

# Aspects of homogeneous vs. heterogeneous transmission

Ping Yan, Research Manager

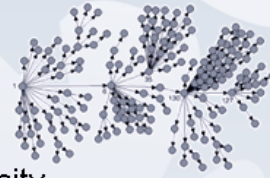
Surveillance and Risk Assessment Division  
Centre for Communicable Diseases and Infection Control

Infectious Diseases Prevention and Control Branch

## Highlight

### 1. Phenomenological observation

“A picture is worth 1000 words.”



#### 1.1. Implications of phenomenological heterogeneity

*At the very beginning: as an infectious individual seeded in an infinitely large susceptible population ... ..*

### 2. Heterogeneity in the agent-host-environment interface

- Conceptual assumptions vs. tactical assumptions

2.1. Heterogeneity in tactical assumptions, implications on dynamics

2.2. Heterogeneity in conceptual assumptions, implications on dynamics

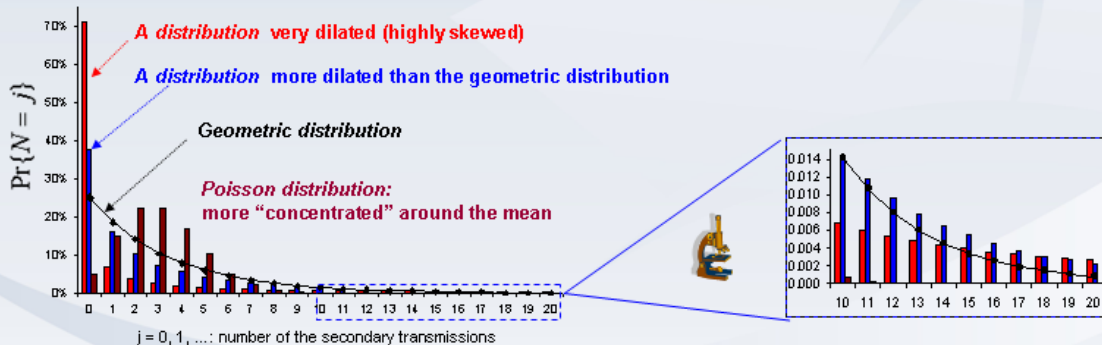
# 1. Phenomenological observation

Let random variable:

$N$  = # of infections produced during ones entire infectious period

The probability:  $\Pr\{N = j\}$ ,  $j = 0, 1, \dots$ ; the mean (weighted average):  $E[N] = \sum_{j=0}^{\infty} j \Pr\{N = j\}$

The follow plots correspond to 4 probability distributions, all with mean = 3



- **Heterogeneous:** *a few infect many; many infect a few*
  - o *the probability of producing zero transmission is higher (skewed);*
  - o *the probability of producing very large number of transmissions is also higher ;*
  - o *more likely to take “extreme” values*

Heterogeneity is a relative term to homogeneity.

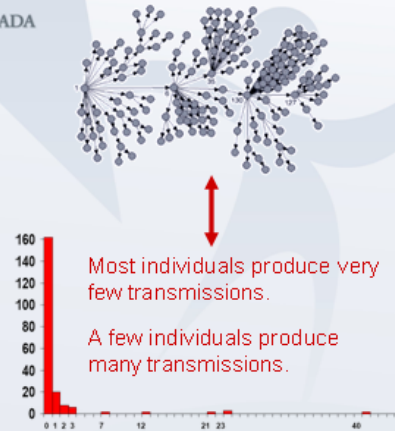
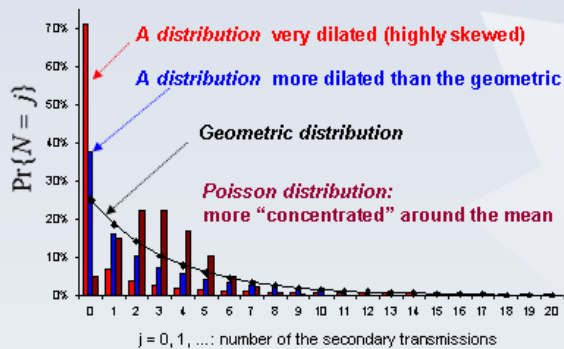
The ODE models for SIR and SEIR and the corresponding stochastic compartment models with Markov property generate the geometrically distributed infectious contacts, and those models are often attached with the word “homogeneity”.

Therefore, we use the geometric distribution as a benchmark. In the plots, the geometric distribution is illustrated as a line. All the distributions have the same mean value.

If more heterogeneous than the geometric distribution, the probability of  $N=0$  is higher and the probability of  $N$  taking large values is also higher (see the magnified image). Hence, the more likely it takes extreme values.

On the other hand, when more homogeneous than the geometric distribution, the distribution is more concentrated around the mean. The Poisson distribution is one of such examples.

# 1. Phenomenological observation



Heterogeneous: relative to homogeneous

## Verbal description:

"the more heterogeneous, the more likely it takes "extreme" values."

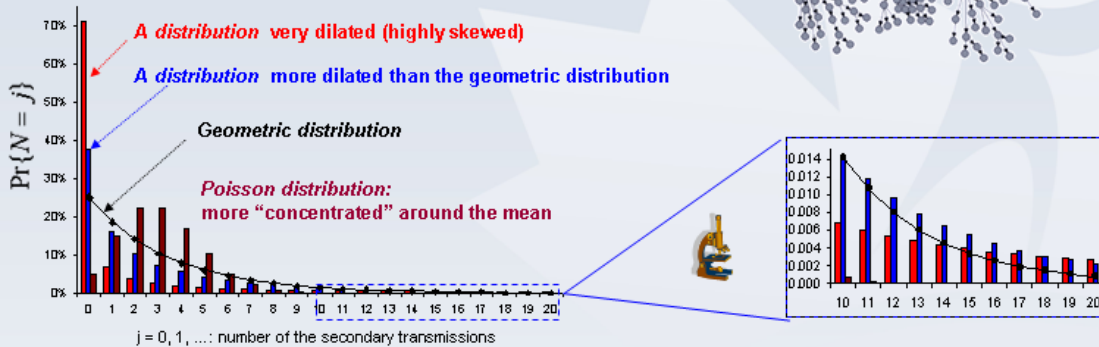
Many different aspects of heterogeneity may produce the same phenomenon:

Could be due to: large variation of the infectious period among infected individuals even when the environment is "homogeneous";

Could be due to: highly heterogeneous environment in which the transmission occurs even if little variation of individuals' infectious periods.

The two "could be due to" aspects will be Part 2 of this presentation.

# 1. Phenomenological observation



**Heterogeneous: relative to homogeneous**

**Mathematical translation: convex ordering of two random variables**

$X_1$  is smaller than  $X_2$  in convex order, denoted as  $X_1 \leq_{cx} X_2$

if  $E[\Phi(X_1)] \leq E[\Phi(X_2)]$  for all convex function  $\Phi(x)$ .

*Convex functions are functions that take on their relatively large values over the regions of the form  $(-\infty, a)$   $(b, \infty)$  for  $a < b$ .*

The geometric distribution is implied in ODE and Markov S(E)IR models, associated with "homogeneity." Hence, we may use the geometric distribution as a benchmark, according to convex order, for "heterogeneity".

For more on convex order, I recommend

Shaked, M and Shanthikumar, J.G (2007) Stochastic Orders. Springer.

It is the mathematical language for the verbal description:

“The more heterogeneous, the more likely it takes “extreme” values.”

## 1.1. Implications of phenomenological heterogeneity

Questions and answers regarding: “as an infectious individual **seeded** in an infinitely large susceptible population”, at the very beginning of an epidemic.

**Characterize a small outbreak versus a large outbreak**  
(along the line of Kendall 1956)

Let  $C(\infty)$  be the cumulative number of infected individuals as  $t \rightarrow \infty$ , random, with mean  $E[C(\infty)]$

sometimes, a **small outbreak**: a handful cases followed by extinction

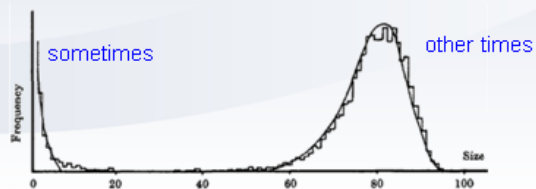
- The expected number of infected individuals by the end of the outbreak is finite even if the population size can be infinitely large:  $\frac{E[C(\infty)]}{n} \rightarrow 0, \text{ as } n \rightarrow \infty$

other times, a **large outbreak**:

- The expected cumulative number of infected individuals scales linearly with the size of the susceptible population  $\frac{E[C(\infty)]}{n} \rightarrow \eta > 0, \text{ as } n \rightarrow \infty$

$C(\infty)$  follows a bi-modal distribution (shown by simulation)

**From Anderson and Watson (1980):**  
simulation based on  $n=100$  individuals.  
**Bi-modal distribution with one mode at zero, and another mode around 0.8.**



The original Kendall paper is hard to find. It is

Kendall, D. (1956) Deterministic and stochastic epidemics in closed populations. Proc. Fifth Berkeley Symp. Math. Statist. Probab. 4. University of California Press 149-165.

