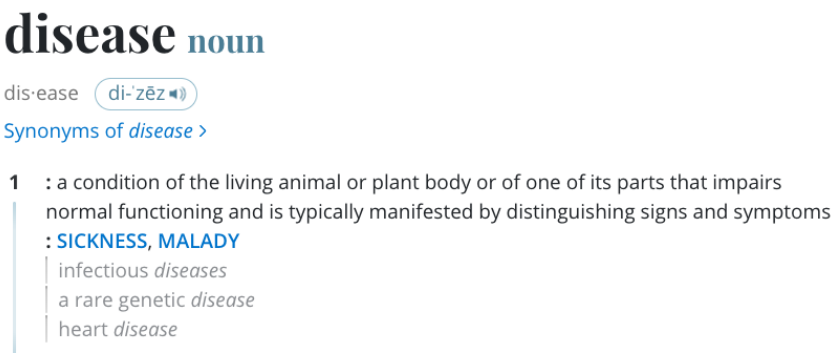
Introduction to Infectious Disease Modelling

James Mba Azam, PhD

An overview of infectious diseases

What are infectious diseasess



disease noun

dis·ease di-ˈzēz

Synonyms of *disease* ›

1 : a condition of the living animal or plant body or of one of its parts that impairs normal functioning and is typically manifested by distinguishing signs and symptoms : SICKNESS, MALADY

- infectious *diseases*
- a rare genetic *disease*
- heart *disease*

Figure 1: Generic definition of a disease (source: Merriam Webster)

Depending on what we care about, we can classify diseases according to:

- Cause (e.g., infectious, non-infectious)
- Duration (e.g., acute, chronic)
- Mode of transmission (direct or indirect)
- Impact on the host (e.g., fatal, non-fatal)

- Note that these classifications are not mutually exclusive. Hence, a disease can be classified under more than one category at a time.

How are infectious diseases controlled?

- In general, infectious disease control aims to reduce disease transmission.
- The type of control used depends on the disease and its characteristics.
- Broadly, there are two main types of control measures:
  - Pharmaceutical interventions (PIs)
  - Non-pharmaceutical interventions (NPIs)

Pharmaceutical interventions (PIs)

- Pharmaceutical Interventions are medical interventions that target the pathogen or the host.
- Examples:
  - Vaccines,
  - Antiviral drugs, and
  - Antibiotics.

Vaccination

- Most effective way to prevent infectious diseases.
- Activates the host's immune system to produce antibodies against the pathogen.
- Generally applied prophylactically to susceptible individuals (before infection); this reduces the risk of infection and disease.

- Challenges with vaccination:
  - Take time to develop for new pathogens
  - Never 100% effective and limited duration of protection
  - Adverse side effects
  - Some individuals cannot be vaccinated or refuce vaccination
  - Some pathogens mutate rapidly (e.g., influenza virus)

Non-pharmaceutical interventions (NPIs)

- Non-pharmaceutical interventions are measures that do not involve medical interventions.
- Examples:
  - Quarantine,
  - Physical/social distancing, and
  - Mask-wearing.

Quarantine

- *Isolation of individuals who may have been exposed to a contagious disease*.
- Advantage is that it's simple and its effectiveness does not depend on the disease.
- Disadvantages include:
  - Infringement on individual rights
  - Can be difficult to enforce
  - Can be costly
  - Can lead to social stigma

Contact tracing

- Contact tracing is used to identify exposed individuals, i.e., individuals who might been in contact with an infected/infectious individual.
- It involves identifying, assessing, and managing people who have been exposed to a contagious disease to prevent further transmission.
- It is a critical component of infectious disease surveillance and is often used in combination with other control measures.

Discussion

- How can we quantify the impact of a control measure?

Infectious disease models

What are infectious disease models?

- *Models* generally refer to conceptual representations of an object or system.
- *Mathematical models* use mathematics to describe the system. For example, the famous $E = mc^2$ is a mathematical model that describes the relationship between mass and energy.
- *Infectious disease models* use mathematics/statistics to represent dynamics/spread of infectious diseases.

- Mathematical models can be used to link the biological process of disease transmission and the emergent dynamics of infection at the population level.
- Models require making some assumptions and abstractions.

- By definition, “all models are wrong, but some are useful” (Box 1979).
  - Good enough models are those that capture the essential features of the system being studied.

