A Campus Simulation Model Supporting the Safe Blues Experiment

A report by Thomas Graham

In collaboration with Azam Asanjarani, Shane G. Henderson, Yoni Nazarathy

Current Safe Blues Contributors:
Keng Chew, Marijn Jansen, Christopher Rackauckas,
Kirsty Short, Peter G. Taylor, Aapeli Vuorinen, Ilze Ziedins

May 28, 2021

1 Abstract

The Safe Blues project aims to provide real-time caseload predictions during epidemics, bypassing the lag induced by disease incubation. This is accomplished by correlating the transmission of biological diseases with the transmission of virtual safe virus-like tokens that spread between participating smartphones. An experiment at the University of Aukland's City Campus will test the efficacy of this framework; however, the range of strand parameters needed to produce a diverse spectrum of epidemic trajectories is unknown. Here, with the goal of understanding the roles of important strand parameters, we develop a simulation that roughly models the planned experiment. This simulation is used to recommend suitable ranges of strand parameters that might achieve the desired diversity in epidemic trajectories.

2 Introduction

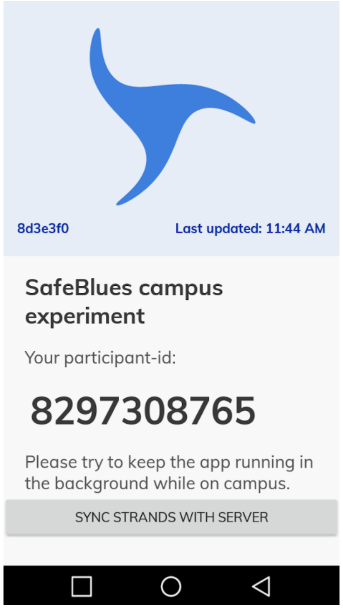
The recent COVID-19 pandemic has highlighted our global vulnerability to diseases, which are able to spread rapidly both within and between densely populated areas. The response from many governments has been to enact various public health measures—for example, mandatory mask wearing, social distancing, or travel restrictions—in an effort to reduce caseloads and fatalities. Consequently, because each of these policies carry different social and economic costs, governments must balance the public health impacts of COVID-19 against the broader societal impacts of taking action [24].

Surely, to effectively compare different public health measures, we must understand how they affect the future progression of an epidemic. This is facilitated by modeling tools within the field of mathematical epidemiology, which applies quantitative techniques to study health and disease at a population level [20]. For example, many specialised models have been designed and used for predicting the course of COVID-19 outbreaks in different countries (see [2], [18], and [19] for instance). A disadvantage of these standard epidemiological models is their reliance on historical caseload statistics, which can often be lagging due to a disease's incubation period. Specifically, because some diseases do not produce symptoms until their incubation period has elapsed, infected individuals are initially unaware of the infection and are unlikely to be diagnosed until after they become symptomatic. Khalili et al. [15] estimate that COVID-19 has an average time between infection and symptom presentation of 5.68 days (99% CI: 4.78, 6.59) and an average time between symptom presentation and first clinical visit of 4.92 days (95% CI: 3.95, 5.90). Overall, since the time of a COVID-19 patient's first clinical visit is an average of 10.60 days behind their time of infection, both the caseload statistics and the predictions of mathematical models are almost certainly lagging.

The Safe Blues project aims to overcome these lagging statistics and, in turn, serve as a method for the real-time measurement, prediction, and control of disease outbreaks (open-source code is available in [5]). Dandekar et al. [4] propose a framework in which the simulated spread of multiple safe virtual viruses (called Safe Blues strands) is used to predict the status of real-world epidemics (see also [6]). A strand spreads between nearby smartphones via Bluetooth communications and its dynamics (transmissibility, incubation duration, and infection duration) are specified by a collection of parameters. Note that, since proximity is also a central factor in the spread of biological pathogens, their epidemiological trajectories should be correlated with those of Safe Blues strands [4]. An ensemble of strands with different parameters, which do not necessarily resemble biological diseases, are simulated simultaneously with the goal of capturing a variety of observable trajectories. Then, a deep neural network is trained on historical data to predict the state of a real-world epidemic using the states of the Safe Blues strands. This means that, because the state of a Safe Blues strand can be inspected without lag at any time, a current prediction of the real-world epidemic can be obtained by feeding the latest strand information into this deep neural network. Hence, the Safe Blues framework can give up-to-date estimates of a disease's caseload, which can be used to more rapidly evaluate the efficacy of various public health measures [4].

Although its underlying mechanism of Bluetooth communications between smartphones is shared with contact-tracing applications for COVID-19 (for example, the Singaporean government's [3] BlueTrace protocol implemented in the OpenTrace application), the Safe Blues project differs in both its purpose and privacy footprint. Given that the purpose of OpenTrace is to assist in contact tracing, it must record a user's interactions alongside other personally identifiable information. Although its designers attempt to mitigate privacy concerns by adopting a model based on "decentralised proximity data collection and centralised contact tracing capability", information must still be transferred to public health authorities after a positive diagnosis of COVID-19 [3]. The Safe Blues project, in contrast, aims to simulate virtual epidemics in a decentralised manner, meaning that the collection of interactions or significant personally identifiable information is not necessary [4]. Thus, while Safe Blues is not intended as a replacement for smartphone-based contact tracing applications, its differences allow it to have a comparatively smaller privacy footprint.

A physical experiment to explore the predictive ability of the Safe Blues framework



8d3e3f0

Last updated: 11:44 AM

SafeBlues campus experiment

Your participant-id:

8297308765

Please try to keep the app running in the background while on campus.

Figure 1: A screenshot of the Android application used within the Safe Blues experiment.

is proposed in [4]. This experiment will be conducted on the University of Aukland's City Campus from May 1, 2021 to November 3, 2021. It involves using smartphones to simulate both red strands, which are stand-ins for biological diseases, and blue strands, which are used for prediction. Note that in Figure 1 a participant ID is used to identify individuals to ensure that only a small amount of personally identifiable information is collected—in fact, only an email is required. A 10-day forecast of the red strands—whose caseloads are artificially delayed to mimic the lag observed in real-world epidemics—is estimated by feeding the blue strands' states into the deep neural network model. The predictions of this model are compared against those of traditional methodologies to assess its relative efficacy [4].

Previously, others have conducted similar epidemiological experiments, including for the BBC documentary "Contagion! The BBC Four Pandemic" [16] and for the FluPhone project [25]. The Contagion experiment simulated a disease outbreak within the United Kingdom by having a smartphone application collect participants' physical locations (GPS) and close contacts (self-reported) throughout a 24-hour period [16]. Similarly, the FluPhone experiment used participants' smartphones to collect their locations (GPS), close contacts (Bluetooth), and influenza-like symptoms (self-reported) [25]. Unlike these previous experiments, the Safe Blues methodology does not need to record participants' locations or interactions because it simulates transmission in a decentralised manner [4].

Although there is a wide range of possible strand parameter values that could be used within the experiment, there is upper bound on the number of strands that can be simulated. So, to maximise the model's predictive ability subject to limited strand capacity, it is desirable to have a good understanding of suitable parameter ranges within which a diversity of epidemic trajectories may be observed. A strand that is too weak will be extinguished quickly due to its inability to maintain sufficient transmission. Similarly, a strand that is too strong will also be extinguished quickly due to it rapidly exhausting the susceptible population. We want to find a "Goldilocks range" so that the strand parameters are neither too weak or too strong. This project aims to develop a simulation tool that approximately describes the spread of these virtual viruses on the University of Auckland's City Campus. Then, we will use this tool to recommend potential parameter ranges depending on the experiment's initial number of participants. Of course, since this simulation is unlikely to perfectly reflect the dynamics of the real-world experiment, these recommendations serve only as an initial "best guess' for the use during its calibration phase.

