Farr's Law and Social Risk Aversion in the Study of Epidemics

Michael Brimacombe, Ph.D.

CT Children's

Department of Pediatrics, University of Connecticut School of Medicine

Abstract

With the recent COVID-19 epidemic, the modeling of the frequency of new cases of infection within a given time period is of interest. Farr's Law and the related Brownlee-Farr condition on infection counts implies a specific functional form for the related incidence function $I(t)$ when the epidemic is under control. This form includes the Gaussian distribution, but also other distributions such as the exponential distribution. It also includes functions that are not probability densities. The Brownlee-Farr condition is shown here to correspond as a mathematical function to the Pratt-Arrow risk avoidance ARA measure and the hazard rate. This closely connects $I(t)$ to the concept of social risk avoidance with higher levels of infection posing a direct risk to healthy individuals and society, resulting in risk avoidance behavior. It is suggested that the degree of social control over the epidemic can be measured using the entropy of the observed $I(t)$ density.

Keywords: Brownlee-Farr condition, epidemics, entropy, exponential family, Gaussian distribution, hazard rate, IDEA model, mathematical models, Pratt-Arrow relatrive risk, relative risk, risk avoidance, statistical models.

1 Introduction

The modeling of epidemics is a source of much interest in light of the COVID-19 pandemic [1]. The ability to model relative infection incidence rates $I(t)$ as a function of time allows epidemiologists and public health officials to predict possible scenarios and better prepare needed resources. Typically cases are counted and aggregated into weekly or monthly counts and these are standardized in varous ways to provide a data based picture of the epidemic, in the form of probability distributions and related predictions.

There is a long history of using probability distributions and stochastic models to summarize such infection-related data patterns through time [2]. There are various distributions available that can be fit or trained to existing data, depending on the type of data. These include the binomial, poisson, exponential and normal distributions usually in the context of time series data or stochastic processes. When these distributions accurately summarize past data patterns, they may also provide useful predictions regarding the continuation of these patterns [3]. It is useful to note that central limit theorems may also apply in these often large-sample settings, simplifying related probability based calculations.

Interestingly, in regard to the modeling of incidence rates in the context of epidemics, early work by Brownlee [4] and Farr [5] identified a useful mathematical condition on the $I(t)$ function. They showed that a Gaussian distribution, written as a function of time, often arose when the pattern of relative incidence $I(t)$ obeyed this condition in the context of epidemics. The condition implied that the incidence of infection shown by $I(t)$ would diminish over time and in this manner would be controlled.

The Gaussian distribution however is only one of several possible probability densities and functions of time satisfying the Brownlee-Farr (BF) condition. The result has been extended to cover various assumed forms of the $I(t)$ relative incidence distribution, includ-

ing compartmental models involving sets of differential equations linking $I(t)$ to specific mechanisms describing how the infection spreads at the individual level. The SIR model [6] is one of these, as is the IDEA model [7] which has been shown to relate to a specific form of the $I(t)$ function satisfying the BF condition [8].

Note that other functional forms have also been suggested for the modeling of pandemics. These include, for example, the nonlinear Richards model which need not satisfy the BF condition [9].

Social Risk Aversion

The BF condition may be more practically interpreted on a broad, society-wide level, linking the functional form of the $I(t)$ function, and the infectious cohort it represents, to perceived social risk and related behavior on the part of the uninfected portion of the population. The $I(t)$ condition and resulting $I(t)$ function that occurs via the BF condition can be interpreted as resulting from risk avoidance behavior in the healthy population. This suggests why the resulting form of the $I(t)$ function often holds across a wide variety of wide-scale infections; it reflects a basic social risk avoidance response to a dangerous infectious situation.

The connection to society-level social risk and risk avoidance behavior is examined here using the Pratt-Arrow absolute risk aversion (ARA) measure of relative risk [10] as a reference, and its mathematical similarity to the BF condition. The condition can be viewed as reflecting the perspective that as the number of infected increase, the social risk they represent to society also increases in respect to the rest of society, a society that is risk averse in terms of threats to health. The shape of the incidence distribution $I(t)$ can be viewed as a risk factor in itself, both to those currently uninfected and society in general, often reflecting government mandated social behavior. This behavior can be modelled using concepts related to utility theory and risk aversion.

It is shown here that the general form of $I(t)$ resulting from the BF condition can be interpreted as a measure of risk, similar in structure to the ARA measure of relative risk avoidance, interpreted here as a societal level measure in a setting where similar governmental restrictions apply to all individual citizens. This may provide a general context for interpretation of the BF condition in modeling and in understanding when epidemics are controlled within specific time frames, and how this can be measured, for example, by using the entropy of the resulting $I(t)$ distribution.