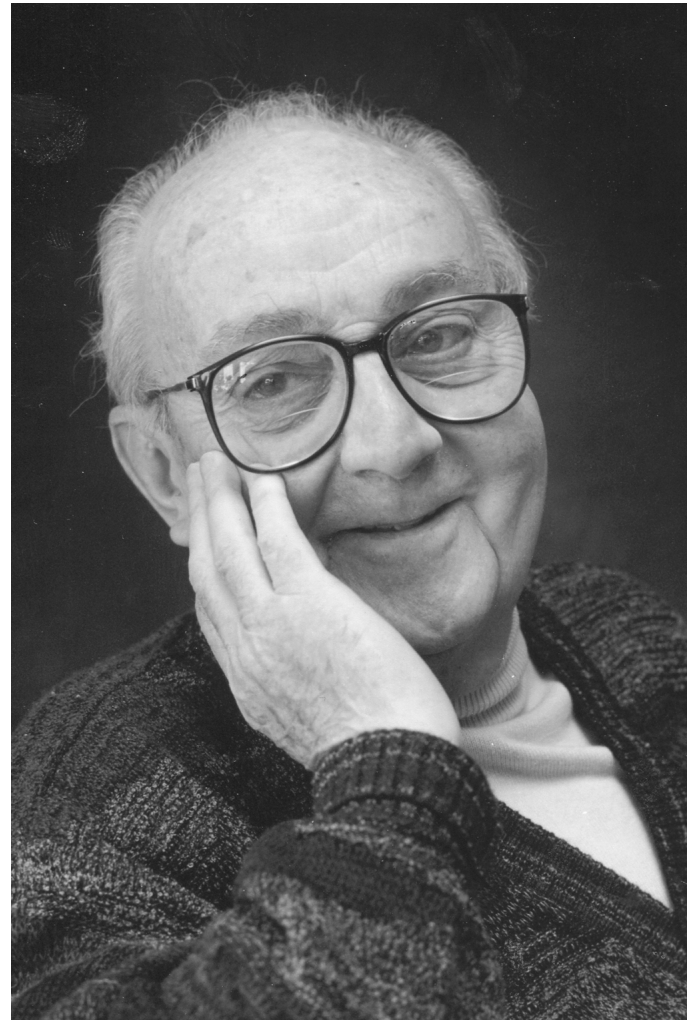


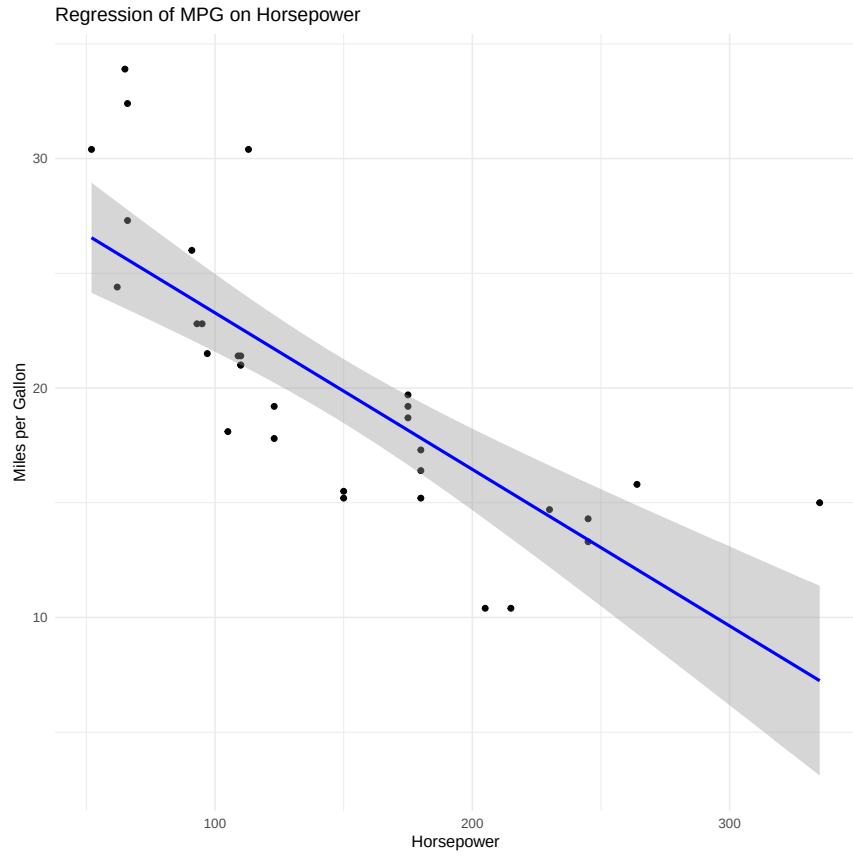
Figure 2: By DavidMCEddy at en.wikipedia, CC BY-SA 3.0, https://commons.wikimedia.org/w

- “Wrong” here means that models are simplifications of reality and do not capture all the complexities of the system being studied. It does not mean that models are useless.

"All models are wrong, but some are useful"

For example, linear regression models:

- Assume a linear relationship between the dependent and independent variables.
- Assume that the residuals are normally distributed.



Factors that influence model formulation/choice

- Accuracy: how well does the model to reproduce observed data and predict future outcomes?
- Transparency: is it easy to understand and interpret the model and its outputs? (This is affected by the model's complexity)
- Flexibility: the ability of the model to be adapted to different scenarios.

What are models used for?

- Generally, models can be used to predict and understand/explain the dynamics of infectious diseases.

Discussion

How are these two uses impacted by accuracy, transparency, and flexibility?

Prediction of the future course

- Must be accurate else they will provide an incorrect outlook of the future.
- Example of the prediction of Ebola deaths during the outbreak in West Africa in 2014/2015.
- "But the estimate proved to be off. Way, way off. Like, 65 times worse than what ended up happening."