Moreover, the goal of this project has also been to support the Safe Blues experiment more broadly, including working with both the front-end and back-end systems. Although this report focuses on discussing the simulation and early experimental results, some contributions have also been made to the Safe Blues Android application (see Figure 1), the front-end participant interface, and (more recently) the collection and analysis of early experimental results.

3 Model Description

Here, we will briefly describe the model used to simulate the spread of Safe Blues strands within the University of Aukland's City Campus. This simulation is based on the standard compartmental models used within mathematical epidemiology, specifically the SEIR differential equation model (see, for instance, [20, §5.1]). The population is divided into four distinct compartments—susceptible, exposed (or incubating), infected (or infectious), and recovered—and an epidemic evolves by moving individuals between these compartments. Unlike an agent-based simulation, which operates on an individual level, a compartmental simulation operates on a population level by tracking only the number of individuals belonging to each compartment.

First, to give a brief description of the basic deterministic SEIR model, suppose $S(t)$, $E(t)$, $I(t)$, and $R(t)$ give the number of susceptible, exposed, infected, and recovered individuals at the time point $t \in [0, \infty)$. Then, this model describes the progression of an epidemic using the dynamics

$$\begin{aligned} S'(t) &= -\beta S(t) I(t), \\ E'(t) &= \beta S(t) I(t) - \eta E(t), \\ I'(t) &= \eta E(t) - \alpha I(t), \\ R'(t) &= \alpha I(t), \end{aligned} \tag{1}$$

for all $t \in [0, \infty)$. Here, α is the infected-to-recovered transition rate, η is the exposed-to-infected transition rate, and β is the transmission rate [20]. Note that the $\beta S(t) I(t)$ term in (1) appeals to the law of mass action, which states that the number new infections is proportional to the number of interactions between susceptible and infected persons [20, §3.2].

We will forego these usual deterministic dynamics and, inspired by Barlett's general stochastic epidemic model [1], instead describe movement and infection as random processes. Although these stochastic epidemics are usually more computationally expensive to simulate, they are sometimes able to better capture the diversity of trajectories that may be observed.

Mathematically, we can describe a Safe Blues strand via a collection of parameters $(\pi, \sigma, \rho, \mu_E, \kappa_E, \mu_I, \kappa_I)$ where $\pi \in [0, 1]$ is its seeding probability, $\sigma \in \mathbb{R}^+$ (min^{-1}) is its infection strength, $\rho \in \mathbb{R}^+$ (m) is its infection radius, $\mu_E \in \mathbb{R}^+$ (h) is its mean incubation duration, $\kappa_E \in \mathbb{R}^+$ is its incubation duration shape, $\mu_I \in \mathbb{R}^+$ (h) is its mean infection duration, and $\kappa_I \in \mathbb{R}^+$ is its infection duration shape. The number of participants that are initially infected by the strand is determined randomly, with each participant having a π probability of being infected. Then, the future infections must occur due to close-contacts and have a successful transmission probability depending on the parameters σ and ρ. Specifically, if a susceptible individual is r metres away from an infected individual for t minutes, then the probability of successful transmission is

$$p(r, t; \sigma, \rho) = 1 - e^{-\sigma t\left(1-\frac{r \wedge \rho}{\rho}\right)}. \tag{2}$$

After an individual has been exposed to the strand, they remain in the exposed compartment for a gamma distributed duration with mean μ_E and shape κ_E. Similarly, once this incubation period has elapsed, they remain in the infected compartment for a gamma distributed duration with mean μ_I and shape κ_I. Note that this parameterisation of a gamma distribution has probability density function f where, for each $x \in (0, \infty)$, we set

$$f(x; \mu, \kappa) = \frac{1}{\Gamma(\kappa)(\frac{\mu}{\kappa})^{\kappa}} x^{\kappa-1} e^{-\frac{\kappa x}{\mu}} \tag{3}$$

where μ is the mean parameter, κ is the shape parameter, and Γ is the gamma function.

Now, having described the properties of a strand, we will explain the simulation's movement and infection dynamics. The simulation evolves over a collection of stages $T = \{0, 1, \ldots, 840\}$ representing each hour within a five week period, matching the duration of the experimental phases. The state at stage $t \in T$ is a tuple $s_t = (S_t, E_t, I_t, R_t, \mathcal{A}_t, \mathcal{Y}_t)$ with

- S_t being the current number of susceptible individuals,
- E_t being the current number of exposed individuals,
- I_t being the current number of infected individuals,
- R_t being the current number of recovered individuals,
- $\mathcal{A}_t$ being a collection of times at which exposed individuals will become infected, and
- $\mathcal{B}_t$ being a collection of times at which infected individuals will become recovered.

Note that the $\mathcal{A}_t$ and $\mathcal{B}_t$ components are necessary because a gamma distribution is not necessarily memoryless, so we cannot always simulate these transitions without knowing an individual's initial exposure or infection time. If a population of $N \in \mathbb{Z}^+$ is used, then the initial state is $s_0 = (S_0, E_0, I_0, R_0, \mathcal{A}_0, \mathcal{B}_0)$ where

$$S_0 = N - I_0, \qquad E_0 = 0, \qquad I_0 \sim \mathsf{Bin}(N, \pi), \qquad \text{and} \qquad R_0 = 0. \tag{4}$$

Moreover, we have $\mathcal{A}_0 = \varnothing$ and $\mathcal{B}_0 = \{B_{0,k} \sim \mathsf{Gamma}(\mu_I, \kappa_I) : k = 1, 2, \ldots I_0\}$. Next, fixing a stage $t \in T$, we will explain the three possible transition types that may occur between the state s_t and s_{t+1}: susceptible to exposed, exposed to infected, and infected to recovered.

Susceptible to Exposed The transition from susceptible to exposed is the only transition type that depends on the simulation's movement dynamics. Essentially, at each stage, an individual may be on campus and running the application, on campus and not running the application, or not on campus. The stage-dependent probability of being on campus (or the attendance probability) is denoted $\alpha_t \in [0, 1]$ and the probability of running the application (or the compliance probability) is denoted $\beta \in [0, 1]$. Note that the process $(\alpha_t)_{t \in T}$ can be used to capture the periodic nature of campus attendance over each day and week. Given that strand transmission can only occur between the subsets of susceptible and infected individuals who are active (or both on campus and running the application), we draw the random variables $S_t^* \sim \mathsf{Bin}(S_t, \alpha_t\beta)$ to be the number of active susceptible participants and $I_t^* \sim \mathsf{Bin}(I_t, \alpha_t\beta)$ to be the number of active infected participants. Furthermore, to determine their locations on campus, we draw the collections of two-dimensional points $\{X_{t,k} \sim \mathsf{Dist} : k = 1, 2, \ldots, S_t^*\}$ and $\{Y_{t,k} \sim \mathsf{Dist} : k = 1, 2, \ldots, I_t^*\}$, which contain the coordinates of the susceptible and infected individuals, respectively. The distribution Dist is defined by a heatmap representing the approximate distribution of participants on the University of Auckland's City Campus (see Figure 2). Note that certain buildings—for example, the Science Centre, the Engineering Blocks, and the Owen G Glenn Building—are given greater weights because they are frequented by several cohorts who will be targeted for enrollment in the experiment.

