Unifying the effective reproduction number, incidence, and prevalence under a stochastic age-dependent branching process

Tresnia Berah[*,2], Thomas A. Mellan[*,1], Xenia Miscouridou[*,2], Swapnil Mishra[*,1], Kris V Parag[1], Mikko S. Pakkanen[2,†] and Samir Bhatt[1,†]

[1] *Department of Infectious Disease Epidemiology, Imperial College London*

[2] *Department of Mathematics, Imperial College London*

[*] *Joint Authorship*

[†] *Corresponding Author: s.bhatt@imperial.ac.uk, m.pakkanen@imperial.ac.uk*

Abstract

Renewal equations are a popular approach used in modelling the number of new infections (incidence) in an outbreak. A unified set of renewal equations where incidence, prevalence and cumulative incidence can all be recovered from the same stochastic process has not been attempted. Here, we derive a set of renewal equations from an age-dependent branching process with a time-varying reproduction number. Our new derivation utilises a measure-theoretic approach and yields a fully self-contained mathematical exposition. We find that the renewal equations commonly used in epidemiology for modelling incidence are an equivalent special case of our equations. We show that these our equations are internally consistent in the sense that they can be separately linked under the common back calculation approach between prevalence and incidence. We introduce a computationally efficient discretisation scheme to solve these renewal equations, and this algorithm is highly parallelisable as it relies on row sums and elementwise multiplication. Finally we present a simple simulation example in the probabilistic programming language Stan where we jointly fit incidence and prevalence under a single time-varying reproduction number and generation interval.

1 Introduction

Mathematical descriptions of infectious disease outbreaks are fundamental to forecasting and simulating the dynamics of epidemics, as well as to understanding the mechanics of how transmission occurs. The quantities of interest are incidence (the number of new infections at a given time point), cumulative incidence (the total number of infections up to a given time point) and prevalence (the number of infected/infectious individuals at a given time point). Taking a somewhat reductive perspective, it can be said that two main popular frameworks co-exist when modelling an infectious disease outbreak, individual-based simulations on one hand, and a single set of governing equations on the other hand. Individual-based simulations are not only simple to understand in terms of their fundamental assumptions but have also proven extremely impactful [14]. However, mathematical tractability is absent, rates of convergence to an average are unknown, inference can be challenging, and sensitivity analysis can be ambiguous. In contrast, governing equations tend to have a stronger physical interpretation, are easier to perform inference over, and have provable convergence properties.

Code available at `https://github.com/ImperialCollegeLondon/BellmanHarris_simulation_and_inference`

The most widely known set of governing equations was presented in the seminal work of Kermack and McKendrick [21], where they studied the number and distribution of infections of a transmissible disease as it progresses through a population over time. They constructed classes, called compartments, and modelled the propagation of infectious disease via interactions among these compartments. The result is the popular susceptible-infected-recovered (SIR) model, variants of which are widely used in epidemiology. SIR models provide an intuitive mechanism for understanding disease transmission, and in the original derivation of [21], they were noted to be similar to the Volterra equation [27]. The Volterra equation (of the second kind) or more commonly, the renewal equation, is another popular governing equation [16, 11, 26, 8]. A large body of work in infectious disease epidemiology is based around the renewal equation and many modifications exist [10, 1, 17, 29]. A renewal process generalises a Poisson process to allow for arbitrary (instead of exponential) holding or event waiting times. From an infectious disease perspective, these holding times model how new infections are generated across an epidemic. There is a connection between compartmental models and renewal equations [9, 28], but this link has only been established under simplified conditions such as constant transmission rates (the basic reproduction number, R_0). Furthermore, within renewal frameworks the link and distinction between prevalence (number of infected people at any given time) and incidence (number of newly infected individuals arising at a given time) is often unclear. Other derivations of the renewal equation in epidemiology [16] also ignore the relationships between prevalence and incidence and rely primarily on heuristic arguments to derive a renewal equation where transmission rates are not constant (effective or time-varying reproduction number $R(t)$).

In 1948, Bellman and Harris [3] elegantly captured an underlying infection mechanism by formulating an age-dependent branching proces. Their framework connected the two worlds of individual based modelling and governing equations by first constructing a stochastic process where infected individuals behave by simple rules, and then deriving a governing equation for the average behaviour. In general, branching processes describe how particles stochastically propagate over time. In epidemiology, age-insensitive branching processes, such as the fundamental Galton-Watson process, which discretise the propagation process into generations, have provided tractable yet intuitive ways of modelling the spread of an infectious disease [2, 18]. However, these approaches, while useful, lack the realism of the more general Bellman-Harris processes [3, 4]. Bellman-Harris processes generalise the Galton-Watson framework of fixed inter-generational time and introduce a more flexible stochastic process, where the time between successive generations follows a non-negative random variable from an arbitrary distribution. The age-dependence assumption of Bellman-Harris

allows for the variable time between exposure to a pathogen and subsequent transmission to be properly modelled and provides a framework encoding useful information on the biology of the infecting pathogen, such as incubation periods and non-monotonic infectiousness.

In overview of the original Bellman–Harris process [3, 4], it is assumed that *particles* live, independently of each other, for random periods of time, and transform into a random number of new particles at the end of their life time (after which they disappear). In what follows we will replace a given *particle* with an *infected individual*. To understand the Bellman-Harris process we consider a homogeneous or well-mixed population, in which members can randomly infect one another. Let $t \in \mathbb{R}^+$, be a positive real representing calendar time. We define the generation time (also commonly and interchangeably referred to as generation distribution) as the random variable representing the period between the time when an individual is first infected and the time when this individual infects others. We introduce $G(\cdot)$ as the cumulative distribution function for the generation time. The function G therefore satisfies $G(0^+) = 0$, $G(\infty) = 1$ and it can also be written as $G(\cdot) = \int_0^{\cdot} G(du)$, where G is the associated probability measure. If G admits a density function g, then we can express it as $G(t) = \int_0^t g(u)du$. The generation time is often parameterised as a probability distribution, but it should be noted that the only constraints on its density $g(\cdot)$ are that it be non-negative and $\int_0^\infty g(u)du = 1$.

Now after some random period $L \sim g(\cdot)$, an infected individual can infect $n \in \mathbb{N}^+$ (positive integers) other individuals with probability q_t. We are interested in characterising the total number of infected individuals at time t, which we denote for now $Z(t)$. However, as each time trajectory of $Z(t)$ is one possible reality (or sample-path) from the epidemic process, we more broadly aim to calculate and subsequently model the average number of infections $\mathbb{E}[Z(t)]$ present at time t in the population. Direct calculation of $\mathbb{E}[Z(t)]$ involves the manipulation of generally intractable point probabilities. Consequently, we adopt an approach based on generating functions, which replaces the need to manipulate intractable integral equations by the simpler requirements of evaluating partial derivatives at a certain point. Figure 1 shows simulations for a age dependent branching process, and the correspondence of the closed form formulas we derive in this paper alongside the mean estimate across a limited number of simulations. The central quantities we are interested in is incidence, cumulative incidence and prevalence.

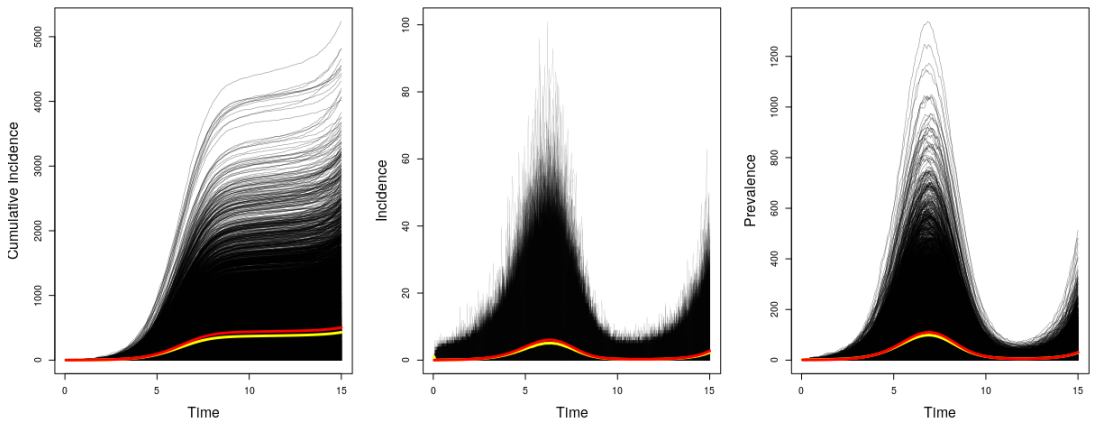

Figure 1: Simulations of an age dependent branching process computing summaries of cumulative incidence, incidence, and prevalence. Black lines are an individual simulation, yellow line is the expected value across these simulations, and red is the theoretical expected value derived in this paper. In this simulation $R(t) = 1.3 + \sin(0.5\,t)$ and $g \sim \text{Gamma}(3, 3)$.

The original formulation [4] by Bellman and Harris of the renewal equation for prevalence, and subsequent work by Harris [20], along with epidemic perspectives [5] all focused on the simple case of a constant reproduction number R_0. The form of this renewal equation when only considering R_0 is exactly what is commonly used in epidemic modelling where the incidence of infections $I(t)$ follows the integral equation given by: $I(t) = R_0 \int_0^\infty I(t-u)g(u)du$. It could be tempting to think that by introducing a time-varying reproduction number $R(t)$, it would simply suffice to switch to $R(t)$ from R_0 in the renewal equation; indeed this is near ubiquitous practice in epidemiology [15, 13, 11, 16]. While justifications based on heuristic arguments such as Lotka's [24] (used in tracking the numbers of females in an age-structured population) or as per [16] are valid within their respective contexts, these arguments lose their validity and fail when considering a stochastic age-dependent branching process with a time-varying reproduction process [23, 22]. Not only does the functional form become incorrect, but confusion then also arises as to whether the measure under consideration is prevalence or incidence. In all of Bellman and Harris' original work [3, 4, 20], prevalence has been the quantity of interest, but other perspectives re-deriving the same equations as in [5], as well as the vast majority of epidemiological applications, considered in fact incidence.

In this paper, we first re-derive two previous results on branching processes due to Kimmel [23], who adopted a point process approach, by following here a measure-theoretic approach more akin

to the classical work of Harris [20] and giving a fully self-contained mathematical exposition. These two results consist of renewal equations respectively for prevalence and cumulative incidence in an age-dependent branching process with a time-varying reproduction number. As a novel extension, we then derive a new renewal equation descri,bing incidence. We explain the technical details of these results in Appendix D. To outline our argument, we first show how to rigorously construct an age-dependent branching process with time-varying reproduction number, we then highlight how Harris's principle of first-generation enables recursive definitions, and consequently we derive integral equations governing generating functions. Next we derive explicit integral equations for prevalence, cumulative incidence and incidence. We demonstrate numerically that the newly derived equation for incidence agrees up to high precision to the commonly used renewal equation for incidence (Eq. (32)), leading us to conjecture that these two equations are equivalent. We also show that the newly derived renewal equations for prevalence and incidence satisfy and are consistent with the well known back calculation relationship [12] that prevalence is equal to incidence convolved with the survival function $\big(1 - G(t)\big)$. We also present a simple transformation to enable efficient computation of the newly derived equations using straightforward matrix operations. Finally, we include a simulation study fitting three likelihoods for prevalence, cumulative incidence and incidence and show that key parameters such as the generation interval can be identifiable.

2 Methods and Theory

2.1 Constructing an age-dependent branching process

We show how to construct a branching process in the context of time-varying offspring distribution. Following Harris [20], our proof involves measure-theoretic arguments, and we therefore provide an explicit description of the probability spaces over which the random quantities related to the branching process will live. For a point process derivation we refer the reader to Kimmel [23]. It should be noted however, that while our model shares many assumptions with [20], the main departing assumption of time-varying reproduction number leads to a very different, and novel path to constructing the probabilistic framework of the branching process.