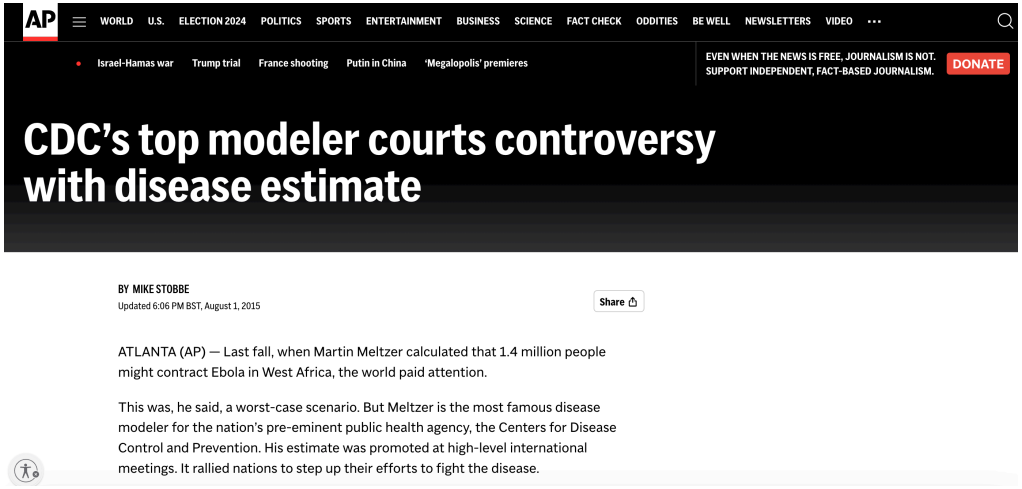


Figure 3: Controversy over estimates of Ebola deaths

< Insert examples of models that have made accurate predictions of the future course of an outbreak >

Understanding or explaining disease dynamics

- Models can be used to understand how a disease spreads and how its spread can be controlled.
- The insights gained from models can be used to:
  - inform public health policy and interventions.
  - design interventions to control the spread of the disease, for example, randomised controlled trials.
  - collect new data.
  - build predictive models.

< Insert examples of models that explain the spread and dynamics of infectious diseases >

< Insert examples of models that evaluate the impact of interventions and determine the next course of action >

Limitations of infectious disease models

- Host behaviour is often difficult to predict.
- The pathogen often has unknown characteristics or known characteristics that are difficult to model.
- Data is often not available or is of poor quality.

- Models:
  - Simplifications of reality and do not capture all the complexities of the system being studied.
  - Only as good as the data used to parameterize them.
  - Can be sensitive to the assumptions made during their formulation.
  - Can be computationally expensive and require a lot of data to run.
  - Can be difficult to interpret and communicate to non-experts.

Introduction to compartmental models

What are compartmental models?

- Compartmental models:
  - divide populations into compartments (or groups) based on the individual's infection status and track them through time (Blackwood and Childs 2018).
  - are mechanistic, meaning they describe processes such the interaction between hosts, biological processes of pathogen, host immune response, and so forth.

- Individuals in a compartment:
  - are assumed to have the same features (disease state, age, location, etc)
  - can only be in one compartment at a time.
  - move between compartments based on defined transition rates.

- Common compartments:
  - Susceptible (S) - hosts are not infected but can be infected
  - Infected (I) - hosts are infected (and can infect others)
  - Removed (R) - hosts are no longer infected and cannot be re-infected

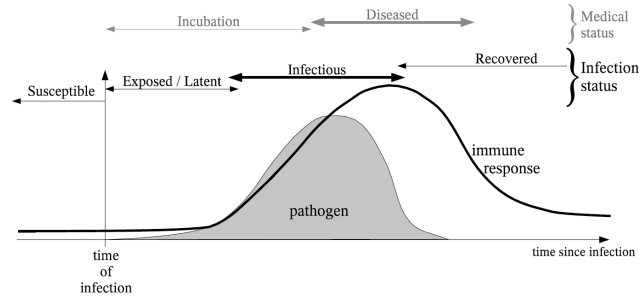


Figure 4: Infection timeline illustrating how a pathogen in a host interacts with the host's immune system (Source: Modelling Infectious Diseases of Humans and Animals)

- Other compartments can be added to the model to account for important events or processes (e.g., exposed, recovered, vaccinated, etc.)
- It is, however, important to keep the model simple, less computationally intensive, and interpretable.

- Compartmental models either have *discrete* or *continuous* time scales:
  - Discrete time scales: time is divided into discrete intervals (e.g., days, weeks, months).
  - Continuous time scales: time is continuous and the model is described using differential equations.

- Compartmental models can be *deterministic* or *stochastic*:
  - Deterministic models always return the same output for the same input.
  - Stochastic models account for randomness in the system and model output always varies. Hence, they are often run multiple times to get an average output.

- The choice of model type depends on the research question, data availability, and computational resources.
- In this introduction, we will focus on deterministic compartmental models with continuous time scales.

- Now, back to the models, we are going to consider infections that either confer immunity after recovery or not.
- The simplest compartmental models for capturing this are the SIS and SIR models.