2 Brownlee-Farr Condition

The pandemic or epidemic setting provides an interesting case study of how time-indexed relative count data values can approximately follow a normal density, when subject to simple restrictions. These patterns may arise multiple times if the infection is, for example, respiratory and subject to seasonal effects [11].

The BF "law" [5] in its simplest form shows that an approximate bell-curve shape can describe the pattern of incidence at time t, $I(t)$, for a designated time period. If $I(t)$ is subject to the restriction:

$$\frac{\frac{I(t+3)}{I(t+2)}}{\frac{I(t+1)}{I(t)}} = K$$

where K is a constant then the incidence rate in the epidemic will occur in a manner described by a function of the general exponential form:

$$I(t) = \exp(-At^2 + Bt + c)$$

which was observed by Brownlee [4]. Note that the ratios involved in the BF condition can be viewed as first and second order percent differences in the time series of monthly

counts, a ratio of re-scaled acceleration to initial percent difference.

To demostrate that this definition of $I(t)$ actually satisfies the condition, write the expression

$$\frac{\frac{\exp(-A(t+3)^2+B(t+3)+c)}{\exp(-A(t+2)^2+B(t+2)+c)}}{\frac{\exp(-A(t+1)^2+B(t+1)+c)}{\exp(-At^2+Bt+c)}} = K \tag{1}$$

and simplify the algebra, showing the left hand side is independent of t.

The BF condition occurs in relation to many possible functions for $I(t)$, not all of them probability density functions. Practically, when the Gaussian distribution does arise, the relative counts $I(t)$ may be directly modeled as:

$$I(t) = f(\theta; t) = N(\mu(t), \sigma^2(t)) \tag{2}$$

with absolute counts given by $n \cdot f(\theta; t)$ at the observed values of t. Note this is only approximately a Gaussian distribution as $I(t)$ is usually a discrete count (daily, weekly or monthly) and the normal curve an approximation.

The BF restriction, often most applicable within a chosen time period, has been observed in several epidemic settings [12, 13], and indicates the epidemic in question is under control. It can be viewed as arising due to (i) the number of potentially infected decreasing over time due to previous infection and vaccination and (ii) patterns of infection-avoidance as non-infected members of society and governments become aware and modify behaviours in a manner affecting all individuals in society. In a sense, healthy members of the population will behave as a non-risk takers or risk avoiders in these settings, investing effort in avoiding infection and perceived negative outcomes (illness and death), often through the use, for example, of enforced isolation of those infected or, more recently, vaccine development.

2.1 COVID-19 Pandemic

To see this type of summary pattern in relation to the recent COVID-19 pandemic, the Gaussian distribution can be seen fitting the $I(t)$ functions for a variety of countries for the period 2020 - 2022 [14]. The $I(t)$ functions defined across time periods and countries in the early stages of the pandemic do not show stable patterns. However, over time, as societal level risk avoidance behaviors, including vaccines, were applied in several ways, the $I(t)$ function stabilized into approximately Gaussian waves corresponding to the onset of the BF condition. Relevant data and images can be found at the Worldometer website and other data sites [14].

These patterns can also be examined for different time periods, corresponding to repeated waves of the epidemic, reflecting perhaps new strains of the COVID-19 infection and other factors. Note that with overlapping infection waves, the resulting overall Gaussian distribution, a mixture of Gaussian distributions, need not be symmetric.

2.2 Measles Epidemics

Measles epidemics reflect the same basic BF patterns in the past [15], but are currently less reflective of epidemic related risk avoidance behavior. Measles is much more common than COVID-19 and has a long history. Currently, measles is controlled through ongoing vaccination programs [16]. Such ongoing immunizations allow for greater control of the infection and this is shown in the resulting stable distribution for $I(t)$ which, in many countries, is close to a uniform distribution for long periods of time with only few serious outbreaks and epidemic-like behavior. See for example [17].

3 Probability Distributions

Assuming $I(t)$ can be modeled directly using a probability distribution with parameter θ, the BF condition can be examined:

$$\frac{\frac{I(t+3;\theta)}{I(t+2;\theta)}}{\frac{I(t+1;\theta)}{I(t;\theta)}} = K \tag{3}$$

Typically the distribution $I(t;\theta)$ is discrete, for example monthly counts, but can be approximated using continuous distributions.

3.1 Exponential Family Distributions

Many of the distributions related to the modeling of epidemics, such as the binomial, geometric, gamma, Poisson and Normal distributions are members of a statistical family of distributions, the exponential family of distributions. These are useful as application of standard large sample results often apply directly to such distributions [18].

