Infectious Disease Modeling

1 October 2025

Achangwa Chiara, M.Sc., Ph.D.
Research Assistant Professor
ciaraacha@gmail.com
Department of Preventive Medicine,
The Catholic University of Korea, Seoul, Korea.

1

Lecture Content

1. Introduction to Infectious Disease Modelling

2. Estimation of key epidemiological Parameters using models

3. Probability Distribution and Maximum Likelihood Estimation

4. Practice in R

5. Practice with a developed Shiny application

2

Requirements

R

R studio

Internet browser

3

Basic Concepts

Infectious/communicable/transmissible diseases are illnesses caused by pathogens such as viruses, bacteria, fungi, and parasites that can spread from one person (or animal) to another.

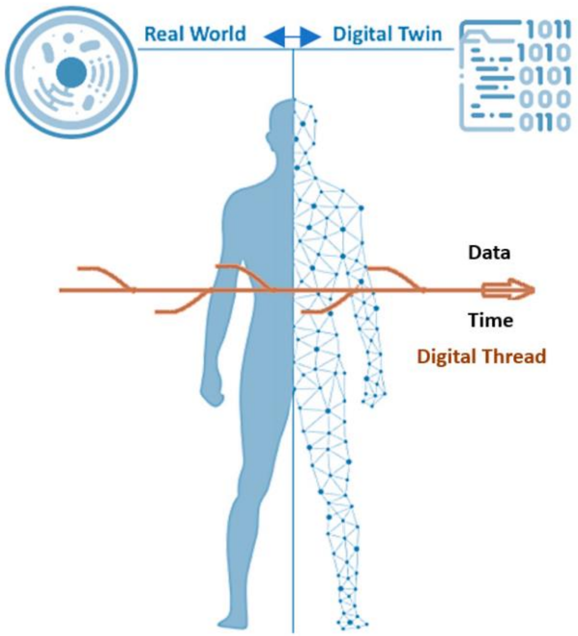
Disease	Pathogen Type	Transmission route
Tuberculosis	Bacteria	Airborne droplets
HIV/AIDS	Virus	Blood, sexual contact
Malaria	Protozoa	Mosquito bite
Influenza	Virus	Airborne/contact
Hepatitis B & C	Virus	Blood, bodily fluids
Measles	Virus	Airborne droplets
Cholera	Bacteria	Contaminated water/food

4

Basic Concepts

Modelling

The use of mathematical and computational tools to simulate how diseases spread, predict future outbreaks, and evaluate the impact of interventions like vaccines or quarantines.



5

Basic Concepts

Infectious disease + Modelling = Infectious Disease Modelling

01

Understanding transmission: Estimates epidemiologic disease parameters e.g Incubation period, serial interval, R0 and Rt

02

Policy planning: Informs decisions on lockdowns, vaccination campaigns, and resource allocation.

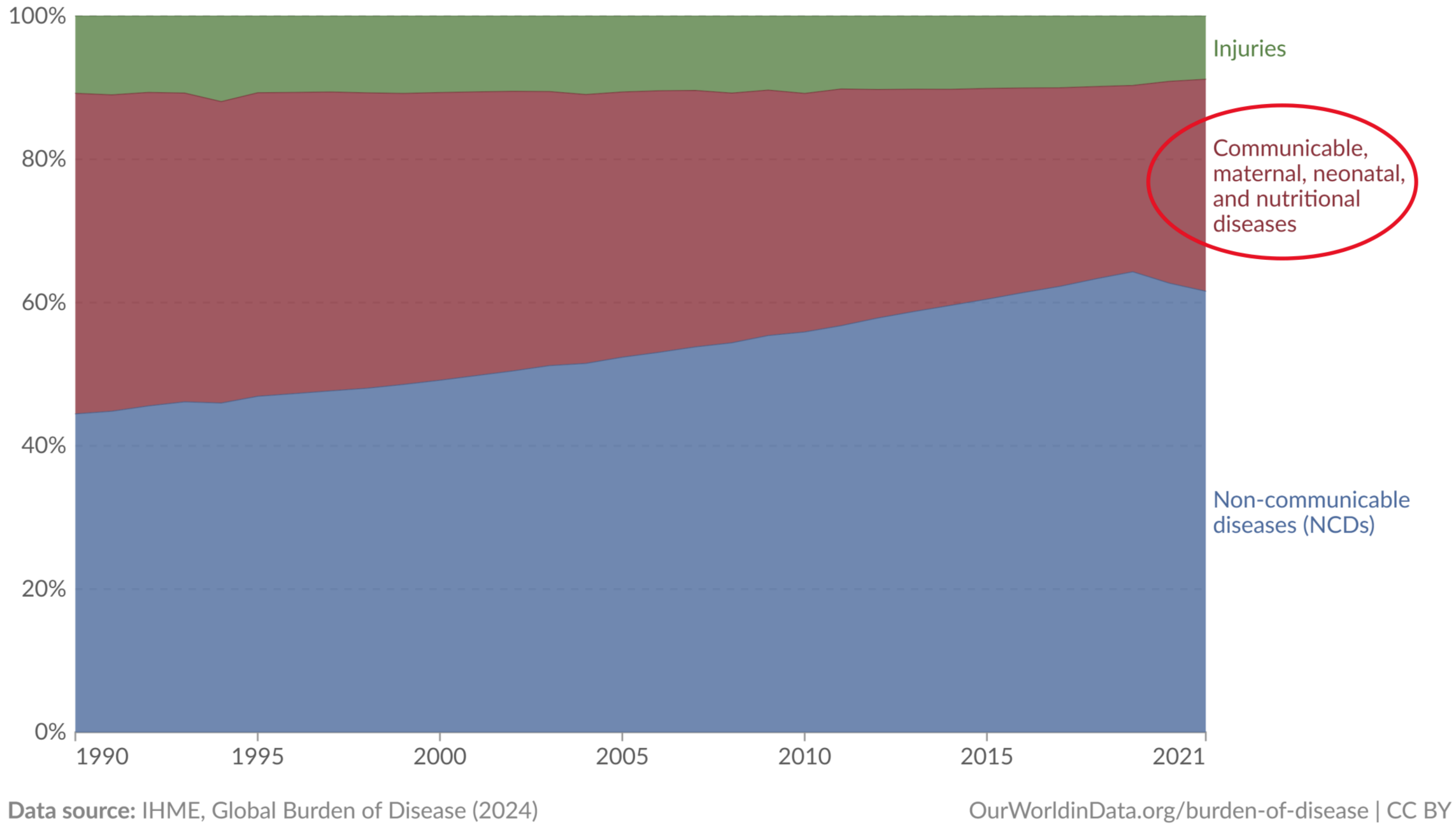
03

Forecasting outbreaks: Helps anticipate how fast and far a disease might spread.

6

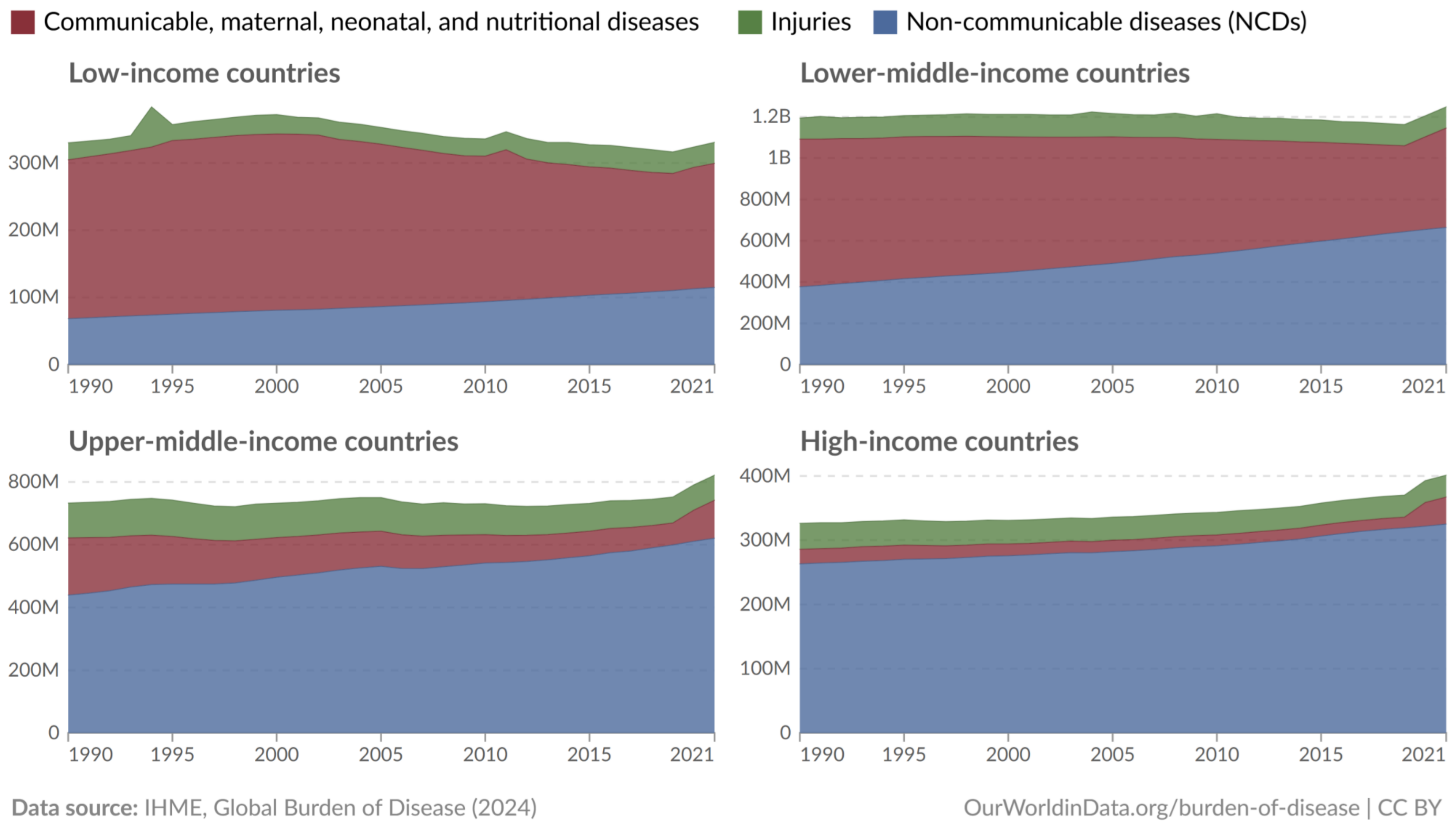
Introduction to infectious disease modelling