Alternatively, in Diekmann and Heesterbeek (2000), the authors provide the same description.

Diekmann, O. and Heesterbeek, J.A.P. (2000). Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation. Wiley Series in Mathematical and Computational Biology.

The simulation example is taken from

Anderson, D. and Watson, R. (1980) On the spread of a disease with gamma distributed latent and infectious periods. Biometrika, 67, 1, 191-198.

# 1.1. Implications of phenomenological heterogeneity

Questions and answers regarding: "as an infectious individual **seeded** in an infinitely large susceptible population", at the very beginning of an epidemic.

## A. What is the risk of a large outbreak? *Denote this probability as $1-\pi$*

$N$  = # of infections produced during ones entire infectious period.

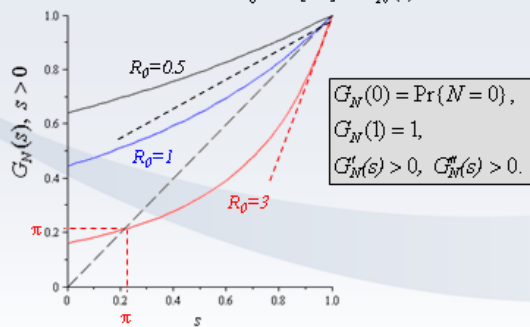
$\Pr\{N = j\}$ ,  $j = 0, 1, \dots$  uniquely determined by prob. generating function

$$G_N(s) = \sum_{j=0}^{\infty} s^j \Pr\{N = j\} = E[s^N], \quad s > 0$$



The basic reproduction number :

$$R_0 = E[N] = G'_N(1)$$

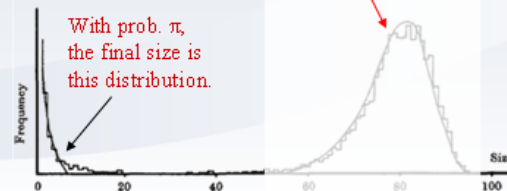


If  $R_0 \leq 1$ , then  $\pi = 1$ , with certainty.

Zero risk of large outbreak.

If  $R_0 > 1$ , then  $\pi < 1$ .

The risk of large outbreak =  $1 - \pi$ .



$\pi$  is the smallest root of the equation  $G_N(s) = s$

The probability generating function (p.g.f.) is a very powerful tool to study non-negative integer values discrete random variable. The probabilities can be uniquely defined through the p.g.f.

The use of p.g.f. to study the extinction probability (i.e. the risk a small outbreak) can be found in every textbook on branching processes.

Relating it explicitly to the risk of a small outbreak, as well as the expression of  $R_0$  as the slope of the p.g.f. evaluated at  $s=1$ , along with all the strictly convex property of the p.g.f., can be found in Diekmann and Heesterbeek (2000).

## 1. 1. Implications of phenomenological heterogeneity

### B. What does heterogeneity do to the risk of a large outbreak $1-\pi$ ?

**Convex ordering:**  $X_1 \leq_{cx} X_2$  if  $E[\Phi(X_1)] \leq E[\Phi(X_2)]$  for all convex function  $\Phi(x)$ .

**Variance by definition:**  $\text{var}[X] = E[(X - \mu)^2]$ , where  $\mu = E[X]$   
 $(x - \mu)^2$  is a convex function of  $x$

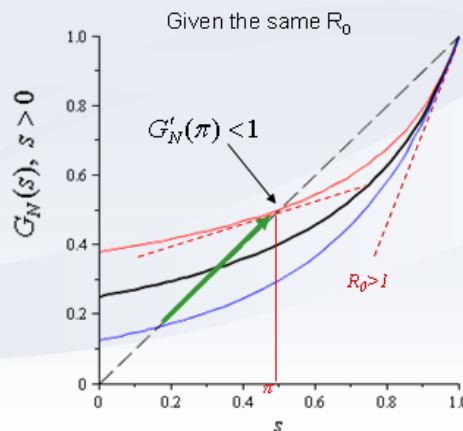
**Prob. generating function:**  $G_X(s) = E[s^X]$ ,  $s \in [0,1]$

$s^x$  is a convex function of  $x$

*Heterogeneity* with respect to the random variable  $N$ , in the sense that,  
 "the more heterogeneous, the more likely it takes 'extreme' values"

then the more heterogeneous:

- the larger is the variance  $\text{var}[N]$ ;
- the larger is the value  $G_N(s)$  for all  $s \in [0,1]$
- the larger is the probability,  $\pi$ , of a small outbreak, as the smallest root of  $G_N(s) = s$



Heterogeneity is described by convex ordering.

Larger in convex order gives larger variance, which is a more intuitive measure.

Larger in convex order gives ordering of the probability generating functions.

Later in the presentation, we shall see a correspondence between the p.g.f. and the Laplace transform function of the infectious period.



# 1. 1. Implications of phenomenological heterogeneity

## C. If a small outbreak, how many generations it takes to become extinct and what does heterogeneity affect it?

$T_g = 1, 2, \dots$  = number of generations to extinction as a discrete random variable

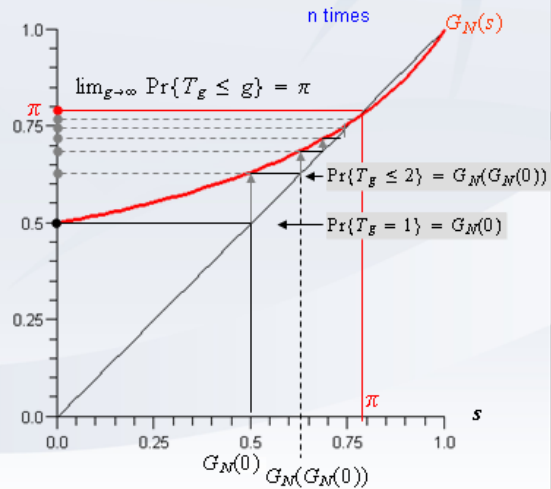
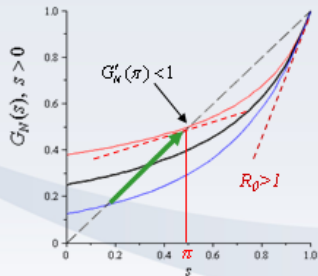
The cumu. prob. can be calculated recursively  $\Pr\{T_g \leq g\} = G_N(G_N(\dots G_N(0) \dots))$

The following are exercises in textbooks.

i. For a suitable positive constant  $A$ ,

$$\pi - \Pr\{T_g \leq g\} \sim A[G'_N(\pi)]^g$$

ii. It is always true  $G'_N(\pi) < 1$



Heterogeneity not only increases the prob. of a small outbreak, but also makes it to become extinct quickly with fewer generations.

This slide is based on Ch. 10 (Yan, P.) of Springer Lecture Notes (Ed. Brauer, van den Driessche and Wu).

Note that if  $R_0 > 1$ ,  $\lim_{g \rightarrow \infty} \Pr\{T_g \leq g\} < 1$ , which defines a sub-distribution, rather than a proper distribution function.

The graph showing  $\lim_{g \rightarrow \infty} \Pr\{T_g \leq g\} = \pi$  is the line of thinking that leads to the proof that  $\pi$  is the smallest root of the fixed-point-equation  $G_N(s) = s$ .

# 1.1. Implications of phenomenological heterogeneity

## D. If a small outbreak, what is the distribution of its final size?

$C(\infty) = 1, 2, \dots$  is a discrete random variable.

