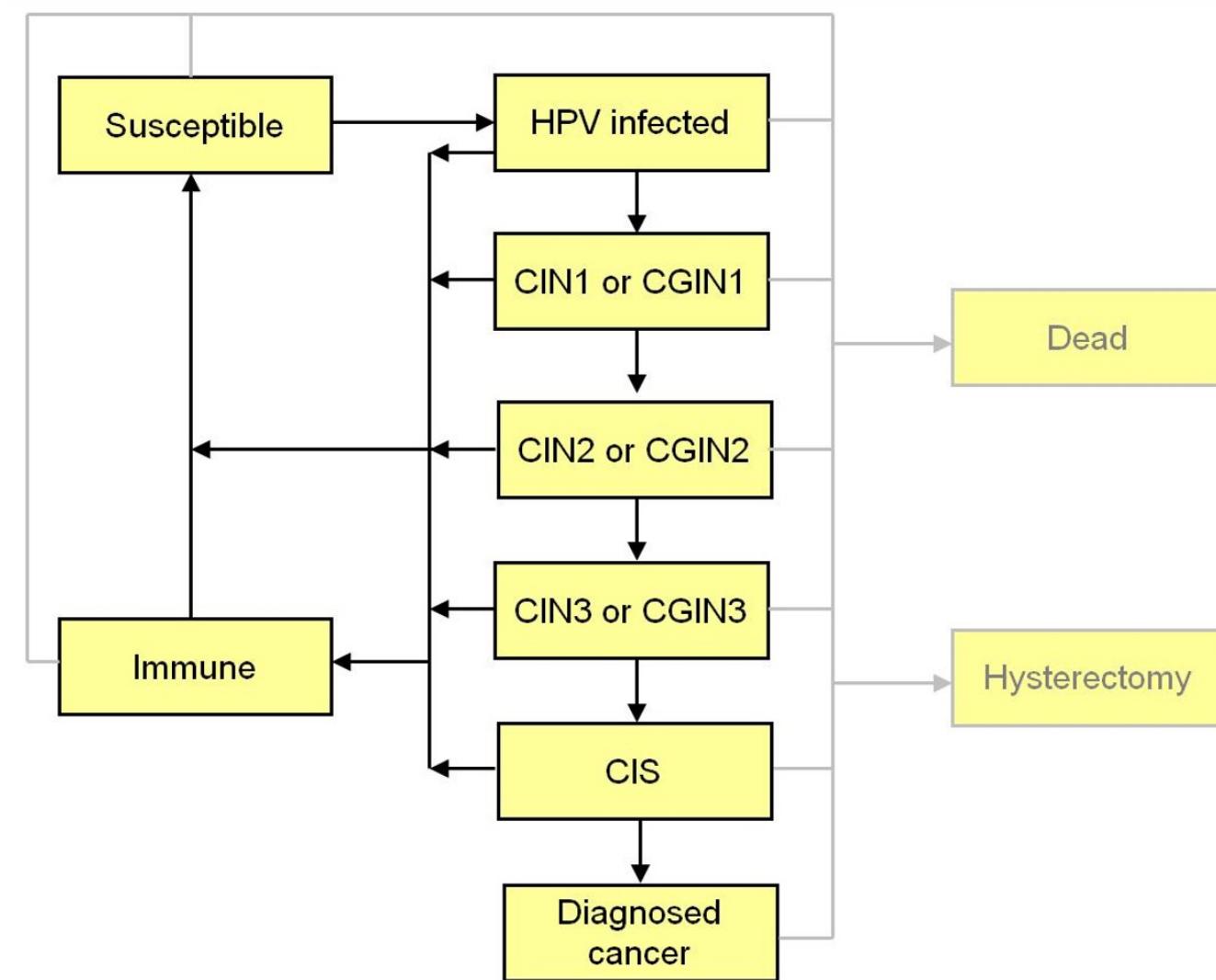


NB Often referred to as the “Infected” or “Exposed” category in the modelling literature...

BUT....this can be misleading : everyone is “Exposed” and infectious individuals can be considered “infected”.

Construct  
transmission  
model  
based on  
current  
dynamics

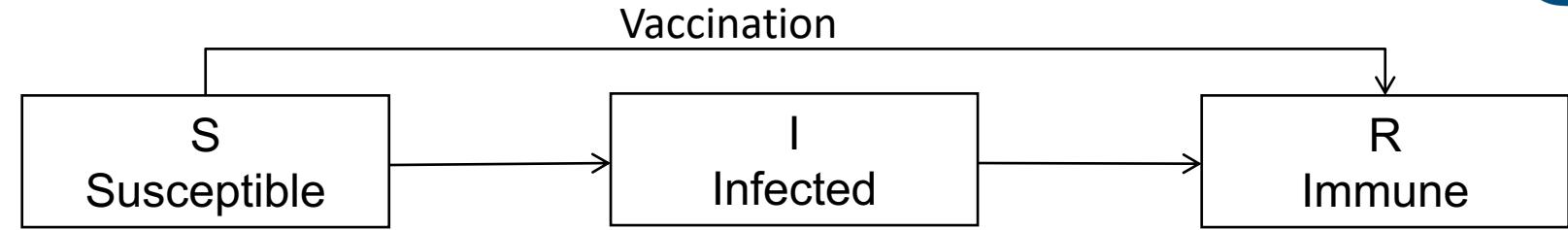
# Model structure



Construct transmission model based on current dynamics

# Types of models

Compartmental transmission dynamic model



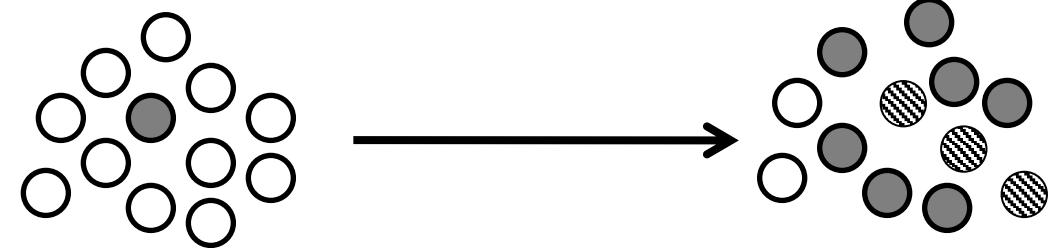
Structured compartmental model (e.g., age or sex structured)

Meta-population model (i.e., spatially structured)

Network model

Multi-scale model

Individual-(agent-) based model:



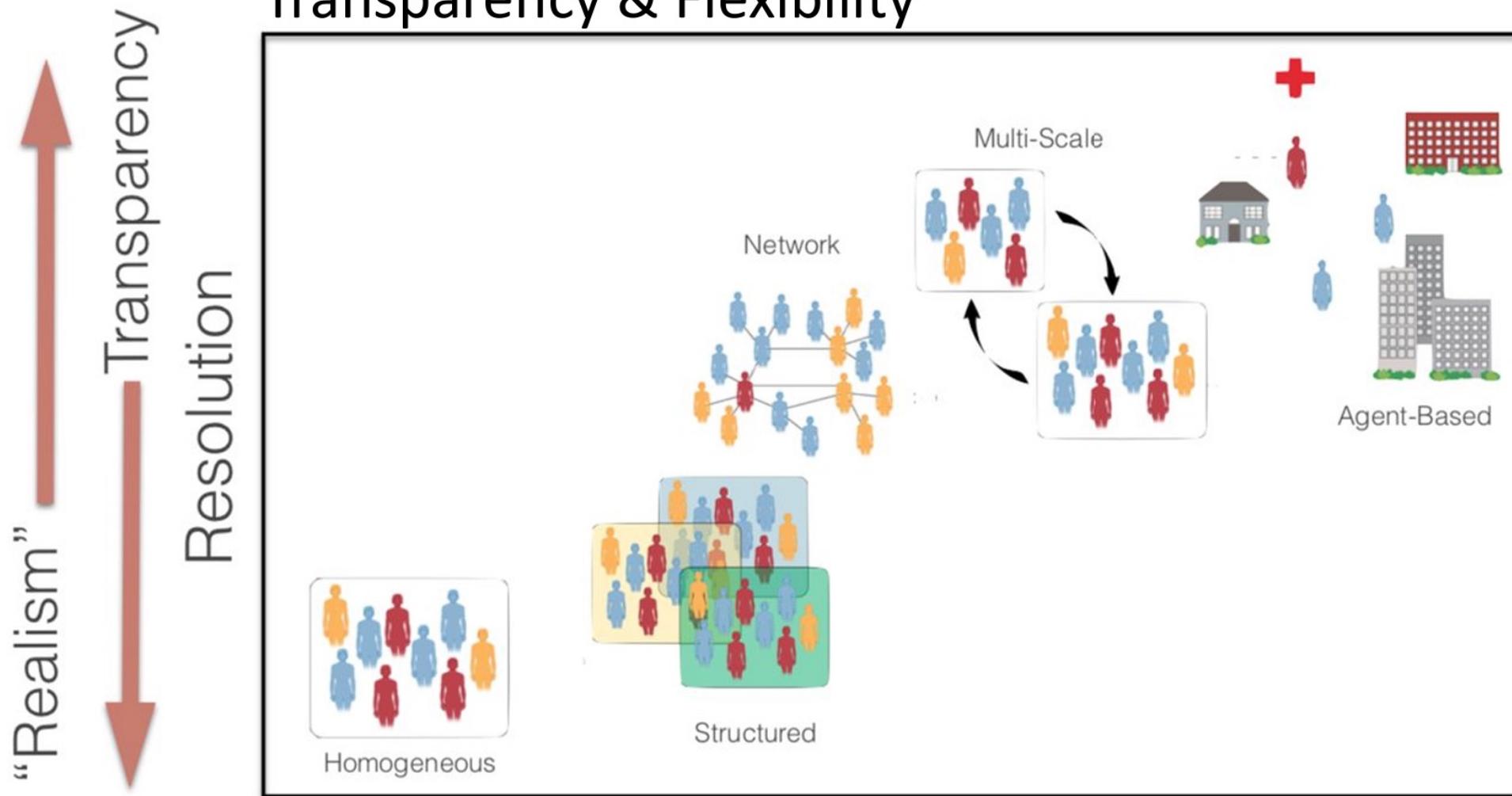
**Non-homogenous mixing**

- Susceptible
- Infected
- ▨ Immune

Construct transmission model based on current dynamics

# Model type depends on modelling objective

## Realism & Complexity vs. Transparency & Flexibility



Construct  
transmission  
model  
based on  
current  
dynamics

# What is a good model?

## Judging a Model...

Three fundamental features of models, often opposing forces:

**Accuracy:** ability to capture observed patterns (qualitative or quantitative) and make predictions

**Transparency:** Ability to understand model components. Decreases with model complexity

**Flexibility:** How easily the model can be adapted to new scenarios. Decreases with model complexity

# 2 Types Modelling Methods

## 2 main types:

### Stochastic

- incorporate chance variation
- provide the probability of a given outcome or range in which the outcome is likely to occur eg
  - probability that transmission ceases
  - 95% certain that 10-15 cases will be seen

To be discussed in detail in block 2

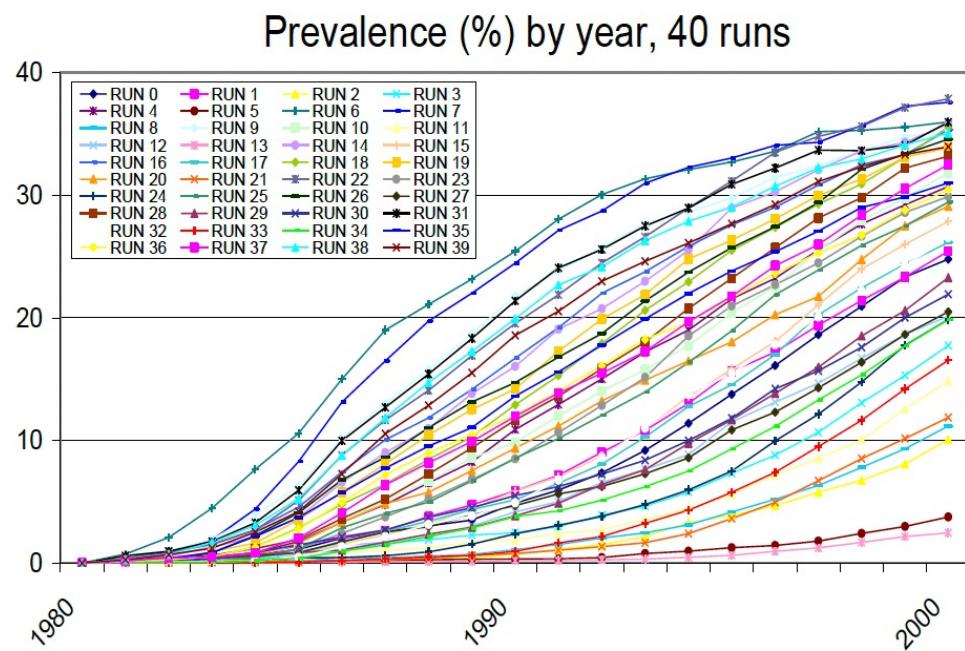
### Deterministic models

- describe what will happen on average in a population
- individuals are subdivided into categories (“compartments”)
- describe transitions between compartments

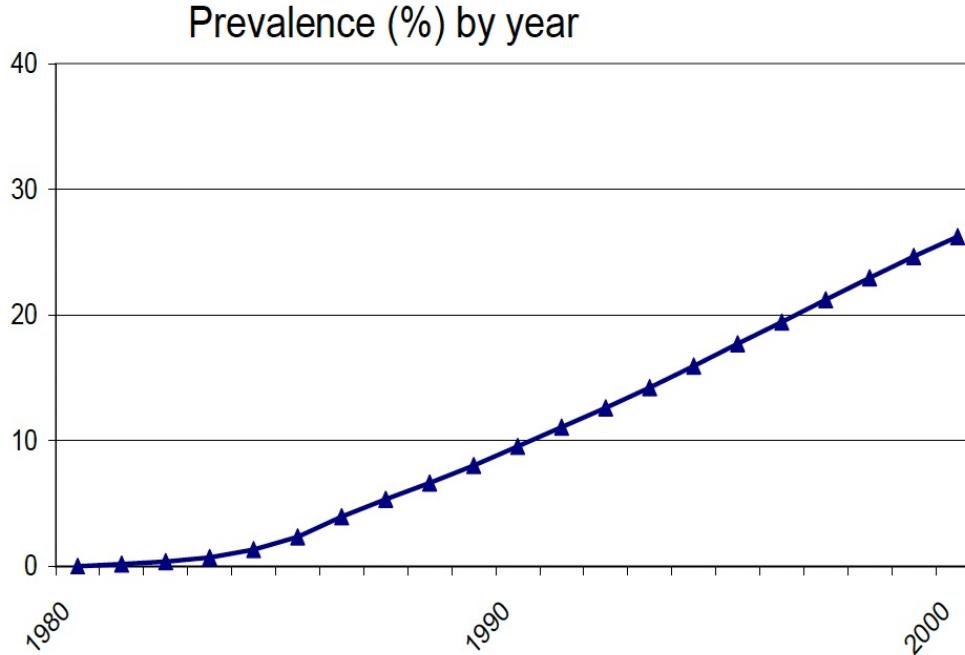
# 2 Types Modelling Methods

## Examples of output from a deterministic and a stochastic model

## Stochastic model



## Deterministic model



# Deterministic Models

Deterministic models are set up using either difference or differential equations

**Difference equations** calculate the number in each infection category using **discrete time** steps e.g. 1, 2, 3 days etc

number of cases tomorrow =  
number of cases today  
+ number of new cases with onset between today and tomorrow  
– the number of cases who recover between today and tomorrow

**Differential equations** calculate the number in each infection category using time steps which are “infinitesimally” small, i.e. in continuous time

# How to implement a mathematical model?

---

## Analytical Models

Concentrate on problems that can be expressed and analyzed fully using analytical approaches (*i.e., pure math*)

## Computational Models

Construct an appropriate model for the system and use a combination computational methods for analysis and scenario analysis (*i.e., computer simulations*)

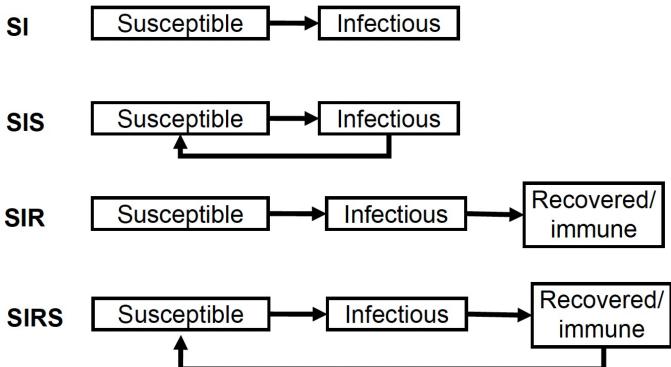


# Overview

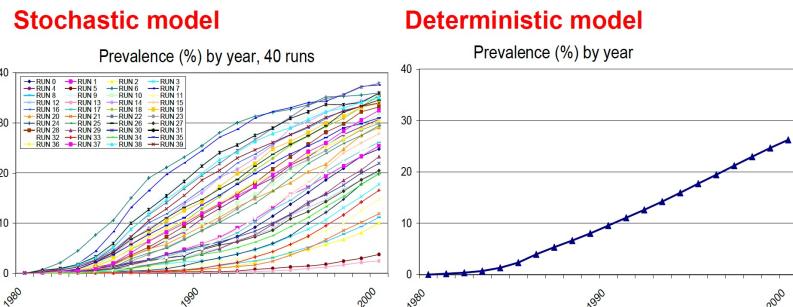
## Model Scale

→ Social, political, economic, health systems  
↓  
Populations ("between host models")  
↓  
→ Individuals, physiological systems  
↓  
Cells, genes, proteins ("within host models")

## Model Structure



## Model Processes



## Model Type

Compartmental transmission dynamic model

Structured compartmental model (e.g., age or sex structured)

Meta-population model (i.e., spatially structured)