The probability that the k^{th} active susceptible participant is exposed to the strand during this stage is denoted $p_{t,k}$ and, because the transmission events are independent, can be calculated as

$$p_{t,k} = \sum_{\ell=1}^{I_t^*} \left((-1)^{\ell-1} \sum_{\substack{K \subseteq \{1,\ldots,I_t^*\} \\ |K| = \ell}} \prod_{k' \in K} p\big(\|X_{t,k} - Y_{t,k'}\|, 60; \sigma, \rho\big) \right). \tag{5}$$

Then, for each $k = 1, 2, \ldots, I_t^*$, we define a random variable $\delta_{t,k}^{S \to E} \sim \mathsf{Ber}(p_{t,k})$ indicating whether the k^{th} active susceptible participant is exposed. This allows us to randomly determine the total number of new exposures as

$$\Delta_t^{S \to E} = \sum_{k=1}^{I_t^*} \delta_{t,k}^{S \to E}. \tag{6}$$

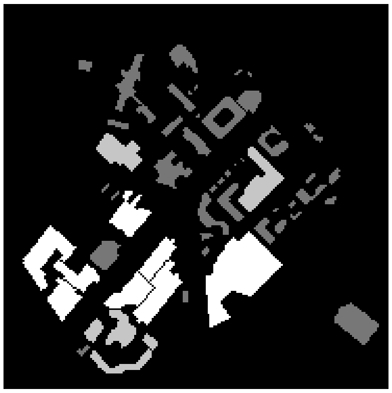


Figure 2: The approximate distribution of participants on the University of Aukland's City Campus represented as a heatmap from black (weighted zero) to white (weighted one).

The set $\mathcal{A}_t^{S\to E} = \{A_{t,k} \sim 60t{+}\mathsf{Gamma}(\mu_E, \kappa_E) : k = 1, 2, \ldots, \Delta_t^{S\to E}\}$ stores the infection transition times of these newly exposed individuals; that is, $A_{t,k}$ is the time at which the k^{th} newly exposed individual becomes infectious and is moved into the infected compartment.

Exposed to Infected If the simulation's state at the current stage is already known, then the number of newly infectious individuals is deterministic since the set $\mathcal{A}_t$ stores the previously generated exposed-to-infected transition times. Thus, the number of participants transitioning from the exposed compartment to the infected compartment is

$$\Delta_t^{E\to I} = \left|\left\{A \le t : A \in \mathcal{A}_t\right\}\right|. \tag{7}$$

We retain the remaining exposed-to-infected transition times $\mathcal{A}_t^{E\to E} = \{A > t : A \in \mathcal{A}_t\}$ and generate the newly infectious participants' infected-to-recovered transition times $\mathcal{B}_t^{E\to I} = \{B_{t,k} \sim 60t + \mathsf{Gamma}(\mu_I, \kappa_I) : k = 1, 2, \ldots, \Delta_t^{E\to I}\}$.

Infected to Recovered The infected-to-recovered transitions behave almost identically to the exposed-to-infected transitions. The number of participants who are moved from the infected compartment to the recovered compartment at the current stage is computed by counting the number of elapsed recovery times:

$$\Delta_t^{I\to R} = \left|\left\{B < t : B \in \mathcal{B}_t\right\}\right|. \tag{8}$$

Moreover, the remaining infected-to-recovered transition times are $\mathcal{B}_t^{I\to I} = \{B > t : B \in \mathcal{B}_t\}$.

Finally, after computing the number of individuals transitioning between compartments, we are able to update the state for the subsequent stage $t + 1$. This state

$s_{t+1} = (S_{t+1}, E_{t+1}, I_{t+1}, R_{t+1}, \mathcal{A}_{t+1}, \mathcal{B}_{t+1})$ is given by

$$\begin{aligned} S_{t+1} &= S_t - \Delta_t^{S\to E}, \\ E_{t+1} &= E_t + \Delta_t^{S\to E} - \Delta_t^{E\to I}, \\ I_{t+1} &= I_t + \Delta_t^{E\to I} - \Delta_t^{I\to R}, \\ R_{t+1} &= R_t + \Delta_t^{I\to R}, \\ \mathcal{A}_{t+1} &= \mathcal{A}_t^{S\to E} \cup \mathcal{A}_t^{E\to E}, \text{ and} \\ \mathcal{B}_{t+1} &= \mathcal{B}_t^{E\to I} \cup \mathcal{B}_t^{I\to I} \end{aligned} \tag{9}$$

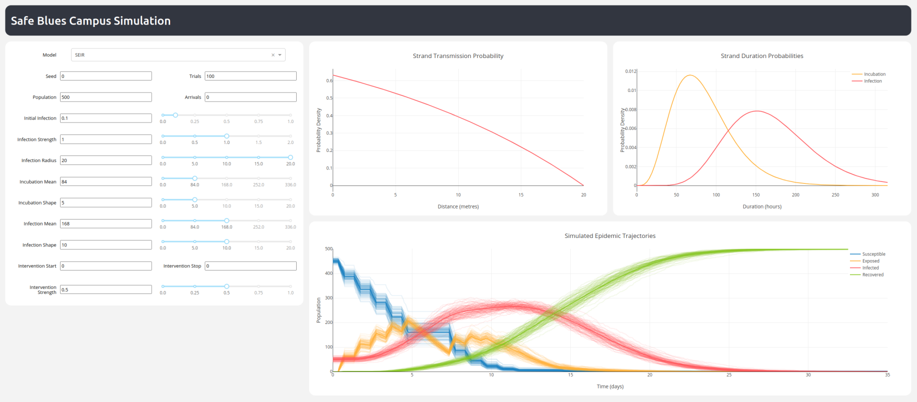
where the elements of $\mathcal{A}_{t+1}$ and $\mathcal{B}_{t+1}$ are relabelled as appropriate. The simulation starts at the initial state s_0 and, for each stage $t \in T$, applies this updating process to determine the subsequent state. Then, we can plot this realisation of the processes $(S_t)_{t\in T}$, $(E_t)_{t\in T}$, $(I_t)_{t\in T}$, and $(R_t)_{t\in t}$ to view the epidemic's trajectory and assess the strand's suitability.

4 Implementation Details

The previously described epidemic model has been implemented in Julia to take advantage of the language's fast development and execution time [8]. A user is able to run these simulations either through the Julia REPL (a command-line tool) or an interactive dashboard (see Figure 3). The interactive dashboard has a variety of controls (see Figure 5) that set the model, strand, and population parameters. Specifically, in addition to the SEIR model discussed previously, the dashboard supports various other models includng SIR (zero incubation duration), SI (zero incubation duration and infinite infection duration), and SEIS (where reinfection may occur). It also allows for "artificial social distancing" that weakens the strand's infection strength over a specified time interval. This is important because the experiment proposed by Dandekar et al. [4] intends to test how the Safe Blues protocol adapts to changing public health policies. Then, the dashboard uses these parameters to produce real-time predictions of the corresponding epidemic trajectories within the campus experiment (see Figure 4). This allows the Safe Blues researchers to view how a strand might behave before introducing it to the experiment's real-world population.