The distribution of  $C(\infty) = 1, 2, \dots$  can be also precisely calculated using a recursive formula based on  $G_N(s), s > 0$

$$G_{C(\infty)}(s) = sG_N(G_{C(\infty)}(s))$$

where  $G_{C(\infty)}(s)$  stands for the prob. generating function of  $C(\infty)$ , which can be used to calculate

$$\Pr\{C(\infty) = j\}, j = 1, 2, \dots$$

The mean and variance of the final size:

Evaluate  $G'_N(\pi)$  and  $G''_N(\pi)$

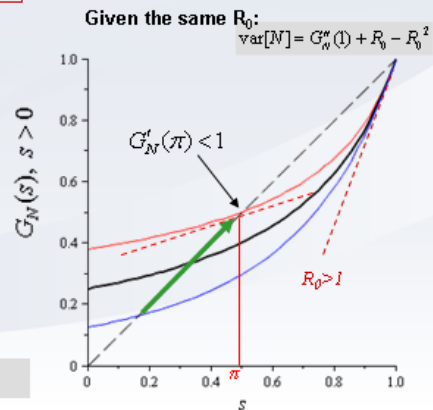
$$E[C(\infty) | \text{small outbreak}] = \frac{1}{1 - G'_N(\pi)}$$

$$\text{var}[C(\infty) | \text{small outbreak}] = \frac{G''_N(\pi) + G'_N(\pi) - G'_N(\pi)^2}{(1 - G'_N(\pi))^3}$$

If  $R_0 \leq 1, \pi = 1, G'_N(1) = R_0$ :

$$E[C(\infty)] = \frac{1}{1 - R_0}, \text{var}[C(\infty)] = \frac{\text{var}[N]}{(1 - R_0)^3}$$

Heterogeneity  $\rightarrow$  large  $\text{var}[N]$ , hard-to-predict final size



This slide is also based on Ch. 10 (Yan, P. ) of Springer Lecture Notes (Ed. Brauer, van den Driessche and Wu).

The case  $R_0 \leq 1$  leads to large variance.

In the case  $R_0 > 1$ ,  $G'_N(\pi)$  replaces  $R_0$ , and  $G''_N(\pi)$  replaces  $G''_N(1)$ . Hence  $G''_N(\pi) + G'_N(\pi) - G'_N(\pi)^2$  replaces  $\text{var}[N]$ . Some theory that leads to the statement of large variance of the final size needs to be fixed.

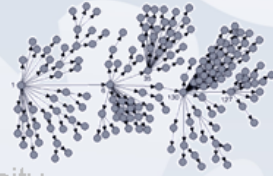
## Highlight

### 1. Phenomenological observation

“A picture is worth 1000 words.”

#### 1.1. Implications of phenomenological heterogeneity

*At the very beginning:* as an infectious individual seeded  
in an infinitely large susceptible population ... ..



### 2. Heterogeneity in the agent-host-environment interface

- **Conceptual assumptions vs. tactical assumptions**

2.1. Heterogeneity in tactical assumptions, implications on dynamics

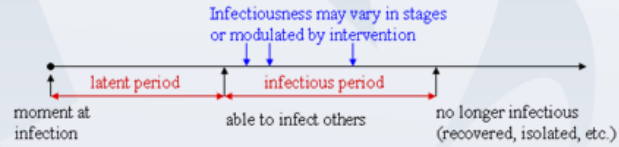
2.2. Heterogeneity in conceptual assumptions, implications on dynamics

## 2. Heterogeneity in agent-host-environment interface

- Tactical assumptions vs conceptual assumptions

### Example of tactical assumptions

*within an infected individuals*



1. *Is there a latent period? If yes, how long on average? How variable?*
2. *Besides the average infectious period  $\mu_I$ , how variable is it distributed?*

**Example of conceptual assumptions:** *environment, population, among individuals,...*

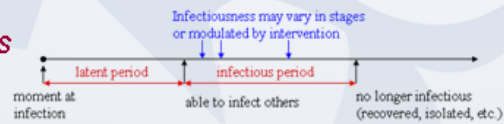
1. **The force of infection** onto a susceptible individual  $\propto \frac{I(t)}{n(t)}$  : % infectious individuals  
 $\longrightarrow \beta \frac{I(t)}{n(t)} S(t)$
2. **The instantaneous rate** of passing the infection from a typical infectious individual to another  $\propto \frac{S(t)}{n(t)}$  : % of susceptible individuals  
 $\longrightarrow \beta \frac{S(t)}{n(t)} I(t)$

One of the (hidden) conceptual assumptions about homogeneity:  $\beta^* = \beta$  (bilinearity)

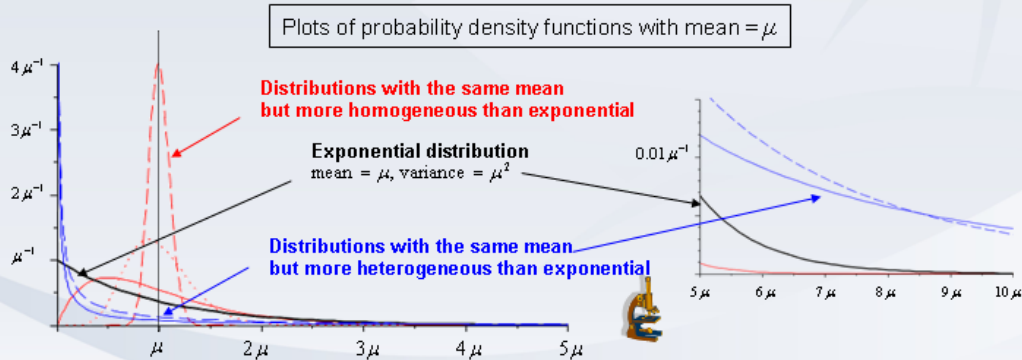
The tactical assumptions are made with an infected individuals, otherwise known as the “natural history”. The conceptual assumptions are made about how individuals interact.

## 2. Heterogeneity in agent-host-environment interface

### Variability of random variables for durations (Tactical assumptions)



- Some models: the latent and the infectious periods are constants (variance = 0).
- ODE or Markov SEIR models: both periods are exponentially distributed (variance = mean<sup>2</sup>).



### Heterogeneous: relative to homogeneous

*The exponential distribution is implied in ODE or Markovian S(E)IR models.*

We use the exponential distribution as a benchmark, according to convex order for heterogeneity.

These are continuous time distributions, used to describe time, which is a non-negative continuous random variable. For continuous distributions, the probabilities are described by probability density functions.

In Part 1, we have seen plots of probabilities, used to describe the distribution of a discrete random variable  $N$ .

The exponential distribution presented here share many similar properties of the geometric distribution for discrete random variables.

We shall see more similarities in the next slide.

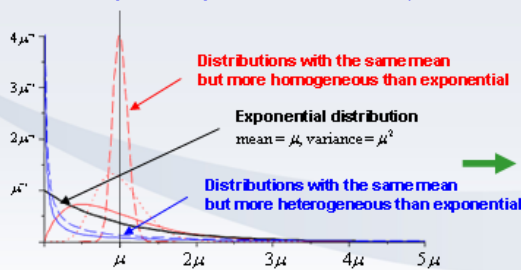
## 2. Heterogeneity in agent-host-environment interface

- Large variance of infectious period leads to heterogeneous phenomenon:

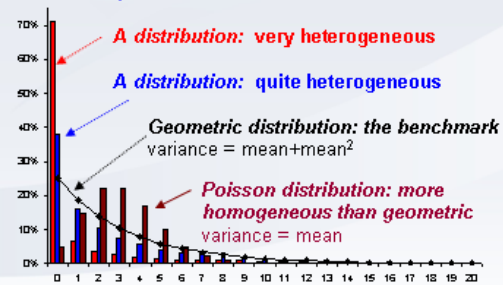


Assuming homogeneity in the environment, population and contacts, the shape of the (discrete) distribution of the secondary transmissions (by a typical infectious individual) resembles the probability density function of the infectious period. (Lynch, J., *Scan. J. Stat.* 1988).

Probability density of the infectious period



Probability function of N:  $\Pr\{N = j\}$ ,  $j = 0, 1, \dots$



$$\text{var}[N] = R_0 + \beta^2 \text{var}[\text{infectious period}]$$

The shapes of the probability density functions of the infectious period (as a continuous random variable) correspond to similar shapes of the probability mass functions for the discrete random number N (such that  $R_0 = E[N]$ ). There is a probability theory for this, but beyond the scope of this presentation. For reference,

Lynch, J. (1988) Mixtures, generalized convexity and balayages. *Scandinavian Journal of Statistics*. 15, 203-210.

The geometric distribution for N corresponds to the exponential distribution for the infectious period.

## 2. Heterogeneity in agent-host-environment interface

- Large variance of infectious period leads to heterogeneous phenomenon:



- But if the infectious period is constant, environmental factors (e.g. social network) can produce infectious contacts that lead to exactly the same phenomenon (same distribution).