Network model

Multi-scale model

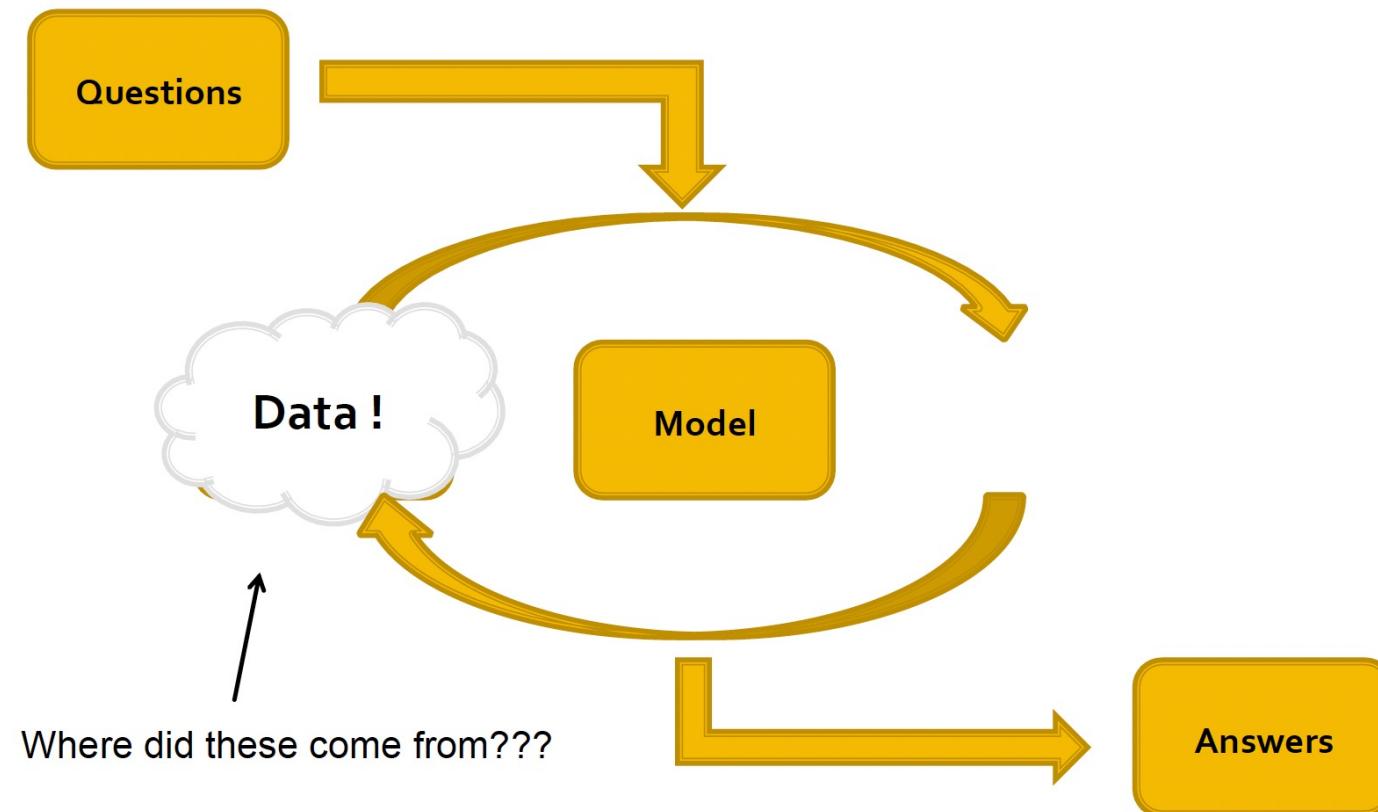
Individual-(agent-) based model:

# Infectious Disease Data

# Models and Data

Discovery comes from testing ideas (models) against observations (data)

Models drive empirical development and vice versa



# Case Definition

“a set of standard criteria for deciding whether a person has a particular disease [or infection]”

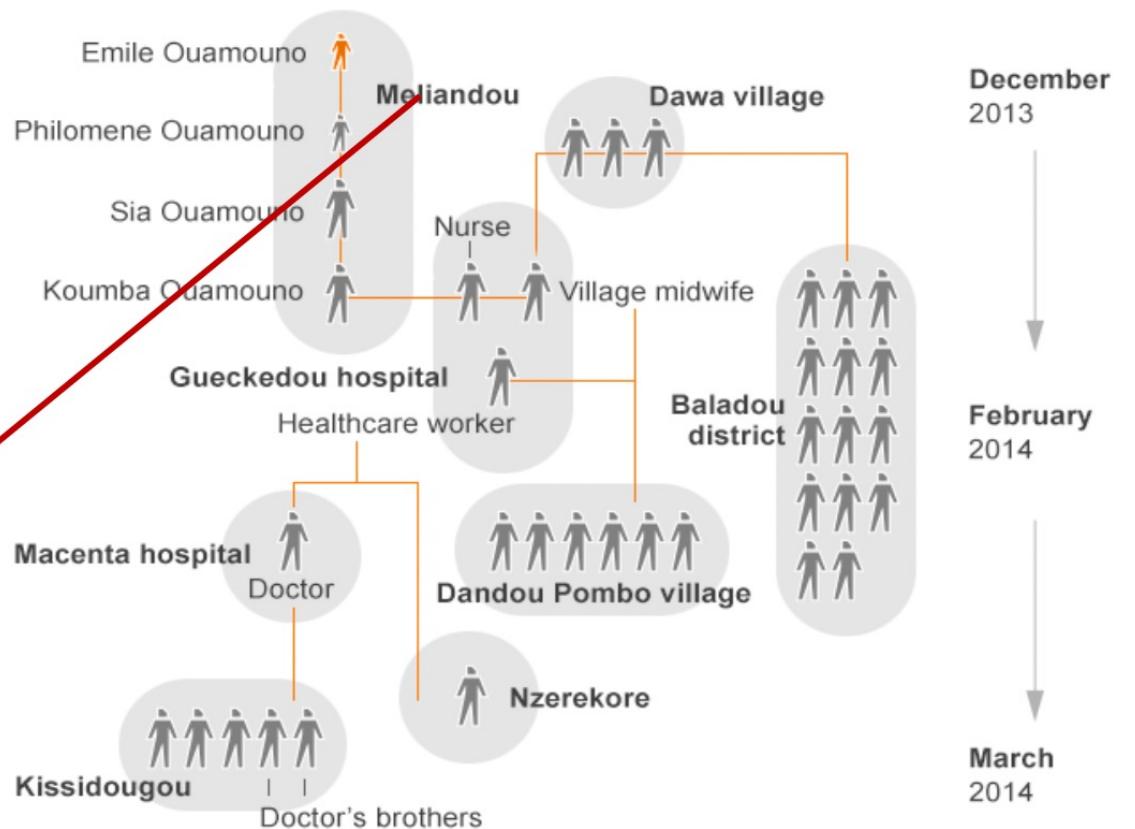
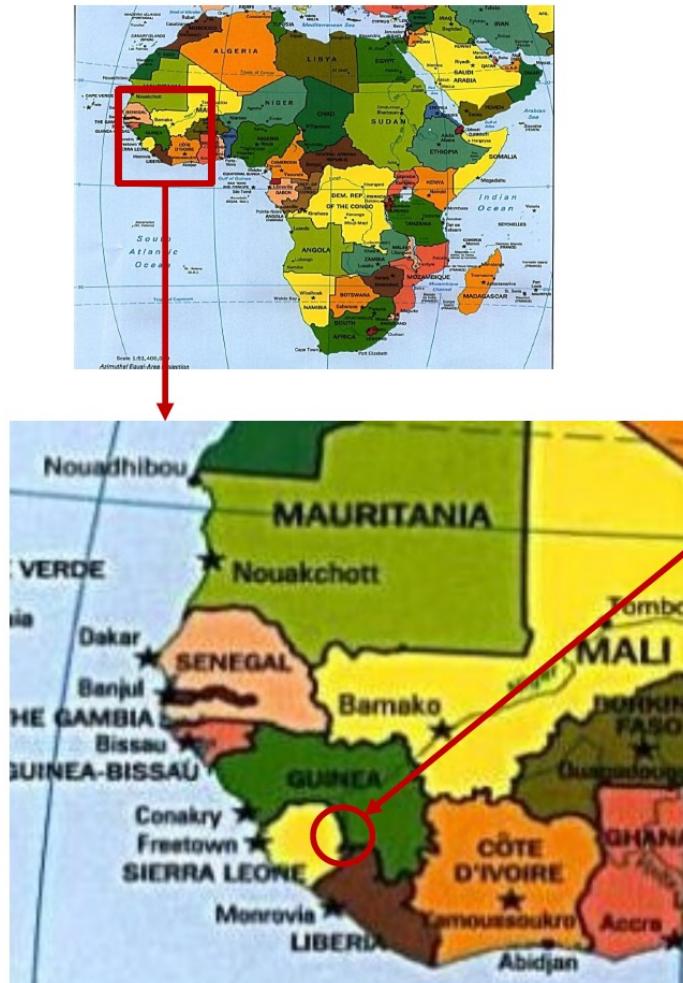
**Person:**

**Place:**

**Time:**

**Clinical description:**

# Ebola case study



# Ebola case study

“a set of standard criteria for deciding whether a person has a particular disease [or infection]”

**Person:** Residents of Meliandou, recent visitors to Meliandou

**Place:** West Africa, Guinea

**Time:** On or after November 15, 2013

**Clinical description:** Elevated body temperature or subjective fever or symptoms, including severe headache, fatigue, muscle pain, vomiting, diarrhea, abdominal pain, or unexplained hemorrhage

