Original research article

An epidemic model for an evolving pathogen with strain-dependent immunity

Adam Griffin [a],[*], Gareth O. Roberts [b], Simon E.F. Spencer [b]

[a] UK Centre for Ecology & Hydrology, Benson Lane, Wallingford, Oxfordshire, UK
[b] Department of Statistics, University of Warwick, Coventry, UK

ARTICLE INFO

ABSTRACT

Between pandemics, the influenza virus exhibits periods of incremental evolution via a process known as antigenic drift. This process gives rise to a sequence of strains of the pathogen that are continuously replaced by newer strains, preventing a build up of immunity in the host population. In this paper, a parsimonious epidemic model is defined that attempts to capture the dynamics of evolving strains within a host population. The 'evolving strains' epidemic model has many properties that lie in-between the Susceptible–Infected–Susceptible and the Susceptible–Infected–Removed epidemic models, due to the fact that individuals can only be infected by each strain once, but remain susceptible to reinfection by newly emerged strains. Coupling results are used to identify key properties, such as the time to extinction. A range of reproduction numbers are explored to characterise the model, including a novel quasi-stationary reproduction number that can be used to describe the re-emergence of the pathogen into a population with 'average' levels of strain immunity, analogous to the beginning of the winter peak in influenza. Finally the quasi-stationary distribution of the evolving strains model is explored via simulation.

1. Introduction

Epidemic models have become important tools for understanding, predicting and developing mitigation strategies for public health planners dealing with infectious diseases. Recent advances in genetic epidemiology have greatly accelerated our understanding of the complex interactions between host immunity and pathogen evolution, and emphasised the important role that pathogen evolution can have on the dynamics of infection. However, it remains extremely challenging to combine together these two interacting processes within the same mathematical framework [1]. In this paper we develop a parsimonious epidemic model that describes the transmission dynamics of a multi-strain pathogen with evolutionary dynamics similar to the influenza A virus evolving via antigenic drift.

Multi-strain models have become increasingly popular due to the rise in availability of pathogen genetic analyses. Many models have been based on ordinary differential equations (ODE), despite the fact that stochastic effects play an important role in mutation [see [1] for a review]. Bichara et al. [2] develop an epidemic model with competition between finitely many pathogen strains, and include vertical transmission and immunity from maternal antibodies in the infection dynamics. Meehan et al. [3] analyse multi-strain epidemic models with mutation between strains within an ODE framework. However since their focus is on drug-resistance, they do not consider the effect of immunity. In

the multi-strain models discussed in Gog et al. [4], there is assumed to be a finite number of possible strains, and each individual may be infected with one or more of such strains. Evolution was been modelled by a random jump process on a finite strain space using a nearest neighbour jump process. Models involving a countable number of infectious statuses have been discussed in the past [5], but these typically only use the previously mentioned nearest-neighbour evolution. In [5] this is expressed as a model for parasitic infections where the "type" of an individual is defined by the quantity of parasites in a host. Despite the many modelling papers on multi-strain epidemics, the methodology required to fit these models to data is only just emerging [6].

Between pandemics, the 4 main sub-types of the influenza virus evolve according to a process called antigenic drift [7]. Antigenic drift arises due to the fact that infection with a particular strain of influenza provides the host with a long-lasting immunity to future infection by the same strain. Once immunity to a particular strain has built up in the population, there is a selection advantage to strains that do not elicit the same immune response. To capture within a mathematical model the complex processes driving the evolution of the influenza virus is extremely challenging due to the interactions between host immunity and viral evolution [8]. Nonetheless, simple models can give rise to surprisingly complex dynamics [9,10]. The H3N2 subtype of influenza A, in particular, exhibits a narrow spread in its evolutionary tree,

---

with all strains a short genetic distance from a single branch [11,12]. Each strain persists for a relatively short amount of time before being replaced.

In this paper we define a novel epidemic model with countably infinite, evolving strains that sits between the traditional susceptible–infected–susceptible (SIS) and susceptible–infected–removed (SIR) epidemic models, in that each individual may be infected many times with the pathogen, but only once by a strain. The model is designed to reflect the narrow pattern of evolution observed in pathogens undergoing antigenic drift, such as seasonal influenza. By introducing an equivalence relation on the state space, we are able to describe the equilibrium behaviour of the model prior to elimination of the pathogen. In Section 5, coupling arguments are used to make precise the relationship between our new model and the traditional SIS and SIR models and to explore the large population limit. In Section 6 we discuss three reproduction numbers for the novel model. Finally in Section 7 we explore simulations from the quasi-equilibrium distributions.

2. Definition of the model

2.1. SIRS with evolving strains (E-SIRS)

Consider a closed population of N individuals which are classified as *susceptible*, *infective* or *removed*. For each time t, we denote the number of susceptibles by $S(t)$, the number of infectives by $\mathcal{I}(t)$, and the number of removed individuals by $\mathcal{R}(t)$. An infective remains in this class for a random period of time known as their infectious period, after which they become removed. Similarly, removed individuals become susceptible again after their immune period, during which they cannot be infected by any strain (even a new one). We assume the durations of infectious periods are i.i.d. draws from $L_I \sim \text{Exp}(\gamma)$ and immune periods are i.i.d draws from $L_R \sim \text{Exp}(\delta)$.

To capture dynamics of competing and evolving strains, every individual has a strain index $k \in \mathbb{Z}$ which denotes the most recent strain with which an individual was (or is currently) infected. We denote the number of susceptibles, infectives and removed individuals respectively with strain index k by $S_k(t)$, $\mathcal{I}_k(t)$ and $\mathcal{R}_k(t)$. Finally we denote by $K^*(t)$ the largest strain index observed up to time t and use $K^* := K^*(t^-)$ where the time is clear from context.

As in the standard SIRS model [13], we assume homogeneous mixing of individuals, and so each pair of individuals makes contact at the points of a Poisson process with rate $\frac{\beta}{N} > 0$. New strains are introduced into the population in the following way. Each time an infective makes contact with a susceptible individual, we assume that with some probability $\theta \in [0,1]$, there is a successful infection of the susceptible with a previously unseen strain, which is given strain index $K^*(t^-) + 1$. With probability $1 - \theta$, the original strain in the infective attempts to infect the susceptible; the success of this infection depends on the strain index of the susceptible. For simplicity, we assume that immunity is cumulative: a susceptible with strain index k is immune to all strains with index $j \leq k$. Removed and susceptible individuals retain the strain index of the strain they have most recently recovered from. All contact processes, mutation events, infectious periods and immune periods are assumed to be independent from each other.

To summarise, the epidemic proceeds according to the following events.

- *Infection without mutation:*
  $(S_j(t), \mathcal{I}_k(t)) \mapsto (S_j(t) - 1, \mathcal{I}_k(t) + 1)$
  for all $j < k$, with rate $\beta(1 - \theta)N^{-1}S_j(t)\mathcal{I}_k(t)$.
- *Infection with mutation:*
  $(S_j(t), \mathcal{I}_{K^*+1}(t) = 0) \mapsto (S_j(t) - 1, \mathcal{I}_{K^*+1}(t) = 1)$ for $j \in \mathbb{Z}$ with rate $\beta\theta N^{-1}S_j(t)\mathcal{I}(t)$.
- *Recovery:*
  $(\mathcal{I}_k(t), \mathcal{R}_k(t)) \mapsto (\mathcal{I}_k(t) - 1, \mathcal{R}_k(t) + 1)$ for $k \in \mathbb{Z}$ with rate $\gamma\mathcal{I}_k(t)$.

- *Loss of global immunity:*
  $(\mathcal{R}_k(t), S_k(t)) \mapsto (\mathcal{R}_k(t) - 1, S_k(t) + 1)$ for $k \in \mathbb{Z}$ with rate $\delta\mathcal{R}_k(t)$.

The state space of the E-SIRS model is given by $\Omega' = \{(\mathbf{s}, \mathbf{i}, \mathbf{r}) : \sum_{k \in \mathbb{Z}}(s_k + i_k + r_k) = N\}$, with $\mathbf{s}, \mathbf{i}, \mathbf{r}$ being infinite sequences taking values in $\{0, 1, \ldots, N\}$. A natural initial condition might be $\mathcal{I}_1(0) = 1, \mathcal{I}_k(0) = 0$ for $k \neq 1$, $S_0(0) = N - 1$, $S_k(0) = 0$ for $k \neq 0$, and $\mathcal{R}_k(0) = 0$ for all k. This equates to a single currently infective individual infected with a strain to which all other individuals are susceptible, and to which no-one is currently recovering.

2.2. SIS with evolving strains (E-SIS)

Consider a second model where following an infectious period, an individual becomes immediately susceptible, corresponding to the E-SIRS model where $\delta = \infty$. In this model there are no periods of immunity and so recovery events generate susceptibles:

- *Recovery:* $(\mathcal{I}_k(t), S_k(t)) \mapsto (\mathcal{I}_k(t) - 1, S_k(t) + 1)$ for $k \in \mathbb{Z}$ with rate $\gamma\mathcal{I}_k(t)$.

plus infection transitions as above. The E-SIS model evolves over the subspace $\{(\mathbf{s}, \mathbf{i}) : \sum_{k \in \mathbb{Z}}(s_k + i_k) = N\}$. We will refer to both spaces by Ω', the meaning will always be clear from context.