Regarding small outbreaks when an infectious individual is seeded at the beginning, it is the phenomenon itself that leads to: ***the more heterogeneous,***

- the smaller is the risk of a large outbreak;
- the smaller is the mean final size of a small outbreak with larger variance;
- the smaller is generation-to-extinction in stochastic order for a small outbreak.

Regarding large outbreaks, their dynamics over time (e.g. growth, peak, duration), their final outcomes, and effectiveness of control measures to mitigate the outcomes, different aspects of heterogeneity at tactical level and conceptual level, have different impacts.

## 2. 1. Heterogeneity of latent/infectious periods and implications

**Laplace transform** of a non-negative r.v.:  $L_X(s) = E[e^{-sX}] = \int_0^{\infty} e^{-sx} dF_X(x), s > 0.$  (always exists)

$L_{Latent}(s)$  and  $L_{Infectious}(s)$  are the Laplace transforms for the latent and the infectious periods.

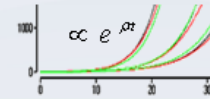
The beginning: "an infectious individual is seeded in a large susceptible population"

$\pi$  is the smallest root of  $G_N(s) = s$   
With prob.  $1 - \pi$ , a large outbreak

The early phase: "depletion of  $S(t)$  is negligible and the infected number is very small"

Under suitable conceptual assumptions about homogeneity:  $i(t) = \beta \frac{S(t)I(t)}{n(t)}$

Exponential growth, Malthusian number  $\rho$



The distribution of  $N$ : determined by the infectious period distribution  $G_N(s) = L_{Infectious}(\beta(1-s))$

On the other hand, if SIR,  $\rho$  satisfies  $\beta \frac{1 - L_{Infectious}(\rho)}{\rho} = 1$  and is unique  
Yan (J. of Theoretical Biology, 2008)

$$\rightarrow L_{Infectious}(\beta(1-\pi)) = \pi$$

$$\rightarrow \beta \frac{1 - L_{Infectious}(\beta(1-\pi))}{\beta(1-\pi)} = 1$$

$$1 - \pi = \frac{\rho}{\beta}$$

Initial growth rate in a "realized" large outbreak vs. the risk of a large outbreak in a repeated "experiment".

The correspondence between the p.g.f. and the Laplace transform function of the infectious period is given here. This correspondence is only true under "suitable conceptual assumptions of homogeneity" which will be discussed in detail later. This correspondence implies that the distribution of  $N$ , the risk of a large outbreak, as well as  $R_0$ , do not depend on whether there is a latent period.

On the right side of the slide, the formula is a special case of a more general formula that depends on the latent period as well. It is found in

Yan, P. (2008) Separate roles of the latent and infectious periods in shaping the relation between the basic reproduction number and the intrinsic growth rate of infectious disease outbreaks. Journal of Theoretical Biology 251, 238-252.

The key result in the slide only works if the underlying model is SIR. The implication is that one can use the observed initial growth rate in a large outbreak to assess the risk of a large outbreak in a similar community, under similar initial conditions (regarding each outbreak as a "random experiment" by nature).

It also implies to "patch models" with an "infected patch" defined as the one with observed large outbreak. Movement of individuals to a susceptible patch do not necessarily result in another infected patch, but with a probability of it. This probability can be modelled, if both "beta" and "rho" are estimated.



## 2. 1. Heterogeneity of latent/infectious periods and implications

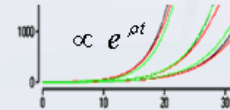
$L_{Latent}(s)$  and  $L_{Infectious}(s)$  are the Laplace transforms for the latent and the infectious periods.

Under suitable conceptual assumptions about homogeneity :  $i(t) = \beta \frac{S(t)I(t)}{n(t)}$

**Regarding the initial growth:** Special case of the Euler-Lotka equation  $\int_0^{\infty} e^{-\rho x} \beta(x) A(x) dx = 1$

**Statement** If SEIR,  $\rho$  satisfies  $\beta L_{Latent}(\rho) \frac{1 - L_{Infectious}(\rho)}{\rho} = 1$  Yan (J. of Theoretical Biology, 2008)

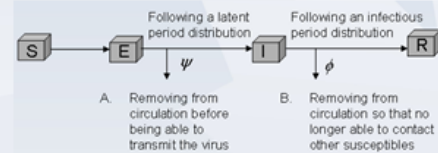
$\rho$  is separately ranked according to Laplace transform orders of the latent and the infectious periods.



**Regarding the controlled reproduction number:** Yan and Feng (Mathematical Biosciences, 2010)  
Limit to measures on infected, not susceptible, individuals

**Statement** Assuming both actions are "perfect" (100% effective)

$$R_c(\psi, \phi) = \beta L_{Latent}(\psi) \frac{1 - L_{Infectious}(\phi)}{\phi}$$



$R_c$  is separately ranked according to Laplace transform orders of the latent and the infectious periods.

(quite unlike  $R_0$ )

The general formula on the initial growth is from

Yan, P. (2008) Separate roles of the latent and infectious periods in shaping the relation between the basic reproduction number and the intrinsic growth rate of infectious disease outbreaks. Journal of Theoretical Biology 251, 238-252.

The formula on the controlled reproduction number is from

Yan, P. and Feng, Z. (2010) Variability order of the latent and the infectious periods in a deterministic SEIR epidemic model and evaluation of control effectiveness. Mathematical Biosciences 224, 43-52.

## 2. 1. Heterogeneity of latent/infectious periods and implications

$\rho$  is separately ranked wrt. Laplace transforms of the latent and the infectious periods.

$R_c$  is separately ranked wrt. Laplace transforms of the latent and the infectious periods.

### Recall: Heterogeneity

“the more heterogeneous, the more likely it takes “extreme” values.”

**Convex order:**  $X_1 \leq_{cx} X_2$  if  $E[\Phi(X_1)] \leq E[\Phi(X_2)]$  for all convex function  $\Phi(x)$ .

**Dilation order** of two non-negative r.v.'s with mean values  $\mu_1, \mu_2$

$X_1$  is smaller than  $X_2$  in dilation order, denoted as  $X_1 \leq_{dil} X_2$

if  $E[\Phi(X_1 - \mu_1)] \leq E[\Phi(X_2 - \mu_2)]$  for all convex function  $\Phi(x)$ .

If comparing two non-negative r.v.'s with equal mean  $\mu_1 = \mu_2 = \mu$ ,

$$X_1 \leq_{cx} X_2 \Leftrightarrow X_1 \leq_{dil} X_2$$

$$X_1 \leq_{cx} X_2 \Rightarrow L_{X_1}(s) = E[e^{-sX_1}] \leq E[e^{-sX_2}] = L_{X_2}(s)$$

$$\Downarrow$$

$$\text{var}[X_1] = E[(X_1 - \mu)^2] \leq E[(X_2 - \mu)^2] = \text{var}[X_2]$$

**Heterogeneity (by convex order) ranks Laplace transforms and variances.**

There are two excellent textbooks about different kinds of variability orders. The results displayed here are only a small part of a broader theory, that are directly related to the current subject.

Marshall, A.W. and Olkin, I. (2007) Life Distributions, Structure of Nonparametric, Semiparametric and Parametric Families. Springer.

Shaked, M and Shanthikumar, J.G (2007) Stochastic Orders. Springer.

## 2. 1. Heterogeneity of latent/infectious periods and implications

Homogeneity / heterogeneity of the *latent period* and implications

Of latent periods of equal mean values.

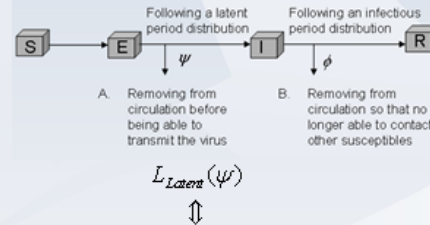
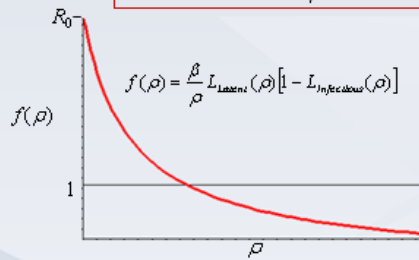
homogeneous: good; heterogeneous: bad.

**Regarding the initial growth:**

**Regarding the controlled reproduction number:**

$$\beta L_{Latent}(\rho) \frac{1 - L_{Infectious}(\rho)}{\rho} = 1$$

$$R_c(\psi, \phi) = \beta L_{Latent}(\psi) \frac{1 - L_{Infectious}(\phi)}{\phi}$$



Given  $\beta$  and the infectious period distribution,

the *more homogeneous* the latent period (smaller in convex order),

- the *smaller* is  $L_{Latent}(\rho)$ ,
- the *smaller* is the initial growth rate  $\rho$ ;
- the *slower* is the growth.