# Ebola case study



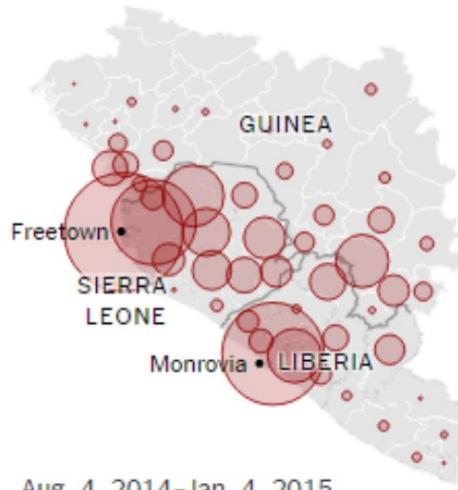
Dec. 30, 2013–May 4, 2014

**The Outbreak Begins**



May 5–Aug. 3, 2014

**New Cases Rise Rapidly**



Aug. 4, 2014–Jan. 4, 2015

**W.H.O. Sounds the Alarm**

**Average new cases each week**

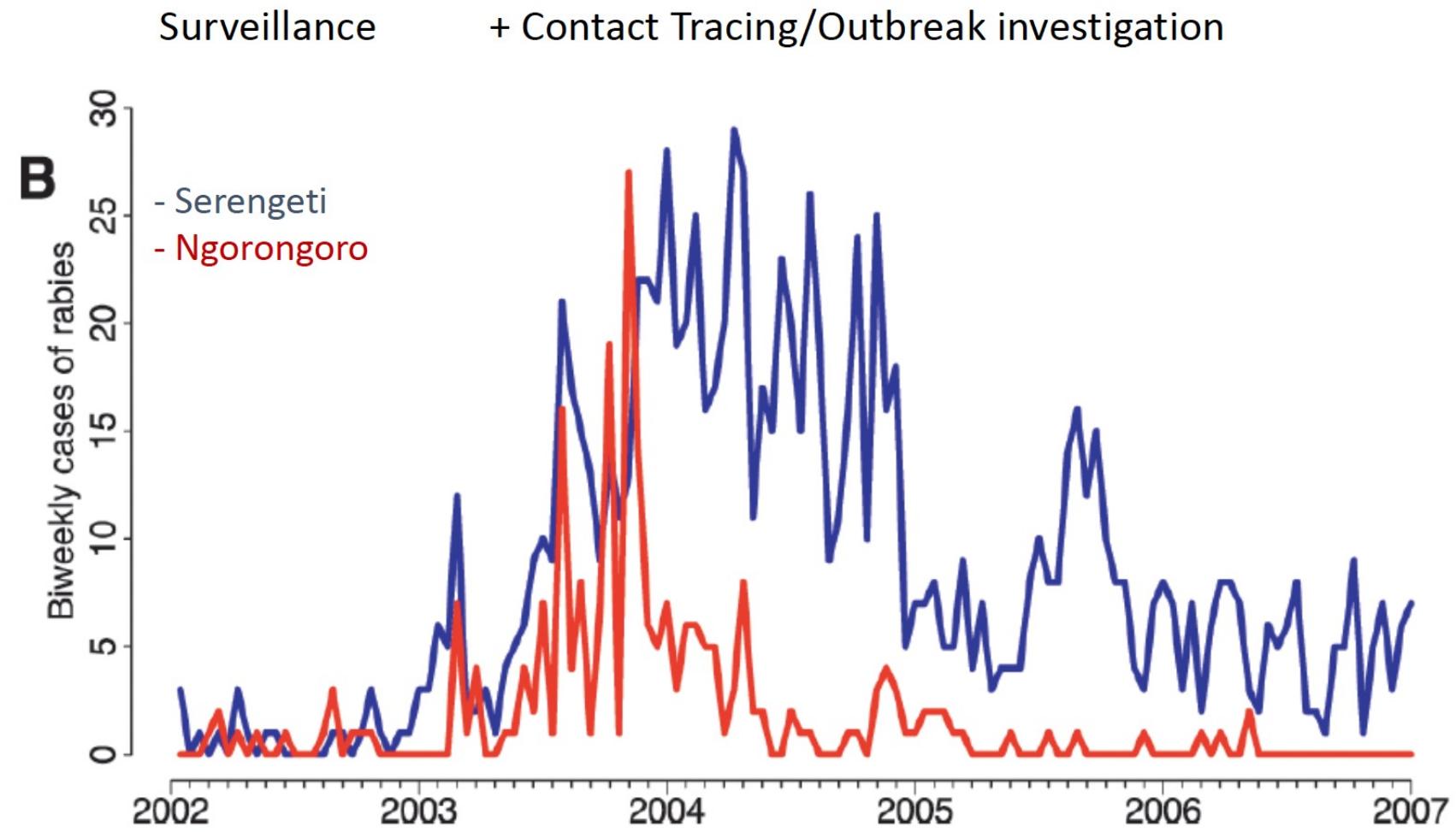
5    20    50

Source: World Health Organization

# Updated case definition for Ebola

<b>Person:</b>	Residents of and recent visitors to West Africa, including Senegal, Guinea, Sierra Leone and Liberia, as well as their close contacts or others in their community
<b>Place:</b>	Worldwide
<b>Time:</b>	On or after November 15, 2013
<b>Clinical Description:</b>	Illness with onset of fever and no response to treatment for usual causes of fever in the area, and at least one of the following signs: bloody diarrhoea, bleeding from gums, bleeding into skin (purpura), bleeding into eyes and urine.

