Research article

Modeling sub-exponential epidemic growth dynamics through unobserved individual heterogeneity: a frailty model approach

Ping Yan[1] and Gerardo Chowell[2,3,*]

[1] Department of Statistics and Actuarial Science, University of Waterloo, Ontario, Canada

[2] School of Public Health, Georgia State University, Atlanta, Georgia, USA

[3] Department of Applied Mathematics, Kyung Hee University, Yongin 17104, Korea

* **Correspondence:** Email: gchowell@gsu.edu.

Abstract: Traditional compartmental models of epidemic transmission often predict an initial phase of exponential growth, assuming uniform susceptibility and interaction within the population. However, empirical outbreak data frequently show early stages of sub-exponential growth in case incidences, challenging these assumptions and indicating that traditional models may not fully encompass the complexity of epidemic dynamics. This discrepancy has been addressed through models that incorporate early behavioral changes or spatial constraints within contact networks. In this paper, we propose the concept of "frailty", which represents the variability in individual susceptibility and transmission, as a more accurate approach to understanding epidemic growth. This concept shifts our understanding from a purely exponential model to a more nuanced, generalized model, depending on the level of heterogeneity captured by the frailty parameter. By incorporating this type of heterogeneity, often overlooked in traditional models, we present a novel mathematical framework. This framework enhances our understanding of how individual differences affect key epidemic metrics, including reproduction numbers, epidemic size, likelihood of stochastic extinction, impact of public health interventions, and accuracy of disease forecasts. By accounting for individual heterogeneity, our approach suggests that a more complex and detailed understanding of disease spread is necessary to accurately predict and manage outbreaks.

Keywords: epidemic modeling; sub-exponential growth; frailty model; parameter identifiability; heterogeneity; transmission dynamics; generalized growth model; infectious diseases; stochastic modeling; forecasting

1. Introduction

The scaling of growth during the initial phase of an infectious disease outbreak is influenced by a complex interplay of factors, many of which are not directly observable. These factors include the nature of the contact network over which the disease spreads [1–5], variability in individuals' risk of infection, and changes in behavior or the implementation of control measures [6]. To place our study in context within the evolution of epidemiological modeling, it is important to understand how these factors contribute to the non-linear dynamics of disease spread. Under the assumption of homogeneous mixing, a scenario where each individual has an equal chance of interacting with others, the incidence curve is expected to grow exponentially during the initial epidemic phase, provided there are no changes in susceptibility, behavior, or interventions [7]. This exponential growth presumes a constant rate intrinsic growth rate, a condition rarely met in real-world scenarios. In contrast, empirical data frequently reveal a sub-exponential growth in the early transmission phase, suggesting a more complex and heterogeneous interaction pattern among individuals. This pattern can emerge from spatial limitations within the contact network partly associated with the transmission mode of the disease, early behavioral modifications, or the introduction of public health interventions [6]. Moreover, the sub-exponential dynamic is intricately linked to the concept of "generation interval contraction", a phenomenon observed when the average time between successive cases in a chain of transmission shortens as the epidemic progresses [8]. This occurs when a susceptible individual is exposed to multiple infectious contacts, creating a competitive scenario for transmission. Such dynamics underscore the critical role of underlying network structures and behavioral responses in shaping the trajectory of an epidemic, leading to the investigation of complex network effects in epidemiological modeling [9–14].

The crucial task of accurately characterizing epidemic growth is underscored by the varied growth patterns observed in empirical data from both historical and recent outbreaks, as documented in the prior studies [15]. These diverse patterns challenge the conventional exponential growth assumption, underscoring the need for developing alternative modeling approaches that reflects the heterogeneity of epidemic processes. The two-parameter generalized-growth model (GGM) offers a nuanced method for estimating the epidemic growth scaling, employing the scaling of growth parameter $0 < p < 1$ [15]. The exploration of this parameter space enables a deeper understanding of the underlying mechanisms driving different epidemic trajectories. Recent research further validates the utility of incorporating these nuanced growth dynamics into epidemic models. Studies have shown that models accounting for the possibility of early sub-exponential growth, rather than just assuming a simple exponential increase, significantly enhance the models' fit to empirical data and improve the accuracy of short-term epidemic forecasts [16–20]. This advancement in modeling aligns more closely with observed epidemic patterns and provides a more reliable foundation for predicting the course of outbreaks, thereby informing more effective public health interventions.

While the generalized-growth model captures the nuances of early sub-exponential growth patterns identified in empirical outbreak data, this model also suggests simple underlying mechanisms for these patterns. However, a comprehensive mathematical framework that integrates the classic mechanistic theories of infectious diseases with the observed variability in early growth dynamics

remains elusive. To bridge this gap, our work introduces the concept of frailty from survival analysis into epidemic modeling, aiming to reflect the observed heterogeneity in individual disease susceptibility and transmission. By incorporating individual-level unobserved heterogeneity into established mechanistic transmission models, we aim to provide a more accurate and nuanced understanding of epidemic spread, better aligning theoretical predictions with empirical observations

2. Model initial epidemic growth in the presence of unobservable heterogeneous infectious contacts through random effect

Understanding the dynamics of epidemic growth is crucial for developing effective public health interventions. Traditional compartmental models often assume homogeneous mixing and uniform susceptibility within a population, leading to predictions of exponential growth during the initial phase of an epidemic. However, empirical observations frequently show sub-exponential growth, indicating that these assumptions may not capture the full complexity of real-world epidemic dynamics. One approach to addressing this discrepancy is the use of phenomenological models, which describe the observed growth patterns without necessarily detailing the underlying transmission mechanisms. To bridge the gap between phenomenological and mechanistic approaches, we introduce a frailty model, which incorporates individual-level heterogeneity into the epidemic growth dynamics. Frailty, a concept from survival analysis, represents unobserved variability in individual susceptibility and transmission potential. This variability can significantly impact the overall epidemic growth rate.

Population growth is described by the instantaneous increase in the total count $C(t)$ at any given time t. This is mathematically represented as:

$$C'(t) = r(t)C(t). \tag{2.1}$$

Within the realm of infectious diseases, $C(t)$ signifies the cumulative count of infections at time t. In the context of our model, $r(t)$ represents the instantaneous growth rate of the epidemic at time t. Unlike the constant growth rate r in the exponential model, $r(t)$ can vary over time to reflect changes in the population's susceptibility and interaction patterns. This variation is influenced by the distribution of frailty among individuals, leading to a more nuanced understanding of epidemic growth.

2.1. The growth rate r(t) in phenomenological and in transmission models

The expression (2.1) has been connected with empirical models to describe population-based phenomena without "mechanically" modelling the transmission dynamics. For instance,

1) The exponential growth equation $C'(t) = rC(t)$ is commonly used to approximate the early stages of an outbreak, assuming a very large population and during the first few generations of disease transmission.
2) The General Growth Model (GGM) as described in [6, 15, 21, 22] is given by the growth equation

$$C'(t) = rC(t)^p, \ 0 \le p \le 1 \tag{2.2}$$

where the rate $r(t) = r/C(t)^{1-p}$ is a function that decreases over time t. Assuming the outbreak began with just one infected individual, meaning $C(0) = 1$, there is a solution given by $C(t) = (1 + r\theta t)^{1/\theta}$, where $\theta = 1 - p$, leading to $r(t) = r/(1 + r\theta t)$. The GGM returns to the exponential growth model when $\theta \to 0$.

