National Defence Défense nationale

DEFENCE RESEARCH AND DEVELOPMENT CANADA (DRDC)
RECHERCHE ET DEVELOPPEMENT POUR LA DÉFENSE CANADA (RDDC)

A branching process and simulation model to evaluate the spread of severe acute respiratory syndrome coronavirus 2 (SARS-Cov-2) in various environments

Enabling simultaneous mitigation strategies including social distancing, masks, symptomatic self isolation, testing, contact tracing and vaccination

Katerina Biron*
Pierre-Luc Drouin*
DRDC – Ottawa Research Centre

Lynne Serré
DRDC – Centre for Operational Research and Analysis

* These authors contributed equally to the work.

Defence Research and Development Canada
Scientific Report
DRDC-RDDC-2023-R025
March 2023

Canada

IMPORTANT INFORMATIVE STATEMENTS

This document was reviewed for Controlled Goods by Defence Research and Development Canada (DRDC) using the Schedule to the *Defence Production Act*.

Disclaimer: This publication was prepared by Defence Research and Development Canada, an agency of the Department of National Defence. The information contained in this publication has been derived and determined through best practice and adherence to the highest standards of responsible conduct of scientific research. This information is intended for the use of the Department of National Defence, the Canadian Armed Forces ("Canada") and Public Safety partners and, as permitted, may be shared with academia, industry, Canada's allies, and the public ("Third Parties"). Any use by, or any reliance on or decisions made based on this publication by Third Parties, are done at their own risk and responsibility. Canada does not assume any liability for any damages or losses which may arise from any use of, or reliance on, the publication.

Endorsement statement: This publication has been peer-reviewed and published by the Editorial Office of Defence Research and Development Canada, an agency of the Department of National Defence of Canada. Inquiries can be sent to: Publications.DRDC-RDDC@drdc-rddc.gc.ca.

Abstract

Starting from existing severe acute respiratory syndrome coronavirus 2 (SARS-Cov-2) propagation branching processes, the work presented in this Scientific Report expands on these earlier models through the concept of interaction events, the introduction of additional classes of infected individuals, and the use of a latent phase for the disease. These changes enable the modelling of mitigation measures, including vaccination, and to evaluate their effects on the propagation of SARS-Cov-2 through the addition of a probability of infection for interaction events. Other features were also added to the model to simulate scenarios more realistically, including some that make the model deviate from a branching process, such as the modelling of contact tracing and isolation.

In this Report, a generalized branching process for the model, providing analytical results, is first formally derived, and modifications going beyond a branching process are introduced. A presentation of the input and output parameters for the model follows. The algorithm implementing the model is also provided. Cases studies (impact of the different parameters from the interaction model, the Diamond Princess cruise ship outbreak using the modified branching process and the impact of the parameters from the vaccination model) are then presented for three different applications. The results show that this model can be used in the study of early outbreaks of the coronavirus disease 2019 (COVID-19), and in the study of mitigation strategies against the spread of its virus.

Significance for defence and security

Following the appearance of SARS-Cov-2 and its worldwide propagation, there was a need by the Canadian Armed Forces to evaluate the risks posed by COVID-19 for its operations. The model presented in this Report was developed to perform simulations of SARS-Cov-2 propagation for different scenarios, prior, but also following, the availability of vaccines. The software implementation of the model is particularly useful to evaluate the consequences of one or multiple undetected COVID-19 infected individuals, and, as such, simulation results obtained with the model for specific ship scenarios were briefed to the Commander of the Royal Canadian Navy (RCN) and to the Surgeon General [1, 2], and were used to decide on the pre-embarkation protocols for the RCN. These analyses were done in conjunction with the estimation, by other models, of the probability of missed infections [3], through the evaluation of the prevalence of the virus within the general population [4], but also specifically within the military personnel after following different pre-embarkation protocols. The studies and briefings were performed multiple times, as the virus and the availability of vaccines evolved.

Résumé

Basé sur des processus de ramification existants pour la propagation du SARS-Cov-2 (*severe acute respiratory syndrome coronavirus 2*), le travail présenté dans ce Rapport scientifique généralise ces précédents modèles à travers le concept d'événements d'interactions, l'introduction de classes additionelles d'individus infectés, ainsi que l'utilisation d'une phase latente pour la maladie. Ces changements permettent la modélisation de mesures de mitigation, incluant la vaccination, et d'évaluer leurs effets sur la propagation de SARS-Cov-2 à travers l'ajout d'une probabilité d'infection pour les événements d'interactions. D'autres éléments ont aussi été ajoutés au modèle afin de simuler des scénarios de manière plus réaliste, incluant certains qui font dévier le modèle d'un processus de ramification, tel que la modélisation de la recherche des contacts et l'isolation.

Dans ce rapport, un processus de ramification généralisé pour le modèle, fournissant des résultats analytiques, est d'abord formellement dérivé, et des modifications qui vont au-delà d'un processus de ramification sont introduites. Une présentation des paramètres d'entrée et de sortie pour le modèle suit. L'algorithme mettant le modèle en œuvre est aussi fourni. Des études de cas (impact des différents paramètres du modèle d'interaction, la flambée d'infections sur le navire de croisière Diamond Princess en utilisant le processus de ramification modifié et l'impact des paramètres du modèle pour la vaccination) sont ensuite effectuées pour différentes applications. Les résultats montrent que ce modèle peut être utilisé pour l'étude de flambées initiales de la maladie à coronavirus 2019 (COVID-19), ainsi que pour l'étude de mesures de mitigation contre la propagation du virus.

Importance pour la défense et la sécurité

Suite à l'apparition de SARS-Cov-2 et à sa propagation mondiale, les Forces armées canadiennes ont eu besoin d'évaluer les risques posés par la COVID-19 pour ses opérations. Le modèle présenté dans ce rapport a été développé afin d'effectuer des simulations de propagation du SARS-Cov-2 pour différents scénarios, précédant, mais aussi suivant, la disponibilité de vaccins. La mise en œuvre logicielle de ce modèle est particulièrement utile pour évaluer les conséquences de un ou de plusieurs individus infectés et non détectés de la COVID-19, et, ainsi, les résultats de simulation obtenus avec le modèle pour des scénarios spécifiques à bord de navires ont été présentés au Commandant de la Marine royale canadienne (MRC) ainsi qu'au Médecin-général [1, 2], et ont été utilisés afin de prendre des décisions par rapport aux protocoles de pré-embarquement pour la MRC. Ces analyses ont été effectuées conjointement avec l'estimation, par d'autres modèles, de la probabilité d'infections non détectés [3], par le biais de l'évaluation de la prévalence du virus dans le public [4], mais aussi de façon spécifique dans le personnel militaire suivant différents protocoles de pré-embarquement. Ces études et ces présentations ont été effectuées plusieurs fois suivant l'évolution du virus et de la disponibilité des vaccins.

Table of contents

List of figures

List of tables

1 Introduction

Defence Research and Development Canada (DRDC)'s modelling efforts related to coronavirus disease 2019 (COVID-19), the disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-Cov-2), began in early March 2020 [5]. The first modelling efforts were done by the North American Aerospace Defence Command (NORAD) team in DRDC's Centre for Operational Research and Analysis (CORA) where a network model was developed to study the potential spread of SARS-Cov-2 in NORAD mission critical teams [6]. DRDC's modelling efforts quickly expanded to include support to the Canadian Forces Intelligence Command, Canadian Forces Health Services Group (CF H Svcs Gp), Canadian Joint Operations Command and the Canadian Special Operations Forces Command [7]. The types of epidemiological models being developed by DRDC scientists also grew quickly and include a compartmental susceptible-exposed-infectious-recovered (SEIR)[1] model of large population SARS-Cov-2 spread and the impact of non-pharmaceutical interventions[2] [9], a stochastic compartmental SEIR-style model of SARS-Cov-2 spread between two or more heterogeneous sub-populations [10], and an agent-based model (ABM) from the Public Health Agency of Canada (PHAC) being adapted for Canadian Armed Forces (CAF) scenarios [11].

In spring 2020, DRDC scientists began exploring the use of branching processes to study how likely a single SARS-Cov-2 infection is to lead to an outbreak. For instance, if an individual arrives at an office building infected, how likely is the virus to spread, causing a major disease outbreak? What are the chances that the virus stops spreading, causing a disease outbreak to go extinct? The advantage of a branching process, over other types of models, is its analytical representation of the reproduction number and its variance, which can be indicative of if a disease outbreak is likely to go extinct or not. In fact, the reproduction number represents the number of infections an infected individual produces, on average [12]. A value smaller than one would indicate that disease transmission is low and that the spread of a virus would eventually stop, causing the disease outbreak to go extinct.

Back-to-work policy questions were being asked by the CAF, not limited to traditional office environments. The CAF had to understand the likelihood of a large outbreak for a wide range of environments, such as for military field exercises and naval vessels, where it could be harder to implement physical distancing. This led to the development of a model that can evaluate the risks associated with outbreaks introduced in CAF environments.

Reviewed existing branching processes were limited and could not be applied to conditions experienced within some CAF scenarios. This Scientific Report presents a model that addresses these limitations. A detailed explanation of the constraints with previous branching processes is provided in Section 2.

A generalized branching process that addresses most of these limitations is first formally derived in Sections 3.1 to 3.3, along with its parameters and underlying assumptions. The resulting analytical expressions provided by this branching process represent useful metrics when evaluating a scenario, but additional parameters can also be required. In particular, studying the evolution of an outbreak over time, including its statistical distribution, as well as the probability of extinction, is often desired. An algorithm allowing the simulation of the model was thus developed, which also includes features that can make the model to deviate from a branching process, such as the interruption of transmission through isolation of infected individuals identified through contact tracing. Section 3.4 presents such output parameters

[1] Compartmental epidemic models divide the population under study into compartments with assumptions about the nature and time rate of transfer from one compartment to another, which is often formulated as a system of differential equations [8]. In a SEIR model, the compartments are susceptible (S), exposed (E), infectious (I) and recovered (R).
[2] Non-medical masks, isolation and physical distancing are examples of non-pharmaceutical interventions.

of the model that require simulation to be assessed, and conditions under which the model goes beyond a branching process are presented in Section 3.5. The algorithm itself is available in Annex A, and the software implementation of the model is available online [13].

The parameters for the final model are presented in in Section 4. Section 5 presents different case studies that were performed, illustrating various applications for the model. Finally, Section 6 lists future work and recommendations for the model.

2 Background

Branching processes have been widely applied to study the spread of infectious diseases, including mumps [14], measles [15], Ebola [16], the Middle East respiratory syndrome (MERS) [17], and the severe acute respiratory syndrome (SARS) [17]. This type of branching process, in its simplest form, assumes that the total number of infected individuals at any given time is the sum of independent secondary infections generated by all previously infected individuals [18]. If the average number of secondary infections is above one (1), the branching process grows the number of new infections exponentially; otherwise, below one (1), the branching process slowly grows and may even stop producing new infections [19]. In epidemiology, the average number of secondary infections produced by a typical case of one infection introduced in a large population where everyone is susceptible is referred to as the *basic reproduction number* (R_0) [19, 20]. In a population, some may not be susceptible, and mitigation strategies can reduce infections. The *effective reproductive number* (R_{eff}) is the average number of secondary infections per infectious individual in a population where some mitigation measures are in place.

Several branching processes have been developed to study COVID-19 outbreaks and to study the impact of mitigation policies. Some of the earliest examples focused on applying branching processes to study the probability of major outbreaks as SARS-Cov-2 began entering new countries via international travellers [21–23] and to project future case counts while still in the early stages of community transmission [24]. More recent examples include the application of branching processes to investigate containment and elimination scenarios for SARS-Cov-2 as the severity of population-wide control measures are eased [25], to understand the impact of digital contact tracing systems[3] on reducing transmission [27], and to explore the potential of combining backward contact tracing[4] with more conventional forward contact tracing for controlling transmission [28].

Branching processes have been shown to do well in approximating how an infectious disease spreads when a population is homogenously mixing[5] and the number of infectious individuals is small compared to the total size of the susceptible population [14]. The latter condition is what makes branching processes well suited for studying the early stages of an outbreak. This is in contrast to deterministic compartmental models, which assume that a disease outbreak has already become established[6] and are therefore not applicable at the beginning of an outbreak [8].

Beginning with a single infected individual, Levesque et al.'s [29] model used a Crump-Mode-Jagers (CMJ) branching process to model the propagation of SARS-Cov-2. Generally speaking, the CMJ branching process evaluates the reproduction and growth of populations, where each individual lives for a random lifetime (distributed according to a random variable) and reproduces at intervals according to a point process [14]. The model developed by Levesque et al. [29] evaluates the reproduction of SARS-Cov-2 and growth of the COVID-19 disease, where the random lifetime represents an infected individual's infectious period (i.e., the time during which they can transmit the virus to others) and reproduction occurs through *transmission events* (i.e., when new individuals get infected). Highlighting the design philosophy of parsimony by these authors, the model has just four parameters:

[3] The Government of Canada's COVID Alert app [26] is an example of a digital contact tracing system.
[4] For a confirmed index case, backward contact tracing aims to identify who infected the index case while forward contact tracing aims to identify who the index case may have infected [28].
[5] A homogenously mixing population is one where contact between any two individuals occurs randomly with equal probability.
[6] Deterministic compartmental models are formulated in terms of the derivatives of the sizes of each compartment; this assumes that the number of individuals in a compartment is a differentiable function of time, which is only a valid assumption once a disease outbreak has progressed beyond the early stages [8].

- one to specify the mean number of infected people per transmission event;
- one to specify the average arrival rate (per day) of transmission events; and
- two to specify the distribution (mean and variance) of the infectious period of an individual.

This produces a simple representation of the infection process: the number of secondary infections (or reproduction number) depends on the number of transmission events (which itself depends on the average arrival rate multiplied by the duration of the infectious phase) and the number of individuals infected at each event. Events in their formulation can be interpreted as meetings or gatherings, which offers a direct physical interpretation and the parameters decision-makers have control over (e.g., they can reduce the frequency or size of meetings). The model can also simulate the probability of extinction of an outbreak, average outbreak size at extinction, and the exponential growth rate (Malthusian parameter). Along with the properties of those outbreaks (e.g., size, growth rate, or extinction time), the model can help military and civilian leaders in the CAF and Department of National Defence (DND) to understand the risks if a new individual is introduced into an otherwise infection-free work environment.

The model by Levesque et al. was inspired by a branching process developed by Hellewell et al. [22]. In Hellewell et al., the number of secondary infections produced by each infectious individual is drawn from a negative binomial distribution with a mean equal to the basic reproduction number and the time of infection for each secondary infection is drawn from a serial interval distribution.[7] Several other branching process studies, such as [21,23,24] and [28], also use a negative binomial distribution based on the reproduction number. The model by Levesque et al. is also based on a negative binomial process, but unlike the Hellewell et al. model, it does not simulate SARS-Cov-2 propagation in the presence of contact tracing that interrupts the branching process when contacts are successfully isolated. Rather, Levesque et al. extended their original model to study the impact of an alternate infectious period, where such alternate infectious period could represent the infectious period of contacted and/or isolated individuals. Their extended model requires additional, but minimal, parameters to specify the probability of occurence for the alternate infectious period and the distribution of the alternate infectious period. Since there is no interruption in the branching process, this allows for an analytical expression for the reproduction number and its variance, given two infectious periods: one for the disease infectious period and another for the alternate infectious period. An analytical representation of the reproduction number and its variance provides insight into the likelihood that a disease outbreak goes extinct.

While the analytical representation from Levesque et al. offers interpretative and computational advantages, some of the disease characteristics were simplified. In particular, individuals are assumed to be infectious as soon as they are infected. That is, in this model, transmission and infectious events are synonymous: all individuals at an event become infected and can infect others immediately. In reality, a latent period exists: there is a delay between the time a person is infected with SARS-Cov-2 (exposed) and is able to infect others (infectious). This simplification is not uncommon, the susceptible-infectious-recovered (SIR) model is one of the most commonly used frameworks for epidemiological systems [31]. However, excluding the latent period or choosing a poor distributional representation of the latent period has been shown to underestimate the basic reproductive number of an infectious disease from outbreak data [31]. Still, parsimonious models, such as the SIR model and basic branching processes, are "particularly well suited to isolating key features of the pandemic and

[7] In Hellewell et al., the serial interval represents the time between an infector becoming infectious and an associated infectee becoming infectious. In other studies, such as Zhao [30], this time period is known as the time interval between the transmission generations and the serial interval instead represents the time between the onset of symptoms in an infector and the onset of symptoms in an associated infectee.

to developing policy-relevant insights" [32]. However, including an exposed period in a compartmental model decreases the initial exponential growth rate of the outbreak [8]. Depending on the application, and outcomes of importance to decision makers, the exclusion of a latent period can be a limitation of the model.

The exclusion or inclusion of a latent period also has implications on modelling the impact of contact tracing on controlling transmission. For instance, if an exposed individual is traced and contacted before they become infectious, isolation prevents onward transmission. If an infectious individual is traced and contacted before they exhibit symptoms, then isolation minimizes onward transmission.[8] In some studies, such as Hellewell et al. [22] and James et al. [33], individuals are assumed to isolate upon symptom onset (with some delay) whether contact tracing is successful or not. However, not all infected individuals develop symptoms (i.e., some individuals are asymptomatic).[9] Hellewell et al. [22] found that outbreak control is more difficult to achieve when the delay between symptom onset and isolation increases, the proportion of pre-symptomatic transmission increases, or the proportion of asymptomatic infections increases [22]. James et al. [33] additionally considered imperfect isolation (e.g., individuals may not comply fully with the isolation) and found that the effectiveness of isolation is a crucial determinant of the ability of a contact tracing system to reduce the reproductive number. The branching process of James et al. [33] includes a detailed model of symptom onset, testing, contact tracing and isolation. This level of depth is beyond the scope of the model by Levesque et al., which can only be used to provide a broad sense of the impact of a single isolation strategy that shortens the infectious period (e.g., due to symptom onset or contact tracing, but not both) by identifying an appropriate distribution for the alternate infectious period.

[8] For context, Zhao [30] approximated the COVID-19 latent period to have a mean duration of 3.3 days and the pre-symptomatic infectious period to have a mean duration of 2.2 days.

[9] For context, some DRDC studies have modelled CAF scenarios assuming 20% [34] to 40% [35] of individuals are asymptomatic.

3 Model

A branching process, in its simplest form, can assume that the total number of infected individuals is the sum of independent secondary infections generated by all previously infected individuals [18] (as depicted in Figure 1). If the average number of secondary infections generated by an infectious individual is less than one, the infections will eventually go extinct as a sufficiently large fraction of infectious individuals will not produce secondary infections.

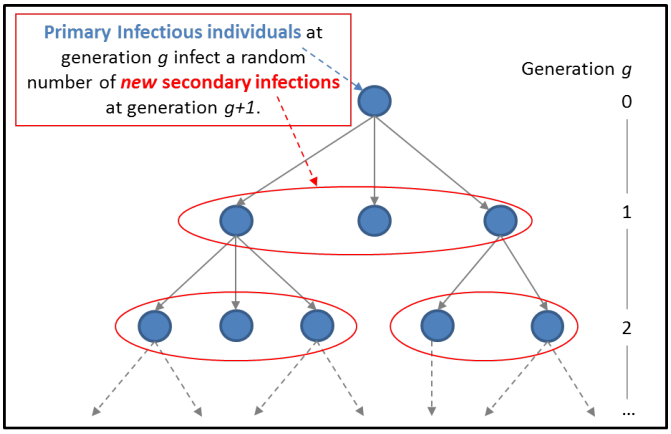

Figure 1: *A simple branching process that starts with one infectious individual at generation $g = 0$. Each primary infectious individual at generation g will generate a random number of new secondary infections at generation $g + 1$.*

One advantage of the branching process is the analytical representation for the average number of secondary infections, or reproduction number R, and its variance which provides insight into the probability that an outbreak self-extinguishes. In fact, it has been demonstrated that the average number of secondary infections in a COVID-19 outbreak can be represented as the multiplication of the average number of transmission events (i.e., where new individuals get infected), the average number of infected individuals at each event, and the average infectious period [29, 36]. The impact of these factors on the reproduction number can be evaluated using the analytical representation, without the need to compute a large number of simulations (as an R below one [1] indicates that an outbreak will, on average, go extinct).

The analytical representation of the reproduction number gives insight on the average outcome of the disease for each infectious individual, but it may be desirable to study the impact of the progression of the outbreak over time. This is especially true in CAF environments where operations are time sensitive and populations are isolated (such as on vessels). To produce an appropriate analysis of the progression of the outbreak over time, a latent period was added to the model for more accurate time scaling of the infectious events. This not only allows for the evaluation of different time-dependent metrics, but also for the modelling of testing and contact tracing by an interrupted branching process.

A branching process is the most useful to study early outbreaks or whenever the number of infections is small relatively to the number of susceptible individuals (which includes when there is fast reinfection due to a lack of immunity). This is due to the assumption of branching processes that the statistical distributions they use are static over time. As an outbreak deviates away from this assumption, a branching process will tend to overestimate the number of infections, given a set of input parameters that reflect the conditions at the beginning of the outbreak. For scenarios where such an assumption is not appropriate (an outbreak in a late stage, or a scenario involving a very small population, for example), a model which does consider the finite size of the population should be considered.

This work modifies the branching process by Levesque et al. [29, 36] so that transmission occurs during interaction events, where not all attendees may become infected. This allows for the representation of contacts and the introduction of a probability of infection. As vaccination rolled out, the CAF had an urgent need to examine the impact that vaccination could have on SARS-Cov-2 transmission. The model was thus further modified to incorporate vaccination. For clarity, the transmission branching process (TBP) refers to the model presented by Levesque et al. [29], whereas the interaction branching process (IBP) and the vaccination branching process (VBP) refer to models presented in this work, all of which can be derived and represented analytically.

The derivation of the analytical expressions for the reproduction number and its relative variance for the TBP, the IBP and the VBP are presented in Sections 3.1, 3.2 and 3.3, respectively. These expressions show the impact of different variables and mitigation strategies on the reproduction number. Time scaling of the model is presented in Section 3.4, and modifications of the model beyond a branching process are finally presented in Section 3.5.

3.1 Expansion of the transmission branching process

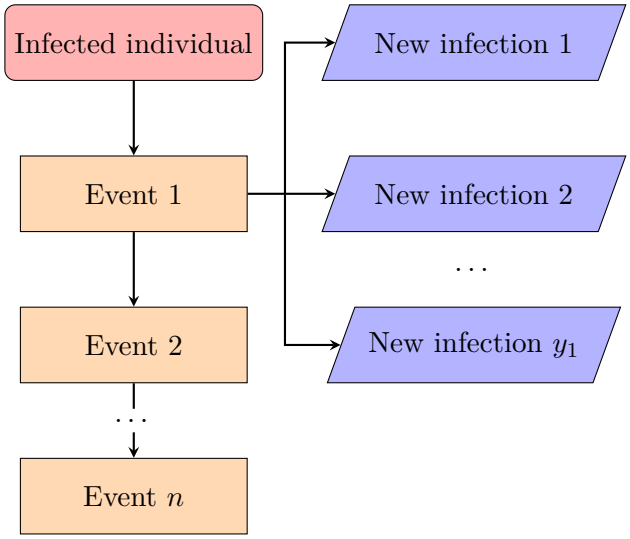

Figure 2: *Infection of susceptible individuals during one event with the original transmission branching process, where n is a realization for $N(\lambda t)$ and y_1 a realization for Y_1.*

In this section, the model of the TBP is rederived and expressed using a format more suitable for further expansion into the IBP.