✓ Total burden by cause



7

Introduction to infectious disease modelling

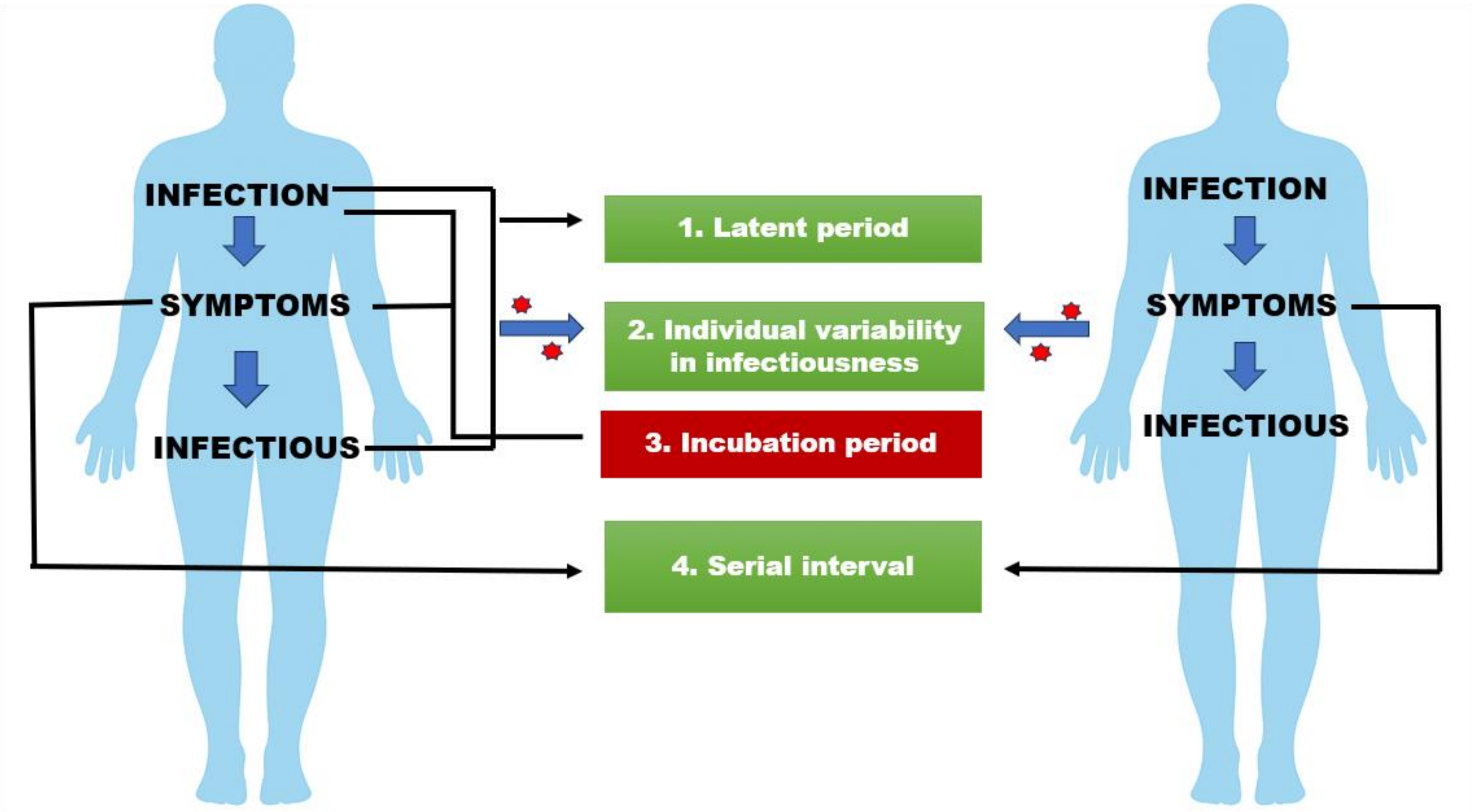


8

Estimation of Infectious Disease Parameters through Models

1. Incubation period

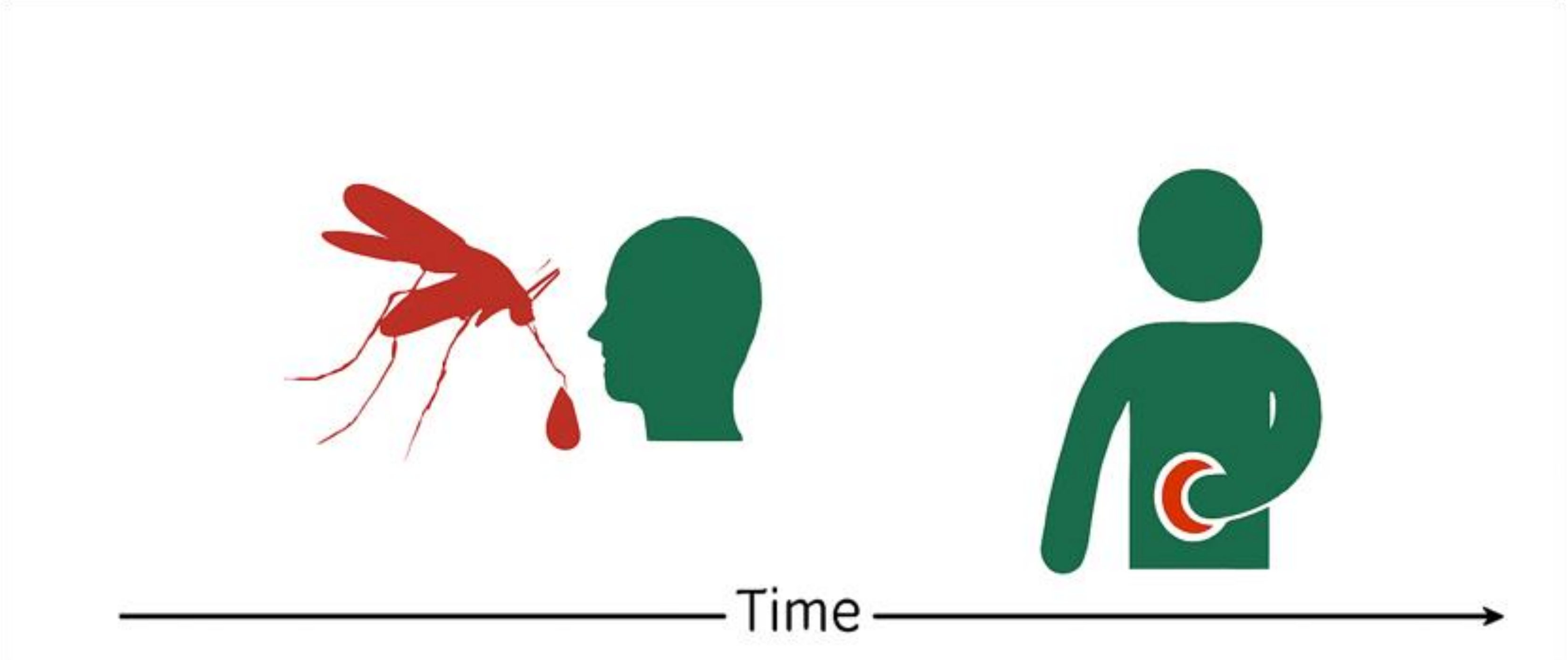
Time between exposure to a given pathogen and the development of signs and symptoms of a disease.



Kraemer MUG et al., 2022 Nat Med.

9

Intrinsic Vs Intrinsic Incubation Period



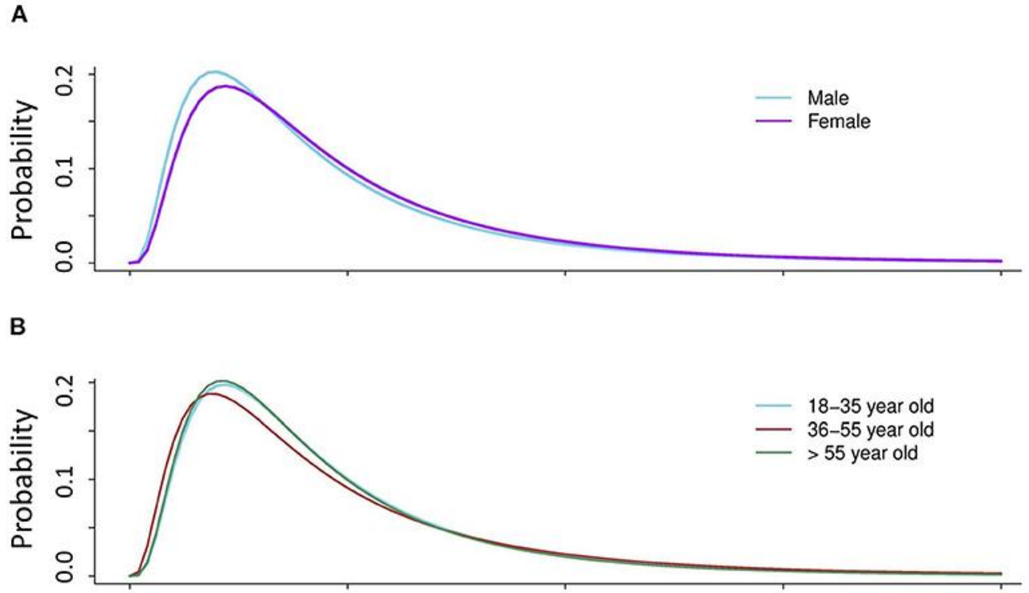
Extrinsic IP: Pathogen development in the Vector

Intrinsic IP: Exposure to symptom development in the host

10

The incubation period of an infectious disease can vary widely among individuals based on;

- ✓ Host (individual) factors: Age, sex, underlying conditions
- ✓ Pathogen-related factors: Dose of exposure, virulence, and strain differences
- ✓ Environmental factors: Route of transmission, temperature, humidity
- ✓ Behavioral factors: concurrent exposures or coinfections
- ✓ Epidemic period

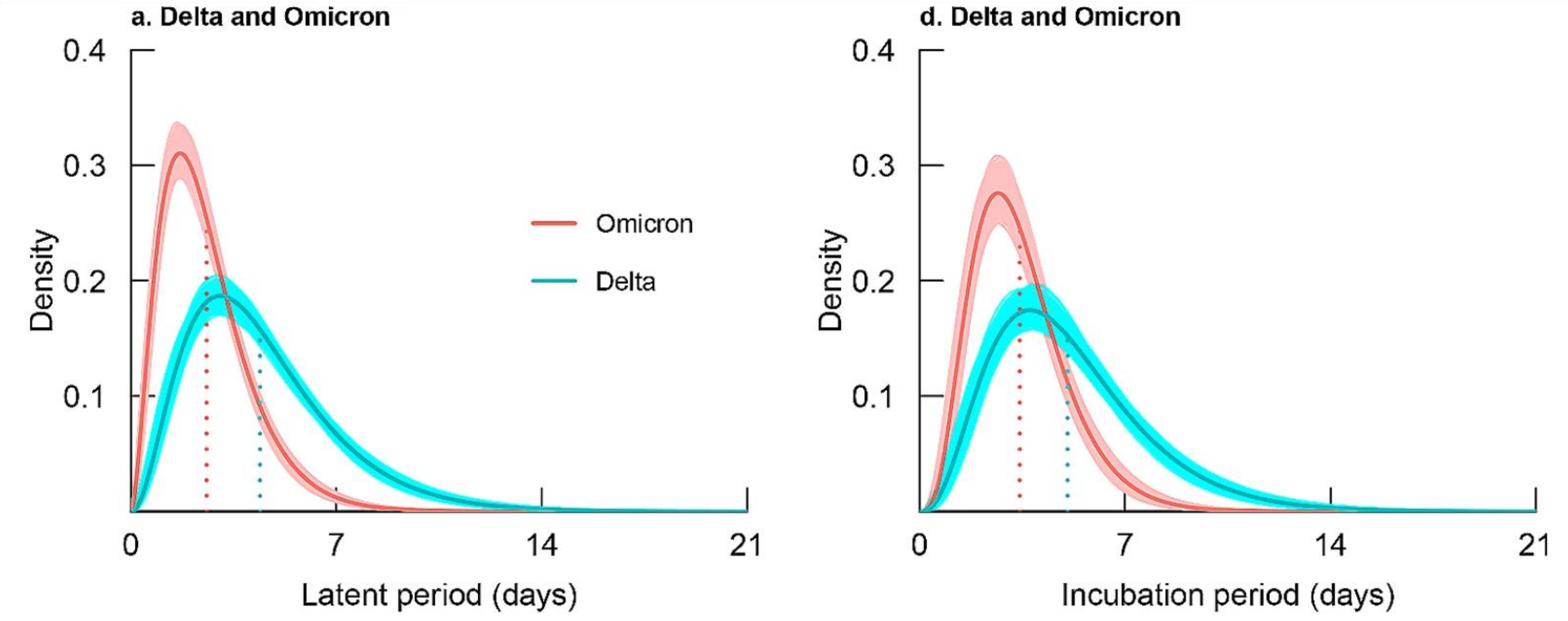


Achangwa et al *Front. Pub Health, 2022*

11

The incubation period of an infectious disease can vary widely among individuals based on;

- ✓ Host (individual) factors: Age, sex, underlying conditions
- ✓ Pathogen-related factors: Dose of exposure, virulence, and strain differences
- ✓ Environmental factors: Route of transmission, temperature, humidity
- ✓ Behavioral factors: concurrent exposures or coinfections
- ✓ Epidemic period

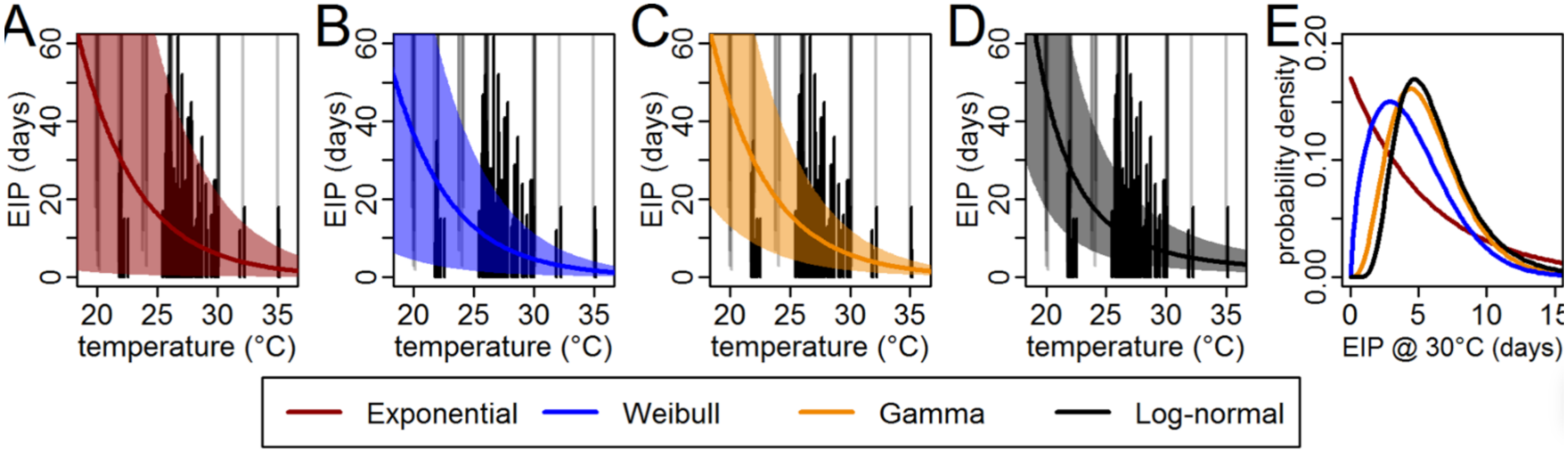


Li et al BMC Inf Dis 2024

12

The incubation period of an infectious disease can vary widely among individuals based on;