These distributions can all be expressed in exponential family canonical form for a single parameter θ:

$$f(t;\theta) = h(t)\exp\{c(\theta) + \theta s(t)\}$$

where $h(t)$ is a norming constant, $c(\theta)$ a function of θ and $s(t)$ sufficient statistic.

The Central Limit Theorem (CLT) applies to standardized versions of the Poisson, Binomial and other useful distributions and allows for Gaussian approximation, typically written in relation to the standardized maximum likelihood estimator $\widehat{\theta}(t)$ for the parameter θ:

$$T = \frac{\widehat{\theta}(t) - \theta}{\sigma_{\widehat{\theta}(t)}} \sim N(0,1) \tag{4}$$

which amounts to writing the large sample density function of T in the form:

$$I(t) = \exp(-At^2 + Bt + c) \tag{5}$$

The large-sample arguments are applied here to a stochastic process, often monthly counts, and can reflect various assumptions. See for example Feller [19].

4 Some Examples

4.1 Uniform Distribution

The uniform distribution defined across a specific time period, $I(t) = k$, $t \in [t_0, t_1]$ trivially satisfies the BF condition. An incidence function $I(t)$ having a uniform distribution is typically not used to describe pandemic situations. It often reflects settings where the infection in question is under control.

4.2 Exponential Distribution

Consider the exponential distribution which can be written:

$$I(t) = e^{-t} \tag{6}$$

where $t > 0$. The BF condition here is given by:

$$\frac{\frac{I(t+3;\theta)}{I(t+2;\theta)}}{\frac{I(t+1;\theta)}{I(t;\theta)}} = K \tag{7}$$

and basic algebra shows this to be true.

A more flexible format is:

$$I(t;\theta) = e^{-\theta t} \tag{8}$$

and this also satisfies the BF condition.

4.3 Poisson Distribution

The Poisson distribution is commonly used to model case counts that occur within a given time period. The Poisson postulates can be used to guide its application [19], though the possibly correlated nature of pandemic counts across subsequent time periods may limit its usefulness. Formally $\theta > 0$, $I(0) = 0$ and the number of events in any interval of length t is a Poisson random variable with mean θt where θ is a constant.

Here this is used as a density formula to be fitted to the actual counts $I(t)$. In general, a Poisson counting process can be written:

$$I(t;\theta) = P(I(t) = n) = e^{-\theta t}\frac{(\theta t)^n}{n!} \tag{9}$$

The BF condition is given by:

$$\frac{\frac{I(t+3;\theta)}{I(t+2;\theta)}}{\frac{I(t+1;\theta)}{I(t;\theta)}} = K \tag{10}$$

and it follows that:

$$\frac{\frac{e^{-\theta(t+3)}\frac{(\theta(t+3))^n}{n!}}{e^{-\theta(t+2)}\frac{(\theta(t+2))^n}{n!}}}{\frac{e^{-\theta(t+1)}\frac{(\theta(t+1))^n}{n!}}{e^{-\theta(t)}\frac{(\theta t)^n}{n!}}} = \frac{\frac{((t+3))^n}{((t+2))^n}}{\frac{((t+1))^n}{(t)^n}} \tag{11}$$

which remains a function of t, not satisfying the BF condition.

Note that as $n \to \infty$, this distribution takes on an approximate Gaussian $N(0,1)$ density once standardized. This version then technically satisfies the BF condition. The central limit theorem here takes the form:

$$W_1(t) = \frac{\widehat{\lambda}(t) - \lambda}{\sqrt{\widehat{\lambda}(t)}} \sim N(0,1) \tag{12}$$

and the standardized incidence function $I(t)$ can be expressed as:

$$I(t) = \frac{1}{\sqrt{2\pi}} \exp(-1/2W_1(t)^2) \tag{13}$$

However, with large samples, as discussed below, a context supporting the onset of social level risk avoidance behavior should be present to claim the BF condition is related to the onset of a Gaussian distribution.

Other distributions such as the gamma and Weibull distributions, once standardized, can also be used to model incidence data in the form of monthly counts and, in large samples, will provide sampling distributions that are approximately Gaussian. This again does not necessarily imply the BF condition holds in such cases.

5 The IDEA Model

Interestingly, the BF condition and result, as it includes functions that are not probability densities, can be related to compartmental models such as the IDEA model for infection based on differential equations [6]. As the IDEA model can be algebraically related to the SIR model, the BF result can be extended as well to the SIR model [8].

