Epidemiology (8) Chapter 6. Infectious Disease Epidemiology

Minato NAKAZAWA, Ph.D.

<minato-nakazawa@umin.net>

What is human infectious disease?

- Symbiosis: Interactive relation between different species
  - Mutualism: Both species get benefits
  - Commensalism: One side gets benefits, the other gets no benefit but no damage
  - Parasitism: One side (parasite) gets benefits (nutriment) from the other (host), which gets damaged.
- Features of parasites
  - Tend to reproduce more quickly than hosts
  - Tend to evolve rapidly in response to host defenses
    - Defense = **innate immune system** (nonspecific inflammatory reaction → recruits a variety of blood cells such as mast cells, phagocytes, neutrophils, ...) + **adaptive immune system** (*cytotoxic T cells and* **antibodies** from specialized B cell lymphocytes and helper T cells, which record antigenic pattern, enable faster response in the next invasion of the same antigen: <u>antigenic memory = immunity</u> ← after the natural infection and vaccination)
    - Emerging infectious diseases: by changing social conditions, contact of human host with newly met pathogens occurs, when human has no immunity against such pathogens
      - Not quite new: common European diseases (smallpox, measles, typhus, cholera) brought to the New World in Middle Ages were emerging, catastrophic diseases for the native Americans who had no defense against such diseases, conversely syphilis (common for native Americans) caused catastrophic damage in Europeans
  - Infectious parasite is typically much smaller than host
  - Agents infecting human host include:
    - Microorganisms (pathogen = *micro parasite*): Virus, bacteria, fungi, protozoa, ...
    - Larger animals (parasite = *macro parasite*): Helminths, …
- Public health burden from infectious disease began to diminish after the acceptance of germ-theory and the arrival of greatly improved sanitation and hygiene. Figure 6-1 shows steady decline in mortality in the US over 20th C. Spike in 1919 was due to pandemic H1N1 influenza (so-called Spanish flu), which brought same damage with Black death (bubonic plague) in Europe in 14th C. Since mid-20th C, antibiotics contributed to the mortality decline.

What makes pandemic?

- Definition of pandemic
  - An epidemic of unusually high occurrence of disease (Chapter 4)
  - An epidemic occurring worldwide or over a very wide area, crossing boundaries of several countries and usually affecting a large number of people (Dictionary of epidemiology)
  - An influenza pandemic occurs when a new influenza virus appears **against which the human population has no immunity**, resulting in several, simultaneous epidemics worldwide <u>with enormous numbers of deaths and illness</u> (WHO, before pandemic H1N1 flu in 2009)
  - An influenza pandemic occurs when a new influenza virus appears against which the human population has no immunity, resulting in several, simultaneous epidemics worldwide (WHO, after pandemic H1N1 flu in 2009)
- Change of definition of pandemic flu by WHO was to answer to the critics that the pandemic declaration for H1N1 flu in 2009 was motivated by the ties between WHO and pharmaceutical industry (though WHO denied such ties)
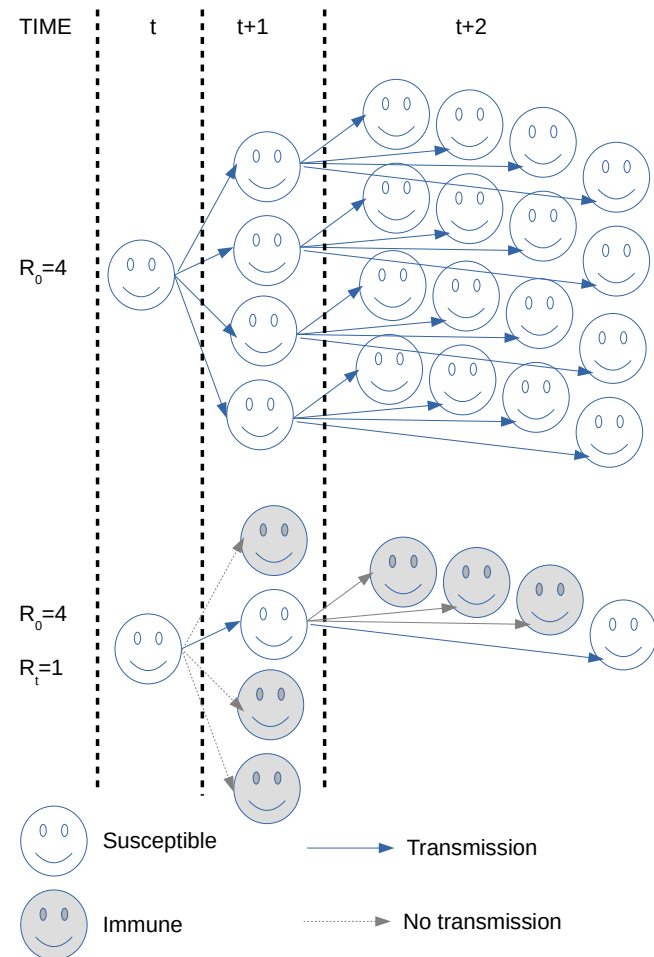
TYPES OF TRANSMISSION