3) The logistic growth model is described by the equation:

$$C'(t) = r\left(1 - \frac{C(t)}{K}\right)C(t), \text{ with } r > 0. \tag{2.3}$$

Here, $K = \lim_{t\to\infty} C(t)$ represents the maximum possible value, known as the carrying capacity. It is evident that the growth rate, $r(t) = r\left(1 - \frac{C(t)}{K}\right)$ also diminishes over time in accordance with the logistic rule. This can be explicitly expressed as:

$$r(t) = r\frac{(K - 1)e^{-rt}}{1 + (K - 1)e^{-rt}}.$$

The exponential growth model is the special case when K approaches infinity.

4) The Richards model is represented by the equation:

$$C'(t) = r\left(1 - \left[\frac{C(t)}{K}\right]^{\theta}\right)C(t), \ \theta > 0. \tag{2.4}$$

Beginning with an initial condition of $C(0) = 1$, the model can also be solved analytically for $C(t)$, leading to the expression:

$$r(t) = r\frac{\left(K^{\theta} - 1\right)e^{-r\theta t}}{1 + (K^{\theta} - 1)e^{-r\theta t}}.$$

In this model, when $K \to \infty$, the growth rate decreases over time as a negative exponential function $r(t) \to re^{-r\theta t}$. It approaches constant value r as $\theta \to 0$.

In these models, the growth rate $r(t)$ phenomenologically incorporates a range of influences from both individual behaviors and environmental conditions that impact how the disease spreads. However, it is noteworthy that distinct "mechanical epidemic models" can yield identical phenomenon in terms of the cumulative growth. For instance, the logistic growth rate above may be derived from various deterministic transmission models. This includes, among others, the SI (Susceptible-Infected) and SIS (Susceptible-Infected-Susceptible) models [23].

2.2. The exponential growth approximation

We begin our discussion by looking into the exponential growth dynamics, which is defined by a steady growth rate, denoted as $r(t) = r$. This approximation of exponential growth is typical for the initial phase of an epidemic in various deterministic and stochastic models, under the assumption that the population size is infinitely large, making the depletion in susceptible individuals within the population negligible. These models further assume that mixing and transmission among the population are uniform. The growth rate, r, is then a function of additional parameters that are specific to particular mechanistic models. We will consider the following examples.

In the *Simple Epidemic* model, as outlined by Bailey in [24], also known as the SI model, the count of infectious persons at any specific time t is expressed by the equation $I'(t) = \beta(1 - I(t)/N)I(t)$. In this context, $I(t)$ is equivalent to $C(t)$, and N denotes the entire population size. The parameter β is the transmission rate per contact between susceptible and infectious individuals. As N tends toward infinity, $C(t)$ converges to $i_0 e^{-\beta t}$. Thus, $r = \beta$.

In the SIS and SIR (Susceptible-Infected-Recovered) models, which are framed using ordinary differential equations, the rate of change in the number of infected individuals at any given moment (time t) is described by the formula $I'(t) = \beta S(t)I(t)/N - \gamma I(t)$. In this context, N represents the total population size, and β has the same interpretation as it does in the SI model, indicating the transmission rate. $S(t)$ and $I(t)$ stand for the counts of susceptible and infected individuals at time t, respectively. The γ parameter represents the rate at which infected individuals recover, moving either back to being susceptible or gaining immunity. If the reduction in susceptible individuals is minor, the fraction $S(t)/N$ can be approximated as 1. This simplification leads to the equation $I'(t) = (\beta - \gamma)I(t)$. If $\beta > \gamma$, it suggests that the number of infected individuals, $I(t)$, could increase exponentially without limit. Many standard texts denote $r = \beta - \gamma$, although this assumes conditions of time-homogenity, often described as being "under equilibrium conditions". In this paper, we delve deeper into these equilibrium conditions and examine whether it is accurate to model the initial growth of a disease outbreak, represented by $C(t)$ with the equation $C'(t) = (\beta - \gamma)C(t)$, from the outbreak's onset.

By introducing a new compartment, $E(t)$, which tracks the number of people who are infected but not yet capable of spreading the infection as of time t, the SIS and SIR models are expanded into the SEIS and SEIR models, respectively. These models use ordinary differential equations to calculate how the numbers of latent and infectious individuals evolve over time. Specifically, the equations $E'(t) = \beta\frac{S(t)I(t)}{N} - \alpha E(t)$ and $I'(t) = \alpha E(t) - \gamma I(t)$ describe the rates of change for latent and infectious individuals at any moment. Here, the new parameter α represents the rate at which latent individuals progress to being infectious. Under the same set of equilibrium conditions as before, there exists an implicit function between r and the parameters (β, α, γ) through the equation $(r + \alpha)(r + \gamma) = \beta\alpha$. This relationship is also a recognized finding in standard mathematical epidemiology textbooks such as [25] and others.

These examples assume homogeneity so that the infection spread rate, denoted as β, remains the same for all individuals. Additionally, the natural increase rate, r, is less than or equal to β. The equation $(r + \alpha)(r + \gamma) = \beta\alpha$ can be re-written as $r = \beta\frac{\alpha}{r+\alpha}\frac{r}{r+\gamma} < \beta$. The exception is the SI model, where there is no latent period and once individuals are infected, they stay infectious indefinitely, which means $\alpha = \gamma = 0$, leading to $r = \beta$.

2.3. From stochastic and sampling perspectives

From a stochastic viewpoint, considering $C(t)$ as a counting process, the immediate increase of $C(t)$ given that $C(t) = n$ is explained as follows:

$$\Pr\{C(t + \delta) = n + 1 | C(t) = n\} = h(t)\delta + o(\delta)$$
$$\Pr\{C(t + \delta) = n | C(t) = n\} = 1 - h(t)\delta + o(\delta) \tag{2.5}$$
$$\Pr\{C(t + \delta) > n + 1 | C(t) = n\} = o(\delta).$$

In this context, we posit that every one of the n infected persons independently possesses a positive rate of $r_i(t) > 0$, $i = 1, ..., n$, which contributes to the emergence of a new infection. At the level of

the entire population, the rate of immediate growth is represented by $h(t)$, which is $h(t) = \sum_{i=1}^{n} r_i(t)$. Assuming $r_i(t) = r(t)$ for all $i = 1, ..., n$, we arrive at $h(t) = nr(t)$.

We further investigate the scenario of uninterrupted growth when the rate of change is constant $r(t) = r$. This situation is linked to an exponentially distributed random variable Y associated with a typical infected individual, characterized by the survival function $\overline{G}(t) = \Pr(Y > t) = e^{-rt}$. The random variable Y is well-defined and can be quantified based on the timing for each person who gets infected, thanks to its exponential distribution and the property of being memoryless.

The stochastic counterpart of the exponential growth $C'(t) = rC(t)$ is the conditional probability

$$\Pr\{C(t + h) = n + 1 | C(t) = n\} = nrh + o(h).$$

The marginal distribution of $C(t)$ at time t is negative binomial, with mean value satisfying the exponential growth $E[C(t)] = i_0 e^{rt}$, where $i_0 = C(0)$ is the initial condition. Consequently, we arrive at the formulas:

$$r = \frac{d}{dt} \log C(t) = -\frac{d}{dt} \log \overline{G}(t), \tag{2.6}$$

where the first formula can be interpreted as r representing the growth rate of the total number of infected individuals; and the second formula views r as the hazard rate of the exponential distribution, considering it from the viewpoint of an average individual who is infected.