2.3. Link to single-strain models

Consider the E-SIS model with $\theta = 1$. All contacts are mutation contacts and hence successful, and so $(S(t), \mathcal{I}(t))$, the total numbers of susceptibles and infectives, follow a traditional single-strain SIS model as defined in [14]. We can also perform a similar identification between $(S(t), \mathcal{I}(t), \mathcal{R}(t))$, the number of susceptibles, infectives and immune individuals in the E-SIRS model and the single-strain SIRS model as defined in [13].

On the other hand, consider the E-SIS model with $\theta = 0$. Since no contacts are mutations, no individual can be infected more than once. If the population starts with strain index 0 except for the initial infectives with strain 1, $(S_0(t), \mathcal{I}_1(t), S_1(t))$ behaves as a traditional single-strain SIR model as defined in [14].

3. Equivalence relation

We wish to study the long-term average behaviour of characteristics such as the levels of immunity and pathogen diversity, however the constant emergence and extinction of strains means that the evolving epidemic process has no steady-state. To counter this we introduce an equivalence relation to fix the process against the most recently emerged strain.

Definition 1. The *active strain set* of a state $(\mathbf{s}, \mathbf{i}, \mathbf{r}) \in \Omega'$ is given by $\mathbb{K} = \{k \in \mathbb{Z} : i_k > 0\}$. Let elements of this set be indexed from 1 to $|\mathbb{K}|$ in ascending order, so for $k_a, k_b \in \mathbb{K}$, we have $k_a < k_b$ whenever $a < b$.

Definition 2. Two states $(\mathbf{s}, \mathbf{i}, \mathbf{r})$ and $(\mathbf{s}', \mathbf{i}', \mathbf{r}') \in \Omega'$ are equivalent if and only if the following conditions hold.

1. The total numbers of susceptibles, infectives and removed individuals are equal: $|\mathbf{s}| = \sum_{k \in \mathbb{Z}} s_k = \sum_{k \in \mathbb{Z}} s'_k = |\mathbf{s}'|$, and similarly $|\mathbf{i}| = |\mathbf{i}'|$ and $|\mathbf{r}| = |\mathbf{r}'|$.
2. The numbers of active strains are equal: $|\mathbb{K}| = |\mathbb{K}'|$.
3. Each active strain has the same number of infectives: $i_{k_a} = i'_{k'_a}$, for $a = 1, \ldots, |\mathbb{K}|$.
4. The numbers of individuals that are susceptible to the ath active strain are equal: $\sum_{k < k_a} s_k = \sum_{k < k'_a} s'_k$, for $a = 1, \ldots, |\mathbb{K}|$.
5. The numbers of removed individuals that will become susceptible to the ath active strain are equal: $\sum_{k < k_a} r_k = \sum_{k < k'_a} r'_k$ for $a = 1, \ldots, |\mathbb{K}|$.

In order to easily refer to the equivalence classes, we define the following representative of each equivalence class.

Definition 3. The *representative* of the equivalence class containing $(\mathbf{s}, \mathbf{i}, \mathbf{r})$, denoted $(\mathbf{s}^*, \mathbf{i}^*, \mathbf{r}^*)$ is defined as follows. If $\mathbb{K} \neq \emptyset$, denote the active strains for the representative by $\mathbb{K}^* = \{1 - |\mathbb{K}|, \ldots, 0\}$. Let $\phi : \mathbb{K} \to \mathbb{K}^*$ be a bijection defined by $\phi(k_a) = a - |\mathbb{K}|$ for $a = 1 \ldots, |\mathbb{K}|$. Then the representative $(\mathbf{s}^*, \mathbf{i}^*, \mathbf{r}^*)$ is given by:

$$
\mathbf{i}_k^* = \begin{cases} \mathbf{i}_{\phi^{-1}(k)} & \text{for } k \in \mathbb{K}^*, \\ 0 & \text{otherwise.} \end{cases}
$$

$$
\mathbf{s}_k^* = \begin{cases} \sum_{j=\phi^{-1}(k)}^{\phi^{-1}(k+1)-1} \mathbf{s}_j & \text{for } k \in \{1 - |\mathbb{K}|, \ldots, -1\}, \\ \sum_{j=\phi^{-1}(0)}^{\infty} \mathbf{s}_j & \text{for } k = 0, \\ \sum_{j=-\infty}^{\phi^{-1}(1-|\mathbb{K}|)-1} \mathbf{s}_j & \text{for } k = -|\mathbb{K}|, \\ 0 & \text{otherwise.} \end{cases}
$$

$$
\mathbf{r}_k^* = \begin{cases} \sum_{j=\phi^{-1}(k)}^{\phi^{-1}(k+1)-1} \mathbf{r}_j & \text{for } k \in \{1 - |\mathbb{K}|, \ldots, -1\}, \\ \sum_{j=\phi^{-1}(0)}^{\infty} \mathbf{r}_j & \text{for } k = 0, \\ \sum_{j=-\infty}^{\phi^{-1}(1-|\mathbb{K}|)-1} \mathbf{r}_j & \text{for } k = -|\mathbb{K}|, \\ 0 & \text{otherwise.} \end{cases}
$$

If $\mathbb{K} = \emptyset$ then $\mathbf{i}_k^* = 0$ for all $k \in \mathbb{Z}$ and $\mathbf{s}_0^* = \sum_{j \in \mathbb{Z}} \mathbf{s}_j$ and $\mathbf{s}_k^* = 0$ for $k \neq 0$, and similarly for $\mathbf{r}$. Let $\{\mathbf{0}\}$ denote the set of all these absorbing states.

In the rest of this paper, the process of representatives on the space of equivalence classes (states described with starred states as in Definition 3) will be referred to as the normalised process, and will be denoted by $(\mathbf{S}^*, \mathbf{I}^*, \mathbf{R}^*)$ or $(\mathbf{S}^*, \mathbf{I}^*)$ as appropriate. Definitions 2 and 3 remove all strains with no infective individuals, and give index 0 to the most recent strain to have emerged and have infectives. All susceptibles and removed individuals are given the strain index one less than the nearest infective above them in strain order. Any individuals immune to all existing strains are given strain 0, as though they just recovered from the most recently emerged strain.

Example 4. Consider the state $(\mathbf{s}, \mathbf{i}, \mathbf{r}) \in \Omega'$ given by

$$(\mathbf{s}_1, \ldots, \mathbf{s}_7) = (0, 0, 1, 1, 0, 0, 1)$$
$$(\mathbf{i}_1, \ldots, \mathbf{i}_7) = (0, 1, 0, 0, 0, 1, 0)$$
$$(\mathbf{r}_1, \ldots, \mathbf{r}_7) = (1, 0, 0, 1, 1, 0, 0)$$

where all remaining terms of $\mathbf{s}, \mathbf{i}, \mathbf{r}$ are zero. Strains 2 and 6 are active so $\mathbb{K} = \{2, 6\} \Rightarrow \mathbb{K}^* = \{-1, 0\}$ and the representative under the equivalence relation is given by

$$(\mathbf{s}_{-2}^*, \ldots, \mathbf{s}_0^*) = (0, 2, 1),$$
$$(\mathbf{i}_{-2}^*, \ldots, \mathbf{i}_0^*) = (0, 1, 1),$$
$$(\mathbf{r}_{-2}^*, \ldots, \mathbf{r}_0^*) = (1, 2, 0).$$

Notation. Recall that without the equivalence relation, the state space of the epidemic process was Ω'. We denote the state space of the normalised process over the set of equivalence class representatives by Ω.

We will use $\mathbf{x} = (\mathbf{s}, \mathbf{i}, \mathbf{r}) \in \Omega$ with, for example, $\mathbf{s} = (\mathbf{s}_k)_k \in \{0, \ldots, N\}^{\mathbb{Z}}$ to denote a typical element of the state space. We will also use, for example, $|\mathbf{s}| = \sum_{k \in \mathbb{Z}} \mathbf{s}_k$ to denote the total number of susceptibles. A random variable written in calligraphic type, e.g. $\mathcal{R}_k(t)$, refers to a process without the equivalence relation. The corresponding variable written in roman type, e.g. $R_k(t)$, refers to the normalised process evolving over representatives from the equivalence classes.

4. Quasi-stationarity and absorbing states

Like many infectious disease models, the E-SIRS and E-SIS models defined in Section 2 have an absorbing class of states that corresponds to the population containing no infected individuals, $\mathbf{i} = \mathbf{0}$. For finite population models, the absorbing state is reached with certainty in finite time, and so the limiting distribution is degenerate with no mass in non-absorbing states. However, like the single strain SIS model, these processes may not go extinct for a long time (individuals can be reinfected) and the transient quasi-stable behaviour is of interest. The quasi-stationary distribution and limiting conditional distribution conditioned on the epidemic not going extinct, represent the long-term average behaviour for an endemic disease.

4.1. Properties of quasi-stationary distributions for epidemics

In this section and the rest of this paper, $\mathbb{P}_{\mathbf{u}}[A] = \mathbb{P}[A | X(0) \sim \mathbf{u}]$.