In the TBP [36], a compound Poisson process with arrival rate λ and of duration t is defined, with $Q(\lambda t)$ expressed as:

$$Q(\lambda t) = \sum_{i=1}^{N(\lambda t)} Y_i, \tag{1}$$

where $N(\lambda t) \sim \text{Pois}(\lambda t)$ is the number of events from the Poisson process and where Y_i is a random variable for the number of infections occurring during event i (Figure 2). The $Q(\lambda t)$ random variable thus represents the total number of infections generated by an infectious individual through a total of $N(\lambda t)$ transmission events. The probability mass function (PMF) for $Q(\lambda t)$, is written as $P_{Q(\lambda t)}(q)$. The compound Poisson process is then subordinated with a gamma process which has the probability density function (PDF):

$$f_{T(a,b)}(t) = \frac{b^a}{\Gamma(a)} t^{a-1} e^{-bt} \tag{2}$$

and a mean of $\bar{t} = a/b$. Let $Z(\lambda, a, b) \equiv Q(\lambda T(a,b))$, which is defined as the number of new infections, where the duration t from the Poisson process has been smeared by the gamma process with parameters a and b. The probability density for Z is then,

$$\begin{aligned} P_{Z(\lambda,a,b)}(z) &= \int_0^\infty P_{Z(\lambda,a,b),T(a,b)}(z,t')dt' \\ &= \int_0^\infty P_{Z(\lambda,a,b)|T(a,b)=t'}(z) f_{T(a,b)}(t')dt' \\ &= \int_0^\infty P_{Q(\lambda t')}(z) f_{T(a,b)}(t')dt'. \end{aligned} \tag{3}$$

The characteristic function for $Z(\lambda, a, b)$ can thus be expressed as:

$$\begin{aligned} \mathbb{E}\left[e^{iuZ(\lambda,a,b)}\right] &= \sum_z e^{iuz} P_{Z(\lambda,a,b)}(z) \\ &= \int_0^\infty \sum_z e^{iuz} P_{Q(\lambda t')}(z) f_{T(a,b)}(t')dt \\ &= \int_0^\infty \mathbb{E}\left[e^{iuQ(\lambda t')}\right] f_{T(a,b)}(t')dt'. \end{aligned} \tag{4}$$

The probability mass function for $Q(\lambda t)$ can be written as:

$$P_{Q(\lambda t)}(q) = \sum_n P_{Q(\lambda t)|N(\lambda t)=n}(q) P_{N(\lambda t)}(n), \tag{5}$$

such that its characteristic function, assuming independent and identically-distributed Y_1, Y_2, …, Y_n variables, is given by:

$$\begin{aligned} \mathbb{E}\left[e^{iuQ(\lambda t)}\right] &= \sum_q e^{iuq} P_{Q(\lambda t)}(q) \\ &= \sum_n \sum_q e^{iuq} P_{Q(\lambda t)|N(\lambda t)=n}(q) P_{N(\lambda t)}(n) \\ &= \sum_n \sum_{Y_1} \cdots \sum_{Y_n} e^{iu\sum_{i=1}^n Y_i} P_{Y_1,\ldots,Y_n}(y_1,\ldots,y_n) P_{N(\lambda t)}(n) \\ &= \sum_n \mathbb{E}\left[e^{iuY}\right]^n P_{N(\lambda t)}(n) \\ &= \sum_n \mathbb{E}\left[e^{iuY}\right]^n \frac{e^{-\lambda t}(\lambda t)^n}{n!} \\ &= e^{\lambda t\left(\mathbb{E}\left[e^{iuY}\right]-1\right)}. \end{aligned} \tag{6}$$

Using Equations (2), (4) and (6),

$$\mathbb{E}\left[e^{iuZ(\lambda,a,b)}\right] = \left[1 + \frac{\lambda}{b}\left(1 - \mathbb{E}\left[e^{iuY}\right]\right)\right]^{-a}. \tag{7}$$

The characteristic function for $Z(\lambda, a, b)$ is then used to compute the expected value and the relative variance of the resulting smeared Poisson process:

$$\mathbb{E}\left[Z(\lambda, a, b)\right] = \lambda \frac{a}{b} \mathbb{E}\left[Y\right] = \lambda \bar{t} \mathbb{E}\left[Y\right] \tag{8}$$

$$\begin{aligned} \frac{\mathrm{Var}\left[Z(\lambda, a, b)\right]}{\mathbb{E}\left[Z(\lambda, a, b)\right]^2} &= \frac{1}{a} + \frac{1}{\lambda \bar{t}} \left[1 + \frac{\mathrm{Var}\left[Y\right]}{\mathbb{E}\left[Y\right]^2}\right] \\ &= \frac{\mathrm{Var}\left[T\right]}{\mathbb{E}\left[T\right]^2} + \frac{\mathrm{Var}\left[N(\lambda \mathbb{E}\left[T\right])\right]}{\mathbb{E}\left[N(\lambda \mathbb{E}\left[T\right])\right]^2} + \frac{1}{\mathbb{E}\left[N(\lambda \mathbb{E}\left[T\right])\right]} \frac{\mathrm{Var}\left[Y\right]}{\mathbb{E}\left[Y\right]^2}. \end{aligned} \tag{9}$$

The two above expressions thus provide a reproduction number and its relative variance for one infectious individual, as a function of the mean and variance for the number of infections occurring during an event (Y). These expressions are equivalent to the ones derived in the TBP [36], but here the expression for the variance has been rearranged to show the contribution from the variance of each component, that is the contributions from the relative variances on the communicable period (the period during which an individual is infectious and without being isolated), on the number of events for the average period, and on the number of new infections per event.

Expression (9) provides insight into the probability of extinction for an outbreak (given the model that has been described so far). For an outbreak to go extinct, infectious individuals need to eventually cease to generate any new infection. Considering one infectious individual, extinction will most likely occur when either the reproduction number is sufficiently small (primary factor), or when the relative variance (Expression (9)) is close to one. Expression (9) indicates that the latter can occur when either the relative variance on the communicable period is significant (the duration of the communicable period is likely to be negligible compared to its mean), the relative variance on the number of events is significant (the possibility of zero transmission event is likely), and/or when the relative variance on the average number of new infections per event is significant. These cases are represented by the three terms in (9), respectively.

If one wants to define mutually exclusives classes of infectious individuals composing the population, the characteristic function for the total number of new infections W generated by an unspecified infectious individual can be expressed as a function of the characteristic functions of each class. Defining C as a random variable representing the class of a given infectious individual,

$$P_W(w) = \sum_c P_C(c) P_{W|C=c}(w), \quad \sum_c P_C(c) = 1 \tag{10}$$

$$\mathbb{E}\left[e^{iuW}\right] = \sum_w P_W(w) e^{iuw} = \sum_c P_C(c) \mathbb{E}_{W|C=c}\left[e^{iuW}\right] \tag{11}$$

$$\mathbb{E}\left[W\right] = \sum_c P_C(c) \mathbb{E}_{W|C=c}\left[W\right] \tag{12}$$

$$\frac{\mathrm{Var}\left[W\right]}{\mathbb{E}\left[W\right]^2} = \frac{\sum_c P_C(c) \left[\frac{\mathrm{Var}_{W|C=c}[W]}{\mathbb{E}_{W|C=c}[W]^2} + 1\right] \mathbb{E}_{W|C=c}\left[W\right]^2}{\left[\sum_c P_C(c) \mathbb{E}_{W|C=c}\left[W\right]\right]^2} - 1. \tag{13}$$

Let a pair of such classes of infectious individuals be defined as individuals that do not self isolate and that do self isolate (following an infection), with a probability q of self isolation. If these two classes of infectious individuals are differentiated through the parameters of their time gamma process, where the parameters (a, b) with a mean $\bar{t} = a/b$ and the parameters (a', b') with a mean $\bar{m} = a'/b'$ represent the

individuals that do not self isolate and those that self isolate, respectively, then the characteristic function of $W(\lambda, q, a, b, a', b')$ for any infectious individual can be expressed as:

$$\mathbb{E}\left[e^{iuW(\lambda,q,a,b,a',b')}\right] = (1-q)\mathbb{E}\left[e^{iuZ(\lambda,a,b)}\right] + q\mathbb{E}\left[e^{iuZ(\lambda,a',b')}\right]. \tag{14}$$

Using this function to calculate the expected value and the relative variance of the resulting process leads to:

$$\mathbb{E}\left[W(\lambda, q, a, b, a', b')\right] = \lambda\left[(1-q)\bar{t} + q\bar{m}\right]\mathbb{E}\left[Y\right] = \lambda\mathbb{E}\left[\mathcal{T}\right]\mathbb{E}\left[Y\right] \tag{15}$$

$$\begin{aligned}\frac{\mathrm{Var}\left[W(\lambda, q, a, b, a', b')\right]}{\mathbb{E}\left[W(\lambda, q, a, b, a', b')\right]^2} &= \frac{1}{\mathbb{E}\left[\mathcal{T}\right]^2}\left[(1-q)\frac{\bar{t}^2}{a} + q\frac{\bar{m}^2}{a'} + q(1-q)(\bar{t}-\bar{m})^2\right] + \\ &\quad \frac{1}{\lambda\mathbb{E}\left[\mathcal{T}\right]}\left[1 + \frac{\mathrm{Var}\left[Y\right]}{\mathbb{E}\left[Y\right]^2}\right] \\ &= \frac{\mathrm{Var}\left[\mathcal{T}\right]}{\mathbb{E}\left[\mathcal{T}\right]^2} + \frac{\mathrm{Var}\left[N(\lambda\mathbb{E}\left[\mathcal{T}\right])\right]}{\mathbb{E}\left[N(\lambda\mathbb{E}\left[\mathcal{T}\right])\right]^2} + \frac{1}{\mathbb{E}\left[N(\lambda\mathbb{E}\left[\mathcal{T}\right])\right]}\frac{\mathrm{Var}\left[Y\right]}{\mathbb{E}\left[Y\right]^2},\end{aligned} \tag{16}$$

where $\mathcal{T}(q, a, b, a', b')$ is defined as the random variable representing the duration of an individual's infectious period:

$$P_{\mathcal{T}(q,a,b,a',b')}(\tau) = (1-q)P_{T(a,b)}(\tau) + qP_{T(a',b')}(\tau) \tag{17}$$

$$\mathbb{E}\left[\mathcal{T}(q, a, b, a', b')\right] = (1-q)\mathbb{E}\left[T(a, b)\right] + q\mathbb{E}\left[T(a', b')\right] \tag{18}$$

$$\begin{aligned}\mathrm{Var}\left[\mathcal{T}(q, a, b, a', b')\right] &= (1-q)\mathrm{Var}\left[T(a, b)\right] + q\mathrm{Var}\left[T(a', b')\right] + \\ &\quad q(1-q)\left\{\mathbb{E}\left[T(a, b)\right] - \mathbb{E}\left[T(a', b')\right]\right\}^2.\end{aligned} \tag{19}$$

In the above expressions, $T(a, b)$ and $T(a', b')$ are gamma-distributed durations for the two classes of infectious individuals. The variance on $\mathcal{T}(q, a, b, a', b')$ is expressed as a sum of contributions from the separate variances on $T(a, b)$ and $T(a', b')$, and from the discrepancy of their means.

Expressions (15) to (16) thus correspond to (8) to (9) (the reproduction number and its relative variance for one infectious individual), respectively, when a given infectious individual is randomly drawn from two classes with different durations. Note that the latter expressions reduce to the former ones when $q = 0$, $q = 1$ or $(a', b') = (a, b)$, as expected. As it was shown previously for (9), (16) was written as a sum of contributions from the relative variances on the communicable period, on the number of events for the average communicable period and on the number of new infections for each event. The Expression (16) is the same as (9), except for the random communicable $\mathcal{T}(q, a, b, a', b')$ replacing $T(a, b)$.

If $T(a, b)$ and $T(a', b')$ provide the communicable periods for individuals that do not and that do self isolate, respectively, as a mitigation measure, then (15) can provide the expression for an effective reproduction number. This expression is simply the product of the Poisson rate λ with the average duration $\mathbb{E}\left[\mathcal{T}\right]$ and the average number of new infections per event $\mathbb{E}\left[Y\right]$, as $\mathcal{T}$ and Y are independent.

3.2 Interaction branching process

In the TBP described in [36], the number of new infections Y occurring during one event is provided directly by a logarithmic distribution. Although a minimal set of parameters is required (which is desirable when little information is available), it does not allow for the modelling of group dynamic, the modelling of contact tracing, or for the modelling of measures that affect the probability of infection (like mask wearing or vaccination, for example).

In the IBP, the distribution for Y is thus modified to decouple infections from interactions, allowing for the representation of contacts (where not all may become infected) and the introduction of a probability of infection. N_s is defined as the random number of susceptible individuals attending an event where one and only one infectious individual is also attending (Figure 3). In contrast with the TBP, where the number of infected individuals was drawn from a logarithmic distribution, the total number of individuals attending an event $(N_\mathrm{s}+1)$ does not necessarily follow such a distribution.

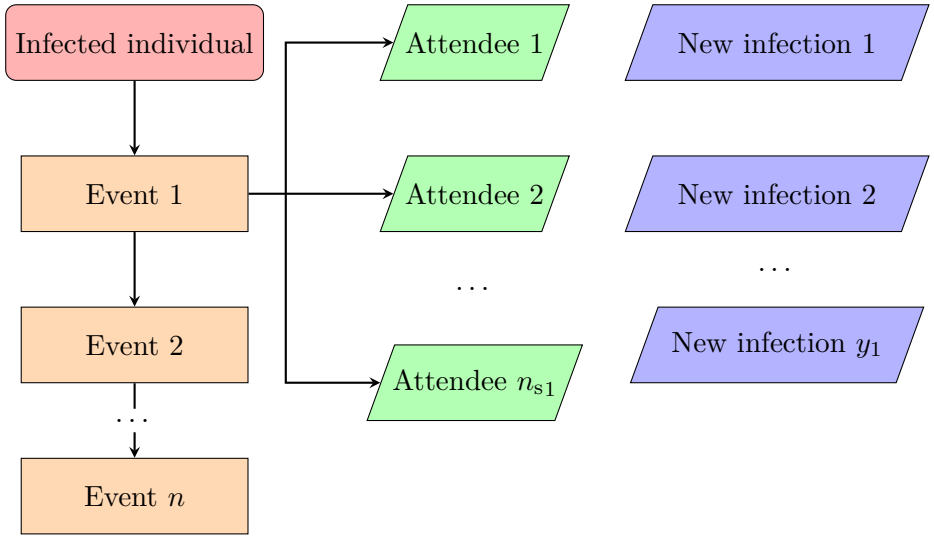

Figure 3: *Infection of susceptible individuals during one event with the interaction branching process, where n, $n_{\mathrm{s}1}$ and y_1 are respectively realizations for N, $N_{\mathrm{s}1}$ and Y_1. The random variable Y_1 is distributed according to a binomial distribution $B(n_{\mathrm{s}1}, p_\mathrm{i})$.*

Given a probability of infection p_i of each attending susceptible individual (by the infectious individual), the probability of Y new infections during an event with n_s susceptible individuals follows a binomial distribution and is thus given by:

$$P_{Y(p_\mathrm{i})|N_\mathrm{s}=n_\mathrm{s}}(y) = \binom{n_\mathrm{s}}{y} {p_\mathrm{i}}^y (1-p_\mathrm{i})^{n_\mathrm{s}-y}, \tag{20}$$

which has a characteristic function of,

$$\mathbb{E}_{Y(p_\mathrm{i})|N_\mathrm{s}=n_\mathrm{s}}\left[e^{iuY}\right] = \left[1+p_\mathrm{i}(e^{iu}-1)\right]^{n_\mathrm{s}}. \tag{21}$$

$P_{Y(p_\mathrm{i})}$ can then be expressed as:

$$P_{Y(p_\mathrm{i})}(y) = \sum_{n_\mathrm{s}} P_{Y(p_\mathrm{i})|N_\mathrm{s}=n_\mathrm{s}}(y) P_{N_\mathrm{s}}(n_\mathrm{s}). \tag{22}$$

The characteristic function for Y is thus,

$$\mathbb{E}_{Y(p_\mathrm{i})}\left[e^{iuY}\right] = \sum_{n_\mathrm{s}} P_{N_\mathrm{s}}(n_\mathrm{s}) \left[1+p_\mathrm{i}(e^{iu}-1)\right]^{n_\mathrm{s}}, \tag{23}$$

which gives:

$$\mathbb{E}\left[Y(p_\mathrm{i})\right] = p_\mathrm{i}\mathbb{E}\left[N_\mathrm{s}\right] \tag{24}$$

$$\begin{aligned}\frac{\mathrm{Var}\left[Y(p_\mathrm{i})\right]}{\mathbb{E}\left[Y(p_\mathrm{i})\right]^2} &= \frac{\mathrm{Var}\left[N_\mathrm{s}\right]}{\mathbb{E}\left[N_\mathrm{s}\right]^2} + \frac{1}{\mathbb{E}\left[N_\mathrm{s}\right]}\frac{1-p_\mathrm{i}}{p_\mathrm{i}} \\ &= \frac{\mathrm{Var}\left[N_\mathrm{s}\right]}{\mathbb{E}\left[N_\mathrm{s}\right]^2} + \frac{\mathrm{Var}_{Y(p_\mathrm{i})|N_\mathrm{s}=\mathbb{E}[N_\mathrm{s}]}\left[Y(p_\mathrm{i})\right]}{\mathbb{E}_{Y(p_\mathrm{i})|N_\mathrm{s}=\mathbb{E}[N_\mathrm{s}]}\left[Y(p_\mathrm{i})\right]^2}.\end{aligned} \tag{25}$$

The relative variance on the number of new infections was thus expressed as a function of the contributions from the relative variances on the number of susceptible attendees and on the number of new infections for the average number of susceptible attendees.

Define $U(\lambda, q, a, b, a', b', p_\mathrm{i})$ as the number of new infections generated from an infectious individual, when the number of new infections per event Y has been expressed as a variable depending on a number of susceptible attendees N_s and a probability of infection p_i. The expected value and the relative variance on U is then,

$$\mathbb{E}\left[U(\lambda, q, a, b, a', b', p_\mathrm{i})\right] = \lambda\left[(1-q)\bar{t} + q\bar{m}\right]p_\mathrm{i}\mathbb{E}\left[N_\mathrm{s}\right] = \lambda\mathbb{E}\left[\mathcal{T}\right]p_\mathrm{i}\mathbb{E}\left[N_\mathrm{s}\right] \tag{26}$$

$$\begin{aligned}\frac{\mathrm{Var}\left[U(\lambda, q, a, b, a', b', p_\mathrm{i})\right]}{\mathbb{E}\left[U(\lambda, q, a, b, a', b', p_\mathrm{i})\right]^2} &= \frac{1}{\mathbb{E}\left[\mathcal{T}\right]^2}\left[(1-q)\frac{\bar{t}^2}{a} + q\frac{\bar{m}^2}{a'} + q(1-q)(\bar{t}-\bar{m})^2\right] + \\ &\quad \frac{1}{\lambda\mathbb{E}\left[\mathcal{T}\right]}\left[1 + \frac{\mathrm{Var}\left[N_\mathrm{s}\right]}{\mathbb{E}\left[N_\mathrm{s}\right]^2} + \frac{1}{\mathbb{E}\left[N_\mathrm{s}\right]}\frac{1-p_\mathrm{i}}{p_\mathrm{i}}\right]\end{aligned} \tag{27}$$

$$\begin{aligned}&= \frac{\mathrm{Var}\left[\mathcal{T}\right]}{\mathbb{E}\left[\mathcal{T}\right]^2} + \frac{\mathrm{Var}\left[N(\lambda\mathbb{E}\left[\mathcal{T}\right])\right]}{\mathbb{E}\left[N(\lambda\mathbb{E}\left[\mathcal{T}\right])\right]^2} + \frac{1}{\mathbb{E}\left[N(\lambda\mathbb{E}\left[\mathcal{T}\right])\right]}\left[\frac{\mathrm{Var}\left[N_\mathrm{s}\right]}{\mathbb{E}\left[N_\mathrm{s}\right]^2} + \right. \\ &\quad \left.\frac{\mathrm{Var}_{Y(p_\mathrm{i})|N_\mathrm{s}=\mathbb{E}[N_\mathrm{s}]}\left[Y(p_\mathrm{i})\right]}{\mathbb{E}_{Y(p_\mathrm{i})|N_\mathrm{s}=\mathbb{E}[N_\mathrm{s}]}\left[Y(p_\mathrm{i})\right]^2}\right].\end{aligned} \tag{28}$$

Expressions (26) to (28) correspond to (15) to (16) (the reproduction number and its relative variance for one infectious individual), respectively, when an infectious individual is randomly drawn amongst two classes with different time duration distributions and transmission events are replaced by interaction events where each susceptible attendee may get infected with a given probability of infection.

3.3 Vaccination branching process

The model presented in the previous sections can also be extended in order to study the effects of vaccination. The following additional parameters are first defined:

f_v: Fraction of the population that has been vaccinated.

$r_{\mathrm{v,inf}}$: Scaling factor, between 0 and 1, for the infection probability of a given vaccinated susceptible individual by a given infectious individual during one event. This parameter captures the vaccine's efficacy against (symptomatic and asymptomatic) infection. A value of 1 indicates that vaccinated and non-vaccinated susceptible individuals are equally likely to become infected, whereas a value of 0 indicates that vaccinated susceptible individuals cannot be infected.

$r_{\mathrm{v,tr}}$: Scaling factor, between 0 and 1, for the infection probability of a given susceptible individual by a given vaccinated infectious individual during one event. This parameter captures the vaccine's efficacy against onward transmission. A value of 1 indicates that vaccinated and non-vaccinated infectious individuals are equally likely to transmit the virus, whereas a value of 0 indicates that vaccinated infectious individuals do not transmit the virus.

q_{v}: Fraction of infectious vaccinated individuals that self isolate.

In order to model vaccination, the definition of the Y random variable must be once again modified to depend on additional random variables. If $V(f_{\mathrm{v}})$ is defined as the number of vaccinated susceptible attendees at an event, given a known number of susceptible attendees, its conditional PMF on N_{s} can be assumed to follow a binomial distribution:

$$P_{V(f_{\mathrm{v}})|N_{\mathrm{s}}=n_{\mathrm{s}}}(v) = \binom{n_{\mathrm{s}}}{v} {f_{\mathrm{v}}}^{v}(1-f_{\mathrm{v}})^{n_{\mathrm{s}}-v}. \tag{29}$$

The total number of new infections Y for an event is then split between the number of new infections Y_{v} for vaccinated susceptible individuals and the number of new infections $Y_{\mathrm{n-v}}$ for non-vaccinated individuals (Figure 4). If ρ_{i} is defined to represent the probability of infection of a non-vaccinated susceptible attendee from the infectious individual (without specifying at this point whether or not the infectious attendee is vaccinated), Y can be expressed as the following expression, based on the above definitions and assuming that $r_{\mathrm{v,inf}}$ has the same value regardless of the vaccination status of the infectious individual:

$$Y(f_{\mathrm{v}}, \rho_{\mathrm{i}}, r_{\mathrm{v,inf}}) = Y_{\mathrm{v}}(f_{\mathrm{v}}, \rho_{\mathrm{i}} r_{\mathrm{v,inf}}) + Y_{\mathrm{n-v}}(f_{\mathrm{v}}, \rho_{\mathrm{i}}). \tag{30}$$

Given a known number of susceptible attendees and a known number of vaccinated individuals amongst them, Y_{v} and $Y_{\mathrm{n-v}}$ can be assumed to follow binomial distributions:

$$P_{Y_{\mathrm{v}}(f_{\mathrm{v}},\rho_{\mathrm{i}} r_{\mathrm{v,inf}})|V(f_{\mathrm{v}})=v}(y_{\mathrm{v}}) = \binom{v}{y_{\mathrm{v}}} (\rho_{\mathrm{i}} r_{\mathrm{v,inf}})^{y_{\mathrm{v}}} (1-\rho_{\mathrm{i}} r_{\mathrm{v,inf}})^{v-y_{\mathrm{v}}} \tag{31}$$

$$P_{Y_{\mathrm{n-v}}(f_{\mathrm{v}},\rho_{\mathrm{i}})|N_{\mathrm{s}}-V(f_{\mathrm{v}})=n_{\mathrm{s}}-v}(y_{\mathrm{n-v}}) = \binom{n_{\mathrm{s}}-v}{y_{\mathrm{n-v}}} (\rho_{\mathrm{i}})^{y_{\mathrm{n-v}}} (1-\rho_{\mathrm{i}})^{n_{\mathrm{s}}-v-y_{\mathrm{n-v}}}. \tag{32}$$

As Y_{v} and $Y_{\mathrm{n-v}}$ are independent when conditioned on N_{s} and $V(f_{\mathrm{v}})$, the characteristic function for $Y(f_{\mathrm{v}}, \rho_{\mathrm{i}}, r_{\mathrm{v,inf}})$, when conditioned on the same variables is simply the product of the conditioned characteristic functions for Y_{v} and $Y_{\mathrm{n-v}}$:

$$\mathbb{E}_{Y(f_{\mathrm{v}},\rho_{\mathrm{i}},r_{\mathrm{v,inf}})|N_{\mathrm{s}}=n_{\mathrm{s}},V(f_{\mathrm{v}})=v}\left[e^{iuY}\right] = \left[1+\rho_{\mathrm{i}} r_{\mathrm{v,inf}}(e^{iu}-1)\right]^{v} \left[1+\rho_{\mathrm{i}}(e^{iu}-1)\right]^{n_{\mathrm{s}}-v}. \tag{33}$$

The characteristic function for $Y(f_{\mathrm{v}}, \rho_{\mathrm{i}}, r_{\mathrm{v,inf}})$ conditioned only on n_{s} can then be expressed as:

$$\begin{aligned}
\mathbb{E}_{Y(f_{\mathrm{v}},\rho_{\mathrm{i}},r_{\mathrm{v,inf}})|N_{\mathrm{s}}=n_{\mathrm{s}}}\left[e^{iuY}\right] &= \sum_{y=0}^{n_{\mathrm{s}}} \sum_{v=0}^{n_{\mathrm{s}}} P_{V(f_{\mathrm{v}})|N_{\mathrm{s}}=n_{\mathrm{s}}}(v) P_{Y(\rho_{\mathrm{i}},r_{\mathrm{v,inf}})|N_{\mathrm{s}}=n_{\mathrm{s}},V(f_{\mathrm{v}})=v}(y) e^{iuy} \\
&= \sum_{v=0}^{n_{\mathrm{s}}} P_{V(f_{\mathrm{v}})|N_{\mathrm{s}}=n_{\mathrm{s}}}(v) \mathbb{E}_{Y(f_{\mathrm{v}},\rho_{\mathrm{i}},r_{\mathrm{v,inf}})|N_{\mathrm{s}}=n_{\mathrm{s}},V(f_{\mathrm{v}})=v}\left[e^{iuY}\right] \\
&= \left\{1+\rho_{\mathrm{i}}\left[1-f_{\mathrm{v}}(1-r_{\mathrm{v,inf}})\right](e^{iu}-1)\right\}^{n_{\mathrm{s}}}.
\end{aligned} \tag{34}$$