Yan and Chowell [23] highlight that, if the growth is exponential with constant rate r, it can be expressed by adjusting the transmission rate β based on the probability of individuals within the population being infectious at any particular moment, by considering a specific sample taken at time t. A binary indicator, $\Delta_i(t)$, is defined to indicate whether an individual i, who was infected prior to time t, is infectious at time t. This is represented by the probability $\Pr(\Delta_i(t) = 1)$, where a value of 1 signifies the individual is infectious, and a value of 0 indicating otherwise. Those individuals for whom $\Delta_i(t) = 1$ are grouped into what is called a "prevalence cohort". Under suitable equilibrium conditions, to be elaborated on in the following subsection, the relationship

$$r = \beta \Pr(\Delta_i(t) = 1) \tag{2.7}$$

is established, indicating that the probability $\Pr(\Delta_i(t) = 1)$ stays constant, regardless of the time of sampling, t.

The expressions $r = \beta r (r + \gamma)^{-1}$ and $r = \alpha \beta r (r + \alpha)^{-1} (r + \gamma)^{-1}$ in the aforementioned SIS, SIR, SEIS and SEIR models are the special cases of (2.7) where the durations of both the latent and infectious phases are assumed to follow exponential distributions with rates α and γ, respectively. This underlying assumption is crucial for models that utilize ordinary differential equations.

We begin by defining the random variable T_I, which symbolizes the duration of infection among infected people. This duration follows a probability distribution with a density function denoted as $f_I(x)$, where x measures time starting from when a person becomes infectious. If we ignore any latent period, assuming individuals are infectious right from the infection's start and Y is independent of T_I, the probability $\Pr(Y > T_I)$ can be calculated using the integral $\int_0^\infty e^{-rt} f_I(t) dt = L[f_I](r)$, where $L[f_I](r)$ is the Laplace transform applied to the probability density function $f_I(x)$. This leads us to understand that, if all infected individuals have the same constant hazard rate r to produce new infections, then the chance of an individual causing a new infection at time t while still infectious is given by

$$\Pr(\Delta_i(t) = 1) = \Pr(Y \leq T_I) = 1 - L[f_I](r).$$

The probability $\Pr(\Delta_i(t) = 1)$ is established through a stochastic comparison between the random variable Y and the infectious period T_I of infected individuals. Specifically, in instances where the infectious period follows an exponential distribution with an average γ^{-1}, the expression $r = \beta(1 - L[f_I](r)) = \beta r/(r + \gamma)$ simplyfies to $r = \beta - \gamma$.

Now we introduce a random variable T_E, representing the period of latency in individuals who are infected but not yet capable of transmitting the infection. We consider T_E and T_I to be sequential and independent of each other. Assuming Y is also independent from both T_E and T_I, we start with $L[f_E](r) = \Pr(Y > T_E)$ and $\Pr(T_E < Y \le T_E + T_I) = \Pr(Y > T_E)\Pr(T_E < Y \le T_E + T_I | Y > T_E)$. Given that the distribution of Y is exponential and memoryless, $\Pr(T_E < Y \le T_E + T_I | Y > T_E) = \Pr(Y \le T_I) = 1 - L[f_I](r)$. Thus, $\Pr(\Delta_i(t) = 1) = \Pr(T_E < Y \le T_E + T_I)$ and

$$r = \beta L[f_E](r)(1 - L[f_I](r)). \tag{2.8}$$

Specifically, in instances where the both T_E and T_I ard exponentially distributed with averages α^{-1} and γ^{-1}, respectively, the formula is expressed as $r = \beta(1 - L[f_I](r)) = \beta r/(r + \gamma)$.

2.4. Multifaceted aspects of homogeneity

We have assumed that all the $C(t) = n$ infected individuals are *homogeneous* in the sense that they carry the same rate to produce a new infection independently from the other $n - 1$ individuals. This rate r is further factorized in (2.7) as the product of transmission rate β and the sampling probability $\Pr(\Delta_i(t) = 1)$.

2.4.1. Time-homogeneity and equilibrium:

The constant growth rate implies that $\Pr(\Delta_i(t) = 1)$ is independent of time t. This is *time-homogeneity*. It requires that the dynamic system is at a state of equilibrium.

In the case without the latent period, (2.8) becomes $r = \beta(1 - L[f_I](r))$ and can be re-written as $\beta(1 - L[f]_I(r))/r = 1$. Integration by parts, one can show that $(1 - L[f_I](r))/r = \int_0^\infty e^{-rt}\overline{F}_I(t)dt$, where $\overline{F}_I(x) = \Pr(T_I > x)$ is the survival function of the infectious period. Thus $r = \beta(1 - L[f_I](r))$ can be also written as $\beta\mu_I \int_0^\infty e^{-rt}\overline{F}_I(x)/\mu_i dt = 1$. Denote $f_w(x) = \overline{F}_I(x)/\mu_i$ and $R_0 = \beta\mu_I$, we have the equation

$$R_0 \int_0^\infty e^{-rt}f_w(t)dt = 1, \text{ where } R_0 > 1. \tag{2.9}$$

The condition $R_0 > 1$ is implied, because $\int_0^\infty f_w(t)dt = 1$ and $L[f_W](r) = \int_0^\infty e^{-rt}f_W(t)dt < 1$.

The Eq (2.9) is recognized as the Euler-Lokta equation in mathematical biology involving two intrinsic parameters, R_0 and r. R_0 is widely known as the basic reproduction number. This equation links these parameters by stabilizing the dynamic system in a state of equilibrium. Furthermore, the function $f_w(x)$ adheres to the characteristics of a probability density function and is widely acknowledged as the equilibrium distribution according to renewal theory in the context of stochastic processes.