To see this for the IDEA model, note that the model effectively assumes:

$$I(t) = \left[\frac{R_0}{(1+d)^t}\right]^t$$

Where d is a constant and R_0 reflects the probability of transmission of the infection at the individual level. It can be shown that the BF condition holds for this functional form of $I(t)$ giving:

$$\frac{\frac{I(t+3)}{I(t+2)}}{\frac{I(t+1)}{I(t)}} = K = \frac{1}{(1+d)^4}$$

As R_0 increases, the number of infected increases and the threat to healthy individuals increases. This leads to behaviour modification and government interventions that look to limit risk and restrain the infection from passing to healthy individuals. The number of sick individuals is directly proportional to the risk occuring and the onset of behavior modifications. Note that while the R_0 value is conceptually useful to consider when modeling the spread of an infection at the individual level, it does not formally appear in the maximization here for the version of the IDEA model given.

The $I(t)$ function, assuming the IDEA model holds, can be written:

$$\begin{aligned} I(t) &= \left[\frac{R_0}{(1+d)^t}\right]^t && (14) \\ &= \exp[t\log(R_0) - t^2\log(1+d)] && (15) \end{aligned}$$

giving the BF result $I(t) = \exp(-At^2 + Bt + c)$ with $A = \log(1+d)$, $B = \log(R_0)$, $C = 0$.

6 Interpretation in Relation to Societal Relative Risk

The general form of the BF restriction is of interest. Practically, it is a function of time and the incidence of infection $I(t)$, which can be viewed as a type of relative risk for those

not yet infected. This risk will modify over time as the pandemic grows, modulating and weakening as the pandemic fades and this is reflected in the Gaussian-type shape of $I(t)$ satisfying the BF condition.

In this way the BF condition can be interpreted as defining a type of global social risk function, where the number of infected cases, usually counted in a specific spatial region, represents increasing social risk in the context of a healthy, risk adverse general population.

While the BF model and restriction does not openly model the immuno-local and individual level infection dynamics of the nation or community in question, leaving out for example an R_0 value reflecting the speed of infection spread at the individual-to-individual level, it can be viewed as implicitly reflecting enforced behavioral changes in overall community behaviour which reflect those underlying infection dynamics and the social risk they represent.

7 Generalized Brownlee-Farr Condition

The BF condition is expressed in terms of a stochastic process $I(t)$ where differences are often represented by differences between lagged values of the sequence.

$$BF(t) = \frac{\frac{I(t+3)}{I(t+2)}}{\frac{I(t+1)}{I(t)}} = K$$

In this setting, the first-order derivative (essentially a time related relative difference or derivative: $\Delta I/\Delta t$) can be written:

$$\frac{I(t+1) - I(t)}{(t+1) - (t)} = [I(t+1) - I(t)] \tag{16}$$

and related second-order changes:

$$\frac{I(t+3)-I(t+2)}{(t+3)-(t+2)} = [I(t+3)-I(t+2)] \tag{17}$$

where the second-order difference in the BF condition was chosen to not overlap time points.

The BF condition itself sets a ratio of such second-order to first-order percent changes on a logarithmic scale equal to a constant. Taking a log transformation of both sides of the BF condition gives:

$$BF(t) = [\log I\,(t+3) - \log I(t+2)] - [\log I(t+1) - \log I(t)] = \log(K) = K' \tag{18}$$

This is the difference of first and second order derivatives on a $\log I(t)$ scale. When this difference is constant across all values of t, then $I(t) = \exp(-At^2 + Bt + c)$ holds and a Gaussian distribution is obtained, implying the epidemic is, in this sense, under control since as $t \to \infty$, $I(t) \to 0$.

In comparison, many large sample Gaussian statistical arguments, for example using the statistical score function [18], are based on a ratio, not a difference, of the first and second order derivatives of $\log(I(t))$ This also underlies the Cramer-Rao bound or Fisher information bound on the sampling variation of maximum likelihood estimates.

Using continuous derivatives in relation to time, and assuming the numerator serves as an approximate second order change or acceleration term for $\log I(t)$, and the denominator an approximate first order change or derivative of $\log I(t)$, a continuous version of the BF condition can be expressed generally as:

$$BF(t) = \frac{d^2}{dt^2}\log I(t) - \frac{d}{dt}\log I(t) = K \tag{19}$$

8 A Related Social Risk Function: Pratt-Arrow Relative Risk

The BF mathematical condition above is similar to a more generally defined risk avoidance measure. The basic Pratt-Arrow measure of absolute relative risk ARA [10] can be defined as:

$$ARA = -\frac{d}{du}\log[\frac{d}{du}I(u)] = -\frac{\frac{d^2}{du^2}I(u)}{\frac{d}{du}I(u)} \tag{20}$$

where $I(u)$ is a function representing utility and $I'(u) > 0$ for all u. While it is formally defined locally and for individual behavior, if it is similar for all individuals, it can be viewed as a societal-level measure. Individuals in an epidemic setting are subject to similar government restrictions, giving the ARA measure both individual and general social relevance. Note that the ARA measure, as a measure of utility, was not defined in order to meet the restrictions of obtaining a Gaussian form for the underlying risk related function $I(u)$.

