A General-Purpose Modeling Framework for Analyzing Infectious Disease Outbreaks

Eva K Lee*[,1], Yifan Liu[1]

Mustapha Debboun[2] and Umair A. Shah[2,3]

[1]NSF-Whitaker Center for Operations Research in Medicine and HealthCare, Georgia Institute of Technology, Atlanta, GA.

[2]Harris County Public Health, Houston, Texas

[3]Department of Health, State of Washington

Abstract

Background: Mathematical models with system dynamics and interactions are powerful tools for analyzing disease outbreaks and supporting real-time decision-making. However, building epidemiology models from scratch during an emergency is not efficient and will cause potential delays in assessing and mitigating risks. A general-purpose framework is needed for real-time epidemiology modeling and analysis.

Methods: We propose a general-purpose modeling framework for contact-based and vector-borne infectious diseases. The design schema categorizes each stage/compartment based on two different criteria: (1) passive or active, and (2) vulnerable, contagious, or unaffected, and establishes relationships between stages. Mathematical expressions for the contact-based and the vector-borne disease models are derived, each of which features a compact matrix formulation to facilitate efficient computation.

Results: All classical epidemiology models can be derived using this modeling framework. We demonstrate the derivation of the susceptible-exposed-infectious-recovered (SEIR) contact-based and SEIR-SID vector-based models; and establish the new Susceptible, Exposed, initial infectious Period, Asymptomatically infectious, systematically Infectious, and Recover (SEPAIR), and SEPAIHQR models that are suitable for modeling the covid-19 pandemic. The associated basic reproduction number of each model is derived using the next-generation matrix concept. The modeling framework describes the dynamics of the system with a short and compact formulation and can easily be generalized and customized based on the biological properties of different infectious diseases. It allows for real-time prediction and analysis. The resulting basic reproduction numbers reveal intimate interplay of dwell times of various contagious stages and host-vectors and their contribution to disease spread. These offer a simple yet elegant framework for decision-makers to objectively and rapidly contrast different interventions and understand their effects in disease mitigation and tradeoffs. **Conclusion:** The proposed modeling framework provides an elegant meta-model for epidemiology and enables decision makers to rapidly build, evaluate, and implement disease models according to the type of outbreak without diving into the interactions among multiple stages and explicitly deriving the ordinary differential equations. This can save time and potentially save lives during pandemic emergencies.

Keywords (3-10 keywords): General purpose disease modeling; epidemiology models; contact-based disease; vector-borne disease; compartmental models; basic reproduction number

Key Messages:

- Building epidemiology models from scratch during an emergency is not efficient and will cause potential delays in assessing and mitigating risks.
- A general-purpose meta-model for infectious diseases will enable decision makers to rapidly build, evaluate, implement and customize disease models according to the type of outbreak.
- The modeling framework presented herein describes the dynamics of a system with a short and compact formulation to facilitate computation.
- All classical epidemiology models can be derived using this modeling framework.
- This modeling framework enables real-time prediction and analysis

Introduction

Infectious diseases continue to be one of the major causes of mortality in many countries in the 21st century. They are the third leading cause of death in the United States and the leading cause worldwide (1). In addition, infectious agents such as bacteria and viruses continue to evolve and pose new threats to human beings (2). The 2014 Ebola virus outbreak in West Africa, the 2016 Zika virus outbreak in South America and the Caribbean, and the current SARS-CoV-2 virus that causes the raging covid-19 global pandemic underscore that combating emerging infectious diseases remains one of the most important tasks in public health.

Some of the major factors driving emerging infectious and vec tor-borne diseases include human population expansion, increased human travel and migrations, urbanization, climate change, lack of vaccines, and the resurgence of the vectors increase the risk of continued global introductions and local outbreaks of vector-borne diseases.

Mathematical models have long played an important role in understanding the underlying mechanisms of the spread and control of infectious diseases. Utilizing available biological and environmental information, and transforming it into knowledge of the disease, these models provide strategies and guidelines for mitigation and containment to prevent it from becoming a global pandemic. Mathematical epidemiology models have been studied extensively for many years. Coupled with computer simulation, they are powerful tools to provide valuable insights into the disease dynamics, understand the transmission characteristics, test hypothesis, assess and evaluate containment strategies. Bernoulli (3) formulated the earliest smallpox model in the 18th century. In 1906, Hamer developed a discrete time model to understand the recurrence of measles in which the number of new infections was assumed to depend on the product of the density of susceptible and infectious populations (4). This idea paved the foundation of the subsequent compartmental models. In 1911, Ross developed a system of differential equations using a host-vector structure for the control of malaria (5). Other deterministic models were also developed for multiple purposes in the early 20th century (6), specifically Kermack and McKendrick (7-9) established the theory of the SIR (susceptible-infectious-recovered) models and other corresponding compartmental models in a series of published articles in the 1920s and 1930s.

Many mathematical models and extensions for pandemics were developed based on the framework of Kermack-McKendrick model. Li and Zou studied the compartmental model in which the infectious disease has a fixed latent period, and formulated the SIR structure for population living in two patches (10). Pathak et al. (11) replaced the constant infection rate with an asymptotically homogeneous transmission function and derived the stability condition of the model. Xu (12)

developed a diffusive Kermack--McKendrick epidemic model with a latent period and determined conditions of the existence of traveling waves solutions of the system. Inaba (13) extended the compartmental model to recognize individual heterogeneity, and expanded the definition of compartments to genetic, psychological, or behavioral characteristics of population. Disraelly et al. (14) extended the methodology of Human Response Injury Profile (HRIP) which uses time-based progression to determine casualty and fatality estimations from infection. Based on the compartmental model, they introduced an injury profile sub-model based on severity of the symptoms to describe the progression of illness at a detailed granularity. Lee et al. (23) introduced asymptomatic infectious within the model to analyze effect infection undercount, and triage errors within mass vaccination facilities on intra-facility disease spread. Daley and Gani (15), Hethcote (16), and Breda et al. (17) provided complete reviews of the compartmental epidemiology models that emerged based on the work of Kermack and McKendrick.

In this study, we introduce a general-purpose modeling framework for the spread of infectious diseases. We first categorize the disease stages according to their roles and properties in the spread of infectious diseases; then we define the essential variables for the model based on the categorization of stages. Utilizing this framework, we derive the mathematical expression of the general-purpose models for contact-based and vector-borne diseases. The resulting modeling framework enables derivation of any existing epidemiology models, allows generation and investigation of new ones, and we demonstrate its application on three disease models.

Methods and Designs

In this section, we introduce and derive the general-purpose framework for modeling contact-based diseases and vector-borne diseases. The schema first categorizes the disease stages according to their roles and properties in the spread of infectious diseases. Based on these categorizations, essential variables are defined. Mathematical expression for the general-purpose models for contact-based and vector-borne diseases are then derived.

Categorization of Disease Stages

The general-purpose modeling framework for infectious diseases is developed based on the compartmental models. Compartments, or disease stages, are the statuses that individuals fall into during the spread of infectious diseases. We assume that the human population is homogeneous. Let Φ denote the collection of all possible stages. For example, for SIR model, $\Phi = \{S, I, R\}$. Then any stage in Φ can be categorized in two different ways:

- *Passive/Active stages*. Individuals in a passive stage will not change their statuses spontaneously. For example, the susceptible stage is passive, as susceptible individuals remain susceptible unless they contact the infectious population and become infected. Let Φ_P denote the collection of passive stages. On the contrary, individuals in an active stage will change their statuses spontaneously. An example of active stages is infectious individuals who will either recover or decease given sufficiently long time. Letting Φ_A denote the collection of active stages, a stage will be either passive or active. Therefore, $\Phi_P \cup \Phi_A = \Phi$. This method of categorization determines if a mean dwelling time is well-defined for a stage.