In its most simplified form, an age-dependent branching process can be thought of as a forward simulation starting with one infected individual at time $t = 0$. No immigration in or out of this system is assumed. At some random time L_0, the initial infected individual transforms/infects

ν_0 new, but identical, infected individuals, all of age 0 at birth. *It is critical to note that when an infected individual transforms/infects/is replaced by newly infected individuals, the original individual ceases to be counted.* The cycle then repeats for each of these newly infected individuals who are fully determined by their random life-length L and their offspring distribution process $\nu(t)$. Intuitively this infection process can be visualised as a branching tree (see Figure 2) or by a vector of times of births. Both the life-lengths L and the offspring distribution $\nu(t)$ are independent and identically distributed respectively with underlying density distributions $g(\cdot)$ and probability mass function $q_t(\cdot)$.

Definitions and proofs of all the following constructions can be found in the Appendix where results are presented with more formal mathematical rigour.

The construction of the branching process with time-varying reproduction number starts at the level of the infected individual for which we create a labelling scheme. The initial infected individual is denoted $\langle 0 \rangle$. Next, the first, second or i-th offspring of the initial infected individual are denoted respectively $\langle 1 \rangle, \langle 2 \rangle, \ldots, \langle i \rangle$. Then again, the j-th offspring of the i-th offspring is denoted $\langle i, j \rangle$. More generally we define an infected individual as an ordered n-tuple (ordered list). This labelling scheme can be seen in Figure 2, where each new branch in the tree is appended from its parent. The space of individuals $\mathcal{I}$ is then defined as the infinite collection of all individuals labelled as follows:

$$\mathcal{I} \coloneqq \{\langle 0 \rangle\} \cup \{\langle i_1, \ldots, i_n \rangle \ : i_1, \ldots, i_n \in \mathbb{N}, n \in \mathbb{N}\}.$$

This space contains all potentially realisable n-tuples corresponding to infected individuals in an age-dependent branching process and $\mathcal{I}$ is by construction countably infinite. As noted earlier, each infected individual is associated to two random variables, a life-length L, such that $\mathbb{P}[0 \leq L \leq l] \coloneqq G(l) = \int_0^l g(u)du$ and the number of offspring $\nu(t)$ it produces at transformation time, such that $\mathbb{P}(\{\nu(t) = r\}) \coloneqq q_t(r)$. In our context, the life-length, more commonly known in epidemiology as the generation distribution or generation time, is not assumed to be time-varying, and common examples in epidemiology are the Rayleigh distribution and Gamma distribution where there is an incubation period to peak infectiousness. However, here the offspring distribution, often known in epidemiology as the secondary distribution or in graph theory as the degree distribution, is assumed to vary in time. The expectation of the offspring distribution $\mathbb{E}[\nu(t)]$ is known in epidemiology as the effective reproduction number, $R(t)$. When simulating an age-dependent branching process, common choices for the offspring distribution are the Poisson and Negative Binomial distributions.

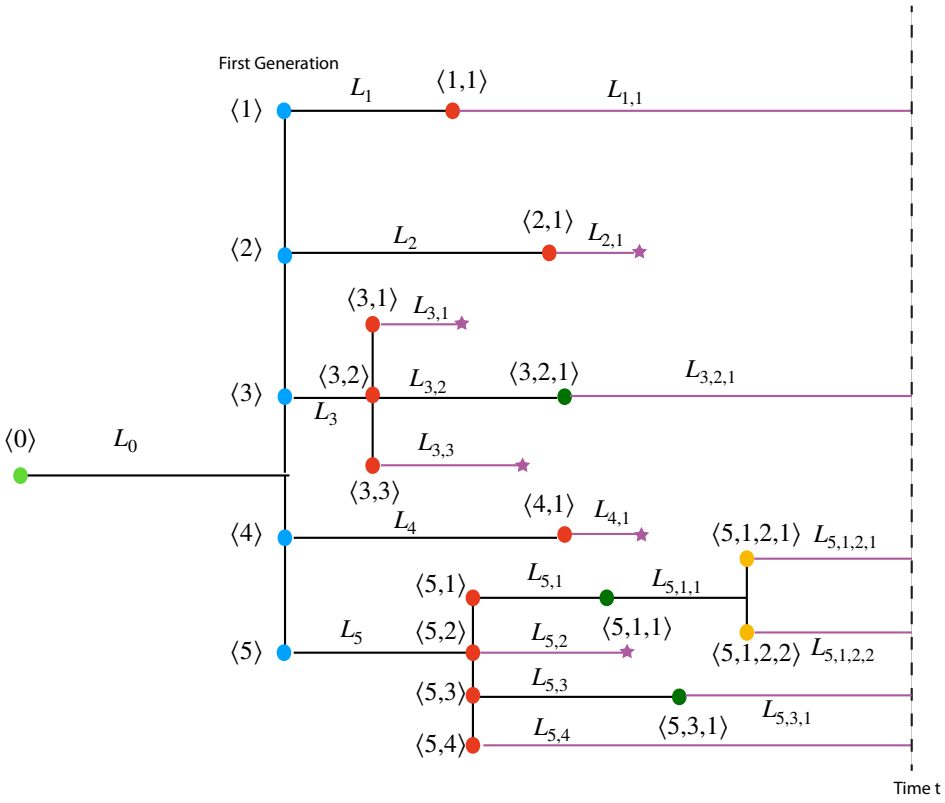

Figure 2: A realisation of a family history X. The labelling scheme shows how the naming of a new offspring simply consists in appending a new *index* to the parent's tuple. For instance, $\langle i, j, k\rangle$ are individuals belonging to the third generation and $L_{i,j,k}$ are their associated life-lengths. The dotted line represents calendar time t. Some infected individuals are still alive at time t : they correspond to the purple lines that meet the dotted line. Stars represent infected individuals who transform up before time t and infected no one else (zero offspring).

Now that we have presented a labelling scheme for infected individuals, and described their associated two fundamental properties, we can introduce family histories. A family history, denoted here by X, consists of an infinite sequence (of pairs) of random variables tracing out the life-lengths and the offspring numbers for every infected individual potentially ever alive in the process. Thus,

$$X \coloneqq \left\{ L_0, \nu_0(L_0), L_{i_1,\ldots,i_k}, \nu_{i_1,\ldots,i_k}\left(L_0 + \sum_{j=1}^{k} L_{i_1,\ldots,i_j}\right) : i_1,\ldots,i_k \in \mathbb{N}, k \in \mathbb{N} \right\}, \tag{1}$$

where, for an infected individual $\langle \iota \rangle = \langle i_1, \ldots, i_k \rangle$, the random variable L_ι represents its life-length, and $\nu_\iota(L_0 + \sum_{j=1}^{k} L_{i_1,\ldots,i_j})$ represents its offspring number *observed at time of death*. A closer look at the definition in Eq. (1) shows that a family history X consists on one hand of the random life-length L_0 and the random offspring number ν_0 of the initial individual, and then subsequently X contains all of the random life-lengths and offspring numbers for all other potential infected individuals. We note here that the only time points that appear/are relevant in the family history are the random times of deaths, namely $L_0 + \sum_{j=1}^{k} L_{i_1,\ldots,i_j}$ for an individual labelled $\langle i_1, \ldots, i_k \rangle$.

Each family history can be thought of as a tree or a realisation of an age-dependent branching process. The space of all possible realisations X is denoted by Ω. An example of a realisation X is shown in Figure 2. For a given realisation X, certain infected individuals will belong to the realised tree, while some others will not. For instance the initial infected individual $\langle 0 \rangle$ belongs by construction to every tree as the process always starts with a single infected individual. Then assuming that $\langle 0 \rangle$ transforms at time L_0 into $\nu_0(L_0) = n_0$ new infected individuals (n_0 offspring), then the infected individual labelled $\langle n_0 + 1 \rangle$ will not belong to this specific realisation X. Note that it is also possible, given the construction of the tree, to define generations $G_k(X)$, whose countable union makes up an entire family $F(X)$ (see Appendix A.2 for more details).

Next we need to introduce the concept of first-generation family history which is essential to facilitate recursive definitions and proofs later on. A *first-generation family history, denoted* X_i, is very simply an infinite sequence of random variables, similar to the sequence defined in Eq. (1) , but *instead of starting at the initial infected individual* $\langle 0 \rangle$, *the family history* X_i *starts at the first-generation offspring labelled* $\langle i \rangle$. Thus we can write:

$$X_i \coloneqq \left\{ L_i, \nu_i(L_0 + L_i), L_{i,i_1,\ldots,i_k}, \nu_{i,i_1,\ldots,i_k}\left(L_0 + L_i + \sum_{j=1}^{k} L_{i,i_1,\ldots,i_j}\right) : i_1,\ldots,i_k \in \mathbb{N}, k \in \mathbb{N} \right\}.$$

The history X_i is therefore the independent branch starting at offspring $\langle i \rangle$ of the initial infected individual. We can now usefully observe that a family history X is equally, and in a more condensed way, described as consisting of the infinite sequence made of the life-length L and offspring process

$\nu(L)$ of the initial particle, followed by all first-generation family histories X_i for $i = 1, 2, \ldots$. This alternative description of a family history on which the first-generation principle described below relies, plays a crucial role in the derivation of the integral equations for prevalence and incidence. So to summarise we also describe X as:

$$X = \{ \underbrace{L, \nu(L)}_{\text{For initial ind. } \langle 0 \rangle} ; \underbrace{X_1; X_2; \ldots; X_i; \ldots}_{\text{First-gen. histories}} \}. \tag{2}$$

2.2 Shifted family histories

First-generation family histories allow us to exploit the self-similarity property of age-dependent branching processes by re-defining a family history via a sequence of first-generation family histories as in Eq. (2). Indeed, considering any infected individual at any time point, the process of infection seeded from this individual is similar to the process originating from the initial infected individual. This fact also holds de facto for the first-generation offspring. The self-similarity property emanates from the assumptions that the life-length and the offspring process are both not only identically distributed but also independent from each other.

Mathematically, however, in the context of time-varying offspring process, it is not straightforward to recreate a full family history X from a set of first-generation families X_i. Strictly speaking, X_i and X do not induce the same probability measure, as they used to in a constant offspring environment described for instance in [20]. To this end we introduce yet another type of family histories, called *a shifted family history* and denoted $X^{(l)}$, for any given shift $l \geq 0$. (We note in passing here that this denomination is slightly loose in the sense that, whilst going from X to $X^{(l)}$, calendar time is not shifted and only the observation times of the offspring process are shifted when comparing $X^{(l)}$ to X). Thus:

$$X^{(l)} := \left\{ L, \underbrace{\nu(l+L)}_{\nu \text{ observed at shifted time}}, L_{i_1,\ldots,i_k}, \underbrace{\nu_{i_1,\ldots,i_k}\left(l + L + \sum_{j=1}^{k} L_{i_1,\ldots,i_j}\right)}_{\nu \text{ observed at shifted time}} : i_1, \ldots, i_k \in \mathbb{N}, k \in \mathbb{N} \right\}. \tag{3}$$

This l-shifted family history $X^{(l)}$ is defined in the same way as the original family history X in Eq. (1), except that the offspring process is observed at a time of death shifted by the quantity $l \geq 0$. Thus the corresponding families $F(X^{(l)})$ and $F(X)$ contain exactly the same collection of infected individuals but differ by the realisations of their respective histories X and $X^{(l)}$. Question can be raised as to why we introduce yet another type of family histories. Previously, as

per [20], introduction of first-generation family histories was based on the need to formalise the *first-generation principle* as it was this principle, combined with the fact that the measures induced by any X_i were the same as the measure induced by the original X, which allowed [20] to derive the renewal equation for the generating function of $Z(t, X)$. However, in our context, the probability measures induced by X_i differ from the measure induced by X. And shifted family histories are needed to connect the probability measure induced by a first-generation family history with the probability measure induced by its parent history X. Indeed we prove in Appendix A.20 that the conditional probability measure of a first-generation family history X_i given $L = l$ (where L is the life-length of the initial infected individual $\langle 0 \rangle$) corresponds exactly to the measure of the l-shifted family history $X^{(l)}$. Specifically, if μ_l denotes the measure induced by $X^{(l)}$ and if μ_i is the measure induced by X_i, then, by additionally defining $\mu_i(X_i) = \mu(X_i|L = l)$, it can be shown that $\mu_i(X_i) \sim \mu_l(X^{(l)})$. This identity between measures is key to deriving the integral equations which rely on the recursive formula that originates from the first-generation principle. Importantly and contrary to [20], it should be noted that in our context, the measures induced by first-generation families μ_i are not identical to $\mathbb{P}$, the measure induced by X. This again reinforces why the l-shifted family histories need to be introduced within a context of a time-varying offspring distribution, even though they were not needed in formulations where the reproduction number was constant.