Figure 1 shows how the equilibrium distribution works, using an example where there are $C(t) = 8$ individuals at time t, with 6 of them being infectious. As a result, the probability $\Pr(\Delta_i(t) = 1) = 0.75$. Each individual is represented by a beginning and an end point. For a given individual i where $\Delta_i(t) = 1$, two time-related random variables, U_i and V_i, are associated with them at time t. In renewal
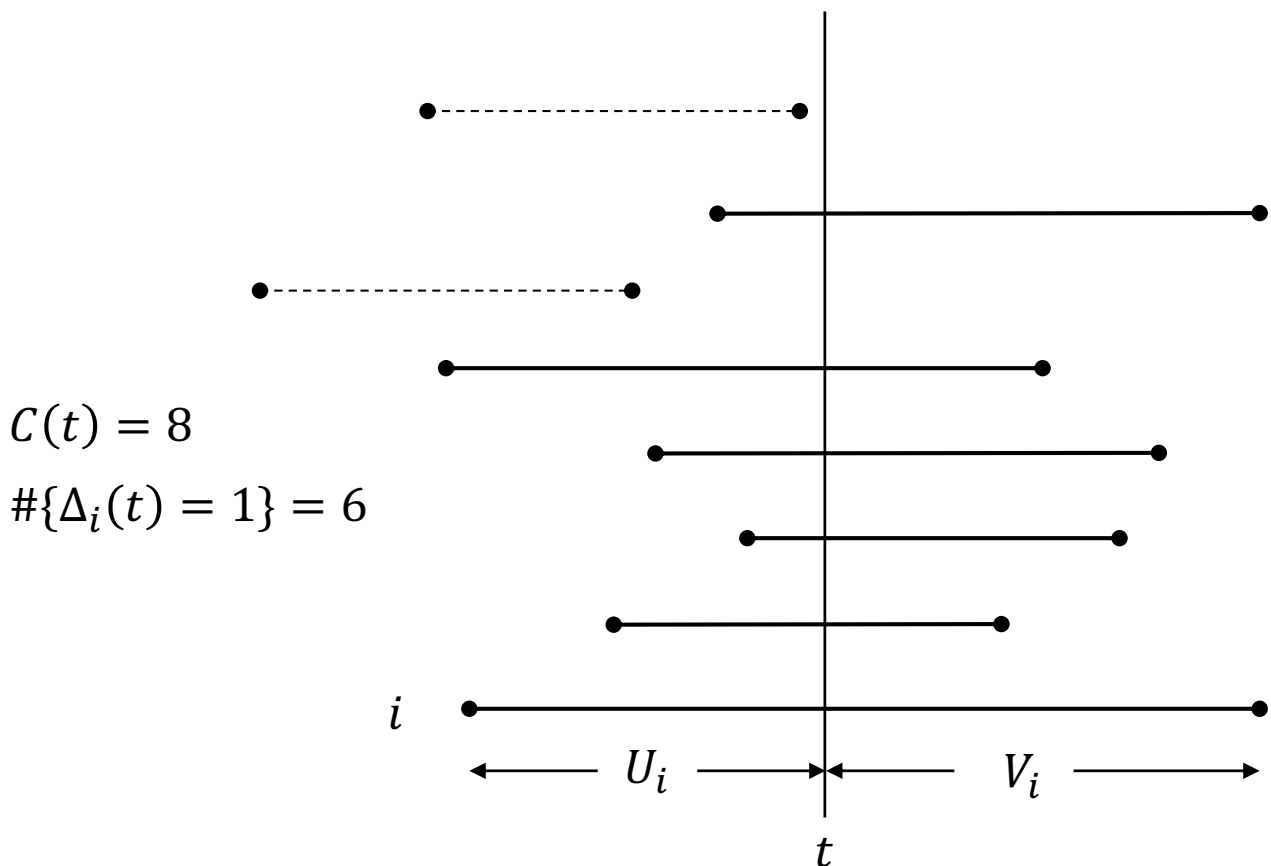
Figure 1. Illustration of $C(t) = 8$ individuals at time t, with 6 of them being infectious. Dotted lines represent the 2 individuals no longer being infectious. Solid lines represent individuals that form the prevalence cohort. This figure demonstrates the concept of the prevalence cohort and highlights the equilibrium distribution in the context of individual infectiousness.

process, U_i refers to the backward recurrence time, which measures the time backward from t to when the individual became infectious. On the other hand, V_i represents the forward recurrence time, measuring the time from t until the individual is no longer infectious. For equilibrium to be achieved, it is necessary that U_i and V_i have the same distribution as the distribution of W_i, with each W_i being independent and identically distributed according to the equilibrium distribution, which has a probability density function denoted by f_w. In addition, the distribution of the sum $T_i^{(B)} = U_i + V_i$ is stochastically longer than that of the infectious period T_I. This occurs because the distribution of infectious periods for individuals in the prevalence cohort at time t is biased towards longer durations. Individuals with longer infectious periods are more likely to be included in the cohort. The probability density of $T_i^{(B)}$ is represented by $x f_I(x)/\mu_I$, which is referred to as the length-biased distribution.

Thus, under equilibrium, the following pair of relationships are equivalent:

$$r = \beta(1 - L[f_I](r)) = \beta \Pr(\Delta_i(t) = 1)$$

$$L[f_W](r) = \int_0^\infty e^{-rt} f_W(t)dt = \Pr(Y > W) = 1/R_0$$

where the second line treats W as the forward recurrent time V_i and each of the infected individual at time t has a probability of $1/R_0$ failing to produce a secondary infection during the infectious period. In the special case of exponentially distributed infectious periods with $\mu_I = \gamma^{-1}$, the equilibrium distribution is identical to the original exponential distribution. Therefore $L[f_W](r) = L[f_I](r) = \gamma/(r + \gamma)$, leading to $r = \beta - \gamma = \gamma(R_0 - 1)$.

Without further elaboration, adding a latent period T_E not only gives (2.8) but also extends (2.9) as

$$R_0 L[f_E](r)L[f_W](r) = 1, \text{ where } R_0 > 1.$$

For further details, we refer to [23].

Challenges to the assumption of equilibrium conditions Even though numerous theoretical findings are based on the assumption of equilibrium conditions, it is improbable that these conditions hold true for data from the initial stages of a major epidemic outbreak. Consider the onset of an outbreak, where nearly every person infected is capable of spreading the disease, resembling the dynamics of an SI model. Here, the likelihood of an individual being sampled, $\Pr(\Delta_i(t) = 1)$, is close to 1, and the rate of infection, r, is approximately equal to β. However, as time goes on and some of the infected individuals recover by time t, the probability $\Pr(\Delta_i(t) = 1)$ drops below 1. Yet, it remains unrealistic to expect that the distributions of U_i and V_i align as they would under equilibrium conditions. When the system is not in equilibrium, the sampling probability not only varies with time t but also the likelihood of each infected individual being included in the prevalence cohort changes based on how recently they were infected relative to time t. In essence, time-nonhomogenity induces one aspect of individual heterogeneity.

2.4.2. Homogeneity assumed in the agent-host-environment interface

The β parameter, often expressed as $\beta = \lambda p$, where λ represents the rate of contact within the social network and p is the probability of transmission during an interaction between an infected and a susceptible person, implies several assumptions about uniformity across the interactions between agents, hosts, and the environment.

Regarding the infectious *agent*, like a virus in viral infections, it is assumed that there are no mutations that alters its infectiousness during the study, ensuring that infections acquired at various times confer the same level of infectiousness at the point of infection.

Concerning the *hosts*, it is assumed that:

1) All susceptible individuals have the same level of susceptibility;
2) An infected individuals maintain a consistent level of infectiousness throughout their infectious period;
3) All infected individuals have equal infectiousness.

In terms of the *environment*, it is assumed that there's uniform mixing, meaning each person is equally likely to interact with every other person in the population. In large populations, this social network can be simplified as a Bernoulli random graph like that described in Erdös and Rényi [26].

Viewing from a stochastic processes angle, the social contact network expands in a manner where the typical individual's number of contacts adheres to a stationary Poisson process with rate λ.

With these homogeneity assumptions, at any given moment during an epidemic, all infected individuals are equally infectious, regardless of when or how long ago they were infected. Each interaction between susceptible and infected individuals carries the same transmission probability p. Together, these assumptions of homogeneity in the agent-host-environment dynamic lead to the formulation $\beta = \lambda p$. This equation defines the rate of new infections in the population as $\beta \frac{S(t)I(t)}{N}$ in deterministic models for disease transmission.

2.5. Frailty as a model to account for non-identifiable and unobservable heterogeneities

The growth rate can be expressed as $r = \lambda p \delta$, where $\delta = \Pr(\Delta_i(t) = 1)$. Data gathered at the group level cannot distinguish λ, p and δ as distinct factors. Each factor reflects specific assumptions about homogeneity. We use a frailty model to describe variability among individuals in a phenomenological manner.

In the field of survival analysis, the proportional hazard model, denoted as $\rho(t|z) = z\rho_0(t)$, accounts for variability in case individuals exhibit varying hazard rates. Here, $\rho_0(t) > 0$ represents a fundamental hazard rate, and $z > 0$ is the frailty variable. This can also be depicted via survival functions, $G(t|z) = e^{-zH_0(t)}$, $H_0(t) = \int_0^t \rho_0(u)du$.. The widely utilized Cox proportional hazard model formulates z as a log-linear regression based on observed variables.