- *Vulnerable/Contagious/Unaffected stages*. Individuals in a vulnerable stage can be infected through contacting the infectious ones. Let Φ_V denote the collection of all vulnerable stages where individuals in a contagious stage will have the ability to infect vulnerable entities. We use Φ_C to denote all the contagious stages. The third category is unaffected stages, in which individuals are neither vulnerable nor contagious, and they are not in the system of infection. We denote them as Φ_U. For example, in the SEIR model, after the initial infection, entities will enter the exposed stage. They are not vulnerable, as they have already been infected; but they are not infectious yet since the density of viruses or bacteria has not reached a level to infect others. Another example of the unaffected stage is the recovered stage in an SIR model with immunity assumed. Every natural stage without intervention will fall into one of these categories, thus $\Phi_V \cup \Phi_C \cup \Phi_U = \Phi$.

Using this modeling schema, each disease stage can be categorized in two different ways, which capture the major characteristics of a disease stage. Figure 1 shows the transition diagram of the 4-stage susceptible-exposed-infectious-recovered (SEIR) model and the categorization of each stage.

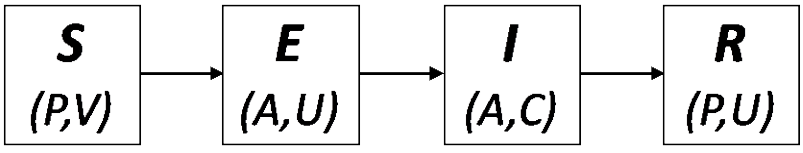

Figure 1: Demonstration of stage categorization of the SEIR model.

Definition of Model Components

Contact-based diseases are transmitted through physical or indirect contacts between humans. The main interest of compartmental models is to understand how human population associated with

each disease stage changes with respect to time t. Let $\phi(t)$ denote the number of individuals in the system at time t, for all $\phi \in \Phi$, and vector $\boldsymbol{y} = \langle \phi(t) \rangle|_{\phi \in \Phi}$ represent the number of entities in each compartment at time t. Similarly, the first-order derivative of the population in each stage in terms of time t can be denoted as $\boldsymbol{y}' = \langle \phi'(t) \rangle|_{\phi \in \Phi}$.

For each active stage, there is a well-defined average transition. This transition rate represents how long an individual will stay in an active stage before transiting to another stage. Therefore, for all $\phi \in \Phi_A$, let μ_ϕ denote the mean transition rate for stage ϕ; for $\phi \in \Phi_P$, define $\mu_\phi = 0$ for completeness. Let $\boldsymbol{\mu} = \langle \mu_\phi \rangle|_{\phi \in \Phi}$ be the vector form of the mean transition rate for each stage.

To understand the destination of the disease transition, define a disease transition matrix $\boldsymbol{D} = \langle d_{ab} \rangle|_{a \in \Phi, b \in \Phi}$ such that d_{ab} represents the disease transition probability from stage b to stage a. The column sum of matrix $\boldsymbol{D}$ is either 0 or 1 where the column sum is 0 for absorbing stages, and the column sum is 1 for all other stages. (An absorbing state is one that, once entered, cannot be left. For example, recovered (assuming immunity) and deceased). Since the transition structure of an infectious disease is determined by its own biological property, the disease transition matrix $\boldsymbol{D}$ can be viewed as a constant parameter. The initial infection is also counted as disease transition.

Let β_ϕ denote the effective baseline infection rate or contact rate adjusted by the total free population in the system for all vulnerable stages $\phi \in \Phi_V$. Again, define $\beta_\phi = 0$ for stages $\phi \in \Phi_C \cup \Phi_U$ for completeness. Its vector form is $\boldsymbol{\beta} = \langle \beta_\phi \rangle|_{\phi \in \Phi}$. Unlike the disease transition matrix $\boldsymbol{D}$, the value of vector $\boldsymbol{\beta}$ may change with respect to time t due to the change in the distribution of

the population. An intuitive example is that people normally do not have contacts with the deceased population, thus it needs to be excluded when calculating the population-adjusted contact rate $\boldsymbol{\beta}$.

In the categorization section, we define the relationship between vulnerable and contagious stages: individuals in contagious stages can infect those in vulnerable stages via direct or indirect contacts, depending on the biological property of the infectious disease. We now represent this relationship using mathematical terms. For simplicity, we define the function $S(\phi)$: $\Phi \to 2^{\Phi}$ such that $S(\phi)$ finds the set of successor stages of ϕ in the disease transition diagram. Since the transition structure can be fully characterized by the disease transition matrix $\boldsymbol{D}$, the function $S(\phi)$ can be expressed equivalently as

$$S(\phi) = \{\psi \in \Phi: d_{\psi\phi} > 0\}$$

Similarly, we can define the function $P(\phi)$: $\Phi \to 2^{\Phi}$ such that $P(\phi)$ finds the set of predecessor stages of ϕ. Using the definition of matrix $\boldsymbol{D}$, the function $P(\phi)$: $\Phi \to 2^{\Phi}$ can be expressed as

$$P(\phi) = \{\psi \in \Phi: d_{\phi\psi} > 0\}$$

With function $S(\phi)$ properly defined, we can now derive the disease contagious matrix $\boldsymbol{C} = \langle c_{ab} \rangle|_{a\in\Phi, b\in\Phi}$ to represent the relationship between contagious and vulnerable stages. In particular, $c_{ab} = -1$ if and only if $a \in \Phi_V$ and $b \in \Phi_C$, and $c_{kb} = 1$ for all $k \in S(a)$ if and only if $a \in \Phi_V$ and $b \in \Phi_C$.

Table 1 summarizes all the model components introduced and discussed thus far. They will be used to establish the general-purpose modeling framework in the following sections.

Table 1: Definition of model components in the generic modeling framework

Symbol	Definition
$\boldsymbol{P}$, $\boldsymbol{A}$	Passive/Active stages with no overlap, $\Phi_P \cup \Phi_A = \Phi$
$\boldsymbol{V}$, $\boldsymbol{C}$, $\boldsymbol{U}$	Vulnerable/Contagious/Unaffected stages with no overlap, $\Phi_V \cup \Phi_C \cup \Phi_U = \Phi$.
$\boldsymbol{y} = \langle \phi(t) \rangle\|_{\phi \in \Phi}$	Number of entities in each compartment at time t.
$\boldsymbol{y}' = \langle \phi'(t) \rangle\|_{\phi \in \Phi}$	The derivative of populations in each stage with respect to time t.
$\boldsymbol{\mu} = \langle \mu_\phi \rangle\|_{\phi \in \Phi}$	Transition rate of active stages. Define $\mu_\phi = 0$ for all $\phi \in \Phi_P$ for completeness.
$\boldsymbol{\beta} = \langle \beta_\phi \rangle\|_{\phi \in \Phi}$	Baseline infection rate for vulnerable stages, adjusted by total population. Define $\beta_\phi = 0$ for all $\phi \in \Phi_C \cup \Phi_U$ for completeness.
$\boldsymbol{C} = \langle c_{ab} \rangle\|_{a \in \Phi, b \in \Phi}$	Disease contagious matrix: $c_{ab} = -1$ if and only if $a \in \Phi_V$ and $b \in \Phi_C$, and $c_{kb} = 1$ for all $k \in S(a)$ if and only if $a \in \Phi_V$ and $b \in \Phi_C$.
$\boldsymbol{D} = \langle d_{ab} \rangle\|_{a \in \Phi, b \in \Phi}$	Disease transition matrix: d_{ab} is the transition probability from stage b to stage a.

Modeling Framework for Contact-Based Diseases

With the model components properly defined in Table 1, we can derive the general-purpose modeling framework for the spread of infectious diseases. In particular, we would like to derive the expression for $\boldsymbol{y}'$ using these modeling components, which gives us a convenient way to compute the population changes in each compartment within a time unit.