Despite the somewhat technical formalism put forward in this exposition so far, we wish to emphasise that it is precisely the introduction of first-generation histories and of shifted family histories which makes the subsequent derivations of renewal equations altogether coherent and relatively straightforward.

2.3 Defining prevalence and incidence

We have now introduced all the essential components of an age-dependent branching process and can formally define prevalence and incidence. Recalling that *prevalence is the average/expected total number of infected individuals alive at any time point t in a branching process*, we start by counting the number of infected individuals whose age is smaller than a given value a at a point in time t. We denote by $Z_\iota(a, t, X)$ the indicator function for the infected individual labelled $\langle \iota \rangle$ belonging to the family $F(X)$. The indicator $Z_\iota(a, t, X)$ is equal to 1, and otherwise zero, if $\langle \iota \rangle$ is alive at time t and if its age at time point t is lower than the value a. From the indicator $Z_\iota(a, t, X)$, it then becomes possible to count the total number of infected individuals in the family

$F(X)$ which were born between two time points $t-a$ and t, denoted by $Z(a,t,X)$:

$$Z(a,t,X) \coloneqq \sum_{\langle\iota\rangle\in F(X)} Z_\iota(a,t,X). \tag{4}$$

Then we take $a\to\infty$ in order to consider all possible ages and denote by $Z_\iota(\infty,t,X) \coloneqq Z_\iota(t,X) \coloneqq \lim_{a\to\infty} Z_\iota(a,t,X)$ the function indicating simply if individual labelled $\langle\iota\rangle$ is alive at time t. Finally, and this is where prevalence emerges once expectation is taken, the total number of infected individuals alive at time t in the family $F(X)$, denoted now $Z(t,X)$, is defined by:

$$Z(t,X) \coloneqq \sum_{\langle\iota\rangle\in F(X)} Z_\iota(\infty,t,X) = \sum_{\langle\iota\rangle\in F(X)} Z_\iota(t,X), \tag{5}$$

and taking expectations, we define more formally:

$$\text{Prevalence} \coloneqq \mathbb{E}[Z(t,X)]. \tag{6}$$

In contrast, *incidence is defined as the rate of new infections at a time point t in a family* $F(X)$. To compute incidence in practice, it has proven convenient to take as a starting point the *cumulative incidence, which is the number of infected individuals born up to time t in the family* $F(X)$, and denoted here $N_b(t,X)$. Then incidence can in principle be derived by first evaluating the difference in cumulative incidence between two points $t-\delta t$ and t, and then by taking the limit as $\delta t\to 0$. In that sense, incidence can be viewed as the intensity function of the counting process $N_b(t,X)$:

$$I\,(t,X) \coloneqq \lim_{\delta t\to 0} \frac{\mathbb{E}[N_b(t,X)] - \mathbb{E}[N_b\,(t-\delta t,X)]}{\delta t}. \tag{7}$$

We stress for the moment that this definition need not be fully rigorous. Technically, the limit may not exist, in which case we can however resort to a distributional derivative.

Finally, a link between expectation of cumulative incidence $\mathbb{E}[N_b(t,X)]$ and prevalence $\mathbb{E}[Z(t,X)]$ can be highlighted by introducing the cumulative number of deaths, $N_d(t,X)$ i.e. the number of infected individuals that transformed (died/replaced), up to time t in the family $F(X)$. Indeed, given that the life-lengths L are known through their assumed density distributions $g(\cdot)$, the cumulative number of deaths is as easy to compute as the cumulative number of births when simulating an age-dependent branching process. Therefore we can alternatively define and compute in practice the number of infected individuals alive at time t, namely $Z(t,X)$, as:

$$Z(t,X) \overset{\Delta}{\coloneqq} N_b(t,X) - N_d(t,X), \tag{8}$$

and retrieve prevalence by taking expectation on both sides.

2.4 The principle of first-generation

Now that formal definitions related to an age-dependent branching processes have been outlined in Section 2.1, shifted families introduced in Section 2.2 and prevalence and incidence more precisely defined in Section 2.3, we have the material to introduce the so-called *principle of first-generation* originally introduced by Harris in [20], which is a critical ingredient to being able to derive the celebrated integral equations.

The first-generation principle relies on conditioning upon the initial infected individual (more precisely upon the value of its life-length and offspring process). We assume that the initial infected individual $\langle 0 \rangle$ transforms at a given time point L into a number $\nu(L) = n_0 > 0$ of offspring. Then, counting the number of infected individuals $Z(t, X)$ alive in the family $F(X)$ at a time point $t \geq L$, is equivalent to counting the number of infected individuals alive at the backwards shifted time point $t - L$ in each of the n_0 first-generation families $F(X_i)$, for $i = 1, \ldots, n_0$. This result is called the *first-generation principle* and as our proofs demonstrate, most derivations obtained in the study of age-dependent branching processes rely on applying it at some point or another.

The intuition behind this principle is that an age-dependent branching process is *self similar*. Each first-generation offspring of the initial infected individual *can be treated as the ancestor/originator of its own process described by* X_i, which is itself a sub-process of the overall family history X. Therefore computing prevalence at time t in X, i.e. counting the total number of infected individuals yet un-transformed in the family history X at time t, is equivalent to summing the numbers of infected individuals in all of the the sub-processes X_i's but at a time point backward shifted by the quantity L. It is important to notice that this accounting is correct only once the first infected individual has transformed at $t \geq L$, i.e. once the first infected individual has transformed (died/replaced). Therefore, we prove in Appendix B.7 that $Z(t, X)$ can be re-written for any $t \geq L$ as:

$$Z(t, X) = \sum_{i=1}^{\nu(L)} Z(t - L, X_i), \qquad \text{(Principle of first-generation)} \tag{9}$$

where, as per Eq. (2), X is composed of the sequence made first of the life-length and offspring process of the initial infected individual and next of all the first-generation family histories: $X = (L, \nu(L); X_1, X_2, \ldots, X_i, \ldots)$. Eq. (9) underscores how every first-generation family history defines in effect its own sub-process $Z(\cdot, X_i)$, but also how the sum of all the first-generation branching processes, after shifting time backwards by the life-length L of the initial individual, makes up the original entire process $Z(t, X)$. Similarly, applying the first-generation family principle to the

cumulative incidence up to time $t \geq L$ in the family history X, we obtain:

$$N_b(t, X) = 1 + \sum_{i=1}^{\nu(L)} N_b(t - L, X_i). \tag{10}$$

2.5 Using probability generating functions to derive an integral equation

Generating functions are very convenient and powerful tools to summarise the entire probability distribution of a process. Before deriving integral equations for the quantities of interest, namely prevalence, incidence and cumulative incidence, we first derive integral equations for the generating functions associated to the distributions of these random processes. We will then use the classical relationship between the partial derivatives of a generating functions evaluated at 1 and the moments of the associated distributions to derive renewal-type equations for prevalence and incidence.

In what follows, and because of the probability-theoretic reasons previously mentioned, we will manipulate shifted family histories $X^{(l)}$ and consider them as our family histories of interest. As a reminder, $X^{(l)}$ is directly related to X as it still represents the original family history X, with the caveat that the offspring process in $X^{(l)}$ is observed at an l-shifted time for any $l \geq 0$. Further ahead, we will see that once we have derived the renewal equation for $\mathbb{E}[Z(t, X^{(l)})], \mathbb{E}[N_b(t, X^{(l)})]$, and for $I(t, X^{(l)})$, it suffices to set $l = 0$ to recover the renewal equations related to the original family history X. Thus to aid readability, we will use the following simplified notations:

$$Z(t, l) \coloneqq Z(t, X^{(l)}), \quad N_b(t, l) \coloneqq N_b(t, X^{(l)}), \quad I(t, l) \coloneqq I(t, X^{(l)}), \tag{11}$$

where we adopt the convention that $X^{(0)} = X$. Now recalling the generic form of a probability generating function for a random process taking integer values $r \in \mathbb{N}$, we associate the generating function $F^{(l)}(s, t)$ to the probability distribution of the shifted branching process $Z(t, l)$ such that for the auxiliary argument $s \in \mathbb{C}, |s| < 1$, we define:

$$F^{(l)}(s, t) \coloneqq \sum_{r=0}^{\infty} \mathbb{P}[Z(t, l) = r] s^r, \tag{12}$$

$$F(t, l) \coloneqq \mathbb{E}[s^{Z(t,l)}], \tag{13}$$

where another simplified notation $F(t, l)$ was introduced with the implicit convention that $F^{(0)}(s, t) = F(s, t)$.

In our context, the first and second moments of the random processes related to the branching process all need to be finite and thus, given the classical connection per [25] and proven in Appendix C.13, between the partial derivatives of the generating functions evaluated at $s = 1$ and the moments of the processes, we assume that the first and second partial derivatives of $F(t, l)$ and subsequently of all the other introduced generating functions are also finite:

$$f(t, l) \coloneqq \partial_s F(t, l)|_{s=1} = \mathbb{E}[Z(t, l)], \qquad \text{(Mean of } Z(t, l)\text{/Prevalence)} \tag{14}$$

$$\partial_s^2 F(t, l)|_{s=1} = \mathbb{E}[Z^2(t, l)] - \mathbb{E}^2[Z(t, l)]. \qquad \text{(Variance of } Z(t, l)\text{)} \tag{15}$$

Note here that we have introduced another simplified functional notation by setting $f(t, l) \coloneqq \partial_s F(t, l)|_{s=1}$.

Similarly the generating function for the cumulative incidence $N_b(t, l)$ is given by:

$$F_b^{(l)}(s, t) \coloneqq \sum_{r=0}^{\infty} \mathbb{P}[N_b(t, l) = r]s^r = \mathbb{E}[s^{N_b(t,l)}]. \tag{16}$$

Next, given that there is no generating function for $I(t, l)$, we will rely on Eq. (31) which connects $N_b(t, l)$ and $I(t, l)$, and directly use the integral equation for the cumulative incidence to then take the limit of a finite difference in order to derive the integral equation for incidence.

Finally, the generating function of the offspring process $\nu(t)$, also called the secondary distribution, is defined by:

$$h(s, t) \coloneqq \sum_{k=0}^{\infty} \mathbb{P}[\nu(t) = k]s^k = \sum_{k=0}^{\infty} q_t(k)s^k. \tag{17}$$

We note here that the expectation of the offspring process $\mathbb{E}[\nu(t)]$ is exactly the time-varying reproduction number $R(t)$, in other words the average number of secondary cases or offspring arising from an infected individual. So again, using the classical relationship between moments and partial derivatives of generating functions, we obtain the relationship:

$$\partial_s h(s, t)_{|s=1} = \mathbb{E}[\nu(t)] = R(t). \tag{18}$$

The usage of generating functions allows to by-pass manipulation of intractable integrals that tend to occur when expectations of processes such as $Z(t, l)$ are taken. Indeed, once an integral equation is derived for the generating function $F(t, l) = \mathbb{E}[s^{Z(t,l)}]$, it suffices for instance to take the first partial derivative of the integral equation and evaluate it at $s = 1$ to deduce another integral equation for the mean of the process, i.e in this case the prevalence $\mathbb{E}[Z(t, l)]$.