# Ways of collecting case data



Hampson et al. 2009

# Ways of collecting case data

---

## Surveillance

- Passive
- Active

## Epidemiological Studies

- Case series
- Case control
- Outbreak investigation

# Levels of Data Aggregation

---

Aggregated data

De-identified data

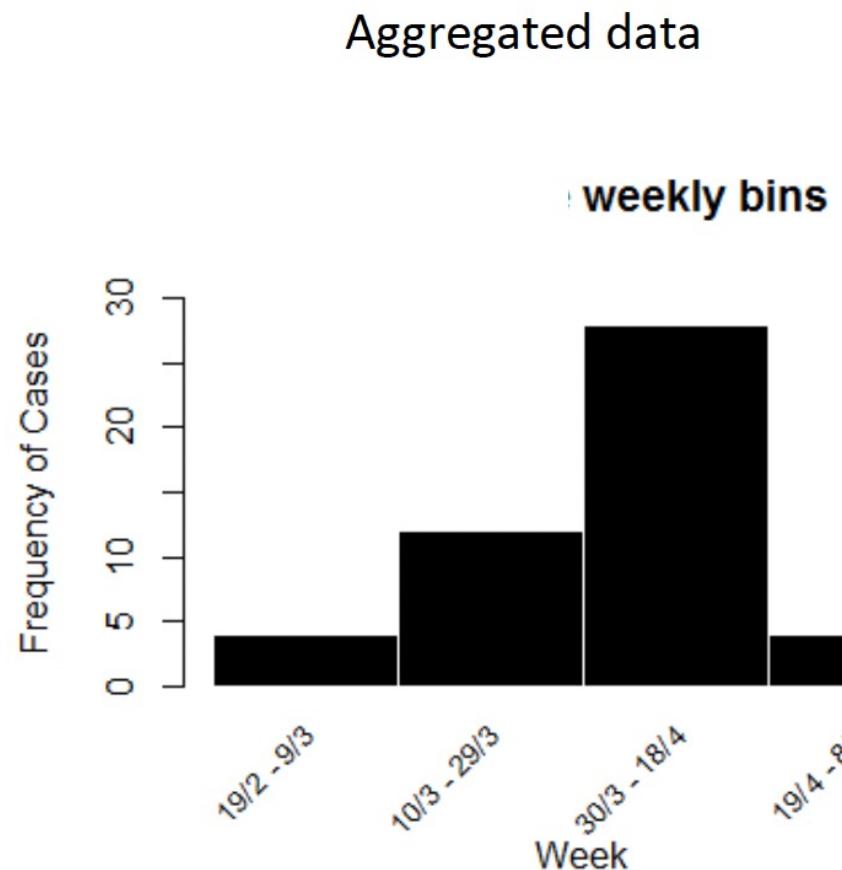
Personally identifying data

# Levels of Data Aggregation

Aggregated data

De-identified data

Personally identifying data



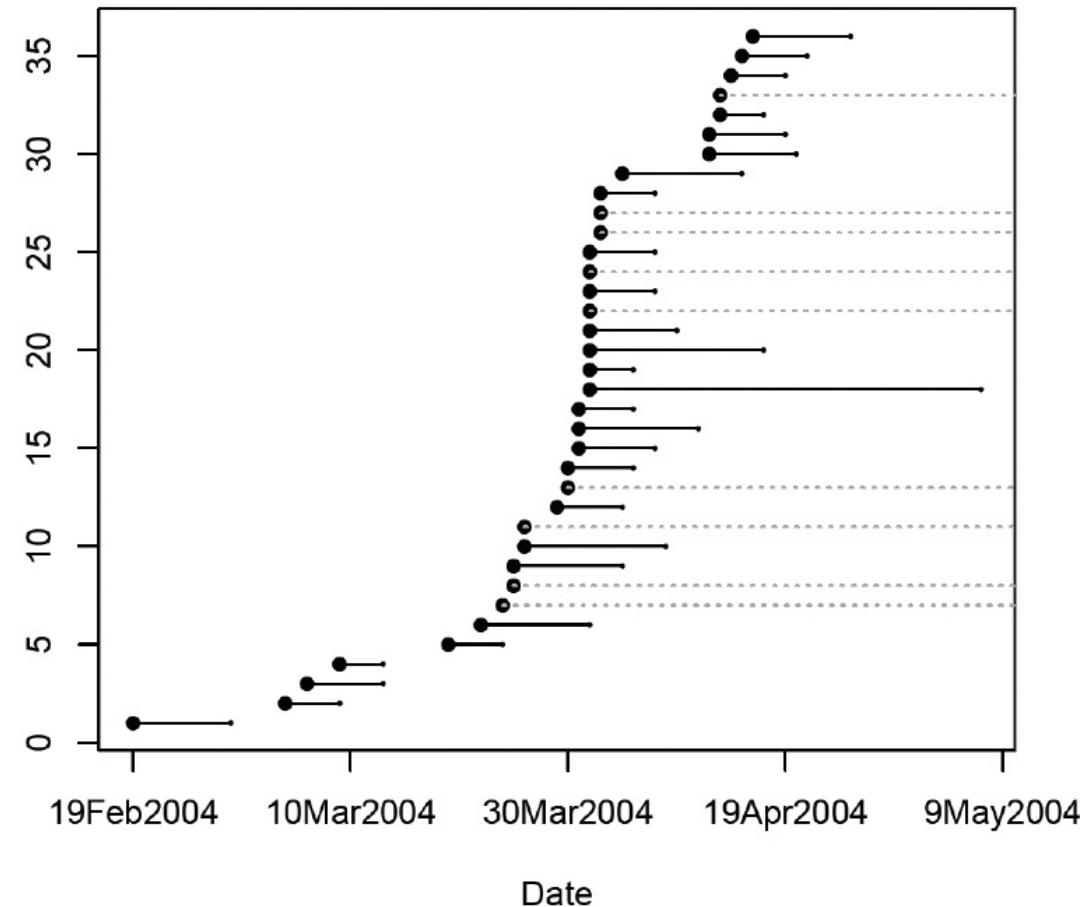
# Levels of Data Aggregation

Aggregated data

De-identified data

Personally identifying data

De-identified data



# Levels of Data Aggregation

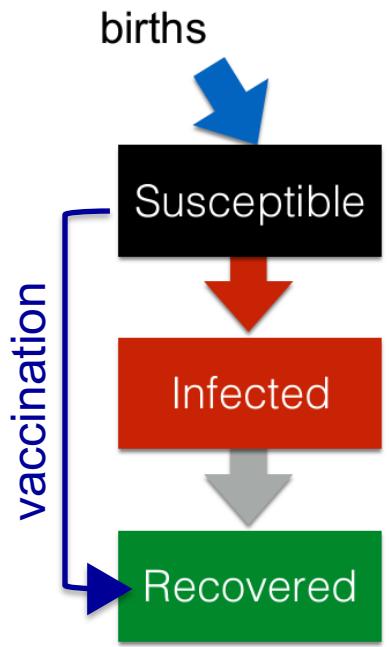
Aggregated data

De-identified data

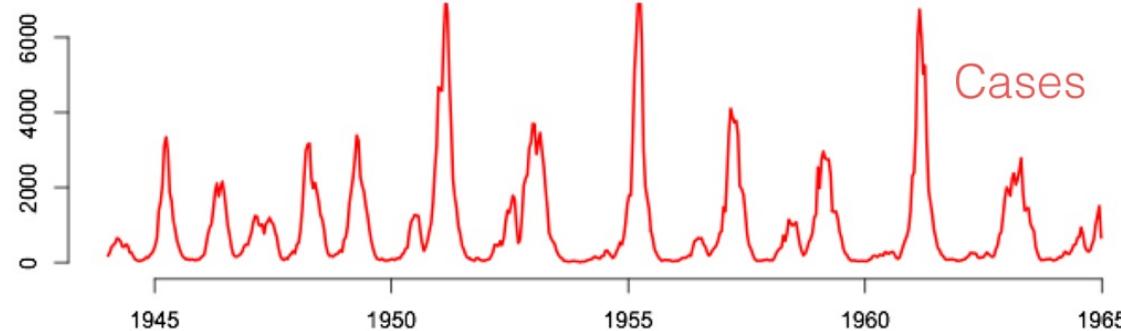
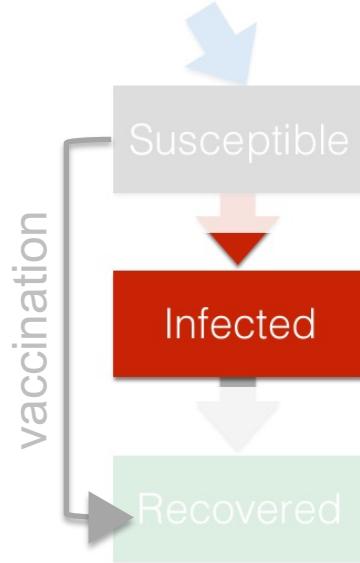
Personally identifying data

CaselD	Date of birth	Date Symptom Onset	Date Lab Received	Lab Test result
543	17 June 2010	4 Jan 2024	10 Jan 2024	Eq
544	9 July 2023	2 Jan 2024	10 Jan 2024	Pos
545	28 Nov 2021	8 Jan 2024	18 Jan 2024	Neg

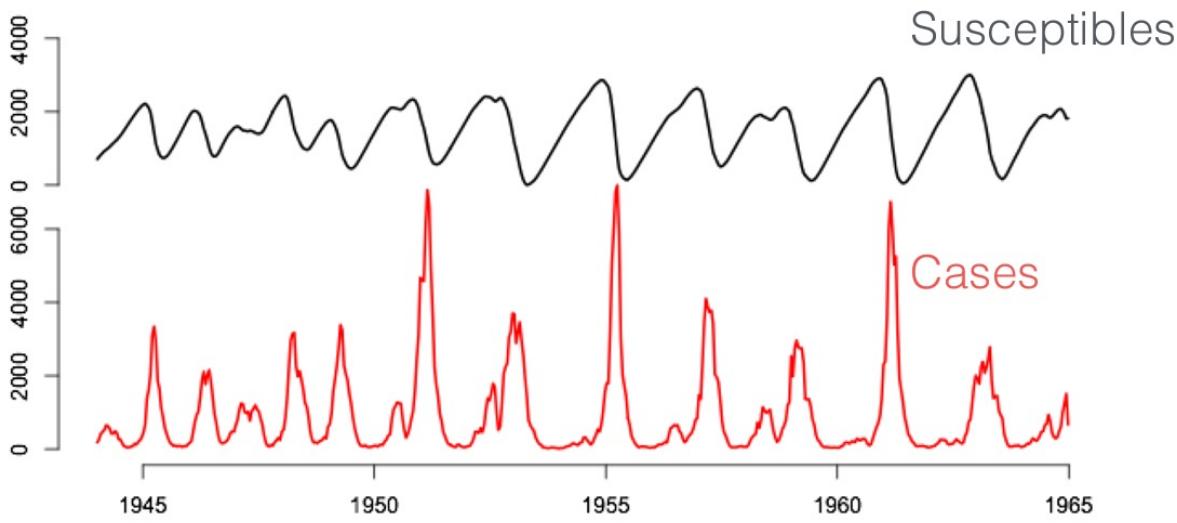
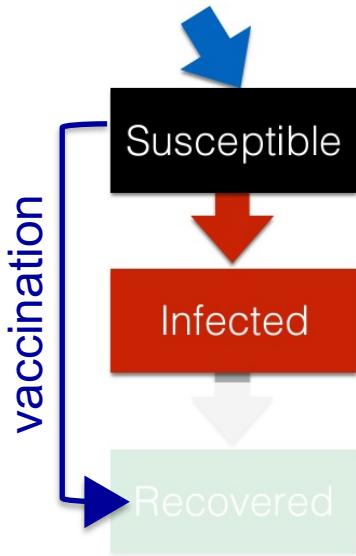
How are data related to ID models?



The **central process** in infectious disease epidemiology

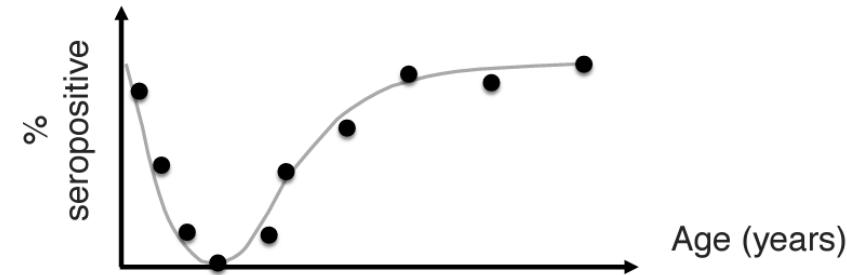
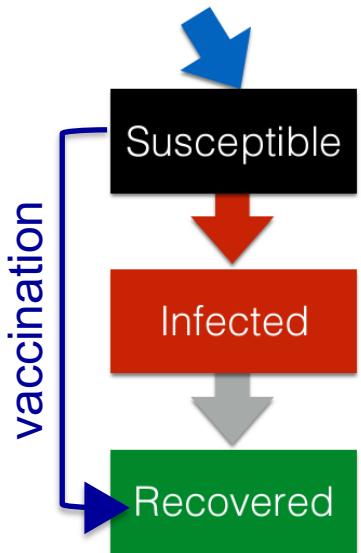


Only **cases** are directly observed



Only **cases** are directly observed  
 Susceptibles can be **inferred** with  
 sufficiently robust data on births,  
 vaccination, and cases through time