The frailty model is a type of random effect model designed to account for unobserved heterogeneity among individuals. It posits that the variable z is random, with an expected mean of $E(z) = 1$ and a probability density function represented by $\xi(z)$. In scenarios where there is no variation among individuals, $\xi(z)$ becomes a singular point where z is exactly 1, leading to no variability. Under these conditions, the survival function is defined as $\overline{G}_0(t) = e^{-H_0(t)}$. Consequently, when the population includes a diverse mix of individuals, the resulting survival function is derived from a combined distribution:

$$\overline{G}^{(mixed)}(t) = \int_0^\infty \overline{G}(t|z)\xi(z)dz = \int_0^\infty e^{-zH_0(t)}\xi(z)dz = L[\xi](H_0(t)) \tag{2.10}$$

where $L[\xi](s) = \int_0^\infty e^{-zs}\xi(z)dz$ is the Laplace transform with respect to $\xi(z)$ and $L[\xi](H_0(t))$ is $L[\xi](s)$ evaluated at $s = H_0(t)$. The hazard function becomes

$$\rho^{(mixed)}(t) = -\frac{d}{dt}\log \overline{G}^{(mixed)}(t) = -\frac{d}{dt}\log L[\xi](H_0(t)). \tag{2.11}$$

We use the frailty framework with a fixed baseline hazard rate, denoted as $\rho_0(t) = r$, and cumulative hazard rate $H_0(t) = rt$. In the frailty model, the hazard rate given frailty z is $\rho(t|z) = zr$, leading to a mixed hazard rate expression $\rho^{(mixed)}(t) = -\frac{d}{dt}\log L[\xi](rt)$. The following general observations are made:

1) For any non-degenerate probability density function $\xi(z)$, the Laplace transform of ξ at rt, satisfies $L[\xi](rt) = \int_0^\infty e^{-zrt}\xi(z)dz \geq e^{-rt}$. Referring to (2.6) when z is constantly 1, $r = \frac{d}{dt}\log C(t) = -\frac{d}{dt}\log \overline{G}(t)$, leading to the relationship $C(t) = \overline{G}(t)^{-1}$. Introducing the random effect (frailty) $\xi(z)$ and assuming $C(0) = 1$, we have $\frac{d}{dt}\log C(t) = -\frac{d}{dt}\log L[\xi](rt)$, resulting in the inequality

$$C(t) = [L[\xi](rt)]^{-1} \leq e^{rt} \text{ for all } t > 0, \tag{2.12}$$

indicating that the growth of $C(t)$ is limited by the exponential function e^{rt} from above.

2) It has been demonstrated [27] that $L[\xi](rt)$ is log-convex. As a survival function, its hazard rate $\rho(t) = -\frac{d}{dt} \log L[\xi](rt)$ decreases monotonically from an initial value of r at time 0. This reflects a decrease in the effective reproduction number, R_t, during the initial phase of epidemic growth, as noted by [21].

3) When comparing the variability of z using convex order (as detailed in the Appendix), it is found that higher variability in z results in a higher value of the Laplace transform $L[\xi](s)$ for all $s > 0$. This, in turn, means that the more variable z is, the higher the value of $L[\xi](rt)$ and thus, the smaller the value of $C(t)$ for any $t > 0$.

2.6. Special cases

2.6.1. Gamma distributed heterogeneity

Let $\xi(z)$ be the p.d.f. of the Gamma distribution with $E(z) = 1$ and variance $var[z] = v > 0$. The Laplace transform is $L[\xi](s) = (1 + sv)^{-1/v}$. Letting $C(0) = 1$, the expression $C(t) = [L[\xi](rt)]^{-1}$ becomes

$$C(t) = (1 + vrt)^{1/v}, \text{ for all } t > 0. \tag{2.13}$$

Meanwhile the instantaneous growth function is

$$r(t) = -\frac{d}{dt} \log L[\xi](rt) = \frac{r}{1 + vrt}.$$

Figure 2 illustrate $r(t)$ and $C(t)$ at different values of $v > 0$. The parameter r is a time-scale parameter and the x-axes in these plots are scaled to $\tau = rt$. We focus on the range $0 < v < 1$. During this range, $(1 + rvt)^{1/v}$ is the same as the GGM (2.2) by letting $v = 1 - p$ and $C'(t) = [r/\{1 + r(1 - p)t\}]C(t)$ can be re-written as $C'(t) = r(1 + r(1 - p)t)^{\frac{p}{1-p}} = rC(t)^p$, corresponding to the phenomenological model (2.2). When $v = 1$, the linear growth $C(t) = 1 + rt$ corresponds to $\overline{G}^{(mixed)}(t) = [L[\xi](rt)] = (1 + rt)^{-1}$ recognized as the survival function of a (continuous time) power-law distribution, analogous to the discrete power-law distribution in social network analysis. $C(t)$ is a convex increasing function bounded by the exponential function from above and the linear function from below. This is the sub-exponential growth feature [6].

We also draw a connection to previous studies [28, 29], which demonstrate that an initial Gamma distribution of agent susceptibility and infectivity in the SIR model leads to power-law sub-exponential growth. These results, although derived through a different theoretical pathways, aligns with our findings on frailty-induced heterogeneity using a Gamma distribution. This similarity suggests a deeper link between different approaches to modeling heterogeneity in epidemics. it is also worth noting the recent work by

2.6.2. Inverse-Gaussian distributed heterogeneity

If $\xi(z)$ be the p.d.f. of an inverse-Gaussian distribution with $E(z) = 1$ and variance $var[z] = v$, the Laplace transform is $L[\xi](s) = e^{(1 - \sqrt{1+2vs})/v}$. Letting $C(0) = 1$, the expression (2.12) becomes

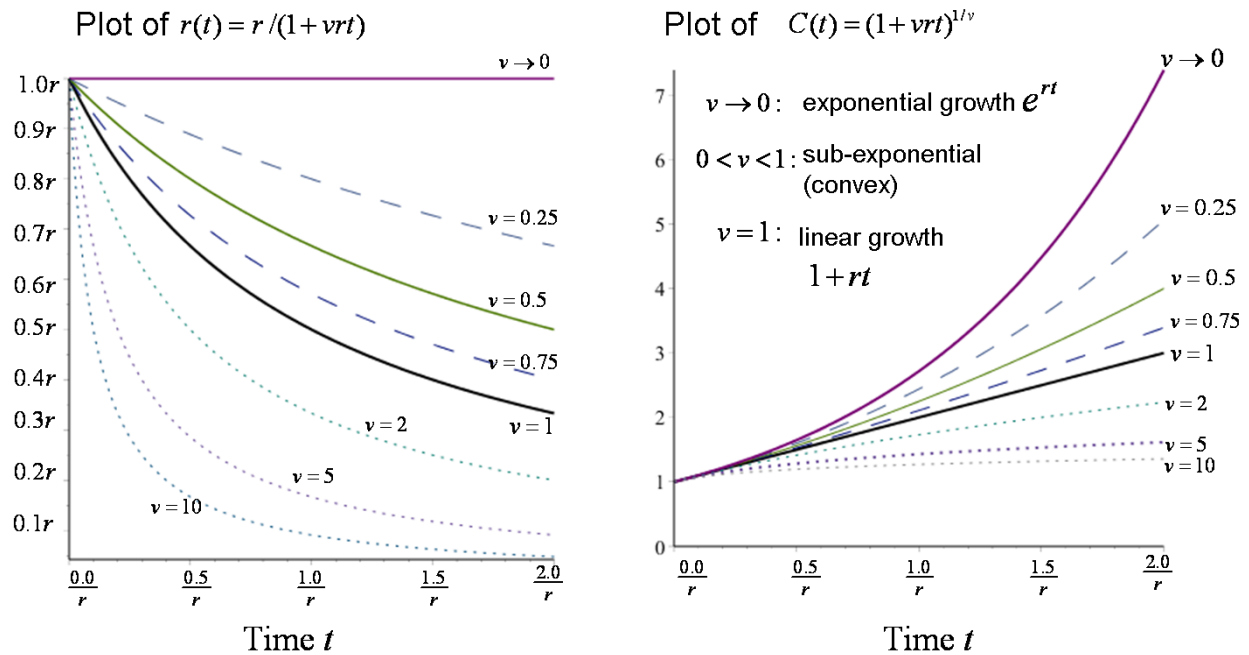$$C(t) = [L[\xi](rt)]^{-1} = e^{(\sqrt{1+2vrt}-1)/v}, \ v > 0 \tag{2.14}$$