Deriving integral equations for the generating functions of $Z(t,l)$ and of $N_b(t,X)$ is rather lengthy. We will thus only detail the derivation for prevalence, bearing in mind that the the logic for cumulative incidence is very similar. Briefly, the derivation highlights why shifted family histories, as well as first-generation family histories, were introduced to ensure that the first-generation principle could indeed be applied. Given the fact that calendar time stays unchanged when manipulating $X^{(l)}$, it is possible to still use the same *trick*, so commonly encountered in the literature on branching processes, i.e. to condition the expectation and split it into two parts, a first part where calendar time t is still before first transformation time L, i.e. $t \leq L$, and a second part where transformation of the initial infected individual has already occurred i.e. where $t > L$. Thus we write:

$$F(t,l) = \mathbb{E}[s^{Z(t,l)}] = \underbrace{\mathbb{E}[s^{Z(t,l)}\mathbf{1}_{\{L>t\}}]}_{A_l} + \underbrace{\mathbb{E}[s^{Z(t,l)}\mathbf{1}_{\{L\leq t\}}]}_{B_l}. \tag{19}$$

The term A_l, corresponding to the scenario where the first transformation has not yet occurred, is straightforward to derive, given that in this scenario, there is probability 1 that a single particle is alive (namely the initial particle):

$$\begin{aligned} A_l &= \mathbb{E}[s^{Z(t,l)}\mathbf{1}_{\{L>t\}}], \\ &= \mathbb{E}[s\mathbf{1}_{\{L>t\}}], \\ &= s\mathbb{P}[L>t], \\ &= s[1-G(t)]. \qquad \text{(Survival term for initial individual)} \end{aligned} \tag{20}$$

The term B_l is more involved, and uses the fact that the offspring process $\nu(\cdot)$ is observed at time $t+l$ in the shifted history $X^{(l)}$:

$$\begin{aligned} B_l &= \mathbb{E}[s^{Z(t,l)}\mathbf{1}_{L<t}], \\ &= \int_0^t \sum_{k=0}^{\infty} \mathbb{E}[s^{Z(t,l)}|L=u,\nu(L+l)=k]\mathbb{P}[\nu(L+l)=k|L=u]\mathbb{P}[L\in du], \\ &= \int_0^t \sum_{k=0}^{\infty} \mathbb{E}[s^{Z(t,l)}|L=u,\nu(L+l)=k]\underbrace{\mathbb{P}[\nu(u+l)=k]}_{q_{u+l}(k)}\underbrace{\mathbb{P}[L\in du]}_{g(u)du}, \\ &= \int_0^t \sum_{k=0}^{\infty} \underbrace{\mathbb{E}[s^{Z(t,l)}|L=u,\nu(L+l)=k]}_{C_l} q_{u+l}(k)g(u)du. \end{aligned} \tag{21}$$

Finally we have to deal with the inner conditional expectation C_l. Here we use the first-generation principle and also introduce an additional simplifying notation, $Z_i(t,l) := Z(t,X_i^{(l)})$. Thus the

first-generation principle can be re-written for a shifted process $X^{(l)}$ using these notations:

$$Z(t,l) = \sum_{i=1}^{\nu(L)} Z_i(t-L,l). \tag{22}$$

And then we derive:

$$\begin{aligned}
C_l &= \mathbb{E}[s^{Z(t,l)}|L=u,\nu(L+l)=k], \\
&= \mathbb{E}[\prod_{i=1}^{\nu(L+l)} s^{Z_i(t-L,l)}|L=u,\nu(L+l)=k], \quad \text{(First-generation principle applied to } X^{(l)}) \\
&= \mathbb{E}[\prod_{i=1}^{k} s^{Z_i(t-u,l)}|L=u], \quad (X_i^{(l)} \text{ does not depend on } \nu(L+l)) \\
&= \prod_{i=1}^{k} \mathbb{E}[s^{Z_i(t-u,l)}|L=u], \quad ((X_i^{(l)}|L=u)_i \text{ are independent}) \\
&= \prod_{i=1}^{k} \mathbb{E}[s^{Z(t-u,l+u)}], \quad (\text{By def. } (X_i^{(l)}|L=u)_i \sim X^{(l+u)}) \\
&= \Big(\mathbb{E}[s^{Z(t-u,l+u)}]\Big)^k, \quad (k \text{ identical terms}) \\
&= \Big(F(t-u,l+u)\Big)^k. \quad (\text{By def. of } F(\cdot,l+u))
\end{aligned} \tag{23}$$

Replacing the expression for C_l from Eq. (23) in Eq. (21), we obtain:

$$F(t,l) = s[1-G(t)] + \int_0^t \underbrace{\sum_{k=0}^{\infty} q_{l+u}(k) \overbrace{\Big(F(t-u,l+u)\Big)^k}^{\text{From Eq. (23)}}}_{\text{We recognise } h[F(t-u,l+u),l+u]} g(u)du,$$

$$F(t,l) = s[1-G(t)] + \int_0^t h[F(t-u,l+u),l+u]g(u)du,$$

which is the integral equation for the generating function $F(t,l)$. As noted, a similar derivation can be applied to derive an integral equation for the generating function $F_b(t,l)$ associated to the cumulative incidence. In summary, for any positive shift l, the final integral equations for the generating functions $F(t,l)$ and $F_b(t,l)$ are proven to be written as:

$$F(t,l) = s[1-G(t)] + \int_0^t h[F(t-u,l+u),l+u]g(u)du, \tag{24}$$

$$F_b(t,l) = s[1-G(t)] + s\int_0^t h[F_b(t-u,l+u),l+u]g(u)du. \tag{25}$$

Notice that the integral equations Eq. (24) and Eq (25) show some similarities with the celebrated integral equation for constant secondary infections originally derived by Bellman and Harris [3], and recalled here: (without the use of the simplified notation for $F(s,t) = F(t,0)$)

$$F(s,t) = s[1-G(t)] + \int_0^t h[F(s,t-u)]g(u)du. \quad \text{(Bellman–Harris generating function)} \tag{26}$$

After first partial derivative is taken with respect to s, the renewal equation for prevalence, assuming $R = R_0$ constant, is classically derived in [20] as well:

$$f(t) = 1 - G(t) + R_0 \int_0^t f(t-u)g(u)du, \qquad \text{(Bellman–Harris prevalence).} \tag{27}$$

Eq. (27) is very similar in form to the renewal equation commonly used in literature for instance in [16, 24, 27] where:

$$f(t) = R_0 \int_0^t f(t-u)g(u)du. \tag{28}$$

Eq. (28) and Eq. (27) are the same, except that the equation commonly used for incidence (28) excludes the $1-G(t)$ term. It can be tempting to simply change R_0 into $R(t)$ in Eq. (28) and assume the renewal equations commonly used in epidemiology arise from an age dependent branching process, but this is not valid. First, Eq. (27) from Bellman and Harris is derived for prevalence, but is approximately the same for incidence *given a constant* R_0. Second is important to therefore note that, although, under the assumption of constant reproduction number, the renewal equation commonly used in practice [15, 11, 26, 16] seems to be consistent mathematically with results coming from age-dependent branching processes, this only holds for R_0 constant. It is thus not possible to simply change R_0 into $R(t)$ in Eq. (28) and have a valid equation for the behaviour of an age-dependent branching process in a context where the reproduction number is time-varying. We will discuss this point further in the next Section 2.6 where we derive renewal equations for prevalence and incidence.

2.6 Renewal equations for incidence, cumulative incidence, and prevalence

As mentioned above, once integral equations are obtained for generating functions, the derivations of integral equations for prevalence $\mathbb{E}[Z(t,l)]$ as well as for the expectation of the cumulative incidence, $\mathbb{E}[N_b(t,l)]$, rely on the usual relationship between the first moment of the processes involved and the partial derivative of their generating functions evaluated at $s = 1$. Thus for instance, prevalence is derived using $\partial_s F(t,l)|_{s=1} = \mathbb{E}[Z(t,l)]$. Whilst we highlight here the three major equations, where we introduced the functional notations: $\mathbb{E}[Z(t,l)] := f_p(t,l)$ for prevalence, $\mathbb{E}[N_b(t,l)] := f_{ci}(t,l)$ for cumulative incidence and $\mathbb{E}[I(t,l)] := f_i(t,l)$ for incidence, full proofs can be found in Appendix (D):

Renewal equations for prevalence, cumulative incidence and incidence

$$f_p(t,l) = 1 - G(t) + \int_0^t R(l+u) f_p(t-u, l+u) g(u) du \qquad \text{(Prevalence)} \qquad (29)$$

$$f_{ci}(t,l) = 1 + \int_0^t R(l+u) f_{ci}(t-u, l+u) g(u) du \qquad \text{(Cumulative incidence)} \qquad (30)$$

$$f_i(t,l) = R(l+t)g(t) + \int_0^t R(l+u) f_i(t-u, l+u) g(u) du. \qquad \text{(Incidence)} \qquad (31)$$

The system of equations (29) to (31) consists of a full set of three independent renewal equations for prevalence (f_p), cumulative incidence (f_{ci}) and incidence (f_i). Technically, these are not exactly renewal equations, but rather *renewal-like equations* —- for simplicity we will refer to these are renewal equations, however. Indeed they can be transformed in order to highlight their renewal aspect by applying a simple change of variable $c := t - l$, on which we shall also rely for implementation of efficient computations in Section (3). These equations, even if this is perhaps not immediately evident, very much mirror the first generation principle underpinning their derivation. For each of these three equations, the first term in front of the integral in effect accounts for the dynamics of the initial infected individual, whilst the second, convolution-like, integral accounts for the renewal of further infected individuals.

It can be helpful to algorithmically view how these equations are approximately solved via a discrete Riemann sum with step size 1: i.e. $t = 1, 2, 3, 4, ..., n$. For further simplicity we will show the solution for cumulative incidence, but the term outside the sum can be readily changed to recover incidence or prevalence. In Algorithm 1 it can be seen that solving Equation 30 for a given t requires a loop across shifts l as well as a loop for the Riemann sum for u. Cumulative incidence is therefore a $n \times n$ matrix, where times points are rows and shifts are columns - the quantity we are ultimately interested in is the first column with no shift. Given that we are interested in $l = 0$, it can be tempting to simply set l to 0 and simplify equation 30, but this is not possible due to the recursive nature of equation 30, to extract f_{ci} when $l = 0$, one needs to solve, or sweep across all l first. Before comparing these renewal equations with those more commonly used in epidemiology, we should note here that we are only able to compare the commonly used incidence renewal equation below (Eq. (32)) to our newly derived equation only for incidence (Eq. (33)) because *to our knowledge, no such comparable self-consistent equations for prevalence or cumulative incidence exist in practice.* Juxtaposing the common incidence renewal equation to the newly derived one, where in Eq. (32) incidence is denoted $f(t)$ in order to distinguish it from the solution $f_i(t)$ of

Algorithm 1 Solving the renewal-type equation for cumulative incidence

Inputs: number of steps n, density $g(t)$, and time varying number $R(t)$
Objective: Approximate $f(t,l) := 1 + \int_0^t R(l+u) f_{ci}(t-u, l+u) g(u) du$
Set: $f[0,:] := h[0,:]$
Set: $c = 0$
for $t = 1, \ldots, n$ **do**
    for $l = 0, \ldots, n-t$ **do**
        $c = 0$
        for $u = 0, \ldots, t$ **do**
            $c = c + R[l+u] f[t-u, l+u] g[u]$ ▷ alternatively $R[l+t-u] f[u, l+t-u] g[t-u]$
        end for
        $f[t,l] = 1 + c$
    end for
end for
Extract $f[:,0]$

Eq. (33):

Commonly used renewal equation for incidence:

$$f(t) = R(t) \int_0^t f(t-u) g(u) du, \tag{32}$$

Equation for incidence derived for an age-dependent branching process with time-varying $R(t)$:

$$f_i(t,l) = R(l+t) g(t) + \int_0^t R(l+u) f_i(t-u, l+u) g(u) du, \text{ for any } l \geq 0, \tag{33}$$

These equations can also be approximated in discrete form as a Riemann sum:

$$\hat{f}_t = R_t \sum_{u=0}^{t} \hat{f}_{t-u} g_u, \tag{34}$$

$$\hat{f}_{i_{[t,l]}} = R_{l+t} g_t + \sum_{u=0}^{t} R_{l+u} \hat{f}_{i_{[t-u,l+u]}} g_u, \text{ for any } \mathbb{N}\ l \geq 0, \tag{35}$$

It can be noted that the usual properties of convolution hold, and in particular, by commutativity, the indices can be changed such that $\sum_{u=0}^{t} \hat{f}_{t-u} g_u = \sum_{u=0}^{t} \hat{f}_u g_{t-u}$, or $\sum_u R_{l+u} \hat{f}_{i_{[t-u,l+u]}} g_u = \sum_u R_{l+t-u} \hat{f}_{i_{[u,l+t-u]}} g_{t-u}$.