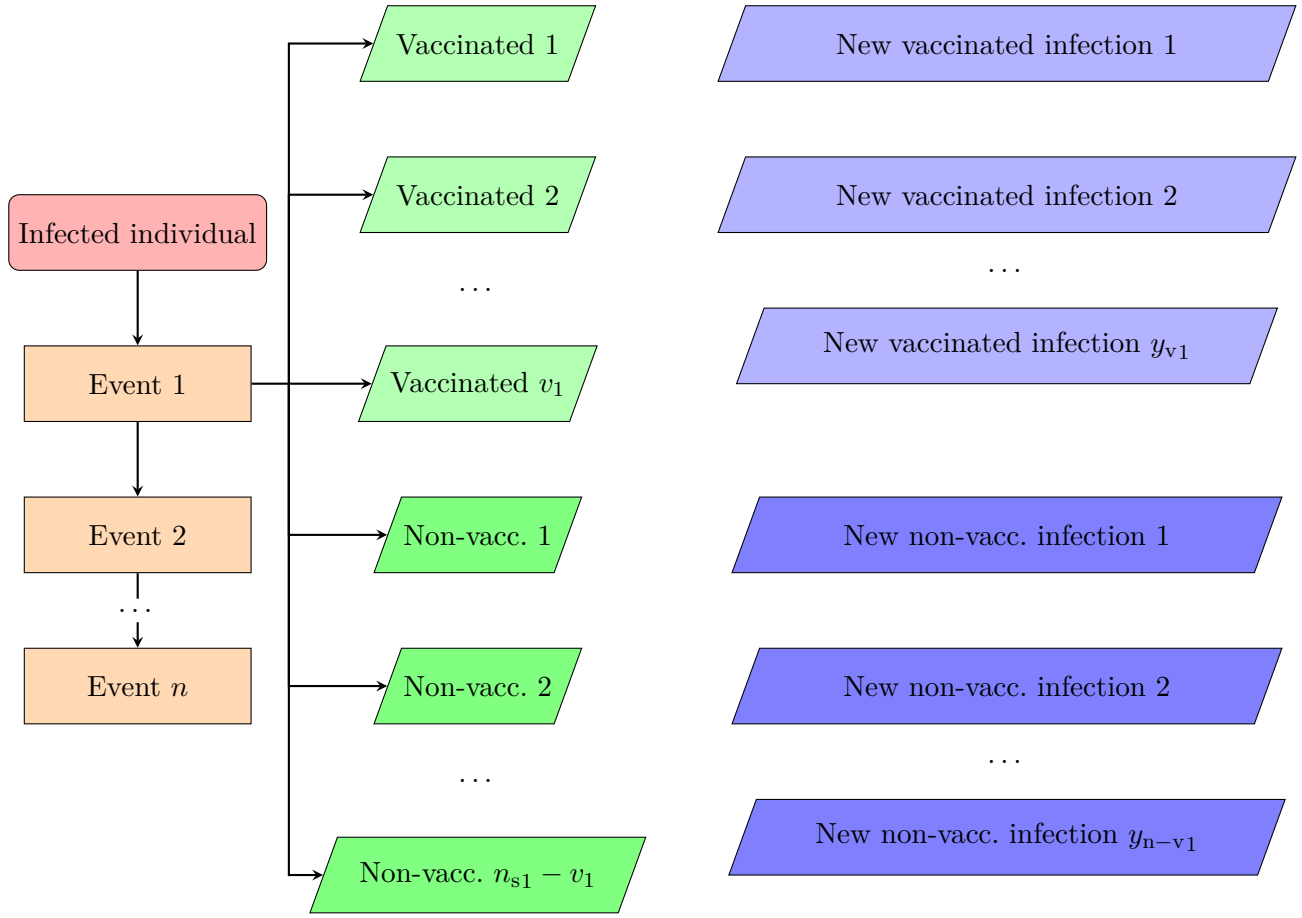

Figure 4: *Infection of susceptible individuals during one event with the vaccination branching process, where n, n_{s1}, v_1, y_{v1} and y_{n-v1} are respectively realizations for N, N_{s1}, V_1, Y_{v1} and Y_{n-v1}. The random variables V, Y_{v1} and Y_{n-v1} are distributed according to binomial distributions $B(n_{s1}, f_v)$, $B(v_1, \rho_i r_{v,inf})$ and $B(n_{s1} - v_1, \rho_i)$, respectively.*

To obtain the above expression, the binomial theorem was used with Equations (29) and (33). The Y random variable conditioned on N_s thus follows a binomial distribution with parameters n_s and $\rho_i [1 - f_v(1 - r_{v,inf})]$.

As the characteristic function for $Y(f_v, \rho_i, r_{v,inf})$ has the same form as Equation (23), but with p_i replaced by $\rho_i [1 - f_v(1 - r_{v,inf})]$, Equations (24), (25) and (28) are valid, using the same substitution, for the vaccination model described. That is, a model where the infectious individual (with a given vaccination status) can infect susceptible individuals from a population that is vaccinated with a fraction f_v, and where vaccinated individuals are less likely to be infected through the scaling factor $r_{v,inf}$.

In order to determine the expected value and the relative variance for the total number of new infections from an infectious individual with an unknown vaccination status, which will be defined as $M(\lambda, q, q_v, a, b, a', b', p_i, f_v, r_{v,inf}, r_{v,tr})$, one can define a set of two new random variables named as ${}^{v}M(\lambda, q_v, a, b, a', b', r_{v,tr} p_i, f_v, r_{v,inf})$ and ${}^{n-v}M(\lambda, q, a, b, a', b', p_i, f_v, r_{v,inf})$, for the total number of new infections from vaccinated and non-vaccinated infectious individuals, respectively. These definitions assume that p_i is defined as the probability of infection of a non-vaccinated susceptible individual by a non-vaccinated infectious individual, that the durations for the communicable periods of self-isolating

and non-self-isolating individuals are the same for vaccinated and non-vaccinated individuals. It also assumes that a fraction q_{v} of vaccinated individuals self isolate, compared to a fraction q for non-vaccinated individuals. Interactions are also assumed to follow the same distributions for vaccinated and non-vaccinated individuals. Based on these definitions and the previous discussion regarding the characteristic function for $Y(f_{\mathrm{v}}, \rho_{\mathrm{i}}, r_{\mathrm{v,inf}})$, the expected values and variances for these two random variables are given by:

$$\begin{aligned}
\mathbb{E}\left[{}^{\mathrm{v}}M\right] &= \lambda\left[(1-q_{\mathrm{v}})\bar{t} + q_{\mathrm{v}}\bar{m}\right] r_{\mathrm{v,tr}} p_{\mathrm{i}}\left[1 - f_{\mathrm{v}}(1-r_{\mathrm{v,inf}})\right]\mathbb{E}\left[N_{\mathrm{s}}\right] \\
&= \lambda\mathbb{E}\left[{}^{\mathrm{v}}\mathcal{T}\right] r_{\mathrm{v,tr}} p_{\mathrm{i}}\left[1 - f_{\mathrm{v}}(1-r_{\mathrm{v,inf}})\right]\mathbb{E}\left[N_{\mathrm{s}}\right] \qquad (35) \\
\frac{\mathrm{Var}\left[{}^{\mathrm{v}}M\right]}{\mathbb{E}\left[{}^{\mathrm{v}}M\right]^2} &= \frac{1}{\mathbb{E}\left[{}^{\mathrm{v}}\mathcal{T}\right]^2}\left[(1-q_{\mathrm{v}})\frac{\bar{t}^2}{a} + q_{\mathrm{v}}\frac{\bar{m}^2}{a'} + q_{\mathrm{v}}(1-q_{\mathrm{v}})(\bar{t}-\bar{m})^2\right] + \\
&\qquad \frac{1}{\lambda\mathbb{E}\left[{}^{\mathrm{v}}\mathcal{T}\right]}\left[1 + \frac{\mathrm{Var}\left[N_{\mathrm{s}}\right]}{\mathbb{E}\left[N_{\mathrm{s}}\right]^2} + \frac{1}{\mathbb{E}\left[N_{\mathrm{s}}\right]}\frac{1 - r_{\mathrm{v,tr}}p_{\mathrm{i}}\left[1-f_{\mathrm{v}}(1-r_{\mathrm{v,inf}})\right]}{r_{\mathrm{v,tr}}p_{\mathrm{i}}\left[1-f_{\mathrm{v}}(1-r_{\mathrm{v,inf}})\right]}\right] \\
&= \frac{\mathrm{Var}\left[{}^{\mathrm{v}}\mathcal{T}\right]}{\mathbb{E}\left[{}^{\mathrm{v}}\mathcal{T}\right]^2} + \frac{\mathrm{Var}\left[N(\lambda\mathbb{E}\left[{}^{\mathrm{v}}\mathcal{T}\right])\right]}{\mathbb{E}\left[N(\lambda\mathbb{E}\left[{}^{\mathrm{v}}\mathcal{T}\right])\right]^2} + \frac{1}{\mathbb{E}\left[N(\lambda\mathbb{E}\left[{}^{\mathrm{v}}\mathcal{T}\right])\right]}\left[\frac{\mathrm{Var}\left[N_{\mathrm{s}}\right]}{\mathbb{E}\left[N_{\mathrm{s}}\right]^2} + \right. \\
&\qquad \left. \frac{\mathrm{Var}_{Y|N_{\mathrm{s}}=\mathbb{E}[N_{\mathrm{s}}]}\left[Y(r_{\mathrm{v,tr}}p_{\mathrm{i}}\left[1-f_{\mathrm{v}}(1-r_{\mathrm{v,inf}})\right])\right]}{\mathbb{E}_{Y|N_{\mathrm{s}}=\mathbb{E}[N_{\mathrm{s}}]}\left[Y(r_{\mathrm{v,tr}}p_{\mathrm{i}}\left[1-f_{\mathrm{v}}(1-r_{\mathrm{v,inf}})\right])\right]^2}\right] \qquad (36)
\end{aligned}$$

and,

$$\begin{aligned}
\mathbb{E}\left[{}^{\mathrm{n-v}}M\right] &= \lambda\left[(1-q)\bar{t} + q\bar{m}\right] p_{\mathrm{i}}\left[1 - f_{\mathrm{v}}(1-r_{\mathrm{v,inf}})\right]\mathbb{E}\left[N_{\mathrm{s}}\right] \\
&= \lambda\mathbb{E}\left[{}^{\mathrm{n-v}}\mathcal{T}\right] p_{\mathrm{i}}\left[1 - f_{\mathrm{v}}(1-r_{\mathrm{v,inf}})\right]\mathbb{E}\left[N_{\mathrm{s}}\right] \qquad (37) \\
\frac{\mathrm{Var}\left[{}^{\mathrm{n-v}}M\right]}{\mathbb{E}\left[{}^{\mathrm{n-v}}M\right]^2} &= \frac{1}{\mathbb{E}\left[{}^{\mathrm{n-v}}\mathcal{T}\right]^2}\left[(1-q)\frac{\bar{t}^2}{a} + q\frac{\bar{m}^2}{a'} + q(1-q)(\bar{t}-\bar{m})^2\right] + \\
&\qquad \frac{1}{\lambda\mathbb{E}\left[{}^{\mathrm{n-v}}\mathcal{T}\right]}\left[1 + \frac{\mathrm{Var}\left[N_{\mathrm{s}}\right]}{\mathbb{E}\left[N_{\mathrm{s}}\right]^2} + \frac{1}{\mathbb{E}\left[N_{\mathrm{s}}\right]}\frac{1 - p_{\mathrm{i}}\left[1-f_{\mathrm{v}}(1-r_{\mathrm{v,inf}})\right]}{p_{\mathrm{i}}\left[1-f_{\mathrm{v}}(1-r_{\mathrm{v,inf}})\right]}\right] \\
&= \frac{\mathrm{Var}\left[{}^{\mathrm{n-v}}\mathcal{T}\right]}{\mathbb{E}\left[{}^{\mathrm{n-v}}\mathcal{T}\right]^2} + \frac{\mathrm{Var}\left[N(\lambda\mathbb{E}\left[{}^{\mathrm{n-v}}\mathcal{T}\right])\right]}{\mathbb{E}\left[N(\lambda\mathbb{E}\left[{}^{\mathrm{n-v}}\mathcal{T}\right])\right]^2} + \frac{1}{\mathbb{E}\left[N(\lambda\mathbb{E}\left[{}^{\mathrm{n-v}}\mathcal{T}\right])\right]}\left[\frac{\mathrm{Var}\left[N_{\mathrm{s}}\right]}{\mathbb{E}\left[N_{\mathrm{s}}\right]^2} + \right. \\
&\qquad \left. \frac{\mathrm{Var}_{Y|N_{\mathrm{s}}=\mathbb{E}[N_{\mathrm{s}}]}\left[Y(p_{\mathrm{i}}\left[1-f_{\mathrm{v}}(1-r_{\mathrm{v,inf}})\right])\right]}{\mathbb{E}_{Y|N_{\mathrm{s}}=\mathbb{E}[N_{\mathrm{s}}]}\left[Y(p_{\mathrm{i}}\left[1-f_{\mathrm{v}}(1-r_{\mathrm{v,inf}})\right])\right]^2}\right], \qquad (38)
\end{aligned}$$

with ${}^{\mathrm{v}}\mathcal{T} \equiv \mathcal{T}(q_{\mathrm{v}}, a, b, a', b')$ and ${}^{\mathrm{n-v}}\mathcal{T} \equiv \mathcal{T}(q, a, b, a', b')$. Vaccinated and non-vaccinated infectious individuals are two mutually exclusive classes, such that Equations (10) to (13) apply, and that the distribution for $M(\lambda, q, q_{\mathrm{v}}, a, b, a', b', p_{\mathrm{i}}, f_{\mathrm{v}}, r_{\mathrm{v,inf}}, r_{\mathrm{v,tr}})$ can be expressed as a linear combination of the distributions for ${}^{\mathrm{v}}M(\lambda, q_{\mathrm{v}}, a, b, a', b', r_{\mathrm{v,tr}}p_{\mathrm{i}}, f_{\mathrm{v}}, r_{\mathrm{v,inf}})$ and ${}^{\mathrm{n-v}}M(\lambda, q, a, b, a', b', p_{\mathrm{i}}, f_{\mathrm{v}}, r_{\mathrm{v,inf}})$, leading to:

$$\begin{aligned}
\mathbb{E}\left[M\right] &= \frac{f_{\mathrm{v}}r_{\mathrm{v,inf}}\mathbb{E}\left[{}^{\mathrm{v}}M\right] + (1-f_{\mathrm{v}})\mathbb{E}\left[{}^{\mathrm{n-v}}M\right]}{1 - f_{\mathrm{v}}(1-r_{\mathrm{v,inf}})} \\
&= \lambda\left\{f_{\mathrm{v}}r_{\mathrm{v,inf}}r_{\mathrm{v,tr}}\left[(1-q_{\mathrm{v}})\bar{t} + q_{\mathrm{v}}\bar{m}\right] + (1-f_{\mathrm{v}})\left[(1-q)\bar{t} + q\bar{m}\right]\right\} p_{\mathrm{i}}\mathbb{E}\left[N_{\mathrm{s}}\right] \\
&= \lambda\bar{t}p_{\mathrm{i}}\mathbb{E}\left[N_{\mathrm{s}}\right]\left\{f_{\mathrm{v}}r_{\mathrm{v,inf}}r_{\mathrm{v,tr}}\left[1 - q_{\mathrm{v}}\left(1 - \frac{\bar{m}}{\bar{t}}\right)\right] + (1-f_{\mathrm{v}})\left[1 - q\left(1 - \frac{\bar{m}}{\bar{t}}\right)\right]\right\}, \qquad (39)
\end{aligned}$$

where $\frac{f_{\mathrm{v}}r_{\mathrm{v,inf}}}{1-f_{\mathrm{v}}(1-r_{\mathrm{v,inf}})}$ is the probability for an infectious individual to be vaccinated. This probability differs from f_{v}, because the vaccinated individuals have a probability of infection scaled down by the factor $r_{\mathrm{v,inf}}$

when compared to non-vaccinated individuals. In other words, vaccinated individuals are less likely to become infected and therefore, fewer infectious individuals are likely to be vaccinated.

In Equation (39), note that $\lambda \bar{t} p_{\mathrm{i}} \mathbb{E}[N_{\mathrm{s}}]$ corresponds to the expected total number of new infections generated from an infectious individual in a population where there is no vaccination ($f_{\mathrm{v}} = 0$) and no self isolation ($q = 0$).

Regarding the relative variance for $M(\lambda, q, q_{\mathrm{v}}, a, b, a', b', p_{\mathrm{i}}, f_{\mathrm{v}}, r_{\mathrm{v,inf}}, r_{\mathrm{v,tr}})$, Equation (13) leads to:

$$\frac{\mathrm{Var}[M]}{\mathbb{E}[M]^2} = \frac{1 - f_{\mathrm{v}}(1 - r_{\mathrm{v,inf}})}{\{f_{\mathrm{v}} r_{\mathrm{v,inf}} \mathbb{E}[{}^{\mathrm{v}}M] + (1 - f_{\mathrm{v}}) \mathbb{E}[{}^{\mathrm{n-v}}M]\}^2} \left\{ f_{\mathrm{v}} r_{\mathrm{v,inf}} \left[\frac{\mathrm{Var}[{}^{\mathrm{v}}M]}{\mathbb{E}[{}^{\mathrm{v}}M]^2} + 1 \right] \mathbb{E}[{}^{\mathrm{v}}M]^2 + (1 - f_{\mathrm{v}}) \left[\frac{\mathrm{Var}[{}^{\mathrm{n-v}}M]}{\mathbb{E}[{}^{\mathrm{n-v}}M]^2} + 1 \right] \mathbb{E}[{}^{\mathrm{n-v}}M]^2 \right\} - 1, \tag{40}$$

which will not be expanded further here.

In this section, an expression for the reproduction number of an outbreak, with a partially vaccinated population and where fractions of vaccinated and non-vaccinated infectious individuals self isolate was derived, for a general model where interactions have been decoupled from infections, through the use of a probability of infection and events where the number of susceptible individuals can be distributed according to any desired distribution. This model encompasses all models presented in earlier sections, as they represent particular cases of the model presented in this section.

The analytical representation of the reproduction number gives insight on the average outcome of the disease for each infectious individual, but it may be desirable to study the impact of the progression of the outbreak over time.

3.4 Time scaling through simulation: latent phase, initial time, and extinction

When analyzing simulated outbreaks, it is often desirable to study different statistics as a function of time, as time is more easily observable for actual outbreaks, as opposed to the infectious individual's generation (as in Figure 1).

When simulating the outbreaks, the modified branching processes define all events in time relative to the start of the outbreak, which is, by default, the time when the first primary infectious individual became infected. The initial time can also be defined as the time when the first primary infectious individual became infectious or when his infectious period ended. Regardless, the choice of time origin *per se* does not modify the model in such a way for it to deviate from a branching process, such that the results previously derived become invalid.

To better reproduce the disease characteristics and the timing of infectious events during simulations, a latent phase is added in the modified branching process. This latent phase is assumed:

- to start as soon as an individual is infected and to end at the beginning of the individual's infectious phase;
- to be statistically independent from its following infectious phase; and
- to be a random variable drawn from the gamma distribution $T(a^l, b^l)$.

In the context of a branching process where the statistical distributions are static over time and across the generations of infected individuals, the introduction of such a latent phase does not affect the branching process (if all other parameters remain the same), but could impact the timing of infectious events.

When observing output metrics during the simulation of an outbreak, since the branching process is a continuous process, special attention needs to be given when evaluating the propagation of the outbreak over a defined and finite time period. That is, as the modified branching process has to be simulated over a finite number of days, proper care has to be taken when processing simulation data to compute metrics. For example, at the end of the simulated time period, only those infected by an infectious individual having completed his infectious period are considered, as this prevents biasing the reproduction number (i.e., if the end time truncates an individual's infectious phase, his number of generated secondary infections would be otherwise underestimated). If on any given simulated day, there is no new infected individual and no individual with an incomplete infectious period (by the end of the day), the outbreak is said to be extinct. The time of extinction is defined as the number of days from the initial time up to that point.

It is worth mentioning that infectious individuals from different generations may end up generating interaction events on the same day, due to individuals' different latent and infectious periods, showing the importance of simulating outbreaks in time, especially when considering deviations from the branching process (as described in Section 3.5).

3.5 Model modifications beyond a branching process

So far, a branching process was generalized in order to be applicable to a wider range of scenarios. This ultimately resulted in Equation (39), which is an analytical expression for the effective reproduction number when the various statistical distributions that are used by the model can be assumed static over time. Some simulation scenarios require additional constraints that may deviate from a branching process. Such constraints are discussed in this section.

3.5.1 Constraints on the primary individual

For some situations, it can be useful to analyse a scenario where the primary infectious individual has a constrained state, such as forcing the individual to be in the self-isolating category or in the non-self-isolating category. This modifies the distribution of the communicable period of the first infectious individual. Consequently, the reproduction number will deviate from the analytical expression, and simulations would be required to evaluate the average impact of the primary individual's initial state.

3.5.2 Constraints on the outbreak path

The model presented so far simulates the number of infected individuals. But when actual outbreaks occur, many infections can be missed because, often, only positive cases are detected and reported. In order for measured statistics from simulated outbreaks to be closer to reality, it is possible to introduce testing of individuals in the model. In this work, testing itself does not impact interactions; therefore, the model will not deviate from a branching process and the analytical expressions still apply. But when studying the statistics that involve only observable outbreak paths (where at least one individual tests positive, as opposed to all simulated outbreak paths), the reproduction number deviates from a branching process. Simulations would be required for statistical analysis.

3.5.3 Outbreak capping

There are cases where outbreak capping may be desirable. That is, the simulated outbreaks are stopped when specific thresholds are reached. There are three capping thresholds supported by the model that could stop the outbreaks:

- once a fixed time duration is reached;
- once a maximum number of infectious individuals is reached; and
- once a number of positive test cases is reached for a given time interval.

The first threshold is typical as simulations are often performed for a fixed time duration. The time is cut so that the duration corresponds to the time of interest at which statistics are computed, without impacting interactions between individuals.[10] This way, the analytical representation of the reproduction number applies.

For the other two capping thresholds, simulations diverge from a branching process. Setting a maximum number of infectious individuals for a given time period avoids runaway outbreaks for scenarios with a sufficiently large reproduction number. This cut prevents new infectious individuals to further propagate the outbreak when they are infected during events that occur at a time when this threshold was exceeded. Such a cut can become effective at a different time depending on the simulation path, and if the cut occurs before the end of the simulation time duration, the model statistics will deviate from a branching process.

A limit on the number of positive tests within a given interval of time can be set to simulate a lock down situation after too many positive tests occurred. Similarly to the previous cut, this cut can also become effective at a different time depending on the simulation path, and if the cut occurs before the end of the simulation time duration, the model statistics will deviate from a branching process.

3.5.4 Contact tracing and subsequent interruption of interactions

The presented model also supports contact tracing and isolation of traced contacts. Here is how contact tracing and isolation is modelled:

- Contact tracing is triggered following a positive test.
- Only symptomatic, self-isolating individuals are assumed to perform testing in the absence of being traced.
- The test is assumed to occur at the same time as the time of self isolation.
- A test can be positive given true positive rates that can be distinct for symptomatic and asymptomatic individuals.
- Test results are assumed to be received after a fixed time delay.
- For the individuals who test positive, contacts that occurred before they self isolated, but within a fixed time window are considered for contact tracing.
- A fixed probability of successful contact identification and tracing is assumed.

[10] The time cut is applied in a way that does not truncate the counting of the new infections generated by a given infectious individual.

- Fixed probabilities of successful infected contact identification, tracing and compliance with self isolation are assumed for symptomatic individuals that self isolate and for asymptomatic individuals.
- The contact tracing process is assumed to require a time following distinct gamma distributions for symptomatic and asymptomatic individuals.

When interactions are interrupted through the contact tracing algorithm, this affects the distribution for the communicable period of infectious individuals, in a manner that differs from one generation to the other, because the contact tracing mechanism relies on a positive test for the parent of the traced infectious individual. Consequently, this algorithm makes the model deviate from a branching process, and the effective reproduction number can no longer be calculated analytically.

3.6 Summary

In Section 3.1, a transmission branching process [36] suitable for two classes of individuals having different communicable periods, such as for infected individuals that self isolate or not, has been generalized for an arbitrary distribution of the number of infections occurring during one event. Events were then generalized in Section 3.2 to interaction events through the introduction of a probability of infection, allowing for a more physical interpretation of the concept of contacts, as well as for the effects of mitigation measures. These two subsections expressed the reproduction number for such a generalized branching process as the product between the event rate, the average value for the duration of the infectious individual's communicable period, the probability of infection of a given susceptible individual attending an event by the infectious individual and the average number of susceptible individuals attending an event. This result is possible thanks to the independence of these variables in the model. The relative variance of the reproduction number for this generalized model was also expressed as a function of the contributions from the relative variances of the different random variables involved in the model.

The model was then expanded further in Section 3.3 to create a branching process suitable for vaccination, following the introduction of four extra parameters. Section 3.4 introduced further expansions to the branching process related to time modelling for the outbreak, including the support for the concept of a latent period. Finally, Section 3.5 presented modifications to the model beyond a branching process, that allow for the modelling of contact tracing, thanks to the concept of interaction events.

4 Input and output parameters for the model

To run simulations, different parameters need to be specified. These can be grouped into the following categories:

1. disease and interaction events parameters;
2. vaccination parameters;
3. testing and contact tracing parameters; and
4. simulation parameters and settings.

This section describes the input parameters (required and optional) to run simulations, and how, where applicable, they relate the the analytical expressions derived in Section 3. This section will also briefly describe the output parameters. Details about the algorithm can be found in Annex A.