Definition 5. Let $X = (X(t))_{t \geq 0}$ be a Markov process on a countable state space Ω with absorbing state 0 from which it cannot escape. Then a distribution $\mathbf{u}$ on $\Omega \setminus \{0\}$ is a *quasi-stationary distribution* (QSD) if $\mathbb{P}_{\mathbf{u}}[X(t) \in A | X(t) \neq 0] = \mathbf{u}(A)$ for all $t \geq 0$.

Given initial condition $\mathbf{v}$ on $S = \Omega \setminus \{0\}$, $\mathbf{u}$ is a $\mathbf{v}$-*limiting conditional distribution* (LCD) if $\lim_{t \to \infty} \mathbb{P}_{\mathbf{v}}[X(t) \in A | X(t) \neq 0] = \mathbf{u}(A)$. Note that, for processes where S is a single communicating class, every QSD $\mathbf{u}$ is a $\mathbf{u}$-LCD and every LCD is a QSD.

Related to the QSD on irreducible state spaces is the notion of the decay parameter which describes the rate of decay of the transition probabilities.

Definition 6. Let $X = (X(t))_{t \geq 0}$ be an irreducible Markov process on a countable state space Ω with absorbing state 0. Let $\mathbf{u}$ be a QSD associated to X. Then the *decay parameter* α is given by

$$\alpha = \inf\{a \geq 0 : P_{ij}(t) = o(e^{-at})\}.$$

for $i, j \in \Omega \setminus \{0\}$ The *absorption parameter* α_0 is given by

$$\alpha_0 = \inf\left\{a \geq 0 : \int_0^{\infty} \mathbb{P}_i[T > t] e^{at} dt = \infty\right\},$$

for $i \in \Omega \setminus \{0\}$ where T is the extinction time of X starting from state i. Note that for irreducible processes, α is independent of i, j and α_0 is independent of i.

According to Theorem 6 of [15], a necessary condition for the existence of a QSD is that $\alpha > 0$.

Theorem 7. *Conditional on non-absorption, the following hold.*

1. *The QSD for the number of infectives in the single-strain SIS model exists uniquely and gives weight to all states $\{1, \ldots, N\}$.*
2. *The QSD for the number of susceptibles and infectives in the single-strain SIR model exists uniquely and gives full weight to the state $\{(S, I) = (0, 1)\}$.*
3. *The QSD for the number of susceptibles and infectives in the single-strain SIRS model exists uniquely, and gives weight to all non-absorbing states.*

Proof. Theorem 1 of [16] states that QSDs exist and are unique on finite irreducible state spaces, and so there is a unique QSD for the SIS model and for the SIRS model conditional on $\{I > 0\}$, and non-zero weight is given to all non-absorbing states. For reducible processes, Theorem 8 of [16] states that QSDs will give full weight to the communicating class with the longest expect time to leave and any states accessible from this "slowest" communicating class. This characterises the QSD for the SIR model. □

Further work on characterising the QSD for the standard SIS model can be seen in [17–19] making use of recurrent processes and normal approximations.
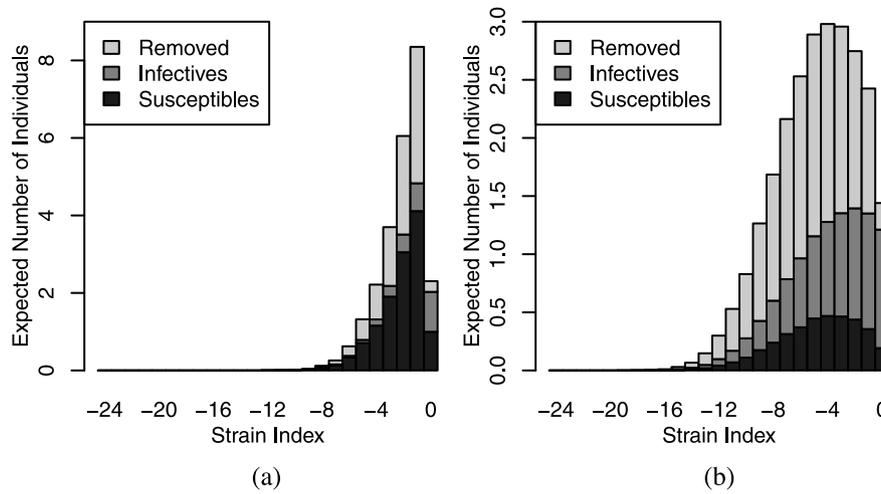
Fig. 1. Comparisons of expected population composition under E-SIRS QSDs with (a) $\beta = 2$, $\theta = 0.9$, $\delta = 0.2$, $N = 25$; (b) $\beta = 2$, $\theta = 0.9$, $\delta = 2$, $N = 25$.

4.2. Existence and uniqueness

Here we will summarise the existence and uniqueness results for the E-SIS and E-SIRS processes.

Theorem 8. *Let the E-SIS model be defined as in Sub Section* 2.2 *with parameters* $\beta, \gamma > 0$ *and* $\theta \in (0, 1]$. *Then for the normalised process, conditional on the events* $\{I(t) > 0\}$, *there exists a unique QSD which equals the unique LCD of the process and gives weight to all non-absorbing (i.e. transient) states,* $\{(\mathbf{s}, \mathbf{i}) \in \Omega : |\mathbf{i}| > 0\}$. *If* $\theta = 0$ *and the process begins with a single infective, then there exists a unique LCD which gives full weight to the state with a single infective with strain index* 0, *and* $N - 1$ *susceptibles with strain index 0.*

Proof. For $\theta \in (0, 1]$, we obtain existence and uniqueness by proving that $S = \Omega \setminus \{\mathbf{0}\}$ is a single finite communicating class, which immediately gives existence, uniqueness and equality of the QSD and LCD by Theorem 3 from [15].

Under the equivalence relation, there can be at most N different strain indices present in the population. This implies that every individual must appear in one of the states $s_{1-N}, \ldots s_0$ or $i_{1-N}, \ldots, i_0$. Therefore we can bound above the size of S, the set of transient (i.e. non-absorbing) states, by $(2N)^N$.

One can see that the transient states form a single communicating class by noting that one can get from a single infective of strain index 0 with $N - 1$ susceptibles of index 0 to any other state through infections (mutation and non-mutation) and recoveries. If all individuals are infected and then all but one recovers, then the process returns to the single infective case mentioned above.

For $\theta = 0$ we consider the E-SIS model starting with a single infective of strain 0 and susceptibles of strain index -1. If $\theta = 0$, then mutation is impossible. As a result, once an individual has become infected and recovered, they join the s_0 class and cannot be reinfected. In this way $\{S_0(t)\}$ behaves identically to the $\{R(t)\}$ class in the SIR model, and we identify the two models in this way. Point 2 in Theorem 7 then gives the required LCD. □

Theorem 9. *Let the E-SIRS model be defined as in Sub Section* 2.1 *with parameters* $\beta, \gamma, \delta > 0$ *and* $\theta \in (0, 1]$. *Then, conditional on having at least one infective, there exists a unique QSD. If* $\theta = 0$ *and there is initially one infective, a QSD still exists and gives full weight to the state with one infective with strain index* 0 *and* $N - 1$ *susceptibles with strain index 0.*

Proof. For $\theta \in (0, 1]$, one follows the same argument as in Theorem 8, this time bounding the size of the state space by $(3N)^N$, since individuals may also reside in classes $r_{1-N}, \ldots, r_0$. The fact that

the transient states form a single communicating class also follows as in Theorem 8. For $\theta = 0$, we see that each state that can be reached is a transient communicating class; there is no way to return to a state once left. As such, we need to consider the decay parameter on leaving each state, which equals the exponential rate of leaving such a state: $\beta s_{-1} i_0 / N + \gamma i_0 + \delta r_0$. The decay parameter for the process is therefore the minimal such value across all non-absorbing states. Therefore $s_{-1} = 0$, $i_0 = 1$ and $r_0 = 0$ minimise the decay parameter. According to Theorem 1 of [16] this forces the QSD to have full mass on this state where $i_0 = 1, s_0 = N - 1$, since the only state accessible from this state is an absorbing one. □

4.3. Sampling the quasi-stationary distribution

Samples from quasi-stationary distributions can be produced using the sequential Monte Carlo (SMC) sampler with stopping time resampling methods developed in [20]. In brief, M realisations of the model (referred to as particles) are simulated forward in time. Absorbed particles (with no infected individuals) are given weight zero and non-absorbed particles are given weight 1 initially. The distribution of weights converges to the limiting conditional distribution. Once the total weight drops below a proportion λ of the initial weight, the particles are replenished via a resampling step. Combine-split resampling [20] was used, which prevents any occupied states from being lost and has the advantage that the distribution of weights remains unchanged after resampling. This resampling method combines particles in the same location together, draws new particle locations uniformly from the existing locations and equalises the weight on particles in the same location. In our implementation, after a burn-in of $T_b = 1$, weighted samples were drawn every $T_d = 1$ time units until time $T_{max} = 100$.

Fig. 1 shows the expected number of individuals in each class under the QSD. In this example we used $M = 1000$ particles and a resampling threshold of $\lambda = 0.6$. The Figure shows that when δ is smaller there is a larger proportion of globally immune individuals in the removed classes, and so the population can support fewer strains.