The Susceptible-Infected-Recovered (SIR) model

- Used to model diseases that confer immunity after recovery. For example, measles and chicken pox.
- Popularised by Kermack and McKendrick in 1927 (Kermack and McKendrick 1927)
  - A must-read paper for budding infectious disease modellers.

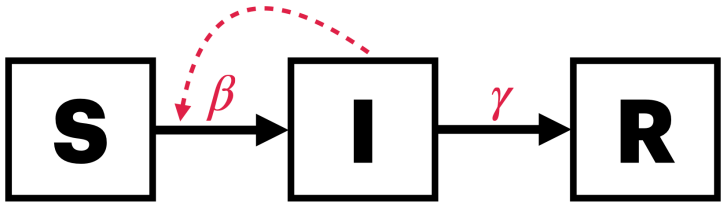


Figure 5: Diagram of an SIR model

- The SIR model groups individuals into three *disease states*:
  - Susceptible (S): individuals who are not infected and can be infected.
  - Infected (I): individuals who are infected and can infect others.
  - Removed (R): individuals who have recovered from the infection and are immune.

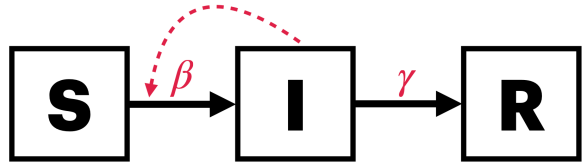


Figure 6: Diagram of an SIR model

How do individuals move between compartments?

- There are two processes that govern the transitions between compartments:
  - Transmission, governed by the transmission rate, β (rate at which susceptible individuals become infected).
  - Recovery, governed by the recovery rate, γ (rate at which infected individuals recover and become immune).

Model assumptions

- There are no inflows or outflows from the population, i.e., no births, deaths, or migration. This is often described as a *closed population*.
  - That is, the epidemic occurs much faster than the time scale of births, deaths, or migration.
- Individuals are infectious immediately after infection and remain infectious until they recover.

- Mixing is *homogeneous*, i.e., individuals mix randomly and have an equal probability of coming into contact with any other individual in the population.
- Transition rates are constant and do not change over time.
- Individuals acquire “lifelong” immunity after recovery.

Formulating the model equations

- Continuous time compartmental models are often formulated using differential equations that describe the change in the number of individuals in each compartment over time.
- The SIR model can be formulated by identifying the events/processes that cause individuals to move between compartments:
  - Transmission: susceptible individuals become infected.
  - Recoveries: infected individuals recover.
- Let's break these two processes down further.

Process 1: Transmission

Discussion

- What factors drive transmission?

- Transmission is driven by several factors:
  - The prevalence, I, i.e., number of infected individuals at the time.
  - The number of contacts, C, susceptible individuals have with infected individual.
  - The probability, p, a susceptible individual will become infected when they contact an infected individual.

- The transmission term is often defined through the force of infection (FOI), λ:
  - the per capita rate at which susceptible individuals become infected.
- The rate at which new infecteds are generated is given by λS, where S is the number of susceptible individuals.
- The force of infection is proportional to the number of infected individuals and the transmission rate, β.
- β is the product of the contact rate and the probability of transmission per contact.

- The FOI can be formulated in two ways, depending on how the contact rate is expected to change with the population size:
  - *Frequency-dependent/mass action transmission*: The rate of contact between individuals is proportional to the population size. Here, $\lambda = \beta \times \frac{I}{N}$.
  - *Density dependent transmission*: The rate of contact between individuals is independent of the population size. Here, $\lambda = \beta \times I$.

- Frequency-dependent transmission assumes that that the number of contacts an individual has does not depend on the population size:
  - Here, the interaction between susceptible and infected individuals does not change with the population size.
  - Often used to model sexually-transmitted diseases and diseases with heterogeneity in contact rates.
  - Sexual transmission in this case does not depend on how many infected individuals are in the population.

- Density-dependent transmission assumes that the number of contacts an individual has is proportional to the population size:
  - It impacts the interaction between susceptible and infected individuals.
  - Can be used to model airborne and directly transmitted diseases, for example, measles.

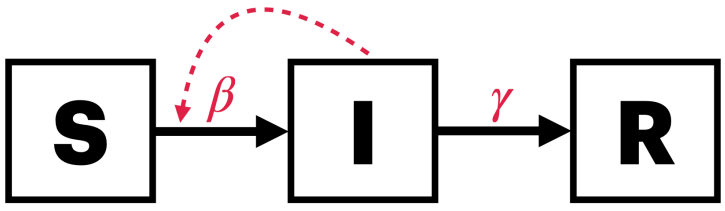
Note

- The differences become important when dealing with population sizes that change significantly over time.
- We will mostly use the density-dependent formulation in this course.

Process 2: Recovery

- Individuals recover after having been infected for the duration of infection.
- The recovery rate, γ, is the reciprocal of the average duration of the infection.
- Infected individuals spend $1/\gamma$ days in the infected compartment before recovering.
- The average infectious period is often estimated from epidemiological data:
  - You will learn about parameter estimation in a future lecture.