4.1 Disease and interaction events input parameters

The disease infectious period needs to be configured. For clarity, this work defines the first class of infectious individuals (with infectious period drawn from the gamma distribution $T(a, b)$) as non-self-isolating individuals (or *asymptomatic*), and the second class of infectious individuals (with infectious period drawn from the gamma distribution $T(a', b')$) as self-isolating individuals at symptom onset (or *symptomatic*). Their combined gamma distribution $\mathcal{T}(q, a, b, a', b')$ can be characterised by their means ($\bar{t}$, $\bar{m}$) and 95th percentiles (t_{95}, m_{95}) or by their gamma shape parameters. In the former case, the gamma shape parameters are obtained by an iterative method (using Newton's method). In the latter case, they are obtained from means and different shape parameters (κ_t, κ_m). Note that,

$$\bar{t} \equiv \frac{a}{b}, \qquad \kappa_t \equiv b, \tag{41}$$

$$\bar{m} \equiv \frac{a'}{b'}, \qquad \kappa_m \equiv b'. \tag{42}$$

The following input parameters are used to define $\mathcal{T}(q, a, b, a', b')$:

$\bar{t}$: average duration of the infectious period of non-self-isolating individuals;

t_{95}: the 95th percentile of the infectious period of non-self-isolating individuals;

κ_t: the gamma shape parameter for the infectious period of non-self-isolating individuals;

$\bar{m}$: average duration of the infectious period of self-isolating individuals at symptoms onset;

m_{95}: the 95th percentile of the infectious period of self-isolating individuals at symptoms onset;

κ_m: the gamma shape parameter for the infectious period of self-isolating individuals at symptoms onset; and

q: fraction of infectious individuals that self isolate (as defined previously).

It may also be of interest to configure the disease latent period. Having a latent period is optional but recommended when interested in the time progression of the disease or when evaluating contact tracing. The latent period can also be characterised by its mean ($\bar{l}$) and 95th percentile (l_{95}) or by its gamma shape parameter (κ_l):

$$\bar{l} \equiv \frac{a^l}{b^l}, \qquad \kappa_l \equiv b^l, \tag{43}$$

where:

$\bar{l}$: average duration of the disease latent period;

l_{95}: the 95th percentile of the disease latent period; and

κ_l: the gamma shape parameter for the disease latent period.

The interaction events need to be configured. Transmission happens during interaction events that occur in individuals' infectious period. Timing between successive events is drawn from an exponential distribution with a mean $\frac{1}{\lambda}$ and the number of events over the infectious period is determined by a Poisson process (with rate λ). The event rate can be configured with either:

λ: rate of the interaction events where at least one susceptible individual is present and where transmission and new infections may occur; or

λ_e: rate of all interaction events including those where only the infectious individual is present.

For example, λ should be used when evaluating the impact of meetings or gatherings on the disease progression, whereas λ_e should be used when evaluating the impact of individuals going into the office where they may use different facilities (e.g., their office, the kitchen, the bathroom, etc.) but may or may not come into contact with others.

The number of susceptible attendees at each event is a random number that can be drawn from a user-selected distribution. The distribution can be selected to represent the total number of attendees at an event (including the infectious individual) or only the susceptible attendees (default). In the first case, the user can select from two types of distribution, logarithmic or discretised Gaussian, and these must be truncated below two (2) to ensure that at least two individuals (the single infectious and one susceptible) are present. The truncated logarithmic distribution can be defined using any of the following three parameters:

p: the shape parameter of the logarithmic distribution;

μ: the mean of a corresponding unbounded (non-truncated) logarithmic distribution; or

g_{ave}: the average number of event attendees, including the infectious individual.

The truncated Gaussian distribution can be defined using:

μ: the mean parameter of the corresponding unbounded Gaussian distribution; or

g_{ave}: the average number of event attendees, including the infectious individual; and

σ: the standard deviation of a corresponding unbounded Gaussian distribution; or

σ_{r}: a relative standard deviation with respect to μ.

For the case where the distribution is selected to represent the susceptible attendees only, the logarithmic distribution is the only supported option, and it can be defined using any of the three parameters p, μ, and g_{ave} defined above.

The parameter g_{ave} is used to define the average number of attendees to an event (including the infectious individual), regardless of which one of the three types of distributions is used. This parameter corresponds to $\mathbb{E}[N_{\text{s}}] + 1$.

Not all individuals will become infected as they have a probability p_{i} (assuming no vaccination) of becoming infected. This value may be unknown, but its relationship to the basic reproduction number R_0 satisfies Equation (26) when $q = 0$:

$$R_0 = \lambda(g_{\text{ave}} - 1)p_{\text{i}}\bar{t}. \tag{44}$$

where:

R_0: the basic reproduction number when no mitigation is in place; and/or

p_{i}: the probability that a given non-vaccinated susceptible attendee be infected by the non-vaccinated infectious individual at an event.

There are three independent parameters that must be used to define the part of the model that relates to R_0. They can be chosen amongst R_0, the interaction event rate (λ, λ_e), the parameters necessary to derive g_{ave} (p, μ, σ, σ_{r}), p_{i} or $\bar{t}$. In addition to the parameters previously listed, either R_0 and/or p_{i} can be provided as a fourth independent parameter to determine Expression (44).

Once the existence of the disease is known, it is expected that a fraction of the infected individuals (q) would have symptoms and self isolate, shortening the communicable period to $\bar{m}$ ($\bar{m} < \bar{t}$). Because of self isolation, the resulting effective reproduction number R_{eff} would have a lower value than R_0 ($R_{\text{eff}} < R_0$). Equation (26) can be expressed as:

$$R_{\text{eff}} = \lambda(g_{\text{ave}} - 1)p_{\text{i}}\left[(1-q)\bar{t} + q\bar{m}\right] \tag{45}$$

$$= R_0\left[1 - q\left(1 - \frac{\bar{m}}{\bar{t}}\right)\right], \tag{46}$$

where R_{eff} is the effective reproduction number when a fraction of the infectious individuals self isolate after the appearance of symptoms.

4.2 Vaccination parameters

If immunization is to be simulated, the following input parameters (defined in Section 3.3 and redefined here for clarity) are required:

f_{v}: Fraction of the population vaccinated.

$r_{\text{v,inf}}$: Scaling factor, between 0 and 1, for the infection probability of a given vaccinated susceptible individual by a given infectious individual during one event. This parameter captures the vaccine's efficacy against (symptomatic and asymptomatic) infection ($1 - r_{\text{v,inf}}$).

$r_{\mathrm{v,tr}}$: Scaling factor, between 0 and 1, for the infection probability of a given susceptible individual by a given vaccinated infectious individual during one event. This parameter captures the vaccine's efficacy against onward transmission $(1 - r_{\mathrm{v,tr}})$.

q_{v}: Fraction of infectious vaccinated individuals that self isolate.

The effective reproduction number with vaccination $R_{\mathrm{eff,v}}$, Equation (39), can be expressed as:

$$R_{\mathrm{eff,v}} = R_{\mathrm{eff}} \left\{ 1 - f_{\mathrm{v}} \left[1 - r_{\mathrm{v,inf}} r_{\mathrm{v,tr}} \frac{1 - q_{\mathrm{v}} \left(1 - \frac{\bar{m}}{\bar{t}} \right)}{1 - q \left(1 - \frac{\bar{m}}{\bar{t}} \right)} \right] \right\}, \tag{47}$$

This shows how the effective reproduction number is further reduced within the model when vaccination is included (i.e., $R_{\mathrm{eff,v}} < R_{\mathrm{eff}} < R_0$).

If vaccination has no impact on self isolation (i.e., the fraction of vaccinated and non-vaccinated infectious individuals that self isolate are the same), the above expression reduces further to:

$$R_{\mathrm{eff,v}} = R_{\mathrm{eff}} \left\{ 1 - f_{\mathrm{v}} \left[1 - r_{\mathrm{v,inf}} r_{\mathrm{v,tr}} \right] \right\} \qquad (\text{for } q_{\mathrm{v}} = q). \tag{48}$$

See Annex B for more details on this special case.

4.3 Testing and contact tracing parameters

When evaluating testing, the following parameters must be specified:

$t_{\mathrm{t.p.r.}}$: the true positive test rate of non-self-isolating individuals;

$m_{\mathrm{t.p.r.}}$: the true positive test rate of self-isolating individuals; and

$T_{\Delta t}$: the time delay between performing the test and reporting results.

When evaluating contact tracing, the testing parameters need to be configured as well as the following:

P_{t}: the probability of successful contact tracing;

$P_{\mathrm{i_t}}$: the probability of successful contact tracing and of self isolation once contacted of infected individuals that would otherwise not self isolate;

$P_{\mathrm{i_m}}$: the probability of successful contact tracing and of self isolation once contacted of infected individuals that would otherwise self isolate;

w_{ct}: the contact tracing window which is the period prior to individual isolation during which contacts are considered; and

ω: the delay in isolating an individual from the moment contact tracing is initiated as soon as a positive test is reported.

The latter is characterised by a mean and a 95th percentile or a gamma shape parameter for each type of contact:

$$\bar{t}^{\mathrm{i}} \equiv \frac{a_{\mathrm{i_t}}}{b_{\mathrm{i_t}}}, \qquad \kappa_t^{\mathrm{i}} \equiv b_{\mathrm{i_t}}, \tag{49}$$

$$\bar{m}^{\mathrm{i}} \equiv \frac{a_{\mathrm{i_m}}}{b_{\mathrm{i_m}}}, \qquad \kappa_m^{\mathrm{i}} \equiv b_{\mathrm{i_m}}. \tag{50}$$

The following input parameters are thus effectively used to define the distribution for ω:

$\bar{t}^{\mathrm{i}}$: the average delay in contact tracing and isolating individuals that, if not contacted, would otherwise not self isolate;

t_{95}^{i}: the 95th percentile of the delay in contact tracing and isolating individuals that, if not contacted, would otherwise not self isolate;

κ_t^{i}: the gamma shape parameter of the delay in contact tracing and isolating individuals that, if not contacted, would otherwise not self isolate;

$\bar{m}^{\mathrm{i}}$: the average delay in contact tracing and isolating individuals that, if not contacted, would otherwise self isolate;

m_{95}^{i}: the 95th percentile of the delay in contact tracing and isolating individuals that, if not contacted, would otherwise self isolate; and

κ_m^{i}: the gamma shape parameter of the delay in contact tracing and isolating individuals that, if not contacted, would otherwise self isolate.

4.4 Simulation parameters and options

The number of days from the start of the outbreak X_n are required. The start of the outbreak can be set to either the time at which:

- the first patient is infected (`ro_time_pri_created`, this is the default);
- the first patient is infectious (`ro_time_pri_infectious`);
- the first self-isolated or no longer infectious patient (`ro_time_pri_end_comm`);
- the first test is reported (`ro_time_pri_test_results`); or
- the first positive test is reported (`ro_time_first_pos_test_results`).

It is also possible to specify whether the first individual:

- self isolated at symptom onset (`ro_pricommper_alt`);
- self isolated at symptom onset and tested positive (`ro_pricommper_alt_use_tpr`); or
- was asymptomatic (`ro_pricommper_alt`).

The number of outbreak paths n_{paths} for which statistics and performance metrics are computed also needs to be specified. These n_{paths} can be either:

- all outbreak paths generated at random (`ro_all_paths`, this is the default);
- observable paths with at least one positive test result (`ro_observable_paths_only`); or
- non observable paths without positive test result (`ro_non_observable_paths_only`).

4.5 Output performance metrics

When running numerous simulations over a given number of days $n = 1, 2, 3, ...X_n$, statistical analyses can be performed to determine the means (and variances) of:

- the number of new infected individuals on any given day n,
- the number of new positive tests on any given day n,
- the number of new contacts on any given day n,
- the reproduction number, and
- the time until extinction.

The probability of extinction P_{ext} at day X_n can also be computed as the sum of all paths with no more infectious individuals (on day $n = X_n$) over all N_{path} simulated paths. It is worth mentioning that the reproduction number can be obtained using both the analytical method or the time simulation model. These will match within statistical uncertainties unless the chosen model deviates from a branching process.

4.6 Summary

For scenarios where transmission occurs through events where the number of interactions is not logarithmically distributed, other types of distributions can be used with the modified model. It is also possible to look at observed infections (through positive test results), and to change the time reference when generating output distributions (e.g., time since the first positive test). In the modified model, there are options to select characteristics for the primary infectious individuals, and also to apply filters and cuts when generating results.

Although the analytical expression computes the reproduction number (when the model does not deviate from a branching process), the time simulation allows for the computation of other performance metrics.

5 Case studies

In this section, three different simulation examples for the model are presented. The first simulation shows how the IBP can be used, in a generalized approach, to evaluate the impact of self isolating at symptom onset, of group size, of frequency of event, and of probability of infection on the spread of the virus. The second simulation shows how model modifications beyond a branching process (c.f. Section 3.5) can be used. This simulation attempts to reproduce the COVID-19 outbreak that occurred on the Diamond Princess cruise ship (in January 2020). The third simulation evaluates the use of the model with vaccination enabled. Overall, these case studies evaluate the possible outcome of different mitigation strategies (such as contact tracing, self isolation and vaccination).

As shown in Section 4, the model requires different input parameters which need to be specified for each scenario. Although the input parameters are specific to each scenario, the disease parameters were the same (unless specified otherwise) and presented in Section 5.1.

5.1 Infectious disease assumptions and input parameters

The following assumptions about the infectious disease, summarized in Table 1, were considered:

- The incubation period of self-isolating infectious individuals is drawn from a gamma distribution with a mean of 5.5 days and a 95th percentile of 9.72 days [37].
- The infectious phase of self-isolating infectious individuals is drawn from a gamma distribution with mean of 2.2 days and a 95th percentile of 6.8 days [30].
- The latent phase is set as the difference between the self-isolating individuals' incubation phase (5.5 days) and infectious phase (2.2 days); therefore, the mean is 3.3 days. The duration of the latent phase is assumed fixed, therefore the 95th percentile of the latent phase is also set to 3.3 days. This assumption is made so that the 95th percentile of the combined latent and infectious phases of self-isolating individuals matches the 95th percentile of the non-self-isolating individuals' incubation phase as closely as possible.
- The infectious phase for non-self-isolating infectious individuals was based on assumptions in [35].

The parameters listed in this section were used throughout this work, unless specified otherwise.

Table 1: *Disease time period for self-isolating and non-self-isolating infected individuals used in this work.*

Infected type	Latent phase		Infectious phase	
Self-isolating	$\bar{l}$=3.3	l_{95}=3.3	$\bar{m}$=2.2	m_{95}= 6.8
Non-self-isolating	$\bar{l}$=3.3	l_{95}=3.3	$\bar{t}$=6.79	t_{95}=12.2

5.2 Simulations I: The interaction branching process

When the basic reproduction number, R_0, of a virus is below 1, outbreaks will eventually go extinct [19]. But when $R_0 > 1$, mitigation strategies need to be put in place to help control the spread of a virus. Although the exact value of the COVID-19 R_0 is debatable [24, 38], it is consensus that its value is well above 1; therefore, it is important to identify outbreak control measures that have an effect on the spread of a virus to help prevent transmission [23].

The IBP can mathematically represent the impacts of self isolation when infected with the SARS-Cov-2 virus (q), of group size (g_{ave}) at interaction events, of the frequency of infectious events (λ), and of the probability of infections (p_{i}) on a virus' basic (R_0) or effective (R_{eff}) reproduction number, through Equations (44) or (46), respectively. If $R_{\text{eff}} < R_0$ the control measures are effective at reducing the spread of the virus.

Although the reproduction number (R_0 or R_{eff}) provides some insight on the impact of control measures on the spread of the virus, it is essential to simulate the timing of outbreaks to determine the control measures that facilitate extinction. That is, for similar R_{eff} values, different control measures may provide a better probability of extinction at a given time.

Using the IBP, this section shows the impact of varying μ, λ, and p_{i} on the probability for extinction for specific values of R_{eff} and q.

5.2.1 Methodology, assumptions, and parameters

For this analysis, an initial value of $R_0 = 3.395$ was assumed as this overlaps with a set of estimates published for the early phase of the outbreak, ranging from 2.2 (95% CI, 1.4–3.9) to 3.58 (95% CI, 2.89–4.39) as reported by Zhang et al. [38].

Once COVID-19 was known, self isolation at symptom onset was encouraged. This work assumed that 40% of infected individuals would be asymptomatic (as reported in [39]) and that symptomatic individuals would self isolate at symptom onset, leading to a $q = 0.6$. The disease parameters for this subsection were listed in Table 1. From Equation (46), this leads to an $R_{\text{eff}} = 2.018$.

For constant $R_{\text{eff}} = 2.018$ and $q = 0.6$, two different values of p_{i} were evaluated:

1. $p_{\text{i}} = 1$, and
2. $p_{\text{i}} = 1/8$.

The first assumes that all susceptible individuals attending an infectious event had 100% probability of becoming infected; therefore all individuals at an event become infected. The second assumed that event attendees were 8 times less likely of becoming infected; therefore the chances of fewer getting infected at each event increases, and on average across all outbreaks, only 1/8 attendee becomes infected.

The number of susceptible individuals at infectious events were drawn from a discrete logarithmic distribution. The average number of susceptible individuals at infectious events (i.e., $\mu = g_{\text{ave}} - 1$) evaluated were 1, 2, 4, 8, 16 and 32. The values of λ were set to allow for constant $R_{\text{eff}} = 2.018$, which, from Equation (45), with values of $\bar{t}$ and $\bar{m}$ listed in Table 1, are computed as:

$$\lambda = \frac{R_{\text{eff}}}{\mu p_{\text{i}} \left[(1-q)\bar{t} + q\bar{m}\right]} \tag{51}$$

$$= \frac{R_{\text{eff}}}{4.036 \mu p_{\text{i}}}, \tag{52}$$

For comparison purpose, a similar analysis was performed for a case where reproduction was cut in half and close to one, that is $R_{\text{eff}} = 1.009$. All possible combinations of interaction input parameters are listed in Table 2. From this table, it is noticeable that the frequency of events are cut in half from $R_{\text{eff}} = 2.018$ to $R_{\text{eff}} = 1.009$ for each combination of p_{i} and μ.

Table 2: *Parameter values used in Simulations I for $R_{\text{eff}} = 2.018$ and $R_{\text{eff}} = 1.009$, and for $p_{\text{i}} = 1/8$ and $p_{\text{i}} = 1$ using different values of μ. The values for λ are determined using Equation (45).*

R_{eff}	μ	$p_{\text{i}} = 1/8$ λ day^{-1}	$p_{\text{i}} = 1$ λ day^{-1}
2.018	1	4	1/2
	2	2	1/4
	4	1	1/8
	8	1/2	1/16
	16	1/4	1/32
	32	1/8	1/64
1.009	1	2	1/4
	2	1	1/8
	4	1/2	1/16
	8	1/4	1/32
	16	1/8	1/64
	32	1/16	1/128

Testing parameters were set to the default values of $t_{\text{t.p.r.}} = 70\%$, $m_{\text{t.p.r.}} = 70\%$, and $T_{\Delta t} = 2$ days listed in Table 3. Each simulation was set to run for 60 days relative to the first positive test reported, and 10^4 outbreaks where at least one positive test case occurred were generated. These are referred to as *observable outbreaks* and described in Section 3.5.2. Limiting the analysis to observable paths allows for a realistic time reference that facilitates analysis. Discarding unobservable outbreak paths results in statistics that deviate from the ones of a branching process, however each individual observable path is simulated using the same process as for the branching process.

Table 3: *List of parameters and their default values (when applicable).*

Parameter	Symbol	Default value
Branching parameters		
Probability of infection	p_{i}	1%
Probability of self isolation[11]	q	0
Testing parameters		
Test true positive rate for non-self-isolating individuals	$t_{\text{t.p.r.}}$	70%
Test true positive rate for self-isolating individuals	$m_{\text{t.p.r.}}$	70%
Testing delay	$T_{\Delta t}$	2 days
Contact tracing parameters		
Probability of being contacted	P_{t}	80%
Historical contact window	w_{ct}	2 days

The following performance metrics were evaluated:

1. The probability of extinction (P_{ext}), defined as the number of observable outbreaks that go extinct over the total number of observable outbreak paths, within 60 days from the first positive test.

2. The average number of days until extinction (t_{ext}) from the first positive test, across all outbreak paths that went extinct.

3. The total number of infected individuals averaged across the extinct outbreaks (N_{ext}).

4. The total number of infected individuals after 60 days averaged across the non-extinct observable outbreaks ($N_{@60}$).

5.2.2 Results

Results for the probability of extinction, time of extinction, and total number of infected individuals as μ and λ varies are listed in Table 4. In this table, darker coloured cells indicate worse results. The results show that, for 10^4 observable outbreaks, simulations with $R_{\text{eff}} = 2.018$ obtained lower probabilities of extinction, but faster times to extinction when compared to the simulations with $R_{\text{eff}} = 1.009$. The results also show that for constant R_{eff}, reducing λ improves the probability of extinction (P_{ext}) and the time until extinction (t_{ext}), whereas increasing the number of susceptible attendees (μ) increases the total number of infected individuals for both extinct outbreaks (N_{ext}) and non-extinct outbreaks after 60 days ($N_{@60}$).

Table 4: *Simulations I results with R_0 3.395 and with disease parameters listed in Table 1.*

		$p_{\text{i}} = 1/8$					$p_{\text{i}} = 1$				
R_{eff}	μ	λ day^{-1}	P_{ext} %	t_{ext} days	N_{ext}	$N_{@60}$	λ day^{-1}	P_{ext} %	t_{ext} days	N_{ext}	$N_{@60}$
2.018	1	4	38.5	0.2	2.1	10,945	1/2	38.1	0.3	2.2	10,931
	2	2	40.9	0.3	2.2	11,463	1/4	55.2	0.7	2.5	15,339
	4	1	47.3	0.5	2.3	13,067	1/8	71.7	0.5	2.8	27,836
	8	1/2	58.3	0.6	2.5	16,502	1/16	84.9	0.1	3.2	52,920
	16	1/4	70.7	0.6	2.7	25,216	1/32	92.8	-0.3	3.5	119,132
	32	1/8	82.2	0.3	3.1	40,938	1/64	96.4	-0.9	3.9	279,114
1.009	1	2	86.7	6.2	5.6	86	1/4	86.8	6.4	5.6	85
	2	1	87.6	5.6	5.5	97	1/8	92.7	3.8	5.5	170
	4	1/2	90.8	5.2	5.8	123	1/16	96.2	2.0	5.9	396
	8	1/4	93.3	3.5	5.7	201	1/32	98.3	0.5	5.8	916
	16	1/8	95.6	2.2	6.0	380	1/64	99.0	-0.4	6.0	2,303
	32	1/16	97.7	1.0	6.4	713	1/128	99.6	-1.1	6.5	3,741

5.2.3 Discussion of results

The improvement in P_{ext} as λ becomes smaller, for constant R_{eff}, is expected; as infectious events are spaced out in time, the disease has more chances of becoming extinct before the next infectious event. This also reduces the time it takes for an outbreak to reach extinction (t_{ext}). But when comparing the t_{ext} of $R_{\text{eff}} = 2.018$ to the t_{ext} of $R_{\text{eff}} = 1.009$ (for any combination of μ and p_{i}, for example $\mu = 32$ and $p_{\text{i}} = 1/8$), an increase (from $t_{\text{ext}} = 0.3$ to $t_{\text{ext}} = 1.0$) is noticeable even if the frequency of events is cut by half (from $\lambda = 1/8$ day^{-1} to $\lambda = 1/16$ day^{-1}). This is not necessarily undesirable, as outbreaks for smaller R_{eff} are slower and take more time to take-off. Although this delays extinction, it also provides more time for health authorities to contain an outbreak. In fact, for the case where $R_{\text{eff}} = 2.018$, extinction (that occurs with a probability P_{ext}) will most likely occur before the first positive test is reported (i.e., $t_{\text{ext}} < 0$), and non-extinct outbreaks produce a large number of infections after 60 days (i.e., $N_{@60} > 10^4$), suggesting that for $R_{\text{eff}} = 2.018$ outbreaks either go quickly extinct or take off considerably within 60 days. This also suggests that non-extinct outbreaks could be catastrophic, even if the frequency of events is low and the probability of extinction is high, as for the case with $R_{\text{eff}} = 2.018$, $p_{\text{i}} = 1$, and $\lambda = 1/64$ day^{-1},

where $P_{\text{ext}} = 96.4$ and $N_{@60} = 279,114$. Hence, reducing R_{eff} may increase the time until extinction, but it prevents the chances of very large outbreaks from occurring and reduces the number of individuals infected at 60 days ($N_{@60}$).

It is worth mentioning that reducing R_{eff} results in a higher value for the average total number of infected individuals at extinction (N_{ext}). For a high R_{eff}, outbreaks either take off really quickly or go extinct rapidly. The latter category of outbreak comprises paths where the first infectious individual typically infects one individual or less, so by the time extinction is reached, very few have been infected. When R_{eff} is lower, paths can still go extinct even when the first individual infects more than one person, because there is a higher probability that none of these secondary infected individuals infect others. This can thus results in a higher value for the average total number of infected individuals. For a constant R_{eff}, N_{ext} changes only slightly when varying λ, μ, or p_{i}. The dependency of N_{ext} on these parameters occurs in a non-linear fashion, similarly to the dependency of Equation (27) on λ, $\mathbb{E}[N_{\text{s}}]$ $(= \mu)$ and p_{i}. Note that the dependency of this expression on μ is particularly non linear, through the dependency of $\text{Var}[N_{\text{s}}]$ on μ. These non-linearities can explain how N_{ext} does not always increase with μ.