5. Limiting behaviour

One aspect of interest is how the E-SIS and E-SIRS processes relate to those without mutation. To this end, we consider the limits of the times to extinction of the processes as θ tends to 0 or 1, and the limit, for fixed θ, of the time to extinction as the population size N tends to infinity. Infinite population models can be used to provide approximations for quantities which we cannot obtain analytically for finite population models, such as the decay parameter.

5.1. Limits as mutation probability changes

Theorem 10. *Let T^θ be the time to extinction of the E-SIS model, and T^1 the time to extinction of the standard SIS model, both starting from a single infective (nominally of strain index 1) in a population of N individuals. Then $T^\theta \to T^1$ in distribution as $\theta \to 1$.*

Proof. We make use of a coupling of $(T^\theta : 0 < \theta < 1)$ and T^1. Firstly, let $\mathbf{X}^\theta(t) = (\mathbf{S}^\theta(t), \mathbf{I}^\theta(t))$ be the E-SIS model. We assume the process to be defined over a population indexed by $n = 1, \dots, N$.

- For each individual n, define a sequence of i.i.d. infectious periods $\{L_m^{(n)} \sim \text{Exp}(\gamma) : m \in \mathbb{N}\}$.
- For each *ordered* pair of individuals (n, n'), define a homogeneous Poisson process $\{A^{(n,n')}(t)\}$ on $[0, \infty)$ with rate β/N.
- For each ordered pair (n, n') define a sequence of uniform random variables $U_l^{(n,n')} \sim \text{Unif}[0, 1]$ for $l \in \mathbb{N}$.
- Let all $L_m^{(n)}$, $A^{(n,n')}$ and $U_l^{(n,n')}$ be independent.

Now let $(E, \mathcal{F}, \mathbb{P})$ be the product space of these random processes and variables. Using these building blocks, we construct an E-SIS model $\{\mathbf{X}^\theta(t)\}$ and SIS model $\{\mathbf{Y}(t) = (S(t), I(t))\}$ as follows. Set $S_{-1}^\theta(0) = N - 1$, $I_0^\theta(0) = 1$ for the E-SIS model and set $S(0) = N - 1$, and $I(0) = 1$ for $\mathbf{Y}(0)$. Assume the initial infective individual has index $n = 1$ without loss of generality. In both processes infectious individual n generates contacts with each susceptible individual n' at points of the Poisson process $\{A^{(n,n')}(u)\}$, where u denotes the cumulative time that individual n has been infectious and n' has been susceptible. In other words the Poisson processes are stopped whenever it is not possible for individual n to infect individual n'. At each contact event in the SIS model an infection occurs. The newly infected individual n' stays infected for a period of length $L_{m(n',t)+1}^{(n')}$, where $m(n', t)$ is the number of infections individual n' has recovered from up to the contact time t. In the E-SIS model the ith contact event in $\{A^{(n,n')}(u)\}$ results in a mutation if and only if $U_i^{(n,n')} \leq \theta$, in which case individual n' is infected with a new strain and given the lowest unused strain index. However if $U_i^{(n,n')} > \theta$ then individual n attempts to infect n' with their current strain and the infection is successful if the strain index of individual n' is strictly less than the stain index of n. As in the SIS model, successful infections in the E-SIS model are given infectious period length $L_{m(n',t)+1}^{(n')}$. Notice that under this coupling non-mutation contacts of n' by n are only successful if the strain index of n is strictly greater than that of n' in the E-SIS model. However mutation contacts and all contacts in the SIS model are always successful.

Fix $\omega \in E$, our probability space. For the SIS model, we have $T^1 < \infty$ almost surely. On the interval $[0, T^1(\omega))$, there are two possibilities for the E-SIS model. At each infective–susceptible contact we compare the strain indices. The first possibility is that every contact will always lead to a successful infection, arising from a sequence of infection events which always contact a susceptible of a lower index. In this case, we have $T^\theta(\omega) = T^1(\omega)$ for all $\theta \in [0, 1]$. The second possibility is one or more "potentially unsuccessful" contact events exist, in which if the event were to be non-mutation, it would fail. This failure occurs if the relevant $U_i^{(n,n')} > \theta$. Since we must have a finite number of such events occurring in $[0, T^1)$, we can find θ^* such that $U_i^{(n,n')} \leq \theta^*$ for all such $U_i^{(n,n')}$ corresponding to potentially unsuccessful events. This means that for $\theta \geq \theta^*$ we must have $T^{\theta^*}(\omega) = T^1(\omega)$. So for every $\omega \in E$, there exists $\theta^* \in (0, 1)$ such that $T^\theta(\omega) = T^1(\omega)$ for all $\theta \geq \theta^*$. Hence $T^\theta(\omega) \to T^1(\omega)$ as $\theta \to 1$ for almost every $\omega \in E$, and hence $T^\theta \to T^1$ in distribution by the Skorohod Dudley theorem from, for example, Theorem 3 of [21]. $\square$

Theorem 11. *Let T^0 be the time to extinction of the standard SIR model. Then $T^\theta \to T^0$ in distribution as $\theta \to 0$.*

Intuitively, one can think of identifying the S_1-class for the E-SIS model and the R-class of the SIR model. As mutation events get rarer, the chance of mutation happening before extinction becomes smaller and smaller, and so the two processes are more likely to coincide under a suitable coupling until extinction.

Proof. This proof follows in a similar fashion to Theorem 10. In this version, the coupling is constructed between the E-SIS and the SIR model. The only differences are that in the SIR model individuals enter the removed category after their infectious period and the Poisson process $\{A^{(n,n')}(u)\}$ progresses during any time for which n is infective and n' is susceptible in the E-SIS model (as before); but when n' is susceptible *or removed* in the SIR model.

In the E-SIS model infectious contacts between n and n' are only successful if the event is a mutation or n' is of a strictly lower strain index than n. In the SIR model, only the first infectious contact is successful. This means that the two epidemics must be identical up to the time of the first repeat contact, when one identifies the $\{S_1, S_2, S_3, \dots\}$ classes in the E-SIS model with the R class of the SIR model.

Similar to the proof of Theorem 10, for each $\omega \in E$ one can find a value of θ^* so that $T^{\theta^*}(\omega) = T^0(\omega)$ for all $\theta \leq \theta^*$, and so $T^\theta \to T^0$ almost surely as $\theta \to 0$ and hence $T^\theta \to T^1$ in distribution by the Skorohod Dudley theorem of [21]. $\square$

5.2. Large population limits

In order to obtain some large population limit results, we will consider an "infinite" population model. We will refer to this as an *Evolving Birth–Death Process* (E-BDP). More precisely we assume that infected individuals are a negligible part of an infinite population of individuals that are not immune to any strains at the start of the epidemic, and so all infections will be successful almost surely. This implies infections from a given strain k and recoveries from that strain behave as a linear birth–death process with birth rate β and death rate γ. Additionally, at the point of each infection, with probability $\theta \in [0, 1]$, the new infective is infected with a previously unseen strain, and given the next available strain index $K^* + 1$.

The possible events comprise:

- *Infection with mutation:* $\mathcal{I}_{K^*+1}(t) = 0 \mapsto \mathcal{I}_{K^*+1}(t) = 1$ with rate $\beta\theta\mathcal{I}(t)$.
- *Infection without mutation:* $\mathcal{I}_k(t) \mapsto \mathcal{I}_k(t) + 1$ for $k \in \mathbb{Z}$ with rate $\beta(1-\theta)\mathcal{I}_k(t)$.
- *Recovery:* $\mathcal{I}_k(t) \mapsto \mathcal{I}_k(t) - 1$ for $k \in \mathbb{Z}$ with rate $\gamma\mathcal{I}_k(t)$.

After it emerges, each strain behaves according to a linear birth–death process with birth rate $\beta(1-\theta)$ and death rate γ. The total number of infectives $\mathcal{I}(t)$ also behaves according a birth–death process with birth rate β and death rate γ.

The time to extinction of the E-SIS model converges to that of the E-BDP model, noting that under a suitable coupling, the time to extinction of the E-BDP equals the Linear BDP without mutation. This leads us to the following result.

Theorem 12. *Let $T^{\theta,N}$ be the time to extinction of the E-SIS model, and T the time to extinction of the E-BDP model. Then we have $T^{\theta,N} \to T$ in distribution as $N \to \infty$ when $\beta < \gamma$. If $\beta \geq \gamma$, then on the event $\{T < \infty\}$, a region of probability $1 - \gamma/\beta$, we also have $T^{\theta,N} \to T$ in distribution as $N \to \infty$*

Proof. Using Theorems 10 and 11 we can conclude that for any fixed N that $T^{0,N}$ is the time to extinction for the standard SIR model, and $T^{1,N}$ is equal to the time to extinction for the standard SIS epidemic model. Furthermore, from these theorems we can construct a coupling of the SIS, E-SIS and SIR models using two sets of Poisson processes and mutation indicator variables such that, for any $\theta \in [0, 1]$,

$$T^{0,N}(\omega) \leq T^{\theta,N}(\omega) \leq T^{1,N}(\omega) \tag{1}$$

for almost every $\omega \in E$. From [22], we know that if $\beta < \gamma$ then $T^{0,N}$ converges in distribution to T, the time to extinction of a Linear BDP with the same parameters β and γ. From [23] we obtain that the same thing happens for SIS models, i.e. $T^{1,N} \to T$ in distribution as $N \to \infty$. Using the bounds in Eq. (1), we obtain that $T^{\theta,N} \to T$ as $N \to \infty$ for all $\theta \in [0, 1]$.