Figure 2. Plots of $r(t)$ and $C(t)$ assuming gamma distributed random effect. The plots show how the growth rate $r(t)$ decreases over time, illustrating the impact of frailty on epidemic growth dynamics and resulting in sub-exponential growth patterns

and $C(t) \to e^{rt}$ as $v \to 0$. Meanwhile,

$$r(t) = \frac{C'(t)}{C(t)} = \frac{r}{\sqrt{1 + 2vrt}}.$$

However, the difference between $C(t)$ given in (2.14) and $C(t)$ given in (2.13) is small:

$$e^{\left(\sqrt{1+2vrt}-1\right)/v} - (1 + vrt)^{1/v} = \frac{1}{6}v^2 r^3 t^3 + O\left(t^4\right).$$

It can be shown that, if $E(z) = 1$ and variance $var[z] = v > 0$, $e^{\left(\sqrt{1+2vrt}-1\right)/v} \geq (1 + vs)^{1/v}$ for all $s > 0$. It is a special case of a more general result [30]:

$$e^{\left(1-\sqrt{1+2vs}\right)/v} \leq (1 + v_G s)^{-1/v_G}, \text{ for all } v \leq v_G \text{ and } s > 0 \tag{2.15}$$

where v on the left hand is the variance of an inverse-Gaussian distribution and v_G on the right hand is the variance of a gamma distribution satisfying $v \leq v_G$. In both distributions, it is assumed the mean value $E[z] = 1$. Relating (2.15) to $C(t) = [L[\xi](rt)]^{-1}$, it implies that if one compares two growth curves $C_{IG}(t)$ and $C_{GAM}(t)$, generated by a frailty model with random effect distribution $\xi(z)$ chosen from an inverse-Gaussian distribution with variance v and from a gamma distribution with variance v_G, respectively, if $v \leq v_G$, $C_{IG}(t) \geq C_{GAM}(t)$.

2.6.3. Two distributions for $\xi(z)$ where $L[\xi](s)$ lacks explicit expressions and discussion of sub-exponential growth

If we replace the left hand side in the inequality (2.15) by the Laplace transform of a log-normal distribution with variance v, denoted by $L_{LogN}[\xi](s)$, the following statement is still true.

$$L_{LogN}[\xi](s) \le (1 + v_G s)^{-1/v_G} \text{ , for all } v \le v_G \text{ and } s > 0. \tag{2.16}$$

To show this, we briefly introduce $L\kappa$-class distributions.

Definition 1. *The distribution for $X > 0$ with p.d.f. $f(x)$ and finite mean μ is said to belong to the L_κ-class of distributions, if*

$$L[f](s) = \int_0^\infty e^{-sx} f_X(x) dx \le \left(1 + \frac{\mu s}{\kappa}\right)^{-\kappa}.$$

The p.d.f. of a log-normal distribution of a random variable X can be written as $f_X(x; \gamma, \varsigma) = \frac{\varsigma}{\sqrt{2\pi}x} e^{-\frac{(\varsigma \log(\gamma x))^2}{2}}$ [31] with mean $E[X] = \frac{1}{\gamma} \exp\left(\frac{1}{2\varsigma^2}\right)$ and variance $var[X] = \frac{1}{\gamma^2} e^{\frac{1}{\varsigma^2}} \left(e^{\frac{1}{\varsigma^2}} - 1\right)$. The gamma distribution with mean value μ_G and shape parameter κ has the variance expression $v_G = \mu_G^2/\kappa$. We compare a log-normal distribution with a gamma distribution with equal mean values and let $\mu_G = \frac{1}{\gamma} \exp\left(\frac{1}{2\varsigma^2}\right)$. The condition $var[X] \le \mu_G^2/\kappa$ becomes $e^{\frac{1}{\varsigma^2}} \le 1 + \frac{1}{\kappa_G}$. Immediately following Proposition 3.4 in [30], if $\varsigma^{-2} \le \log\left(1 + \frac{1}{\kappa}\right)$, the log-normal distribution belongs to the $L\kappa$-class.

Applying the above to $\xi(z)$ with mean $\mu = 1$, then $v_G = 1/\kappa$, the condition $\varsigma^{-2} \le \log\left(1 + \frac{1}{\kappa}\right)$ is $v \le v_G$ and the inequality (2.16) holds even though $L_{LogN}[\xi](s)$ does not have explicit expression. Since $C(t) = [L[\xi](rt)]^{-1}$, it implies that if one compares two growth curves $C_{logN}(t)$ and $C_{GAM}(t)$, generated by a frailty model with random effect distribution $\xi(z)$ chosen from a log-normal distribution with variance v and from a gamma distribution with variance v_G, respectively, if $v \le v_G$, $C_{logN}(t) \ge C_{GAM}(t)$.

With respect to the benchmark $v_G = 1$, of the three distributions for $\xi(z)$, when $E(z) = 1$ and variance $var[z] = v \le 1$, they all belong to the L_1−class satisfying $L[\xi](s) \le (1 + s)^{-1}$. This gives $C(t) = [L[\xi](rt)]^{-1} \ge 1 + rt$. We can add the Weibull distribution for $\xi(z)$ into this class because [30] has also shown that the Weibull distribution with $E(z) = 1$ and variance $var[z] = v \le 1$ belongs to the L_1−class. Therefore, for four distributions for $\xi(z)$: the gamma, the inverse-Gaussian, the log-normal and the Weibull distribution, when $E(z) = 1$ and variance $var[z] = v \le 1$, $C(t) = [L[\xi](rt)]^{-1}$ satisfies $1 + rt \le C(t) < e^{rt}$.

It is worth noting that a power-law sub-exponential diffusion growth has been demonstrated on the recent COVID-19 epidemic (for example in [32–36]).

3. Discussion

Most popular compartmental epidemic models treat populations as homogeneous, with only a few key differences considered, such as age, gender, and location [7]. However, capturing all the factors that influence the transmission dynamics of infectious diseases in these models is practically impossible. This is partly because some relevant factors might be unknown, a challenge frequently encountered in epidemiology. Our analysis employs the frailty model from survival analysis, offering

a robust framework for addressing this issue. This model is suitable for incorporating unobserved heterogeneity into the study of epidemic growth dynamics, without the need to define the specific mechanisms. By doing so, it allows us to explore the dynamic impacts of these unaccounted-for variables on disease transmission, providing insights that are not obtained from conventional models.

The frailty model has found increasing use across various applications, notably in epidemiology. In our study, we leverage the concept of frailty to provide a straightforward explanation for the observed diversity in the early growth dynamics of epidemics across different infectious diseases [15]. This approach allows us to delve into the phenomenon of sub-exponential epidemic growth by acknowledging the presence of unseen internal and external factors. These factors contribute to individual-level heterogeneity, a complexity that traditional mechanistic frameworks struggle to fully encompass.