For any combination of λ and p_{i}, reducing the number of susceptible individuals at infectious events by half would reduce the R_{eff} by half (for example, reducing $\mu = 32$ to $\mu = 16$, for $p_{\text{i}} = 1$ and $\lambda = 1/64$ day^{-1} would reduce the $R_{\text{eff}} = 2.018$ to $R_{\text{eff}} = 1.009$). This drastically reduces $N_{@60}$ (from 279,114 to 2,303, for this example), while P_{ext} improves by only a few percentage points (from 96.4% to 99.0%, for this example). This improvement becomes more significant for larger values of λ. The effect on $N_{@60}$ holds for all combinations of λ and p_{i}, suggesting that the risk associated with catastrophic outbreaks could be managed by reducing the number of susceptible individuals at infectious events (regardless of the value of λ). Therefore, the results suggest that allowing smaller gatherings is an effective method to reduce risk.

The risk of a large number of infected individuals can also be managed by reducing p_{i} as this has a direct impact on R_{eff} (as noticeable in Equation (45)). Reducing p_{i} also improves the probability of extinction as each susceptible individual is less likely of becoming infected compared to when $p_{\text{i}} = 1$. The number of susceptible attendees becoming infected is a random process following a (binomial) distribution with a range of possible outcomes. Reducing p_{i} increases the chances that fewer become infected at each infectious event, increasing the chances of extinction. Consequently, the probability of extinction improves. Although reducing p_{i} improves P_{ext}, if the average number of attendees per event is allowed to increase to a point of insufficient reduction on R_{eff}, outbreaks with a large number of infected individuals can occur. This is illustrated in Table 5 that lists three cases with the same value of $\lambda = 1/4$ day^{-1}, but with increasing μ. The first case has $p_{\text{i}} = 1$ and $\mu = 2$, whereas cases B and C have $p_{\text{i}} = 1/8$, and $\mu = 8$ and $\mu = 16$, respectively. Reducing p_{i} resulted in better P_{ext}, regardless of the value of μ. An increase in μ while reducing p_{i} is tolerable, provided that the combined effect of increasing μ and decreasing p_{i} results in a reduction of R_{eff} (as governed by Equation (45)). Such is the situation for case B in Table 5 when compared to case A. R_{eff} is reduced from 2.018 to 1.009 producing better $N_{@60}$ for case B. But a large increase in μ can produce a larger number of infected individuals after 60 days for non-extinct outbreaks, as noticed from case A to case C where μ increased from 2 to 16 and $N_{@60}$ increased from 15,339 to 25,216. This suggests that, although more outbreaks went extinct (70.7% compared to 55.2%), some outbreaks with a lot of transmission are not prevented when reducing p_{i}. This is due to the fact that the number of susceptible attendees follows a logarithmic distribution; increasing μ pushes the tail end of the distribution towards a higher number of susceptible individuals. This means that when $\mu = 16$ more events may have a large number of susceptible individuals that can become infected (with a probability of p_{i}). Having a lot of susceptible individuals sets the conditions for super-spreading events (i.e., events where there are lots of infections). Reducing p_{i} does not prevent these super-spreading events, but rather increases the

chances that fewer get infected. As demonstrated with case B, it is possible to reduce p_i while increasing μ without increasing $N_{@60}$, provided that the combined effect of increasing μ and decreasing p_i results in a sufficient reduction of R_{eff}.

Table 5: *Simulation I results subset demonstrating the impact of p_i on results for constant $\lambda = 1/4$ and increasing μ.*

Case	R_{eff}	p_i	μ	λ day^{-1}	P_{ext} %	$N_{@60}$
A	2.018	1	2	1/4	55.2	15,339
B	1.009	1/8	8	1/4	93.3	201
C	2.018	1/8	16	1/4	70.7	25,216

The two variable that were controlled by health authorities at the beginning of the COVID-19 outbreak were the frequency of events (λ) and the size of gatherings (captured here by μ). These results clearly show that the size of group gathering has a first order impact on the total number of infected individuals (for both extinct and non-extinct outbreaks), and that controlling the frequency of events has a first order impact on the chances of extinction. The results also show that reducing the probability of infection (p_i) can improve the chances of extinction, but if λ or μ are allowed to increase to a level that counters the improvement of p_i on R_{eff}, the improvement on P_{ext} may only be marginal (if λ is too high) and there are risks of large outbreaks (if μ is too high). Finally, the results also show that to determine the outbreak risks associated with a given scenario, it is important to consider several metrics. For instance, a high P_{ext} did not guarantee a low number of infected individuals for non-extinct outbreaks.

5.2.4 Conclusions

The probability of extinction is not linearly proportional to R_{eff}. Therefore, R_{eff} alone (if above 1) is an insufficient metric to describe the spread of a virus. μ, λ and p_i not only impact R_{eff}, but also the probability of extinction, the time of extinction, and the total number of infected individuals. These are all important measures to help describe the spread of the virus.

The results suggest that having a low frequency of infectious events can prevent the transmission of the disease and, in the context of the selected disease parameters (Table 1), it is the most effective course of action to improve the chances of extinction. This may increase the time for outbreaks to reach extinction, but this is not undesirable as a slower outbreak allows authorities more time to react and control an outbreak.

Outbreaks that have high probability of extinction may produce catastrophic results if the number of susceptible individuals at infectious events is large. Reducing the number of susceptible individuals at infectious events may not have a large impact on the probability of extinction, especially when the rate of events is low, but it sets up an environment that prevents events where super-spreading may occur. Reducing p_i reduces the chances of susceptible individuals from becoming infected, which improves the chances that fewer individuals become infected (at each event). Although this improves the probability of extinction, if the susceptible population is large, a single non-extinct outbreak could be catastrophic as reducing p_i reduces the chances of, rather than prevents, susceptible individuals from becoming infected.

5.3 Simulations II: Demonstration of the modified branching process using a real-world scenario.

The previous section evaluated the IBP's theoretical application. As mentioned in Section 3.5, there are scenarios that require additional constraints that make the simulation and/or statistics deviate from a branching process, and the IBP was modified to enable some of these constraints. This section shows how the model can be applied to this type of scenario, and how to determine the impacts of different mitigation strategies that require the application of some of these constraints.

5.3.1 The Diamond Princess: Chronology of events and factors contributing to disease transmission.

The Diamond Princess cruise ship departed Japan on January 20th 2020 (Figure 5) [40]. On January 25th, the presumed primary infected individual left the ship. Some report that this individual was coughing before boarding the ship [38] while others report that coughing did not begin until the 23rd of January [41]. This individual tested positive on February 1st, while on February 3rd the ship was quarantined from the rest of the world. Initial testing was performed on individuals with symptoms and their close contacts. Passengers were isolated from each other starting February 5th. Crew members, on the other hand, continued working, even if they had been in direct contact with a confirmed case. On February 4th, 10 COVID-19 cases were confirmed on board the Diamond Princess.

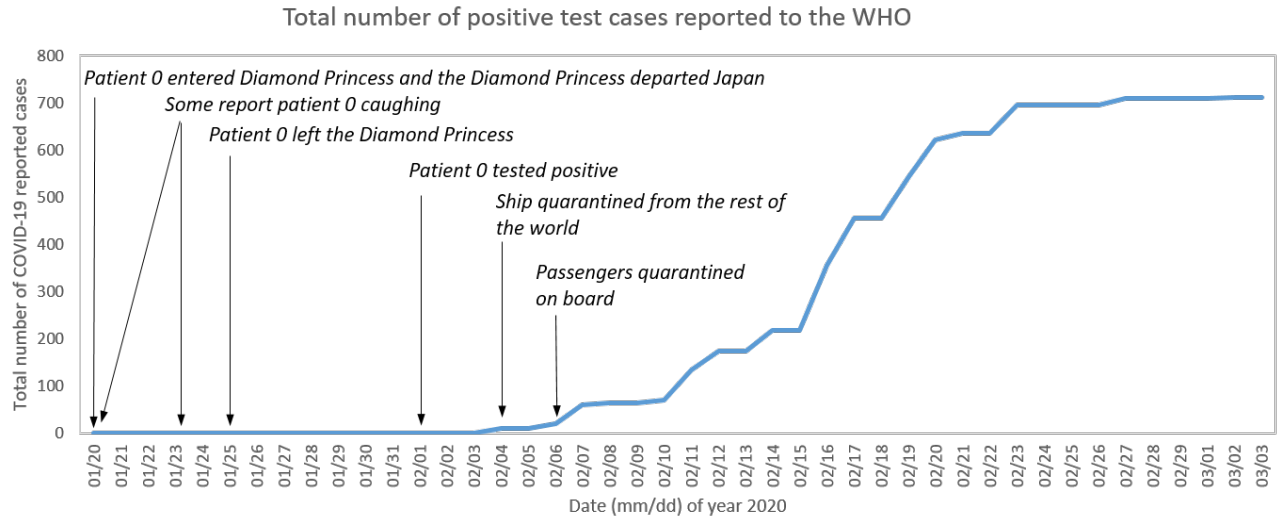

Figure 5: *Total number of infected individuals with a positive test reported to the WHO on the Diamond Princess over the course of the outbreak, along with marked events obtained from the literature.*

By definition, a reported case is a "person with laboratory confirmation of SARS-Cov-2 infection, irrespective of clinical signs and symptoms" [42]. There were a total of 2666 passengers and 1045 crew members (for a total of 3711 people) on board the Diamond Princess [40]. A total of 706 were confirmed cases. The outbreak's chronology of events as well as the number of reported cases to the World Health Organization (WHO) are well illustrated and documented in [40].

There are reports that mitigation strategies to control the spread of SARS-Cov-2 on the Diamond Princess were not properly executed [43]. Although the ship was quarantined 3 days after the primary individual

reported positive to COVID-19 and that contact tracing was executed on the ship to help control the outbreak among those on board, it has been reported that the following may have contributed to the spread of the virus:

- The suspected primary individual did not self isolate on board of the Diamond Princess, even if he/she may have had symptoms while on the ship [38, 41].
- Crew members that were in direct contact with a confirmed case did not self isolate [40, 43].
- A red zone with proper decontamination was not set into place [43].

The modified branching process was set up to simulate the outbreak that occurred on the Diamond Princess and to evaluate the impact of mitigating the three items listed above.

5.3.2 Assumptions and parameters of the Diamond Princess outbreak

5.3.2.1 Disease and interaction events input parameters

The disease parameters for this scenario are the same as the previous example and are summarized in Table 1.

Since the outbreak on the Diamond Princess had already begun once the first positive test case was reported, the interaction parameters (i.e., μ, λ, and p_{i} described in Section 4) were selected to represent the social dynamics of the individuals on the ship before mitigation strategies were set in place.

Because of the following information and assumptions, this work assumed that $\mu = 4$ (i.e., $\mu = g_{\mathrm{ave}} - 1$), that $\lambda = 4$, and that the susceptible attendees were drawn from a logarithmic distribution:

- Since passengers can attend various events on the Diamond Princess, and that there are several dining rooms, and exercise rooms, it was assumed that an infected passenger could produce four different infectious events per day (including meal time, and bed time).
- Per cabin, the average number of passengers was reported to be 1.98, whereas the average number of crew members was reported to be 1.73 [40]. This suggests that most passengers travelled in pairs, and that the majority of crew members were roomed with another crew member.
- It was assumed that passengers that travelled in pairs congregated with other passengers that travelled in pairs (for a total of four individuals), and that these groups were often accompanied or requested assistance from a crew member (at meal time, for example). Resulting in an average group size of five, at infectious events.
- The number of susceptible individuals at infectious events are drawn from a logarithmic distribution (`group_log`) to account for large group gatherings (such as entertainment shows).

The probability of infection at each infectious event was set to 7% (i.e., $p_{\mathrm{i}} = 0.07$), because it was assumed that:

- The initial 10 cases reported on the ship by February 4th resulted from the same outbreak.
- These 10 positive cases were most likely to have become infected over the 9-day period ranging between January 21st and January 30th.[12]

[12] Infectious events began somewhere between January 20th (when the ship departed) and January 25th (when the primary individual left the ship). Testing is most effective five days after being infected [44], suggesting that these positively reported individuals were most likely infected on or prior to January 30th.

This suggests that between January 21st and January 30th, an average of 1.11 individuals were infected per day. At 4 events per day, this results in 0.2775 infected individuals per event. Since the distribution of the susceptible population at infections events was assumed to have $\mu = 4$, susceptible individuals have 0.07 probability of becoming infected (i.e., 10 positive cases / 9 days / 4 events per day / 4 susceptible individuals per event = 0.06944 positive case per susceptible individual).

5.3.2.2 Testing and contact tracing parameters

Testing (and contact tracing) began on February 3rd. Since testing and contact tracing were performed in a controlled and constrained environment, it was assumed that:

- Although there may have been delay receiving the first reported case, it was assumed that subsequent tests were performed and results were obtained rapidly with no testing delay (i.e., $T_{\Delta t} = 0$).
- Tests could be performed several times to confirm results. This was simulated by configuring the true positive test rate for all those who tested to 100% (i.e., $t_{\text{t.p.r.}} = m_{\text{t.p.r.}} = 100\%$).
- Those who tested positive were able to recall who they had been in contact with over the last $w_{\text{ct}} = 5$ days.
- The probability of successful contact tracing was set to $P_{\text{t}} = 1.0$.
- The delay in contact tracing and isolating individuals was short with an average of 0.5 days ($\bar{t}^{\text{i}} = \bar{m}^{\text{i}} = 0.5$ day) and 95th percentile of $t^{\text{i}}_{95} = m^{\text{i}}_{95} = 1.5$ days.

Aboard, if in contact with a positive case, asymptomatic crew members were not self isolated but asymptomatic passengers were; therefore P_{i_t} was set to the ratio of passengers to total number of people on board (i.e., $P_{\text{i}_\text{t}} = 2666/3711 = 0.72$). All symptomatic individuals were assumed to self isolate at symptom onset or when in contact with a positive test case (whichever occurred first). Evidently, crew members may not have self isolated once contacted prior to having symptoms, but having been in contact with a positive case would make them more aware of symptoms related to the COVID-19 disease. As a result, $P_{\text{i}_\text{m}} = 1$.

5.3.2.3 Simulation options

To simulate the Diamond Princess outbreak, other assumptions and simulation options were selected.

`pri_no_alt_period`: Since information about the primary individual seems speculative, and 10 individuals were already infected by the time measures were set in place, we assumed that the primary individual was non-self-isolating.

`observable_paths_only`: Since the outbreak was observed once the primary individual tested positive, only outbreaks with at least one positive test were used for the analysis. Those without a reported positive test were discarded.

`time_rel_pri_end_comm`: The simulation time is relative to the end of the primary individual's infectious phase. This time was assumed to be equivalent to the first positive test that was reported to the WHO, with a potential 3 day bias. This is explained by Figure 6 that shows the simulation timeline with respect to some events

reported about the Diamond Princess outbreak.[13] This figure shows that the end of the infectious phase of the primary individual most likely occurred 2 to 3 days before the first test was reported to the WHO.[14]

The parameters of the Diamond Princess outbreak are summarized in Table 6, and are referred to as the *baseline simulation* or `Sim0`. From this baseline simulation, some parameters were modified to simulate different mitigation strategies.

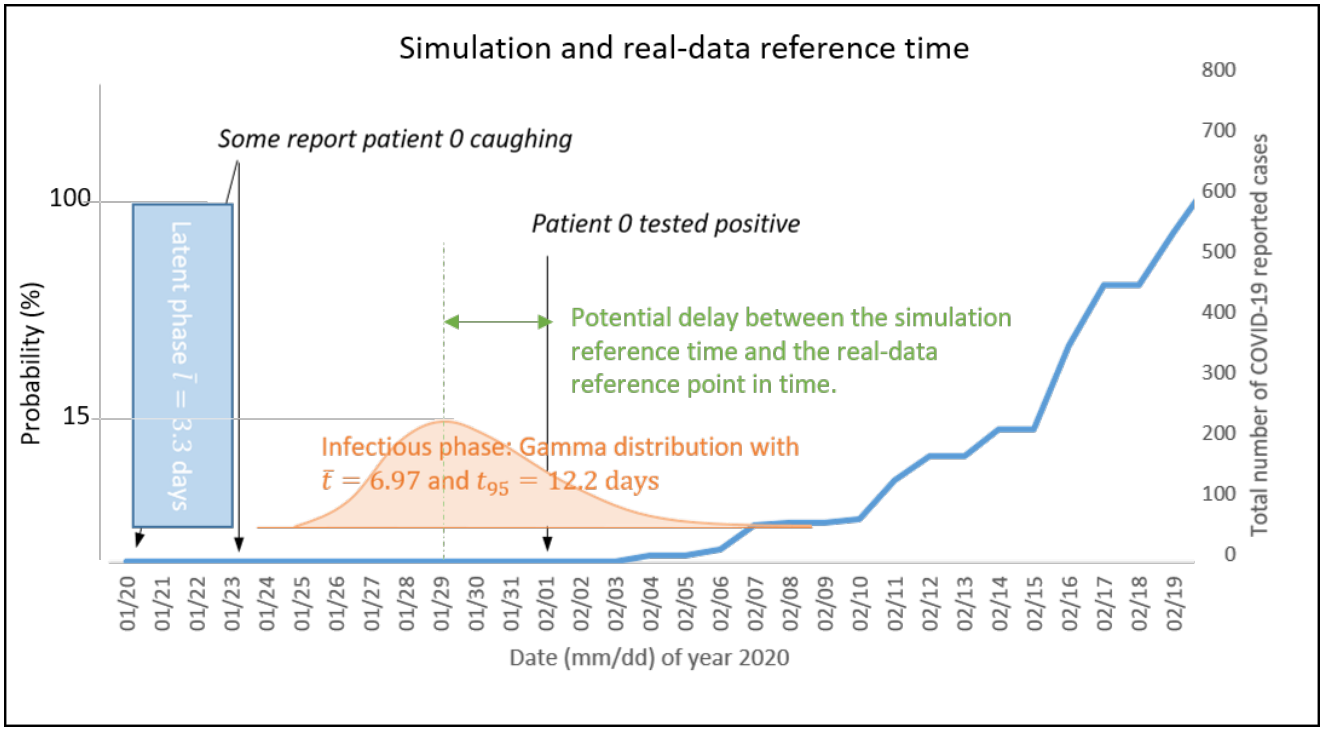

Figure 6: *Total number of infected individuals with a positive test reported to the WHO on the Diamond Princess over the course of the outbreak, along with marked events obtained from the literature.*

5.3.3 Assumptions and parameters of the mitigation strategies

The following five (5) scenarios with different mitigation strategies were simulated from the Diamond Princess baseline simulation:

1. Isolating all asymptomatic individuals (not just passengers) that were in direct contact with a confirmed cases. This is simulated with $P_{\text{i}_\text{t}} = 1.0$ (instead of 0.72). This simulation is referred to as `Sim1`.

2. Minimizing the probability of infection. Although it is unclear how masks, decontamination, social distancing, or red-zones impact the probability of infection, reducing the probability of infection by half, that is $p_\text{i} = 0.035$, was simulated. This simulation is referred to as `Sim2`.

[13] It must be mentioned that the branching model cannot simulate individuals that leave the population, as was the case for the primary individual that left the ship on January 25th. Regardless, other individuals on-board the ship continued infecting others, suggesting that the departure of a single individual may have little impact on the simulation results.
[14] The choice of setting up the simulation time relative to the end of the primary individual's communicable period rather than the first reported test will become apparent in the discussion of the results

Table 6: *List of parameters for the Diamond princess simulations listed in Table 7.* `Sim0` *represents the baseline scenario.*

Parameter		`Sim0`	`Sim1`	`Sim2`	`Sim3`	`Sim4`	`Sim5`
μ		4	4	4	4	4	4
λ	(day^{-1})	4	4	4	4	4	4
p_{i}		0.07	0.07	0.035	0.07	0.07	0.07
$T_{\Delta t}$	(days)	0	0	0	0	0	0
$t_{\text{t.p.r.}}$		1	1	1	1	1	1
$m_{\text{t.p.r.}}$		1	1	1	1	1	1
P_{t}		1	1	1	1	1	1
w_{ct}	(days)	5	5	5	5	5	5
$P_{\text{i}_{\text{t}}}$		0.72	1	0.72	0.72	0	0
$P_{\text{i}_{\text{m}}}$		1	1	1	1	0	0
$\bar{m}^{\text{i}}$, $\bar{t}^{\text{i}}$	(days)	0.5	0.5	0.5	0.5	n/a	n/a
n_{paths}		10^4	10^4	10^4	10^4	10^4	10^4
t_{max}	(days)	30	30	30	30	30	30
`group_log`		✓	✓	✓	✓	✓	✓
`time_rel_pri_end_comm`		✓	✓	✓	✓	✓	✓
`observable_paths_only`		✓	✓	✓	✓	✓	✓
`pri_no_alt_period`		✓	✓	✓		✓	
`pri_no_main_period`					✓		✓

3. Self-isolating the primary individual as soon as he/she had symptoms. This is simulated with option `pri_no_main_period`. For such a simulation, the time would be relative to the one at which the primary individual self isolated (`time_rel_pri_end_comm`). This simulation is referred to as `Sim3`.

4. Disabling all contact tracing efforts at the beginning of, and throughout, the outbreak, for comparison purposes. This is simulated with $P_{\text{i}_{\text{t}}} = P_{\text{i}_{\text{m}}} = 0$.This simulation is referred to as `Sim4`.

5. Disabling all contact tracing efforts at the beginning of, and throughout the outbreak, but the primary individual self isolated as soon as he/she had symptoms. This is simulated with $P_{\text{i}_{\text{t}}} = P_{\text{i}_{\text{m}}} = 0$ and with option `pri_no_main_period`. For such a simulation the time would be relative to the one at which the primary individual self isolated (`time_rel_pri_end_comm`). This simulation is referred to as `Sim5`.

The simulation names and descriptions are summarised in Table 7; their respective parameters are summarized in Table 6. The parameters that differ from the baseline simulation are shown in red font.

5.3.4 Methodology

All simulations were performed on 10^4 observable outbreaks for 30 days.

For each simulation, the average number of confirmed cases (i.e., positively reported tests) across all outbreaks was computed over time and compared to the number of positively reported cases to the WHO. The average number of infections over all paths and over non-extinct paths was also computed over time. The average number of infections for extinct paths is irrelevant to this study as the branching model is simulating the initial outbreak that occurred on the Diamond Princess well before it reached extinction.

Table 7: *List of Diamond Princess simulations.*

Name	Description
`Sim0`	Baseline
`Sim1`	Contact tracing of both passengers and crew members in direct contact with a confirmed positive case
`Sim2`	Lowered probability of infection
`Sim3`	Primary self isolated as symptoms appeared
`Sim4`	No contact tracing was performed at the beginning of the outbreak
`Sim5`	No contact tracing was performed at the beginning of the outbreak, but the primary individual self isolated

Finally, the probability of extinction at 30 days[15] and the time of extinction was also computed for each simulation (`Sim0` to `Sim5`).

5.3.5 Results

The total number of confirmed cases, and of infected individuals over time averaged across all 10^4 outbreaks is illustrated in Figure 7 and in Figure 8, respectively, whereas the total number of infected individuals over time averaged across the non-extinct paths is illustrated in Figure 9. These plots also show the real data representing the number of reported cases to the WHO for the first 30 days of the outbreak. In Figure 7, the real data can be compared to the simulation results, but in Figure 8 and 9, the real data is plotted as a reference as it represents only the positive reported cases rather than the number of infections.

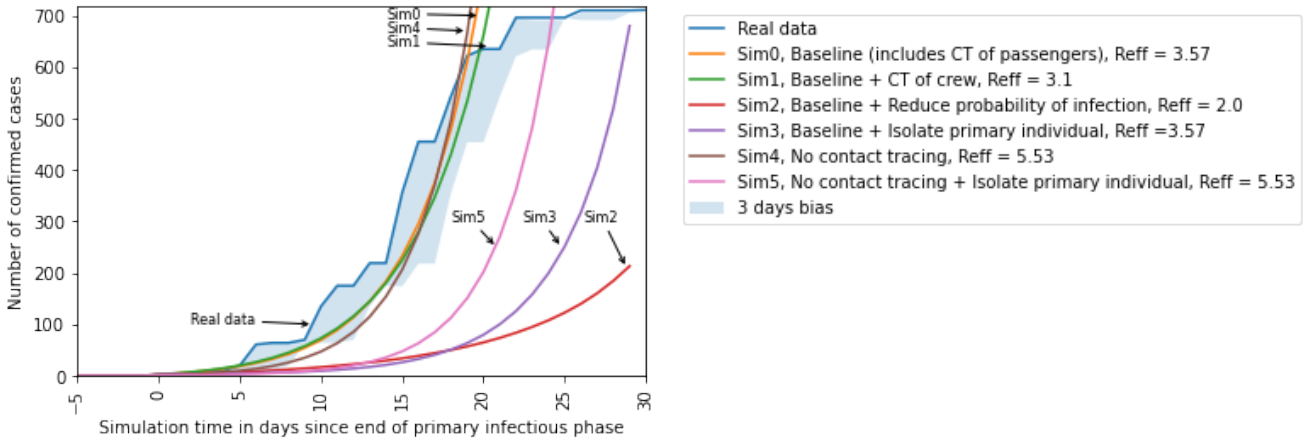

Figure 7: *Total number of confirmed cases over time averaged across all* 10^4 *observable outbreaks.*

As previously mentioned, simulations are relative to the end of the primary individual's communicable period, and the real data is relative to the first case reported by the WHO. Also, from Figure 6, there may be up to a 3-day difference between the two time references, illustrated through the light blue area in Figure 7. Regardless, the baseline simulation (`Sim0`) follows the real data, and `Sim2`, `Sim3` and `Sim5` are shifted in time from the baseline simulation. In Figure 8 and Figure 9, the results of `Sim0` suggest

[15] Simulation of 30 days represents the time frame of the available data obtained from the WHO. It is also sufficient to capture extinct outbreaks across most simulations, with the exception of `Sim2` as will be described in the Section 5.3.6.