- Host population constitutes **reservoir** for the pathogen
  - Primary habitat for pathogen
  - Pathogen can survive and spread via other hosts than human
- Highly virulent pathogen cannot survive and spread because of early death of host
  - Variety of transmission pathway evolution
    - Direct, person-to-person (communicable, contagious): measles (host is only human) viable only for 2-3 hours in droplets
    - Via transmitting animal (vector): malaria (from infected human with 5 types *Plasmodium* gametocytes to *Anopheles* mosquitoes, then sporozoites in salivary gland moves to another human by the next biting). Most vectors are arthropods
    - **Zoonoses** can spread animal reservoirs to humans
      - Vector-borne: Equine encephalitis, plague
      - Directly from animal to human: Toxoplasmosis (from cat), ebola virus (from bat), flu (hosts are human, birds, and pigs), rabies (hosts are all warm-blooded animals)
        - **CFR** (Number of death due to that disease divided by the number of diagnosed patients) of rabies is 100% if untreated (human is dead-end host), but the virus can survive within other animals than humans

Transmission	Route	Examples
Direct	Airborne	Anthrax（炭疽）, chicken pox, common cold, influenza, measles, mumps, rubella, tuberculosis, whooping cough
	Direct contact	Athlete's foot（水虫）, impetigo（とびひ）, warts（いぼ）
	Fecal-oral	Cholera, hepatitis A, rotavirus, salmonella (=typhoid fever)
	Maternal-fetal	Hepatitis B, syphilis
	Sexual	Chlamydia, gonorrhea, hepatitis B, herpes, syphilis, HPV
Indirect	Intermediate host	Tapeworm (from eating inadequately cooked pork)
	Vector-borne	Bubonic plague (by fleas), malaria (by Anopheles mosquitoes), typhus (by lice), West Nile encephalitis (by Culex mosquitoes), yellow fever (by Culex mosquitoes), dengue fever (by Aedes mosquitoes)

HERD IMMUNITY AND BASIC REPRODUCTION NUMBER (R_0), effective reproduction number (R_t)

- The relative proportions of immune and susceptible persons in a population can determine whether the infection will take hold in the community or die out quickly
- When substantial proportion is immune (**herd immunity** situation), an infected person will be less likely to spread the pathogen
- R_0 (basic reproduction number) is the average number of secondary cases that occur from a single index case in a susceptible population
  - If $R_0 < 1$, the outbreak will die out unless fueled by external re-infections
- R_t (effective reproduction number) is the value of reproduction number that takes into account the mix of immunity and social interaction at any point in time as an outbreak progresses

Disease	Primary mode of transmission	R_0
Measles	Airborne	15
Pertussis (whooping cough)	Airborne droplet	15
Diphtheria	Saliva	6
Smallpox	Social contact	6
Polio	Fecal-oral	6
Rubella	Airborne droplet	6
Mumps	Airborne droplet	5
HIV/AIDS	Sexual contact	3
SARS	Airborne droplet	3
Ebola	Bodily fluids	2
1918 flu	Airborne droplet	2
2009 flu	Airborne droplet	1.5
COVID-19	Airborne droplet	1.4-3?

Note: If average R_0 is same, control efficacy may largely differ by variance.

The nature of R_0 and R_t

- The reproduction number reflects the biologic potential of the infectious agent and the social intercourse that leads to situations in which transmission might occur
  - If directly transmitted disease patient is too sick to move, there will be few contact with susceptible host, results in low reproduction number
- R_0 varies by population (due to behavioral difference by age and so on)
- Even if R_0 is low, some social networks within a population may form a subset with rapid spread of infection. (eg. a few "**superspreaders**" such as needle-sharers transmitting a blood-borne infection can suffice to spark an outbreak)
- Superspreading is not always an attribute of person, sometimes a condition of the field setting (in the case of COVID-19).

- While $R_t>1$, epidemic continues
- Eventually R_t becomes 1 or below, because the proportion of susceptible people decreases or control measures are implemented
- If $R_t=1$ (**endemic equilibrium**), the prevalence of infection remains level over time as new susceptibles are added to the population to balance those who acquire immunity
  - $R_t = 1 = R_0$ x p_s, where p_s is the proportion of the population susceptible to infection at equilibrium, thus $R_0 = 1/p_s$
- Basic strategy to reduce transmission is isolation of infected persons.
- Related strategy is quarantine to restrict contacts among people who are not yet ill but already contacted with infected persons
- (For bioterrorism by smallpox, ring-vaccination is to be conducted to reduce R_t)
- In Japan, restriction of behavior to fill the conditions for superspreading events was taken to reduce R_t as a countermeasure against COVID-19 outbreak.