Putting it all together, the SIR model can be formulated as a set of differential equations:



$$\frac{dS}{dt} = -\beta \frac{SI}{N} \tag{1}$$

$$\frac{dI}{dt} = \beta \frac{SI}{N} - \gamma I \tag{2}$$

$$\frac{dR}{dt} = \gamma I \tag{3}$$

where β is the transmission rate, and γ is the recovery rate.

With initial conditions $S(0) = N - 1$, $I(0) = 1$, and $R(0) = 0$.

where N is the total population size.

Some things to note

- We represent the compartments as population sizes.:
  - Some modellers often use proportions instead of population sizes as a way to remove the dimensions from the equations.

What questions can we answer with the SIR model?

- The SIR model can be used to understand the dynamics of an epidemic:
  - How long will the epidemic last?
  - How many individuals will be infected (final epidemic size)?
  - When will the epidemic reach its peak?

Solving the SIR model

- Compartmental models cannot be solved analytically.
- Other methods are used to understand the dynamics of the model:
  - Qualitative analysis of the equations; threshold phenomena.
  - Numerical simulations.

Threshold phenomena

- Here, we study the conditions under which an epidemic will grow or die out using the model equations.
- Consider the case where $I(0) = 1$ individual is introduced into a population of size N at time $t = 0$.
- That means in a completely susceptible population, we have $S(0) = N - 1$ susceptible individuals.

- At time 0, the disease will not spread if the rate of change of infections is negative, that is $\frac{dI}{dt} < 0$.
- Recall from the SIR model that $\frac{dI}{dt} = \beta \frac{SI}{N} - \gamma I$
- Let's solve this equation at $t = 0$ by setting $S = N - 1$ and $I = 1$, assuming $\frac{dI}{dt} < 0$.

- At $t = 0$, we have

$$\frac{dI}{dt} = \beta \times \frac{N-1}{N} \times 1 - \gamma I < 0$$

- Re-arranging the equation, we get

$$\frac{\beta}{\gamma} \times \left(\frac{N-1}{N}\right) < 1$$

- For very large populations, $\frac{N-1}{N} \approx 1$, and the equation simplifies to $\frac{\beta}{\gamma} < 1$.
- This implies that the epidemic will not grow if $\frac{\beta}{\gamma} < 1$.
- This is the threshold condition for the epidemic to die out. The epidemic will grow if $\frac{\beta}{\gamma} > 1$.
- For the SIR model, the quantity $\frac{\beta}{\gamma}$ is called the reproduction number, $R0$ (pronounced "R naught" or "R zero").

The basic reproduction number, R0

- The basic reproduction number, $R0$, is the average number of secondary infections generated by a single primary infection in a completely susceptible population.
- The basic reproduction number is a key quantity in infectious disease epidemiology.

- It is often represented as a single number or a range of high-low values.
  - For example, the $R0$ for measles is popularly known to be 12 - 18 (Guerra et al. 2017).
- For an insightful historical account of the evolution of $R0$ derivations, see (Heesterbeek 2002).

- $R0$ is often used to express the threshold phenomena in infectious disease epidemiology:
  - If $R0 > 1$, the epidemic will grow.
  - If $R0 < 1$, the epidemic will decline.
- A pathogen's $R0$ value is determined by biological characteristics of the pathogen and the host's behaviour.

- Conceptually, $R0$ is given by

$$\begin{aligned}\text{Probability of infection given contact}\times\\ \text{Number of contacts per unit time}\times\\ \text{Duration of infectiousness}\end{aligned}$$

and mathematically,

$$R0 \propto \frac{\text{Infection}}{\text{Contact}} \times \frac{\text{Contact}}{\text{Time}} \times \frac{\text{Time}}{\text{Infection}}$$

- $R0$ is therefore unitless and dimensionless.

The next-generation matrix approach

- The method used to derive $R0$ in the SIR model is possible because there is only one compartment with infected individuals.
- In more complex models, $R0$, this method becomes impossible to use.
- A more general method would be to use the next-generation matrix approach.

<Insert slides on next-generation matrix approach for deriving $R0$ for SEIR>

Numerical simulations

- Numerical simulations can be performed with any programming language.
- This course focuses on the R programming language:
  - In R, we can use the `deSolve` package to solve the differential equations.

- To solve the model in R, we will always need to define at least three things:
  - The model equations.
  - The initial parameter values.
  - The initial conditions (population sizes).

- Let's start by defining the model equations.

```
# Define the SIR model
sir_model <- function(t, state, parameters) {
  with(as.list(c(state, parameters)), {
    dS <- -beta * S * I
    dI <- beta * S * I - gamma * I
    dR <- gamma * I
    return(list(c(dS, dI, dR)))
  })
}
```

Next, we will define the parameter values and initial conditions.

```r
# define parameters we know
N  <- 1
R0 <- 10
infectious_period <- 7

# Remember gamma <- 1/ infectious_period as discussed earli
gamma <- 1/infectious_period

# We will use R0 = beta N / gamma  instead because it is ea
# beta is not directly interpretable.

params <- c(beta = R0 * gamma / N, gamma = gamma)

# Initial conditions for S, I, R
# Why is S = N - 1?
inits <- c(S = N - 0.001, I = 0.000001, R = 0)

# Time steps to return results
dt <- 1:70
```

Finally, we will solve the model using the `ode()` function from the `deSolve` package. For now, we will use the default values of the function. You are encouraged to explore the documentation to understand the function better.