We observe that Eq. (34) corresponds exactly to the one used ubiquitously in epidemiology [11, 26, 15, 16, 19]. However, in both the discrete and continuous forms of the commonly used equation,

there is an erroneous lack of a term outside the convolution. This *missing term* outside the convolution in the common renewal equation is in some ways highly problematic, as it implies there could be at most one solution, and a trivial one, with zero infection. It is straightforward to notice that Eq. (32) provides no means to start an epidemic, i.e. does not provide with initial conditions. Thus in practice, this lack of initial condition within the commonly used equation is usually remedied by either specifying a seeding period (e.g. [15]) or more simply by setting $f_1 = 1$, and summing from $j = 2$.

One way to put Eq. (32) on somewhat more rigorous grounds would be to include a delta function, in order to represent the first infection arising at a fixed time $t = 0$:

$$f(t) = \underbrace{\delta_0(t)}_{\text{First inf.}} + R(t) \underbrace{\int_{(0,t]} f(t-u)g(u)du}_{\text{Convolution part}} \tag{36}$$

We note here that integration is over $(0, t]$, and not $[0, t]$, to ensure that incidence at time 0 is really a unit delta mass. Now introducing the indicator function $1_t(t) := 1$, and $1_t(u) := 0$ for all $t \neq u$, we can rewrite the convolution part of Eq. (36) for $t > 0$ as:

$$\begin{aligned}
\int_{(0,t]} f(t-u)g(u)du &= \int_{(0,t)} f(t-u)g(u)du + \int_{\{t\}} f(t-u)g(u)du, \\
&= \int_{(0,t)} f(t-u)g(u)du + \int_{\mathbb{R}} 1_t(u)f(t-u)g(u), du \\
&= \int_{(0,t)} f(t-u)g(u)du + g(t). \qquad \text{(By def. of } \delta_t(\cdot)\text{)}
\end{aligned} \tag{37}$$

For $t > 0$, the term $\delta_0(t)$ in Eq. (36) is no longer relevant, so the common renewal equation can be re-written for $t > 0$ as

$$f(t) = R(t)g(t) + R(t) \int_0^t f(t-u)g(u)du. \tag{38}$$

Eq. (38) can be explained intuitively: indeed, the first infected individual infects, with probability $g(t)$ (i.e. its generation time), a certain number of infected individuals at time t, $R(t)$, then, subsequently, the rest of the epidemic proceeds via the convolution term. The form of this equation also happens to be similar to the newly derived renewal equation (33) for incidence.

The key difference between the newly derived renewal equations and the common renewal equation is thus the inclusion of a shift, l, that originally arose due to the time varying nature of the reproduction process. Reconciling the commonly used Equations (32) or (38) with the newly derived Eq. (33) is not a straightforward mathematical task. First, as it was noted earlier, there exist no *commonly used* renewal equations for prevalence or cumulative incidence of a similar form

as (32), so the comparison between commonly used and newly derived equations must be restricted to incidence. We then note that the initial condition for both renewal equations is different in both the discrete case (Eq. (34)) and in the continuous case in Eq. (38). While the initial conditions in the continuous case do at first glance appear similar, the presence of l in the new renewal equations means the uncertainty when the initial infection happens is *propagated* through the entire infection process. This difference can be counter intuitive. To help illustrate the importance of this point, consider an initial infection $I = 1$ at $t = 0$, and then solve equations:

$$\hat{f}_0 = \hat{f}_{i_{[0,l]}} = I = 1, \quad l \in \{0, \dots, n\} \tag{39}$$

$$\hat{f}_t = R_t \sum_{u=1}^{t} \hat{f}_{t-u} g_u, \tag{40}$$

$$\hat{f}_{i_{[t,l]}} = \sum_{u=1}^{t} R_{l+u} \hat{f}_{i_{[t-u,l+u]}} g_u \tag{41}$$

We find numerically that the solutions $\hat{f}_t$ and $\hat{f}_{i_{[t,l]}}$ agree. This result has not yet been rigorously proven analytically, but both, simulations and closed form expressions of a recursion back to I, show that these two equations have an identical set of terms when expanded and just *differ by the way their respective terms factorise*. Visually, we can observe the factorisation difference in Figure 3, where for a sinusoidal $R(t)$ and Gamma distributed $g(t)$, the individual summands that make up Eq. (41) do not exactly match between the new renewal equations (Figure 3, Top Left) and the common equations (Figure 3, Middle Left) - however their solution (red lines in Figure 3, Top Right and Middle Right), or approximation to the continuous integral, agree to machine precision. These numerical results suggest that the commonly used and the newly derived equations for incidence are equivalent in the discrete case and under the aforementioned assumptions. Future work should investigate the properties, necessary and/or sufficient, for the equivalence to hold.

All in all, the correspondence of these discrete equations, both analytically and numerically, leads us to conjecture the relationship still holds in the continuous case.

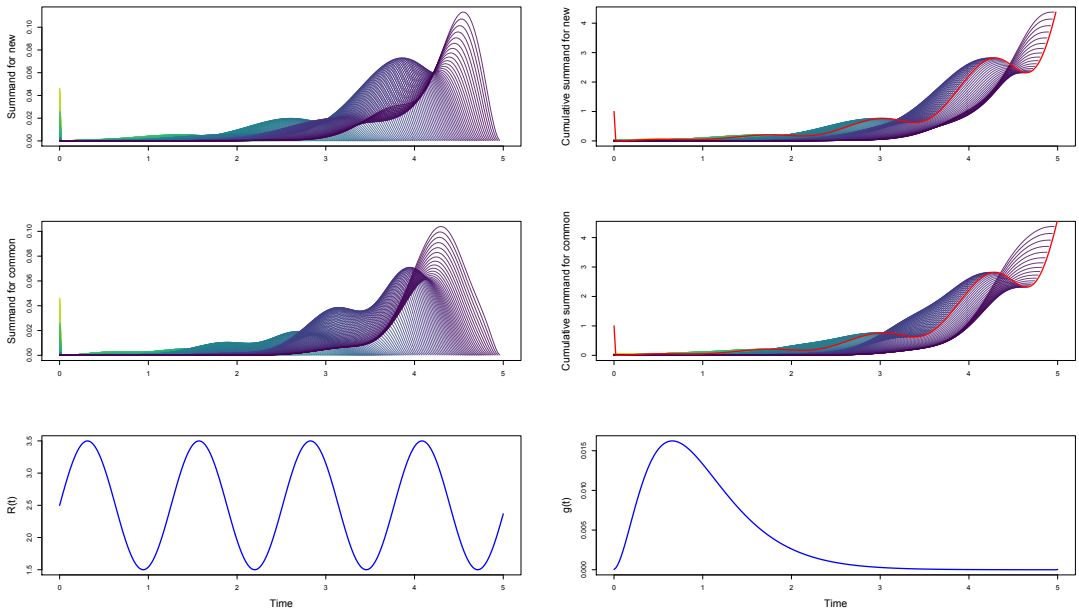

Figure 3: Simulation analysis of discrete renewal Equations. (35) and (34). **Top left:** individual summands for the new renewal equation. **Top Right**, cumulative/Riemann sum of the new renewal equation with solution in red. **Middle left:** individual summands for the old renewal equation, **Middle Right:** cumulative/Riemann sum of the old renewal equation with solution in red. **Bottom Left:** $R(t)$ as the sinusoidal function $2.5 + \sin(5t)$. **Bottom Right:** Gamma distributed $g(t)$ with shape and scale 3

Practically, the exact correspondence of Equations (41) is extremely important as it leads us to conjecture that, under a specific and simple initial condition, *the common renewal equation used in epidemiology is likely a special case of our equations.* Following from these results we can further conjecture that as $t \to \infty$, and $g(t) \to 0$, the common and new renewal equations will only ever differ by the initial conditions, and that this error will not accumulate as t becomes large.

This correspondence means all of the epidemiological theory using the current renewal equation (e.g. accounting for under ascertainment) can readily be applied to our equations. We are not aware of a mathematical way to exactly match the initial conditions between the common and the newly derived renewal equations without the use of a shift l. However, if this was possible, then simpler equations for incidence, prevalence and cumulative incidence may exist. At this point, we do not believe this reconciliation to be trivial. Indeed, consider on one hand changing from incidence to cumulative incidence, in the newly derived equations, one simply needs to change this

initial condition to 1, and the result functions become monotonic — there is no simple way to achieve this using the common equations.

2.7 Back calculating prevalence from incidence

As we have shown reconciliation of Eqs. (32) and (33) requires additional study and is interesting area for further developments. However, one aspect is clear: the commonly used renewal equation for incidence in a context of time-varying reproduction number does not provide us with a fully consistent system allowing to compute independently incidence or prevalence. Instead, in common practice, prevalence needs to be computed via a cumulative sum or alternatively by convolving incidence with the survival function $1-G(\cdot)$ — also called back calculating [12, 6]. From a practical standpoint, this traditionally implies that we rely on either incidence-based or on prevalence-based models, and that the other quantity can only be implicitly, but not independently, derived. In contrast, the system of equations in Eq. (33) allows to derive both prevalence and incidence independently and consistently.

We can reassuringly, we show in Appendix F that our equations reconcile with the back calculating of prevalence through incidence convoluted with the survival function of the generation interval [12], i.e. we are able to check that our equations are consistent with:

$$f_p(t,l) = [(1-G) * f_i(\cdot,l)](t). \tag{42}$$

It is useful here to describe some of the intuition behind this relationship. Indeed on one hand, incidence quantifies how many new infections arise at a given point in time t. On the other hand, we also know a statistical distribution describing how long a single infection survives/is being *active*, which is given by the survival function: $1 - G(t) = 1 - \int_0^t g(u)du$, representing the probability for an infection to *survive* between time 0 and t. Thus, to compute prevalence, which is the total number of infected individuals expected to be alive at time t, and therefore to have survived at time t, we need to *sum/integrate* the survival distribution against the number of newly born infected individuals at each time point up to t, which is mathematically equivalent to performing a convolution.

More precisely, in Appendix F, we show that the textbook relationship in Eq. (42) still holds for our renewal equations by proving first a general result for what could look like the convolution of

three non-negative Lebesgue measurable functions $f_1(\cdot,\cdot), f_2(\cdot)$, and $f_3(\cdot)$:

$$\int_0^t \left(\int_0^u f_1(u-s,s) f_2(s) ds \right) f_3(t-u) du = \int_0^t \left(f_1(\cdot,s) * f_3 \right)(t-s) f_2(s) ds. \tag{43}$$

Applying Eq. (43) to $I(\cdot,\cdot), R(\cdot)$ and $1-G(\cdot)$ allows us to prove consistency of the newly derived equations with the back calculation formula.

3 Efficient computation

In the literature as well as in practice, one can distinguish two main popular approaches to solving a renewal equation, namely on one hand the ones using numerical integration, and on the other hand, approaches based on Laplace transform (where convolution is transformed into a product). Here we present a computationally efficient procedure based on discretisation and vectorisation to help solve Eqs. (29) to (31). This algorithm relies on a discretisation method that is commonly used to solve renewal equations [11]. Furthermore, pragmatically, most epidemiological data is already discretised (hours, days, weeks), which only adds to the usefulness of this choice of scheme.