In the case where $\beta \geq \gamma$ we note that on a set of probability $1 - \gamma/\beta$, the time to extinction of the linear BDP is infinite, as discussed in Chapter 3.2 of [24]. From [23], we know that $T^{1,N} \to T$ almost surely (and hence in distribution) on the event $\{T < \infty\}$. From [22] we know that on this event, $T^{0,N} \to T$ in distribution. Therefore we must have that $T^{\theta,N} \to T$ as $N \to \infty$ for all $\theta \in [0, 1]$ here too. $\quad\square$

It should be noted, that on the event $\{T = \infty\}$ we do not have $T^{0,N} \to \infty$. Instead a suitably rescaled version of $T^{0,N}$ converges to an extreme-value distribution as mentioned in Theorem 8.1 [14].

Next we show existence of a QSD for the E-BDP model.

Theorem 13. *Let* $\mathbf{X}(t)$ *be the E-BDP with parameters* $\gamma > \beta > 0$ *and* $\theta \in [0, 1]$. *Then, under the equivalence relation described in Section* 3 *and conditional on the event* $\{I(t) > 0\}$, *there exists a unique QSD.*

Proof. To prove existence, we first show that the state space we are interested in is countable. To do this we use the following construction. Starting with a single infective of strain 0, we can define a method of constructing the state space. By having a birth in strain 0, or a mutation event, one can systematically arrive at any state in the state space. Given these two possible events, one can encode each state according to a finite binary sequence, which corresponds to a unique integer which we can use to enumerate the space. Given that there exists a lower bound $l \in \mathbb{Z}$ such that $I_k = 0$ for all $k \leq l$, we construct the state as follows. Starting with the lowest non-zero strain index $l + 1$ consider I_{l+1} strain 0 within-strain-infection events. Then for each higher strain k, we choose a mutation event followed by $I_k - 1$ within-strain infection events. Note that only considering finite sequences gives countability, unlike the uncountability of the infinite paths on this binary tree.

To obtain existence of a QSD, we now introduce a coupling. Let $\mathbf{X}(t) = (X_j(t))_{j \in \mathbb{Z}}$ be the E-BDP. Let α^X be the decay parameter for $\mathbf{X}(t)$. Let $(Y(t))_{t \geq 0}$ be the process defined on the same probability space, given by $Y(t) = \sum_{j \in Z} X_j(t)$. Since the mutations do not affect whether infections are successful or not, $Y(t)$ can be seen to be a single-strain linear BDP with birth rate β, and death rate γ. As discussed in Example 1 of [25], $Y(t)$ has the decay parameter $\alpha^Y = \gamma - \beta$. Let T_X be the extinction time of $\mathbf{X}(t)$ and T_Y for $Y(t)$.

Letting α_0^Y be the absorption parameter for $Y(t)$, and α_0^X for $\mathbf{X}(t)$, we also know that $\alpha_0^Y = \gamma - \beta$. Since $T_X = T_Y$ under the coupling, we use the definition of the decay parameter to deduce that $\alpha_0^X = \gamma - \beta$, and hence $\alpha^X \geq \alpha_0^X > 0$. Using Theorem 13 of [15] we get existence of a QSD. Moreover, using Theorem 3.3.2 of [26], we must have $\alpha_X = \gamma - \beta$ since there is only one state from which extinction can occur: one must have 1 infective before extinction, which must be of strain 0 under the equivalence relation. This leads to the uniqueness of the QSD. $\quad\square$

6. Reproduction numbers

To characterise the dynamics of the models, we look to a number of key statistics which are related to the commonly used *basic reproduction number*, R_0, that illustrates whether or not an epidemic is likely to infect a large proportion of the population. The basic reproduction number is defined as the number of individuals infected by a single typical infective in a large, otherwise susceptible population [27]. In the E-SIRS model, we still have $R_0 = \beta/\gamma$. One issue with R_0 is that it fails to take into account the likely immunities present in the population, or how much the pathogen evolves during the opening phase of the epidemic.

6.1. Modified household reproduction number R_*

In [28], an epidemic is considered which spreads through a population grouped into households, such that individuals in the same household make contact at a different rate to individuals in different households. The households reproduction number R_* is shown in [28, Section 2.3] to be equal to $R_* = \mu R_H$ where μ is the expected number of individuals infected in a single household epidemic (including the initial infective), and R_H is the mean number of contacts an infective individual makes with individuals in other households during a single infectious period.

For the E-SIRS model, we consider each strain as a "household" which has countably many individuals, and mutations are considered contacts between households. In this case μ is the expected total population of a birth–death process with birth rate $\beta(1 - \theta)$ and death rate γ, including the initial infective. One can use the branching property to compute $\mu = \gamma/(\gamma - \beta(1 - \theta))$ and note that the between household reproduction rate $R_H = \beta\theta/\gamma$ and so,

$$R_* = \begin{cases} \frac{\beta\theta}{\gamma - \beta(1-\theta)} & \beta(1-\theta) < \gamma \\ \infty & \beta(1-\theta) \geq \gamma. \end{cases}$$

To recontextualise this in terms of strains and mutations, one can think of R_H as the expected number of new strains originating from a single individual during one infectious period, and μ as the expected number of individuals that ever get infected by a specific strain.

One could consider R_0 to be the "intra-strain" reproduction number, and μ to be the "inter-strain" reproduction number. With these we obtain one of three regimes:

- If $R_0 = \beta/\gamma < 1$, then the whole population would die out with certainty, and no large epidemic would occur.
- If $R_0 \geq 1$ and $\mu < \infty$ then a large epidemic occurs with positive probability, but each individual strain dies out quickly.
- If $R_0 \geq 1$ and $\mu = \infty$, then each strain has a positive probability of producing a large outbreak.

This can be compared to, for example, the multi-type epidemics of [29, 30], where epidemics in specific groups can die out but the overall epidemic survives, or where the survival of an epidemic somewhat depends on a large epidemic occurring within a single group. Fig. 2 shows realisations of the genetic trees under the E-SIS model under the two supercritical regimes. For small θ, we obtain only a small number of strains, and the epidemic is more likely to die out. Moreover, in a finite population, this low θ leads to high immunity in the population and hence shorter epidemics. The trees highlight how only a small number of the strains survive for a long time, particularly in Fig. 2(a). Fig. 2 also show similarities to the tree for H3N2 in Extended Data Figure 9(c) of [12], a paper which specifically looks to model influenza.

6.2. Quasi-stationary reproduction number R_Q

One drawback to R_0 is that it only usefully describes the initial behaviour of an epidemic in a naive population and does not take into account the build up of immunity in the E-SIRS model. One alternative is to consider the *effective reproduction number*, denoted R_t, defined as $R_t = R_0 \frac{S(t)}{N}$ in a population of size N. Much work has been done in trying to evaluate R_t for specific infections such as influenza by [31] and Ebola by [32]. However, R_t is time-dependent and can therefore be difficult to compute and interpret. At a quasi-stable equilibrium the number of new infections balances the recoveries and so $R_t \approx 1$, and hence R_t is not informative about the disease characteristics. Ideally, we would like a reproduction number that adjusts for the build-up of immunity in the population, but remains informative about the infectivity of a disease.

We offer an alternative reproduction number, based on the QSD, which aims to describe the infectiousness of strains of an endemic disease in a population with 'average' levels of historical immunity. The
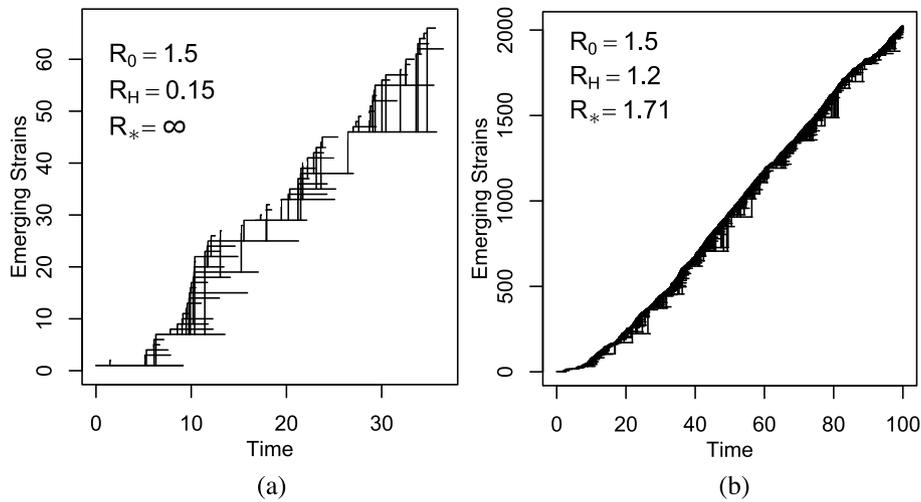
Fig. 2. Comparisons of emergence of strains under different R_0, R_H under the E-SIS model with $N = 25$, $\gamma = 1$, $\beta = 1.5$ with (a) $\theta = 0.1$; (b) $\theta = 0.8$.

quasi-stationary reproduction number (R_Q) is the average number of secondary infections caused by a single typical infective introduced into an otherwise uninfected (S status) population with levels of immunity (strain indexes) drawn from the quasi-stationary distribution, so each other individual may or may not be immune to the current strain of the infective. By typical infective, we mean an individual with strain index sampled from the distribution of strain indexes of infectives in the QSD. Under the E-SIRS model, the total number of infectives is always less than the SIS model without evolving strains, and so $R_Q \leq R_0$.