As time affects the level of infection and thus relative risk and risk avoidance behavior for those who are not infected, this might also be more directly expressed as a function of infected counts through time, which are the primary sources of risk:

$$ARA(t) = -\frac{d}{dt}[\log\frac{d}{dt}I(t)] = -\frac{\frac{d^2}{dt^2}I(t)}{\frac{d}{dt}I(t)} \tag{21}$$

If the ARA measure of relative risk is constant, a condition similar to the BF condition is obtained:

$$ARA(t) = -\frac{\frac{d^2}{dt^2}I(t)}{\frac{d}{dt}I(t)} = K \tag{22}$$

with risk avoidance behavior occuring where $K < 1$. Such a constant value for the

ARA measure implies a stabilization in terms of risk avoidance behavior. Again, even though the ARA measure is defined at the individual level, the unique setting of shared pandemic restrictions for all individuals extends this restriction to society-level application.

9 Entropy Considerations

The concept of entropy [20] can sometimes provide a useful perspective when examining the shape of probability densities. Where the form of $I(t)$ resulting from the BF condition is given by a formal probability distribution, the entropy of that distribution can provide a summary of the overall shape of the $I(t)$ probability density. As well, the idea of maximum entropy may be relevant in expressing an optimal form for $I(t)$ under the BF restriction. Note that if $I(t)$ has a parametric form, the observed value of the maxiumum likelihood estimate $\widehat{\theta}(t)$ should be used to fit the distribution to the data: $I(t) = I(t; \widehat{\theta}(t))$.

The observed entropy of $I(t)$ in general is defined here as:

$$Ent = -\sum_t I(t) \log I(t) \tag{23}$$

and if the BF condition holds, then:

$$I(t) = \exp(-At^2 + Bt + c) \tag{24}$$

and

$$Ent = -\sum_t [\exp(-At^2 + Bt + c)](-At^2 + Bt + c) \tag{25}$$

Note that in the specific case of the normal density, $N(\mu, \sigma^2)$ entropy is given by:

$$\frac{1}{2} \log(2\pi\sigma^2) + \frac{1}{2} \tag{26}$$

and can be estimated by:

$$\frac{1}{2}\log(2\pi s^2) + \frac{1}{2} \tag{27}$$

where s^2 is the sample variance of the observed data.

In the case of the exponential distribution, the distribution and entropy are given by:

$$\begin{aligned} f(t;\lambda;t) &= \lambda e^{-\lambda t}, t \geq 0 && (28)\\ E[T] &= 1/\lambda; Var[T] = 1/\lambda^2 && (29)\\ Ent &= 1 - \ln\lambda && (30) \end{aligned}$$

With the estimated entropy given by:

$$1 - \ln \overline{x} \tag{31}$$

Interestingly, for distributions having unknown mean and variance parameters, the Normal distribution represents a distribution having maximum entropy [21]. As well, for distributions having only an unknown mean parameter, the exponential distribution represents a maximum entropy distribution. As well, for simple discrete distributions, the uniform distribution represents a maximum entropy distribution. This shows a link between the onset of the BF restriction, a stable level of social risk avoidance, and the relative incidence density function $I(t)$ having maximum entropy.

Note that the Gibbs inequality [22] applies here using the log sum inequality. If $J(t)$ is another probability distribution describing the incidence of infection, or the $I(t)$ distribution for an earlier wave of the pandemic, then:

$$-\sum I(t)\log\left(\frac{I(t)}{J(t)}\right) \geq 0 \tag{32}$$

This is typically defined as the cross-entropy [23] of $I(t)$ in relation to another incidence density $J(t)$. This can also be used to assess the expected entropy of $I(t)$ under the BF condition on relation to the observed incidence distribution $J(t)$. The cross entropy will be closer to zero as $I(t)$ and $J(t)$ become more similar.

Since the BF condition can be interpreted in relation to relative risk, the entropy of the $I(t)$ distribution can be interpreted in relation to social risk avoidance behavior, with the epidemic becoming more controlled as maximum entropy of $I(t)$ is achieved. Social restrictions alter behavioral patterns of all individuals, limiting the spread of the infection and placing a limit on the possible entropy of $I(t)$.