Estimating R_0 from data

- Using SIR model, let the populations in each compartment at time t S(t), I(t), and R(t). Let the transmission coefficient (proportionate to the probability of transmission per contact) β, and the rate of recovery/isolation γ,
  - $dS(t)/dt = -\beta S(t)I(t)$
  - $dI(t)/dt = \beta S(t)I(t) - \gamma I(t)$
  - $dR(t) = \gamma I(t)$
- The total population at time t is denoted as N(t)=S(t)+I(t)+R(t), and then N(t) is constant because dN(t)/dt=0
- In the beginning, S(0)+R(0)=N
- Since $dI(t)/dt = (\beta S(0) - \gamma)I(t)$, infected population exponentially increases in the beginning, and then $I(t)=I(0)\exp\{(\beta S(0)-\gamma)t\}$
- Using the coefficient of exponential growth, R_0 can be simply estimated as $\beta S(0)/\gamma = R_0$. This **Exponential Growth method** is the most simple ("EG" in **R0** package)
- If we let the generation time (average interval from the infection to primary case to the infection to secondary case, approximated by serial interval) T, $R_0=1+(\beta S(0)-\gamma)T$
  - If we know or can assume T, R_0 can be obtained using the **Maximum Likelihood method** ("ML" in **R0** package)

- For the flu epidemic data at the boarding school in UK, I made the R code to estimate R_0 using R0 package.
  https://minato.sip21c.org/tiid/R0.R
- In R0 package, implemented estimation methods are, in addition to "EG" and "ML", "AR", "TD", and "SB", though "AR" is not applicable to this data.
  - "AR" is based on the equation by Dietz (2013). Let the incidence rate per total population AR,
    $R_0 = -\ln((1-AR)/S(0))/(AR-(1-S(0)))$
  - "TD" assumes time-dependency in the reproduction number, suggested by Cauchemez et al. *AJE* (2006).
  - "SB" is **Sequential Bayes method**, which also assumes time-dependent reproduction number, suggested by Bettencourt & Ribeiro (2008).
- R_0 estimates by "EG" → 8.91 [7.57, 10.52]
- R_0 estimates by "ML" → 6.83 [6.10, 7.62]
- Time-dependent R_0 series estimates
  - by "SB" → 3.96, 4.57, 4.15, 4.22, ...
 - by "TD" → 11.78, 8.82, 6.22, 3.65, ...

Example of SARS, and if vaccine would be available?

- The strategy of isolation and quarantine worked well against SARS
  - SARS nearly became pandemic in 2003, rapidly spread from China to 37 countries (infected more than 8000 people, CFR was almost 10%).
  - Canadian officials quarantined more than 23000 people who had been in contact with SARS cases, about 100 persons for every identified case of SARS. Movement of those under quarantine was restricted until 10 days after the last contact
  - SARS was emerging disease in 2003 and thus no vaccine existed
- If vaccine would be available, R_t depends on vaccine efficacy (V_e) and coverage (V_c)
  - $R_t = R_0 (1 - V_e \times V_c) \leftrightarrow R_t/R_0 = 1 - V_e \times V_c \leftrightarrow V_c = (1 - R_t/R_0)/V_e$
  - $R_t < 1 \leftrightarrow V_c > (1 - 1/R_0)/V_e$
  - When R_0 is large, to succeed in curtailing the epidemic, high efficacy and coverage are needed (If R_0 is 10 and V_e is 95%, V_c has to be larger than 95% needed to reduce R_t below 1; $(1 - 1/10)/0.95 = 0.947\ldots \approx 0.95$)
  - In the case of measles, R_0 is 15. Even if V_e is 100%, V_c has to be larger than 93% to reduce R_t below 1; $(1 - 1/15)/1 = 0.933\ldots \approx 0.93$
  - If R_0 is 2 and V_e is 95%, V_c needs to be larger than 53% to reduce R_t below 1; $(1 - 1/2)/0.95 = 0.526\ldots \approx 0.53$.
- The same relationship may stand for not only vaccination but also naturally acquired immunity after infection. Vaccine efficacy corresponds to the proportion of immunized by single infection and coverage corresponds to the proportion of people ever infected and recovered (**herd immunity threshold**)

Estimation methods of Rt

- Cori et al. (2013) https://doi.org/10.1093/aje/kwt133 and https://doi.org/10.1140/epjp/s13360-021-01339-6: EpiEstim package of R

- Rough and simple approximation (Niid, 2021): (Newly reported cases within the recent 7 days) / (Generation time earlier newly reported cases within 7 days) https://www.niid.go.jp/niid/ja/diseases/ka/corona-virus/2019-ncov/2502-idsc/iasr-in/10465-496d04.html

- Toyo Keizai (https://toyokeizai.net/articles/-/351826?page=3): {(Newly reported cases within recent 7 days) / (Newly reported cases within 14-8 days before)}^{(average generation time)/(reporting interval)}