The changes in population can be decomposed into two parts: 1) the changes due to natural progression of infectious diseases, $\boldsymbol{y}'_{\boldsymbol{dev}}$, and 2) the changes due to new infections, $\boldsymbol{y}'_{\boldsymbol{inf}}$. To determine $\boldsymbol{y}'_{\boldsymbol{dev}}$, the change of population in stage ϕ due to the natural development can be written as $-\mu_\phi \phi(t)$ and rewritten more compactly in matrix form: $\boldsymbol{y}'_{\boldsymbol{1}} = -diag(\boldsymbol{\mu})\boldsymbol{y}$. This only considers the populations flowing out of the active stages. Since disease transition matrix $\boldsymbol{D}$ captures the

transition structure, it can be used to describe the destinations of these populations, i.e., the change of population in each stage due to the inflow caused by natural progression of active stage can be written as $\boldsymbol{y_2'} = \boldsymbol{D}diag(\boldsymbol{\mu})\boldsymbol{y}$. Combining the population changes in $\boldsymbol{y_1'}$ and $\boldsymbol{y_2'}$, the change of populations in each stage due to the natural development of infectious disease can be written as:

$$\boldsymbol{y'_{dev}} = (\boldsymbol{D} - \boldsymbol{I})diag(\boldsymbol{\mu})\boldsymbol{y}$$

To determine $\boldsymbol{y'_{inf}}$, the cause of changes in population is new infections due to direct or indirect contacts between vulnerable and contagious individuals. In this case, the generation of new infections is determined by the population of both vulnerable and contagious stages. Consider a vulnerable individual in stage $\phi \in \Phi_V$ who makes b_ϕ contacts with others on average in a unit time. Assume the total free population, i.e., the population that can be reached by this individual at time t is $N(t)$. Among these $N(t)$ individuals, $C(t)$ of which are contagious. Therefore, the average size of contagious population contacted by the considered individual in a unit time is $b_\phi C(t)/N(t)$. Since the total size of population in stage ϕ is $\phi(t)$, the total new infections introduced to the system in a unit time from vulnerable stage ϕ is $b_\phi \phi(t) C(t)/N(t)$ and the baseline infection rate $\boldsymbol{\beta}$ defined in Table 1 is already adjusted by the total free population $N(t)$, the size of new infections from vulnerable stage ϕ can be written as $\beta_\phi \phi(t) C(t)$.

Therefore, for a vulnerable stage $\phi \in \Phi_V$, the change in population due to new infections is $-\beta_\phi \phi(t) \sum_{\psi \in \Phi_C} \psi(t)$. The infection structure between vulnerable and contagious stages has already been captured by disease contagious matrix $\boldsymbol{C}$, thus, this change of population can be rewritten more compactly in matrix form: $\boldsymbol{y_3'} = diag(diag(\boldsymbol{\beta})\boldsymbol{y})\boldsymbol{C}\boldsymbol{y}$. Similarly, this equation only considers the outflow of population from vulnerable stages due to new infections. To capture the

destinations of these outflows, disease transition matrix $\boldsymbol{D}$ is used. The inflow in each stage due to new infection is then given as

$$\boldsymbol{y}_4' = diag(\boldsymbol{D}diag(\boldsymbol{\beta})\boldsymbol{y})\boldsymbol{C}\boldsymbol{y}$$

The expression for $\boldsymbol{y}_4'$ can also be deduced using the expressions for the predecessors and successors discussed earlier: for a vulnerable stage v that can be infected by individuals in stage u, $c_{\psi u} = 1$ for all $\psi \in S(v)$. Then, it is trivial that for each $\psi \in S(v)$,

$$\boldsymbol{e}_\psi^T \boldsymbol{C}\boldsymbol{y} = \sum_{\omega \in \Phi_C} \omega(t)$$

where $\boldsymbol{e}_\psi$ is the standard unit vector with the element corresponding to stage ψ being 1. On the other hand, for each $\psi \in S(v)$,

$$\boldsymbol{e}_\psi^T \boldsymbol{D}diag(\boldsymbol{\beta})\boldsymbol{y} = \sum_{\phi \in P(\psi) \cap \Phi_V} d_{\psi\phi}\beta_\phi \phi(t)$$

Combining the two equations, for a stage ψ that is a successor of any vulnerable stage,

$$\boldsymbol{e}_\psi^T diag(\boldsymbol{D}diag(\boldsymbol{\beta})\boldsymbol{y})\boldsymbol{C}\boldsymbol{y} = \sum_{\phi \in P(\psi) \cap \Phi_V} d_{\psi\phi}\beta_\phi \phi(t) \sum_{\omega \in \Phi_C} \omega(t)$$

Since the column sum of disease transition matrix $\boldsymbol{D}$ is 1 for any vulnerable stage, adding up this value for all stages with predecessors of vulnerable stages, we have

$$\sum_{\psi : P(\psi) \cap \Phi_V \neq \emptyset} \boldsymbol{e}_\psi^T diag(\boldsymbol{D}diag(\boldsymbol{\beta})\boldsymbol{y})\boldsymbol{C}\boldsymbol{y} = \sum_{\psi : P(\psi) \cap \Phi_V \neq \emptyset} \sum_{\phi \in P(\psi) \cap \Phi_V} d_{\psi\phi}\beta_\phi \phi(t) \sum_{\omega \in \Phi_C} \omega(t)$$

$$= \sum_{\phi \in \Phi_V} \beta_\phi \phi(t) \sum_{\omega \in \Phi_C} \omega(t)$$

which is the negative of the summation of outflows in vulnerable stages. Therefore, the validity of $\boldsymbol{y}_4' = diag(\boldsymbol{D}diag(\boldsymbol{\beta})\boldsymbol{y})\boldsymbol{C}\boldsymbol{y}$ is justified.

Combining the population changes in $\boldsymbol{y_3'}$ and $\boldsymbol{y_4'}$, the change of population in each stage due to the introduction of new infections in the system can be written as:

$$\boldsymbol{y_{inf}'} = diag((\boldsymbol{D} + \boldsymbol{I})diag(\boldsymbol{\beta})\boldsymbol{y})\boldsymbol{C}\boldsymbol{y}$$

We have derived the mathematical expressions for $\boldsymbol{y_{dev}'}$ to model the population change due to natural progression of diseases and $\boldsymbol{y_{inf}'}$ to describe the population change due to new infections. Summing up the two, we obtain a compact expression of $\boldsymbol{y}'$. Specifically, our general-purpose modeling framework for contact-based diseases in the matrix form is given by:

$$\boldsymbol{y}' = \boldsymbol{y_{dev}'} + \boldsymbol{y_{inf}'} = (\boldsymbol{D} - \boldsymbol{I})diag(\boldsymbol{\mu})\boldsymbol{y} + diag((\boldsymbol{D} + \boldsymbol{I})diag(\boldsymbol{\beta})\boldsymbol{y})\boldsymbol{C}\boldsymbol{y} \qquad \textbf{(GModel-CB)}$$

This equation does not include the natural birth and death of the population. An additional term $\lambda(\boldsymbol{y})$ can be added to the earlier equation to describe the change of population in each stage due to natural birth and death.

Modeling Framework for Vector-Borne Diseases

Vector-borne diseases are transmitted by vector bites and do not usually transmit directly between humans. Examples of vector-borne diseases include dengue, West Nile, yellow fever, chikungunya, and Zika. where numerous compartmental models have been developed. In this section, we develop the general-purpose modeling framework for vector-borne infectious diseases.

For simplicity, we assume there are only two groups of populations in the system: humans and vectors. Susceptible humans get infected through contact with infectious vectors, and infectious humans can infect susceptible vectors through contact as well. We assume an SEIR structure for

human population, and an SID structure for vector population, where D stands for deceased. Figure 2 demonstrates a disease transition diagram for this setting, where subscript H stands for the human stages, and subscript V stands for the vector stages. The black solid arrows refer to natural progression of diseases, and the red dashed arrows to contact between humans and vectors that cause infections.

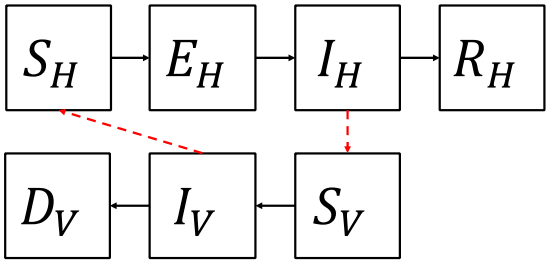

Figure 2: Demonstration of stage transitions for a vector-borne disease model.