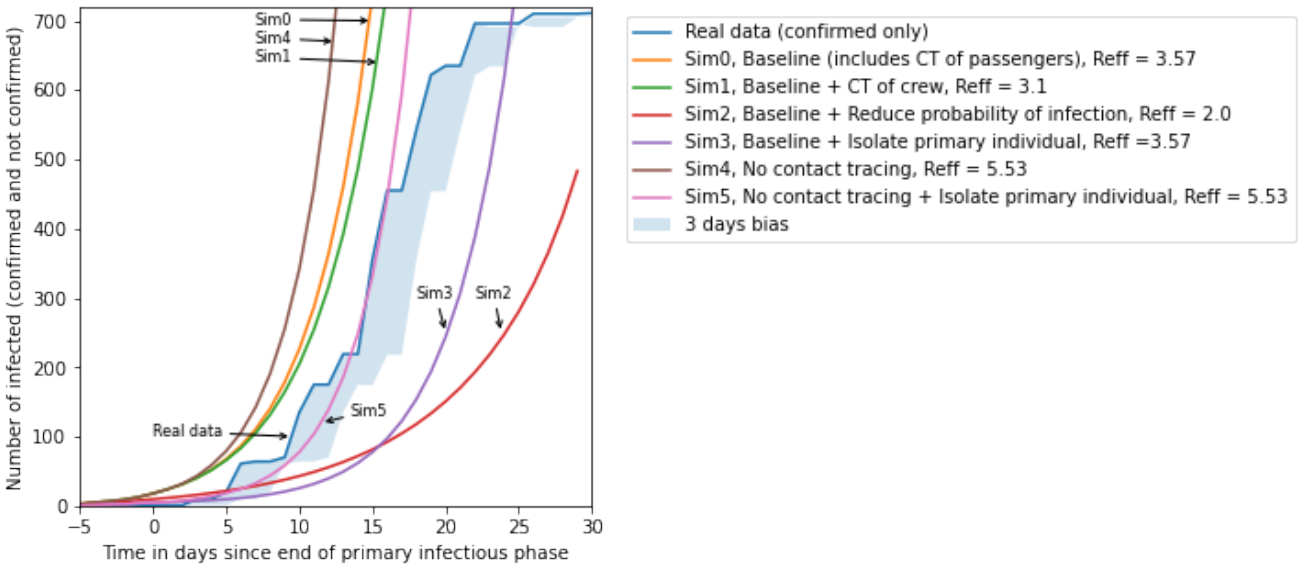

Figure 8: *Total number of infected individuals over time averaged across all* 10^4 *observable outbreaks.*

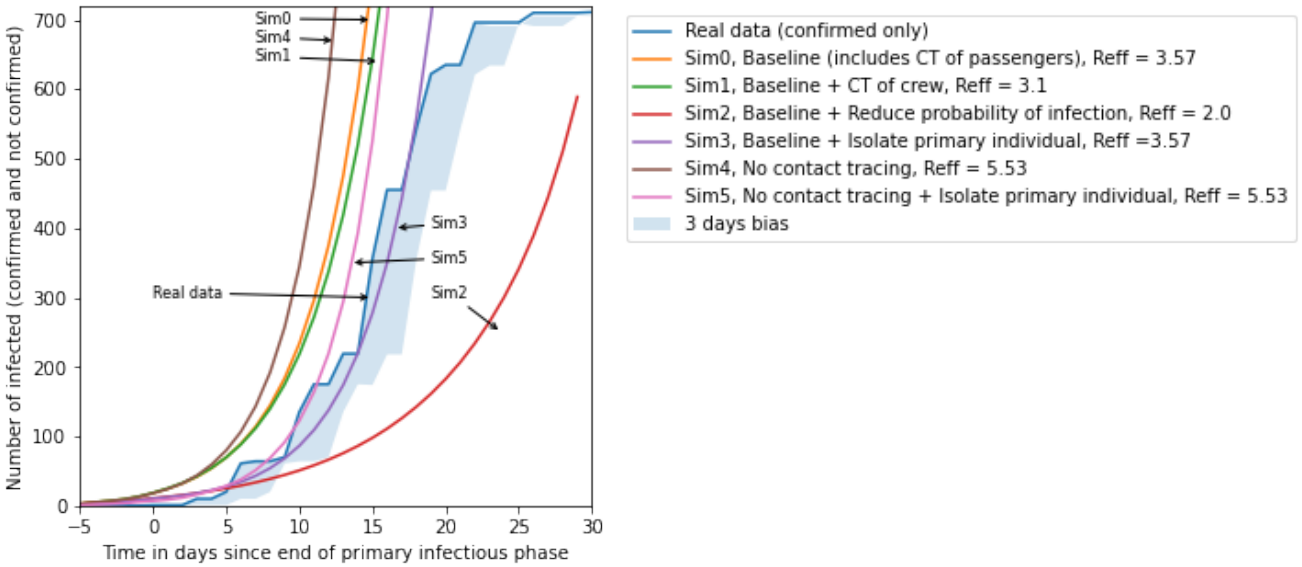

Figure 9: *Total number of infected individuals over time averaged across all outbreaks that are non extinct.*

that by the time the first test was reported (i.e., when the real data begins to climb), there were a lot more infections than reported cases. The results also show that `Sim4` obtained the highest numbers of infected individuals and the fastest infection rate. `Sim5` is shifted from `Sim4` in time. `Sim3` and `Sim0` are also shifted in time from each other. The results also show that `Sim2` obtained the slowest infection rate. However, in Figure 8, `Sim3` had the least number of infected individuals in the first 15 days.

It is worth mentioning that if the simulations were relative to the first reported test (as opposed to the end of the primary individual's communicable period), the simulation curves for `Sim2`, `Sim3` and `Sim5` would be shifted to the left in Figure 7. Also, the number of infected individuals would have increased

for all simulations in Figure 8 and Figure 9, while the relative difference between each would have been similar. As a result, the impact of the different mitigation strategies on the number of reported cases would have been less obvious, and the number of infected cases would have appeared higher.

For clarity, only averages are reported, although spread in the results are expected. To understand this spread in the results, the distribution in the average number of infected individuals illustrated in Figure 10 can be considered. The distribution is averaged over all infectious individuals across all outbreaks for a given simulation. When this distribution is larger (i.e., longer along the x-axis), the spread in the number of reported or infected individuals is expected to be larger in Figures 7 to 9.

In Figure 10, relative counts below 0.1% are not shown. The results show a peak for 0 secondary infection, that is, between 47% (`Sim0`, `Sim2` and `Sim3`) and 54% (`Sim1`) of infectious individuals did not infect others. For `Sim4` and `Sim5`, 20% of infectious individuals did not infect others. The corresponding cumulative distribution (shown in Figure 11) illustrates that for `Sim2`, close to 0% of the infectious individuals infected more than 15 other people, whereas for the other simulations between 5% and 8% of the infectious individuals infected more than 15 other people.

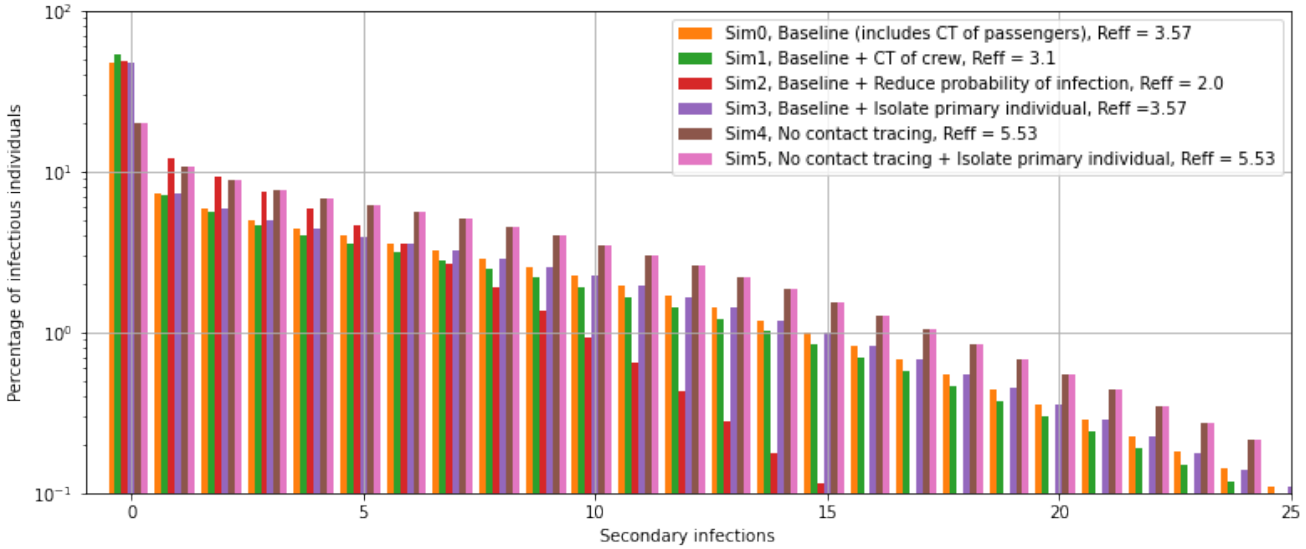

Figure 10: *The distribution of the average number of secondary infections caused by an infectious person. The y-axis shows the probability of occurrence between 0.1% and 100% across all* 10^4 *simulated outbreaks.*

The probability of extinction after 30 days as well as the time of extinction for each simulation are listed in Table 8.[16] These results show that `Sim3` obtained the highest probability of extinction followed by `Sim5` and then `Sim2`. `Sim4` obtained the lowest probability of extinction. Extinct outbreaks reached extinction within 0.456 and 5.389 days on average from the end of the primary individual's infectious phase, which is, for `Sim0`, `Sim1`, `Sim2`, and `Sim4`, the time at which the primary individual left the ship, and for `Sim3` and `Sim5`, the time at which the primary individual would have self isolated.

[16] The average days to extinction is relative to the end of the primary individual's communicable period. For `Sim0`, `Sim1`, `Sim2`, and `Sim4` this is analogous to the time at which the primary individual is non infectious. For `Sim3` and `Sim5` this is analogous to the time at which the primary individual would have self isolated. A time of extinction shorter than one (1) day indicates that if extinction were to occur, it would have occurred before the end of the primary individual's communicable period.

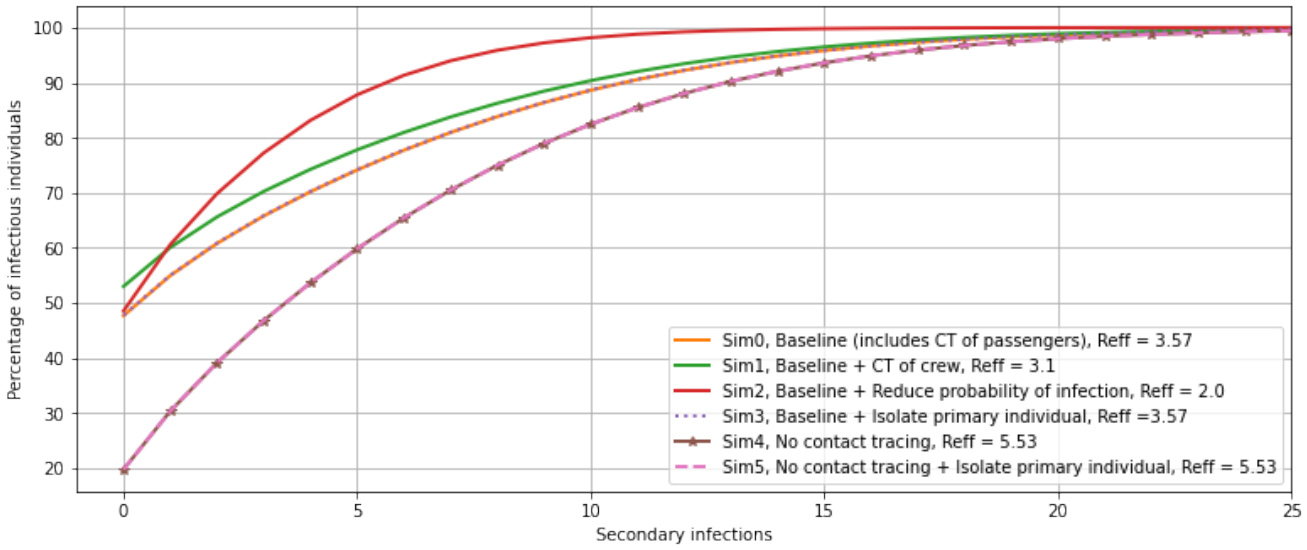

Figure 11: *The cumulative distribution of the average number of secondary infections caused by an infectious person across all 10^4 simulated outbreaks, in percentage points.*

Table 8: *Probability of extinction (after 30 days) and average time to extinction across all extinct outbreaks, for each simulation.*

Metric	Sim0	Sim1	Sim2	Sim3	Sim4	Sim5
Simulated R_{eff}	3.575	3.101	2.000	3.567	5.527	5.525
Probability of extinction (%)	3.20	6.60	18.2	72.8	0.80	37.0
Average days to extinction	3.048	4.653	5.389	0.573	1.400	0.456
± (days)	3.407	4.177	5.996	1.280	2.254	1.456

5.3.6 Discussion of results

The results from Table 8 suggest that the probability of the outbreak going extinct within 30 days would have increased by 70% (`Sim3` compared to `Sim0`) if the primary infected individual that boarded the Diamond Princess would have self isolated once symptoms appeared. Self isolation of the primary individual alone without the measures taken on the Diamond Princess during the outbreak (`Sim5`) may have also improved the probability of extinction (from 3.2% to 37.0%). These results suggest that, for the Diamond Princess scenario, the infectious phase of the primary individual impacted the probability of extinction. This is expected because the outbreak has more chances of taking off uncontrollably if the first individual produces more secondary infections, which is more likely to occur when the infectious phase is longer. Self isolating at symptom onset can shorten the infectious phase. Self isolation of the primary individual does not only improve the chances of extinction, it also shortens the time until extinction. In fact, for `Sim3` and `Sim5`, extinction occurred, respectively, at 0.573 and 0.456 days on average from the end of the primary individual's communicable period, suggesting that extinction occurs with the self isolation of the primary individual most of the time.

Although self isolation of the primary individual increases the likelihood of extinction, this control measure does not necessarily suppress the total number of infected individuals, but rather shifts the outbreak in time, as noticeable when comparing `Sim0` (baseline without primary individual self isolating) and `Sim3` (baseline with primary individual self isolating), and when comparing `Sim4` (no control measures) and `Sim5`

(primary individual self isolating only) in Figures 8 and 9. Such a time shift could have given authorities more time to better respond to the outbreak, but this time difference may be less than five days for non-extinct paths (when comparing the number of infected individuals in non-extinct paths for `Sim0` and `Sim3` [Figure 9]). The fact that the distribution for the secondary number of infected individuals (Figures 10 and 11) are identical for `Sim0` and `Sim3`, and for `Sim4` and `Sim5`, suggests that the outbreaks' dynamics are actually similar, although shifted in time. This is expected because the simulations have the same R_{eff} value (3.57 for `Sim0` and `Sim3`, and 5.53 for `Sim4` and `Sim5`). These results indicate that self isolation of the primary individual could have given authorities more time to respond to the outbreak, but additional measures would have been required to change the dynamic of the outbreak in order to mitigate the number of infected individuals. This is especially true when outbreaks do not go extinct.

The dynamics of the outbreak are changed when contact tracing is applied, as it causes the distribution for the number of secondary infections to change. In fact, the number of individuals causing zero secondary infections for simulations with contact tracing (`Sim0`, `Sim1`, `Sim2` and `Sim3`) increases between 48% and 54% from 20% for simulations without contact tracing (`Sim4` and `Sim5`), in Figures 10 and 11. This is expected, as contact tracing can isolate individuals before they become infectious, reducing the R_{eff} from 5.53 for simulations without contact tracing (`Sim4` and `Sim5`) to 3.57 for comparable simulations with contact tracing enabled (`Sim0` and `Sim3`). Contact tracing starts reducing the number of infected individuals in non-extinct outbreaks 5 days after the end of the primary individual's communicable period, as illustrated in Figure 9 (when comparing `Sim1` to `Sim4`), but when only looking at reported cases, the impacts of contact tracing may only be observable 15 days after the end of the primary individual's communicable period (when comparing `Sim1` to `Sim4` in Figure 7). Unfortunately, contact tracing has little impact on the initial stages of the outbreak, and by the time the benefits of contact tracing are observed, there are several infected individuals. In reality, contact tracing efforts may become less productive as more resources are required for an increasing number of infected individuals. The simulations do not model depleting contact tracing resources, but rather assumes that the contact tracing efforts are executed consistently. Additionally, when comparing `Sim1` to `Sim5` in Figure 9, it is noticeable that, at the initial stages of a non-extinct outbreak, self isolation of the primary individual will help minimize the number of infected individuals more efficiently than consistent contact tracing. But as mentioned previously, over time, self isolation of the primary individual would have been insufficient to slow down the number of infected individuals, especially in the case of a non-extinct outbreak.

Reducing the probability of infection could have slowed down the infection rate over time, as noticeable in Figure 9 for `Sim2`. This reduction in the rate of infections can be explained by the fact that there are fewer infected individuals with more than 15 secondary infections, as noticed in Figures 10 and 11. From these figures, it is interesting to note that, contrary to contact tracing, reducing the probability of infection (`Sim2`) had no noticeable impact on the number of individuals with 0 secondary infections. This suggests that reducing the number of super spreaders, rather than increasing the number of individuals with 0 secondary infections, could have helped reduce the rate of infection during the Diamond Princess outbreak. However, reducing the probability of infection produced more infected individuals at the start of the outbreak when compared to simulations where the primary individual self isolated (`Sim3` and `Sim5`) in Figures 8 and 9, and only provided 18% probability of extinction (Table 8). Note that the latter effect is time dependent. For 30 day simulations, the average number of days to extinction was relatively large (5.389 days) and had a large variation (5.996 days), suggesting that increasing the simulation time would increase both the probability of extinction, and the time of extinction. These results indicate that, reducing the probability of infection may not significantly impact the start of the outbreak (like self isolating the primary individual), but effectively improves the probability of extinction and slows down the outbreak, even though it may take longer to reach extinction because the outbreaks occur at a slower rate.

5.3.7 Conclusion

The simulated results suggest that if the primary infected individual would have self isolated, the outbreak's likelihood of going extinct would have increased from 3% to 73% at 30 days after the end of the primary individual's communicable period. In the event of a non-extinct outbreak, the simulations show that such a mitigation strategy is insufficient at reducing the number of infected individuals over time, as it only slows down the outbreak at the beginning. Although this allows health authorities more time to respond to an outbreak, it may not prevent large outbreaks over time.

Contact tracing efforts will improve the R_{eff} of an outbreak, reducing the number of infected individuals, as some individuals are isolated before the start of their infectious period. However, the impact of contact tracing efforts takes time to be observable, and may be over-estimated in these simulations as they do not account for depleting contact tracing resources as the number of infected individuals increases.

These simulations suggest that reducing the susceptible population's probability of infection would have been the most effective method at slowing down the spread of the virus once it had been observed.

5.4 Simulation III: Vaccination

In January 2021, COVID-19 vaccines started to be administered in Canada. It is well known that vaccines can provide:

- Efficacy against infection for individuals that are vaccinated. That is, vaccinated individuals are less likely to get infected when in contact with an infectious individual.
- Efficacy against onward transmission for individuals that are vaccinated. That is, vaccinated individuals that are infected are less likely to become infectious or to pass on the virus to others.

The impact of vaccination on the transmission of SARS-Cov-2 in a ship environment was simulated in support of the CAF. At the time of simulation, the vaccine's efficacy against onward transmission and against infection was unknown; therefore, this work sweeps all possible values. Also, at the time of simulation when vaccination was starting to roll out and was not mandatory, the vaccine uptake of the CAF was estimated to be around 60%, whereas the desired uptake was in the 95% range. Both of these vaccination uptake levels were evaluated.

5.4.1 Methodology, assumptions, and parameters

The disease time periods used for this simulation are presented in Table 1.

The RCN operates small to medium sized vessels. As such, the mitigation strategies on these ships are limited (due to limited space), and contacts between individuals can be high. Therefore, the parameters in Equation (47) are chosen assuming that self isolation is to be limited to only severe symptomatic individuals (corresponding to approximately $q = 5.9\%$ of infected individuals). The effective reproduction number, without vaccine, of the scenario is selected to be $R_{\text{eff}} = 2.5$, which corresponds to the value for navy vessel scenarios [2] used at the time of simulation. The simulated event rate λ and group size distribution parameters p and g_{ave} for the described scenario are presented in Table 9. These parameters indicate that an infectious individual comes into close contact[17] with, on average, $(g_{\text{ave}} - 1)\lambda \approx 9.6$ susceptible individuals per day, and each individual has a probability of becoming infected of $p_{\text{i}} \approx 4\%$ at

[17] A close contact is defined here as less than 2 metres for more than 15 minutes.

each event. The scenario also assumes that vaccination has no impact on self isolation (i.e., $q = q_v$), nor on the disease communicable period. Note that these values were assumptions based on the information available to the RCN at the time of simulation.

Simulations with vaccination were performed for different combinations of efficacy against infection (E_i) and efficacy against onward transmission (E_t). The former is equivalent to $E_i = 1 - r_{v,inf}$ whereas the latter is equivalent to $E_t = 1 - r_{v,tr}$. Therefore, simulations with a set of $r_{v,inf}$ and $r_{v,tr}$ values ranging between 0 and 1 with a 0.025 step size were performed. The values of vaccination uptake, f_v, equal to 60% and 95% where evaluated to match the vaccination uptake of the CAF at the time of simulation and the desired vaccination uptake. This produced a total of 3362 different vaccination scenarios. One million simulations[18] were performed, each one for a 30-day period, were generated for each of these 3362 scenarios. For all simulations, the analytical reproduction number with vaccination, or $R_{eff,v}$ (Equation (47)), was confirmed to match the corresponding computed value from simulation within statistical uncertainties. The probability of extinction at 30 days was computed, and the total number of infected individuals within 30 days were averaged across all one million simulated outbreaks.

Table 9: *Parameters used for the CAF ship scenario with vaccination. A logarithmic distribution with parameter p was chosen for the number of contacts at each event. Shown values for g_{ave} and p_i parameters are approximated as they are computed from the other parameters (as described in Section 4).*

Parameter	Value	
λ	6.5249	events/day
p	5.174	
g_{ave}	2.4715	individuals
p_i	0.0400	
q	0.059	
q_v	0.059	
f_v	60% and 95%	
$r_{v,inf}$	0 to 1	steps of 0.025
$r_{v,tr}$	0 to 1	steps of 0.025

5.4.2 Results and discussion

This section presents and discusses the results for the ship scenario described in the previous section. It is worth mentioning that these results are specific to this scenario, with values of $R_{eff} = 2.5$, and a 5.9% self isolation rate amongst both vaccinated and non-vaccinated individuals (i.e., $q = q_v$). A value of 5.9% for q is much smaller than for an office scenario, for example, where self isolation would be the norm for symptomatic individuals. For a different scenario, the value of q_v could also be expected to be noticeably smaller than q. A change in any of these or any other of the simulation parameters would alter the results.

Figure 12 presents the probability of extinction for the outbreaks (on the y-axis), as a function of the $R_{eff,v}$ (on the lower x-axis) and of the reduction factor in the probability of infection due to vaccination efficacy against infection and against onward transmission combined (on the higher x-axis). The latter is equivalent to $r_{v,inf}r_{v,tr}$. For a given vaccine uptake in the population (i.e., f_v= 0.60 and f_v= 0.95) and $q_v = q$, the figure shows how the probability of extinction is mostly driven by the value of $R_{eff,v}$, which is

[18] Since R_{eff} of this case study was relatively low, compared to the previous case studies, more simulations could be performed in a reasonable amount of time.

a function of $r_{\mathrm{v,inf}} r_{\mathrm{v,tr}}$ as demonstrated in Equation (47). These results clearly show that when a vaccine scales down infections or transmission, the expected number of infected individuals per each infectious person decreases ($R_{\mathrm{eff,v}}$), improving the probability of extinction.