10 Social Utility and Entropy

The issue of optimizing behavior in the presence of widespread illness has been examined in the literature [24, 25]. These papers tend to modify the SIR or SEIR model to account for additional considerations related to maximizing utility at the individual level subject to illness related adjustments and additional restrictions to the SIR model.

Social utility functions are typically defined as weighted averages of individual actions, where these can be given a ranking across a set of possible actions [26]. Here the risk averse actions of individuals living through a pandemic under similar governmental restrictions can be defined as expected risk-averse responses to the level of infection in the society and measured by $I(t)$. Whatever the actual values of the social utility function, measured in relation to ordinal rankings of possible actions and values, these will be a function of $I(t)$ in the context of the epidemic, the observed relative incidence of infection in the society.

Where the BF condition leads to a normal distribution, it can be interpreted as maximizing the entropy of the $I(t)$ density. If the empirical probabilities for the histogram of observed $I(t)$ values are available, and the context is related to social risk, then the entropy

of the observed $I(t)$ density can be obtained and may be compared to the entropy expected from the Gaussian distribution.

An approximate measure of social control of the epidemic, within a given time frame, can be defined as Δ_c, where:

$$\Delta_c = Ent - \left(\frac{1}{2}\log(2\pi s^2) + \frac{1}{2}\right) \tag{33}$$

where Ent is the observed entropy of $I(t)$. Values close to zero imply the BF condition is approximately true and the epidemic may be viewed as being under control within the time frame of interest. In settings where multiple Gaussian waves describe $I(t)$, the entropy value can appropriately adjusted, defined as a weighted average.

11 Hazard Rate and Risk

The concept of risk, typically in the form of the onset of a fatal event, has been studied for many years in relation to insurance life tables and the related topic of survival analysis [27, 28]. A mathematical structure similar to the BF and Pratt-Arrow constructs can be applied in relation to risk defined as the instantaneous infection rate, given non-infection to time t.

The hazard rate (the risk of infection at any point in time, given non-infection up to that point) can be defined in terms of the survivor function $S(t) = 1 - F(t)$, where $F(t)$ is the cumulative probability density for $I(t)$. Letting T denote the time of infection, the hazard rate can be defined in relation to the onset of infection:

$$h(t) = P(T = t|X > t) = \frac{P(T = t)}{P(T > t)} = \frac{I(t)}{1 - F(t)} = -\frac{d}{dt}[\log(S(t))] \tag{34}$$

and a condition fairly similar to the BF condition can be defined using the instantaneous

infection rate, reflecting a constant, controlled hazard:

$$-\frac{d}{dt}[\log(S(t))] = K \tag{35}$$

For example, with the exponential distribution, we have $I(t) = \lambda e^{-\lambda t}$ and $S(t) = e^{-\lambda t}$. Its hazard rate $h(t)$ can be shown to equal λ, a constant. This implies an assumption of constant instantaneous risk or hazard if we assume the exponential distribution.

Note that in a mathematically similar context the BF condition uses the instantaneous first and second order derivatives of the incidence function $\log(I(t))$, the hazard rate uses the cumulative probability density as a baseline and its first derivative relative to this baseline.

For example, in the case of the Weibull distribution for infection, $f(t) = \lambda\gamma t^{\gamma-1}\exp(-\lambda t^{\gamma})$, its survivor function is given by $S(t) = \exp(-\lambda t^{\gamma})$. It follows that its hazard function is given by $h(t) = -\frac{d}{dt}\{\log(S(t))\} = \lambda\gamma t^{\gamma-1}$. If $\gamma > 1$, the hazard or risk of infection is increasing over time and if $\gamma < 1$, it is decreasing. If $\gamma = 1$ the hazard is equal to λ, showing the exponential distribution to be a specific case of the Weibull distribution. This approach, along with Kaplan-Meier curves, can be used to compare times to infection across demographic groups.

12 Multiple Epidemic Waves

A spatial assessment of incidence counts can be assessed with the overall incidence being a mixture of locally defined $I_i(t)$. If the BF condition for a given time period holds locally, for each spatial region or country considered, when social restrictions have imposed the condition, then the overall incidence function can be modeled as the weighted average of local distributions:

$$I(t)_{overall} = \sum_i w_i(t) I_i(t) = \sum_i w_i(t) \exp(-A_i t^2 + B_i t + c_i) \tag{36}$$

Practically, as a pandemic proceeds, the wavelength and frequency of the infection-waves decreases, often with telescoping weights. There is a societal learning effect as more waves are encountered. The imposition of social restrictions and resulting risk avoidance will have greater effects on later pandemic waves which will relatively weaken, subject to the caveat that the infectious agent does not genetically mutate. These will potentially differ by area allowing for different weighting schemes $w_i(t)$ and may overlap. Such waves have a long history in the epidemiologic and related gene wave literature [29].