Let Φ^H denote the collection of all disease stages in the human population and let Φ^V denote all stages in the vector population. Let $\boldsymbol{y}_H$ and $\boldsymbol{y}_V$ denote the size of each stage in the human and vector populations respectively. The categorization of disease stages introduced in the previous section still works with the separation of human and vector stages. Since the symptoms and outcomes of infection is different between humans and vectors, all model parameters need to be distinguished by population groups. Let $\boldsymbol{\mu}_H$ and $\boldsymbol{\mu}_V$ denote the mean transition rate for active stages in human population and vector population respectively, and let $\mu_\phi = 0$ for passive stages in both human and vector populations. Similarly, the disease transition matrices for human and vector populations are denoted by $\boldsymbol{D}_H$ and $\boldsymbol{D}_V$.

Let $\boldsymbol{\beta}_{VH}$ and $\boldsymbol{\beta}_{HV}$ denote the baseline infection rates for human and vector populations due to the inter-species contacts adjusted by the total free human population in the system. Although the

contact rate between the two populations should be the same for both human and vector populations, the probabilities of getting infected may differ. Consider the scenario in which the overall contact rate between the two groups is b, and the probabilities of getting infectious from these contacts in each disease stage are $\boldsymbol{p}_H$ for human, and $\boldsymbol{p}_V$ for vector respectively, then $\boldsymbol{\beta}_{VH} = b\boldsymbol{p}_H$ and $\boldsymbol{\beta}_{HV} = b\boldsymbol{p}_V$. In addition to inter-species contacts, intra-species contacts may also introduce new infections to the system, depending on the biological properties of the disease. To model this, let $\boldsymbol{\beta}_{HH}$ and $\boldsymbol{\beta}_{VV}$ be the baseline infection rates due to the contacts within the same population group. These two parameters have the same meaning as the baseline infection rate $\boldsymbol{\beta}$ in the model for contact-based diseases.

Since the infection structure involves the interaction of two populations, there should be two different disease contagious matrices for each direction of the infection as well. Let $\boldsymbol{C}_{VH} = \langle c_{ab} \rangle|_{a\in\Phi^H, b\in\Phi^V}$ denote the disease contagious matrix for vectors infecting humans, where $c_{ij} = -1$ if $i \in \Phi_V^H, j \in \Phi_C^V$, and $c_{kj} = 1$ for all $k \in S(i)$ if $i \in \Phi_V^H, j \in \Phi_C^V$. Similarly, let $\boldsymbol{C}_{HV} = \langle c_{ab} \rangle|_{a\in\Phi^V, b\in\Phi^H}$ be the disease contagious matrix for humans infecting vectors, where $c_{ij} = -1$ if $i \in \Phi_V^V, j \in \Phi_C^H$, and $c_{kj} = 1$ for all $k \in S(i)$ if $i \in \Phi_V^V, j \in \Phi_C^H$. If infections within human or vector population group are possible, the intra-group disease contagious matrices $\boldsymbol{C}_{HH}$ and $\boldsymbol{C}_{VV}$ can be defined n the same manner as in the model for contact-based diseases.

Similar to the modeling framework for contact-based diseases, the model for vector-borne disease can be split into two parts: 1) the population change in stages due to natural disease progression, and 2) the population change due to new infections. Since the natural progression of diseases does

not involve the interaction between humans and vectors, this part can be written independently for human and vector population:

$$\boldsymbol{y'_{H,dev}} = (\boldsymbol{D}_H - \boldsymbol{I})diag(\boldsymbol{\mu}_H)\boldsymbol{y}_H, \qquad \boldsymbol{y'_{V,dev}} = (\boldsymbol{D}_V - \boldsymbol{I})diag(\boldsymbol{\mu}_V)\boldsymbol{y}_V$$

The change in populations due to new infections can be further split into two parts: the intra-species infections and inter-species infections. The approach to model intra-species infections is the same as the model for contact-based diseases, since this change does not involve interactions between different species:

$$\boldsymbol{y'_{H,intra-inf}} = diag\big((\boldsymbol{D}_H + \boldsymbol{I})diag(\boldsymbol{\beta}_{HH})\boldsymbol{y}_H\big)\boldsymbol{C}_{HH}\boldsymbol{y}_H$$

$$\boldsymbol{y'_{V,intra-inf}} = diag\big((\boldsymbol{D}_V + \boldsymbol{I})diag(\boldsymbol{\beta}_{VV})\boldsymbol{y}_V\big)\boldsymbol{C}_{VV}\boldsymbol{y}_V$$

To model the inter-species infections, we assume the total human and vector populations at time t is $N_H(t)$ and $N_V(t)$, respectively. Consider a vulnerable stage ϕ for the human population. Since the baseline contact rate of vectors infecting humans adjusted by the human population in stage ϕ is $\beta_{VH,\phi}$, the number of infectious contacts a human makes with the vector population in a unit time is $\beta_{VH,\phi}N_V(t)$. Among the $N_V(t)$ vectors, $C_V(t)$ of them are contagious. Therefore, the exposure rate to infections for an individual human in stage ϕ is $\beta_{VH,\phi}C_V(t)$.

Therefore, for a vulnerable stage ϕ in the human population, the change of size due to new infection is $-\beta_{VH,\phi}\phi(t)\sum_{\psi\in\Phi_C^V}\psi(t)$. Rewrite it into a compact form using the disease contagious matrix $\boldsymbol{C}_{VH}$, the reductions of population in all human stages due to new infections can be written as $diag(diag(\boldsymbol{\beta}_{VH})\boldsymbol{y}_H)\boldsymbol{C}_{VH}\boldsymbol{y}_V$. Again, the destinations of the outflows due to new infections are captured in the disease transition matrix for human $\boldsymbol{D}_H$, the inflows for all human stages can be

written as $diag(\boldsymbol{D}_H diag(\boldsymbol{\beta}_{VH})\boldsymbol{y}_H)\boldsymbol{C}_{VH}\boldsymbol{y}_V$, following the same reasoning as in the contact-based disease model. Therefore, the change in population in the human group due to inter-species infections is:

$$\boldsymbol{y'}_{\boldsymbol{H,inter-inf}} = diag\big((\boldsymbol{D}_H + \boldsymbol{I})diag(\boldsymbol{\beta}_{VH})\boldsymbol{y}_H\big)\boldsymbol{C}_{VH}\boldsymbol{y}_V$$

Similarly, for a vulnerable stage ϕ in the vector population, the baseline contact rate of humans infecting vectors in stage ϕ adjusted by the human population is $\beta_{HV,\phi}$ per unit time. Assume the total contagious human population at time t is $C_H(t)$, thus the exposure rate to infections for vectors in stage ϕ is $\beta_{HV,\phi}C_H(t)$. Then for the same vulnerable stage ϕ in the vector population, the change of size due to new infection is $-\beta_{HV,\phi}\phi(t)\sum_{\psi\in\Phi_C^H}\psi(t)$. Using the disease contagious matrix of human infecting vectors $\boldsymbol{C}_{HV}$, the reductions of populations in all vector stages due to new infections is $diag(diag(\boldsymbol{\beta}_{HV})\boldsymbol{y}_V)\boldsymbol{C}_{HV}\boldsymbol{y}_H$. With $\boldsymbol{D}_V$ capturing the information of the destinations of the outflows from stages due to new infections, the inflows for all vector stages is $diag(\boldsymbol{D}_V diag(\boldsymbol{\beta}_{HV})\boldsymbol{y}_V)\boldsymbol{C}_{HV}\boldsymbol{y}_H$. Summing up outflows and inflows, the changes in population in the vector group due to inter-species infections are:

$$\boldsymbol{y'}_{\boldsymbol{V,inter-inf}} = diag\big((\boldsymbol{D}_V + \boldsymbol{I})diag(\boldsymbol{\beta}_{HV})\boldsymbol{y}_V\big)\boldsymbol{C}_{HV}\boldsymbol{y}_H$$