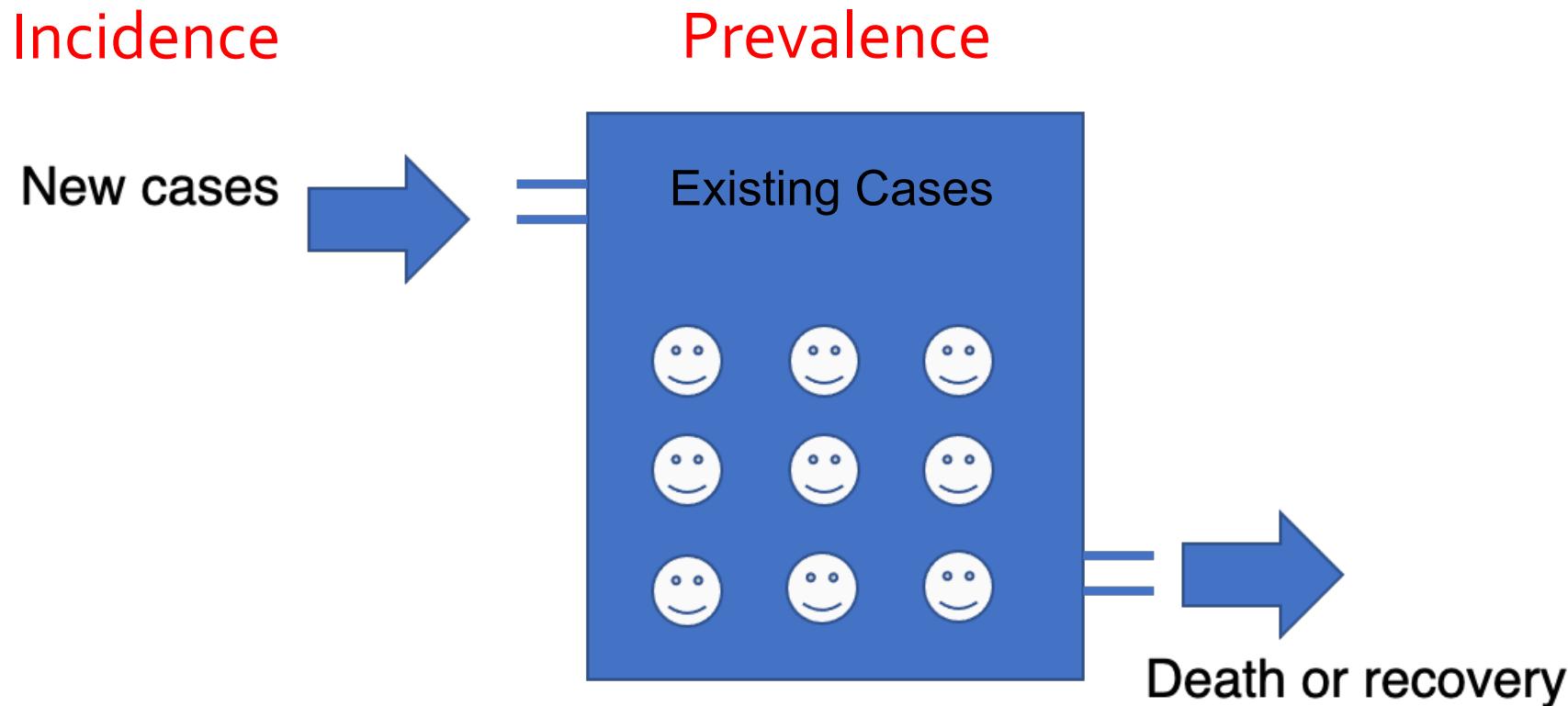
**Serology** can directly measure susceptibles



Only **cases** are directly observed

Susceptibles can be **inferred** with sufficiently robust data

# Measures of Disease Frequency



Disease **prevalence** is influenced by:

- Incidence of disease
- Duration of disease (time to recovery for non-fatal disease or survival time for fatal disease)

# Measures of Disease Frequency

$$\text{Incidence risk} = \frac{\text{Number of new cases in a time period}}{\text{Population at risk at the start}}$$

Cumulative incidence  
Attack rate

$$\text{Incidence rate} = \frac{\text{Number of new cases}}{\text{Total person-time at risk}}$$

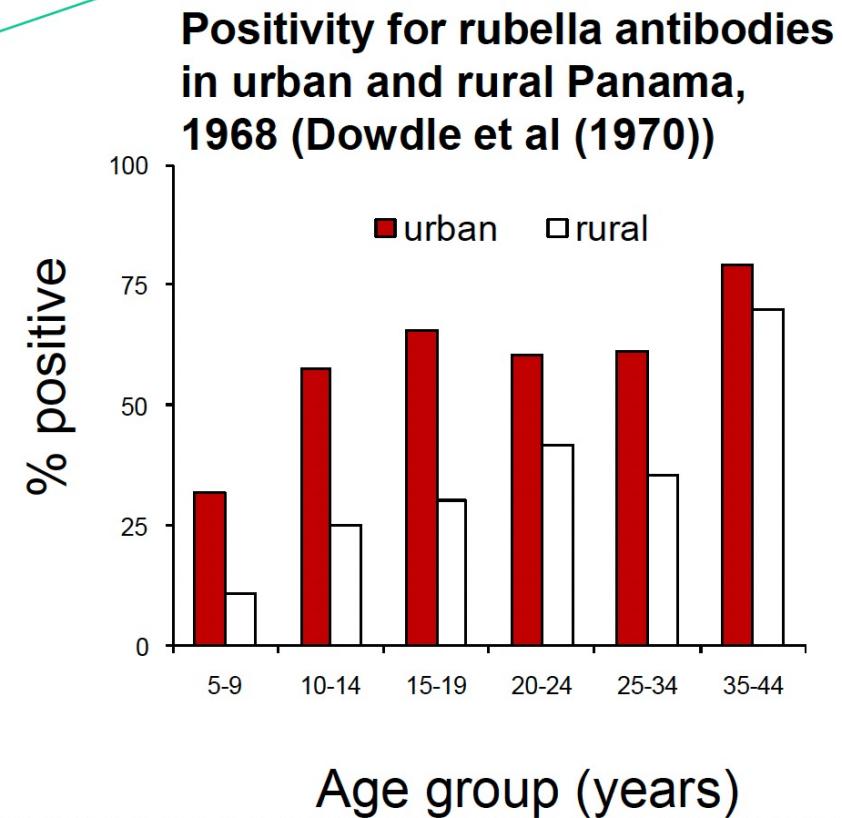
Incidence density rate

# Force of Infection

- Depends on the number of infectious individuals and rate at which susceptible individuals come into **effective contact** with an infectious individual

Defined as contact that is sufficient to lead to infection if it occurs between a susceptible and an infectious person

**NB depends on the infection and the setting**

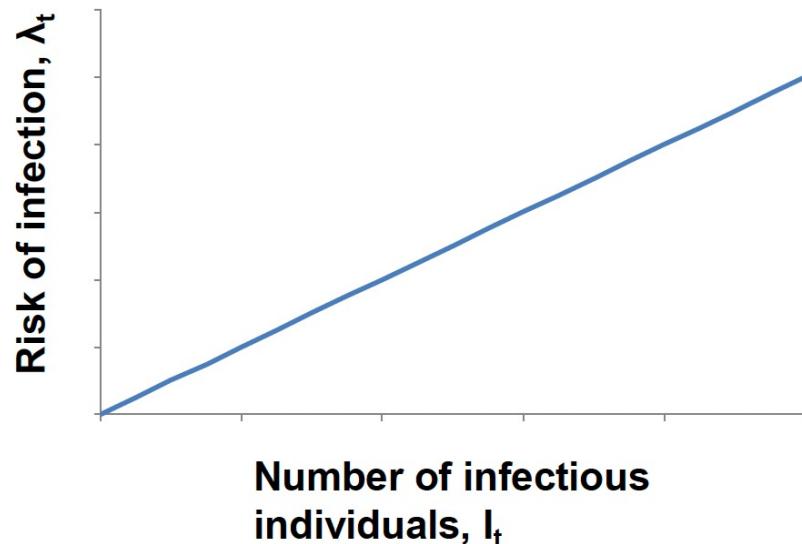


# Force of Infection

The simplest assumption: individuals mix randomly or the “mass action principle”

i.e. Risk of infection is proportional to the number of infectious individuals,  $I_t$ :

$$\lambda_t = \beta I_t$$



**Precise definition of  $\beta$ :**

The per capita (or per person) rate at which two specific individuals come into effective contact per unit time

Also known as “transmission rate”, “transmission parameter”, “transmission probability” etc

**During the course, it will usually be referred to as the “rate at which two specific individuals come into effective contact per unit time”**

# Risk vs Rate

The parameters which go into difference equations should be risks

e.g. the number of individuals who are infectious at time  $t+1$  =  
{number who were infected at time  $t$ }

$\times$

{proportion who became infectious between  $t$  and  $t+1$ }

However, under for most situations, the risk  $\approx$  rate

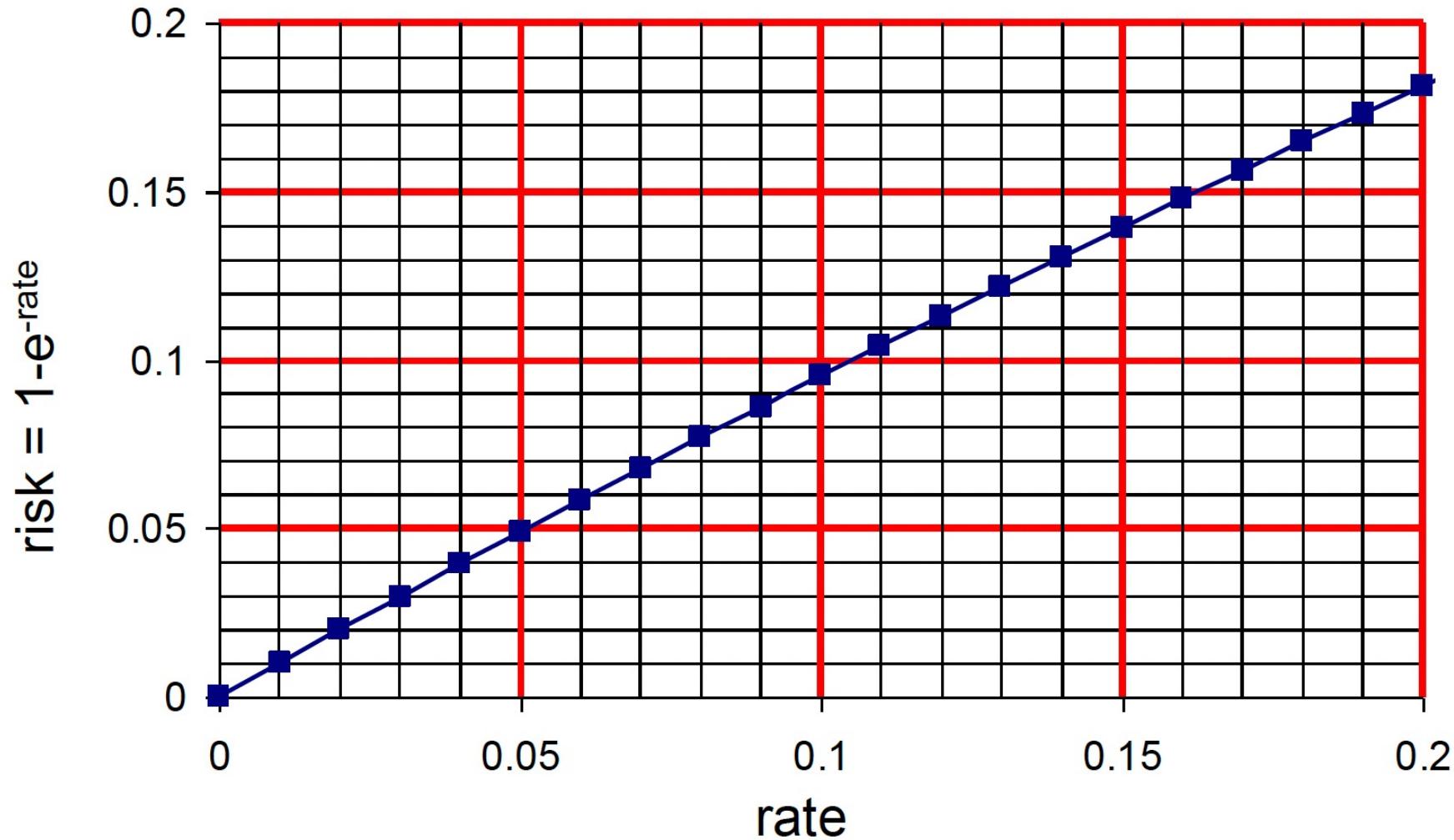
**From previous training you may recall:**