- ✓ Host (individual) factors: Age, sex, underlying conditions
- ✓ Pathogen-related factors: Dose of exposure, virulence, and strain differences
- ✓ Environmental factors: Route of transmission, temperature, humidity
- ✓ Behavioral factors: concurrent exposures or coinfections
- ✓ Epidemic period

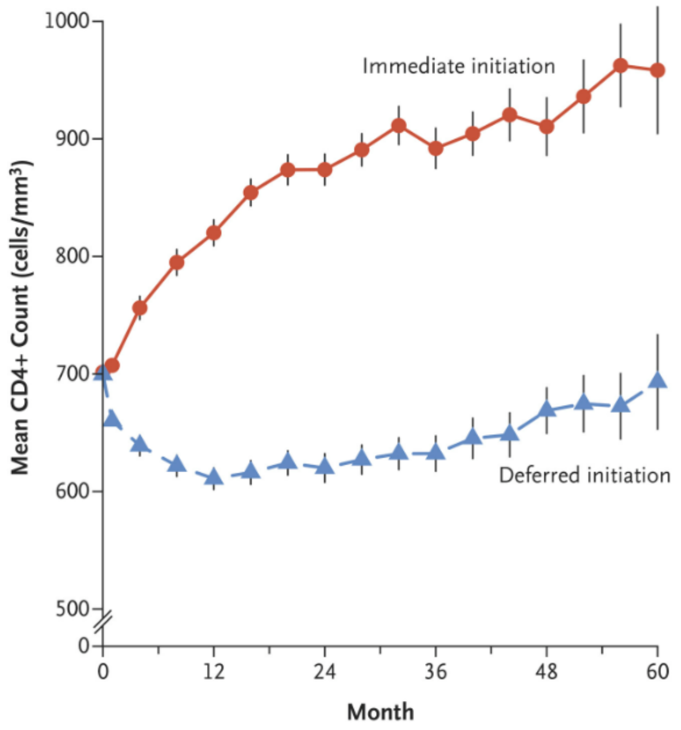


Chan et al *Plos One* 2012

13

The incubation period of an infectious disease can vary widely among individuals based on;

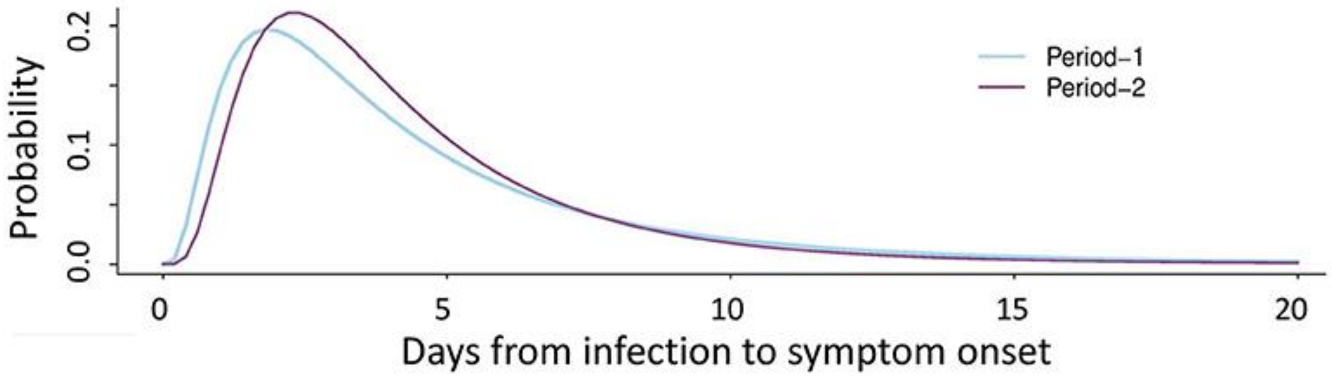
- ✓Host (individual) factors: Age, sex, underlying conditions
- ✓Pathogen-related factors: Dose of exposure, virulence, and strain differences
- ✓Environmental factors: Route of transmission, temperature, humidity
- ✓Behavioral factors: concurrent exposures or coinfections
- ✓Epidemic period



14

The incubation period of an infectious disease can vary widely among individuals based on;

- ✓Host (individual) factors: Age, sex, underlying conditions
- ✓Pathogen-related factors: Dose of exposure, virulence, and strain differences
- ✓Environmental factors: Route of transmission, temperature, humidity
- ✓Behavioral factors: concurrent exposures or coinfections
- ✓Epidemic period



Achangwa et al *Front. Pub Health, 2022*

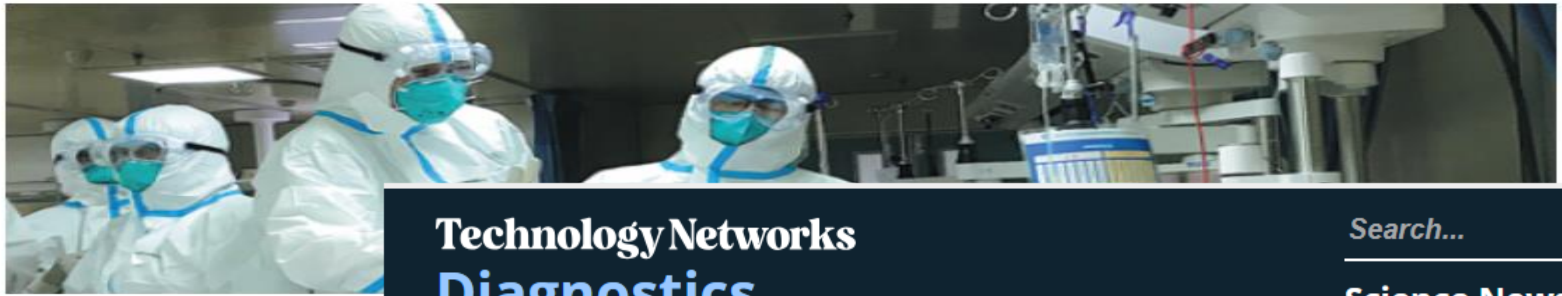
15

Researcher clarifies reported 24-day incubation period for coronavirus

China Daily, February 12, 2020 0 Comment(s) Print E-mail

Adjust font size:

Only one patient infected with the new strain of the coronavirus was found to have an incubation period of as much as 24 days, a scientist said Tuesday.



Technology Networks
Diagnostics

Search...

Science News Communities

Home > Diagnostics > News > Content Piece

Incubation Period for COVID-19 Now Estimated at 5.1 Days

NEWS

What is the incubation period for Covid - and how long is Omicron's incubation period?

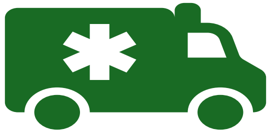
Scientists in South Africa have suggested that Omicron has a shorter incubation period than earlier coronavirus variants

16

Why estimate the incubation period of a disease?

1. For clinical management

To predict disease severity, e.g. shorter incubation time is associated with more severe outcome



2. For public health control

1- To estimate the duration necessary for quarantining suspected or contacts of cases to ensure that they are not infected upon release

2- To identify the origin of common-source outbreaks

3- Backcalculate the *incidence of infection* from the *incidence of report*

17

Estimating the incubation period of a disease

1-Parametric Models

- ✓ Assume a known distribution (log-normal, Weibull, or gamma).
- ✓ Use maximum likelihood estimation (MLE) to fit the model to observed data

2-**Non-Parametric Models**

- ✓ Make no assumptions about the underlying distribution.
- ✓ Useful when data is sparse or highly variable.

3-**Bayesian models**

- ✓ Incorporate prior knowledge and uncertainty.
- ✓ Markov Chain Monte Carlo (MCMC) to estimate incubation densities from coarse data.
- ✓ These are especially powerful when exposure windows are imprecise or censored.

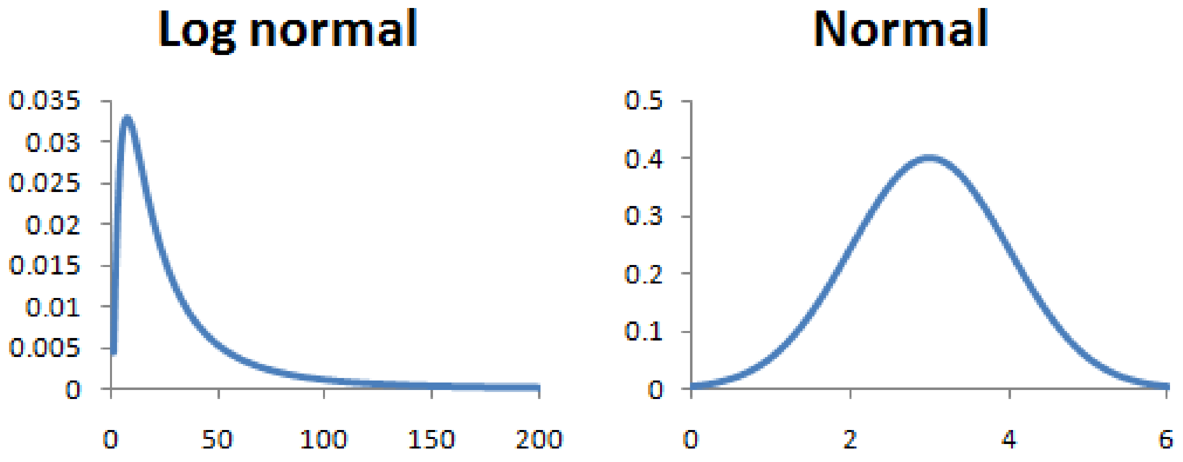
Parametric Models

Feature	Log-normal	Weibull	Gamma
Typical Use	Right-skewed data where log of time is approximately normal	increasing or decreasing symptom-onset rate	Sum of several independent exponential stages (e.g., multiple latent steps)
Parameters	μ: mean of $\ln T$ $\sigma > 0$:SD of $\ln T$	$k > 0$: shape $\lambda > 0$: scale	$\alpha > 0$: shape $\beta > 0$: rate (scale $\theta = 1/\beta$)
PDF $f(t)$	$\frac{1}{t\,\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln t - \mu)^2}{2\sigma^2}\right]$	$\frac{k}{\lambda}\left(\frac{t}{\lambda}\right)^{k-1} e^{-(t/\lambda)^k}$	$\frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\beta t}$
CDF $F(t)$	$\Phi\left(\frac{\ln t - \mu}{\sigma}\right)$	$1 - e^{-(t/\lambda)^k}$	$\frac{\gamma(\alpha, \beta t)}{\Gamma(\alpha)}$
Mean $E[T]$	$e^{\mu + \sigma^2/2}$	$\lambda\,\Gamma\left(1 + \frac{1}{k}\right)$	$\frac{\alpha}{\beta}$
Variance $Var[T]$	$(e^{\sigma^2} - 1)e^{2\mu + \sigma^2}$	$\lambda^2[\Gamma\left(1 + \frac{2}{k}\right) - \Gamma^2\left(1 + \frac{1}{k}\right)]$	$\frac{\alpha}{\beta^2}$
Median	e^μ	$\lambda(\ln 2)^{1/k}$	/
Parameter Estimation	MLE on log-transformed data	MLE on raw times	MLE or method-of-moments ($\hat{\alpha} = \bar{t}^2/s^2$, $\hat{\beta} = \bar{t}/s^2$)

The log-normal model

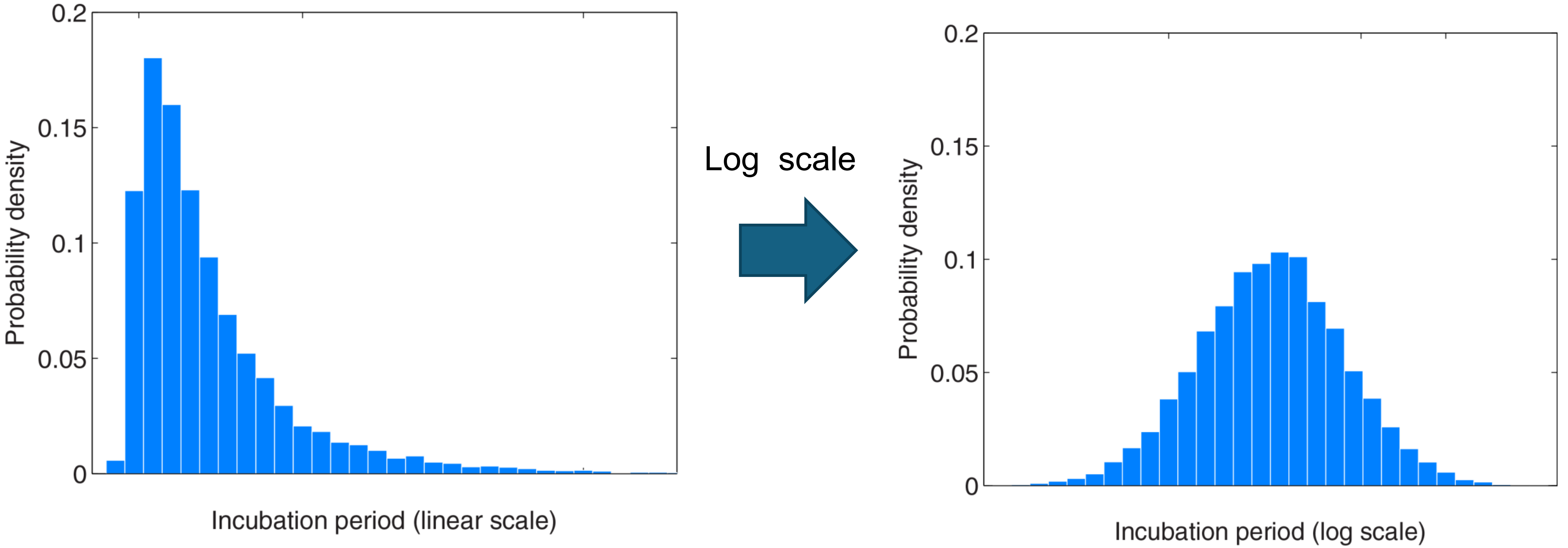
✓ Log-normal distribution if its logarithm is normally distributed.