- See, https://minato.sip21c.org/epispecial/libuseTokyoRt.R

東京のCOVID-19について4つの方法で推定した実効再生産数(Rt)の推移
[Data] Guidotti E, Ardia D (2020) COVID-19 Data Hub.
Working paper https://doi.org/10.13140/RG.2.2.11649.81763
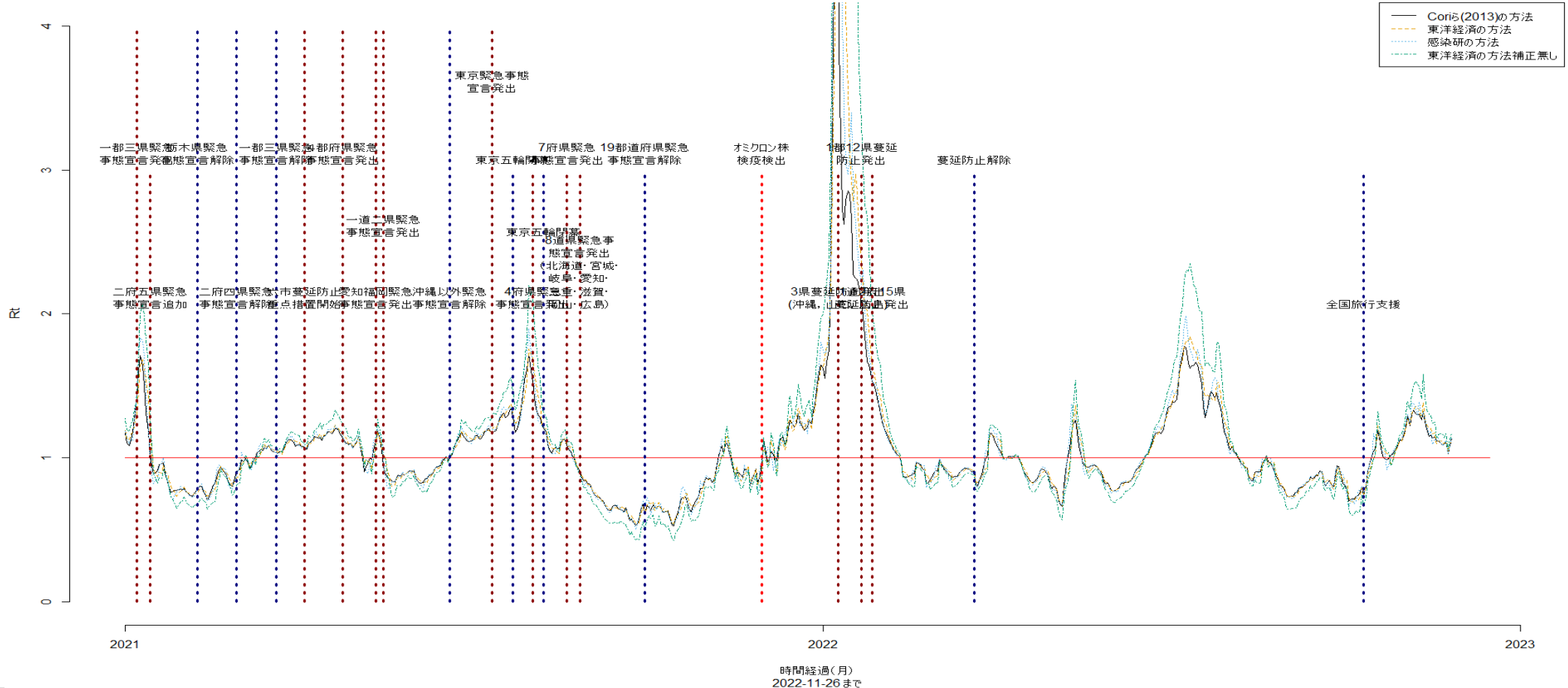
Glossary of key terms (slightly modified from the textbook and some added)

- **Case Fatality Risk (CFR)**: Same as Case Fatality Ratio. Number of death due to a disease divided by the number of infected cases with diagnosis. Indicates severity. Rabies: 100%, Ebola virus infection: 60%, Spanish flu: 2-2.5%, seasonal flu: 0.01-0.02%, COVID-19: 0.6-6%. If the denominator is symptomatic cases instead of confirmed cases, it should be written as sCFR (Nishiura et al., 2010).
- **Infection Fatality Risk (IFR):** In the pandemic of COVID-19, transmission of the disease occurred from asymptomatic cases and the patients in latent period and thus CFR largely varied by the situation of diagnostic testing. Therefore, Nishiura (2020) suggested to use IFR as the number of death due to a disease divided by the total number of infected cases including asymptomatic cases. Based on CFR estimated in China where mostly moderate or severe cases were tested (3-6%) and the total number of infected people was 10 times more, IFR of COVID-19 was estimated as 0.3-0.6%.
- **Communicable**: capable of person-to-person transmission
- **Generation time**: the time interval between one person getting infected and another person getting infected from the first
- **Herd immunity**: a prevalence of susceptibles in a population low enough so that transmission cannot be sustained
- **Immunity**: resistance to infection
- **Inapparent infection**: infected without any symptoms. (eg.) Almost 100% of JC virus infection causes no symptoms
- **Incubation period** (**Latent period**): the time interval between getting infected and developing <u>symptoms</u>
- **Index of infection to symptomatic cases**: The proportion of symptomatic cases among all infected. Untreated rabies: 100%, Measles: 99%, Seasonal flu: 60%, Poliomyelitis: 0.1-1%, Japanese encephalitis: 0.1-3%.
- **Reproduction number**: the average number of infected persons resulting from contact with a single infected person. Incl. R_0 and R_t
- **Reservoir**: the host population for an infectious agent
- **Secondary cases**: cases of infection that occur from contact with a primary case
- **Secondary attack rate**: risk of infection among susceptibles exposed to an infected source
- **Susceptibility**: at risk of contracting disease (**lack of immunity**)
- **Transmission probability**: probability of transmission from an infected person to susceptible person during a contact
- **Vector**: an animal that transmits disease from an infected person to an uninfected person
- **Virulence**: the degree to which a pathogen can cause disease and death

# Basic reproduction number for several infectious diseases

Table 1.2 Approximate serial intervals, basic reproduction numbers and implied crude herd immunity thresholds (as $1 - 1/R_0$) for common potentially vaccine-preventable diseases.