Risks and rates are related through the following expression:

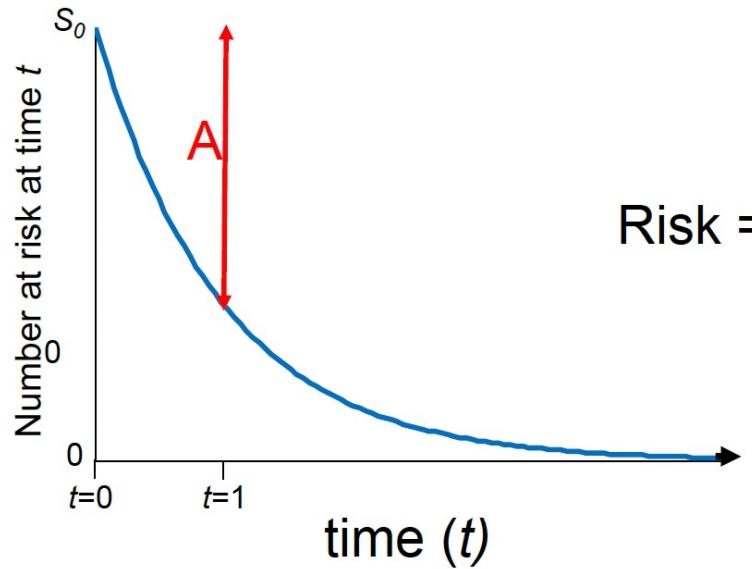
$$\text{risk} = 1 - e^{-\text{rate}}$$

If the rate is small, then  $e^{-\text{rate}} \approx 1 - \text{rate}$ , and so risk  $\approx$  rate

# Risk vs Rate



# Risk vs Rate



Risk =  $\frac{\text{number who became cases (A)}}{\text{number at risk at the start } (S_0)}$

Rate =  $\frac{\text{Number of new cases (A)}}{\text{Total person-time at risk}}$

Accounts for fact that those who develop disease by point A would have become cases in the meantime

# Rate to Average Time

The rate at which something occurs

$$= 1/\{\text{average time to the event}\}$$

The rate at which individuals become infectious

$$= 1/\{\text{average pre-infectious period}\}$$

The rate at which individuals recover from being infectious

$$= 1/\{\text{average duration of infectiousness}\}$$

The mortality rate

$$= 1/\{\text{average "duration of life" or the life expectancy}\}$$

# Metrics to measure disease presence and spread of disease

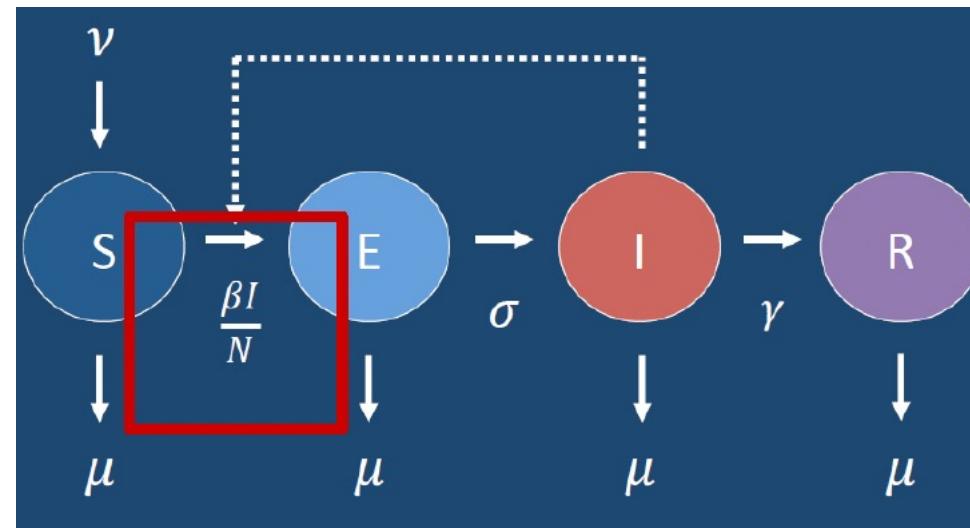
---

- Incidence of infection
- Prevalence of infection
- Seroprevalence (prevalence of antibody)

# Metrics to measles presence and spread of disease

- Incidence of infection
- Prevalence of infection
- Seroprevalence (prevalence of antibody)

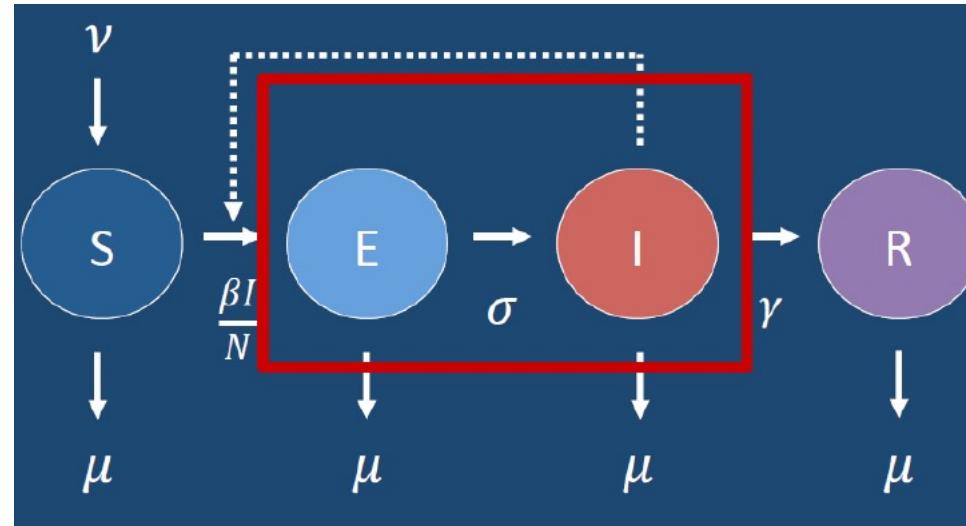
Reminder: E (exposed) class = infected but pre-infectious  
 $\beta SI/N$



# Metrics to measles presence and spread of disease

- Incidence of infection
- Prevalence of infection
- Seroprevalence (prevalence of antibody)

$(E+I) / N$

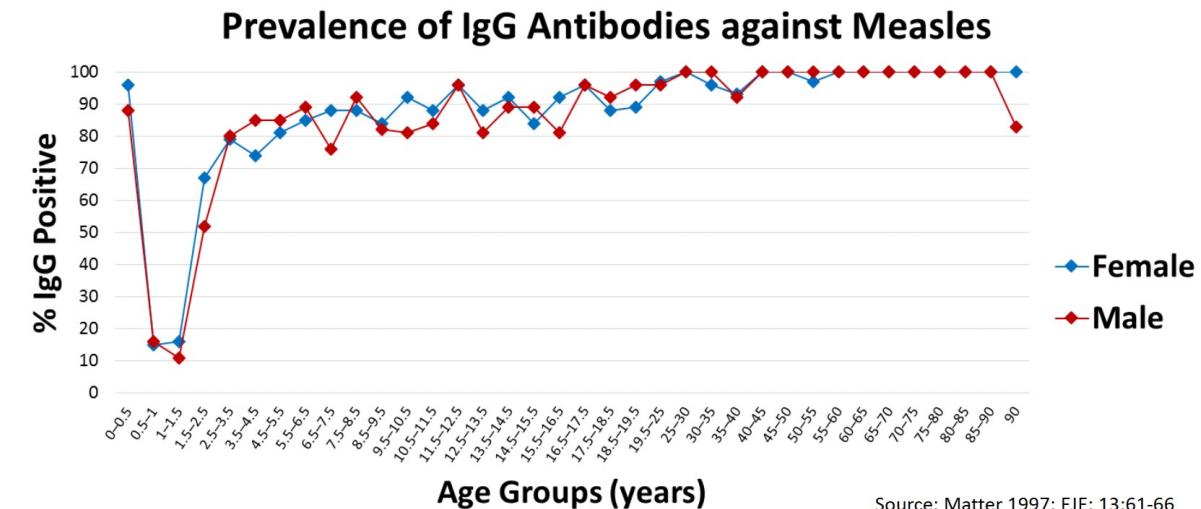
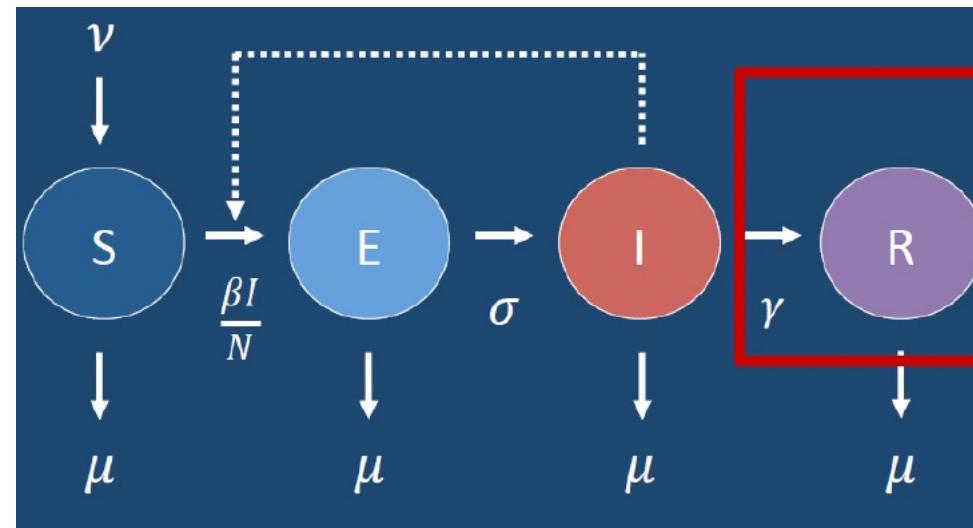