✓ Log of the time from exposure to symptom onset is modelled as a normal distribution.



Why use the log-normal model?

- ✓Incubation times are always positive and often skewed, making log-normal a natural fit.
- ✓It captures the long tail of delayed symptom onset better than a normal distribution.



21

Structure of the data

id	date_of_exposure	regions	date_of_onset	Source.of.infection	type	country	Transmission.type	Age.group.	sex
#1	2021-11-24	incheon	2021-11-25	11.24 Arrivals from Nigeria(ET672)	abroad	Nigeria	Imported	41	M
#2	2021-11-23	incheon	2021-11-25	Patient #1	National	Domestic	Domestic	45	F
#3	2021-11-23	incheon	2021-11-25	Patient #1	National	Domestic	Domestic	43	M
#4	2021-11-24	incheon	2021-11-25	Patient #1	National	Domestic	Domestic	35	F
#5	2021-11-24	incheon	2021-11-25	Patient #1	National	Domestic	Domestic	51	M
#6	2021-11-25	incheon	2021-11-26	Patient #1	National	Domestic	Domestic	43	F
#7	2021-11-25	incheon	2021-11-26	Patient #1	National	Domestic	Domestic	61	M
#8	2021-11-25	incheon	2021-11-26	Patient #1	National	Domestic	Domestic	33	M
#9	2021-11-25	incheon	2021-11-26	Patient #1	National	Domestic	Domestic	42	F
#10	2021-11-25	incheon	2021-11-26	Patient #1	National	Domestic	Domestic	40	F
#11	2021-11-25	incheon	2021-11-26	Patient #1	National	Domestic	Domestic	34	F
#12	2021-11-25	incheon	2021-11-26	Patient #1	National	Domestic	Domestic	55	F
#13	2021-11-25	incheon	2021-11-26	Patient #1	National	Domestic	Domestic	44	M
#14	2021-11-25	incheon	2021-11-26	Patient #1	National	Domestic	Domestic	46	M
#15	2021-11-25	Daegu	2021-11-26	Patient #1	National	Domestic	Domestic	31	F
#16	2021-11-26	Daegu	2021-11-26	Patient #15	National	Domestic	Domestic	47	M
#17	2021-11-25	Daegu	2021-11-27	Patient #1	National	Domestic	Domestic	39	F
#18	2021-11-25	Daegu	2021-11-27	Patient #1	National	Domestic	Domestic	48	M
#19	2021-11-25	Daegu	2021-11-27	Patient #1	National	Domestic	Domestic	51	F
#20	2021-11-25	Daegu	2021-11-27	Patient #1	National	Domestic	Domestic	46	M
#21	2021-11-25	Daegu	2021-11-27	Patient #1	National	Domestic	Domestic	53	M
#22	2021-11-25	Daegu	2021-11-27	Patient #1	National	Domestic	Domestic	48	F
#23	2021-11-26	Daegu	2021-11-29	Patient #1	National	Domestic	Domestic	56	F
#24	2021-11-26	Daegu	2021-11-29	Patient #1	National	Domestic	Domestic	44	F
#25	2021-11-26	Daegu	2021-11-29	Patient #1	National	Domestic	Domestic	21	F
#26	2021-11-26	Daegu	2021-11-29	Patient #1	National	Domestic	Domestic	25	M
#27	2021-11-26	Daegu	2021-11-29	Patient #1	National	Domestic	Domestic	47	M
#28	2021-11-26	Daegu	2021-11-29	Patient #1	National	Domestic	Domestic	49	F
#29	2021-11-26	Daegu	2021-11-29	Patient #1	National	Domestic	Domestic	37	M
#30	2021-11-26	Daejeon	2021-11-29	Patient #29	National	Domestic	Domestic	42	F
#31	2021-11-26	Daejeon	2021-11-29	Patient #1	National	Domestic	Domestic	43	M
#32	2021-11-26	Daejeon	2021-11-30	Patient #1	National	Domestic	Domestic	59	F
#33	2021-11-26	Daejeon	2021-11-30	Patient #1	National	Domestic	Domestic	51	M

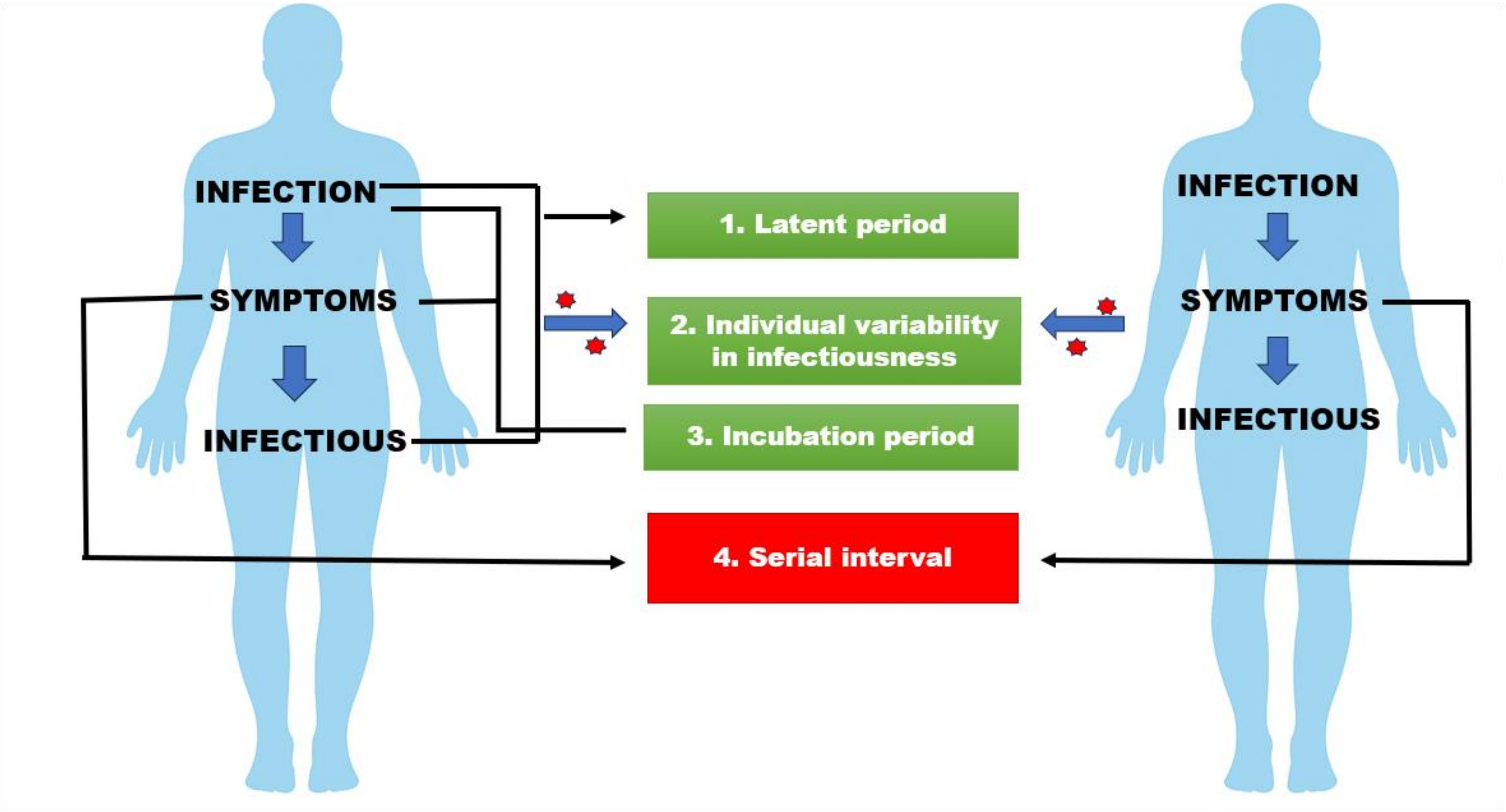
Lee et al., 2020 J Infect Chemo.

22

Estimation of key epidemiological Parameters

2. Serial interval

Time between symptom onset in a primary case (infector) and symptom onset in a secondary case (infectee).

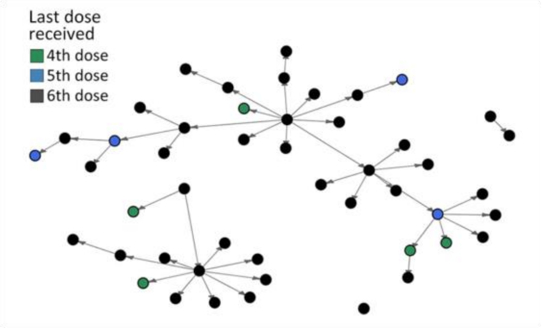


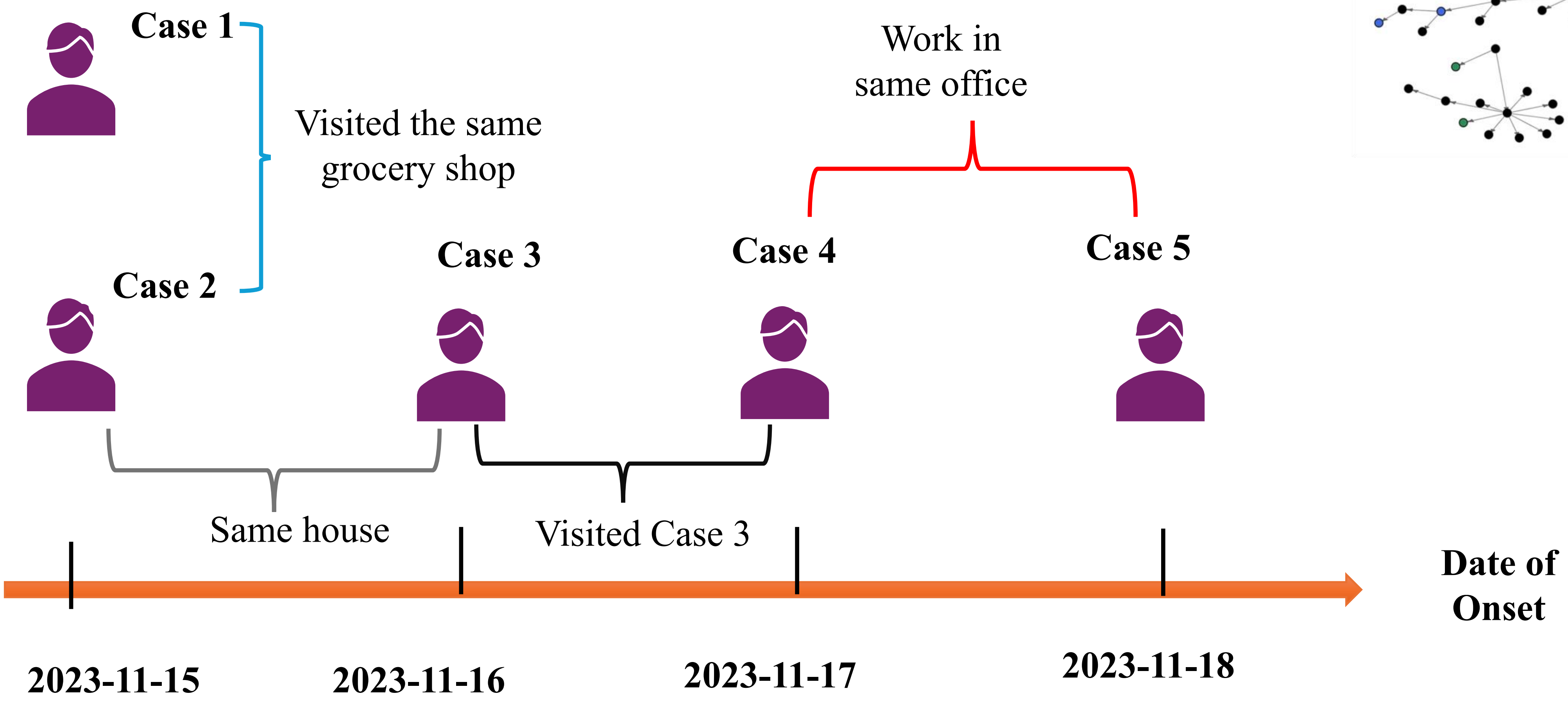
if Person A infects Person B, the serial interval is the time between when A shows symptoms and when B does.