Here, we will review some of the implementation decisions that were made to improve the simulation's performance, including optimising the processes of generating locations, transmitting strands, and maintaining the collections $\mathcal{A}_t$ and $\mathcal{B}_t$. Overall, combined with the naturally fast execution time of Julia, this allowed the dashboard to operate almost in real-time on small populations of less than 1000, which is certainly sufficient for the number of expected participants in the experiment (see Table 1 in Appendix).

First, we will focus on the process of generating the participants' locations within the campus. Recall that these locations are drawn from a probability distribution Dist that is defined using a heatmap showing likely areas of congregation (see Figure 2). Precisely, sampling a single individual's position $(x, y) \in \mathbb{R}^2$ occurs in two stages:



Safe Blues Campus Simulation

Model SEIR

Seed 0 Trials 100

Population 500 Arrivals 0

Initial Infection 0.1

Infection Strength 1

Infection Radius 20

Incubation Mean 84

Incubation Shape 5

Infection Mean 168

Infection Shape 10

Intervention Start 0 Intervention Stop 0

Intervention Strength 0.5

Figure 3: A snapshot of the web-based dashboard that can be used for interacting with the simulation tool.

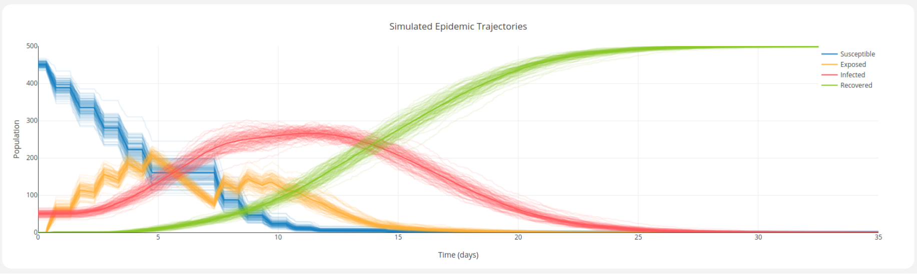
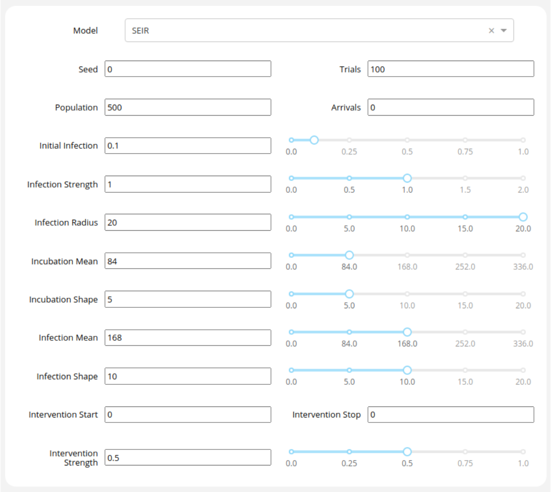


Figure 4: A snapshot of the interactive dashboard's simulation output.



Model SEIR

Seed 0 Trials 100

Population 500 Arrivals 0

Initial Infection 0.1

Infection Strength 1

Infection Radius 20

Incubation Mean 84

Incubation Shape 5

Infection Mean 168

Infection Shape 10

Intervention Start 0 Intervention Stop 0

Intervention Strength 0.5

Figure 5: A snapshot of the interactive dashboard's control panel.

(a) generating the position of a pixel's lower-left corner from a discrete distribution, and

(b) generating a position within the pixel uniformly at random.

Given that the campus heatmap may need to be changed in future to better reflect the population dynamics, the process in (a) must be able handle arbitrary discrete distributions. This was initially accomplished using the `sample` function from the `StatsBase.jl` Julia package, which includes basic probability and statistics tools [11]. Precisely, this function implements a poly-algorithm that, for taking a weighted sample with replacement, selects from either direct sampling or alias sampling depending on its arguments. Suppose a weighted selection of $k \in \{1, 2, \ldots, n\}$ elements from an array of length $n \in \mathbb{Z}$ is needed. The direct sampling method, which has a (worst-case) time complexity of $O(nk)$, accomplishes this by iterating through the weights array k times. Comparatively, after the alias method completes an $O(n \log n)$ set-up stage, it is able to generate the weighted sample with time complexity $O(k)$. The set-up stage creates a pair of distribution-dependent tables: a probability table $p : \{1, \ldots, n\} \to [0, 1]$ that maps indices to probabilities and an alias table $a : \{1, \ldots, n\} \to \{1, \ldots, n\}$ that maps indices to other indices. Then, to sample from this discrete distribution, we simply generate $i \in \{1, 2, \ldots, n\}$ uniformly at random and return the i^{th} element with probability $p(i)$ and the $a(i)^{\text{th}}$ element with probability $1 - p(i)$ [17, §3.1.4]. The `sample` method must be called at least once every stage to generate the locations of the susceptible and infected individuals; however, because it does not reuse the previously constructed tables, the set-up process is repeated multiple times. We can improve on this by noting that the underlying distribution of participants does not change throughout the course of the simulation. So, after constructing a probability table and alias table once, we can reuse these to generate locations without having repeat the set-up process. Given that an open-source implementation of the alias method that allows for this reusability does not appear to exist in Julia, the set-up and sampling functions had to be written from scratch. Note that this implementation is based on the pseudocode in [17, Algorithm 3.7, Algorithm 3.8] where the set-up method has time complexity $O(n)$.

Next, note that the calculation in (5) of a susceptible participant's exposure probability requires iteration over each active infected individual. If there are m active susceptible participants and $n \in \mathbb{Z}^+$ active infected participants, then computing their exposure probabilities is an $O(mn)$ operation. We cannot expect to improve upon this time complexity because the law of mass action, which is a fundamental assumption within epidemic modelling, states that the number of transmission events is proportional to the product mn (see [20, §3.2]). Instead, to slightly improve the execution time of the simulation, we can reduce the cost of computing the probability that an infected individual successfully transmits a strand to a susceptible individual. Notice that, in the right-hand side of (5), the probability $p(\|X_{t,k} - Y_{t,k}\|, 60; \sigma, \rho)$ of transmission between the k^{th} susceptible participant and k'^{th} infected participant depends on their distance $\|X_{t,k} - Y_{t,k}\|$. This requires the use of the square root operation. However, when the distance is greater than the strand's infection radius, the transmission probability is always zero and has no further dependence on the participants' distance. Essentially, from the definition in (2), we have

$$p\big(\|X_{t,k} - Y_{t,k'}\|, 60; \sigma, \rho\big) = \begin{cases} 1 - e^{-60\sigma\left(1 - \frac{\|X_{t,k} - Y_{t,k'}\|}{\rho}\right)}, & \|X_{t,k} - Y_{t,k'}\| \leq \rho, \\ 0, & |X_{t,k} - Y_{t,k'}\| > \rho, \end{cases} \tag{10}$$

but

$$p\big(\|X_{t,k} - Y_{t,k'}\|, 60; \sigma, \rho\big) = \begin{cases} 1 - e^{-60\sigma\left(1 - \frac{\|X_{t,k} - Y_{t,k'}\|}{\rho}\right)}, & \|X_{t,k} - Y_{t,k'}\|^2 \leq \rho^2, \\ 0, & \|X_{t,k} - Y_{t,k'}\|^2 > \rho^2, \end{cases} \tag{11}$$

would suffice. Observe that the conditions in (11) do not actually require the computation of a square root, so implementing this function instead of (10) will result in performing fewer square root operations overall. Indeed, making this change results in a small improvement to the simulations performance.