Disease name	Serial interval (range)	R_0	Herd immunity threshold (%)
Diphtheria	2-30 days	6-7	85
Influenza	2-4 days	2-4	50-75
Malaria	20 days	5-100	80-99
Measles	7-16 days	12-18	83-94
Mumps	8-32 days	4-7	75-86
Pertussis	5-35 days	12-17	92-94
Polio	2- 45 days	2-4, 8-14 (a)	Controversial (immunity to infection is not solid)
Rubella	7-28 days	6-7	83-85
Smallpox	9-45 days	5-7	80-85
Tuberculosis	Months-Years	Not established because contacts over time may largely change during long serial interval	

(a) R_0 of polio depends on hygienic condition (if poor, it becomes large).

Cited from EMILIA VYNNYCKY and RICHARD G WHITE (2010) "AN INTRODUCTION OF INFECTIOUS DISEASE MODELLING".Oxford Univ. Press
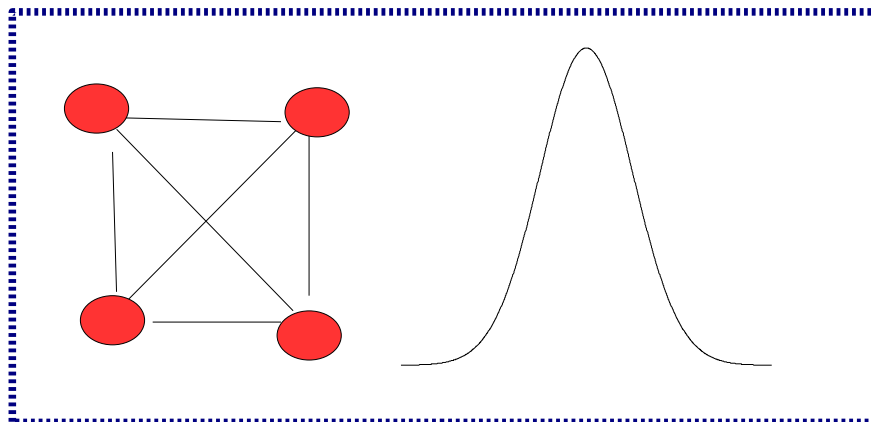
Terms for individual infection history
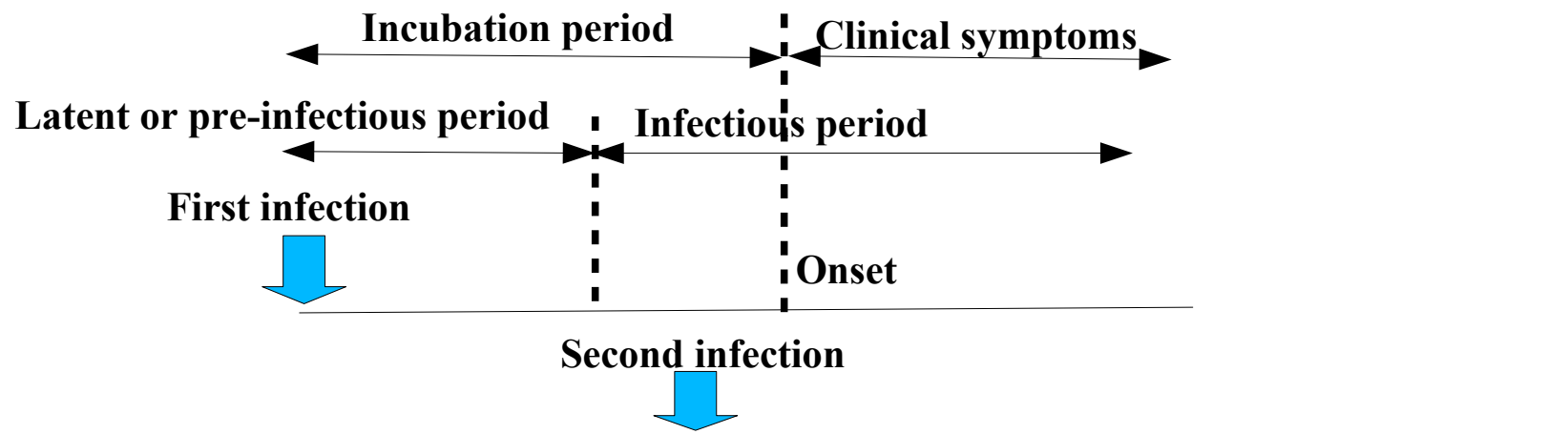
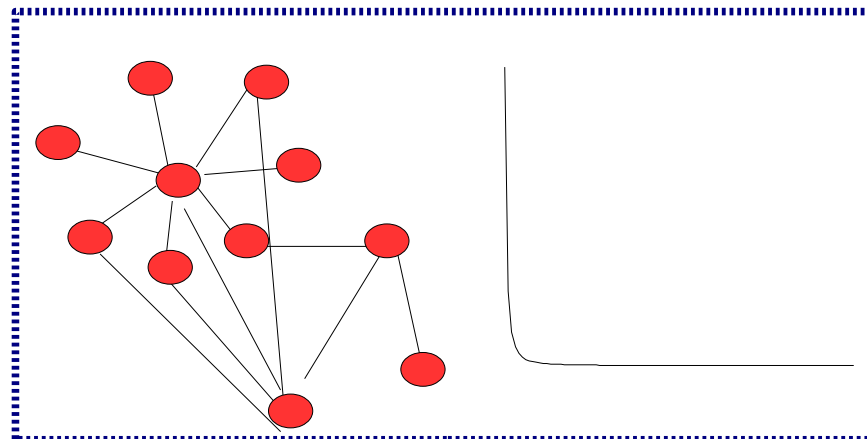Two general types of infection network topology
(1) Random link network
* equal infection probability for each
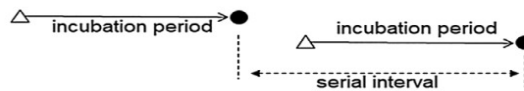* distribution of infection frequency
  is unimodal