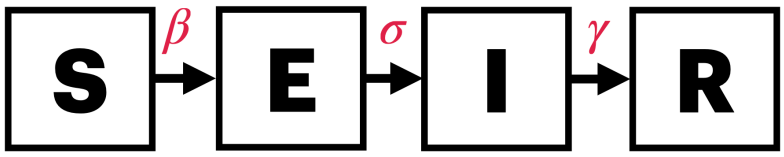
```r
# Solve the model
results <- deSolve::ode(
  y = inits,
  times = dt,
  func = sir_model,
  parms = params
)

# Make it a data.frame
results <- as.data.frame(results)
```

Now, let's plot the results.

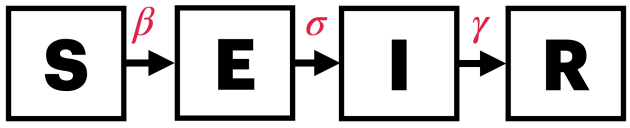
```
# Load the necessary libraries
library(dplyr)
library(tidyr)
library(ggplot2)
# Create data for ggplot2 by reshaping
results_long <- results |>
  pivot_longer(
    cols = c(2:4),
    names_to = "compartment",
    values_to = "value"
  )

plot <- ggplot(
  data = results_long,
  aes(
    x = time,
    y = value,
    color = compartment
  )
```

The Susceptible-Exposed-Infected-Recovered (SEIR) Model

- Many diseases have an incubation period during which individuals are infected but not yet infectious. Examples include COVID-19 and Ebola.
- Disease transmission can occur during the incubation period, but the individual is not yet infectious.
- Capturing this incubation period is essential for understanding the dynamics of the disease and the impact of control measures.



- The SEIR model extends the SIR model to include an exposed compartment, E.
- The exposed compartment represents individuals who have been infected but are not yet infectious.
- Individuals stay in the exposed compartment for a period of time $(1/\sigma)$ before moving to the infectious compartment.



Model equations:

$$\frac{dS}{dt} = -\beta \frac{SI}{N} \tag{4}$$

$$\frac{dE}{dt} = \beta \frac{SI}{N} - \sigma E \tag{5}$$

$$\frac{dI}{dt} = \sigma E - \gamma I \tag{6}$$

$$\frac{dR}{dt} = \gamma I \tag{7}$$

The $R0$ of the SEIR Model

Numerical simulations

- We can use the same approach as the SIR model to simulate the SEIR model.
- Let's do a code walk though in R.

Controlling epidemics

Note

- For a recap on the various control measures, refer to Section 4.

Modelling vaccination

Note

- For a background on vaccination, refer to Section 6.

- Conceptually, vaccination works to reduce the number of susceptible individuals, S.
- The SIR and SEIR model can be extended to include vaccination by adding a new compartment, V.
- Let's consider the SEIR model with vaccination.

The Susceptible-Exposed-Infected-Recovered-Vaccinated (SEIRV) Model

- The SEIRV model is simply the SEIR model with a vaccinated compartment, V.
- The vaccinated compartment represents previously susceptible individuals who have been vaccinated and are immune to the disease.
- The vaccinated compartment is not infectious and does not move to the exposed or infectious compartments.
- The vaccinated compartment is replenished by the rate of vaccination, η.

The model diagram and equations are as follows:

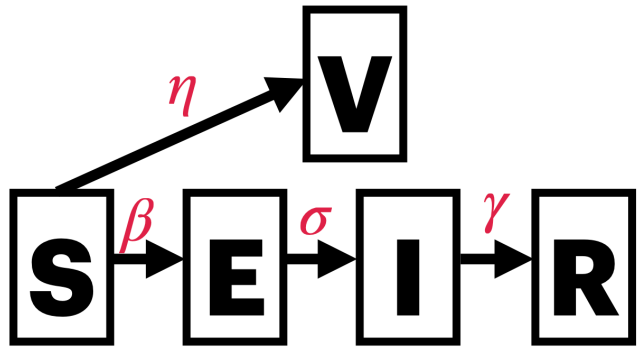
$$\frac{dS}{dt} = -\beta\frac{SI}{N} - \eta S \quad (8)$$

$$\frac{dE}{dt} = \beta\frac{SI}{N} - \sigma E \quad (9)$$

$$\frac{dI}{dt} = \sigma E - \gamma I \quad (10)$$

$$\frac{dR}{dt} = \gamma I \quad (11)$$

$$\frac{dV}{dt} = \eta S \quad (12)$$



where η is the rate of vaccination.

Discussion

- What are some of the assumptions of the SEIRV model?
- What are the implications of these assumptions for the model's predictions?