Lastly, to optimise the performance of the exposed-to-infected and infected-to-recovered transitions, we need to select the correct data structures for $\mathcal{A}_t$ and $\mathcal{B}_t$. These were initially represented as arrays whose elements were ordered from lowest to highest. If an ordered array is of length $n \in \mathbb{Z}^+$, then correctly inserting a new element takes time $O(n)$, deleting an element takes time $O(n)$, and finding the minimum element takes time $O(1)$ (only because the array is ordered) [21, §3.2]. We can improve upon this performance by using the minimum binary heap implementation from the `DataStructures.jl` package [10]. Given a binary heap of size $n \in \mathbb{Z}^+$, adding a new element takes time $O(\log n)$, deleting an element takes time $O(\log n)$, and finding the minimum element takes time $O(1)$ [21, §6.1]. So, it is preferable to use a heap data structure when maintaining the collections $\mathcal{A}_t$ and $\mathcal{B}_t$ throughout the simulation.

5 Results

Now, when using the simulation to explore the role of strand parameters within the Safe Blues experiment, we will focus on a strand's infection strength σ and infection radius ρ. This narrows the number of varying parameters and, as we shall show, we can capture a sufficiently diverse range of epidemic trajectories by changing only these parameters. Specifically, we will explore infection strengths in the range $\sigma \in [0, 0.1]$ and infection radii in the range $\rho \in [0, 20]$. Moreover, throughout our future simulations, we will assume that the seeding probability is $\pi = 0.1$, the mean incubation duration is $\mu_E = 24.0$h (or a single day), the incubation duration shape is $\kappa_E = 5.0$, the mean infection duration is $\mu_I = 168.0$h (or a single week), and the infection duration shape is $\kappa_I = 5.0$. Note that the number participants enrolled in the Safe Blues experiment, as of the beginning of its calibration phase on May 1st 2021, is roughly $N = 100$. So, we will focus on modelling epidemics for population sizes $N \in \{100, 200, 500\}$, which includes potential for growth during future phases.

How should the "success" of a strand be quantified. We know that an unsuccessful epidemic would spread to a small portion of the cohort and a successful epidemic would spread to a large portion of the cohort. Thus, we are interested in the number of cumulative exposures throughout the simulation's entire time horizon. This can be computed as either

$$E_t + I_t + R_t \qquad \text{or} \qquad N - S_t \tag{12}$$

where $t = 840$. Here, to explore the effect of varying a strand's infection strength and infection radius on its cumulative exposures, we plot this dependence for strengths $\sigma \in \{0, 0.005, \ldots, 0.1\}$ and radii $\rho \in \{0.0, 1.0, \ldots, 20.0\}$. The average total exposures, after repeating 100 simulations of each combination of infection strength and infection radius, is shown in Figure 6 for $N = 100$, Figure 7 for $N = 200$, and Figure 8 for $N = 500$. Moreover, for a selected subset of parameters, Table 1 (in Appendix) shows their average execution time to quantify the performance of the implementation.

6 Discussion

Analysis and Interpretation Next, we want to analyse and understand several of the prominent features appearing in Figure 6, Figure 7, and Figure 8. Clearly, these graphs share some similar structural elements, including a plateau of high cumulative exposures around $(\sigma, \rho) = (0.1, 20.0)$, a plateau of low cumulative exposures around $(\sigma, \rho) = (0.0, 0.0)$, and a transition between these regions. Given that the Safe Blues framework described in [4] uses a diverse ensemble of strands to predict the caseloads of biological epidemics, we are mostly interested in this transitional region. If the parameters within this region are targeted, then we should expect to see strands that spread with various degrees of success.

Observe that there appears to be a minimum infection radius before any significant transmission can occur in the simulation results. This minimum infection radius is roughly $\rho = 10.0$ in Figure 6, $\rho = 5.0$ in Figure 7, and $\rho = 2.0$ in Figure 8. Why is this the case? Recall that, from the transmission probability function in (2), the k^{th} susceptible individual can only be infected by the k'^{th} infected individual when a close contact (within ρ metres) occurs. Thus, if $\mathbb{P}(\|X_{t,k} - Y_{t,k'}\| \leq \rho)$ is the probability of a close contact, then the transmission probability can be rewritten as

$$p\big(\|X_{t,k} - Y_{t,k'}\|, 60; \sigma, \rho\big) = \mathbb{P}\big(\|X_{t,k} - Y_{t,k'}\| \leq \rho\big)\left(1 - e^{-\sigma 60\left(1 - \frac{\|X_{t,k} - Y_{t,k'}\|}{\rho}\right)}\right). \tag{13}$$

The close-contact probability $\mathbb{P}(\|X_{t,k} - Y_{t,k'}\| \leq \rho)$ depends only on ρ and serves as an upper bound on the probability of transmission

$$p\big(\|X_{t,k} - Y_{t,k'}\|, 60; \sigma, \rho\big) \leq \mathbb{P}\big(\|X_{t,k} - Y_{t,k}\| \leq \rho\big). \tag{14}$$

Accordingly, for a sufficiently small infection radius, we should expect the transmission probability to always be too small to result in a successful epidemic. Note that this minimum infection radius for successful spread is dependent on the total population because increasing the density of participants also increases the rate of close contacts, making transmissions more frequent overall. This explains why the simulated epidemics fail to spread beyond their initial exposed populations when the strand's infection radius is small.

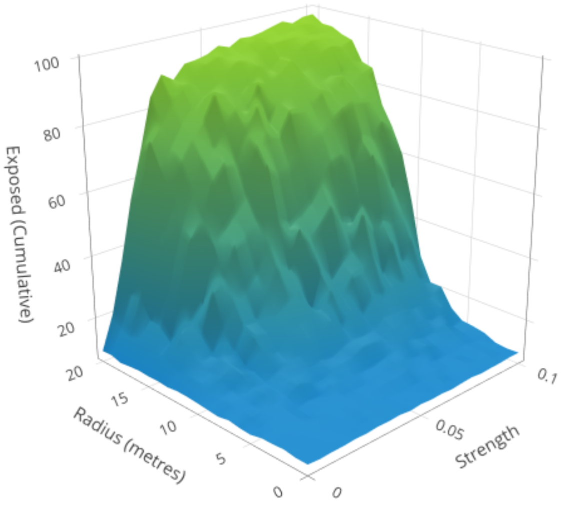


Figure 6: The dependence of cumulative exposures on a strand's infection radius and infection strength after 100 simulations with $N = 100$.

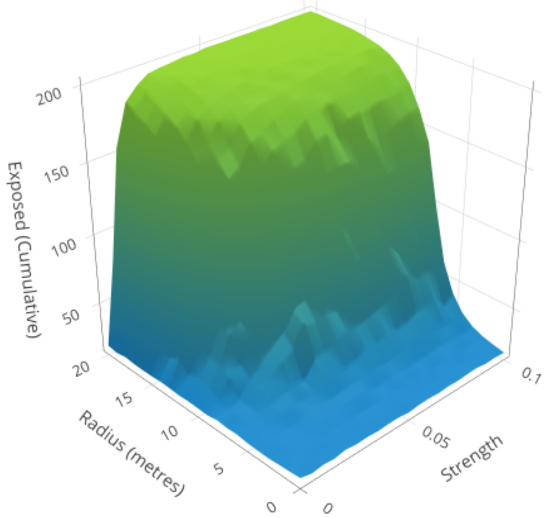


Figure 7: The dependence of cumulative exposures on a strand's infection radius and infection strength after 100 simulations with $N = 200$.