For the purpose of calculations, often $I(t)$ needs to be weighted by the total number of cases defined within a given time period, for example a year, and for a respiratory illness this often overlaps with defined waves of occuring cases. The COVID pandemic examined overall and within specific countries often shows this pattern [14].

13 Mathematical Similarities

It is interesting to examine epidemiologic phenomena as risk events at the aggregate societal and governmental level, rather than focus on individual-level transmission dynamics as the onset of risk events often prompts common risk avoidance behavior in the entire unaffected human population.

The mathematics to describe the basic structure of risk and risk avoidance in very different risk-related settings and disciplines is often, as shown here, very similar. The BF condition and approach allows for the modeling of statistical properties that can be related to social utility based measures of risk avoidance, and also more standard measures of risk in the form of hazards, the hazard of becoming infected.

Consider the three basic representations of risk considered here:

$$BF(t) = \frac{d^2}{dt^2} \log I(t) - \frac{d}{dt} \log I(t) \tag{37}$$

and

$$ARA(t) = -\frac{d}{dt} \log[\frac{d}{dt} I(t)] \tag{38}$$

andd

$$h(t) = -\frac{d}{dt}[\log(S(t))] \tag{39}$$

The differences in these measures of relative relate to changes in scale and the use of differences rather than ratios. The hazard rate $h(t)$ measure uses changes in the cumulative probability density with $I(t)$ on a log scale. The $BF(t)$ measure uses the difference in second and first order derivatives with $I(t)$ on a log scale. The $ARA(t)$ measure uses a ratio of first and second order derivatives of $I(t)$ without a log scale.

Note that these definitions of risk and related risk avoidance conditions come from very different areas of research and application: epidemiology, economics, actuarial science. In each, the study and modeling of perceived risk using the properties of the $I(t)$ distribution has lead to similar mathematical structures of measurement and interpretation.

14 Discussion

Epidemics and pandemics are common in the history of societies. While they occur in real time, there is a period of rising incidence, sickness for the infected and rising peril for the non-infected. In such settings the increasing number of infected motivates risk avoidance behavior patterns on the part of society that remains healthy, often at the level of government restriction applying to all individuals. These patterns can be observed and

modeled.

The BF restriction can be viewed as holding as the result of societal risk avoidance behavior in the presence of pandemic waves as individuals attempt to maximize risk avoidance in a given time period. It is mathematically similar in structure to the Pratt-Arrow ARA utility based risk avoidance behavioral measure. The condition also supports viewing the shape of $I(t)$ in relation to entropy, with many of the probability distributions meeting the condition being interpretable as having maximum entropy.

Given the impact of an epidemic or pandemic on overall society, social risk and related utility functions are of interest. These will reflect the shape or entropy of the $I(t)$ distribution, which reflects the overall infection rate affecting all of society. The shape of this function and the degree to which it differs from the applicable BF distribution or function, for example the Gaussian distribution reflecting maximum entropy, assuming the BF condition applies, may provide a measure of the extent to which the pandemic is not yet controlled.

To better understand the social processes affecting the controlling an epidemic, it may be useful to examine the shape of the relative incidence distribution via entropy and formally model the onset of maximum entropy in the relative incidence function $I(t)$, utilizing variables related to the measuring of social risk avoidance, and the formal structure of social utility functions. This macro-level or aggregate approach allows for the application of statistical and probability related concepts to better understand how the various risk measures relate to each other. Applications and comparisons of these measures to various infectious diseases and other related data are the subject of ongoing and future work.

15 Conclusions

Farr's Law can be understood, at a macro-level, as reflecting risk averse social behavior in the setting of an epidemic where the distribution of the relative incidence of infection, $I(t)$, represents a general level of risk. This perspective allows for the resulting BF condition to be compared to other measures of risk and risk avoidance with notable similarities. As the shape of the $I(t)$ distribution, especially its first two derivatives, are important aspects, the use of entropy is helpful in measuring convergence to the BF condition.

16 References

1. Organization WH. Coronavirus disease 2019 (COVID-19) Situation Report—83. World Health Organization, 2020 April 12, 2020. Report No.: Contract No.: 83.

2. Anderson RM and May RM. Infectious Diseases of Humans: Dynamics and Control. Oxford University Press, Oxford, 1991.

3. Andersson H. and Britton T. Stochastic epidemic models and their statistical analysis, volume 151 of Lecture Notes in Statistics. Springer, New York, 2000.