% of latent individuals who escape from being removed (under constant rate  $\psi$ ) and become infectious

The *smaller* the  $L_{Latent}(\psi)$ , the *larger* is the prob. of latent individuals being removed; the *easier* it is to use A. to control the epidemic.

This is a synthesis among the two papers:

Yan, P. (2008) Separate roles of the latent and infectious periods in shaping the relation between the basic reproduction number and the intrinsic growth rate of infectious disease outbreaks. *Journal of Theoretical Biology* 251, 238-252.

Yan, P. and Feng, Z. (2010) Variability order of the latent and the infectious periods in a deterministic SEIR epidemic model and evaluation of control effectiveness. *Mathematical Biosciences* 224, 43-52.

and theories from

Marshall, A.W. and Olkin, I. (2007) *Life Distributions, Structure of Nonparametric, Semiparametric and Parametric Families*. Springer.

Shaked, M and Shanthikumar, J.G (2007) *Stochastic Orders*. Springer.

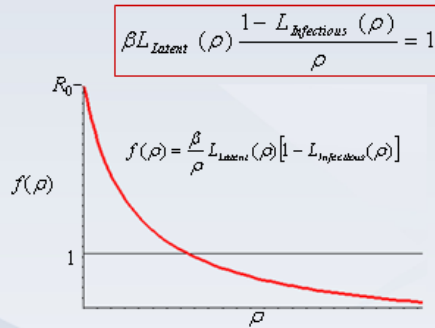
## 2. 1. Heterogeneity of latent/infectious periods and implications

Homogeneity / heterogeneity of the *infectious period* and implications

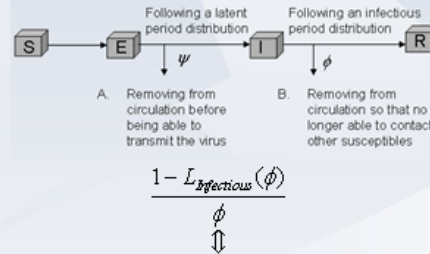
Of infectious periods of equal mean values. homogeneous: bad; heterogeneous: good.

**Regarding the initial growth:**

**Regarding the controlled reproduction number:**



$$R_c(\psi, \phi) = \beta L_{\text{Latent}}(\psi) \frac{1 - L_{\text{Infectious}}(\phi)}{\phi}$$



Given  $\beta$  and the latent period distribution,

- the *more homogeneous the infectious period*, smaller  $L_{\text{Infectious}}(\rho) \Rightarrow$  larger  $\frac{1 - L_{\text{Infectious}}(\rho)}{\rho}$ ,
- the *larger is the initial growth rate  $\rho$* ;
- the *faster is the growth*.

the average duration in the I-class before recovery or removal by control measure (under constant rate  $\phi$ ).

The *smaller* the  $L_{\text{Infectious}}(\phi)$ , the *longer* is the average infectiousness duration; the *harder* to use B. to control.

Ditto: This is a synthesis among the two papers:

Yan, P. (2008) Separate roles of the latent and infectious periods in shaping the relation between the basic reproduction number and the intrinsic growth rate of infectious disease outbreaks. *Journal of Theoretical Biology* 251, 238-252.

Yan, P. and Feng, Z. (2010) Variability order of the latent and the infectious periods in a deterministic SEIR epidemic model and evaluation of control effectiveness. *Mathematical Biosciences* 224, 43-52.

and theories from

Marshall, A.W. and Olkin, I. (2007) *Life Distributions, Structure of Nonparametric, Semiparametric and Parametric Families*. Springer.

Shaked, M and Shanthikumar, J.G (2007) *Stochastic Orders*. Springer.

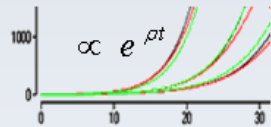
## 2. 1. Heterogeneity of latent/infectious periods and implications

*Some further implications*

Under suitable conceptual assumptions about homogeneity so that  $i(t) = \beta \frac{S(t)I(t)}{n(t)}$

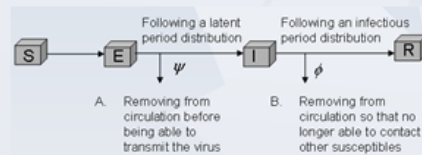
**Regarding the initial growth:**

$$\beta L_{Latent}(\rho) \frac{1 - L_{Infectious}(\rho)}{\rho} = 1$$



**Regarding the controlled reproduction number:**

$$R_c(\psi, \phi) = \beta L_{Latent}(\psi) \frac{1 - L_{Infectious}(\phi)}{\phi}$$



If one can set control objectives  $\psi \geq \psi_c$  and  $\phi \geq \phi_c$  in order to achieve  $R_c(\psi, \phi) \leq 1$ , then one can successfully prevent a large outbreak from taking place.

Ideally, it is achievable if  $\psi_c = \phi_c = \rho$ , assuming both actions are "perfect" (100% effective).

Lessons can be learned from observed initial growth rate in large outbreak that have happened elsewhere to set control targets to prevent a large outbreak from happening in a similar community, under similar initial conditions.

This is a synthesis between the two papers:

Yan, P. (2008) Separate roles of the latent and infectious periods in shaping the relation between the basic reproduction number and the intrinsic growth rate of infectious disease outbreaks. *Journal of Theoretical Biology* 251, 238-252.

Yan, P. and Feng, Z. (2010) Variability order of the latent and the infectious periods in a deterministic SEIR epidemic model and evaluation of control effectiveness. *Mathematical Biosciences* 224, 43-52.

Clearly, if "psi" and "phi" both take the value "rho", then  $R_c = 1$ .

In other words, control measures are race against time.

## 2. 1. Heterogeneity of latent/infectious periods and implications

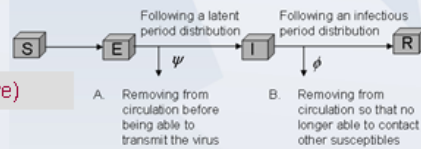
Extension for: **the controlled reproduction number**

Under suitable conceptual assumptions about homogeneity :  $i(t) = \beta \frac{S(t)I(t)}{n(t)}$

Before:

$$R_c(\psi, \phi) = \beta L_{Latent}(\psi) \frac{1 - L_{Infectious}(\phi)}{\phi}$$

assuming both actions are "perfect" (100% effective)



Extension: "leaky situations"

1. For action B, infected individuals may be put into "leaky isolation", with reduced transmissibility  $(1 - \sigma_I)\beta$ .
2. For action A, latent individuals may be put into "leaky isolation", and when they become infectious, they have reduced transmissibility  $(1 - \sigma_L)\beta$ .

**Quantitative:**  $R_c(\psi, \phi | \sigma_L, \sigma_I) = (1 - \sigma_L)R_0 + (\sigma_L - \sigma_I)L_{Latent}(\psi)R_0 + \sigma_I \beta L_{Latent}(\psi) \frac{1 - L_{Infectious}(\phi)}{\phi}$

Yan and Feng (*Mathematical Biosciences*, 2010)

**Qualitative: variability of the latent / infectious periods on control measures remains**

Latent period: homogeneous good; heterogeneous bad;  
 Infectious period: **homogeneous bad; heterogeneous good.**

For details:

Yan, P. and Feng, Z. (2010) Variability order of the latent and the infectious periods in a deterministic SEIR epidemic model and evaluation of control effectiveness. *Mathematical Biosciences* 224, 43-52.

## 2. 1. Heterogeneity of latent/infectious periods and implications

*Extension for: **the controlled reproduction number***

Under suitable conceptual assumptions about homogeneity :  $i(t) = \beta \frac{S(t)I(t)}{n(t)}$

**Qualitative: variability of the latent / infectious periods on control measures remains**

Latent period: homogeneous good; heterogeneous bad;  
Infectious period: **homogeneous bad; heterogeneous good.**

**Also applicable to other measures applied to individuals during their latent and infectious periods:**

1. Control measures, such as contact tracing for exposed individuals with subsequent quarantine and/or pharmaceutical interventions ( prophylaxis), work well if there is a significantly long latent period, and not so well if the latent period is very short.

Add: Such measures work well if the latent period is a long and not very variable (homogeneous). They may not work well if there is large variation (heterogeneous), even if the latent period is long on average.

2. Isolating infectious individuals and/or treating them using antiviral drugs that may reduce transmission, work better if the natural infectious period is short.

Add: Such measures work well if the infection period has large variation (heterogeneous), even when the infectious period is long on average.