The essence of the frailty framework lies in its ability to capture information about elusive individual heterogeneity through the structure of conditional hazards and the distribution of frailty (or heterogeneity) among individuals. However, a significant challenge we face is that of identifiability. Specifically, the hazard function (2.11) when the population arises from a random mixture of heterogeneous individuals involves a baseline hazard $h_0(t)$ and a random effect distribution $\xi(z)$. Thus, different combinations of $h_0(t)$ and $\xi(z)$ may produce the same marginal hazard rate $h^{(mixed)}(t)$. To overcome this issue, we propose limiting the baseline hazard rate to a constant value, $h_0(t) = r$ and also confining $\xi(z)$ to a specific family of distributions, such as the gamma distribution.

By incorporating individual unobserved heterogeneity into standard mechanistic compartmental epidemic models through the frailty framework from survival analysis offers a significant opportunity to understand the consequences of heterogeneity on infectious disease dynamics and control. Moreover, the use of different statistical distributions to represent frailty highlights the flexibility and robustness of our approach. The connection to previous work [28, 29] further underscores the relevance and potential of incorporating heterogeneity in epidemic modeling. For instance, Novozhilov (2008) examined the spread of epidemics in a closed heterogeneous population by introducing a model that incorporates individual variability in susceptibility and infectivity through an initial Gamma distribution. He demonstrated that this variability leads to power-law sub-exponential growth, highlighting how heterogeneity can significantly alter epidemic trajectories.

While incorporating individual heterogeneity into epidemic models enhances their accuracy and predictive power, it also increases their complexity. This complexity can pose challenges in terms of model availability and computational feasibility. Therefore, a careful balance must be struck between the level of detail included and the practical utility of the model. Our approach aims to provide a nuanced understanding of epidemic dynamics through the concept of frailty, while maintaining a level of simplicity that ensures the model's applicability and ease of use.

In this paper we have distinguished the growth rate r as a scale parameter of time in the growth curve function from the instantaneous growth function $r(t)$ given in (2.1). Importantly, our work here and a recent study [37] suggests that mathematical analyses that rely on the integration of mechanistic

epidemic models and the frailty model are often feasible. For instance, in [37] authors incorporated the frailty framework into mechanistic models to explore the impact of individual variability in the level of compliance on a control intervention. Immediate research in this direction include studies that extend specific mechanistic models with the frailty framework to better understand the impact of individual unobserved heterogeneity on estimates of the reproduction number, epidemic size, stochastic extinction as well as the accuracy of epidemic forecasts.

The proposed frailty model provides a comprehensive mathematical framework to incorporate unobserved heterogeneity into epidemic growth models. By doing so, it bridges the gap between phenomenological observations and mechanistic theories, offering a more nuanced understanding of epidemic spread. This approach aligns with related work in the field, such as Novozhilov (2008) [28], and suggests promising directions for future research in modeling epidemic heterogeneity.

In summary, we demonstrate how the variability in early epidemic growth, ranging from sub-exponential to exponential, can be mathematically derived by incorporating individual unobserved heterogeneity into simple mechanistic models. This is achieved using the frailty model from survival analysis. By adopting the standard assumption that heterogeneity is gamma-distributed, we have derived a generalized growth equation based on the variance of the frailty distribution. These findings are in line with previous work [6], emphasizing the importance of including detailed individual-level characteristics, including dynamic contact networks, in transmission models to aim to capture sub-exponential epidemic growth dynamics.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. L. Meyers, Contact network epidemiology: Bond percolation applied to infectious disease prediction and control, *Bull. Am. Math. Soc.*, **44** (2007), 63–86. https://doi.org/10.1090/S0273-0979-06-01148-7

2. J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, W. M. Getz, Superspreading and the effect of individual variation on disease emergence, *Nature*, **438** (2005), 355–359. https://doi.org/10.1038/nature04153

3. L Danon, A. P. Ford, T. House, C. P. Jewell, M. J. Keeling, G. O. Roberts, et al., Networks and the epidemiology of infectious disease, *Interdiscip. Perspect. Infect. Dis.*, **2011** (2011), 284909.

4. M. J. Keeling, K. T. Eames, Networks and epidemic models, *J. R. Soc. Int.*, **2** (2005), 295–307.

5. S. Towers, Z. Feng, Social contact patterns and control strategies for influenza in the elderly, *Math. Biosci.*, **204** (2012), 241–249.

6. G. Chowell, L. Sattenspiel, S. Bansal, C. Viboud, Mathematical models to characterize early epidemic growth: A Review, *Phys. Life Rev.*, **18** (2016), 66–97. https://doi.org/10.1016/j.plrev.2016.07.005

7. O. Diekmann, J. Heesterbeek, *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*, Wiley, 2000.

8. E. Kenah, M. Lipsitch, J. M. Robins, Generation interval contraction and epidemic data analysis, *Math. Biosci.*, **213** (2008), 71–79. https://doi.org/10.1016/j.mbs.2008.02.007

9. R. Pastor-Satorras, C. Castellano, P. Van Mieghem, A. Vespignani, Epidemic processes in complex networks, *Rev. Mod. Phys.*, **87** (2015), 925–979. https://doi.org/10.1103/RevModPhys.87.925

10. M. J. Keeling, The effects of local spatial structure on epidemiological invasions, *Proc. R. Soc. London Ser. B Biol. Sci.*, **266** (1999), 859–867. https://doi.org/10.1098/rspb.1999.0716

11. M. Barthélemy, A. Barrat, R. Pastor-Satorras, A. Vespignani, Dynamical patterns of epidemic outbreaks in complex heterogeneous networks, *J. Theor. Biol.*, **235** (2005), 275–288. https://doi.org/10.1016/j.jtbi.2005.01.011

12. Y. Wang, Z. Wei, J. Cao, Epidemic dynamics of influenza-like diseases spreading in complex networks, *Nonlinear Dyn.*, **101** (2020), 1801–1820. https://doi.org/10.1007/s11071-020-05867-1

13. P. Ji, J. Ye, Y. Mu, W. Lin, Y. Tian, C. Hens, et al., Signal propagation in complex networks, *Phys. Rep.*, **1017** (2023), 1–96.

14. O. Artime, M. Grassia, M. De Domenico, J. P. Gleeson, H. A. Makse, G. Mangioni, Robustness and resilience of complex networks, *Nat. Rev. Phys.*, **6** (2024), 114–131. https://doi.org/10.1038/s42254-023-00676-y

15. C. Viboud, L. Simonsen, G. Chowell, A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks, *Epidemics*, **15** (2016), 27–37. https://doi.org/10.1016/j.epidem.2016.01.002

16. R. Bürger, G. Chowell, L. Y. Lara-Díaz, Measuring differences between phenomenological growth models applied to epidemiology, *Math. Biosci.*, **334** (2021), 108558. https://doi.org/10.1016/j.mbs.2021.108558

17. A. Smirnova, L. deCamp, G. Chowell, Forecasting epidemics through nonparametric estimation of time-dependent transmission rates using the SEIR model, *Bull. Math. Biol.*, **81** (2019), 4343–4365.

18. D. Shanafelt, G. Jones, M. Lima, C. Perrings, G. Chowell, Forecasting the foot-and-mouth disease epidemic in the United Kingdom, EcoHealth, forthcoming.