Based on Eqs. (29) to (31), we observe that the generic form of the integral equations for which we need numerical approximations is given by:

$$f(t,l) = h(t,l) + \int_0^t R(l+u) f(t-u, l+u) g(u) du, \tag{44}$$

where the function $h(t,l)$ in Eq. (44) translates respectively into $h(t,l) = 1-G(t)$ for prevalence, then into $h(t,l) = 1$ for cumulative incidence and finally into $h(t,l) = R(t+l)g(t)$ for incidence (see Section (2.6)). The first step in the discretisation procedure is to transform this generic integral equation into a simpler renewal-like equation. For this purpose we write:

$$l \coloneqq c - t, \tag{45}$$

where $c \in \mathbb{R}$ is a newly introduced parameter. And indeed, by also introducing the function $f_c(t)$ such as:

$$f_c(t) \coloneqq f(t, c-t) \tag{46}$$

we can re-write the generic integral equation Eq. (44) and highlight its renewal aspect:

$$f_c(t) = \underbrace{h(t,c-t)}_{h_c(t)} + \int_0^t R(c-(t-u)) f(t-u, c-(t-u)) g(u) du, \qquad \text{i.e.}$$

$$f_c(t) = h_c(t) + \int_0^\infty R[c-(t-u)] f_c(t-u) g(u) du. \tag{47}$$

Next, in order to discretise Eq. (47), we introduce $t := i\Delta$ to solve approximately for $f_c(i\Delta)$, for any $i = 0, 1, \ldots, n$ and for a step size $\Delta > 0$. We call the associated approximate numerical solution $\hat{f}_c(i\Delta)$. For $i = 0$, we need to set $\hat{f}_c(0) := h_c(0)$. Then for $i \geq 1$ and $0 \leq j \leq i$, we define recursively:

$$\hat{f}_c(i\Delta) := h_c(i\Delta) + \sum_{j=1}^{i} R[c - (i-j)\Delta]\hat{f}_c[(i-j)\Delta]g[(j-1)\Delta]\Delta. \tag{48}$$

Next we vectorise these equations in order to significantly decrease computation time. Letting n be the number of time steps, and choosing n values for c, namely, $c_1, \ldots, c_n \in \mathbb{R}$, for $0 \leq i \leq n$, we define the three n-dimensional vectors:

$$\hat{\boldsymbol{f}}(i\Delta) := \begin{bmatrix} \hat{f}_{c_1}(i\Delta) \\ \vdots \\ \hat{f}_{c_n}(i\Delta) \end{bmatrix} \in \mathbb{R}^n, \ \boldsymbol{h}_c(i\Delta) := \begin{bmatrix} h_{c_1}(i\Delta) \\ \vdots \\ h_{c_n}(i\Delta) \end{bmatrix} \in \mathbb{R}^n, \quad \boldsymbol{R}(i\Delta) := \begin{bmatrix} R(c_1 - i\Delta) \\ \vdots \\ R(c_n - i\Delta) \end{bmatrix} \in \mathbb{R}^n. \tag{49}$$

We can then rewrite Eq. (48) using this vectorisation:

$$\hat{\boldsymbol{f}}(i\Delta) := \boldsymbol{h}_c(i\Delta) + \sum_{j=1}^{i} g[(j-1)\Delta]\boldsymbol{R}[(i-j)\Delta] \odot \hat{\boldsymbol{f}}[(i-j)\Delta]\Delta, \tag{50}$$

where $\odot$ stands for element-wise (Hadamard product) vectors multiplication (and below matrices). Vectorising one dimension further, we then define the $n \times (i+1)$ matrices for $i = 1, \ldots, n$:

$$\begin{aligned} \boldsymbol{A}_i &:= \begin{bmatrix} \boldsymbol{R}(0) & \boldsymbol{R}(\Delta) \ldots & \boldsymbol{R}(i\Delta) \end{bmatrix} \in \mathbb{R}^{n \times (i+1)}, \\ \boldsymbol{B}_i &:= \begin{bmatrix} \hat{\boldsymbol{f}}(0) & \hat{\boldsymbol{f}}(\Delta) & \ldots & \hat{\boldsymbol{f}}(i\Delta) \end{bmatrix} \in \mathbb{R}^{n \times (i+1)}, \\ \boldsymbol{C}_i &:= \Delta \begin{bmatrix} g(i\Delta) & \ldots & g(\Delta) & g(0) \\ \vdots & & & \\ g(i\Delta) & \ldots & g(\Delta) & g(0) \end{bmatrix} \in \mathbb{R}^{n \times (i+1)}, \end{aligned}$$

such that Eq. (50) can be re-written as a product of matrices for $i = 1, \ldots, n$:

$$\hat{\boldsymbol{f}}(i\Delta) = \boldsymbol{h}_c(i\Delta) + \text{RowSum}(\boldsymbol{C}_{i-1} \odot \boldsymbol{A}_{i-1} \odot \boldsymbol{B}_{i-1}). \tag{51}$$

The RowSum in Eq. (51) can be replaced by a matrix-vector multiplication using a vector of ones. The matrix $\boldsymbol{B}_i$ is then built up recursively using:

$$\boldsymbol{B}_i = \text{Concatenate}\Big(\boldsymbol{B}_{i-1}, \boldsymbol{h}_c(i\Delta) + \text{RowSum}(\boldsymbol{C}_{i-1} \odot \boldsymbol{A}_{i-1} \odot \boldsymbol{B}_{i-1})\Big) \in \mathbb{R}^{n \times (i+1)}. \tag{52}$$

Finally it suffices to notice that by setting $c_i := i\Delta$, and extracting for $i = 1, \ldots, n$ the diagonal terms $\hat{f}_{i\Delta}(i\Delta)$, from $\boldsymbol{B}_n$, we obtain the approximate solution of the renewal-type equation we were

looking to solve for:

$$\hat{f}_{i\Delta}(i\Delta) \approx f_{i\Delta}(i\Delta) = f(i\Delta, i\Delta - i\Delta) = f(i\Delta, 0). \tag{53}$$

Algorithm 2 Discretisation of renewal-type equation

Inputs: step size Δ, number of steps n, $c_i = i\Delta$, for $i = 0, \ldots, n$, density $g(t)$, cdf $G(t)$, offspring number $R(t)$ and generic function $h_c(t)$

Objective: Compute $\hat{f}_c(i\Delta) := h_c(i\Delta) + \sum_{j=1}^{i} R(c - (i-j)\Delta)\hat{f}_c((i-j)\Delta)g((j-1)\Delta)\Delta$.

Set: $\hat{f}_c(0) := h_c(0)$

for $i = 0, \ldots, n$ **do**

    Compute $\boldsymbol{h}_c(i\Delta) := [h_{c_1}(i\Delta), \ldots, h_{c_n}(i\Delta)]$ and $\boldsymbol{R}(i\Delta) = [R(c_1 - i\Delta), \ldots, R(c_n - i\Delta)]$

end for

Set: $A_n \in \mathbb{R}^{n\times(n+1)}$, $B_n \in \mathbb{R}^{n\times(n+1)}$ and $C_n \in \mathbb{R}^{n\times(n+1)}$

for $i = 0, \ldots, n$ **do**

    Compute $A_n[:, i] = R[c - i\Delta]$

    Compute $C_n[:, i] = g[(n + 1 - i)\Delta]\Delta$

end for

Set: $\boldsymbol{B}_n[:, 0] = \boldsymbol{h}_c(0)$

for $i = 1, \ldots, n$ **do**

    Compute $\boldsymbol{B}_n[:, i] = \boldsymbol{h}_c(i\Delta) + \text{rowSums}\left(\boldsymbol{C}_n[:, (n + 1 - i) :] \odot \boldsymbol{A}_n[:, : i] \odot \boldsymbol{B_n}[:, : i]\right)$

end for

for $i = 0, \ldots, n$ **do**

    Set $c_i := i\Delta$

    Extract $\hat{f}_{i\Delta}(i\Delta) = B_n[i, i]$

    Use $\hat{f}_{i\Delta}(i\Delta) \approx f_{i\Delta}(i\Delta)$

end for

Algorithm 2 summarises the proposed discretisation procedure which proves very efficient in practice. Note that the indices and notations in Algorithm 2 are compatible with Python conventions. To further aid understanding, we include code in a GitHub repository

`https://github.com/ImperialCollegeLondon/BellmanHarris_simulation_and_inference`.

Note finally that this algorithm does not provide us with the most efficient implementation possible. We recover the approximate solution by taking the diagonal of the matrix $\boldsymbol{B}$, which means that

the matrices $\boldsymbol{A}$ and $\boldsymbol{C}$ would only need to be lower triangular and all multiplications and row sums could be only done over lower triangular matrices. Using lower triangular matrices would not only speed up computation time but also greatly reduce storage needs for large values of n. However here, for simplicity, we have not yet optimised this algorithm to exploit these savings. Also note that, in practical computation, dealing with full matrices instead of lower triangular requires us to set $R(t) := R_0$ for negative t as otherwise the upper triangular parts of the matrix $\boldsymbol{A}$ would treat cases where $c - i\Delta < 0$.

4 A simple Bayesian simulation case study

We perform a simulation case study to demonstrate fitting incidence and prevalence jointly in a probabilistic programming framework [7].

We simulate prevalence ($p_t := f_p(t)$) and incidence ($i_t := f_i(t)$) from Eqs. (29) to (31) over the time interval $[0 - 50]$ with a step size $dt = 0.5$. We use a generation interval $g \sim \text{Gamma}(3, 2)$ and a time varying reproduction number $R(t) = R_0 + \sin(\frac{1}{4}t)$ with $R_0 = 1.3$. This choice of $R(t)$ results in a common two wave epidemic [13]. We then add noise to the prevalence and incidence simulations as follows:

$$p_t^* = f_p(t) + \text{Normal}(0, \frac{2}{5}p_t) \tag{54}$$

$$i_t^* = f_i(t) + \text{Normal}(0, \frac{2}{5}i_t) \tag{55}$$

Note the prevalence's and incidences are rounded up to integers. We then use the known generation interval g (cdf G) and attempt to reconstruct R from the noisy observations of p_t^* and i_t^* using the following Bayesian hierarchical model

$$\beta \sim \text{Normal}(0, 1) \tag{56}$$

$$R(t) = S\beta^T \tag{57}$$

$$f_p(t, l) = 1 - G(t) + \int_0^t R(l + u) f_p(t - u, l + u) g(u) du \tag{58}$$

$$f_i(t, l) = R(l + t)g(t) + \int_0^t R(l + u) f_i(t - u, l + u) g(u) du \tag{59}$$

$$p_t^* \sim \text{Poisson}(f_p(t, 0)) \tag{60}$$

$$i_t^* \sim \text{Poisson}(f_i(t, 0)), \tag{61}$$

where S is a B-Spline basis over x with 10 degrees of freedom. We us our approximation algorithm in Section (3) and optimise (L-BFGS) a maximum a posteriori probability (MAP) estimate for spline coefficients β. We see in Figure (4) that our model captures the true prevalence and incidence closely despite the added noise. Our B-Spline basis also captures the sinusoidal pattern in the true $R(t)$, but there is uncertainty initially due to low case numbers. Using our framework, we can use two likelihoods to perform inference on $R(t)$. Fitting the model specified above with two likelihoods (Eq.(60) and Eq.(61)) versus just a single likelihood for prevalence results in a more accurate recovery of prevalence, incidence and $R(t)$. In the two likelihood model the mean squared error between the true and predicted prevalence was 226.7, incidence was 30.0 and $R(t)$ was 0.033. In the one likelihood model for prevalence,the mean squared error between the true and predicted prevalence was 436.5, incidence was 55.4 and $R(t)$ was 0.052

This simple example shows that joint fitting of both prevalence and incidence can be used to perform data synthesis that is more challenging using current renewal equation formulations. Further more, in this example, the simulated data or prevalence and incidence are independently augmented with noise. This is a scenario expected with real data that are separately collected; the sources of noise and amount of noise will be different. Using our approach, we can use our internally consistent equations to be more robust as we can utilise two likelihoods with the same mechanism.

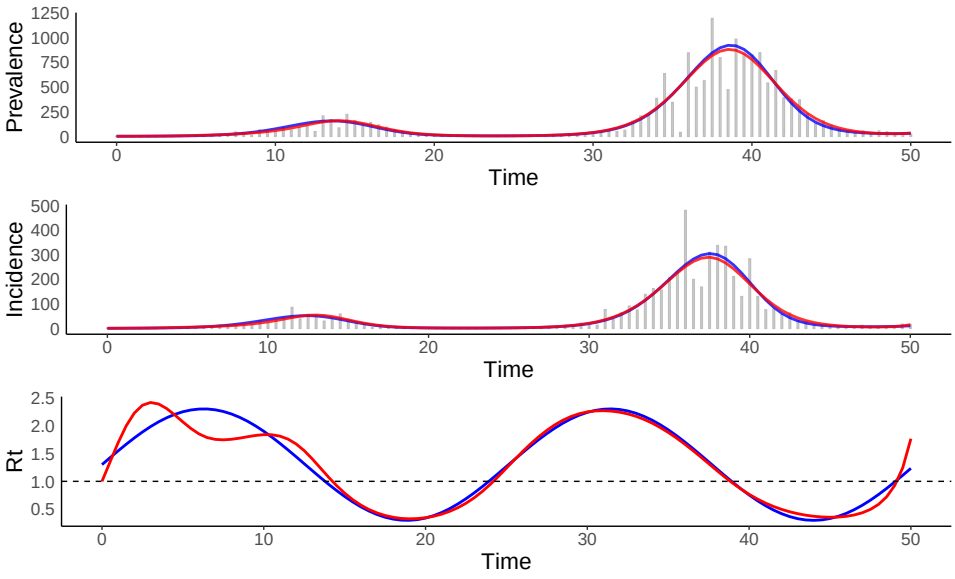

Figure 4: Simulation and fitting example. Top plot shows prevalence. The blue line is the true prevalence (p), the bars represent noisy prevalence levels (p^*), and the red line is the fitted mean. The middle plot shows incidence. The blue line is the true incidence (i), the bars are noisy incidences (i^*), and the red line is the fitted mean. Bottom plot is the simulated R_t in blue, and the B-Spline estimated R_t in red

Funding

This work was supported by Centre funding from the UK Medical Research Council under a concordat with the UK Department for International Development, the NIHR Health Protection Research Unit in Modelling Methodology and Community Jameel. This research was also partly funded by the Imperial College COVID-19 Research Fund. S. Bhatt acknowledges The UK Research and Innovation (MR/V038109/1), the Academy of Medical Sciences Springboard Award (SBF004/1080), The MRC (MR/R015600/1), The BMGF (OPP1197730), Imperial College Healthcare NHS Trust-BRC Funding (RDA02), The Novo Nordisk Young Investigator Award (NNF20OC0059309) and The NIHR Health Protection Research Unit in Modelling Methodology.