Adding up the changes in population due to disease progression, intra-species infections and inter-species infections, the general-purpose modeling framework for vector-borne diseases **(GModel-VB)** can be written as:

$$\boldsymbol{y'_H} = (\boldsymbol{D}_H - \boldsymbol{I})diag(\boldsymbol{\mu}_H)\boldsymbol{y}_H + diag\big((\boldsymbol{D}_H + \boldsymbol{I})diag(\boldsymbol{\beta}_{HH})\boldsymbol{y}_H\big)\boldsymbol{C}_{HH}\boldsymbol{y}_H$$
$$+ diag\big((\boldsymbol{D}_H + \boldsymbol{I})diag(\boldsymbol{\beta}_{VH})\boldsymbol{y}_H\big)\boldsymbol{C}_{VH}\boldsymbol{y}_V$$

$$\boldsymbol{y_V'} = (\boldsymbol{D}_V - \boldsymbol{I})diag(\boldsymbol{\mu}_V)\boldsymbol{y}_V + diag\big((\boldsymbol{D}_V + \boldsymbol{I})diag(\boldsymbol{\beta}_{VV})\boldsymbol{y}_V\big)\boldsymbol{C}_{VV}\boldsymbol{y}_V$$
$$+ diag\big((\boldsymbol{D}_V + \boldsymbol{I})diag(\boldsymbol{\beta}_{HV})\boldsymbol{y}_V\big)\boldsymbol{C}_{HV}\boldsymbol{y}_H$$

This modeling framework for vector-borne diseases can be further expanded to include multiple species of vectors by introducing proper model parameters and interaction terms between each combination of human and vector species. Additionally, the effect of alternative hosts for vector-borne diseases other than humans can also be introduced, by adding their own compartments, or use their populations to adjust the baseline infection rate.

Basic Reproduction Number

The basic reproduction number R_0, defined as the average number of secondary infections caused by one infectious individual in a system consists of only susceptible population, is one of the major concepts in epidemiology. It measures the transmission potential of a disease, and it also serves as a threshold for stability of a disease-free equilibrium of the ODE system (18). If $R_0 < 1$, the number of new infections introduced to the system will fail to replace themselves, and the outbreak containment will begin; if $R_0 > 1$, the infectious population will increase and the disease will spread.

Castillo-Chavez et al.(19), Van den Driessche and Watmough (20, 21) and Diekmann et al. (22) used the approach of next generation matrix at equilibrium to define the basic reproduction number R_0. We will also derive the basic reproduction number for our modeling framework using the next generation matrix approach. First, we assume that the system has a disease-free equilibrium $\boldsymbol{y}_\infty$ such that $\phi(\infty) = 0$ for all $\phi \in \Phi_A$, i.e., there is no population in active stages, thus no new

infections will be introduced to the system spontaneously. Assume a new infectious individual is introduced to the system at this equilibrium, and the populations in all passive stages are treated as constants, then we may only consider the population changes in active stages. Let $\boldsymbol{y}_A$ denote the partial population vector $\boldsymbol{y}$ with only active stages. Let $\boldsymbol{A}$ be a $|\Phi_A| \times |\Phi|$ matrix such that $a_{\phi\phi} = 1$ if and only if $\phi \in \Phi_A$, while all other elements are 0. For any $|\Phi| \times |\Phi|$ matrix $\boldsymbol{B}$, $\boldsymbol{ABA}^T$ gives a sub-matrix of $\boldsymbol{B}$ which consists of only rows and columns in Φ_A. Then, following the general-purpose modeling framework, the population changes in active stages after introducing a new infection at equilibrium $\boldsymbol{y}_\infty$ is expressed as

$$\boldsymbol{y}'_A = \boldsymbol{A}(\boldsymbol{D} - \boldsymbol{I})diag(\boldsymbol{\mu})\boldsymbol{A}^T\boldsymbol{y}_A + \boldsymbol{A}diag((\boldsymbol{D} + \boldsymbol{I})diag(\boldsymbol{\beta})\boldsymbol{y}_\infty)\boldsymbol{C}\boldsymbol{A}^T\boldsymbol{y}_A$$

Following the notation in Van den Driessche and Watmough (20, 21), we denote

$$\boldsymbol{V} = -\boldsymbol{A}(\boldsymbol{D} - \boldsymbol{I})diag(\boldsymbol{\mu})\boldsymbol{A}^T, \quad \boldsymbol{F} = \boldsymbol{A}diag((\boldsymbol{D} + \boldsymbol{I})diag(\boldsymbol{\beta})\boldsymbol{y}_\infty)\boldsymbol{C}\boldsymbol{A}^T$$

Then $\boldsymbol{y}'_A$ can now be written as $\boldsymbol{y}'_A = (\boldsymbol{F} - \boldsymbol{V})\boldsymbol{y}_A$.

The number of infections produced by the new infectious individual is the product of the expected duration of this individual staying infectious and the rate of introducing new infections. The rate of introducing new infections is already captured in matrix $\boldsymbol{F}$. To calculate the expected duration of staying infectious, consider the following system which only involves disease progression with a specified initial value:

$$\boldsymbol{y}'_A = -\boldsymbol{V}\boldsymbol{y}_A, \quad \boldsymbol{y}_A(0) = \boldsymbol{y_0}$$

The solution of this system is $h(t, \boldsymbol{y_0}) = e^{-\boldsymbol{V}t}\boldsymbol{y_0}$, where each component in this solution can be interpreted as the probability that the infectious individual represented by $\boldsymbol{y_0}$ introduced at $t = 0$

is in the corresponding disease stage at time t. Therefore, the total number of new infections introduced by $\boldsymbol{y_0}$ is

$$\int_0^{\infty} \boldsymbol{F}h(t, \boldsymbol{y_0})dt = \boldsymbol{F}\int_0^{\infty} e^{-\boldsymbol{V}t}dt\boldsymbol{y_0} = \boldsymbol{FV}^{-1}\boldsymbol{y_0}$$

where the matrix $\boldsymbol{K} = \boldsymbol{FV}^{-1}$ is the next generation matrix for the system at the disease-free equilibrium $\boldsymbol{y}_{\infty}$. The (ϕ, ψ) entry of matrix $\boldsymbol{K}$ is the expected number of new infections in stage ϕ produced by infectious individuals initially in stage ψ. Van den Driessche and Watmough (20, 21) showed that $\boldsymbol{K}$ has nonnegative eigenvalues, and the basic reproduction number of the system R_0 is given by $R_0 = \rho(\boldsymbol{K})$, where $\rho(\boldsymbol{K})$ is the spectral radius of matrix $\boldsymbol{K}$, i.e., the maximum of the moduli of the eigenvalues of $\boldsymbol{K}$; and the eigenvector $\boldsymbol{\omega}$ associated with R_0 is also nonnegative. An interpretation of this definition of basic reproduction number is that if the distribution of the infectious individual introduced at the equilibrium follows the eigenvector $\boldsymbol{\omega}$, then the maximal number of typical secondary infections produced by the initial infection will be R_0.

Application of Our Disease Modeling Framework

By design, our general-purpose disease modeling framework can accommodate different types of transmission mechanisms and allows for incorporation of multiple hosts and vectors. All classical compartmental disease models can readily be derived using this modeling framework. Moreover, it enables derivation and investigation of new disease models. Below, we demonstrate its application to two classical compartmental disease models and establish new results for a new model.