Kraemer MUG et al., 2022 *Nat Med.*

23

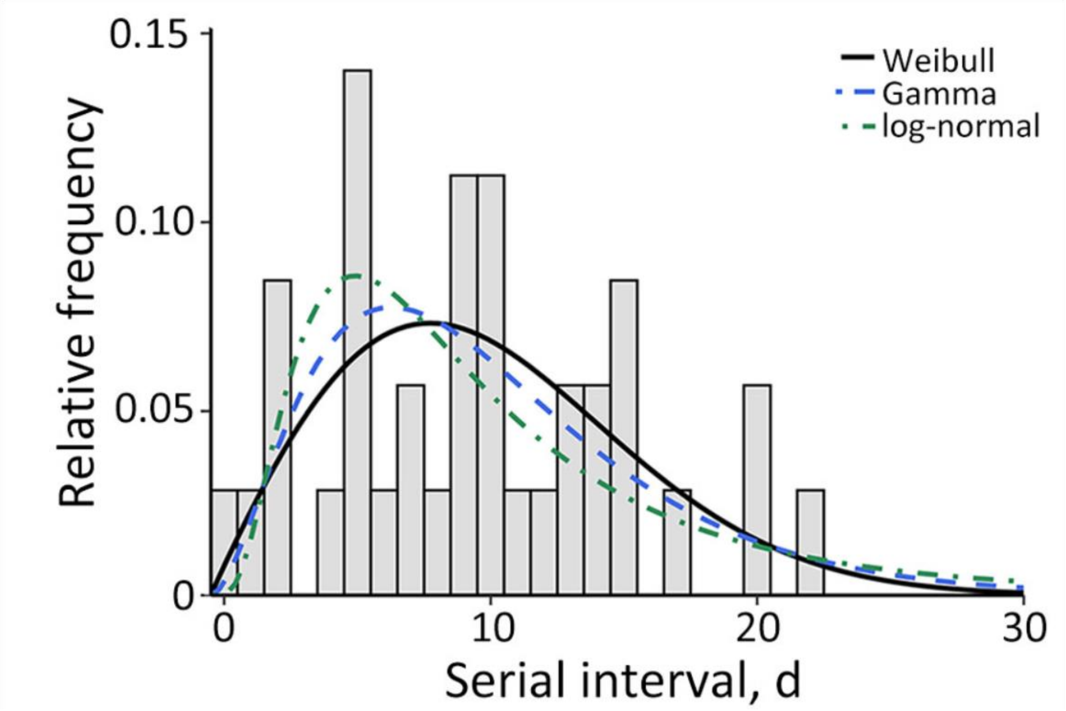
Serial interval



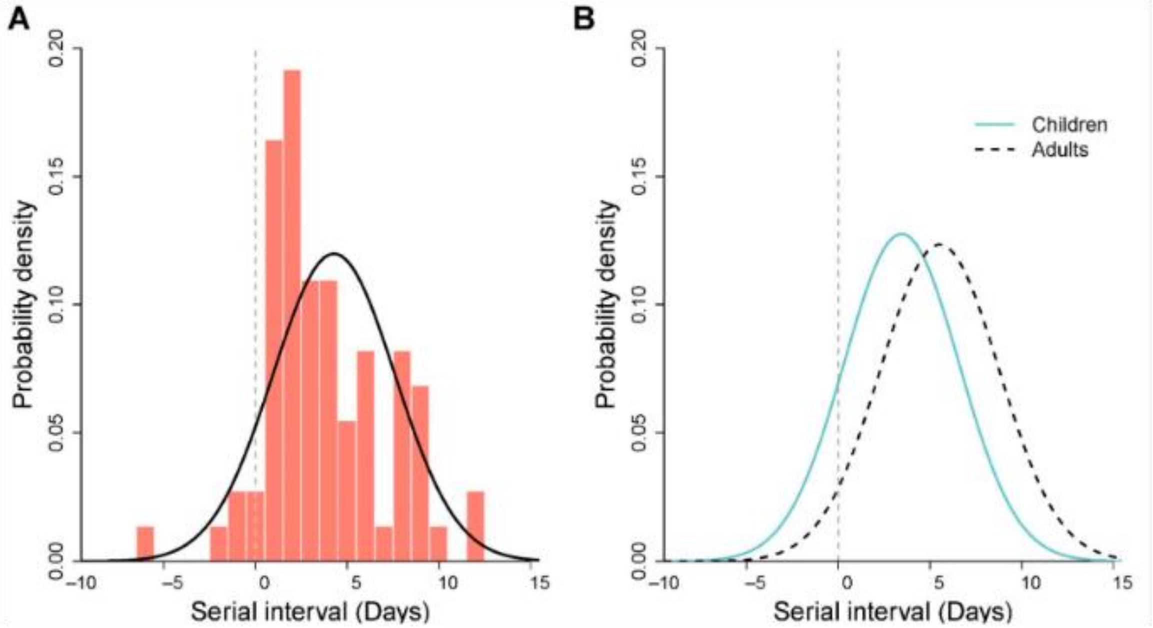


24

Serial interval



Cho et al. 2025 EID



Kim et al. 2022 Viruses

25

Why estimate the serial interval

1- Understanding Transmission patterns in a population

2- Calculating the Reproduction Number (R_0 and R_t)

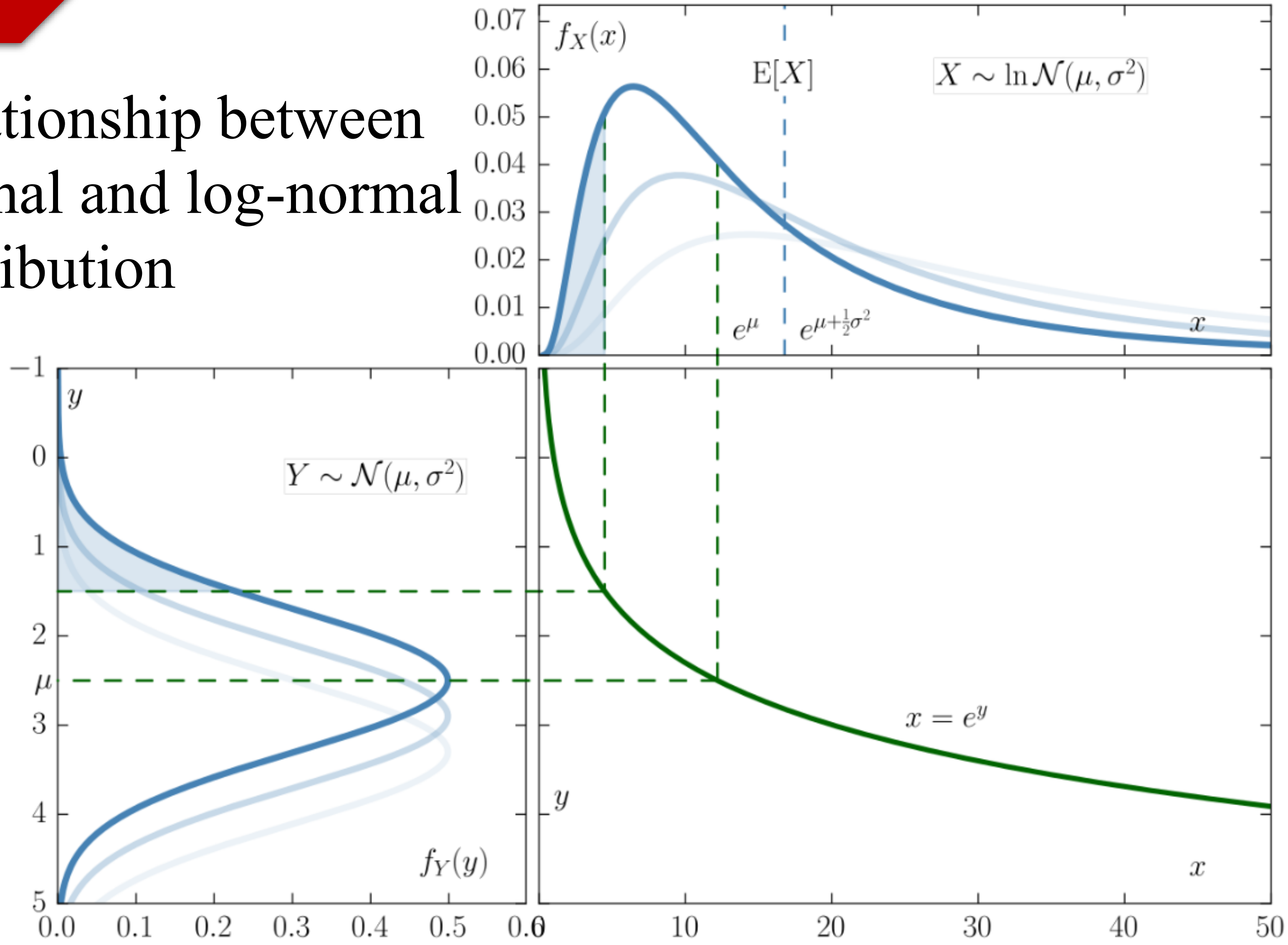
3- Modelling Epidemic Curves

4. Evaluating Pre-Symptomatic Transmission

5. Tailoring Public Health Responses

Probability distribution

Relationship between normal and log-normal distribution



- Log-normal distribution
- Mean

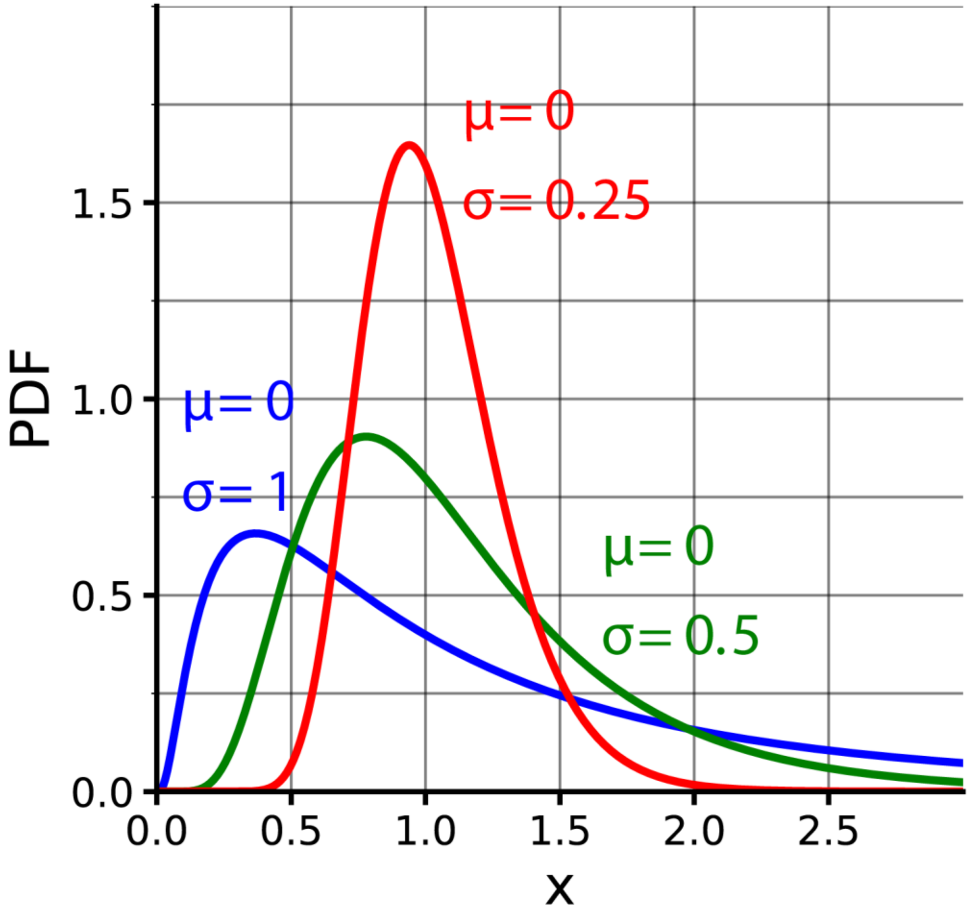
  $\exp\left(\mu + \frac{\sigma^2}{2}\right)$
- Variance

  $\exp[(\sigma^2) - 1] \times \exp(2\mu + \sigma^2)$

https://en.wikipedia.org/wiki/Log-normal_distribution

Probability distribution

Log-normal distribution



- Continuous probability distribution
  - logarithm is normally distributed

- Probability density function (PDF):

$$log\, f(x_i|\theta) = \frac{1}{x\sigma\sqrt{2\pi}} e^{\left(\frac{-(\ln(x)-\mu)^2}{2\sigma^2}\right)}$$