References

[1] G. K. Aldis and M. G. Roberts. "An integral equation model for the control of a smallpox outbreak". en. *Math. Biosci.* 195.1 (May 2005), pp. 1–22.

[2] R. Bartoszynski. "Branching Processes and the Theory of Epidemics". *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (1967).

[3] R. Bellman and T. E. Harris. "On the Theory of Age-Dependent Stochastic Branching Processes". *Proceedings of the National Academy of Sciences* (1948). DOI: 10.1073/pnas.34.12.601.

[4] R. Bellman and T. Harris. "On Age-Dependent Binary Branching Processes". *The Annals of Mathematics* (1952). DOI: 10.2307/1969779.

[5] A. T. Bharucha-Reid. "On the Stochastic Theory of Epidemics". *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 4: Contributions to Biology and Problems of Health*. Berkeley, Calif.: University of California Press, 1956, pp. 111–119. URL: https://projecteuclid.org/euclid.bsmsp/1200502550.

[6] R. Brookmeyer and M. H. Gail. "A Method for Obtaining Short-Term Projections and Lower Bounds on the Size of the AIDS Epidemic". *J. Am. Stat. Assoc.* 83.402 (June 1988), pp. 301–308.

[7] B Carpenter et al. "\emph{Stan}: A Probabilistic Programming Language". *Journal of Statistical Software* 76.1 (2017), pp. 1–32. DOI: 10.18637/jss.v076.i01. URL: http://www.jstatsoft.org/v76/i01/.

[8] S. Cauchemez et al. “Unraveling the drivers of MERS-CoV transmission”. *Proceedings of the National Academy of Sciences of the United States of America* (2016). DOI: 10.1073/pnas.1519235113.

[9] D. Champredon, J. Dushoff, and D. J. Earn. “Equivalence of the Erlang-distributed SEIR epidemic model and the renewal equation”. *SIAM Journal on Applied Mathematics* (2018). DOI: 10.1137/18M1186411.

[10] D. Champredon et al. “Two approaches to forecast Ebola synthetic epidemics”. en. *Epidemics* 22 (Mar. 2018), pp. 36–42.

[11] A. Cori et al. “A new framework and software to estimate time-varying reproduction numbers during epidemics”. *American Journal of Epidemiology* (2013). DOI: 10.1093/aje/kwt133.

[12] R. E. Crump and G. F. Medley. “Back-calculating the incidence of infection of leprosy in a Bayesian framework”. en. *Parasit. Vectors* 8 (Oct. 2015), p. 534.

[13] N. R. Faria et al. “Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil”. en. *Science* (Apr. 2021).

[14] N Ferguson et al. *Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand*. Tech. rep. Mar. 2020.

[15] S. Flaxman et al. “Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe”. en. *Nature* 584.7820 (Aug. 2020), pp. 257–261.

[16] C. Fraser. “Estimating individual and household reproduction numbers in an emerging epidemic”. *PLoS ONE* (2007). DOI: 10.1371/journal.pone.0000758.

[17] C. Fraser et al. “Factors that make an infectious disease outbreak controllable”. en. *Proc. Natl. Acad. Sci. U. S. A.* 101.16 (Apr. 2004), pp. 6146–6151.

[18] W. Getz and J. Lloyd-Smith. “Basic methods for modeling the invasion and spread of contagious diseases”. 2006. DOI: 10.1090/dimacs/071/05.

[19] N. C. Grassly and C. Fraser. *Mathematical models of infectious disease transmission*. 2008. DOI: 10.1038/nrmicro1845.

[20] T. E. Harris and Others. *The theory of branching processes*. Vol. 6. Springer Berlin, 1963.

[21] W. O. Kermack and A. G. McKendrick. “A Contribution to the Mathematical Theory of Epidemics”. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 115.772 (1927), pp. 700–721. DOI: 10.1098/rspa.1927.0118. URL: http://rspa.royalsocietypublishing.org/cgi/doi/10.1098/rspa.1927.0118.

[22] M Kimmel and D. Axelrod. *Branching processes in biology., 2002.*

[23] M. Kimmel. "The point-process approach to age- and time-dependent branching processes". *Advances in Applied Probability* 15.1 (1983), pp. 1–20.

[24] A. J. Lotka. "Studies on the mode of growth of material aggregates". *American Journal of Science* (1907). DOI: 10.2475/ajs.s4-24.141.199.

[25] J. C. Miller. "A primer on the use of probability generating functions in infectious disease modeling". *Infectious Disease Modelling* 3 (2018), pp. 192–248.

[26] P. Nouvellet et al. "A simple approach to measure transmissibility and forecast incidence". *Epidemics* (2018). DOI: 10.1016/j.epidem.2017.02.012.

[27] A. Polyanin. *Handbook of Integral Equations.* 1998. DOI: 10.1201/9781420050066.

[28] M. A. Rizoiu et al. "Expecting to be HIP: Hawkes intensity processes for social media popularity". *26th International World Wide Web Conference, WWW 2017.* 2017. DOI: 10.1145/3038912.3052650.

[29] M. G. Roberts. *Modelling strategies for minimizing the impact of an imported exotic infection.* 2004.

Appendices

A Constructing an age-dependent branching process with time varying offspring distribution

We consider a single-type branching process initialised by the *birth* of a unique infected individual at time $t = 0$. We assume no migration in or out. The random life-length of the initial infection is denoted L_0. This infected individual transforms at the end of its life into a random number ν_0 (offspring number) of new, but identical, infected individuals, all of age 0 at birth. Each infected individual is then fully determined by its random life-length L and its offspring distribution process $\nu(t)$. We show here how to construct a branching process based on those assumptions. As in [20], since most proofs involve measure-theoretic arguments, we start by giving an explicit description of the probability spaces on which the random quantities related to the branching process will live. However, even though the model under consideration shares many assumptions with [20], the main departing assumption of time-varying offspring number leads to a different path to constructing the probabilistic framework of the branching process.

In this section most notations and definitions come from [20] and, when needed, are then adapted to a time-varying offspring process.

A.1 Probability space of infected individuals

A.1.1 Collection of infected individuals

The initial infected individual is denoted $\langle 0 \rangle$. Next, the first, second or i-th child of this initial infected individual are denoted respectively $\langle 1 \rangle, \langle 2 \rangle, \ldots, \langle i \rangle$. Then again, the j-th child of the i-th child is denoted $\langle i, j \rangle$. More generally we define an infected individual as an ordered n-tuple:

Definition A.1 (Infected individual)**.** Any infected individual (besides the initial one $\langle 0 \rangle$) is defined for $n \in \mathbb{N}$ by an ordered n-tuple of strictly positive integers. This ordered n-tuple is denoted: $\langle \iota \rangle = \langle i_1, \ldots, i_n \rangle$, where $i_k \in \mathbb{N}$, for $k = 1, \ldots, n$.

This definition leads to the construction of the entire space of infected individuals.

Definition A.2 (Space $\mathcal{I}$ of infected individuals)**.** We denote by $\mathcal{I}$ the infinite collection of all infected individuals as per Def. A.1, i.e. $\mathcal{I}$ is the collection, for any $n \in \mathbb{N}$, of all ordered n-tuples

$\langle i_1, \ldots, i_n \rangle$, including the initial infected individual represented by the singleton $\{0\}$.

$$\mathcal{I} := \{\langle 0 \rangle\} \cup \{\langle i_1, \ldots, i_n \rangle \ : i_1, \ldots, i_n \in \mathbb{N}, n \in \mathbb{N}\}.$$

Each infected individual is then assumed to be fully determined by its random life-length and its offspring process, both of which are more formally defined below.

A.1.2 Life-length distribution

The life-length of an infected individual is a real-valued random quantity.

Definition A.3 (Life-length L)**.** The life-length L of an infected individual is a positive, real valued, continuous, random variable defined on $\mathcal{I}$ such as $L : \mathcal{I} \mapsto \mathbb{R}^+, \langle \iota \rangle \mapsto L_\iota$. The random variable L induces a probability measure $g(\cdot)$ defined on the Borel sigma-algebra $\mathcal{B}(\mathbb{R}^+)$.

Assumption A.4 (Assumptions related to L)**.** The random variable L forms an i.i.d sequence. All infected individuals follow the same life-length distribution. And as such, we will often use the same notations for the realisations L_ι of the random life-length of an infected individual $\langle \iota \rangle$ and for the random variable L. All the life-lengths are independent of each other.

A.1.3 Time-varying offspring process

Departing from the assumptions in [20], the random offspring number ν is assumed to depend on calendar time t, as in [23]. As such, it becomes a stochastic process.

Definition A.5 (Offspring process $\nu(t)$)**.** The offspring number $\nu(t)$ is the positive, integer valued, random process, defined on $\mathcal{I} \times \mathbb{R}^+$, such as: $\nu : \mathcal{I} \times \mathbb{R}^+ \mapsto \mathbb{N} \cup \{0\}, (\langle \iota \rangle, t) \mapsto \nu_\iota(t)$.

Assumption A.6 (Assumptions related to $\nu(t)$)**.** For any $t \in \mathbb{R}^+$, $\nu(t)$ form an i.i.d sequence. For any $t > 0$, a probability measure denoted $q_t(dy)$ describes the one-dimensional (marginal) distribution of $\nu(t)$. Additionally for any Borel set $A \in \mathcal{B}(\mathbb{R}^+)$, the map $A \mapsto q_t(A)$ is measurable which makes the marginal distribution $q_t(dy)$ a valid conditional distribution. The stochastic process $\{\nu(t)\}_{t \geq 0}$ and the random variables L are independent from each other.

A.1.4 Induced probability measure on $\mathcal{I}$

Having formalised the definition of infected individuals and their associated life-length random variable as well as their offspring process, we can now equip the space of infected individuals $\mathcal{I}$ with a probability measure $\mathbb{P}$ consistent with the assumptions that L and $\{\nu(t)\}_{t>0}$, are both i.i.d and independent from each other.

Definition A.7 (Probability measure $\mathbb{P}$ on $\mathcal{I}$)**.** The probability measure $\mathbb{P}$ on $\mathcal{I}$ is defined by the underlying assumptions:

- L_ι is i.i.d. and induce a probability measure $\mathbb{P}[L \in dl]$ on $\mathcal{B}(\mathbb{R}^+)$. $\mathbb{P}$ is further defined by a cumulative distribution function, $G(\cdot)$, assumed to be right continuous, with $G(0^-) = 0$ and $G(0+) < 1$. As in [20], it is also assumed that $G(\cdot)$ has density $g(\cdot)$ such as:
$$\mathbb{P}[0 \leq L \leq l] := G(l) = \int_0^l g(u)du.$$
- The process $\{\nu(t)\}_{t\geq 0}$ is i.i.d. and independent of all L_ι. It induces a marginal probability measure given by:
$$\mathbb{P}[\nu(t) \in dy] := q_t(dy).$$
Additionally the map: $\mathcal{B}(\mathbb{R}^+) \to [0,1], A \mapsto q_t(A)$ is measurable, which makes $q_t(dy)$ a valid conditional distribution. Finally, $\mathbb{P}$ is further defined by a probability mass function given by: $\mathbb{P}(\{\nu(t) = r\}) := q_t(\{r\}) := q_t(r)$.