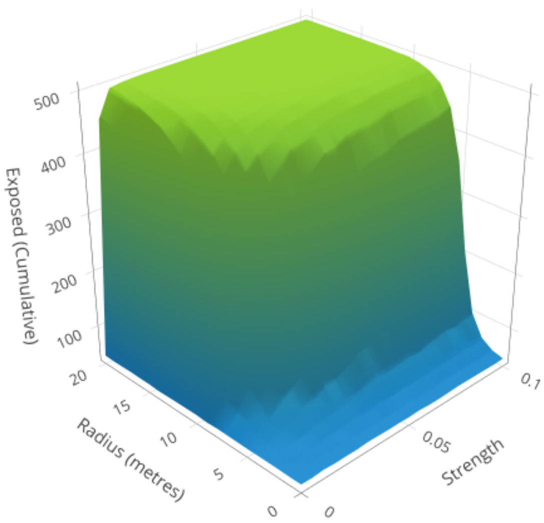


Figure 8: The dependence of cumulative exposures on a strand's infection radius and infection strength after 100 simulations with $N = 500$.

A consequence of our previous observation, within the context of the Safe Blues experiment, is that the strands' infection radii must be carefully chosen dependent on the number of enrolled participants. Unfortunately, an added complication is introduced in the form of range limitations of the underlying Bluetooth technology. The original implementation of OpenTrace—the contact tracing project upon which the Safe Blues application for Android is based—in [3] uses the received signal strength indicator (RSSI) of Bluetooth to infer distances between interacting smartphones. Bay et al. [3] note that the range of Bluetooth in indoor environments is 10 metres, suggesting that the transmission of strands with $\rho > 10.0$ might be unreliable or impossible. So, it is especially important that the Safe Blues experiment has sufficiently many participants to ensure that the minimum required infection radius for widespread transmission can be achieved subject to the technological limitations of Bluetooth.

Additionally, another prominent feature in Figure 6, Figure 7, and Figure 8 is that the steepness of the transitional region increases with the population size. This means that, by raising the total number of enrolled participants, we will narrow the parameter ranges necessary to produce a diversity of epidemic trajectories. A straightforward explanation for this phenomenon is that, by increasing the density of the population within the fixed campus dimensions, we increase the expected number of new infections caused by a single carrier over their infectious period. Note that this quantity is known as the effective reproduction number (often denoted as $\mathcal{R}_t$) and, with the added condition that the surrounding population is completely susceptible, it is known as the basic reproduction number (often denoted as $\mathcal{R}_0$). The basic reproduction number is a fundamental constant associated with an epidemic as it can often be used to predict the degree to which it will spread. Indeed, in many common epidemiological models, $\mathcal{R}_0 < 1$ implies that an epidemic will not occur and $\mathcal{R}_0 > 1$ implies that an epidemic will occur (to some extent) [9]. Kermack and McKendrick [12, 13, 14], alongside their introduction of the original deterministic SIR epidemiological model, show an equivalent result phrased in terms of population densities. Precisely, they show that an infectious disease has a threshold population density where (a) an epidemic does not occur below this density and (b) an epidemic does occur above this density. Moreover, they note that the severity of an epidemic increases as the population density increases. This threshold result is extended to the stochastic setting by Whittle [23] by proving that an epidemic does not occur when below the threshold population density (or $\mathcal{R}_0 < 1$) and an epidemic does occur with non-zero probability above the threshold population density (or $\mathcal{R}_0 > 1$). Clearly, these threshold density observations explain the expanding plateaus in Figure 6, Figure 7, and Figure 8 on which the cumulative exposures are maximised. Specifically, as the population density rises, we are seeing the expansion of the subset of strands whose threshold densities have been exceeded. This steepens the transitional region as the expanding plateau grows towards occupying the entire parameter space. The consequence of this observation, within the context of the Safe Blues experiment, is that larger cohorts demand a more concentrated distribution of strand parameters to attain the desired diversity.

Now, after exploring the structural features of the simulation results, we are prepared to make parameter recommendations for the Safe Blues experiment. Recall the three important points from our previous discussion:

(a) there exists a minimum radius for sustained transmission, which decreases as population size increases, and

(b) there exists a region of transitional parameters, which narrows as population size increases,

We apply these observations to conclude that $\sigma \in [0.0, 0.05]$ or $\rho \in [10.0, 20.0]$ is the optimal parameter range for $N = 100$, $\sigma \in [0.0, 0.05]$ or $\rho \in [5.0, 15.0]$ is the optimal parameter range for $N = 200$, and $\sigma \in [0.0, 0.04]$ or $\rho \in [2.0, 12.0]$ for $N = 500$. In all three cases, it is important that the infection radius is greater than the observed minimum infection radius for sufficient transmission: 10m for $N = 100$, 5m for $N = 200$, and 2m for $N = 500$. Evidently, given that [3] states 10 metres is the effective Bluetooth range, achieving the greatest levels of cumulative exposures within these ranges might not be possible.

Errors and Improvements Finally, it is important to discuss potential errors within the simulation and the effects of these errors on our recommendations. We cannot expect this simulation to be a perfect predictor of reality due to the assumptions that have been made. These assumptions can be categorised into two groups: those that address transmission dynamics and those that address movement dynamics. The remaining elements of the simulation—for example, the exposed-to-infected and infected-to-recovered transitions—do not need to be approximated because their counterparts in the smartphone application could be closely replicated.

First, the transmission mechanism implemented in the simulation is not constrained by the same technological or environmental limitations as the experimental smartphone application. We have already identified that limited communication range is an important factor in the spread of the virtual viruses [3]; this limitation is not modelled in the simulation where the infection range is unbounded. Additionally, the simulation does not explicitly model the relationship between obstructions and the transmission of strands via Bluetooth signals. Recall that the Safe Blues implementation is based on the OpenTrace application introduced in [3], which uses the received signal strength indicator (RSSI) to estimate the distance between smartphones. Although this signal strength decreases with distance, it can also be limited by obstructions common in indoor environments [7]. Given that the simulation does not differentiate between indoor (or high obstruction) and outdoor (or low obstruction) environments, we should expect this to contribute as a source of error. We could avoid these errors by explicitly modelling obstructions and the RSSI process used to compute distance; however, this would necessarily increase the complexity and execution time of the model.

Second, the movement dynamics are only a rough approximation of individuals' behaviour on the university campus. The simulation does not describe a participant's day as a continuous process in which they move around campus according to a schedule. Instead, at every stage, they are randomly placed somewhere on campus according to a distribution that is intended to reflect the average participant's schedule. This allows movement to be modelled at a population level and sufficiently captures the tendency to congregate in certain buildings, but it misses other elements that may play a role in epidemic growth during the experiment. Specifically, since the entire population is functionally identical, the increased likelihood of strand transmission within certain settings—for examples, lectures, tutorials, or social gatherings—are ignored. Although this could be improved by using an agent-based model in which the population was modelled on an individual level, but this would also increase the simulation's complexity.

Ultimately, due to the assumptions that have been made throughout the development of this simulation, the recommended parameter ranges serve only as a starting point for the experiment's calibration phase. The information collected during this initial phase will then inform the future choices of parameters during the remaining experimental phases.