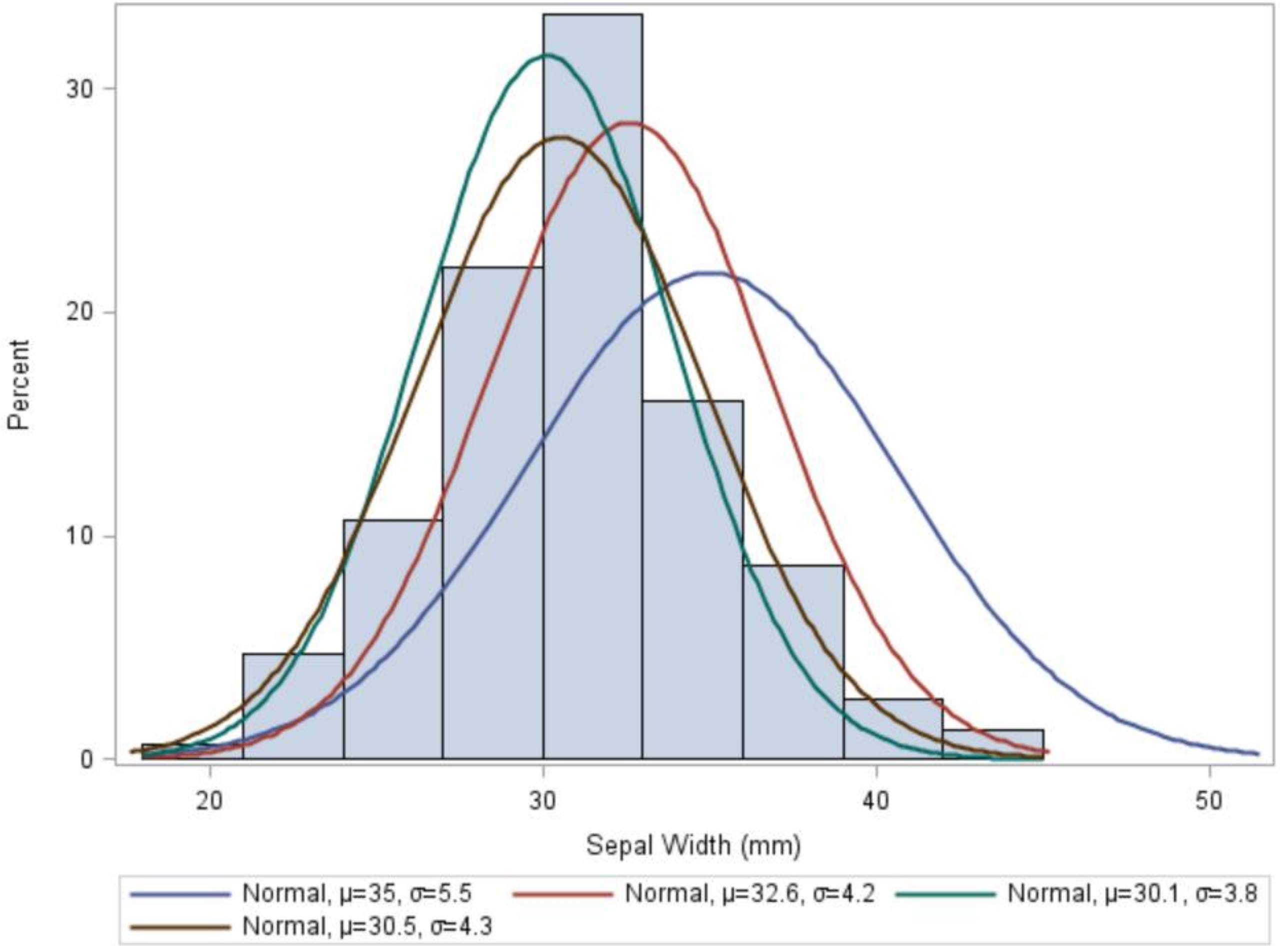
Mean: $\exp\left(\mu + \frac{\sigma^2}{2}\right)$

Variance: $\exp[(\sigma^2) - 1] \times \exp(2\mu + \sigma^2)$

https://en.wikipedia.org/wiki/Log-normal_distribution

28

Maximum Likelihood Estimation



- Maximum likelihood estimation (MLE)

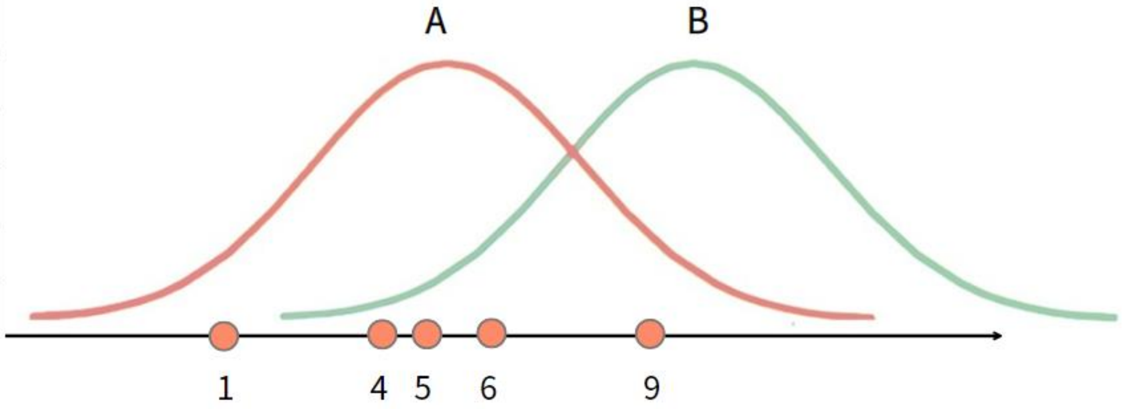
Estimate the most appropriate parameters among the assumed probability distribution types.

29

Maximum Likelihood Estimation

"Which distribution, A or B, is more likely to have generated the data?"

ID	X
1	1
2	4
3	5
4	6
5	9



- MLE

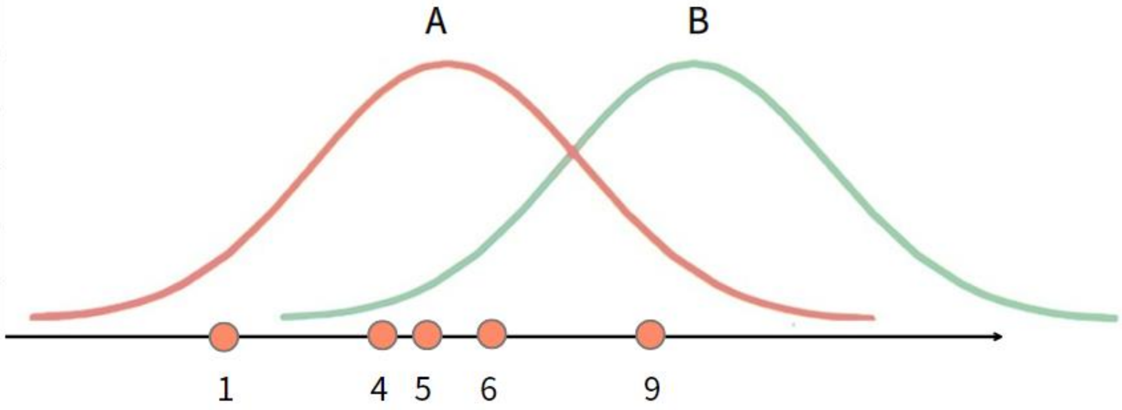
① Assume distribution type: Normal distribution

② Plot PDF

$$f(x_i|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Maximum Likelihood Estimation

ID	X
1	1
2	4
3	5
4	6
5	9



- L_i: likelihood contribution of i^{th} observation.
- Likelihood function
  - Multiply the likelihood contribution of all the observations.
- $L = \prod_{i=1}^{n} L_i$

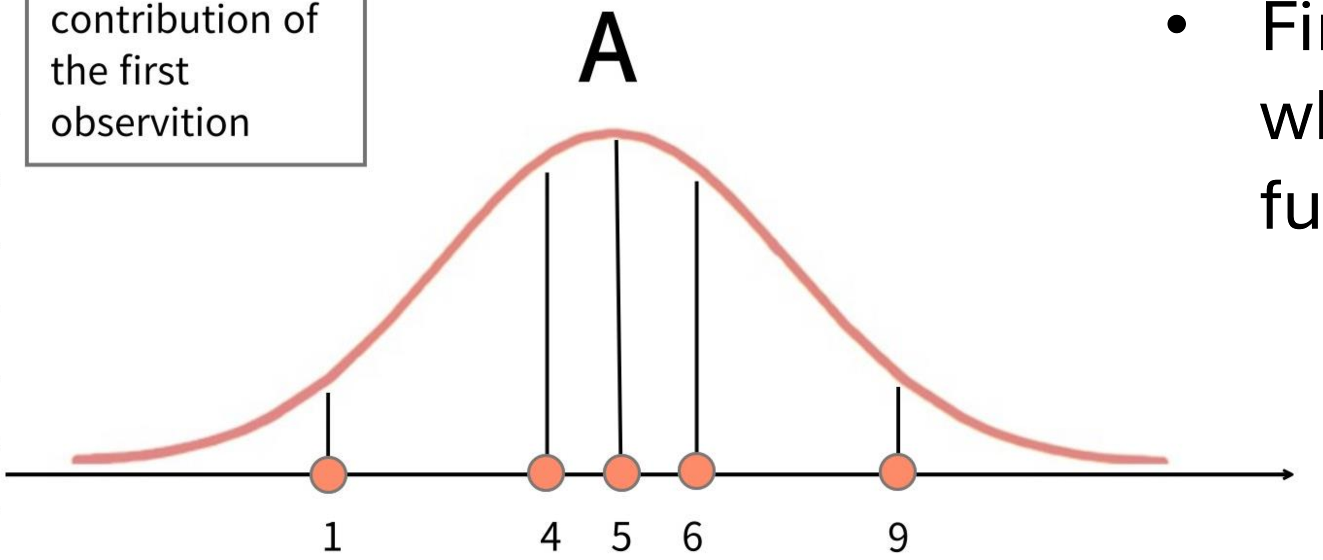
$$f(x_i|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

31

Maximum Likelihood Estimation

The likelihood contribution of the first observition

ID	X
1	1
2	4
3	5
4	6
5	9



- $L(\mu, \sigma) = \prod_{i=1}^{5} L_i$
- Find the values of μ and σ which maximize the likelihood function.

$$L_1 \times L_2 \times L_3 \times L_4 \times L_5$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(1-\mu)^2/\sigma^2} \times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(4-\mu)^2/\sigma^2} \times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(5-\mu)^2/\sigma^2} \times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(6-\mu)^2/\sigma^2} \times \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(9-\mu)^2/\sigma^2}$$

Maximum Likelihood Estimation

- The density function of normal distribution with mean μ and variance σ^2 is given by :

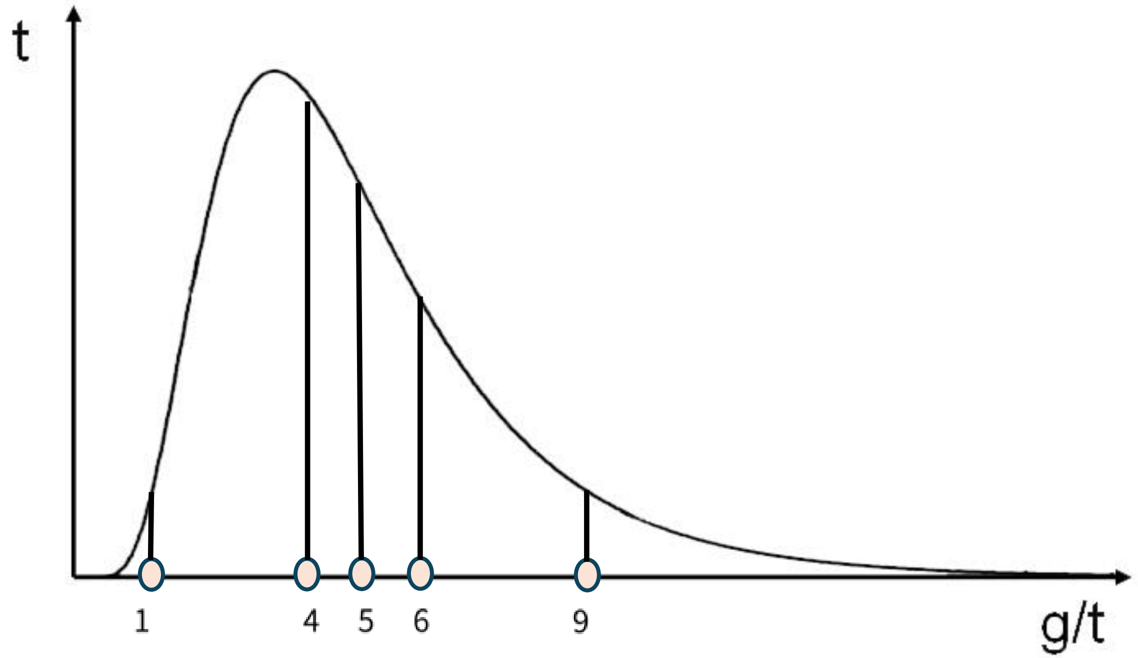
- $f(x_i|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(x-\mu)^2/2\sigma^2}$ for $-\infty < x < \infty$

- The density function of log normal distribution with mean $(\exp(\mu + \frac{\sigma^2}{2}))$, and variance, $(\exp[(\sigma^2) - 1] \times \exp(2\mu + \sigma^2))$ is given by :

- $log\, f(x_i|\theta) = \frac{1}{x\sigma\sqrt{2\pi}}\, e\left(\frac{-(\ln(x)-\mu)^2}{2\sigma^2}\right)$ for $-\infty < x < \infty$

33

Maximum Likelihood Estimation



- Log Likelihood function
  - a logarithmic transformation of the likelihood function

$$l = log \prod_{i=1}^{n} l_i$$

- The log likelihood function is like

$$l(\mu, \sigma) = \prod_{i=1}^{5} l_1 \times l_2 \times l_3 \times l_4 \times l_5$$

$$= \frac{1}{1\sigma\sqrt{2\pi}} e\left(\frac{-(\ln(1) - \mu)^2}{2\sigma^2}\right) \times \frac{1}{4\sigma\sqrt{2\pi}} e\left(\frac{-(\ln(4) - \mu)^2}{2\sigma^2}\right) \times \frac{1}{5\sigma\sqrt{2\pi}} e\left(\frac{-(\ln(x) - \mu)^2}{2\sigma^2}\right)$$
$$\times \frac{1}{6\sigma\sqrt{2\pi}} e\left(\frac{-(\ln(6) - \mu)^2}{2\sigma^2}\right) \times \frac{1}{9\sigma\sqrt{2\pi}} e\left(\frac{-(\ln(9) - \mu)^2}{2\sigma^2}\right)$$

34

Practice and Application

35

Estimating Epidemiological parameters in R

Scenario: Estimating the incubation period and serial interval

You are part of the frontline surveillance team in your country. You have been asked to estimate the incubation period and serial interval of the newly discovered COVID-19. Using the line list data provided;

1) Estimate the incubation period

2) Estimate the serial interval

Note: Use R for your analysis.