SEIR Model for Contact-Based Diseases

The SEIR model (Figure 1) is widely used in epidemiology. We will show that it can be derived readily using our modeling framework. Specifically, we arrange the four stages as $\{S, E, I, R\}$, then the disease transition matrix is

$$\boldsymbol{D} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

and the disease contagious matrix is

$$\boldsymbol{C} = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Now assume the baseline infectious rate adjusted by the total human population is β, and the mean dwelling times at the exposed and infectious stage are $1/\mu_E$ and $1/\mu_I$. Then, the vectors $\boldsymbol{\beta}$ and $\boldsymbol{\mu}$ can be written as:

$$\boldsymbol{\beta} = [\beta, 0, 0, 0]^T, \boldsymbol{\mu} = [0, \mu_E, \mu_I, 0]^T$$

Plugging these parameters to our general-purpose modeling framework for contact-based disease (GModel-CB) yields the following system of ordinary differential equations:

$$\frac{d}{dt}S = -\beta SI$$

$$\frac{d}{dt}E = \beta SI - \mu_E E$$

$$\frac{d}{dt}I = \mu_E E - \mu_I I$$

$$\frac{d}{dt}R = \mu_I I$$

To calculate the basic reproduction number of this system, we observe that there is no closed-form disease-free equilibrium. Therefore, assume this equilibrium is $\boldsymbol{y}_{\infty} = \{S^*, 0, 0, R^*\}$. Then following the definition of next-generation matrix,

$$\boldsymbol{F} = \begin{pmatrix} 0 & \beta S^* \\ 0 & 0 \end{pmatrix}, \boldsymbol{V} = \begin{pmatrix} \mu_E & 0 \\ -\mu_E & \mu_I \end{pmatrix}, \boldsymbol{K} = \boldsymbol{F}\boldsymbol{V}^{-1} = \begin{pmatrix} \beta S^*/\mu_I & \beta S^*/\mu_I \\ 0 & 0 \end{pmatrix},$$

the basic reproduction number of the system is given by

$$R_0 = \rho(\boldsymbol{K}) = \max\left\{0, \frac{\beta S^*}{\mu_I}\right\} = \beta S^*/\mu_I \qquad \text{(R0-1)}$$

SEPAIR Model for Contact-Based Diseases

Lee et al. (23) first proposed a **S**usceptible, **E**xposed, initial infectious **P**eriod, **A**symptomatically infectious, symptomatically **I**nfectious, and **R**ecover (SEPAIR) model to describe the propagation of contact-based diseases outside and within point-of-dispensing (POD) facilities. The analysis is essential since during mass vaccination, PODs could become hot-spots for disease transmission. The SEPAIR model extends the classic SEIR model with two additional disease stages: asymptomatically infectious and symptomatically infectious. This 6-stage propagation model provides more opportunities to examine the interaction between POD layout design and disease propagation. For example, one can gain a better understanding of triage accuracy on the degree of disease spread. Specifically, during triage, providers may fail to identify infectious individuals who exhibit no symptoms, thus leading to increase in intra-facility disease spread during mass vaccination events (23). Careful POD layout design to ensure short queues, small crowds and social distancing can be carried out to mitigate the effect. The SEPAIR model has proven to be critical in analyzing the current covid-19 pandemic since early clinical reports have shown that (undetected infected covid-19) individuals without symptoms are shedding SARS-CoV-2 RNA

comparably to symptomatic patients (31, 32). Lee et al. further extended the 6-stage model to include also post-recovery infectious (Q) and hospitalization (H) (30, 31). The 8-stage model couples hospital resources availability and their effect on the overall disease mitigation.

We will establish new results related to the basic reproduction number for SEPAIR using our general-purpose modeling framework. In the SEPAIR model, every individual will enter an initial infectious period (P) after exposure (E). Then the individual will either become asymptomatically infectious (A) with probability $1 - p_S$ or symptomatically infectious (I) with probability p_S. Individuals in stages P, A, or I are contagious. In the derivation, for brevity, we consider only outer-POD disease propagation without vaccination effect and adapt the notations and parameters from the original articles(23, 30).

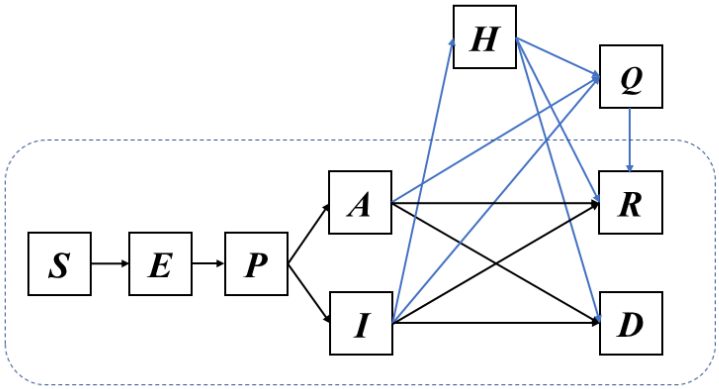

Figure 3: Disease stage transition diagram for a SEPAIR model (23), and its extension to a 8-stage model for covid-19 analysis (30). Here, individuals start to be infectious during late stage of incubation, and some recovered individuals may remain infectious (German studies). Hence in our disease models, we split these individuals, (Q and R), so that they belong to a unique category.

To calculate the basic reproduction number, assume that the equilibrium is achieved at $\boldsymbol{y}_\infty = \{S^*, 0, 0, 0, 0, R^*\}$, i.e., all active stages will have no population by the end of the outbreak. Then following the definition of the next-generation matrix and ordering the stages as $\{E, P, A, I\}$ in the matrix representations, we have

$$\boldsymbol{F}=\begin{pmatrix}0 & \beta S^* & \beta S^* & \beta S^*\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{pmatrix}, \boldsymbol{V}=\begin{pmatrix}\mu_E & 0 & 0 & 0\\ -\mu_E & \mu_P & 0 & 0\\ 0 & -(1-p_S)\mu_P & \mu_A & 0\\ 0 & -p_S\mu_P & 0 & \mu_I\end{pmatrix},$$

$$\boldsymbol{K}=\boldsymbol{F}\boldsymbol{V}^{-1}=\begin{pmatrix}\left(\frac{1}{\mu_P}+\frac{1-p_S}{\mu_A}+\frac{p_S}{\mu_I}\right)\beta S^* & \left(\frac{1}{\mu_P}+\frac{1-p_S}{\mu_A}+\frac{p_S}{\mu_I}\right)\beta S^* & \frac{\beta S^*}{\mu_A} & \frac{\beta S^*}{\mu_I}\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\end{pmatrix}$$

The basic reproduction number of the SEPAIR system is given by

$$R_0=\rho(\boldsymbol{K})=\max\left\{0,0,0,\left(\frac{1}{\mu_P}+\frac{1-p_S}{\mu_A}+\frac{p_S}{\mu_I}\right)\beta S^*\right\}=\left(\frac{1}{\mu_P}+\frac{1-p_S}{\mu_A}+\frac{p_S}{\mu_I}\right)\beta S^* \qquad \text{(R0-2)}$$

Using the same approach, we can deduce the basic reproduction number for the 8-stage SEPAIHQR system as $R_0=\left(\frac{1}{\mu_P}+\frac{1-p_S}{\mu_A}+\frac{p_S}{\mu_I}+\frac{(1-p_S)\,p_{AQ}+p_S\,p_{IQ}}{\mu_Q}\right)\beta S^*$ where p_{AQ}, and p_{IQ} corresponds to the probability of (a)symptomatic patient recovers but remains infectious. The results demonstrate clear interplay of the dwell times of contagious stages and their effect and extent of contribution to disease spread.