Numerical simulations

- We can use the same approach as the SIR and SEIR models to simulate the SEIRV model.
- Let's do a code walk though in R.

Modelling NPIs

Note

- For a background on non-pharmaceutical interventions, refer to Section 8.

- Conceptually, NPIs usually act to either reduce the transmission rate, β or prevent infected individuals from transmitting.
- Let's consider two scenarios that will extend the SIR model:
  1. Reducing the transmission rate, β.
  2. Preventing infected individuals from transmitting through isolation.

Reducing the transmission rate

- NPIs such as social distancing, mask-wearing, and hand hygiene can reduce the transmission rate, β.
- We can model this by introducing a reduction factor, α, that reduces transmission rate, β.

The modified SIR model with a reduced transmission rate is as follows:

$$\frac{dS}{dt} = -\alpha\beta\frac{SI}{N} \tag{13}$$

$$\frac{dI}{dt} = \alpha\beta\frac{SI}{N} - \gamma I \tag{14}$$

$$\frac{dR}{dt} = \gamma I \tag{15}$$

where α is the reduction factor, and $0 \leq \alpha \leq 1$.

Scenarios with reduced transmission rate

- We can use the same approach as the SIR model to simulate the model with a reduced transmission rate.
- Let's do a code walk though in R.

Modelling quarantine

- Isolation is a key NPI that prevents infected individuals from transmitting the disease.
- We can model this by introducing a new compartment, Q, for quarantining infected individuals.
- Infected individuals move to the quarantine compartment at a rate, δ.
- Infected individuals in the quarantine compartment do not transmit the disease.

The modified SIR model with quarantine is as follows:

$$\frac{dS}{dt} = -\beta \frac{SI}{N} \tag{16}$$

$$\frac{dI}{dt} = \beta \frac{SI}{N} - \gamma I - \delta I \tag{17}$$

$$\frac{dR}{dt} = \gamma I + \tau Q \tag{18}$$

$$\frac{dQ}{dt} = \delta I - \tau Q \tag{19}$$

where δ is the rate at which infected individuals move to the quarantine compartment, and τ is the rate at which individuals recover from quarantine.

Scenarios with quarantine

- We can use the same approach as the SIR model to simulate the model with quarantine.
- Let's do a code walk though in R.

Incorporating host heterogeneity

- The models we have discussed so far assume that all individuals in the population are identical.
- However, in reality, individuals differ in their susceptibility to infection and their ability to transmit the disease.
- It is essential to capture this heterogeneity in the models in order for the models to be more realistic and useful for decision-making.

- This can be captured by incorporating heterogeneity into the models.
- Heterogeneity is often captured by stratifying the population into different groups.

Incorporating age structure

- For many infectious diseases, the risk of infection and the severity of the disease vary by age.
- Hence, it is essential to capture age structure in the models.
- To do this, we divide the population into different age groups and model the disease dynamics within each age group.
- Let's extend the SIR model to include age structure.

- We will divide the population into n age groups.
- Because the homogeneous model has 3 compartments, the age structured one will have $3n$ compartments: $S_1, I_1, R_1, S_2, I_2, R_2, ..., S_n, I_n, R_n$.

- The (compact) model equations are as follows:

$$\frac{dS_i}{dt} = -\sum_{j=1}^{n} \beta_{ji} S_i I_j \tag{20}$$

$$\frac{dI_i}{dt} = \sum_{j=1}^{n} \beta_{ji} S_i I_j - \gamma I_i \tag{21}$$

$$\frac{dR_i}{dt} = \gamma I_i \tag{22}$$

- The model can be used to study the impact of age structure on the dynamics of the epidemic.
- For example, we can study the impact of vaccinating different age groups on the dynamics of the epidemic.

Other Heterogeneities

- Other forms of heterogeneity that can be incorporated into the models include:
  - Spatial heterogeneity
  - Temporal heterogeneity
  - Contact heterogeneity

Wrap-Up and Q&A

Takeaways

In the last two days, we have covered a lot of ground. Here are some key takeaways:

- Infectious diseases are a major public health concern that can have devastating consequences.
- Mathematical models are essential tools for studying the dynamics of infectious diseases and informing public health decision-making.
- Models are simplifications of reality that help us understand complex systems.

- We have discussed several compartmental models, including the SIR, SEIR, and SEIRV models.
- The SIR model is a simple compartmental model that divides the population into three compartments: susceptible, infected, and recovered.
- The SEIR model extends the SIR model by adding an exposed compartment.
- We can model various pharmaceutical and non-pharmaceutical interventions (NPIs) by modifying the transmission rate or adding new compartments.

- We have discussed the basic reproduction number, $R0$, which is a key parameter in infectious disease epidemiology.
- $R0$ is the average number of secondary infections produced by a single infected individual in a completely susceptible population.
- If $R0 > 1$, the disease will spread in the population; if $R0 < 1$, the disease will die out.

- Deriving the basic reproduction number, $R0$, is an essential step in understanding the dynamics of infectious diseases.
- Deriving $R0$ for the simple SIR model is simple as we just need to study the threshold phenomena.
- For more complex models, we can use the next-generation matrix approach to derive $R0$.