4. Brownlee J. Historical Note on Farr's Theroy of The Epidemic. Br Med J. 1915; 2(2850):250–2. https://doi.org/10.1136/bmj.2.2850.250 PMID: 20767766.

5. Farr W. (1840). Appendix to the second annual report of the registrar-general. London: UK. General Register Office. Online Historical Population Reports. Available at http://histpop.org/ohpr/servlet/PageBrowser2?ResourceType=Registrar.

6. W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society of London Series A, 115:700–721, 1927.

7. Tuite AR and Fisman DN. The IDEA model: A single equation approach to the Ebola forecasting challenge. Epidemics. 2018 Mar;22:71-77. doi: 10.1016/j.epidem.2016.09.001. Epub 2016 Sep 28. PMID: 27717616.

8. Santillana M., Tuite A., Nasserie T., Fine P., Champredon D., Chindelevitch L., Dushoff J., Fisman D. Relatedness of the incidence decay with exponential adjustment (IDEA) model, "Farr's law" and SIR compartmental difference equation models. Infectious Disease Modeling 3 (2018) 1-12.

9. Lee SY, Lei B, and Mallick B (2020) Estimation of COVID-19 spread curves integrating global data and borrowing information. PLoS ONE 15(7): e0236860. https://doi.org/10.1371/journal.pone.0236860.

10. Pratt JW. Risk Aversion in the Small and in the Large. Econometrica, Vol. 32, No. 1/2 (Jan. - Apr., 1964), pp. 122-136.

11. Mikolai J., Dorey P., Keenan K. and Kulu H. Spatial patterns of COVID-19 and non-COVID-19 mortality across waves of infection in England, Wales and Scotland. Social Science and Medicine, Vol. 338, 2023, 116330, ISSN 0277-9536.

12. Bregnman DJLA. Farr's law applied to AIDS projections. JAMA. 1990; 263(11):1522–5. PMID: 2308183.

13. Darakjy S, B J, DiMaggio CJ, Li Guohua. Applying Farr's Law to project the drug overdose mortality epidemic in the United States. Inj Epidemiol. 2014; 1(31). https://doi.org/10.1186/s40621-014-0031-2 PMID: 27747664.

14. https://www.worldometers.info/coronavirus/.

15. Bartlett MS. Stochastic population models in ecology and epidemiology, volume 4

of Methuen's Monographs on Applied Probability and Statistics. Spottiswoode, Ballantyne, London, 1960.

16. Bester JC. Measles and Measles Vaccination: A Review. JAMA Pediatr. 2016;170(12):1209-1215. doi:10.1001/jamapediatrics.2016.1787.

17. IIDDA. The International Infectious Disease Data Archive, http://iidda.mcmaster.ca.

18. Casella G. and Berger RL (2002), Statistical Inference, 2nd ed. Wadsworth Group, CA.

19. Feller W. (1968). An Introduction to Probability Theory and Its Applications. John Wiley and Sons, New York.

20. Shannon C. A Mathematical Theory of Communication. The Bell System Technical Journal, Vol. 27, pp. 379–423, 623–656, July, October, 1948.

21. Buck B. and Macauley VA. Maximum Entropy in Action. Clarendon Press, Oxford, 1991.

22. Shannon CE and Weaver W. The Mathematical Theory of Communication. University of Illinois Press, Urbana IL, 1949.

23. Cover TM. and Thomas JA. Elements of Information Theory, 2nd ed., Wiley.

24. Morsky B., Magpantayb F., Dayb T., Akçayc E. The impact of threshold decision mechanisms of collective behavior on disease spread. PNAS 2023 Vol. 120 No. 19 e2221479120.

25. Funk S., Salathe M. and Jansen VAA. Modelling the influence of human behaviour on the spread of infectious diseases: a review. J. R. Soc. Interface (2010) 7, 1247–1256 doi:10.1098/rsif.2010.0142 Published online 26 May 2010.

26. Bossert W. and Weymark JA. (2004). Utility in Social Choice, Chapter 20 in the Handbook of Social Utility Theory, Barbera et al (eds.), Springer Science-Business Media, New York, p. 1099-1177.

27. Collett D. (2023). Modelling Survival Data in Medical Research, 4th ed. Chapman and Hall, New York.

28. Brimacombe M. (2019). Likelihood Methods in Biology and Ecology: A Modern Approach to Statistics. CRC Press.

29. Tanaka H., Stone HA. and Nelson DR. Spatial gene drives and pushed genetic waves. PNAS July 25, 2017 114 (32) 8452-8457 https://doi.org/10.1073/pnas.1705868114.