# Metrics to measles presence and spread of disease

- Incidence of infection
- Prevalence of infection
- Seroprevalence (prevalence of antibody)

Can be related to:  
• Prevalence of infection or  
• Past infection

R / N



# Measles and Rubella Surveillance Data

# Measles

**Table 5.3. Overview of relevant outcomes from all other data sources: World Health Organization (WHO), Centers for Disease Control and Prevention (CDC), Red Book (RB), Quick Reference guide (RG), Richardson 2001 (R2001)**

Disease/agent	Incubation period	Period of infectiousness	Period of shedding	Exclusion period	Reference
<b>Vaccine preventable diseases</b>					
Measles	<b>WHO:</b> 7–18 (often, 10–12) days from exposure to the onset of fever [121]; <b>RB:</b> Generally 8–12 days from exposure to onset of symptoms; <b>CDC:</b> 7–21 (14) days to a rash [122–124]; 10–12 days to prodrome [123, 125] <b>R2001:</b> 6–19 (13) days	<b>WHO:</b> 4 days before the rash until 1–2 days after a rash [121]; <b>RB:</b> 4 days before the rash to 4 days after a rash; <b>CDC:</b> 4 days before to 4 days after a rash onset [123–125] <b>R2001:</b> ND (but at least 1–2 days before a rash)	<b>WHO:</b> Following a rash onset, measles excretes for very short period (about 5 days) [126]; <b>CDC:</b> Beginning with the prodrome until 3–4 days after a rash onset ; <b>R2001:</b> -2 to +3 days	<b>RB:</b> Until 24 hours after treatment has been initiated; <b>RG:</b> At least 2 weeks after a rash in the last case for unimmunised people who have been exempted from measles immunised within 72 hr of exposure; <b>CDC:</b> 4 days after a rash for cases; 21 days after a rash in the last case for persons who have been exempted from measles vaccination within the appropriate time; <b>R2001:</b> 5 days from onset of a rash	[121–126]

Reference: European Centre for Disease Prevention and Control. Systematic review on the incubation and infectiousness/shedding period of communicable diseases in children.

Stockholm: ECDC; 2016

# Measles Disease Description

Measles is an acute viral illness caused by a virus in the family paramyxovirus, genus *Morbillivirus*. Measles is characterized by a prodrome of fever (as high as 105°F) and malaise, cough, coryza, and conjunctivitis, followed by a maculopapular rash. [1] The rash spreads from head to trunk to lower extremities. Measles is usually a mild or moderately severe illness. However, measles can result in complications such as pneumonia, encephalitis, and death. Approximately one case of encephalitis.[2] and two to three deaths may occur for every 1,000 reported measles cases.[3]

One rare long-term sequelae of measles virus infection is subacute sclerosing panencephalitis (SSPE), a fatal disease of the central nervous system that generally develops 7–10 years after infection. Among persons who contracted measles during the resurgence in the United States (U.S.) in 1989–1991, the risk of SSPE was estimated to be 7–11 cases/100,000 cases of measles.[4] The risk of developing SSPE may be higher when measles occurs prior to the second year of life.[4]

The average incubation period for measles is 11–12 days[5], and the average interval between exposure and rash onset is 14 days, with a range of 7–21 days. [1, 6] Persons with measles are usually considered infectious from four days before until four days after onset of rash with the rash onset being considered as day zero.

# Rubella Disease Description

Rubella is a viral illness caused by a togavirus of the genus *Rubivirus* and is characterized by a mild, maculopapular rash. The rubella rash occurs in 50%–80% of rubella-infected persons and is sometimes misdiagnosed as measles or scarlet fever. Children usually develop few or no constitutional symptoms, but adults may experience a 1–5-day prodrome of low-grade fever, headache, malaise, mild coryza, and conjunctivitis. Postauricular, occipital and posterior cervical lymphadenopathy is characteristic and precedes the rash by 5–10 days. Arthralgia or arthritis may occur in up to 70% of adult women with rubella. Rare complications include thrombocytopenic purpura and encephalitis.<sup>[1–3]</sup> Rubella is transmitted through direct or droplet contact from nasopharyngeal secretions and has an average incubation period of 17 days (range: 12–23 days). Persons with rubella are most infectious when rash is erupting, but they can shed virus from 7 days before to 7 days after rash onset.

When rubella infection occurs during pregnancy, especially during the first trimester, serious consequences can result. These include miscarriages, fetal deaths/stillbirths, and a constellation of severe birth defects known as congenital rubella syndrome (CRS). The most common congenital defects are cataracts, heart defects and hearing impairment. See [Chapter 15, “Congenital Rubella Syndrome,”](#) for more details.

# Congenital Rubella Syndrome Description

Congenital rubella syndrome (CRS) is an illness in infants that results from maternal infection with rubella virus during pregnancy. When rubella infection occurs during early pregnancy, serious consequences—such as miscarriages, stillbirths, and a constellation of severe birth defects in infants—can result. The risk of congenital infection and defects is highest during the first 12 weeks of gestation and decreases thereafter; defects are rare after infection in the 20th week (or later) of gestation.<sup>[1–3]</sup> Common congenital defects of CRS include cataracts, congenital heart disease, hearing impairment, and developmental delay. Infants with CRS often present with more than one of these signs but may also present with a single defect, most commonly hearing impairment.

See [Chapter 14](#), “Rubella,” for more information on rubella infection.

# Fever-Rash Surveillance System



**Under-reporting (non-constant)**

**NB Rubella: Up to 40% of infections are sub-clinical**

# What is a measles “case”

**Suspected Case** = **Clinically compatible case**: fever, rash, 3 Cs (cough, corzya, conjunctivitis)

**Epidemiologically linked**: clinically compatible +

**Confirmed case**: clinically compatible + measles-IgM test positive

**Other**: clinically compatible + measles-IgM test negative

# What is a rubella "case"?

**Suspected Case** = Clinically compatible case: fever, rash, 3 Cs (cough, corzya, conjunctivitis)

**Epidemiologically linked:** clinically compatible + contact with suspected measles cases

**Confirmed case:** clinically compatible + rubella-IgM test positive

# Bias in measles suspected cases

Other acute infections that may be confused with measles:

- Rubella!
- Scarlet fever
- Human herpes virus type 6 and 7
- Entero- and adenoviruses
- Epstein-Barr virus
- Coxsackievirus
- Parvovirus B19
- Dengue viruses
- Cytomegalovirus
- Chikungunya-virus
- Zika-virus
- West-Nile-virus
- Ross River-virus
- Sindbis-virus

# Measles and Rubella Line List

Individual level data made up of the following variables (plus or minus)

- Sex
- Age
- DateOfBirth
- DateSpecimenCollected
- DateSpecimenSentLab
- DateLastVaccination
- DateOfOnset
- DateSpecimenRecInLab
- DistrictOfResidence
- SpecimenCondition
- ReportingFacilty
- StateOfResidence
- DistrictOfResidence
- MeaslesIgM
- RubellaIgM
- FinalClassification

# MR Vaccination Coverage

Survey data –

- Population-based survey design
- Include denominator and numerator

Administrative vaccination data –

- Number of doses administered
- Numerator only

Administrative vaccination coverage –

- # doses / population size of target birth cohort
- Denominator from census projections

# MR Vaccination Coverage Biases

- Migrant populations captured? If so, where?
- Numerator issues: 2 doses to one child, doses to children not in the target cohort, mis-reporting
- Denominator issues: census mis-reporting, projection uncertainty, population movement

# Citations – These Slides were modified from the following:

Thunbi SM, Bruce F, Bellan SE, Pulliam JRCP.  
“Introduction to Infectious Disease Data” Clinic on the  
Meaningful Modeling of  
Epidemiological Data.  
<https://doi.org/10.6084/m9.figshare.5044603.v4>

[https://sismid2023.callumarnold.com/L01\\_intro-to-modeling](https://sismid2023.callumarnold.com/L01_intro-to-modeling)

Bellan, Steven; Borcherding, Rebecca; Bruce, Faikah;  
Dushoff, Jonathan; Grebe, Eduard; Hargrove, John; et al.  
(2017). International Clinics on Infectious Disease  
Dynamics and Data. figshare. Collection.  
<https://doi.org/10.6084/m9.figshare.c.3788224.v13>

*Mark Jit, LSHTM*