These are ad hoc arguments, without rigorous proof.

## 2. 1. Heterogeneity of latent/infectious periods and implications

Some further implications

$$\beta L_{\text{Latent}}(\rho) \frac{1 - L_{\text{Infectious}}(\rho)}{\rho} = 1;$$

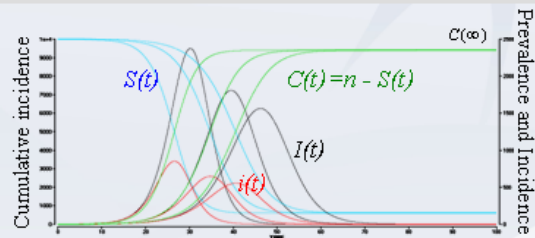
$\rho$  is separately ranked according to the Laplace transform orders (i.e. variability) of the latent and the infectious periods.

The distributions of the latent and the infectious periods determines the dynamics over time.

However,  $R_0$ ,

$$C(\infty) = \int_0^{\infty} i(t) dt, \quad \int_0^{\infty} I(t) dt = \mu_I \times C(\infty)$$

- independent from the latent period;
- dependent on the infectious period only through its mean.

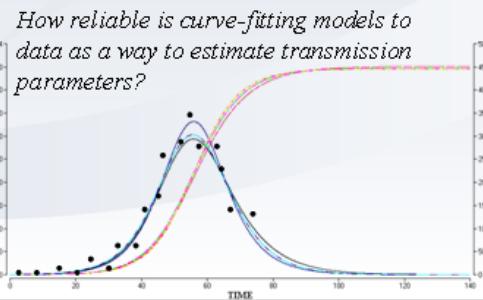


Same  $R_0 \rightarrow$  same total areas

different paths for  $i(t)$ : incidence;  
different paths for  $I(t)$ : prevalence

Conversely, for very different  $R_0$ :

different distributions for the latent and infectious periods can produce the same or very similar curves for  $i(t)$  or  $I(t)$ .



This is my long standing opinion about estimating key epidemic parameters such as  $R_0$ , based on observed exponential growth rate.

There are various formulae in the literature, but each of them is crucially dependent on the underlying (and mostly hidden) tactical assumptions about the natural history of an infected individual.

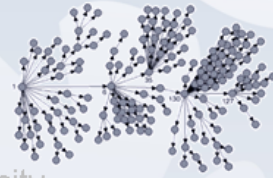
However, I have not been able to provide a better method than those widely used in the literature and understand the importance of estimating  $R_0$  at the very early stage of an epidemic.



## Highlight

### 1. Phenomenological observation

“A picture is worth 1000 words.”



#### 1.1. Implications of phenomenological heterogeneity

*At the very beginning:* as an infectious individual seeded  
in an infinitely large susceptible population ... ..

### 2. Heterogeneity in the agent-host-environment interface

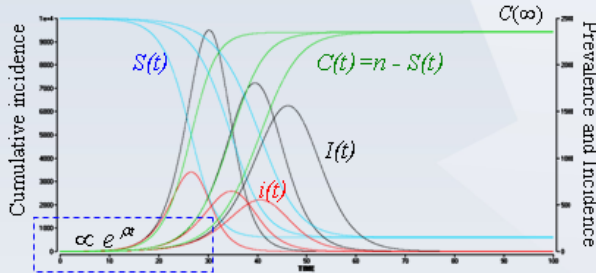
- Conceptual assumptions vs. tactical assumptions

2.1. Heterogeneity in tactical assumptions, implications on dynamics

2.2. Heterogeneity in conceptual assumptions, implications on dynamics

## 2. 2. Heterogeneity in conceptual assumptions and implications

We have emphasized: under suitable conceptual assumptions on homogeneity:  $i(t) = \beta \frac{S(t)I(t)}{n(t)}$



- prevalence = # individuals in a "state":  $S(t), I(t)$ .

- (instantaneous) incidence  $i(t) = \text{instantaneous infection}$

- cumulative incidence  $C(t) = \int_0^t i(u) du$ .

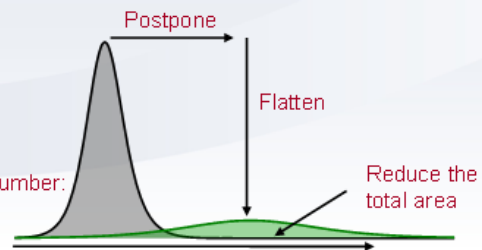
1. Distributions for the latent / infectious periods lead to different paths for  $i(t), I(t), S(t)$  and  $C(t)$ .
2. Given the basic reproduction number  $R_0$  (the beginning of the epidemic), as long as there is no intervention, the total areas under  $i(t)$  and  $I(t)$  remain unchanged.

Final size:  $C(\infty) = \int_0^{\infty} i(t) dt,$

Total person-time:  $\int_0^{\infty} I(t) dt = \mu_I \times C(\infty)$

3. A public health measure that reduces  $R_0$  to  $R_c < R_0$ , changes both the paths and the areas.

Reduce the reproduction number:  $R_0 \rightarrow R_c$



The very key discussion point from this page onwards is the transcendental relationship between  $R_0$  and the final size.

This relationship is under “the suitable conceptual assumptions on homogeneity”.

We shall tease these assumptions apart and show to what extent these assumptions can be relaxed while the transcendental relationship is still valid.

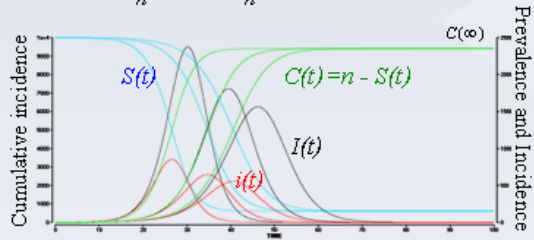
## 2. 2. Heterogeneity in conceptual assumptions and implications

We have emphasized: *under suitable conceptual assumptions on homogeneity* :  $i(t) = \beta \frac{S(t)I(t)}{n(t)}$

This assumption yields a simple expression  $R_0 = \beta\mu_I$ , and a transcendental relationship

$$1 - \eta = (1 - \varepsilon) \exp(-R_0\eta)$$

where  $\varepsilon = \frac{I(0)}{n}$  and  $\frac{E[C(\infty)]}{n} \rightarrow \eta > 0$ , as  $n \rightarrow \infty$ .



Without intervention,

Final size:  $C(\infty) = \int_0^{\infty} i(t) dt,$

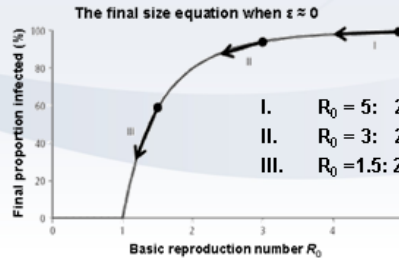
Total person-time:  $\int_0^{\infty} I(t) dt = \mu_I \times C(\infty)$

Both are determined at the very beginning of the epidemic given  $R_0$ .

The transcendental relationship

$$1 - \eta = (1 - \varepsilon) \exp(-R_0\eta)$$

is extremely useful



This shows why it is so important to estimate  $R_0$  at the very early stage of an epidemic.

## 2. 2. Heterogeneity in conceptual assumptions and implications

Bulletin of Mathematical Biology (2006) 68: 679–702  
DOI 10.1007/s11538-005-9047-7

ORIGINAL ARTICLE

Generality of the Final Size Formula for an Epidemic of a Newly Invading Infectious Disease

Junling Ma<sup>a</sup>, David J.D. Earn

*Department of Mathematics & Statistics, McMaster University, Hamilton, ON, Canada L8S 4K1*

As we have seen: distributions of latent and infectious periods change the paths, but not the final outcomes (the total areas).

McKendrick's formula is valid. We show that the final size formula is unchanged if there is a latent stage, any number of distinct infectious stages and/or a stage during which infectives are isolated (the durations of each stage can be drawn from any integrable distribution). We also consider the possibility that the transmission rates of infectious individuals are arbitrarily distributed—allowing, in particular, for the existence of super-spreaders—and prove that this potential complexity has no impact on the final size formula. Finally, we show that the final size formula is unchanged even for a general class of spatial contact structures. We conclude that whenever a new respiratory pathogen emerges, an estimate of the expected

We shall discuss,

- To which extent we can relax some *assumptions on homogeneity*  $i(t) = \beta \frac{S(t)I(t)}{n(t)}$  such that  $1 - \eta = (1 - \varepsilon) \exp(-R_0 \eta)$  is still valid;
- To which extent ..... so that a transcendental relationship still exists (not this equation) such that the beginning determines the end;
- To which extent ..... so that the transcendental relationship fails.