(2) Scale free network
* host preferences
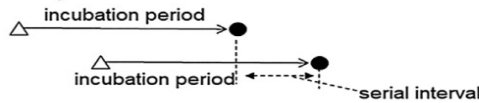* distribution obeying power law
* superspreader exists
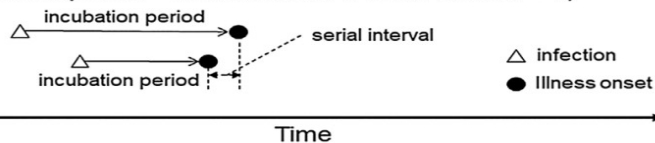
Serial interval and generation time

Figure 2. The relationship between the incubation period and serial interval. If the transmission takes place during the symptomatic period of the primary case, the serial interval is longer than the incubation period. However, this relationship can be reversed when pre-symptomatic transmission takes place. Furthermore, it is possible that the secondary case may even experience illness onset prior to onset in their infector.

TABLE 2

Parameter estimates and credible intervals of generation and serial interval distributions of COVID-19 with missing serial intervals only allowed to be positive by different incubation periods, Singapore, 21 January–26 February 2020; Tianjin, China, 14 January–27 February 2020

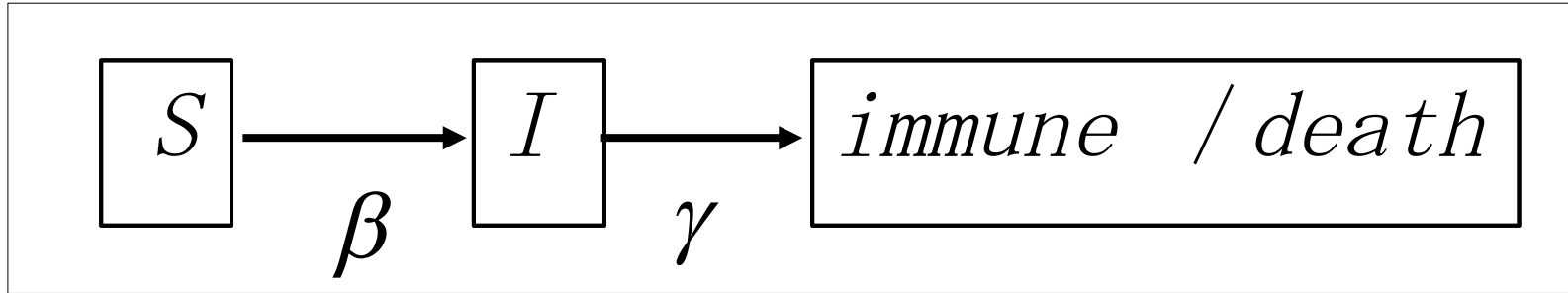| Dataset	Assumed incubation period (days)	Interval	Estimate (95% credible interval) (days)	
			Mean	SD
Singapore[a]	Mean 6.4, SD 2.3	GI	5.29 (3.89 - 6.77)	2.08 (0.97 - 4.07)
		SI	5.29 (−2.13 - 13.16)	3.86 (3.40 - 5.21)
	Mean 4.8, SD 2.6	GI	5.19 (3.82 - 6.74)	1.77 (0.91 - 4.11)
		SI	5.19 (−2.86 - 13.45)	4.08 (3.79 - 5.51)
Tianjin (China)[b]	Mean 6.4, SD 2.3	GI	4.02 (3.11 - 5.00)	2.29 (1.02 - 3.80)
		SI	4.02 (−4.83 - 13.45)	3.98 (3.41 - 5.00)
	Mean 4.8, SD 2.6	GI	3.95 (3.05 - 4.93)	1.75 (0.77 - 3.35)
		SI	3.95 (−4.60 - 12.73)	4.07 (3.76 - 4.97)