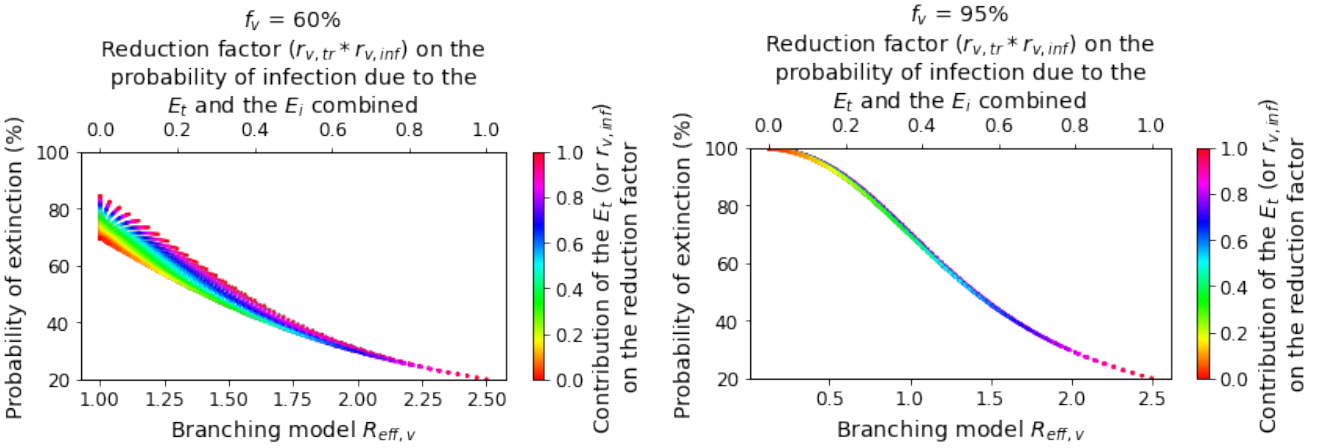

Figure 12: *Percentage of extinct outbreaks vs $R_{\mathrm{eff,v}}$, $r_{\mathrm{v,inf}} r_{\mathrm{v,tr}}$ and $r_{\mathrm{v,inf}}$, for the different simulated immunization rates ($R_{\mathrm{eff}} = 2.5$, $q = q_{\mathrm{v}} \approx 5.9\%$).*

Figure 12 also shows the separate $r_{\mathrm{v,inf}}$ parameter (on the colour bar). For f_{v}= 60% and a given value of $R_{\mathrm{eff,v}}$, it is noticeable that varying the vaccine's efficacy against infection (through $r_{\mathrm{v,inf}}$) has an impact on the probability of extinction. A vaccine that has a lower efficacy against infection (i.e., larger value for $r_{\mathrm{v,inf}}$) leads to an increase in the probability of extinction, and consequently, so does a vaccine that has a higher efficacy against onward transmission (i.e., smaller value for $r_{\mathrm{v,tr}}$). This results in a further improvement of the probability of extinction, compared to another vaccine where the efficacy against infection or onward transmission were swapped. The situation differs for $f_{\mathrm{v}} = 95\%$, where the curves for different values of $r_{\mathrm{v,inf}}$ are superposed.

This effect can also be seen in Figure 13, where the impact of varying $r_{\mathrm{v,tr}}$ (x-axis) and $r_{\mathrm{v,inf}}$ (y-axis), individually, on the probability of extinction (in the colour bar) is illustrated. For both $f_{\mathrm{v}} = 0.6$ and 0.95, when $r_{\mathrm{v,inf}} = 0$, the probability of extinction is independent of $r_{\mathrm{v,tr}}$. However, for $f_{\mathrm{v}} = 0.6$ the probability of extinction is larger when $r_{\mathrm{v,tr}} = 0$ and $r_{\mathrm{v,inf}} = 1$. This is caused by the vaccination status of the primary individual. If non-vaccinated, there is a higher expected number of new infections caused by this primary infectious individual (assuming $r_{\mathrm{v,tr}} < 1$), resulting in a more rapid take off of the outbreak producing a lower probability of extinction. If vaccinated, the expected number of new infections caused by the primary infectious individual can be lower, or even nil if $r_{\mathrm{v,tr}} = 0$. The primary infectious individual would be guaranteed to be non-vaccinated if $r_{\mathrm{v,inf}} = 0$ (i.e., a vaccinated person cannot become infected). This explains why, for low values of $r_{\mathrm{v,tr}}$, the overall average probability of extinction can be lower when $r_{\mathrm{v,inf}}$ is also small ($r_{\mathrm{v,inf}} < 30\%$), as these scenarios involve outbreaks where the infectious individual at the origin of the outbreak has a probability significantly less than f_{v} of being vaccinated. If the immunization status of the primary infectious individual was known, such an effect would not be observed.

Figure 14 presents the total number of infected individuals, including the primary individual, within the 30 days of the simulation. It shows how the level of sensitivity for the number of infections on $r_{\mathrm{v,inf}}$ and $r_{\mathrm{v,tr}}$ increases with the percentage of vaccination, with a sharper increase of the number of infections for large $r_{\mathrm{v,inf}}$ and $r_{\mathrm{v,tr}}$ values.

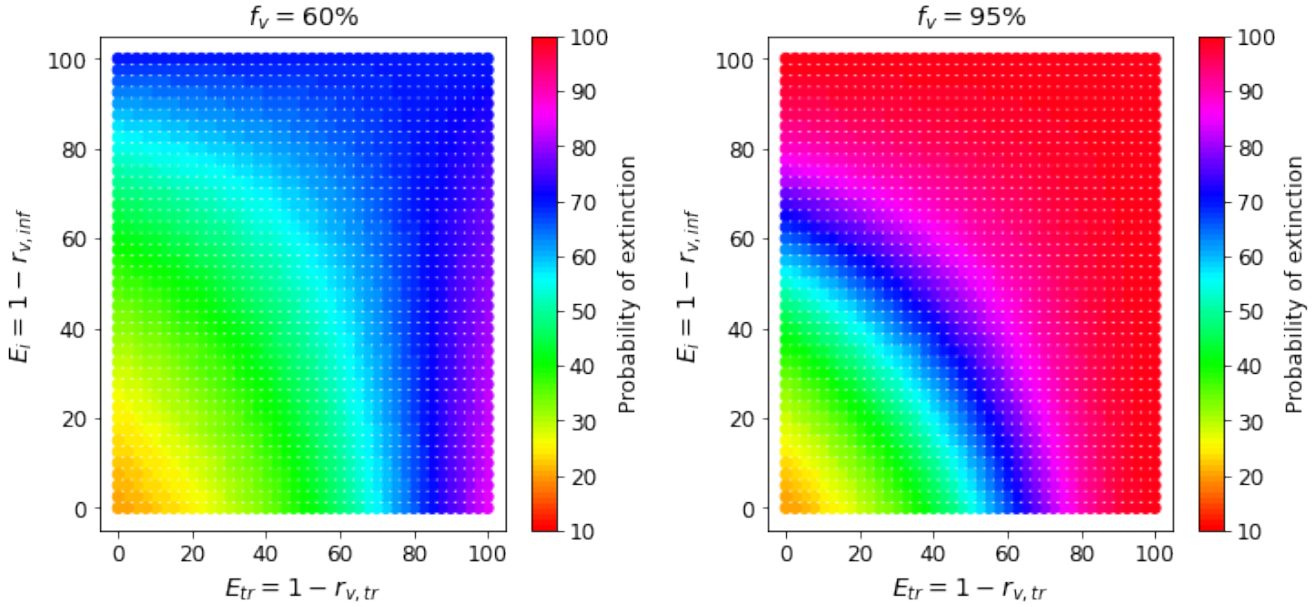

Figure 13: *Percentage of extinct outbreaks vs* $E_{\mathrm{i}} = 1 - r_{\mathrm{v,inf}}$ *and* $E_{\mathrm{t}} = 1 - r_{\mathrm{v,tr}}$*, for the different simulated immunization rates* ($R_{\mathrm{eff}} = 2.5$, $q = q_{\mathrm{v}} \approx 5.9\%$).

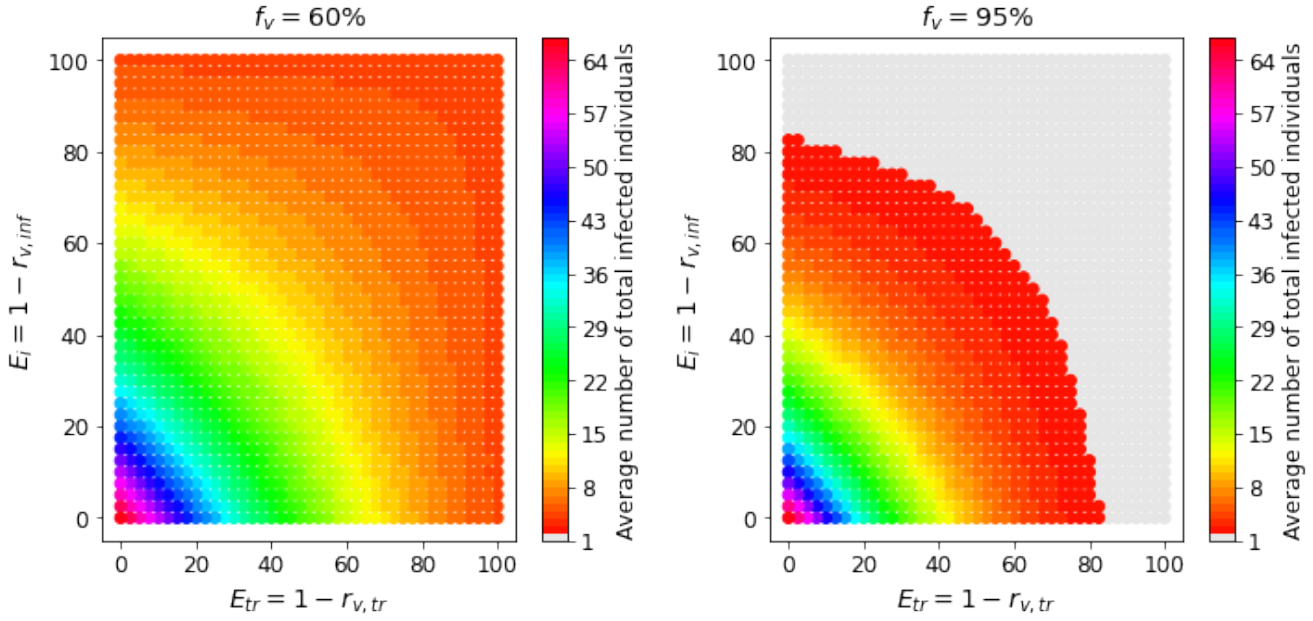

Figure 14: *Total number of infected individuals, including the primary individual, within the 30 simulation days vs.* $r_{\mathrm{v,inf}}$ *and* $r_{\mathrm{v,tr}}$*, for the different simulated immunization rates* ($R_{\mathrm{eff}} = 2.5$, $q = q_{\mathrm{v}} \approx 5.9\%$)*. Scenarios that lead to an average of less than two infections are shown in grey.*

5.4.3 Conclusion

The model that assesses the effects of vaccination on the development of COVID-19 outbreaks was applied to ship-board outbreak scenarios, assuming that:

- mitigation strategies were constrained,
- only severe symptomatic individuals self isolate,
- those infected contaminated 2.5 other individuals on average, and
- vaccination had no impact on self isolation nor on the disease communicable period.

For this specific case, the results showed that vaccination can reduce the number of infections caused by each infectious individual, which improves the outbreak's probability of extinction. The vaccine's combined efficacy against onward transmission and infection drives the probability of extinction, and the effect of one can only constitute an improvement beyond what is achieved by the other. However, for a vaccine with an efficacy against infection higher than 70%, the impact of the vaccine's efficacy against onward transmission is important when a smaller portion of the population is vaccinated, and becomes less important when a larger portion of the population is vaccinated.

6 Conclusions

In this Report, a generalized branching process suitable for the assessment of the effects of mitigation strategies on the early propagation of SARS-Cov-2 was presented. Based on previous work done for a more specific model [29], a branching process that separates interactions from transmission and that models vaccination and latent periods, was developed and implemented to perform simulations with negligible statistical uncertainties. The resulting tool can be used to simulate scenarios using this branching process, but it can also include different conditions that cause the model to deviate from a branching process. After deriving the model and presenting the input and output parameters that it involves, three simulation scenarios were presented in the Report: a branching process scenario leveraging the interaction events in the model, another based on a real scenario going beyond a branching process and a vaccination scenario. For these three scenarios, the flexibility of the model proved to be useful to study the dynamics for the early stages of outbreaks.

As branching processes rely on the assumption that probability distributions are static over time, the presented model is less suitable to evaluate propagation when a significant fraction of the population of interest is infected. Several of the distributions, along with their associated parameters that were presented in this report, could however be used by a non-branching process, while using time-varying distributions to model the interactions within the population.

References

[1] Biron, K. and Drouin, P.-L. (2021), The impact of vaccination on COVID-19 outbreaks as simulated with a branching process, Defence Research and Development Canada, Scientific Letter, DRDC-RDDC-2021-L128.

[2] Biron, K., Drouin, P.-L., Guillouzic, S., MacLeod, M. R., Schofield, S., and Jonasson, J. (2021), Analytic input to relaxation of pre-embarkation protocols for Naval vessel, Defence Research and Development Canada, Scientific Letter, DRDC-RDDC-2021-L147.

[3] Mirshak, R. and Guillouzic, S. (2021), Adjusting the probability of an unknown COVID-19 infection for differently vaccinated populations: Application at NORAD/USNORTHCOM, Defence Research and Development Canada, Scientific Letter, DRDC-RDDC-2021-L095.

[4] Drouin, P.-L., Guillouzic, S., Hunter, D. G., and MacLeod, M. R. (2022), Reactivating international locations in the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) point prevalence estimation tool, Defence Research and Development Canada, Scientific Letter, DRDC-RDDC-2022-L150.

[5] Guillouzic, S., MacLeod, M. R., Waller, D., and Bourdon, S. (2021), How flexible OR&A teams provided decision advantage through pandemic uncertainty, *Proceedings of the 15th NATO Operations Research & Analysis (OR&A) conference: Emerging and disruptive technology.*

[6] Mirshak, R. and Cazzolato, F. (2020), Analysis to support managing the spread of COVID-19 within a Mission Capability Preservation Team (MCPT), Defence Research and Development Canada, Scientific Letter, DRDC-RDDC-2020-L099.

[7] Waller, D. (2020), COVID-19 modeling summary. Briefing at US INDOPACOM meeting, 23 July 2020.

[8] Brauer, F. and Castillo-Chavex, C. (2012), Mathematical Models for Communicable Diseases, Society for Industrial and Applied Mathematics.

[9] Beech, T. (2020), Forecasting the spread of COVID-19 and the impact of non-pharmaceutical interventions, Defence Research and Development Canada, Scientific Letter, DRDC-RDDC-2020-L092.

[10] Van Den Hoogen, J. and Okazawa, S. (2020), Heterogeneous & Meta-population modelling of COVID-19. Briefing to members of Canadian Forces Health Services.

[11] Mirshak, R. and Pearce, T. (2020), Modelling the incidence of coronavirus disease 2019 (COVID-19) in Building 2 due to community spread, Defence Research and Development Canada, Scientific Letter, DRDC-RDDC-2020-L229.

[12] Rothman, K., Greenland, S., and Lash, T. (2011), Modern epidemiology: Third edition.

[13] Drouin, P.-L., randoutbreaksim (online), https://github.com/pldrouin/randoutbreaksim (Access date: 01-03-2023). Tag v1.0.

[14] Ball, F., González, M., Martínez, R., and Slavtchova-Bojkova, M. (2014), Stochastic monotonicity and continuity properties of functions defined on Crump–Mode–Jagers branching processes, with application to vaccination in epidemic modelling, *Bernoulli*, 20(4), pp. 2076–2101.

[15] Farrington, C. P., Kanaan, M. N., and Gay, N. J. (2003), Branching process models for surveillance of infectious diseases controlled by mass vaccination, *Biostatistics*, 4(2), pp. 279–295.

[16] Drake, J., Kaul, R., Alexander, L., O'Regan, S., Kramer, A., Pulliam, J., Ferrari, M., and Park, A. (2015), Ebola cases and health system demand in Liberia, *PLoS Biol*, 13(1).

[17] Chowell, G., Abdirizak, F., Lee, S., Lee, J., Jung, E., Nishiura, H., and Viboud, C. (2015), Transmission characteristics of MERS and SARS in the healthcare setting: a comparative study, *BMC Medicine*, Vol. 13.

[18] Jacob, C. (2010), Branching processes: their role in epidemiology, *International Journal of Environmental Research and Public Health*, 7(3), pp. 1186–1204.

[19] Heffernan, J. M., Smith, R. J., and Wahl, L. M. (2005), Perspectives on the basic reproductive ratio, *Journal of the Royal Society*, 2(4), pp. 281–293.

[20] EPI-WIN: WHO Information Network for Epidemics (2020), COVID-19—a global pandemic: What do we know about SARS-CoV-2 and COVID-19. World Health Organization (WHO).

[21] Boldog, P., Tekeli, T., Vizi, Z., Dénes, A., Bartha, F. A., and Röst, G. (2020), Risk Assessment of Novel Coronavirus COVID-19 Outbreaks Outside China, *Journal of Clinical Medicine*, 9(2), p. 571.

[22] Hellewell, J., Abbott, S., Gimma, A., Bosse, N., Jarvis, C., Russell, T., Munday, J., Kucharski, A., Edmunds, W., Funk, S., Eggo, R., Sun, F., Flasche, S., Quilty, B., Davies, N., Liu, Y., Clifford, S., Klepac, P., Jit, M., and Zandvoort, K. (2020), Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts, *The Lancet Global Health*, Vol. 8.

[23] Kucharski, A., Russell, T., Diamond, C., Liu, Y., Edmunds, J., Funk, S., Eggo, R., Sun, F., Jit, M., Munday, J., Davies, N., Gimma, A., Zandvoort, K., Gibbs, H., Hellewell, J., Jarvis, C., Clifford, S., Quilty, B., Bosse, N., and Flasche, S. (2020), Early dynamics of transmission and control of COVID-19: a mathematical modelling study, *The Lancet Infectious Diseases*, Vol. 20.

[24] Pearson, C., van Schalkwyk, C., Foss, A., O'Reilly, K., and Pulliam, J. (2020), Projected early spread of COVID-19 in Africa through 1 June 2020, *Eurosurveillance*, Vol. 25.

[25] Plank, M., Binny, R., Hendy, S., Lustig, A., James, A., and Steyn, N. (2020), A stochastic model for COVID-19 spread and the effects of Alert Level 4 in Aotearoa New Zealand.

[26] Government of Canada (2020), Download COVID Alert today (online), https://www.canada.ca/en/public-health/services/diseases/coronavirus-disease-covid-19/covid-alert.html (Access date: 14-12-2020).

[27] Plank, M., James, A., Lustig, A., Steyn, N., Binny, R., and Hendy, S. (2022), Potential reduction in transmission of COVID-19 by digital contact tracing systems: a modelling study, *Mathematical medicine and biology: a journal of the IMA*, 39, pp. 156–168.

[28] Endo, A., Leclerc, Q., Knight, G., Medley, G., Atkins, K., Funk, S., and Kucharski, A. (2020), Implication of backward contact tracing in the presence of overdispersed transmission in COVID-19 outbreaks, *Wellcome Open Research*, 5, p. 239.

[29] Levesque, J., Maybury, D. W., and Shaw, R. D. (2020), A model of COVID-19 propagation based on a gamma subordinated negative binomial branching process, *Journal of Theoretical Biology*.

[30] Zhao, S. (2020), Estimating the time interval between transmission generations when negative values occur in the serial interval data: Using COVID-19 as an example, *Mathematical Biosciences and Engineering*, 17, pp. 3512–3519.

[31] Wearing, H., Rohani, P., and Keeling, M. (2005), Appropriate models for the management of infectious diseases, *PLoS Med*, 2, pp. 813–813.

[32] Bertozzi, A., Franco, E., Mohler, G., Short, M., and Sledge, D. (2020), The challenges of modeling and forecasting the spread of COVID-19, *Proceedings of the National Academy of Sciences*, Vol. 117.

[33] James, A., Plank, M. J., Hendy, S., Binny, R., Lustig, A., Steyn, N., Nesdale, A., and Verrall, A. (2021), Successful contact tracing systems for COVID-19 rely on effective quarantine and isolation, *PLOS ONE*, 16(6), pp. 1–14.

[34] Cazzolato, F. (2020), The impact of COVID-19 prevalence, reproductive number, and testing on the Canadian Special Forces Command Headquarters and the possibility of an institutional outbreak, Defence Research and Development Canada, Scientific Letter, DRDC-RDDC-2020-L182.

[35] Guillouzic, S., Mirshak, R., and Sirjoosingh, A. (2020), Likelihood of undetected COVID-19 infection in a group: Effect of quarantining and testing (online), https://covid-app.cloud.forces.gc.ca/calculator (Access date: 24-08-2022).

[36] Levesque, J. and Maybury, D. W. (2020), A model COVID-19 propagation based on a gamma subordinated negative binomial process. Public Services and Procurement Canada, July 6, 2020.

[37] Lauer, S. A., Grantz, K. H., Bi, Q., et al. (2020), The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application, *Annals of Internal Medicine*, 172, pp. 577–582. https://doi.org/10.7326/M20-0504.

[38] Zhang, S., Diao, M., Yu, W., Pei, L., Lin, Z., and Chen, D. (2020), Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis, *International Journal of Infectious Diseases*, 93, pp. 201–204.

[39] Feaster, M. and Goh, Y.-Y. (2020), High Proportion of Asymptomatic SARS-CoV-2 Infections in 9 Long-Term Care Facilities, *Emerg Infect Dis*, 126, pp. 2416–2419. https://doi.org/10.3201/eid2610.202694.

[40] Moriarty, L., Plucinski, M., Marston, B., Kurbatova, E., Knust, B., Murray, E., Pesik, N., Rose, D., Fitter, D., Kobayashi, M., Toda, M., Canty, P., Scheuer, T., Halsey, E., Cohen, N., Stockman, L., Wadford, D., Medley, A., Green, G., and Richards, J. (2020), Public Health Responses to COVID-19 Outbreaks on Cruise Ships—Worldwide, February–March 2020, *Morbidity and Mortality Weekly Report*, 69, pp. 347–352.

[41] Zhang, K. (2020), Coronavirus: Hong Kong resident denies he is 'patient zero' of Diamond Princess cruise ship outbreak (online), South China Morning Post, https://www.scmp.com/news/hong-kong/society/article/3074698/coronavirus-hong-kong-resident-denies-he-patient-zero (Access date: 28-10-2020).

[42] World Health Organization (2020), Public Health Surveillance for COVID-19 (online), https://www.who.int/publications/i/item/who-2019-nCoV-surveillanceguidance-2020.7 (Access date: 28-10-2020).

[43] Tokuda, Y., Sakihama, T., Aoki, M., Taniguchi, K., Deshpande, G. A., Suzuki, S., Uda, S., and Kurokawa, K. (2020), COVID-19 outbreak on the Diamond Princess Cruise Ship in February 2020, *Journal of General and Family Medicine*, 21(4), pp. 95–97.

[44] Kucirka, L., Lauer, S., Laeyendecker, O., Boon, D., and Lessler, J. (2020), Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction–Based SARS-CoV-2 Tests by Time Since Exposure, *Annals of Internal Medicine*, 173(4), pp. 262–267.

Annex A Algorithm

In this annex, the simulation algorithm for the model is described. Figure 1 shows a straightforward representation of the outbreak tree as a function of the generation. There is no computational advantage to simulating the outbreak tree generation by generation when studying the progression of the outbreak as a function of time for a continuous-time branching process. Such a method requires the information for all infectious individual of a current generation g to be retained into memory until all infections have been produced, leading to an exponential use of memory as a function of g if the outbreak grows.

To minimize memory usage, the approach that was taken to implement the model described in this document is quite different, although equivalent from a modelling and simulation standpoint. Rather than simulating the outbreak one generation at a time, the chosen approach follows the lineages that descend from one generation to the other. That is, the outbreak is simulated from a single individual in a generation to his first generated child in the next generation until the last generation is reached, at which point the algorithm goes back to the closest previous generation where a parent's child has not been simulated yet, and it continues to proceed from there until all descendants for all primary infectious individuals have been simulated. This is captured in Figure A.1 for an outbreak comprised of three generations. Such an algorithm generates the outbreak in a very memory-efficient manner as it allows for a stack-based data structure for the simulation, where only minimal data are recorded for up to a single infectious individual of each generation. When the end of a lineage is reached, the algorithm can move back to a previous element of the stack until the desired node is reached (i.e., a parent for which at least one child has not been simulated) and the algorithm can then proceed again into a forward fashion. The data structure for an element in the stack consists of fields such as the individual's latent and communicable periods, the end time of this communicable period, the number of interaction events within this period, the index of the current event, the time of the current event, the number of attendees for the current event, the number of new infections for the current event and the index of the current new infection for the current event.

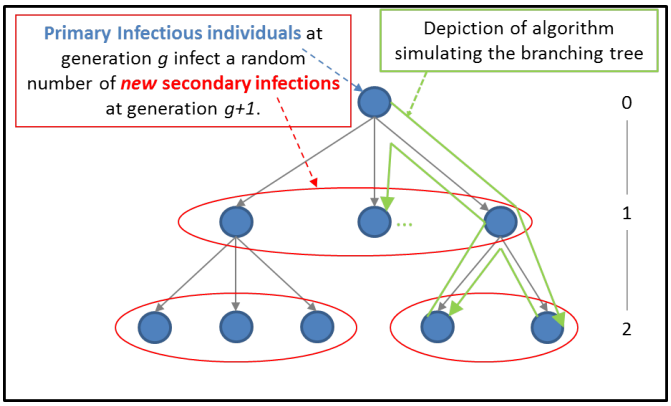

Figure A.1: *A simple branching process and the algorithm's simulation down the branching tree that starts with one infectious individual at generation $g = 0$. All infectious individuals at generation g will generate a random number of new secondary infectious at generation $g + 1$, but only a single individual per generation needs to be saved onto the data stack during simulation.*

Through the structure of the algorithm that uses a set of user-definable functions, the outbreak simulating mechanism is decoupled from the data to be recorded and from the computations to be made for analysis purpose. This also introduces flexibility in the data to be recorded, and allows for direct computation of the desired data, eliminating the need for very costly post-processing (in terms of memory and computation) of

each infection occurrence. For example, if one is interested in the computation of a simulated reproduction number, probability of extinction, or the timeline for a given set of metrics, it is possible to compute them directly using this set of user functions, resulting in memory usage proportional to the number of time periods used for the timeline instead of it growing exponentially with the number of generations. The design of this algorithm results in a collapse of the required memory to a negligible amount for most scenarios.