SEIR-SID Model for Vector-Borne Diseases

The SEIR-SID model shown in Figure 2 is the most basic model setup for vector-borne diseases. It can easily be derived with our general-purpose modeling framework for vector-borne disease. We first arrange the human and vector compartments as $\{S_H, E_H, I_H, R_H\}$ and $\{S_V, I_V, D_V\}$. Based on the disease transition diagram, the disease transition matrices for this model are

$$\boldsymbol{D}_H=\begin{pmatrix}0 & 0 & 0 & 0\\ 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\end{pmatrix}, \boldsymbol{D}_V=\begin{pmatrix}0 & 0 & 0\\ 1 & 0 & 0\\ 0 & 1 & 0\end{pmatrix}$$

and the disease contagious matrices are

$$\boldsymbol{C}_{VH} = \begin{pmatrix} 0 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \boldsymbol{C}_{HV} = \begin{pmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Assume the infectious contact rate between human and vector population is β, the mean dwelling time at the exposed and infectious stage is $1/\mu_E$ and $1/\mu_I$ for humans, and the mean dwelling time at infectious stage is $1/\mu_V$ for vectors. Then the vectors $\boldsymbol{\beta}_{VH}, \boldsymbol{\beta}_{HV}, \boldsymbol{\mu}_H, \boldsymbol{\mu}_V$ can be written as

$$\boldsymbol{\beta}_{VH} = [\beta, 0,0,0]^T, \boldsymbol{\beta}_{HV} = [\beta, 0,0]^T, \boldsymbol{\mu}_H = [0, \mu_E, \mu_I, 0]^T, \boldsymbol{\mu}_V = [0, \mu_V, 0]^T$$

For brevity, we omit the intra-group infections. Plugging these parameters to the general-purpose modeling framework for vector-borne disease (GModel-VB) yields the following system of ordinary differential equations:

$$\frac{d}{dt} S_H = -\beta S_H I_V$$

$$\frac{d}{dt} E_H = \beta S_H I_V - \mu_E E_H$$

$$\frac{d}{dt} I_H = \mu_E E_H - \mu_I I_H$$

$$\frac{d}{dt} R_H = \mu_I I_H$$

$$\frac{d}{dt} S_V = -\beta S_V I_H$$

$$\frac{d}{dt} I_V = \beta S_V I_H - \mu_V I_V$$

$$\frac{d}{dt} D_V = \mu_V I_V$$

This system of ordinary differential equations does not have a closed-form solution for disease-free equilibrium. To calculate the basic reproduction number, assume that the disease-free equilibrium is achieved at $\boldsymbol{y}_{\infty,H} = \{S_H^*, 0, 0, R_H^*\}, \boldsymbol{y}_{\infty,V} = \{S_V^*, 0, D_V^*\}$. Then following the definition of next-generation matrix,

$$\boldsymbol{F} = \begin{pmatrix} 0 & 0 & \beta S_H^* \\ 0 & 0 & 0 \\ 0 & \beta S_V^* & 0 \end{pmatrix}, \boldsymbol{V} = \begin{pmatrix} \mu_E & 0 & 0 \\ -\mu_E & \mu_I & 0 \\ 0 & 0 & \mu_V \end{pmatrix}, \boldsymbol{K} = \boldsymbol{F}\boldsymbol{V}^{-1} = \begin{pmatrix} 0 & 0 & \beta S_H^*/\mu_V \\ 0 & 0 & 0 \\ \beta S_V^*/\mu_I & \beta S_V^*/\mu_I & 0 \end{pmatrix},$$

the basic reproduction number of the system is given by

$$R_0 = \rho(\boldsymbol{K}) = \sqrt{\frac{S_H^* S_V^*}{\mu_I \mu_V}}\beta \qquad \text{(R0-3)}$$

Discussion

In this study, we proposed a general-purpose modeling framework for analyzing the spread of infectious diseases. This modeling framework extends the traditional compartmental models by categorizing each stage or compartment based on two different criteria: (1) passive or active, and (2) vulnerable, contagious, or unaffected, and establishing relationships between stages based on their categorization. Using this framework, we derived the mathematical expression for the contact-based and vector-borne disease models. The modeling technique for vector-borne diseases can depict the system dynamics of multiple species (> 2, for both the hosts and the vectors), given the well-defined inter-group disease contagious matrices. We derived the basic reproduction number of the system using the next-generation matrix approach and demonstrated applications of the modeling framework on the SEIR and SEPAIR models for contact-based diseases and the SEIR-SID model for vector-borne diseases.

We illustrated how to derive new results on the basic reproduction number for the SEPAIR and SEPAIHQR model using the general-purpose framework. The result, $R_0 = \left(\frac{1}{\mu_P} + \frac{1-p_S}{\mu_A} + \frac{p_S}{\mu_I}\right)\beta S^*$, reveals significant contribution of asymptomatic individuals towards the disease spread. For the current covid-19 pandemic, the result offers justification the necessity of early adoption of strategic diagnostic testing (sampling to include no-symptom individuals) for case discovery and biosurveillance, universal facemasks to reduce transmission, and contact-trace and self-quarantine for both symptomatic and asymptomatic patients. In a model developed in February 2020 (30), Lee et al. use p_S= 2/3 for symptomatic cases for non-pharmaceutical intervention analysis.. Recent clinical analysis puts p_S in the range of $38\% \; to \; 63.7\%$ (33,34,35).

From the basic reproduction number equations R0-1 to R0=3, we can observe the intimate interplay of dwell times of various contagious stages and host-vectors and their contribution to disease spread. These offer a simple yet elegant framework for decision-makers to objectively and rapidly contrast different interventions and understand their effects in disease mitigation and tradeoffs.

Using this proposed modeling framework, all classical epidemiology models can be derived. It also allows for derivation and investigation of new disease models. This modeling framework describes the dynamics of the disease transmission system with a short and compact formulation and can easily be generalized and customized based on the biological properties of different infectious diseases.

Infectious diseases remain a major threat to public health. In addition to common infectious diseases, new and re-emerging diseases continually challenge us. The devastation of covid-19 underscores the necessity of multi-level and timely interventions. Accurate prediction and immediate response upon outbreaks are of paramount importance for early containment and mitigation. Mathematical models with system dynamics and interactions are powerful tools for analyzing the trend of disease outbreaks and supporting real-time decision-making. However, building epidemiology models from scratch during the emergency is not efficient and will cause potential delays in assessing and mitigating risks. The proposed modeling framework provides an elegant meta-model for epidemiology and enables epidemiologists and emergency public health responders to rapidly build, evaluate, and implement disease models according to the type of outbreak without diving into the interactions among multiple stages and explicitly deriving the ordinary differential equations. This can save time and lives during public health emergencies.

With the availability of modern computing technologies, the idea of "digital surveillance" using automated and computerized methods to track and predict the spread of infectious diseases is becoming increasingly important (24, 25). Our proposed modeling framework provides a practical backend engine for such a digital disease surveillance system since it is highly general-purpose and can be adapted to fit different types of computations. Furthermore, the matrix formulation is computationally advantageous since most programming languages have libraries with fast matrix and vector multiplication algorithms implemented. We have developed RealOpt-ASSURE, a digital disease surveillance, response, and decision-making system (26). It connects the modeling framework with a graphical user interface and allows users to design a disease transition diagram and input model parameters that fit the needs of an unfolding outbreak; and translates the input

automatically on the backend into a system of ordinary differential equations. Public health users can model, solve, and interpret the results without having to manipulate the complex mathematical equations.

Medical interventions can be introduced into the modeling framework implicitly or explicitly. When the process of medical interventions is treated as a black box, we can create new stages to the intervention process by assigning proper values in all model matrices. For example, hospitalization can be treated as a new stage in the SEIR model with its own transition rates (23, 30). Another solution is to model the interventions in detail by segmenting the population into different groups. For example, to model the effect of vaccination, we can separate the population into two groups: outside and inside vaccine clinics. For the population outside the vaccine clinics, the previous disease propagation model will be used; for the population inside the vaccine clinics, in addition to the natural propagation of the disease, the vaccination effect will also drive the transition between disease stages. This modeling method is suitable for studying the effect of operations at healthcare facilities or PODs during emergency disease outbreaks and identifying the optimal operation strategies. We have discussed this concept in mass dispensing (23, 27) and in vaccine prioritization analysis (28).

Optimization problems can also be formulated based on the modeling framework to determine the optimal allocation of medical resources or to compare the effects of different containment strategies. Instead of treating all model matrices as constants, they can be modeled as time-variant variables whose values are controllable through human interventions at a certain cost. Then optimization problems can be formulated and solved to achieve containment under various

constraints, and assist on-the-ground operations for disaster relief effort. We have successfully applied this type of optimization problem on containing the 2014 Ebola virus outbreaks in West Africa by determining the minimum number of beds required to reach containment (26), and evaluating different intervention strategies to contain the 2016-2017 Zika virus outbreak in Brazil and Puerto Rico (29).