Ma, J. and Earn, D. (2006) Generality of the final size formula for an epidemic of a newly invading infectious disease. Bulletin of Mathematical Biology. 68. 679-702.

The blue underlined texts have been discussed in previous slides.

The red underlined texts will be examined.

## 2. 2. Heterogeneity in conceptual assumptions and implications

About: *suitable conceptual assumptions on homogeneity* :  $i(t) = \beta \frac{S(t)I(t)}{n(t)}$

Conceptual assumptions concern

*environment, population, among individuals,...*

1. **The force of infection** onto a susceptible individual  $\propto \frac{I(t)}{n(t)}$  : % infectious individuals  
 $\rightarrow \beta \frac{I(t)}{n(t)} S(t)$ .

2. **The instantaneous rate** of passing the infection from a typical infectious individual to another  $\propto \frac{S(t)}{n(t)}$  : % of susceptible individuals  
 $\rightarrow \beta \cdot \frac{S(t)}{n(t)} I(t)$ .

One of the (hidden) conceptual assumptions about homogeneity:  $\beta^* = \beta$  (bilinearity)

**In many deterministic models:**

$$\frac{d}{dt} S(t) = -\beta \frac{S(t)I(t)}{n(t)} + (\text{susceptible replacement}) - (\text{non-disease related depletion})$$

**In stochastic, Markov models:**

$$\Pr \left\{ \begin{array}{l} S(t+dt) = s-1 \\ I(t+dt) = i+1 \end{array} \middle| \begin{array}{l} S(t) = s \\ I(t) = i \end{array} \right\} = \beta \frac{S(t)I(t)}{n(t)} dt$$

Another way to describe this conceptual assumption on homogeneity is that the numbers of new infection contacts generated by a typical infectious individual through its infectious period follow a Poisson process.

## 2. 2. Heterogeneity in conceptual assumptions and implications

What does *homogeneity* mean regarding bilinearity  $\beta \frac{S(t)I(t)}{n(t)}$ ?

1. The force of infection onto a susceptible individual:  $\beta \frac{I(t)}{n(t)}$
2. The instantaneous rate of infecting another:  $\beta \frac{S(t)}{n(t)}$



**Agent:** same infectiousness during the study period.

### Host:

- All susceptible individuals are the same: equally susceptible.
- All infectious individuals are the same: equally infectious when infectious period starts.
- An infected individual remains equally infectious throughout its infectious period.

**Environment (homogeneous mixing):** an individual contacts with all other individuals in the population with equal probability. In an infinitely large population, the number of contacts made by a typical individual follows a stationary Poisson process.



$$\beta = \lambda p$$

= contact freq.  $\times$  prob. of infection per contact

- independent of time
- Independent of which contact pair

## 2. 2. Heterogeneity in conceptual assumptions and implications

**Agent:** same infectiousness ( ✓ Yes / No)

**Host:**

All susceptible individuals are the same ( ✓ Yes / No)

All infectious individuals are the same ( ✓ Yes / No)

Equally infectiousness during infectious period ( ✓ Yes / No)



**Environment:** homogeneous mixing ( ✓ Yes / No)

If all the answers are "Yes", then we have the bilinear relationship  $\beta \frac{S(t)I(t)}{n(t)}$ .

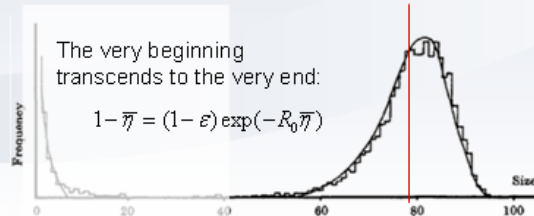
We also have the expression:  $R_0 = \beta \mu_I$ , where  $\mu_I$  = average infectious period.

If the population is closed with size  $n$ , without replacement of susceptible individuals, the **final size** of the epidemic is meaningful:

$$\eta = \frac{C(\infty)}{n}, \text{ with mean } \bar{\eta} > 0.$$

where  $C(\infty)$  is the cumulative number of infected individuals as  $t \rightarrow \infty$ .

The final size distribution of a large outbreak  
Mean value  $\bar{\eta}$



Not explicitly mentioned here is that when the population size is very large (much larger than  $n = 100$  as displayed in the simulation), the final size distribution is asymptotically Gaussian. There is a central limit theorem to prove this.

## 2. 2. Heterogeneity in conceptual assumptions and implications

**Agent:** same infectiousness (✓ Yes / No)

**Host:**

All susceptible individuals are the same (✓ Yes / No)

All infectious individuals are the same (✓ Yes / No)

Equal infectiousness during infectious period (Yes / **X** No)

**Environment:** homogeneous mixing (✓ Yes / No)

*Staging*

Staged infectious period with different infectiousness

$$\beta_j : j = 1, \dots, k$$

assigned to each stage.



**What have changed:**

$$1. \quad \beta \frac{S(t)I(t)}{n(t)} \text{ becomes } (\beta_1 I_1(t) + \dots + \beta_k I_k(t)) \frac{S(t)}{n(t)}$$

where  $I_j(t)$  represents the numbers of infectious individuals in stage  $j = 1, \dots, k$ .

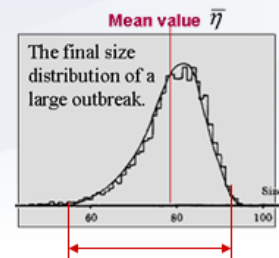
$$2. \quad R_0 = \beta \mu_I \text{ becomes } R_0 = \beta_1 \mu_I^{(1)} + \dots + \beta_k \mu_I^{(k)}$$

$\mu_I^{(j)}$  = average time of the  $j^{\text{th}}$  stage of the infectious period

**What does not change:**  $1 - \bar{\eta} = (1 - \varepsilon) \exp(-R_0 \bar{\eta})$

- no effect to the **mean final size** (*Ma and Earn, 2006*)

**But:** - if the variation among  $\beta_j$ 's is small, the more the staging, the **smaller** is the **variance for the final size**  $\eta$ .





## 2. 2. Heterogeneity in conceptual assumptions and implications

**Agent:** same infectiousness (  Yes / No )

**Host:**

All susceptible individuals are the same ( Yes /  No )

All infectious individuals are the same ( Yes /  No )

Equal infectiousness during infectious period (  Yes / No )

**Environment:** homogeneous mixing (  Yes / No )

### Structured population

multiple types of susceptibles

multiple types of infectives

Theories are separately developed in deterministic compartment models, and in multi-type branching processes frameworks.

For single type of susceptibles and single type of infectives:

$$R_0 = \beta \mu_I, \text{ where } \mu_I = \text{average infectious period}$$

For structured populations: slightly complicated expression but straightforward generalization

$R_0$  = the dominant eigen-value (spectral radius) of the second generation matrix.

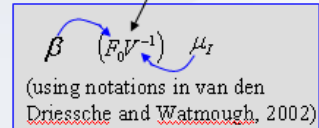
For single type of susceptibles and single type of infectives:

$$\text{Explicit final size equation: } 1 - \bar{\eta} = (1 - \varepsilon) \exp(-R_0 \bar{\eta})$$

For structured populations:

Analogous relationships have been developed, although complicated.

(Ludwig, 1975; Scalia-Tomba, 1986; Ball, 1986; Addy, Longini, et al. 1991; and many others.)



**What does not change: the very beginning transcends to the very end.**

Ludwig, D. (1975) Final size distributions for epidemics. *Mathematical Biosciences* 23, 33-46.

Scalia-Tomba, G. Asymptotic final size distribution for some chain binomial processes. *Advances in Applied Probability*. 17, (1985)477-495.

Ball, F. (1986) A unified approach to the distribution of the total size and total area under the trajectory of infectives in epidemic models. *Advances in Applied Probability* 18, 289-310.

Addy, GL., Longini, IM. and Haber, M. (1991) A generalized stochastic model for the analysis of infectious disease final size data. *Biometrics* 47, 961-974.

## 2. 2. Heterogeneity in conceptual assumptions and implications

**Agent:** same infectiousness (✓ Yes / No)

**Host:**

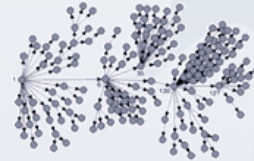
All susceptible individuals are the same (Yes / No ?)

All infectious individuals are the same (Yes / No ?)

Equally infectiousness during infectious period (Yes / No ?)