The quasi-stationary reproduction number provides a measure of the ability of a pathogen to re-invade a population from which it has been eradicated. For diseases like seasonal influenza which have greatly reduced incidence during the summer months, R_Q measures the reproduction number at the beginning of the next influenza season after accounting for the residual immunity left over from last year.

More precisely, we draw the single infective from the marginal number of infectives in the QSD $u_I(k)$: the probability that given an individual is infective, it is of strain index k. For QSD **u** this is given by

$$u_I(k) = \sum_{(\mathbf{s,i,r}) \in \Omega} u_{(\mathbf{s,i,r})} \frac{\mathbf{i}_k}{|\mathbf{i}|}.$$

Under the equivalence relation described in Section 3, we can have a maximum of N strains in a population of size N, and so the strain index ranges over $k \in \mathbb{K}^* = \{0, 1 - N\}$. The susceptible population is drawn from the total strain marginals of the QSD $u_K(k)$: the probability under the QSD that a given individual is of strain index k.

$$u_K(k) = \sum_{(\mathbf{s,i,r}) \in \Omega} u_{(\mathbf{s,i,r})} \frac{\mathbf{i}_k + \mathbf{r}_k + \mathbf{s}_k}{N}$$

Finally, we require the probability that a randomly chosen individual drawn from the strain marginal will be susceptible to strain k (i.e. will have a strain index lower than k):

$$u_L(k) = \sum_{j=1-N}^{k-1} u_K(j).$$

During their infectious period the infective makes infectious contact with each individual at the points of a Poisson process with rate β/N. For large populations the infective is unlikely to contact the same individual twice (or themselves), and so the expected number of contacts is β/γ. With probability θ the contacts are mutations and are successful infections. With probability $(1 - \theta)$ the contacts are non-mutations and are only successful if the individual contacted has a lower strain index, which occurs with probability $u_L(k)$ when the infective has strain index k. To calculate R_Q we condition on the strain

index of the initial infective, hence

$$R_Q = \frac{\beta}{\gamma} \sum_{k=1-N}^{0} (\theta + (1 - \theta)u_L(k))u_I(k)$$
$$= \frac{\beta}{\gamma}(\theta + (1 - \theta)\mathbf{u}_L^T \mathbf{u}_I). \tag{2}$$

Since u_I and u_L are both probability mass functions, $0 \leq \mathbf{u}_L^T \mathbf{u}_I \leq 1$ and so we have that $R_H \leq R_Q \leq R_0$. As $\theta \to 1$ then $R_Q \to \beta/\gamma = R_0$, as does R_H.

The three notions of a reproduction number in this section describe three different facets of the epidemic model, and can be compared in Fig. 3. It shows that R_Q is always less than R_0 due to the effects of immunity, and R_* depends greatly on θ; the unplotted points for R_* are in the regions where it is infinite, namely where $\beta(1 - \theta) \geq \gamma$. Values of R_Q were calculated using the SMC sampler described in Section 4.3 with $M = 100$ particles and a resampling threshold of $\lambda = 0.4$.

7. Simulation study

To further explore the E-SIRS model we use the SMC sampler described in Section 4.3 to investigate numerically features of the QSD which we cannot obtain analytically. We wish to observe how various key properties behave as we vary parameters of the model. To this end we look at the following expectations over the QSD. For brevity we omit time indices and conditioning, and denote expectations under the QSD by $\mathbb{E}_Q$.

- The expected total number of infectives $\mathbb{E}_Q[I]$ and immune individuals $\mathbb{E}_Q[R]$ in the QSD, where $I = \sum_{k=-\infty}^{0} I_k$, $R = \sum_{k=-\infty}^{0} R_k$.
- The expected total number of active strains $\mathbb{E}_Q[K]$ in the QSD where $K = |\{k : I_k > 0\}|$.
- We also look at how varying the model parameters affects strain diversity in infectives and the whole population.

We will focus on the E-SIS model, but also discuss for each statistic how the addition of an immune period, as in the E-SIRS models, changes the number of infectives and strain diversity. Unless otherwise stated, all expectations over the QSD were produced with the SMC sampler described in 4.3 with $M = 100$ particles and resampling threshold $\lambda = 0.4$.

7.1. Expected number of infectives

Fig. 4 shows a heatmap of the expected number of infectives in the population under quasi-stationarity, $\mathbb{E}_Q[I]$, and how this depends on
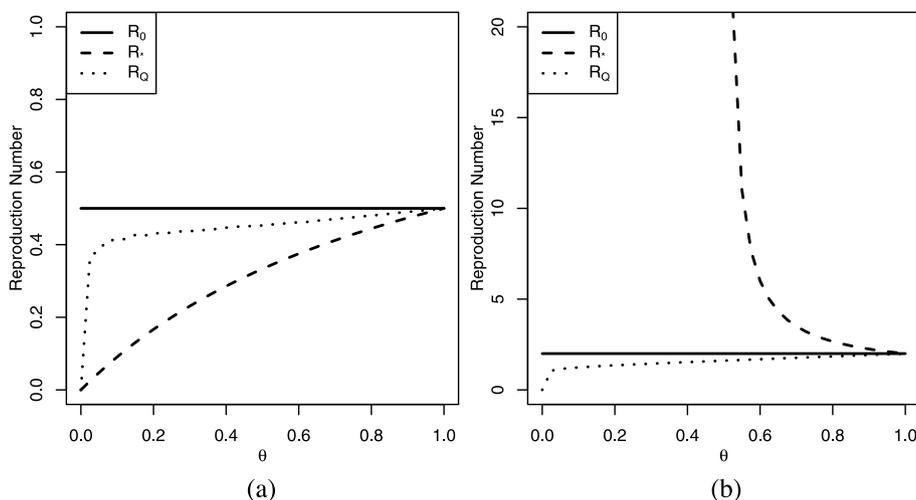
Fig. 3. Comparisons of R_0, R_* and R_Q under varying θ. with (a) $\beta = 0.5$, $\gamma = 1$, $N = 100$; (b) $\beta = 2$, $\gamma = 1$, $N = 100$.
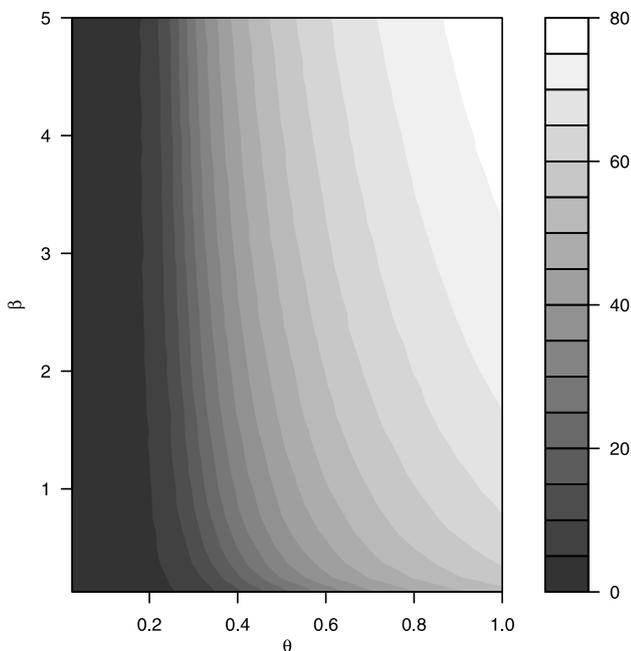
Fig. 4. Expected number of infectives as β and θ change in E-SIS with $\gamma = 1$, $N = 100$.

the contact rate and the probability of mutation. Increasing the contact rate β or mutation probability θ increases the expected number of infectives. However, for a fixed population size (in this case $N = 100$), the number of infectives increases linearly in β when $\mathbb{E}_Q[I]$ is much smaller than N. This can be observed in Fig. 5(a), which shows that the number of infectives grows more slowly as β increases, especially when θ is small and so the probability of failed infections is high. As was also noted in Section 4.3, increased levels of global immunity result in fewer infectives under quasi-stationarity, due to the increased possibility of failed infections.

Fig. 5(b) shows that as N increases the expected proportion of infectives ($\mathbb{E}_Q[I/N]$) decreases in the case where $\beta < \gamma$, whereas in the supercritical case we see that $\mathbb{E}_Q[I/N]$ remains fairly constant. In the E-SIRS model, we see that $\mathbb{E}_Q[I/N]$ is decreased by the introduction of transient global immunity. Furthermore, as δ gets smaller the transient immunity lasts longer and $\mathbb{E}_Q[I/N]$ further decreases.