As in any compartmental models, estimating model parameters are essential for this general-purpose modeling framework. The flexibility of the model enables one to incorporate a large number of parameters to describe the system dynamics in detail. Some of these parameters can be obtained through the biological properties of the diseases, others will require estimation. With high degree of freedom, improper estimations of model parameters may result in over-fitting. In practice, users should start with simple yet robust settings of the model and adjust and enhance as more data become available and a better understanding of the outbreak is obtained. Our previous research shows that simple models have sufficient predicting accuracy in many cases and should be preferred when accurate parameter estimations are not immediately available (26, 29). Sensitivity analysis is also useful to determine the relevant importance of different model parameters (24).

The spread of infectious diseases has intrinsic stochasticity due to the difference among individuals, the randomness in the behavior of individuals, and the heterogeneity among different population groups. Using Langevin dynamics, stochasticity can be introduced into our modeling framework. In this case, a stochastic term can be added to each differential equation in the model

and a confidence interval of the projected trend can be obtained. This can also be easily implemented computationally at the expense of generating random noises during each evaluation.

Funding

This study was partially supported by a grant from the National Institutes of Health, the National Science Foundation and the Centers for Disease Control and Prevention. The findings and conclusions in this manuscript are those of the authors and do not necessarily represent the official position of the funding agencies.

References

1. Binder S, Levitt AM, Sacks JJ, Hughes JM. Emerging infectious diseases: public health issues for the 21st century. Science. 1999;284(5418):1311-3.
2. Morens DM, Fauci AS. Emerging infectious diseases: threats to human health and global stability. PLoS Pathogens. 2013;9(7):e1003467.
3. Bernoulli D, Blower S. An attempt at a new analysis of the mortality caused by smallpox and of the advantages of inoculation to prevent it. Reviews in Medical Virology. 2004;14(5):275-88.
4. Hammer W. Epidemic disease in England: the evidence of variability and the persistence of type. The Lancet. 1906;167(4306):655-62.
5. Ross R. The Prevention of Malaria. London: John Murray; 1911.
6. Bailey NT. The Mathematical Theory of Infectious Diseases and Its Applications. High Wycombe: Charles Griffin & Company; 1975.
7. Kermack WO, McKendrick AG. Contributions to the mathematical theory of epidemics—I. Bulletin of Mathematical Biology. 1991;53(1-2):33-55.
8. Kermack WO, McKendrick AG. Contributions to the mathematical theory of epidemics. II. The problem of endemicity. Bulletin of Mathematical Biology. 1991;53(1-2):57-87.
9. Kermack WO, McKendrick AG. Contributions to the mathematical theory of epidemics. III. Further studies of the problem of endemicity. Bulletin of Mathematical Biology. 1991;53(1-2):89-118.
10. Li J, Zou X. Generalization of the Kermack-McKendrick SIR model to a patchy environment for a disease with latency. Mathematical Modelling of Natural Phenomena. 2009;4(2):92-118.
11. Pathak S, Maiti A, Samanta G. Rich dynamics of an SIR epidemic model. Nonlinear Analysis: Modelling and Control. 2010;15(1):71-81.
12. Xu Z. Traveling waves in a Kermack–Mckendrick epidemic model with diffusion and latent period. Nonlinear Analysis: Theory, Methods & Applications. 2014;111:66-81.

13. Inaba H. On a pandemic threshold theorem of the early Kermack–McKendrick model with individual heterogeneity. Mathematical Population Studies. 2014;21(2):95-111.
14. Disraelly DS, Walsh TJ, Curling CA. A new methodology for estimating contagious biological agent casualties as a function of time. Mathematical and Computer Modelling. 2011;54(1):649-58.
15. Daley DJ, Gani J. Epidemic Modelling: An Introduction. Cambridge: Cambridge University Press; 1999.
16. Hethcote HW. A thousand and one epidemic models. Frontiers in Mathematical Biology. Berlin, Heidelberg: Springer; 1994. p. 504-15.
17. Breda D, Diekmann O, De Graaf W, Pugliese A, Vermiglio R. On the formulation of epidemic models (an appraisal of Kermack and McKendrick). Journal of Biological Dynamics. 2012;6(sup2):103-17.
18. Diekmann O, Heesterbeek JAP. Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation. West Sussex: John Wiley & Sons; 2000.
19. Castillo-Chavez C, Feng Z, Huang W. On the computation of R0 and its role on global stability. In: Castillo-Chavez C, Blower S, Van den Driessche P, Kirschner D, Yakubu A-A, editors. Mathematical Approaches for Emerging and Reemerging Infectious Diseases: An Introduction. New York Springer-Verlag; 2002. p. 229–50.
20. Van den Driessche P, Watmough J. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Mathematical Biosciences. 2002;180(1):29-48.
21. Van den Driessche P, Watmough J. Further notes on the basic reproduction number. In: Brauer F, Van den Driessche P, Wu J, editors. Mathematical Epidemiology. Berlin, Heidelberg: Springer-Verlag; 2008. p. 159-78.
22. Diekmann O, Heesterbeek JAP, Metz JA. On the definition and the computation of the basic reproduction ratio R0 in models for infectious diseases in heterogeneous populations. Journal of Mathematical Biology. 1990;28(4):365-82.
23. Lee EK, Chen C-H, Pietz FH, Benecke BA. Disease propagation analysis and mitigation strategies for effective mass dispensing. AMIA Annual Symposium Proceedings. 2010;2010:427–31.
24. Lee EK, Liu Y, Pietz FH. A computational framework for a digital surveillance and response tool: application to avian influenza. AMIA Annual Symposium Proceedings. 2017;2017:1090-9.
25. Lee EK, Pietz FH, Chen C-H, Liu Y. An interactive web-based decision support system for mass dispensing, emergency preparedness, and biosurveillance. 2017 International Conference on Digital Health; London. New York: ACM; 2017. p. 137-46.
26. Lee EK. Analyzing intervention strategies for containing a pandemic outbreak. AAAS 2017 Annual Meeting; Boston. 2017.
27. Lee EK, Pietz FH, Benecke BA, Mason J, Burel G. Advancing public health and medical preparedness with operations research. Interfaces. 2013;43(1):79-98.
28. Lee EK, Yuan F, Pietz FH, Benecke BA, Burel G. Vaccine prioritization for effective pandemic response. Interfaces. 2015;45(5):425-43.
29. Lee EK, Liu Y, Pietz FH. A compartmental model for Zika virus with dynamic human and vector populations. AMIA Annual Symposium Proceedings. 2016;2016:743-52.
30. Lee EK, "Modeling and Evaluating Intervention Options and Strategies for COVID-19 Containment: Biological-Behavioral-Logistics Computation Decision Framework" Red Dawn Coronavirus Collaborative. Feb 2020.

31. Rothe C, M Schunk, P Sothmann, et al. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. N Engl J Med 2020; 382:970-971. Published on January 30, 2020, at NEJM.org.
32. Zou L, F Ruan, MX Huang, et al. SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients. N Engl J Med 2020; 382:1177-1179. Published on February 19, 2020, at NEJM.org.
33. Lee SJ, T Kim, EJ Lee, et al. Clinical Course and Molecular Viral Shedding Among Asymptomatic and Symptomatic Patients With SARS-CoV-2 Infection in a Community Treatment Center in the Republic of Korea. AMA Intern Med. 2020 August;180(11):1447–1452.
34. Jung CY, HY Park, DW Kim, et al. Clinical Characteristics of Asymptomatic Patients with COVID-19: A Nationwide Cohort Study in South Korea. Int J Infect Dis. 2020 Oct; 99: 266–268.
35. Johansson MA, Quandelacy TM, Kada S, et al. SARS-CoV-2 Transmission From People Without COVID-19 Symptoms. JAMA Netw Open. 2021;4(1):e2035057.