19. B. Pell, Y. Kuang, C. Viboud, G. Chowell, Using phenomenological models for forecasting the 2015 Ebola challenge, *Epidemics*, **22** (2018), 62–70. https://doi.org/10.1016/j.epidem.2016.11.002

20. G. Chowell, D. Hincapie-Palacio, J. Ospina, B. Pell, A. Tariq, S. Dahal, et al., Using phenomenological models to characterize transmissibility and forecast patterns and final burden of Zika epidemics, *PLOS Curr.*, **8** (2016).

21. G. Chowell, C. Viboud, L. Simonsen, S. Moghadas, Characterizing the reproduction number of epidemics with early sub-exponential growth dynamics, *J. R. Soc. Interface*, **13** (2016), 20160659. https://doi.org/10.1098/rsif.2016.0659

22. G. Chowell, C. Viboud, Is it growing exponentially fast?–Impact of assuming exponential growth for characterizing and forecasting epidemics with initial near-exponential growth dynamics, *Infect. Dis. Model.*, **1** (2016), 71–78.

23. P. Yan, G. Chowell, *Quantitative methods for investigating infectious disease outbreaks*, Cham, Switzerland, Springer, 2019.

24. N. T. J. Bailey, *The Mathematical Theory of Infectious Diseases and Its Applications*, Second Edition, The Griffin & Company Ltd, 1975.

25. F. Brauer, C. Castillo-Chávez, *Mathematical Models in Population Biology and Epidemiology*, Springer, 2001.

26. P. Erdös, A. Rényi, On the evolution of random graphs, *Bull. Inst. Int. Stat.*, **38** (1961), 343–347.

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

28. A. S. Novozhilov, On the spread of epidemics in a closed heterogeneous population, *Math. Biosci.*, **215** (2008), 177–185. https://doi.org/10.1182/asheducation-2008.1.177

29. C. Rose, A. J. Medford, C. F. Goldsmith, T. Vegge, J. S. Weitz, A. A. Peterson, Heterogeneity in susceptibility dictates the order of epidemic models, *J. Theor. Biol.*, **528** (2021), 110839. https://doi.org/10.1016/j.jtbi.2021.110839

30. B. Klar, A note on the *L*-class of life distributions, *J. Appl. Prob.*, **39** (2002), 11–19. https://doi.org/10.1239/jap/1019737984

31. J. D. Kalbfleisch, R. L. Prentice, *Statistical Analysis for Failure Time Data*, Second Edition, Wiley & Sons, New York, 2002.

32. N. L. Komarova, L. M. Schang, D. Wodarz, Patterns of the COVID-19 pandemic spread around the world: exponential versus power laws, *J. R. Soc. Interface*, **17** (2020), 20200518.

33. D. G. Xenikos, A. Asimakopoulos, Power-law growth of the COVID-19 fatality incidents in Europe, *Infect. Dis. Modell.*, **6** (2021), 743–750. https://doi.org/10.1016/j.idm.2021.05.001

34. A. Tariq, E. A. Undurraga, C. C. Laborde, K. Vogt-Geisse, R. Luo, R. Rothenberg, et al., Transmission dynamics and control of COVID-19 in Chile, March-October, 2020. *PLoS Neglected Trop. Dis.*, **15** (2021), e0009070. https://doi.org/10.1371/journal.pntd.0009070

35. G. Chowell, S. Dahal, A. Tariq, K. Roosa, J. M. Hyman, R. Luo, An ensemble n-sub-epidemic modeling framework for short-term forecasting epidemic trajectories: Application to the COVID-19 pandemic in the USA, *PLoS Comput. Biol.*, **18** (2022), e1010602. https://doi.org/10.1371/journal.ppat.1010602

36. G. Chowell, R. Rothenberg, K. Roosa, A. Tariq, J. M. Hyman, R. Luo, Sub-epidemic model forecasts during the first wave of the COVID-19 pandemic in the USA and European hotspots, in *Mathematics of Public Health: Proceedings of the Seminar on the Mathematical Modelling of COVID-19*, (2022), 85–137.

37. P. Yan, A frailty model for intervention effectiveness against disease transmission when implemented with unobservable heterogeneity, *Math. Biosci. Eng.*, **15** (2018), 275–298. https://doi.org/10.3934/mbe.2018012

Appendix: Order of variability of non-negative random variables

If T_1 and T_2 have equal mean value $\mu_1 = \mu_2$ (should they exist), a general description of variability is based on "marjorization". Let f_1 and f_2 are the corresponding p.d.f. for T_1 and T_2, the verbal description for T_2 *being more dispersed (spread out) than* T_1 is reflected in Figure A1 about the change of signs between. f_1 and f_2 and their corresponding survival functions $\overline{F}_1$ and $\overline{F}_2$.

$f_2(t) - f_1(t)$ has two sign changes, and the sign sequence is +,-,+

$\overline{F}_1(t) - \overline{F}_2(t)$ has one sign changes, and the sign sequence is +,-.
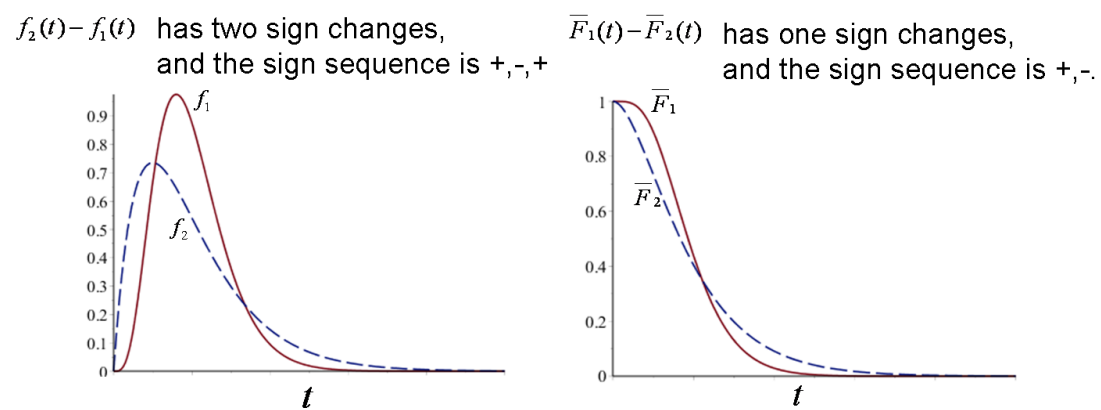
Figure A1. Verbal and graphic presentation for the convex order showing that T_2 is more "spread out" than T_1. This figure explains the concept of the convex order and how it relates to the variability in the distribution of infectious periods.

The verbal and graphic description gives the following definition:

Definition 2. $T_1 \underset{cv}{\leq} T_2$ *if and only if* $\mu_1 = \mu_2$ *plus the following two statements:*

1) $f_2(x) - f_1(x)$ has two sign changes and the sign sequence is: $+, -, +$ (see Figure A1).
2) $\overline{F}_1(x) - \overline{F}_2(x)$ has one sign change and the sign sequence is: $+, -$ (see Figure A1).

The above definition is verbal and has been proven to be equivalent to the following definition in mathematical terms:

Definition 3. $T_1 \underset{cv}{\leq} T_2$ *if* $E[\Psi(T_1)] \leq E[\Psi(T_2)]$ *for all convex functions* $\Psi(x)$ *for which these expectations exist.*

Therefore, the order $T_1 \underset{cv}{\leq} T_2$ is called the convex order.

1) The convex order implies the ordering according to variance $var(T) = E[(T - \mu)^2]$ because x^2 is a convex function.
2) It also implies the ordering according to the Laplace transform $E[e^{-sT}]$ for all $s > 0$, because e^{-sx} is a convex function.