COVID-19: coronavirus disease; GI: generation interval; SD: standard deviation; SI: serial interval.
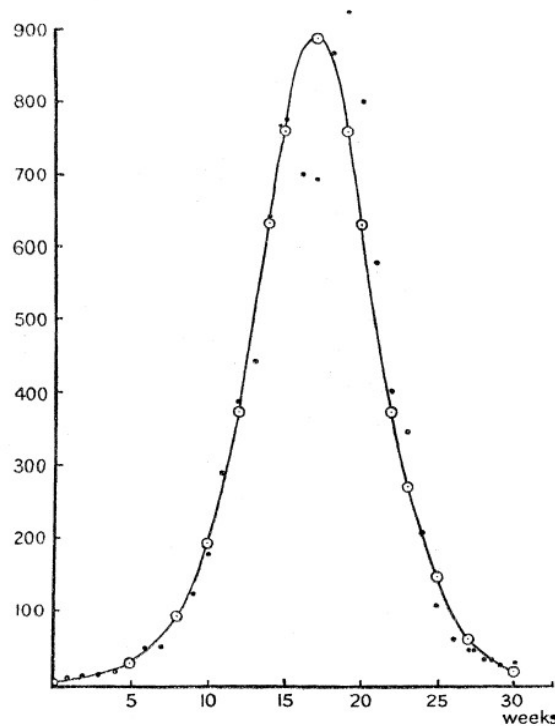[a] Source: Ministry of Health (https://www.moh.gov.sg/news-highlights/, as at 26 February).
[b] Source: Tianjin Municipal Health Commission (http://www.tjbd.gov.cn/zjbd/gsgg/, as at 27 February).

- Generation time (from time at infection to the primary case to average time at infection to the secondary cases) is usually assumed to match with serial interval (from time at onset of symptom in the primary case to average time at onsets of symptoms in the secondary cases).

- However, if the half of transmission is shared by presymptomatic cases like in covid-19 (as suggested by Nishiura et al. (2020) [https://dx.doi.org/10.1016%2Fj.ijid.2020.02.060] (Figure 2 shown top left)), it was not sure *a priori*.

- Ganyani et al. (2020) [https://dx.doi.org/10.2807%2F1560-7917.ES.2020.25.17.2000257] showed the matches of generation time and serial interval in the covid-19 data (top right).

Simple mathematical model (SI)

$$S \xrightarrow{\beta} I \xrightarrow{\gamma} immune \;/\; death$$

$$\frac{dz}{dt} = \frac{l^3}{2x_0\kappa^2}\sqrt{-q}\;\operatorname{sech}^2\!\left(\frac{\sqrt{-q}}{2}lt - \phi\right). \tag{31}$$

$$\frac{dS}{dt} = -\beta SI$$

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

The accompanying chart is based upon figures of deaths from plague in the island of Bombay over the period December 17, 1905, to July 21, 1906. The ordinate represents the number of deaths per week, and the abscissa denotes the time in weeks. As at least 80 to 90 per cent. of the cases reported terminate fatally, the ordinate may be taken as approximately representing dz/dt as a function of t. The calculated curve is drawn from the formula

$$\frac{dz}{dt} = 890\,\operatorname{sech}^2(0\cdot 2t - 3\cdot 4).$$

Kermack-McKendrick model (1927) for plague outbreak in Bombay from December 17, 1905 to July 21, 1906.

SIR model for flu epidemic

$$\frac{dS}{dt} = -\beta SI + \delta R$$

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

$$\frac{dR}{dt} = \gamma I - \delta R$$

Example: The data in English boys boarding school in 1978

```
NI <- c(3, 8, 28, 75, 221, 291, 255, 235, 190, 125, 70, 28,
12, 5)

sir.I <- function(S0=762, I0=1, R0=0, days=1:14,
 beta=0.0026, gamma=0.5, delta=0.0001) {
  S <- I <- R <- double(length(days)+1)
  S[1] <- S0
  I[1] <- I0
  R[1] <- R0
  for (i in days) {
    S[i+1] <- S[i] - beta*S[i]*I[i] + delta*R[i]
    I[i+1] <- I[i] + beta*S[i]*I[i] - gamma*I[i]
    R[i+1] <- R[i] + gamma*I[i] - delta*R[i]
  }
  return(I)
}

plot(0:14, c(1, NI), pch=16, type="p",
 ylim=c(0, 800), xlab="time (days)",  ylab="infected")
lines(0:14, sir.I(762, 1, 0, 1:14, 0.0026, 0.5, 0.0001), co
l="red", lty=1)
```
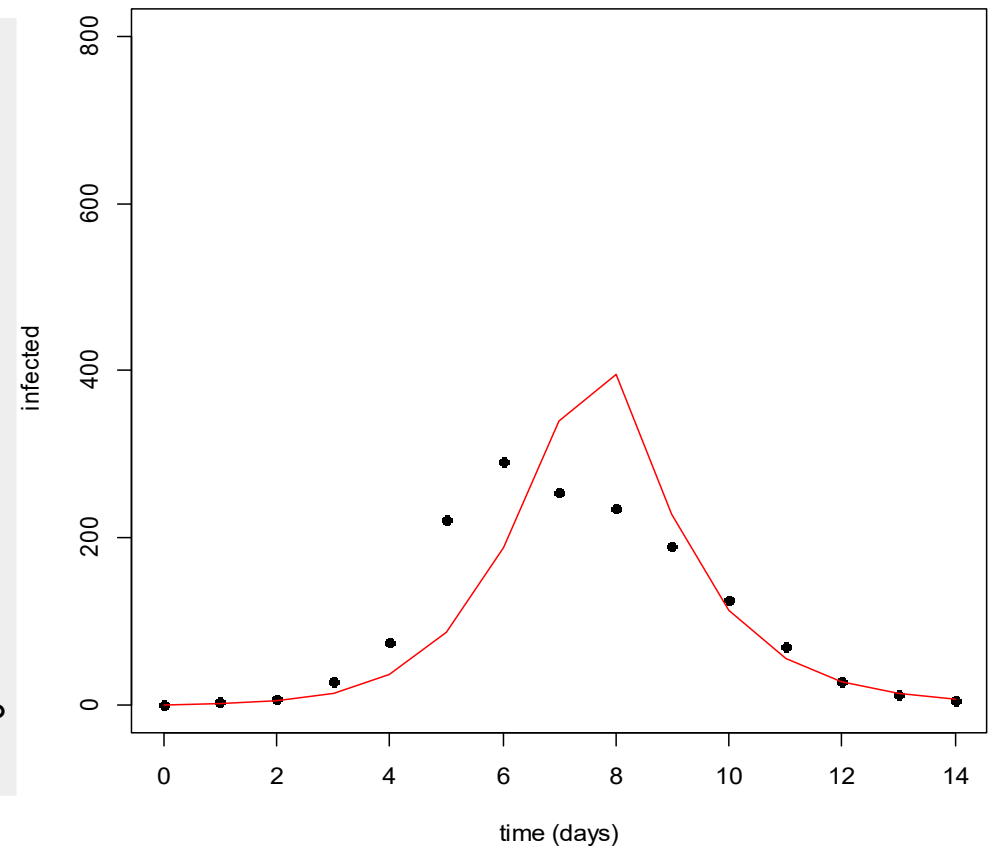
THE REED-FROST EPIDEMIC MODEL