The following presents a generalized and simplified version of the simulation of a single outbreak path for the algorithm that was implemented. It supports both the contact tracing and vaccination features of the model. Considerable efforts have been made to optimize the algorithm, as implemented in C. The code of the modified model is about 1,000 times faster than the one of the original model, and it supports parallel processing, which allows to generate far more statistics in the same amount of time. The actual implementation uses specialized functions generated through preprocessing macros in order to minimize conditional statements. Some components can also be disabled at compilation time in order to achieve further performance gains. Simulation time strongly depends on the effective reproduction number and the chosen time cut for the outbreak. The implementation [13] simulates outbreak paths in a parallel-processing fashion, and it was used to generate billions of outbreak paths on a single workstation within minutes.

Algorithm to simulate a single outbreak path:

1. Call the user-defined function that initializes the data structures for the path to be generated (`path_init_proc_func`).
2. A primary infectious individual is generated. The vaccination status of this individual is randomly generated. The probability of vaccination for this primary individual is $\frac{f_{\mathrm{v}} r_{\mathrm{v,inf}}}{1-f_{\mathrm{v}}(1-r_{\mathrm{v,inf}})}$.
3. The type of communicable period for the infectious individual (self-isolating or not) is randomly generated, based on the applicable probability of self isolation (q or q_{v}).
4. The durations of the latent and communicable periods are randomly generated according to the applicable gamma distributions. The durations are recorded, as well as the end time of the communicable period, which is the sum of the latent and communicable periods for a primary individual if the time origin is the time of infection for this individual. For self-isolating individuals, the presymptomatic period is also set to the duration of the communicable period.
5. If the individual is self isolating and testing is to be performed, the test result is randomly generated based on the true positive rate of the test for self-isolating individuals ($m_{\mathrm{t.p.r.}}$).
6. If the time origin for the simulation differs from the time of infection for the primary individual, this time of infection, as well as the end time of the communicable periods, are updated according to the chosen time origin.
7. A user-defined initialization function (`pri_init_proc_func`) for the primary individual is called.
8. If contact tracing is enabled and the communicable period for the infectious individual is shorter than w_{ct}, pseudo interaction events must be generated during the part of the w_{ct} period that overlaps with the latent period. The number of events during this overlapping period is Poisson-distributed with the same rate λ as for events during the communicable period, but they must result in zero new infection, as the individual is not infectious yet. For each pseudo interaction event that is uniformly generated during this period, the number of event attendees is generated as usual and as defined

by the user, and the number of successfully traced contacts is generated based on the number of susceptible individuals (the number of attendees minus one) and the probability of contact tracing P_t. If the number of such events is non-zero, the user-defined initialization function for new infectious individuals (`new_inf_proc_func`) is called. Also, the user-defined function associated to a new event (`new_event_proc_func`) is called for each one of these events.

9. The number of regular interaction events is randomly generated based on the duration of the communicable period for the current infectious individual and the λ event rate.

10. If the number of events generated in the previous step is zero, look at the number of pseudo events generated in Step 8. If this number of events is also zero, call the user-defined function for an infectious individual that produces no interaction event (`new_inf_proc_func_noevent`). If the number of pseudo events is non-zero, call instead the user-defined function that has to be called once all transmission events for a given infectious individual have been generated (`end_inf_proc_func`). Go to Step 32.

11. Otherwise if the number of regular events is non-zero, look also at the number of pseudo events generated in Step 8. If this number of events is zero, call the user-defined initialization function for new infectious individuals (`new_inf_proc_func`).

12. Reset the event index to zero for the current infectious individual.

13. For the current event, generate the random event time uniformly within the communicable period of the current infectious individual.

14. Generate also the random number of attendees according to the user-selected distribution. Generate the random number of susceptible vaccinated attendees based on the number of attendees minus one and the probability of vaccination in the population f_v. Generate the random number of new infected vaccinated attendees based on the number of susceptible vaccinated attendees and their probability of infection $p_i r_{v,inf}$. This number is used to initialize the remaining number of new infected vaccinated attendees for the current event of the current infectious individual. Generate the random number of new infected non-vaccinated attendees based on the number of susceptible non-vaccinated attendees and their probability of infection p_i. This number is used to initialize the remaining number of new infected non-vaccinated attendees for the current event of the current infectious individual. Store the sum of these two numbers as well, that is the total number of new infected individuals. If the current infectious individual was determined to be self isolating and tested positive, and the current event time is within his contact tracing window having a duration w_{ct}, generate a random number of successfully traced non-infected contacts based on the number of non-infected susceptible attendees and the probability of contact tracing P_t. Also, generate a random number of successfully traced infected contacts based on the number of infected susceptible attendees and P_t.

15. If instead the current individual is either not self isolating, tested negative or the event time is not within his contact tracing window, set the numbers of traced infected and non-infected contacts to zero.

16. Call the user-defined function associated to a new event (`new_event_proc_func`).

17. If the function returned `false`, the susceptible individuals that get infected during the events should not initiate transmissions themselves within the simulation (e.g., because of simulation capping). If this is the case, verify if the current event index corresponds to the last event for the current

infectious individual. If it is, call the user-defined function that has to be called once all transmission events for a given infectious individual have been generated (`end_inf_proc_func`) and go to Step 32. If instead there are still some events left, increment the event index by one and go back to Step 13.

18. If the function returned `true` instead, set the new infection index for the current infectious individual to zero.

19. Move to the next generation of infection: the infected child identified through the new infection index for the current infectious individual becomes the new current infectious individual.

20. The vaccination status of the current infectious individual is randomly generated based on the fraction of vaccinated individuals amongst the remaining newly infected individuals by the infectious individual's parent. The applicable number of remaining newly infected vaccinated or non-vaccinated individuals by the infectious individual's parent is updated based on the outcome of the draw. The probability of self isolation and the probability of infection of the current infectious individual are set to q_{v} and $p_{\mathrm{i}} r_{\mathrm{v,inf}}$, or q and p_{i}, respectively, depending if this individual is drawn to be vaccinated or not.

21. The latent and communicable periods for the current infectious individual are generated, with the distribution for the latter dependent on the outcome of the previous step. This is done similarly to how it was done in Step 4 for a primary infectious individual, with two major differences: the end time of the communicable period is the sum of the durations of the latent and communicable periods added to the time of the event, as recorded by the parent, when the individual was infected, and the communicable period for a non-primary infectious individual can be interrupted through contact tracing. Regarding the latter, this can only occur if the parent tested positive and if the time of the event when the infection of the current infectious individual occurred is within the duration w_{ct} prior to the end of the parent's communicable period. If it is the case, a random number of successfully traced infected contacts was stored in Step 14. The new infection index that was set by the parent in Step 19 is compared to this random number. If it is smaller, that is, if the current infectious individual is successfully traced, and if his communicable period is randomly determined to be interrupted (with probability $P_{\mathrm{i_m}}/P_{\mathrm{t}}$ or $P_{\mathrm{i_t}}/P_{\mathrm{t}}$), the communicable period of this individual is truncated based on the contacting time as drawn from the contact tracing gamma time distribution (with corresponding average duration $\bar{m}^{\mathrm{i}}$ or $\bar{t}^{\mathrm{i}}$) that started once the parent's positive test result was received. If such truncation occurs, but the duration of the adjusted communicable period becomes negative (the individual is contacted before becoming infectious), the latent period of the individual is truncated accordingly and the duration of the communicable period is set to zero. If interruption occurred, the end time of the communicable period is updated, and test results are generated based on the applicable true positive rate ($m_{\mathrm{t.p.r.}}$ or $t_{\mathrm{t.p.r.}}$).

22. If required, pseudo events are generated for the current infectious individual, as it was done in Step 8.

23. The number of regular interaction events is randomly generated based on the duration of the communicable period for the current infectious individual and the λ event rate.

24. If there is at least one regular event, reset the event index to zero for the current infectious individual, and look at the number of pseudo events. If this number of events is zero, call the user-defined initialization function for new infectious individuals (`new_inf_proc_func`).

25. Otherwise if the number of regular events is zero, look at the number of pseudo events. If this number of events is also zero, call the user-defined function for an infectious individual that produces no

interaction event (`new_inf_proc_func_noevent`). If the number of pseudo events is non-zero, call instead the user-defined function that has to be called once all transmission events for a given infectious individual have been generated (`end_inf_proc_func`). Continue to Step 29.

26. Generate the current event the same way as it was done in Steps 13 to 16.

27. If the function `new_event_proc_func` returned `true`, set the new infection index for the current infectious individual to zero. Go to Step 19.

28. Otherwise if the function returned `false`, verify if the current event index corresponds to the last event for the current infectious individual. If it is not, increment the event index by one and go back to Step 26 to generate the next event. Otherwise if the last event index was reached, call the user-defined function that has to be called once all transmission events for a given infectious individual have been generated (`end_inf_proc_func`).

29. All events for the current infectious individual have been generated. If the current infectious individual is a primary infectious individual, go to Step 32.

30. Move back to the previous generation (the parent of the current infectious individual becomes again the current infectious individual).

31. Verify if all new infections for the current event have been generated. If it is the case, verify if all events have been generated for the current infectious individual. If all events have been generated, call the user-defined function that has to be called once all transmission events for a given infectious individual have been generated (`end_inf_proc_func`), and go back to Step 29. Otherwise, if there are still some events left, increment the event index by one and go back to Step 26. If instead there are still some new infections to be generated for the current event, increment the new infection index by one and go back to Step 19.

32. Verify if it was the last primary infectious individual for the current outbreak path. If it was not, generate the next primary infectious individual by going back to Step 2.

33. The current outbreak path has been generated. Call the user-defined function that terminates the processing of the data for the path and that determines if the path should be included or not (`path_end_proc_func`). If the path is rejected, the whole path must be regenerated, by resetting the index for the primary infectious individual and by going back to Step 2.

Annex B When vaccination has no impact on self isolation

Equation (48) depends only on f_{v} and on the product of $r_{\mathrm{v,inf}}$ and $r_{\mathrm{v,tr}}$. The dependency of the reduction factor $R_{\mathrm{eff,v}}/R_{\mathrm{eff}}$ on f_{v} and $r_{\mathrm{v,inf}}r_{\mathrm{v,tr}}$ is shown in Figure B.1. For an R_{eff} of 2.5, below the dotted line are all possible combinations of f_{v} and $r_{\mathrm{v,inf}}r_{\mathrm{v,tr}}$ for $R_{\mathrm{eff,v}} < 1$ (which ultimately guarantees extinction).

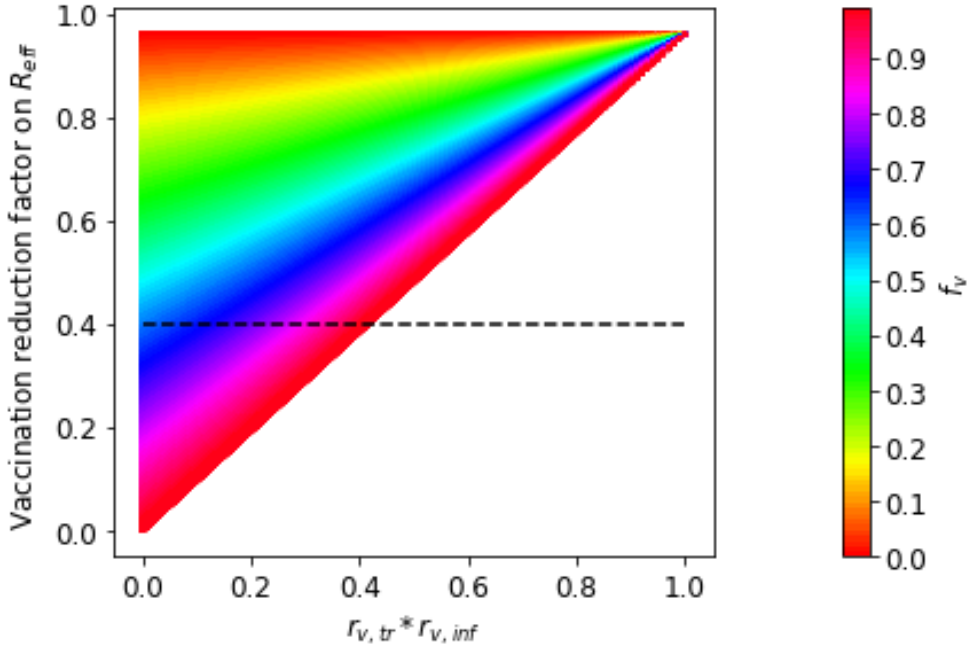

Figure B.1: *Vaccination reduction factor $R_{\mathrm{eff,v}}/R_{\mathrm{eff}}$ for $r_{\mathrm{v,inf}}r_{\mathrm{v,tr}}$ and f_{v}. For an R_{eff} of 2.5, below the dotted line are all possible combinations of f_{v} and $r_{\mathrm{v,inf}}r_{\mathrm{v,tr}}$ that would produce an $R_{\mathrm{eff,v}} < 1$ (which ultimately guarantees extinction).*

List of symbols

R_0	The basic reproduction number is the average number of secondary infections caused by a single infectious individual in a population where all individuals are susceptible.
R_{eff}	The effective reproduction number is the average number of secondary infections caused by a single infectious individual in a population where some mitigation measures are in place. For the branching process, this is the effective reproduction number when a fraction of the infectious individuals self isolate (after the appearance of symptoms, for example), but no vaccination is involved.
$R_{\text{eff,v}}$	The effective reproduction number with vaccination is the average number of secondary infections by a single infectious individual that may be vaccinated in a partially vaccinated population where some mitigation measures are in place.
λ	Rate of the interaction events where at least one susceptible individual is present and where transmission and new infections may occur.
λ_e	Rate of all interaction events including those where only the infectious individual is present.
$\bar{t}$	Average duration of the infectious period of non-self-isolating individuals.
$\bar{m}$	Average duration of the infectious period of self-isolating individuals at symptoms onset.
$\bar{l}$	Average duration of the disease latent period.
t_{95}	The 95th percentile of the infectious period of non-self-isolating individuals.
m_{95}	The 95th percentile of the infectious period of self-isolating individuals at symptoms onset.
l_{95}	The 95th percentile of the disease latent period.
κ_t	The gamma shape parameter for the infectious period of non-self-isolating individuals.
κ_m	The gamma shape parameter for the infectious period of self-isolating individuals at symptoms onset.
κ_l	The gamma shape parameter for the disease latent period.
g_{ave}	The average number of individuals at events, including the infectious individual.
μ	Parameter for the mean of an unbounded logarithmic distribution ($\mu \geq 1$) or of an unbounded Gaussian distribution used to draw number of individuals for one event.
p	The logarithmic shape parameter.
σ	Parameter for the standard deviation of an unbounded Gaussian used to draw the number of individuals for one event.
σ_{r}	Parameter for the standard deviation of an unbounded Gaussian used to draw the number of individuals for one event, relative to the μ parameter.
p_{i}	Probability of infection of a given non-vaccinated susceptible individual by the non-vaccinated infectious individual at an event.
f_{v}	Fraction of the population vaccinated.

$r_{\mathrm{v,inf}}$ Scaling factor, between 0 and 1, for the infection probability of a given vaccinated susceptible individual by a given infectious individual during one event. This parameter captures the vaccine's efficacy against (symptomatic and asymptomatic) infection ($1 - r_{\mathrm{v,inf}}$).

$r_{\mathrm{v,tr}}$ Scaling factor, between 0 and 1, for the infection probability of a given susceptible individual by a given vaccinated infectious individual during one event. This parameter captures the vaccine's efficacy against onward transmission ($1 - r_{\mathrm{v,tr}}$).

q_{v} Fraction of infectious vaccinated individuals that self isolate.

E_{i} The efficacy against infection for individuals that are vaccinated. It is equivalent to $1 - r_{\mathrm{v,inf}}$.

E_{t} The efficacy against onward transmission for individuals that are vaccinated. It is equivalent to $1 - r_{\mathrm{v,tr}}$.

$t_{\mathrm{t.p.r.}}$ The true positive test rate of non-self-isolating individuals.

$m_{\mathrm{t.p.r.}}$ The true positive test rate of self-isolating individuals.

$T_{\Delta t}$ The time delay between performing the test and reporting results.

w_{ct} The contact tracing window which is the period prior to individual isolation during which contacts are considered.

P_{t} The probability of successful contact tracing.

$P_{\mathrm{i_t}}$ The probability of successful contact tracing and of self isolation once contacted of infected individuals that would otherwise not self isolate.

$P_{\mathrm{i_m}}$ The probability of successful contact tracing and of self isolation once contacted of infected individuals that would otherwise self isolate.

$\bar{t}^{\mathrm{i}}$ The average delay in contact tracing and isolating individuals that, if not contacted, would otherwise not self isolate.

$\bar{m}^{\mathrm{i}}$ The average delay in contact tracing and isolating individuals that, if not contacted, would otherwise self isolate.

t^{i}_{95} The 95th percentile of the delay in contact tracing and isolating individuals that, if not contacted, would otherwise not self isolate.

m^{i}_{95} The 95th percentile of the delay in contact tracing and isolating individuals that, if not contacted, would otherwise self isolate.

κ^{i}_{t} The gamma shape parameter of the delay in contact tracing and isolating individuals that, if not contacted, would otherwise not self isolate.

κ^{i}_{m} The gamma shape parameter of the delay in contact tracing and isolating individuals that, if not contacted, would otherwise self isolate.

n_{paths} The number of generated outbreak simulation paths.

t_{max} Maximum simulation time, in days, at which new infectious individuals can be created in the simulation.

List of abbreviations/acronyms/initialisms

ABM	agent-based model
CAF	Canadian Armed Forces
CF H Svcs Gp	Canadian Forces Health Services Group
CMJ	Crump-Mode-Jagers
CORA	Centre for Operational Research and Analysis
COVID-19	coronavirus disease 2019
DND	Department of National Defence
DRDC	Defence Research and Development Canada
IBP	interaction branching process
MERS	Middle East respiratory syndrome
NORAD	North American Aerospace Defence Command
PHAC	Public Health Agency of Canada
PDF	probability density function
PMF	probability mass function
RCN	Royal Canadian Navy
SARS	severe acute respiratory syndrome
SARS-Cov-2	severe acute respiratory syndrome coronavirus 2
SEIR	susceptible-exposed-infectious-recovered
SER	susceptible-infectious-recovered
TBP	transmission branching process
VBP	vaccination branching process
WHO	World Health Organization

Glossary

Asymptomatic individual:

An infected individual that shows no symptoms or that shows very mild symptoms that would not result in them self isolating.

Basic reproduction number (R_0):

The average number of secondary infections (known or unknown) caused by a single infectious individual (known or unknown) in a population where all, but one individual, are susceptible, and where no mitigation measures, including self isolation, are in place.

Disease communicable period:

Period of a disease during which an individual is infectious and not isolated.

Effective reproduction number (R_{eff}):

The average number of secondary infections (known or unknown) caused by a single infectious individual (known or unknown) in a population where some mitigation measures are in place.

Incubation period:

Begins when an individual is exposed and ends when symptoms begins.

Infected individual:

Someone that has contracted the SARS-Cov-2 virus.

Infectious individual:

Someone that has contracted the SARS-Cov-2 virus and that can infect others.

Infectious non-spreading individual:

Someone that has contracted the SARS-Cov-2 virus, but does not infect anyone.

Infectious phase:

Begins at the end of the latent phase, and ends when the individual is no longer infectious or self isolates.

Latent phase:

Begins when an individual is exposed, and ends when the individual is infectious.

Non-self-isolating individuals:

Infected individuals that do not self isolate either because they are asymptomatic, mildly symptomatic, or ignore symptoms.

Observable path:

An outbreak is observable if an individual tests positive to COVID-19, and the outbreak is non-extinct or not yet extinct.

Observable reproduction number:

The average number of secondary infections that tested positive caused by a single infectious individual that tested positive in a population where some mitigation measures are in place.

Outbreak:

Begins as soon as a single individual is infected. The outbreak path can become extinct, if the number of infectious individuals reaches zero, otherwise it is defined as a growing outbreak because the number of individuals infected increases perpetually.

Pre-symptomatic phase:

Begins when a symptomatic individual is exposed, and ends when a symptomatic individual becomes symptomatic.

Probability of extinction:

The probability that an outbreak will reach extinction.

Probability of infection:

The probability that a susceptible individual, present at an infectious event, gets infected.

Self isolating individual:

Infected individuals that self isolate when symptoms occur.

Symptomatic individual:

An infected individual that shows symptoms.

DOCUMENT CONTROL DATA

*Security markings for the title, abstract and keywords must be entered when the document is sensitive.

1. ORIGINATOR (The name and address of the organization preparing the document. A DRDC Centre sponsoring a contractor's report, or a tasking agency, is entered in Section 8.)

DRDC – Ottawa Research Centre
3701 Carling Avenue,
Ottawa ON
K1A 0Z4, Canada

2a. SECURITY MARKING (Overall security marking of the document, including supplemental markings if applicable.)

CAN UNCLASSIFIED

2b. CONTROLLED GOODS

NON-CONTROLLED GOODS
DMC A

3. TITLE (The document title and subtitle as indicated on the title page.)

A branching process and simulation model to evaluate the spread of severe acute respiratory syndrome coronavirus 2 (SARS-Cov-2) in various environments: Enabling simultaneous mitigation strategies including social distancing, masks, symptomatic self isolation, testing, contact tracing and vaccination

4. AUTHORS (Last name, followed by initials – ranks, titles, etc. not to be used. Use semi-colon as delimiter.)

Biron, K.; Drouin, P.-L.; Serré, L.

5. DATE OF PUBLICATION (Month and year of publication of document.)	6a. NO. OF PAGES (Total pages, including Annexes, excluding DCD, covering and verso pages.)	6b. NO. OF REFS (Total cited in document.)
March 2023	70	44

7. DOCUMENT CATEGORY (e.g., Scientific Report, Contract Report, Scientific Letter)

Scientific Report

8. SPONSORING CENTRE (The name and address of the department project or laboratory sponsoring the research and development.)

DRDC – Ottawa Research Centre
3701 Carling Avenue,
Ottawa ON
K1A 0Z4, Canada

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.)	9b. CONTRACT NO. (If appropriate, the applicable contract number under which the document was written.)
PEOPLE_010	

10a. DRDC PUBLICATION NUMBER	10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned to this document either by the originator or by the sponsor.)
DRDC-RDDC-2023-R025	

11a. FUTURE DISTRIBUTION WITHIN CANADA (Approval for further dissemination of the document. Security classification must also be considered.)

Public release

11b. FUTURE DISTRIBUTION OUTSIDE CANADA (Approval for further dissemination of the document. Security classification must also be considered.)

Public release

12. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Use semi-colon as a delimiter.)

COVID-19; SARS-Cov-2; Branching Process; Outbreak Propagation; Mitigation Strategies; Vaccination

13a. ABSTRACT (When available in the document, the English version of the abstract must be included here.)

Starting from existing severe acute respiratory syndrome coronavirus 2 (SARS-Cov-2) propagation branching processes, the work presented in this Scientific Report expands on these earlier models through the concept of interaction events, the introduction of additional classes of infected individuals, and the use of a latent phase for the disease. These changes enable the modelling of mitigation measures, including vaccination, and to evaluate their effects on the propagation of SARS-Cov-2 through the addition of a probability of infection for interaction events. Other features were also added to the model to simulate scenarios more realistically, including some that make the model deviate from a branching process, such as the modelling of contact tracing and isolation.

In this Report, a generalized branching process for the model, providing analytical results, is first formally derived, and modifications going beyond a branching process are introduced. A presentation of the input and output parameters for the model follows. The algorithm implementing the model is also provided. Cases studies (impact of the different parameters from the interaction model, the Diamond Princess cruise ship outbreak using the modified branching process and the impact of the parameters from the vaccination model) are then presented for three different applications. The results show that this model can be used in the study of early outbreaks of the coronavirus disease 2019 (COVID-19), and in the study of mitigation strategies against the spread of its virus.

13b. RÉSUMÉ (When available in the document, the French version of the abstract must be included here.)

Basé sur des processus de ramification existants pour la propagation du SARS-Cov-2 (*severe acute respiratory syndrome coronavirus 2*), le travail présenté dans ce Rapport scientifique généralise ces précédents modèles à travers le concept d'événements d'interactions, l'introduction de classes additionelles d'individus infectés, ainsi que l'utilisation d'une phase latente pour la maladie. Ces changements permettent la modélisation de mesures de mitigation, incluant la vaccination, et d'évaluer leurs effets sur la propagation de SARS-Cov-2 à travers l'ajout d'une probabilité d'infection pour les événements d'interactions. D'autres éléments ont aussi été ajoutés au modèle afin de simuler des scénarios de manière plus réaliste, incluant certains qui font dévier le modèle d'un processus de ramification, tel que la modélisation de la recherche des contacts et l'isolation.

Dans ce rapport, un processus de ramification généralisé pour le modèle, fournissant des résultats analytiques, est d'abord formellement dérivé, et des modifications qui vont au-delà d'un processus de ramification sont introduites. Une présentation des paramètres d'entrée et de sortie pour le modèle suit. L'algorithme mettant le modèle en œuvre est aussi fourni. Des études de cas (impact des différents paramètres du modèle d'interaction, la flambée d'infections sur le navire de croisière Diamond Princess en utilisant le processus de ramification modifié et l'impact des paramètres du modèle pour la vaccination) sont ensuite effectuées pour différentes applications. Les résultats montrent que ce modèle peut être utilisé pour l'étude de flambées initiales de la maladie à coronavirus 2019 (COVID-19), ainsi que pour l'étude de mesures de mitigation contre la propagation du virus.