- Often, homogeneous models are not sufficient to capture the complexity of infectious diseases.
- Incorporating heterogeneity into the models is essential for capturing the complexity of infectious diseases.
- Age structure is a common form of heterogeneity that can be incorporated into the models.
- Other forms of heterogeneity include spatial, temporal, and contact heterogeneity.

What skills are needed to build and use infectious disease models?

Mathematical and statistical skills

- Differential equations.
- Probability and statistics.
- Stochastic processes.
- Numerical analysis
- Time series analyses
- Survival analysis

Programming skills

- Proficiency in at least one programming language (e.g., R, Python, Julia, C++).
- Experience with version control (e.g., Git).
- Experience with data manipulation and visualization.

Domain knowledge in infectious diseases

- Immunology
- Epidemiology
- Virology
- Genomics

Other subject areas

- Communication skills:
  - Ability to communicate complex ideas to non-experts.
  - Ability to write clear and concise reports.
  - Ability to present results to a diverse audience.
- Public health.
- Health economics.
- Policy analysis.

List of Resources

Textbooks

- Modeling Infectious Diseases in Humans and Animals by Matt Keeling and Pejman Rohani -Epidemics: Models and Data Using R by Ottar N. Bjornstad
- Infectious Disease Modelling by Emilia Vynnycky and Richard White
- Infectious Diseases of Humans: Dynamics and Control by Roy M. Anderson and Robert M. May

Papers and Articles

Modelling infectious disease transmission

- Kirkeby, C., Brookes, V. J., Ward, M. P., Dürr, S., & Halasa, T. (2021). A practical introduction to mechanistic modeling of disease transmission in veterinary science. Frontiers in veterinary science, 7, 546651.
- Blackwood, J. C., & Childs, L. M. (2018). An introduction to compartmental modeling for the budding infectious disease modeler.
- Grassly, N. C., & Fraser, C. (2008). Mathematical models of infectious disease transmission. Nature Reviews Microbiology, 6(6), 477–487.

- Cobey, S. (2020). Modeling infectious disease dynamics. Science, 368(6492), 713–714.
- Bjørnstad, O. N., Shea, K., Krzywinski, M., & Altman, N. (2020). Modeling infectious epidemics. Nature Methods, 17(5), 455–456.
- Bodner, K., Brimacombe, C., Chenery, E. S., Greiner, A., McLeod, A. M., Penk, S. R., & Soto, J. S. V. (2021). Ten simple rules for tackling your first mathematical models: A guide for graduate students by graduate students. PLOS Computational Biology, 17(1), e1008539.
- Mishra, S., Fisman, D. N., & Boily, M.-C. (2011). The ABC of terms used in mathematical models of infectious diseases. Journal of Epidemiology & Community Health, 65(1), 87–94.

- James, L. P., Salomon, J. A., Buckee, C. O., & Menzies, N. A. (2021). The Use and Misuse of Mathematical Modeling for Infectious Disease Policymaking: Lessons for the COVID-19 Pandemic. 41(4), 379–385.
- Holmdah, I., & Buckee, C. (2020). Wrong but useful—What COVID-19 epidemiologic models can and cannot tell us. New England Journal of Medicine.
- Metcalf, C. J. E. E., Edmunds, W. J., & Lessler, J. (2015). Six challenges in modelling for public health policy. Epidemics, 10(2015), 93–96.
- Roberts, M., Andreasen, V., Lloyd, A., & Pellis, L. (2015). Nine challenges for deterministic epidemic models. Epidemics, 10(2015), 49–53.

Deriving and Interpreting R0

- Jones, J. H. (2011). Notes On R0. Building, 1–19.
- Diekmann, O., Heesterbeek, J. A. P., & Metz, J. A. J. (1990). On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations. Journal of Mathematical Biology, 28(4), 365–382.
- Diekmann, O., Heesterbeek, J. A. P., & Roberts, M. G. (2010). The construction of next-generation matrices for compartmental epidemic models. Journal of the Royal Society Interface, 7(47), 873–885.

References

Blackwood, Julie C., and Lauren M. Childs. 2018. "An Introduction to Compartmental Modeling for the Budding Infectious Disease Modeler." *Letters in Biomathematics* 5 (1): 195–221. https://doi.org/10.1080/23737867.2018.1509026.

Box, GE. 1979. "All Models Are Wrong, but Some Are Useful." *Robustness in Statistics* 202 (1979): 549.

Guerra, Fiona M, Shelly Bolotin, Gillian Lim, Jane Heffernan, Shelley L Deeks, Ye Li, and Natasha S Crowcroft. 2017. "The Basic Reproduction Number (R0) of Measles: A Systematic Review." *The Lancet Infectious Diseases* 17 (12): e420–28.

Heesterbeek, Johan Andre Peter. 2002. "A Brief History of r 0 and a Recipe for Its Calculation." *Acta Biotheoretica* 50 (3): 189–204.

Kermack, William Ogilvy, and Anderson G McKendrick. 1927. "A Contribution to the Mathematical Theory of Epidemics." *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 115 (772):