- Assumptions
  - There is random mixing, with contact between infected people and susceptible people within the population during each time period
  - There is a uniform, fixed probability that a contact between an infected person and a susceptible person would result in transmission
  - An infection is always followed by immunity
  - The population is isolated from other populations
  - These conditions remain constant with time
- $C(t+1) = S(t) \{1 - (1 - p)^{C(t)}\}$
  - $C(t)$: the number of infected people at time t
  - $S(t)$: the number of susceptible people at time t
  - p: the probability that within one time period an infected person will transmit the infection to a susceptible person with whom there is contact
- See, Figure 6-2: Reed-Frost projection of epidemic curve for infected, susceptible, and immune subpopulations among 100 people with one initial infected person and an effective contact probability of 4% (high R_0 in upper panel) and 1.5% (low R_0 in lower panel). The time scale is measured in generation times.

INFECTIOUS DISEASE EPIDEMIOLOGY INVESTIGATIONS

- Several types of epidemiologic studies are unique to the investigation of infectious disease
- Four types are worthy to mention
  - Contact-tracing studies: In the early stage of an epidemic, it may be possible to interrupt person-to-person transmission enough to bring R_0 below 1 by isolation, treatment and quarantine of patients (shoe-leather approach).
  - Outbreak investigation: When a local epidemic occurs, documenting outbreak and investigating its origin and propagation. Often detective work such as identification of the cause of diarrhea outbreak as the church supper on the potato salad.
    - (Note: AIDS is an acronym of Acquired Immuno-Deficiency Syndrome, not Acute)
  - Seroprevalence surveys (sero-epidemiology): Distribution of prevalence of a specific disease is reflected in the distribution of the antibody titers against that disease agent in the blood, since the antibody titers remain for several months after infection. It can assess the vulnerability of a population to existing infectious agents, for finding susceptible subgroups
  - Vaccine trials: A randomized trial of preventive measure is called a field trial (Chapter 5). It's much more difficult than clinical trials. One of the reasons is the outcome (prevented) is rare. (eg.) Salk vaccine trial. Infection of polio virus was popular but paralysis symptoms were rare.

OUTLOOK FOR INFECTIOUS DISEASE EPIDEMIOLOGY

- In the very beginning period after the invention of antibiotics, human misunderstood that the ultimate defense against infection from bacteria was found
- For vaccines, human also misunderstood that viral illness might be tamed and possible to eradicate as smallpox
- However,
  - High reproductive rate of microorganisms and their ability to mutate have enabled them to evade many of our technologically driven defenses
  - Widespread and unnecessary use of antibiotics produced antibiotic-resistant bacteria
  - Increasing urbanization and intercontinental travel added risk of communicating infectious diseases
  - Social and medical practices opened new routes of transmission
- Infectious disease epidemiology is a frontier that has observed 2 remarkable triumphs
  - Eradication of smallpox
  - Near-elimination of poliomyelitis
- The hope of eradication of other diseases: **malaria** is candidate but challenging
  - Quinine, chloroquine, artemisinine, and other chemotherapy were effective to cure patients but resistance developed
  - DDT and other insecticides were effective to reduce anopheles mosquitoes but those were toxic for environment and mosquitoes got resistance
  - Development of highly effective vaccine is very difficult because life-cycle of malaria parasite is very complex and multi-stage and *P. falciparum* can escape from antibody by distributing junk antigens
  - Previously nonhuman (simian) malaria, *P. knowlesi* switched the host from monkeys to human