Safe Blues Experiment Now, as of May 21st 2021, the initial calibration phase of the Safe Blues experiment has been running for approximately a month. This has involved the release of multiple strand batches with different parameters to further investigate the suitable parameter ranges. A total of five batches have been released labelled from v1.01 to v1.05. Here, we will give a brief overview of the strands released during the calibration phase; a detailed description of these strands is available at [22].

The first batch (v1.01) contained 162 strands consisting of infection strengths $\sigma \in \{0.06, 0.16, 0.48\}$ and infection radii $\rho \in \{5, 15, 500\}$. The remaining parameters varied to include strands with no incubation, no recoveries, short infection duration ($\mu_I = 120$h or 5 days), or long infection duration ($\mu_I = 240$h or 10 days). The simulation was used to verify that the parameters covered a suitable variety of epidemic trajectories.

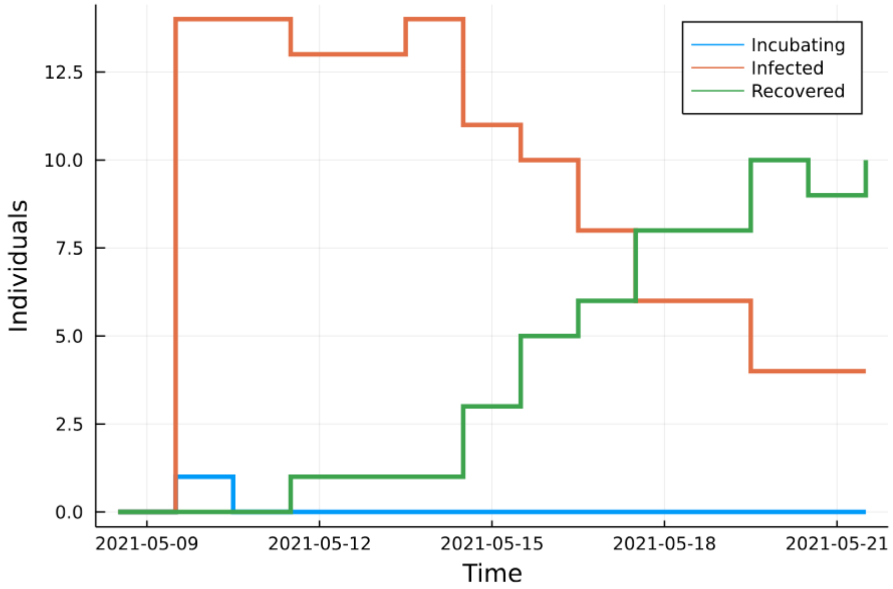


Figure 9: The trajectory of v1.04 Safe Blues Strand #255 with $\pi = 0.1$, $\sigma = 0.48$, $\rho = 15$m, $\mu_E = 12$h, $\kappa_E = 10000$, $\mu_I = 240$h, and $\kappa_I = 3$.

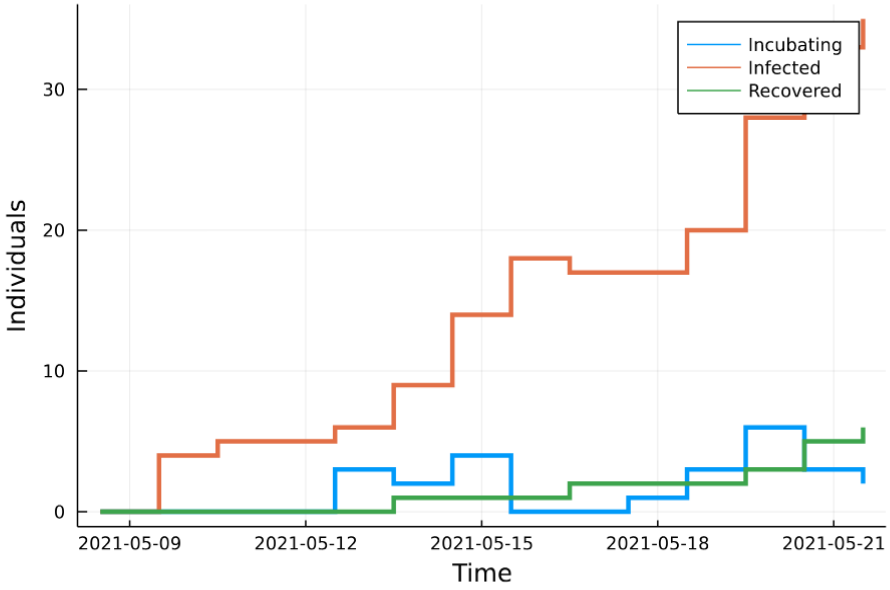


Figure 10: The trajectory of v1.04 Safe Blues Strand #258 with $\pi = 0.1$, $\sigma = 0.48$, $\rho = 500$m, $\mu_E = 12$h, $\kappa_E = 10000$, $\mu_I = 240$h, and $\kappa_I = 3$.

Unfortunately, the release of the v1.01 strands revealed a bug in the Safe Blues Android application. The random number generator used a fixed seed shared between devices and, as a consequence, each strand initially infected almost everybody or almost nobody. The v1.02 and v1.03 strand releases were addressed at fixing this mistake.

After sufficiently many participants had updated to the latest patched application version, the v1.01 strands were re-released as v1.04. Although only a small amount of data has been collected, we can already begin to compare some of the initial real-world trajectories with the simulation's predictions. Particularly, across the 162 strands, the strength parameter appears to have insignificant influence on the transmission rate, whereas, the radius parameter appears to greatly influence transmission. Here, as an example, the trajectory of Strand #255 (with infection radius $\rho = 15\text{m}$) and Strand #258 (with infection radius $\rho = 500\text{m}$) are shown in Figure 9 and Figure 10. These strands share the remaining parameters $\pi = 0.1$, $\sigma = 0.48$, $\mu_E = 12\text{h}$, $\kappa_E = 10000$, $\mu_I = 240\text{h}$, and $\kappa_I = 3$. Note that Strand #258's infection radius of $\rho = 500\text{m}$ is a stand-in for an "umlimited" infection range; it is unlikely that Bluetooth communication can exceed this range (see [3]). Clearly, Strand #255 exhibits very little growth and only manages to infect one additional participant. This causes its caseload to fall from 14 to 4 and its cumulative exposures to reach only 15 during the 12-day period. Strand #258, on the other hand, is able to infect multiple people with its caseload growing from 4 to 35 and its cumulative exposures reaching 43 during the 12-day period. These initial results indicate that there is a "transmissibility slope" between $\rho = 15\text{m}$ and $\rho = 500\text{m}$ on which the transmissibility of the strands increase. Indeed, this aligns with the simulation's results for $N = 100$ participants in Figure 6; however, the exact location of this transitional region appears to be different. Although it did not exactly predict these outcomes, the simulation still serves as a useful tool for interpreting the experimental results and understanding the underlying dependence of transmissibility of the parameters.

Recently, the v1.05 strand batches have been released with the goal of further exploring suitable infection radius values. These strands have various infection radii radii from $\rho \in \{7.5\text{m}, 15\text{m}, 30\text{m}, 60\text{m}, 120\text{m}, 500\text{m}\}$. The remaining parameters are fixed at $\pi = 0.1$, $\sigma = 0.16$, $\mu_E \approx 0\text{h}$, $\kappa_E = 10000$, $\mu_I = 960\text{h}$, and $\kappa_I = 10000$. This ensures that the strands have effectively no incubation and no recoveries. Unfortunately, at the time of writing, these strands have not been circulating within the population for long enough to clearly identify the epidemics' trajectories.