36

Estimating the incubation period in R

fitdistr	*Maximum-likelihood Fitting of Univariate Distributions*

Description

Maximum-likelihood fitting of univariate distributions, allowing parameters to be held fixed if desired.

Usage

```
fitdistr(x, densfun, start, ...)
```

Arguments

`x`	A numeric vector of length at least one containing only finite values.
`densfun`	Either a character string or a function returning a density evaluated at its first argument. Distributions `"beta"`, `"cauchy"`, `"chi-squared"`, `"exponential"`, `"gamma"`, `"geometric"`, `"log-normal"`, `"lognormal"`, `"logistic"`, `"negative binomial"`, `"normal"`, `"Poisson"`, `"t"` and `"weibull"` are recognised, case being ignored.
`start`	A named list giving the parameters to be optimized with initial values. This can be omitted for some of the named distributions and must be for others (see Details).
`...`	Additional parameters, either for `densfun` or for `optim`. In particular, it can be used to specify bounds via `lower` or `upper` or both. If arguments of `densfun` (or the density function corresponding to a character-string specification) are included they will be held fixed.

- library(MASS)
- "fitdistr" function estimates the parameters μ and σ of the log-normal distribution

37

Estimating the incubation period in R

① Add data

② Plot histogram
- for distribution assumption

③ Maximum likelihood fitting
- Estimates the μ and σ

④ Construct PDF

⑤ Plot PDF

⑥ Identify the 95% percentile of incubation period

PDF of log normal distribution
- Mean, $\exp\left(\mu + \frac{\sigma^2}{2}\right)$,
- Variance, $\exp[(\sigma^2) - 1] \times \exp(2\mu + \sigma^2)$

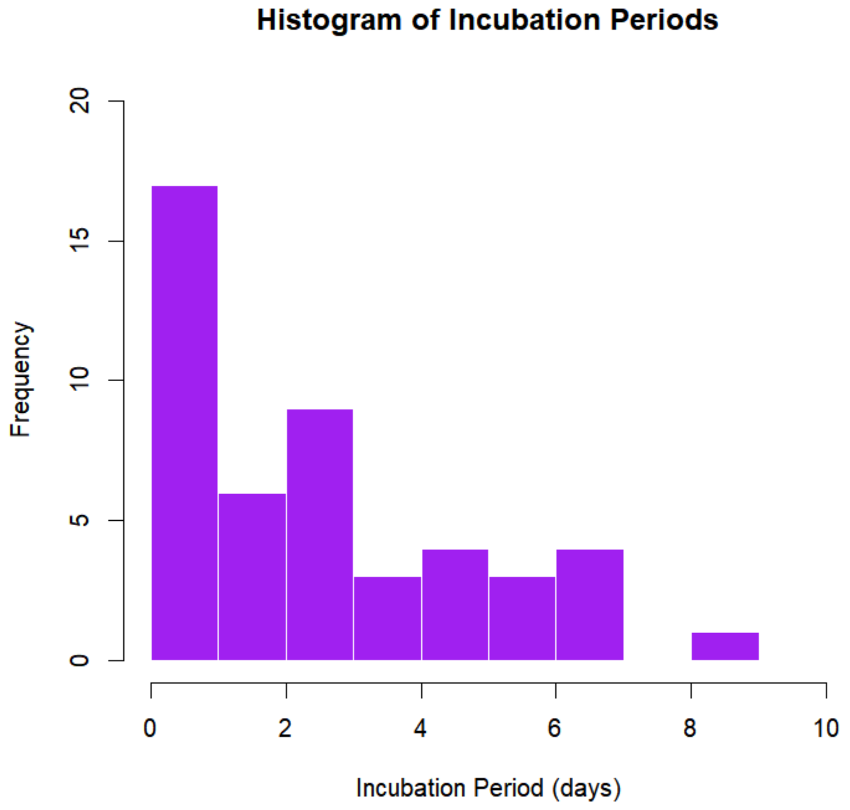
Estimating the incubation period in R

① Add data

incuPeriods = c(rep(0.5,1),rep(0.5,2),rep(1,14),rep(2,6), rep(3,9), rep(3.5,1),rep(4,2),rep(5,4),rep(5.5,1), rep(6,2), rep(6.5,2), rep(7,2),9)

② Plot histogram

```
hist(incuPeriods,
     xlim = c(0, 10),        # adjust to your desired x-axis range
     ylim = c(0, 20),        # adjust to your desired y-axis range
     col  = "purple",       # fill color of the bars
     border = "white",
     xlab = "Incubation Period (days)",
     ylab = "Frequency",
     main = "Histogram of Incubation Periods")
```



39

Estimating the incubation period in R

③ Maximum likelihood fitting

- Estimates the μ and σ

```
library(MASS)
fit_logn<-fitdistr(incuPeriods, "lognormal")
fit_logn
```

```
fit_logn
   meanlog   μ    sdlog  σ
 0.79895685    0.79485577
(0.11594163)  (0.08198311)
```

- Log-normal distribution
- Mean
  $\exp\left(\mu + \frac{\sigma^2}{2}\right)$
- Variance
  $\exp[(\sigma^2) - 1] \times \exp(2\mu + \sigma^2)$

Estimating the incubation period in R

④ Construct PDF

PDF of log normal distribution

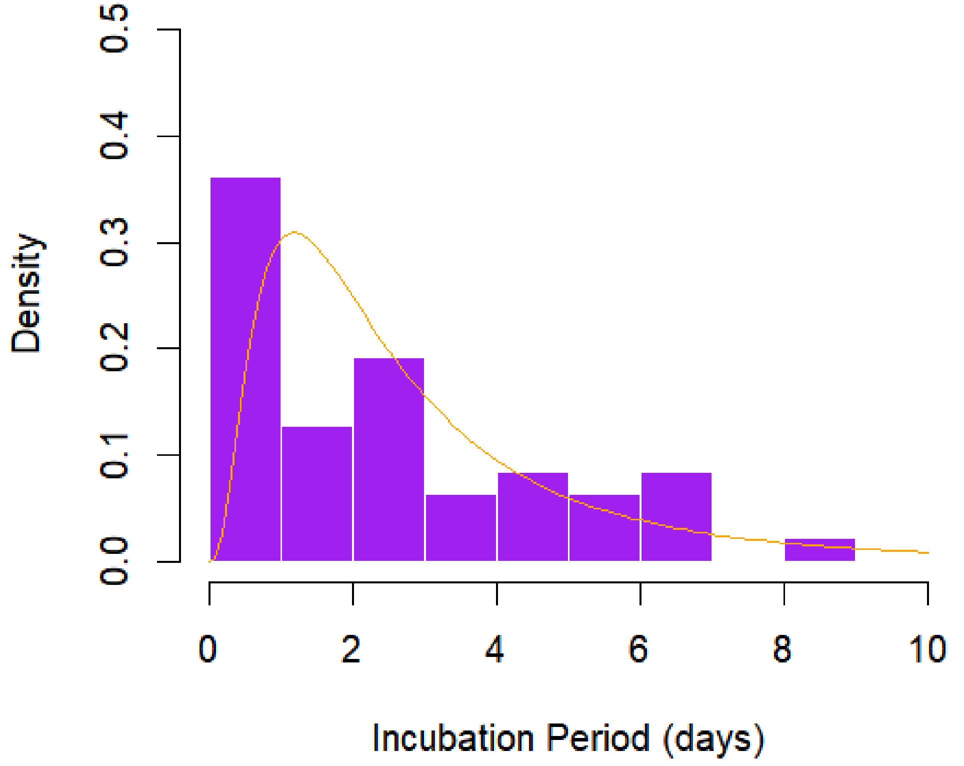
- Mean, $\exp\left(\mu + \frac{\sigma^2}{2}\right)$,
- Variance, $\exp[(\sigma^2) - 1] \times \exp(2\mu + \sigma^2)$

```
fit_logn_mean <-as.numeric (exp(fit_logn$estimate['meanlog']+(fit_logn$estimate['sdlog'])^2/2))

fit_logn_Var<-as.numeric ((exp((fit_logn$estimate['sdlog'])^2)-1)*exp(2*fit_logn$estimate['meanlog’] +
(fit_logn$estimate['sdlog'])^2))
```

Estimating the incubation period in R

⑤ Plot PDF

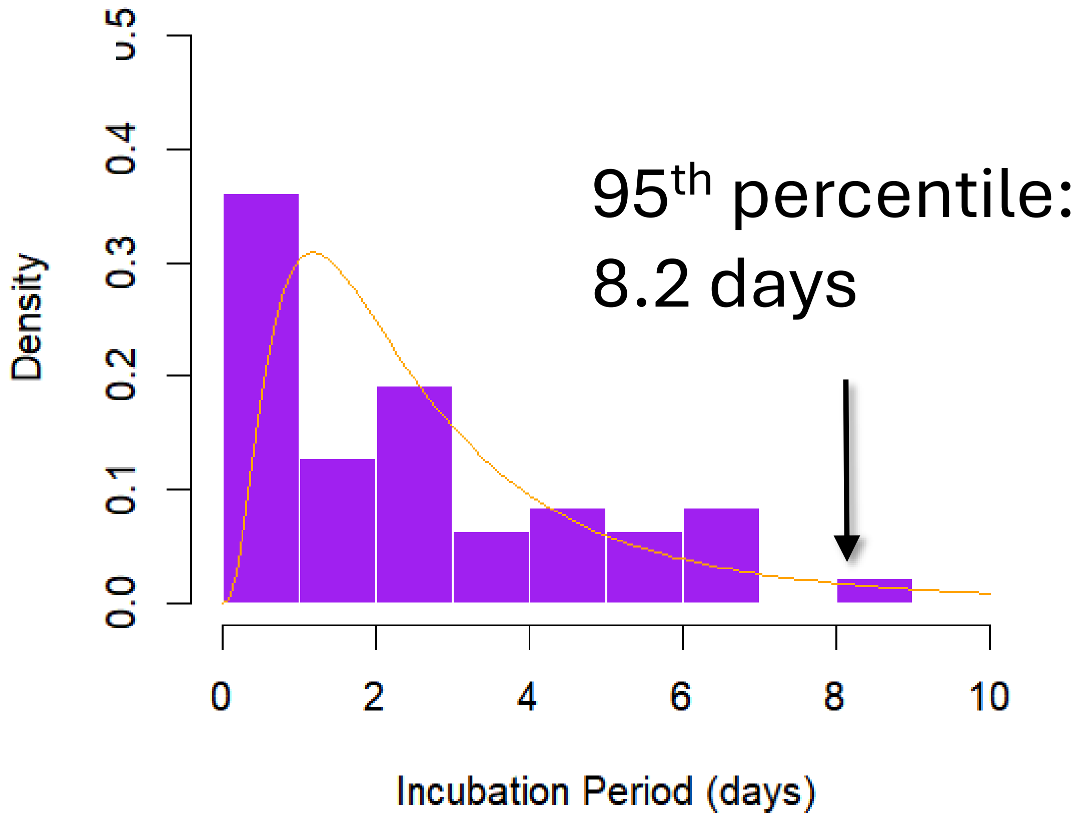
```
hist(incuPeriods,prob=TRUE,
     xlim = c(0, 10),       # adjust to your desired x-axis range
     ylim = c(0, 0.5),
     col  = "purple",       # fill color of the bars
     border = "white",
     xlab = "Incubation Period (days)",
     main = "Histogram of Incubation Periods")
# Fitted curve
curve(dlnorm(x,fit_logn$estimate['meanlog'],fit_logn$estimate['sdlog']),
      add=TRUE, col="orange",lwd=1, lty = 1.5)
```



Estimating the incubation period in R

⑥ Identify the 95% percentile of incubation period

```
> qlnorm(.95, fit_logn$estimate['meanlog'], fit_logn$estimate['sdlog'])
[1] 8.218422
```