7.2. Expected number of strains

We investigated what happens to the expected number of active strains, $\mathbb{E}_Q[K]$ (strains held by infectives) as the parameters change. Under our models, the number of strains is always less than the number of infectives due to the absence of super-infectivity (infection of an individual by multiple strains during a single infectious period). As such, much of the behaviour is similar to that of the expected number of infectives in the previous subsection. For example, the expected number of strains increases linearly with β when $\mathbb{E}_Q[K]$ is much less than N. This follows since we already know that for $\theta = 1$ every infective begins a new strain and so $\mathbb{E}_Q[K] = \mathbb{E}_Q[I]$. Like in single-strain SIS and SIR models with low R_0, the number of infectious contacts behaves like a birth–death process and so increases linearly in β. When $\mathbb{E}_Q[K]$ is closer to N, the chance of having an unsuccessful contact is higher due to the immunity induced by competing strains. At the other end of the scale, we automatically have that $\mathbb{E}_Q[K] = 1$ if $\theta = 0$.

Fig. 6 shows the expected number of strains for fixed $\beta\theta$ (mutation contact rate) and $\beta(1 - \theta)$ (non-mutation contact rate), as β and θ vary. Note that for Fig. 6(b), both θ and β increase from left to right, whereas, to maintain fixed $\beta\theta$, β decreases as θ increases. In the case when $\beta\theta$ is high, one might expect $\mathbb{E}_Q[I]$ and $\mathbb{E}_Q[K]$ to be closer in value, since there is a high probability of mutation leading to a high number of co-circulating strains. This is demonstrated in Fig. 6(a), where we see that for fixed $\beta\theta$, the number of strains is larger (and therefore closer to the number of infectives) for the $\beta\theta = 2$ line than for the $\beta\theta = 0.05$ line. In Fig. 6(a) there is a maximum point for the number of strains as β increases, after which the number of strains decreases. As $\beta(1 - \theta)$ increases in Fig. 6(b), the number of strains becomes more linear in θ.

7.3. Strain diversity

In Fig. 7 we investigate the distribution of immunity across the active strains. The figure shows the expected proportion of infectives $\mathbb{E}_Q[I/N]$ and total individuals for each strain index $\mathbb{E}_Q[I_k + S_k + R_k]$, relative to the most recently emerged strain. The expectations taken over the quasi-stationary distribution were calculated using the SMC sampler described in Section 4.3, with $M = 100$ particles, resampling threshold $\lambda = 0.4$ and burn-in $T_b = 11$.

Fig. 7(a) illustrates that larger values of β greatly increase the diversity of strains in the infectives and the variation in immunity in the population, since there are more infectives and so more chances for mutation contacts. Another point of interest is the lag of the strain diversity: the number of strains between the mode of the infective strains and the mode of the total population. The lag is fairly consistent
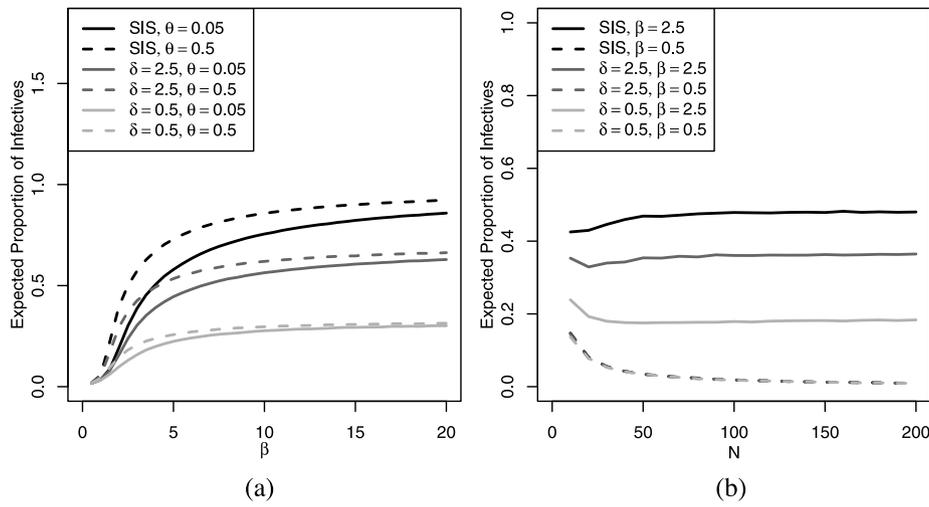
Fig. 5. Expected proportion of infectives in the E-SIS and E-SIRS models: (a) as β varies with $\gamma = 1$, $N = 100$; (b) as N varies with $\gamma = 1$, $\theta = 0.4$.
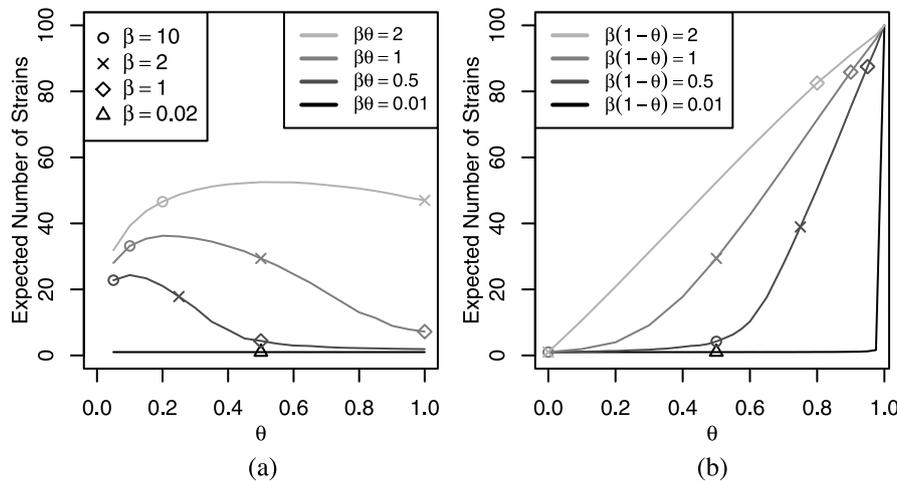
Fig. 6. Expected number of strains under quasi-stationarity for (a) fixed $\beta\theta$; (b) fixed $\beta(1-\theta)$; with $\gamma = 1$ and $N = 100$.

for the different values of β, but does increase slowly in β. Fig. 7(b) shows the change in strain diversity in θ. As θ increases, the number of strains present increases, so the strain diversity curve flattens out. For high values of θ, a larger lag is observed between the infectives and the whole population, due to the higher diversity. In Fig. 7(c), the effect of population size is explored. As N increases, we observe a wider number of strains, as one would expect given $\mathbb{E}_Q[K]$'s behaviour. However, unlike the behaviour as β changes, the peak moves away from 0 but the lag between the infectives and the rest of the population appears more consistent. For the E-SIRS model explored in Fig. 7(d), the immune period reduces strain diversity by reducing the expected number of infectives.

In applications, one might wish to look further into the lag between the strain distribution of the infectives and the immunity in the population. If a pathogen has a long lag, then vaccination can be effective in updating the immunity present in the population. However, if the lag is short, then a vaccine based on a recent strain will have little effect in increasing the levels of immunity in the population, as the most immunity profiles in the population will already represent the currently circulating pathogen.