**Environment:** homogeneous mixing (Yes / No ?)

*Unstructured, randomness in susceptibility, infectivity and mixing (network), with result:*



*Previously addressed:*

- Large variance of infectious period alone leads to heterogeneous phenomenon.
- Even if the infectious period is constant, large variances in other factors questioned above, also produce the same phenomenon.

*Now we address:*

- The "tree" represents a snapshot view of the "infectious contact process" (terminology from Dietz 1995) at a given time, or at the end of an epidemic. It is a snapshot view of the sub-graph of the contact network, of which, infectious contacts occur along its edges.
- It does not contain information on dynamic features on how the infectious contact process, or how the underlying contact network, growing over time.

In stochastic processes terms, for a stochastic process  $\{Y(t):t > 0\}$ , a realization of the process at time  $t$ ,  $Y(t)=y(t)$ , may be represented by the marginal probability has a mar

$\Pr \{Y(t)=y(t)\}$

or the conditional probability

$\Pr \{Y(t)=y(t) \mid \text{given all the past history of } Y(s): s \leq t\}$ .

It is the conditional probability presentation defines the path of the stochastic process over time.

The slide shows an analogous argument, by extending these concepts to a "stochastic graph process".

The snapshot view corresponds to the marginal presentation. It does not contain information of how the graph grows dynamically, which should have been the correspondence of the conditional representation of a stochastic process.

## 2. 2. Heterogeneity in conceptual assumptions and implications

**Agent:** same infectiousness (✓ Yes / No)

**Host:**

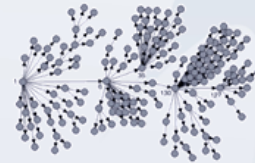
All susceptible individuals are the same (Yes / No ?)

All infectious individuals are the same (Yes / No ?)

Equally infectiousness during infectious period (Yes / No ?)

**Environment:** homogeneous mixing (Yes / No ?)

*Unstructured, randomness in susceptibility, infectivity and mixing (network), with result:*



**Scenario A:**  $\beta = \lambda p$  = contact freq.  $\times$  prob. of infection per contact

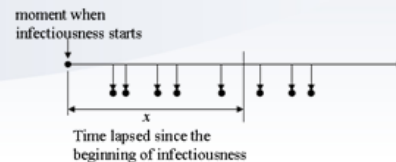
$\lambda$  may be random due to heterogeneous mixing

$p$  may be random due to heterogeneity among individuals susceptibility and/or infectivity

- $\beta$ , as a whole, is no longer a constant, but random, with large (but mostly finite) variation
- The snapshot view of the tree for infectious contacts has very large variance.

**Dynamic of growth:** stationary increment

The expected number of infectious contacts by a typical individual is proportional to the length of time during the infectious period.



The two scenarios in the following slides may give the same snapshot views, but with very different stochastic mechanisms on the dynamic features.

## 2. 2. Heterogeneity in conceptual assumptions and implications

**Agent:** same infectiousness (✓ Yes / No)

**Host:**

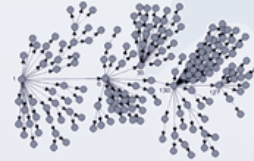
All susceptible individuals are the same (Yes / No ?)

All infectious individuals are the same (Yes / No ?)

Equally infectiousness during infectious period (Yes / No ?)

**Environment:** homogeneous mixing (Yes / No ?)

*Unstructured, randomness in susceptibility, infectivity and mixing (network), with result:*



**Scenario A:**  $\beta$  is random

**Dynamic of growth:** stationary increment

The expected number of infectious contacts by a typical individual is proportional to the length of time during the infectious period.

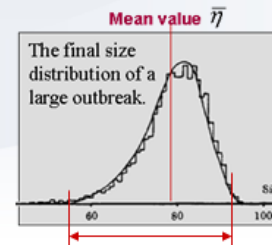
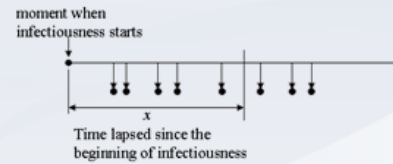
(can be modelled by a mixed-Poisson process)

**Results:**

$R_0 = \bar{\beta} \mu_I$ , where  $\bar{\beta}$  is mean of  $\beta$ ,  $\mu_I$  is mean infectious period

**Still true:**  $1 - \bar{\eta} = (1 - \epsilon) \exp(-R_0 \bar{\eta})$

- but  $\eta$  will have larger variance.



## 2. 2. Heterogeneity in conceptual assumptions and implications

**Agent:** same infectiousness (✓ Yes / No)

**Host:**

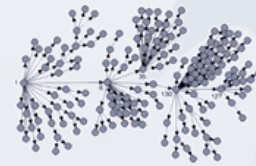
All susceptible individuals are the same (Yes / No ?)

All infectious individuals are the same (Yes / No ?)

Equally infectiousness during infectious period (Yes / No ?)

**Environment:** homogeneous mixing (Yes / No ?)

*Unstructured, randomness in susceptibility, infectivity and mixing (network), with result:*



**Scenario B:** The growth of the tree or the contact network **does not** maintain *stationary increment*. The expected number of infectious contacts by a typical individual is no longer proportional to the length of time during the infectious period.

*Preferential attachment* : the more one attracts others, the larger the probability of making more new connections, the # contacts (in any time interval) follows highly skewed distributions (Yule, Waring, power-law, etc.). (*"scale-free" networks*)

**Variance:** can be extremely large, often does not exist (approaches infinity) !

**Mean:** as defined by mathematical expectation of the form  $\int_0^{\infty} x dF_x(x)$ , may not exit.  
 $R_0$  is no longer meaningful !

**Final size:** No well established transcendental relationship between  $R_0$  (if exists) and the mean value of the final size.

There is an ongoing debate whether preferential attachment actually happen in growing networks. Liljeros, et al. are convinced of preferential attachment as a mechanism for sexual networks, as "people become more attractive the more partners they get." However, Jones and Handcock is skeptical and argues that networks with infinitely large variances but dramatically different structures can manifest the same marginal degree distribution, whereas these different network structures produce different epidemic behaviour.

Liljeros, F., Edling, C.R., and Amaral, L.A.N. Sexual networks: implications for the transmission of sexually transmitted infections. *Microbes and Infection* 3, (2003) 189-196.

Jones, J.H. and Handcock, M.S. An assessment of preferential attachment as a mechanism for human sexual network information. *Proceedings: Biological Sciences, The Royal Society*. 270. (2003). 1123-1128.

The debate between Scenarios A and B has a longer history . In 1919, Greenwood and Woods put forward three hypotheses into the occurrence of accidents:

1. Pure chance, which gives rise to the Poisson process (corresponding to conceptual assumptions on homogeneity in this presentation);

2. True contagion, i.e. initially all individuals have the same probability of incurring an accident, but this probability is modified by each accident sustained to give rise to the linear pure birth process (corresponding to Scenario B);

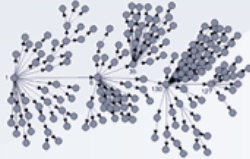
3. Apparent contagion (prone), i.e. individuals have constant but unequal probabilities of having an accident and the resulting process being a mixed-Poisson process (corresponding to Scenario A).

Greenwood, M. and Woods, H.M. On the incidence of industrial accidents upon individuals with special reference to multiple accidents. Report of the Industrial Research Board, No.4. London: His Majesty of Stationery Office. (1919).

PUBLIC HEALTH AGENCY of CANADA | AGENCE DE SANTÉ PUBLIQUE du CANADA

## Future study areas

1. Combine theories of stochastic processes (point processes, counting processes, martingals, etc. ) with random graph theory to address network models for
  - Preferential attachment* : the more one attracts others, the larger the probability of making more new connections, the # contacts (in any time interval) follows highly skewed distributions (Yule, Waring, power-law, etc.).
  - much of such work may have already be published or in development;
  - leave this for more knowledgeable colleagues ([Babak Pourbohloul](#)) to comment.
2. Development of **statistical methodology** to discriminate:
  - Due to large variance of infectious period while other agent-host-environmental factors are homogeneous  
**versus**
  - Due large variances in agent-host-environmental factors ?
  - Scenario A:  $\beta$  is random but the growth of the tree has stationary increment  
**versus**
  - Scenario B: the growth of the tree (or network) has preferential attachment ?



It is important to develop statistical models and methods to distinguish the two scenarios, as well as the kind of data to seek after. The identical marginal (i.e. snapshot) data may arise from two very different stochastic mechanisms which produce quantitatively very different epidemic behaviour.