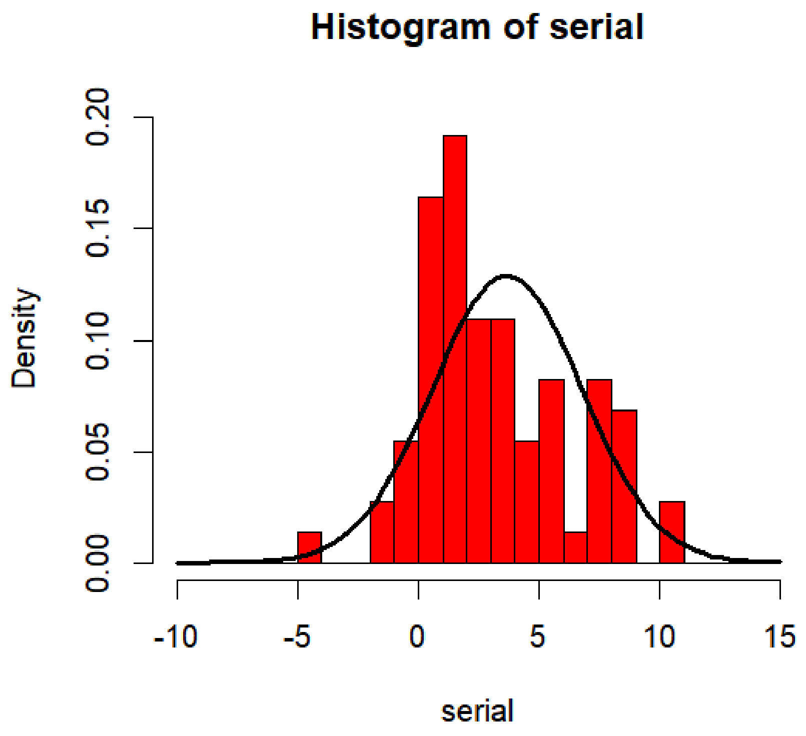


43 Estimating the serial interval in R

```
serial = c(-4, rep(-1,2), rep(0,4), rep(1,12), rep(2,14), rep(3,8), rep(4,8), rep(5,4), rep(6,6), 7, rep(8,6), rep(9,5),
rep(11,2))

fit_norm<-fitdistr(serial, "normal")

fit_normhist(serial, prob=TRUE, ylim=c(0,0.2), breaks=c(-10:15), xlim=c(-10,15),  col  = "red")

curve(dnorm(x,fit_norm$estimate['mean'],
            fit_norm$estimate['sd']),     xlim=c(-10,15),
            add=TRUE, col="black",lwd=3, lty = 1)
```

```
# 95th percentile
qnorm(.95, fit_norm$estimate['mean'], fit_norm$estimate['sd'])
```

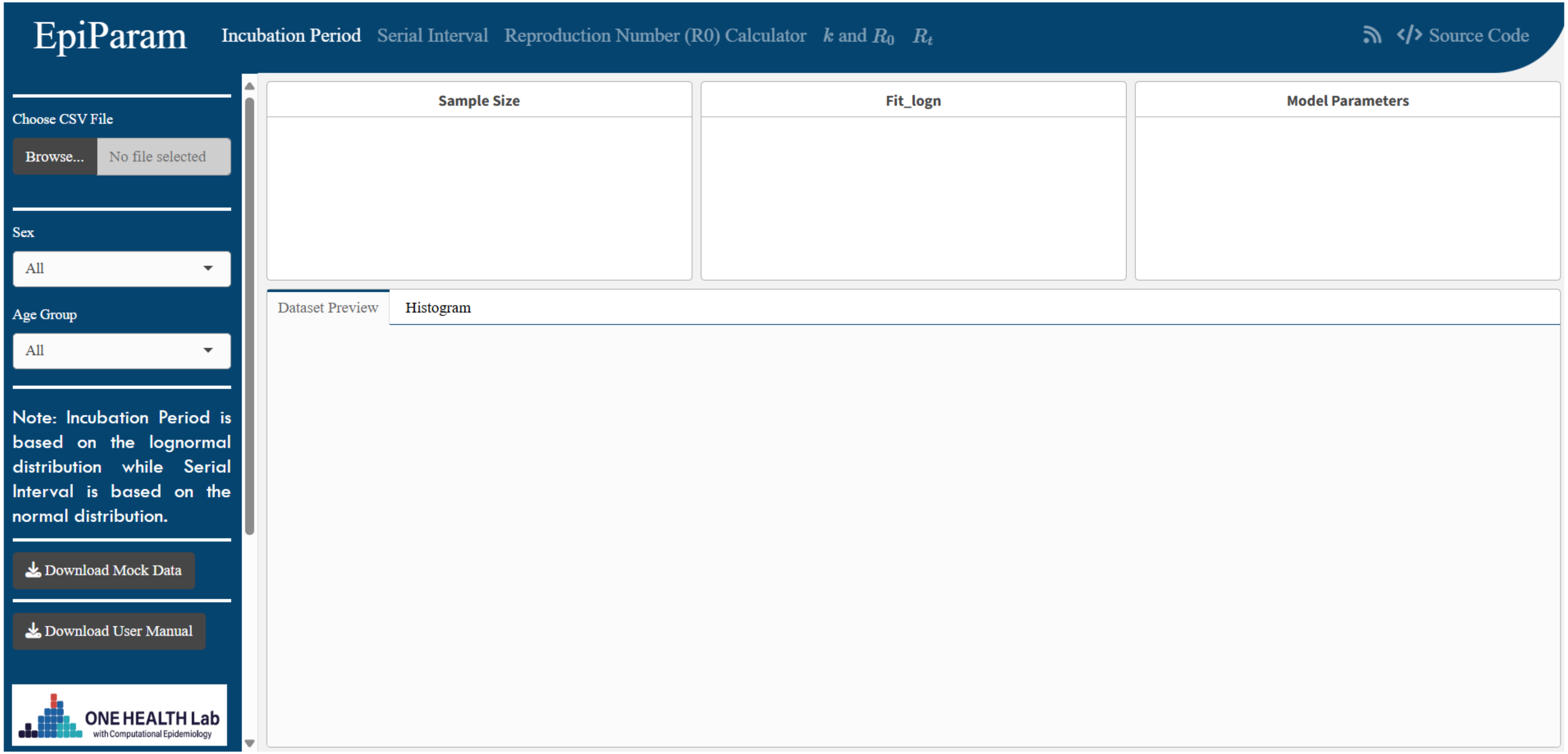
```
> qnorm(.95, fit_norm$estimate['mean'], fit_norm$estimate['sd'])
[1] 8.757833
```



44

Practice with Shiny application

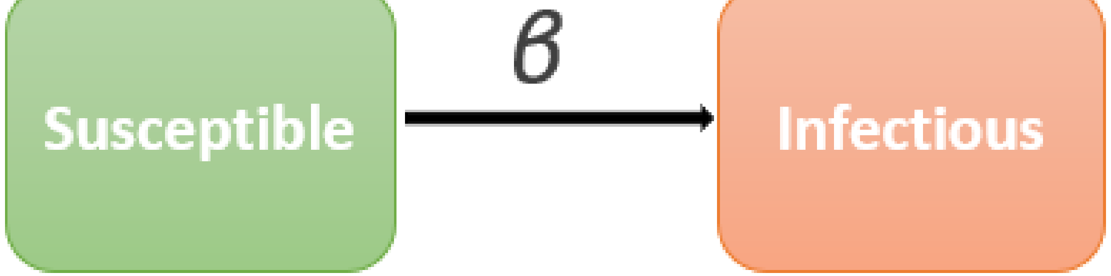
EpiParam: https://achangwa.shinyapps.io/EpiParam/



45

Compartmental models for infectious diseases

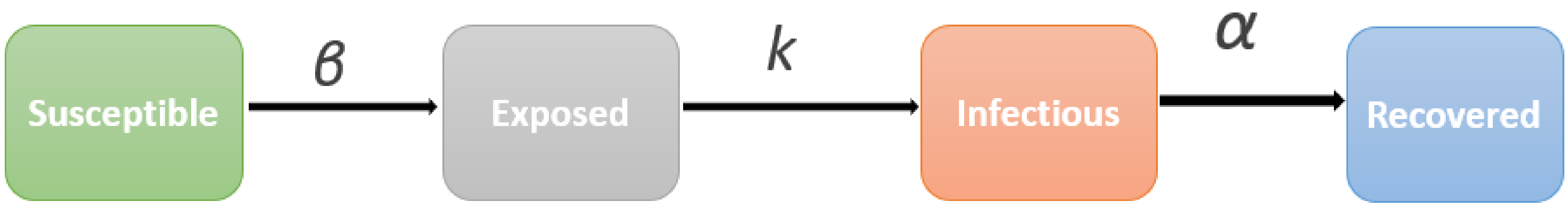
1. Susceptible (S) – Infectious (I): SI Model



2. Susceptible (S) – Infectious (I) – Recovered (R): SIR Model

Compartmental models for infectious diseases

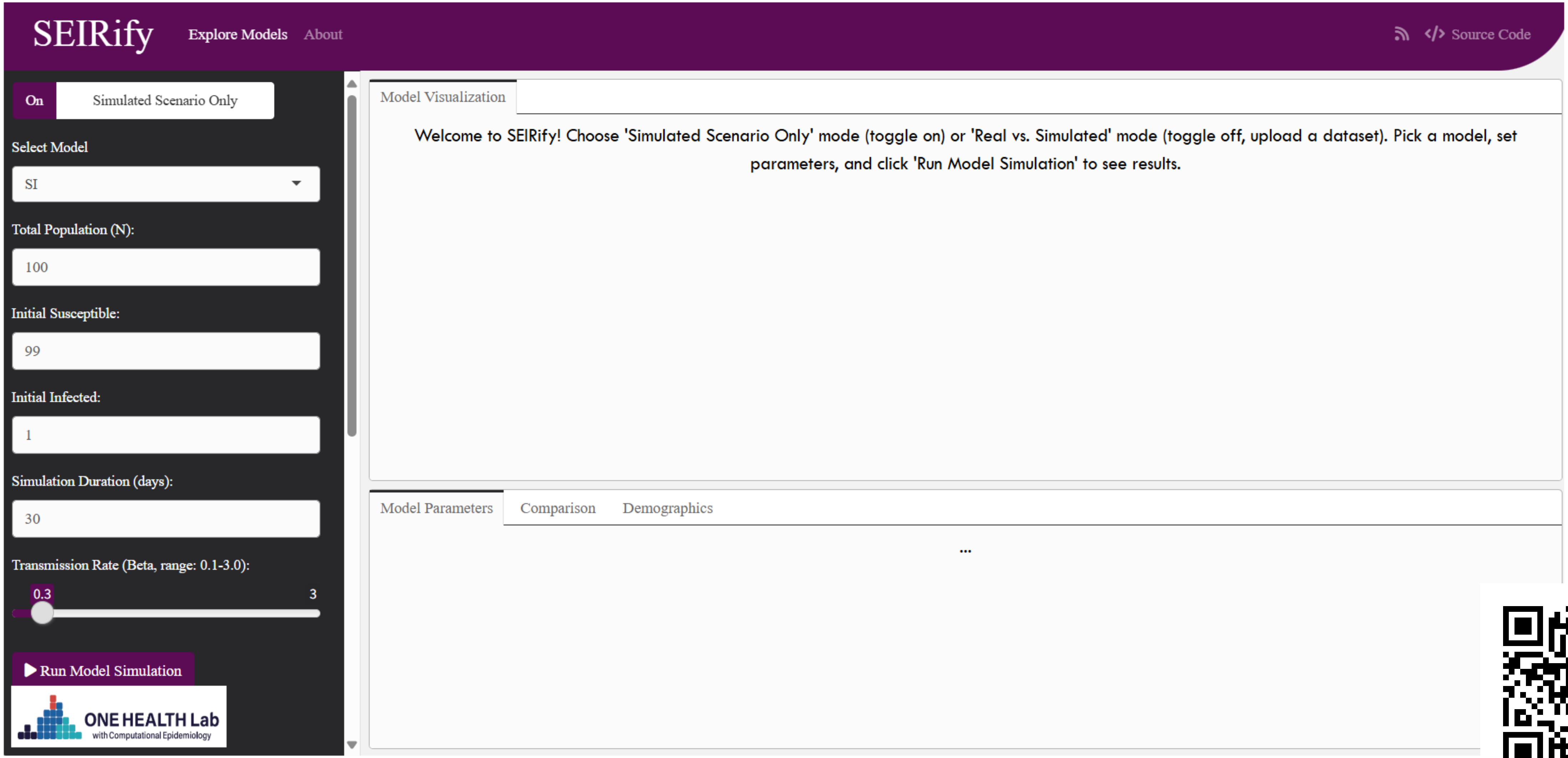
3. Susceptible (S) – Exposed (E) – Infectious (I) – Recovered (R): SEIR Model



47

Compartmental models for infectious diseases

SEIRify: https://achangwa.shinyapps.io/SEIRify/



Key takeaways

- ✓ Infectious disease models are used to estimate epidemiological parameters such as the incubation period and serial interval, which are essential for understanding diseases in the community.

- ✓ Infectious disease modelling provides crucial insights for predicting outbreaks, evaluating interventions, and guiding public health policy.

- ✓ Infectious disease models are only as reliable as the data and assumptions behind them.

ONE HEALTH Lab
with Computational Epidemiology

http://onehealth.or.kr/

The Team

Sukhyun Ryu

ASSISTANT PROFESSOR

Chiara Achangwa

TEAM LEAD

Seonghui Cho

DATA ANALYST

Leah Yeongseon Ko

DATA ANALYST

감사합니다!!

Thank you!!