7 Conclusion

Here, we sought to develop a simulation of the planned Safe Blues experiment to understand the roles of strand parameters and their optimal ranges. The underlying epidemiological model used in this project was a stochastic compartmental SEIR model, which was then implemented in Julia as a command-line tool and interactive dashboard. This simulation was used to understand the role of the infection strength and infection radius in determining the cumulative number of exposures. This served as a starting point for the experiment's calibration phase and was used when designing the v1.01 and v1.04 strands. These strands have been active for approximately 12 days—at the time of writing—and significant transmission was observed for large infection radii (see Figure 10), but no significant transmission was observed for smaller infection radii (see Figure 9). Recently, additional strands have been released to more precisely capture the transition between weak and strong epidemics. Altogether, the simulation has fulfilled its purpose as a starting point for planning strands that achieve sufficiently diverse trajectories.

References

[1] M. S. Barlett. Some evolutionary stochastic processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(2):211–229, 1949.

[2] C. M. Batistela, D. P. F. Correa, Á. M. Bueno, and J. R. C. Piqueira. Sirsi compartmental model for covid-19 pandemic with immunity loss. *Chaos, Solitons & Fractals*, 142, 2021.

[3] J. Bay, J. Kek, A. Tan, C. S. Hau, L. Yongquan, J. Tan, and T. A. Quy. BlueTrace: A Privacy-Preserving Protocol for Community-Driven Contact Tracing across Borders, Singapore, SG, 2020.

[4] R. Dandekar, S. G. Henderson, M. Jansen, J. McDonald, S. Moka, Y. Nazarathy, C. Rackauckas, P. G. Taylor, and A. Vuorinen. Safe blues: the case for virtual safe virus spread in the long-term fight against epidemic. *Patterns*, 2(3), 2021.

[5] R. Dandekar, S. G. Henderson, M. Jansen, S. Moka, Y. Nazarathy, C. Rackauckas, P. G. Taylor, V. A., T. Stace, and J. McDonald. Safe Blues. GitHub Organization, 2020. URL: `github.com/SafeBlues`.

[6] R. Dandekar, S. G. Henderson, M. Jansen, S. Moka, Y. Nazarathy, C. Rackauckas, P. G. Taylor, and A. Vuorinen. Safe blues: a method for estimation and control in the fight against covid-19. *medRxiv*, 2020.

[7] A. B. Dar, A. H. Lone, S. Zahoor, A. A. Khan, and R. Naaz. Applicability of Mobile Contact Tracing in Fighting Pandemic (COVID-19): Issues, Challenges, and Solutions. *Computer Science Review*, 38, 2020.

[8] T. Graham. SafeBluesCampusSimulation. GitHub Repository, 2021. URL: `github.com/tjgraham/SafeBluesCampusSimulation`.

[9] J. A. P. Heesterbeek. The concept of r0 in epidemic theory. *Statistica Neerlandica*, (1):89–110, 1996.

[10] JuliaCollections. DataStructures.jl v0.18.9. GitHub Repository, 2021. URL: `github.com/JuliaCollections/DataStructures.jl`.

[11] JuliaStats. StatsBase.jl v0.33.6. GitHub Repository, 2021. URL: `github.com/JuliaStats/StatsBase.jl`.

[12] W. O. Kermack and A. G. McKendrick. Contributions to the mathematical theory of epidemics ii. *Bulletin of Mathematical Biology*, 53(1–2):33–55, 1991.

[13] W. O. Kermack and A. G. McKendrick. Contributions to the mathematical theory of epidemics iii. *Bulletin of Mathematical Biology*, 53(1–2):57–87, 1991.

[14] W. O. Kermack and A. G. McKendrick. Contributions to the mathematical theory of epidemics iii. *Bulletin of Mathematical Biology*, 53(1–2):89–118, 1991.

[15] M. Khalili, M. Karamouzian, N. Nasiri, S. Javadi, A. Mirzazadeh, and H. Sharifi. Epidemiological characteristics of covid-19: a systematic review and meta-analysis. *Epidemiology & Infection*, 148, 2020.

[16] P. Klepac, S. Kissler, and J. Gog. Contagion! the bbc four pandemic—the model behind the documentary. *Epidemics*, 24:49–59, 2018.

[17] D. P. Kroese, T. Taimre, and Z. I. Botev. *Handbook of Monte Carlo Methods*. Wiley Series in Probability and Statistics. John Wiley & Sons, Hoboken, NJ, 2011.

[18] A. P. Lemos-Paião, C. J. Silva, and D. F. M. Torres. A new compartmental epidemiological model for covid-19 with a case study of portugal. *Ecological Complexity*, 44, 2020.

[19] A. Leontitis, A. Senok, A. Alsheikh-Ali, Y. Al Nasser, T. Loney, and A. Alshamsi. Seahir: a specialized compartmental model for covid-19. *International Journal of Environmental Research and Public Health*, 18(5), 2021.

[20] M. Martcheva. *An Introduction to Mathematical Epidemiology*. Texts in Applied Mathematics. Springer, New York, NY, 1st edition, 2015.

[21] K. Mehlhorn and P. Sanders. *Algorithms and Data Structures*. Springer, 2008.

[22] Y. Nazarathy. SafeBluesStrandMaker. GitHub Repository, 2021. URL: `github.com/yoninazarathy/StrandMaker`.

[23] P. Whittle. The Outcome of a Stochastic Epidemic—A Note on Bailey's Paper. *Biometrika*, 42(1/2), 1955.

[24] A. Wilder-Smith and D. O. Freedman. Isolation, quarantine, social distancing and community containment: pivotal role for old-style public health measures in the novel coronavirus (2019-ncov) outbreak. *Journal of Travel Medicine*, 27(2), 2020.

[25] E. Yoneki and J. Crowcroft. Epimap: towards quantifying contact networks for understanding epidemiology in developing countries. *Ad Hoc Networks*, 13:83–93, 2014.

8 Appendix

Table 1: The average runtime and average number of cumulative exposures after 100 simulations of a variety of strand configurations.

Population	Strength	Radius (m)	Runtime (ms)	Cumulative Exposures
100	0.01	5	0.20	10.23
100	0.01	10	0.21	13.45
100	0.01	20	0.26	32.89
100	0.05	5	0.20	12.46
100	0.05	10	0.25	25.00
100	0.05	20	0.25	94.29
100	0.10	5	0.21	14.24
100	0.10	10	0.28	42.63
100	0.10	20	0.23	97.45
200	0.01	5	0.29	23.46
200	0.01	10	0.37	35.65
200	0.01	20	0.47	166.62
200	0.05	5	0.33	36.84
200	0.05	10	0.47	148.63
200	0.05	20	0.29	199.39
200	0.10	5	0.38	47.36
200	0.10	10	0.42	175.55
200	0.10	20	0.27	199.68
500	0.01	5	0.79	79.04
500	0.01	10	1.27	318.62
500	0.01	20	0.75	494.54
500	0.05	5	1.25	249.38
500	0.05	10	0.75	492.89
500	0.05	20	0.44	500.00
500	0.10	5	1.20	353.59
500	0.10	10	0.64	497.20
500	0.10	20	0.41	500.00