8. Conclusions

In this paper we defined an epidemic model for a pathogen undergoing genetic drift, that lies between the well-studied SIR and SIS epidemic models. The model appears to capture some qualitative aspects of the evolution of strains of influenza A, despite depending on just 4 parameters. Compared to models used by [12] and [33], which require the storage of a whole antigenic history, our model is much simpler, which makes simulation, computation and inference much easier. Despite these simplifications, the simulated genetic trees in Fig. 2 show similarities to the tree for H3N2 in Extended Data Figure 9(c) of [12]. It should be noted that the population is much smaller for the demonstration, and runs over a short time period. This is due to computational limitations with regards to finding an appropriate simulation. However, the overall genetic history of the two pathogens is noticeably similar. The model also allows investigation of different regimes where, for example, the overall epidemic continues while a single strain is only short-lived. The relative simplicity of our model enables analytical insights into model behaviour, such as the relationship between our models and the SIS and SIR models discussed in Theorems 10 and 11. A simulation study showed that there is a non-linear trade-off between mutation and infectivity when trying to estimate the number of co-circulating strains under quasi-stationarity. The development of a novel quasi-stationary reproduction number R_Q allows us to quantify the ability of a pathogen to re-invade a population from which it has been eradicated, for example at the start of the next epidemic season. This work could be further developed to include more complex evolutionary models to describe the emergence of new strains; to account for partial immunity based on the similarity between strains; and to incorporate seasonal changes in infectivity.
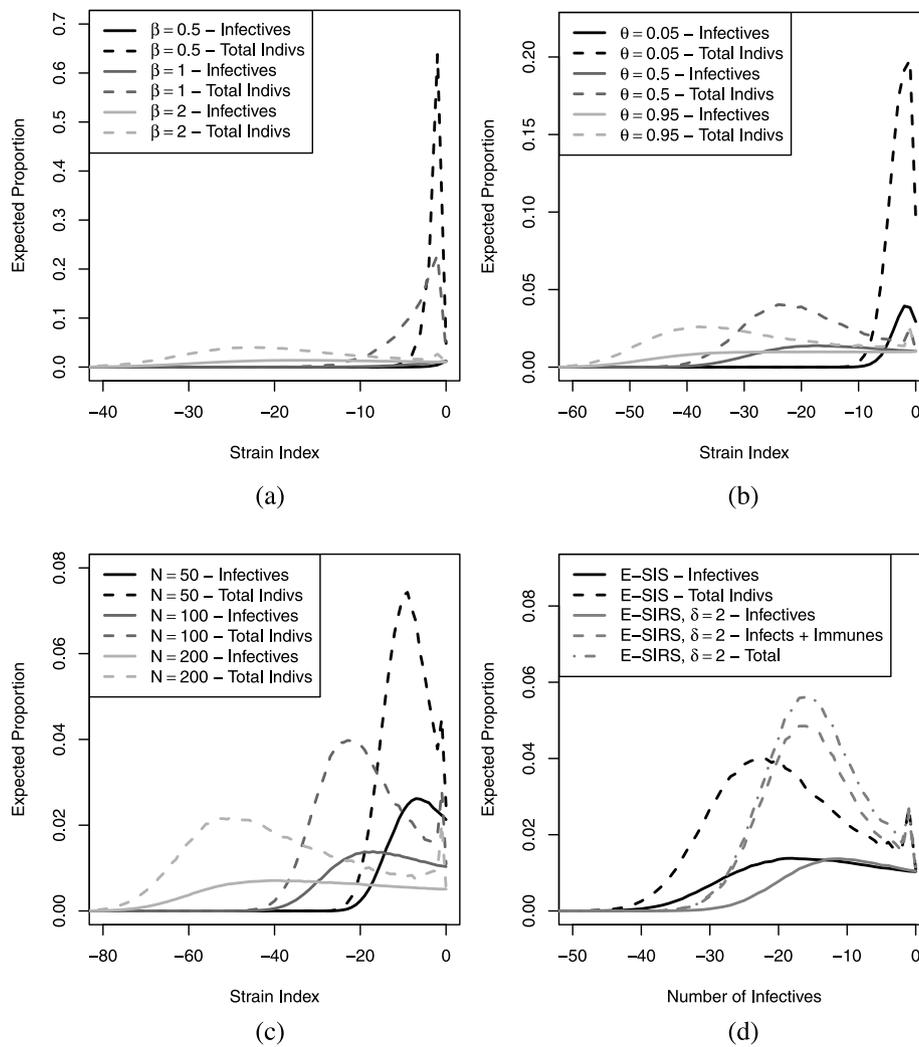
Fig. 7. Strain diversity as a function of the parameters in the E-SIS and E-SIRS models: (a) β varies; (b) θ varies; (c) N varies; (d) δ varies. Unless otherwise stated $\beta = 2$, $\gamma = 1$, $N = 100$, $\theta = 0.5$.

Declaration of competing interest

No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to https://doi.org/10.1016/j.mbs.2020.108480.

References

[1] A.J. Kucharski, V. Andreasen, J.R. Gog, Capturing the dynamics of pathogens with many strains, J. Math. Biol. 72 (1–2) (2015) 1–24.

[2] D. Bichara, A. Iggidr, G. Sallet, Global analysis of multi-strains SIS, SIR and MSIR epidemic models, J. Appl. Math. Comput. 44 (1–2) (2013) 273–292.

[3] M.T. Meehan, D.G. Cocks, J.M. Trauer, E.S. McBryde, Coupled, multi-strain epidemic models of mutating pathogens, Math. Biosci. 296 (2018) 82–92.

[4] J. Gog, B. Grenfell, Dynamics and selection of many-strain pathogens, Proc. Natl. Acad. Sci. 99 (26) (2002) 17209–17214.

[5] S. Moy, Extensions of a limit theorem of everett, ulam and harris on multitype branching processes to a branching process with countably many types, Ann. Math. Stat. 38 (4) (1967) 992–999.

[6] P. Touloupou, B. Finkenstädt, N. French, S. Spencer, Bayesian Inference for multi-strain epidemics with application to *escherichia coli* O157:H7 in feedlot cattle, Ann. Appl. Stat. (2020) in press.

[7] L. E.B.rown, A. Kelso, Prospects for an influenza vaccine that induces cross-protective cytotoxic T lymphocytes, Immunol. Cell Biol. 87 (4) (2009) 300–308.

[8] T. Bedford, M.A. Suchard, P. Lemey, G. Dudas, V. Gregory, A.J. Hay, J.W. McCauley, C.A. Russell, D.J. Smith, A. Rambaut, Integrating influenza antigenic dynamics with molecular evolution, eLife 3 (2014).

[9] M.G. Roberts, R.I. Hickson, J.M. McCaw, L. Talarmain, A simple influenza model with complicated dynamics, J. Math. Biol. 78 (3) (2018) 607–624.

[10] T. Bedford, A. Rambaut, M. Pascual, Canalization of the evolutionary trajectory of the human influenza virus, BMC Biol. 10 (1) (2012).

[11] W. Fitch, R. Bush, C. Bender, N. Cox, Long term trends in the evolution of H(3) HA1 human influenza type A, Proc. Natl. Acad. Sci. 94 (15) (1997) 7712–7718.

[12] T. Bedford, S. Riley, I. Barr, S. Broor, M. Chadha, N. Cox, R. Daniels, C. Gunasekaran, A. Hurt, A. Kelso, A. Klimov, N. Lewis, X. Li, J. McCauley, T. Odagiri, V. Potdar, A. Rambaut, Y. Shu, E. Skepner, D. Smith, M. Suchard, M. Tashiro, D. Wang, X. Xu, L. P., C. Russell, Global circulation patterns of seasonal influenza viruses vary with antigenic drift, Nature 523 (7559) (2015) 217–220.

[13] I. Nåsell, Stochastic models of some endemic infections, Math. Biosci. 179 (1) (2002) 1–19.

[14] H. Andersson, T. Britton, Stochastic Epidemic Models and Their Statistical Analysis, Springer, 2000.

[15] E. van Doorn, P. Pollett, Quasi-stationary distributions for discrete-state models, European J. Oper. Res. 230 (2013) 1–14.

[16] E. van Doorn, P. Pollett, Survival in a quasi-death process, Linear Algebra Appl. 429 (2008) 776–791.

[17] I. Nåsell, On the quasi-stationary distribution of the stochastic logistic epidemic, Math. Biosci. 156 (1999) 21–40.

[18] I. Nåsell, Extinction and Quasi-Stationarity in the Stochastic Logistic SIS Model, Springer Berlin Heidelberg, 2011.

[19] D. Clancy, P. Pollett, A note on quasi-stationary distributions of birth-death processes and the SIS logistic epidemic, J. Appl. Probab. 40 (2003) 821–825.

[20] A. Griffin, P. Jenkins, G. Roberts, S. Spencer, Simulation of quasistationary distributions on reducible state spaces, Adv. Appl. Probab. 49 (2017).

[21] R.M. Dudley, Distances of probability measures and random variables, Ann. Math. Stat. 39 (5) (1968) 1563–1572.

[22] A. Barbour, The duration of the closed stochastic epidemic, Biometrika 62 (2) (1975) 477–482.

[23] H. Andersson, B. Djehiche, A threshold limit theorem for the stochastic logistic epidemic, J. Appl. Probab. 35 (3) (1998) 662–670.

[24] W. Anderson, Continuous-Time Markov Chains: An Applications-Oriented Approach, Springer-Verlag, 1991.

[25] E. van Doorn, Quasi-stationary distributions and convergence to quasi-stationarity of birth-death processes, Adv. Appl. Probab. 23 (4) (1991) 683–700.

[26] S. Jacka, G. Roberts, Weak convergence of conditioned processes on a countable state space, J. Appl. Probab. 32 (4) (1995) 902–916.

[27] R. Anderson, R. May, Infectious Diseases of Humans: Dynamics and Control, Oxford University Press, 1992.

[28] F. Ball, Epidemics in populations with two levels of mixing, Ann. Appl. Probab. 7 (1997) 46–89.

[29] F. Ball, P. Neal, A general model for stochastic SIR epidemics with two levels of mixing, Math. Biosci. 180 (1–2) (2002) 73–102.

[30] F. Ball, P. Neal, Network epidemic models with two levels of mixing, Math. Biosci. 212 (1) (2008) 69–87.

[31] B. Cowling, M. Lau, L. Ho, S. Chuang, T. Tsang, S. Liu, P. Leung, S. Lo, E. Lau, The effective reproduction number of pandemic influenza: prospective estimation, Epidemiology 21 (6) (2010) 842–846.

[32] C. Althaus, Estimating the reproduction number of Ebola virus (EBOV) during the 2014 outbreak in West Africa, PLoS Curr. 6 (2014).

[33] A. Parisi, J. Lopes, A. Nunes, M. Gomes, Heterogeneity in antibody range and the antigenic drift of influenza a viruses, Ecol. Complex. 14 (2013) 157–165.