A.2 Probability space Ω of family histories

A.2.1 Family histories as realisations of a branching process

The theory of age-dependent branching processes as introduced by [20] relies on the notion of family histories which are simply defined as infinite sequences of random variables tracing out the life-lengths and the offspring numbers of every infected individual, including the *dream infected individuals* of the original ancestor $\langle 0 \rangle$.

Definition A.8 (Family history X)**.** A *family history* X is an infinite sequence of pairs of random variables written:
$$X := \Big\{L_0, \nu_0(L_0), L_{i_1,\ldots,i_k}, \nu_{i_1,\ldots,i_k}\Big(L_0 + \sum_{j=1}^{k} L_{i_1,\ldots,i_j}\Big) : i_1, \ldots, i_k \in \mathbb{N}, k \in \mathbb{N}\Big\},$$

where for an infected individual $\langle \iota \rangle = \langle i_1, \ldots, i_k \rangle$, L_ι represents its life length, $\nu_\iota(L_0 + \sum_{j=1}^{k} L_{i_1,\ldots,i_j})$ represents its offspring number *observed at time of death*. The pairs $(L_\iota, \nu_\iota(L_0 + \sum_{j=1}^{k} L_{i_1,\ldots,i_j}))$ appear in a given order of enumeration. The state space of X is identifiable with $(\mathbb{R} \times \mathbb{N})^{\mathbb{N}}$.

Definition A.9 (Space Ω of family histories X)**.** The collection of all family histories X is denoted Ω.

Remark A.10. Each family history $X \in \Omega$ defines a *family tree*. For a given realisation X of this tree, certain infected individuals will belong to the realised tree, while some others will not. For instance the initial infected individual $\langle 0 \rangle$ belongs by construction to every tree. Assuming that for a certain realisation X, the initial infected individual $\langle 0 \rangle$ transforms at time L_0 into $\nu_0(L_0) = n_0$ children, then the infected individual labelled $\langle n_0 + 1 \rangle$ will not belong to this particular realisation X of the family history. The individual labelled $\langle n_0 + 1 \rangle$ is called *a dream child* of $\langle 0 \rangle$.

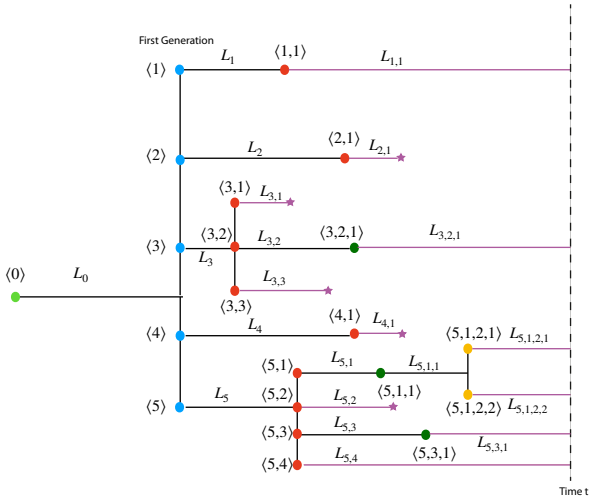

Figure 5: A realisation of a family history X.

Following this, the k-th generation $G_k(X)$ within a family history $X \in \Omega$ can be defined inductively:

Definition A.11 (Generations $G_k(X)$)**.** For a given family history X, the *k-th generation* $G_k(X)$ is defined inductively as follows: $G_0(X) = \langle 0 \rangle$, i.e. the first infected individual represents the 0-th generation. Then the first generation is defined as: $G_1(X) = \{\langle i \rangle, \text{ where } i \in \mathbb{N}, 1 \leq i \leq \nu_0(L_0)\}$, and $\nu_0(L_0)$ is the offspring number of infected individual $\langle 0 \rangle$ at its time of death L_0. Then, for $k \geq 2$, the k-th generation is given by:

$$G_k(X) := \{\langle i_1, i_2, \ldots, i_k \rangle \in \mathbb{N}^k, \text{ s.t. } \langle i_1, i_2, \ldots, i_{k-1} \rangle \in G_{k-1}(X), \nu_{i_1,\ldots,i_{k-1}}(L_0 + \sum_{j=1}^{k-1} L_{i_1,\ldots,i_j}) \geq i_k\}.$$

The countable union of all the generations within a family history X makes up an entire family, $F(X)$.

Definition A.12 (Family $F(X)$)**.** For a given family history X, the *family*, denoted $F(X)$, is defined as the countable union of all the generations $G_k(X)$ making up this family history.

$$F(X) := \bigcup_{k=0}^{\infty} G_k(X).$$

Remark A.13. Notations with respect to the initial infected individual $\langle 0 \rangle$ will now be simplified: $L_0 = L$ and $\nu_0 = \nu$.

A.2.2 First-generation family histories

In order to introduce the *first-generation principle* in section B.3, on which most of the following derivations rely, we define *first-generation family histories.*

Definition A.14 (First-generation family history)**.** For each family history X, and for any $i \in \mathbb{N}$, the *first-generation family history* X_i is defined as the infinite sequence of pairs (L, ν) starting at the child $\langle i \rangle$ of the initial infected individual $\langle 0 \rangle$:

$$X_i := \Big\{ L_i, \nu_i(L_0 + L_i), L_{i,i_1,\ldots,i_k}, \nu_{i,i_1,\ldots,i_k}\Big(L_0 + L_i + \sum_{j=1}^{k} L_{i,i_1,\ldots,i_j}\Big) : i_1, \ldots, i_k \in \mathbb{N}, k \in \mathbb{N} \Big\}.$$

X_i is represented by the branch starting at the child $\langle i \rangle$ of the initial infected individual $\langle 0 \rangle$.

In turn, each of these first-generation family histories defines a new probability sub-space:

Definition A.15 (Sub-space Ω_i of first-generation family histories X_i)**.** For any fixed $i \in \mathbb{N}$, the collection of all first-generation family histories X_i is denoted Ω_i.

Remark A.16. If $L = l$, and $\nu(l) = n$ and, additionally, if $i > n$, i.e. if i is strictly greater than the realised number of offspring of the initial infected individual, then the first-generation family history X_i is never realised and $X_i = \emptyset$ In this case, the infected individual labelled $\langle i \rangle$ is called a *dream-child* of $\langle 0 \rangle$.

A.2.3 Family histories with shifted offspring process

The first-generation principle introduced below is at the basis of all the integral equations derived further in this work. In [20], it is exactly the combination of the first generation principle and

the fact that the measures $\mathbb{P}_i$ on each sub-space Ω_i are all similar to $\mathbb{P}$, that leads to the renewal equation for the generating function of the prevalence. Given the time-varying nature of the offspring process in our context, we will see that the measures $\mathbb{P}_i$ on Ω_i are not similar to $\mathbb{P}$ and will only be similar with each other if they are defined conditionally on the value of the first life-length. This has led us to introduce another type of family histories, somehow loosely named *shifted family histories,* in order to derive the integral equations for the prevalence and incidence of the branching process in a context of time-varying offspring process. It will become apparent that this new type of shifted family histories allows to link the conditional probability measures of the first-generation family histories with the measure of the shifted family histories which we define more formally below.

Definition A.17 (Family history with shifted offspring $X^{(l)}$)**.** For any real valued quantity $l \geq 0$, we define $X^{(l)}$, family history with an l-shifted offspring process, informally abbreviated *l-shifted family history*, as the following infinite sequence of random variables:

$$X^{(l)} := \Big\{L, \underbrace{\nu(l+L)}_{\nu \text{ observed at shifted time}}, L_{i_1,\ldots,i_k}, \underbrace{\nu_{i_1,\ldots,i_k}\Big(l + L + \sum_{j=1}^{k} L_{i_1,\ldots,i_j}\Big)}_{\nu \text{ observed at shifted time}} : i_1, \ldots, i_k \in \mathbb{N}, k \in \mathbb{N}\Big\}.$$

Remark A.18. Both families, $F(X^{(l)})$ and $F(X)$, contain exactly the same set of infected individuals. Their respective histories X and $X^{(l)}$ include the same life-length realisations L_ι. However in the family history X, the offspring process $\nu(\cdot)$ is observed at the exact time of death of the infected individual, whereas in the family history $X^{(l)}$, the offspring process $\nu(\cdot)$ is observed at the shifted time of death, namely: $l + L + \sum_{j=1}^{k} L_{i_1,\ldots,i_j}$.

Shifted family histories being infinite sequences of random variables, they induce their own measure.

Definition A.19 (Shifted Measure μ_l)**.** For any $l \geq 0$, the l-shifted family histories $X^{(l)}$ induce a probability measure μ_l on the state space $\mathbb{R}^{\mathbb{N}}$ via:

$$\mu_l(dx) := \mathbb{P}[X^{(l)} \in dx].$$

A.2.4 Relationship between shifted and first-generation family histories

It will prove important to connect the measures induced by first-generation family histories with the ones induced by shifted-family histories when we derive integral equations for prevalence and cumulative incidence. It will indeed be this specific connection, allied to the first-generation principle, which will be at the basis of all further derivations. Thus we show here that the conditional

probability measure of a first-generation family history X_i given $L = l$ (where L is the life-length of the initial infected individual $\langle 0 \rangle$) corresponds exactly to the measure induced by the l-shifted family history $X^{(l)}$.

Lemma A.20 (Conditional measure of X_i given L)**.** *Given $L = l$ for $\langle 0 \rangle$, the random first-generation family histories $X_1, X_2, \ldots$ are mutually independent of each other, and for any $i \in \mathbb{N}$:*

$$\mu(X_i | L = l) \sim \mu_l.$$

Proof. Given $L = l$, we assume that the first particle $\langle 0 \rangle$ transforms into $\nu(l) = n > 0$ new particles. For any $i = 1, \ldots, n$, first, comparing the distribution of $X_i|L$:

$$\mu\bigg(X_i|L = l\bigg) \sim \mu\Bigg(\Big\{L_i, \nu_i(l + L_i), L_{i,i_1,\ldots,i_k}, \nu_{i,i_1,\ldots,i_k}(l + L_i + \sum_{j=1}^{k} L_{i,i_1,\ldots,i_j}\Big) : i_1, \ldots, i_k \in \mathbb{N}, k \in \mathbb{N}\Big\}\Bigg),$$

with the distribution of $X^{(l)}$:

$$\mu\bigg(X^{(l)}\bigg) \sim \mu\Bigg(\Big\{L, \nu(l + L), L_{i_1,\ldots,i_k}, \nu_{i_1,\ldots,i_k}\Big(l + L + \sum_{j=1}^{k} L_{i_1,\ldots,i_j}\Big) : i_1, \ldots, i_k \in \mathbb{N}, k \in \mathbb{N}\Big\}\Bigg),$$

and next, recalling that L and $\nu(t)$ are each i.i.d., whilst using independence of L from $\nu(t)$, it is straight forward to observe that these two random variables $X_i|L$ and $X^{(l)}$ induce the same probability measure. □

As a consequence, each of the first-generation sub-spaces Ω_i's can be equipped with its own probability measure $\mathbb{P}_i$:

Definition A.21 (Probability measure $\mathbb{P}_i$ on Ω_i)**.** Given $L = l$, the probability measure $\mathbb{P}_i$ of X_i on Ω_i is defined as the conditional probability of X_i on Ω given L, in other words as the measure on Ω of the l-shifted family history $X^{(l)}$.

$$\begin{aligned}
\mathbb{P}_i(dx_i) &:= \mathbb{P}(X_i \in dx_i | L = l), \\
&= \mu_l(dx_i), && \text{(By Lemma } A.20\text{)} \\
&= \mathbb{P}[X^{(l)} \in dx_i]. && \text{(By Def. of } \mu_l\text{)}
\end{aligned}$$

A.2.5 Induced probability measure on Ω

We observe here that a family history X can equally, and more condensely, be described by the infinite sequence of the initial particle life-length and offspring process followed by all its first-

generation family histories X_i. This alternative description of a family history proves crucial in the derivation of the following integral equations for prevalence and incidence.

Lemma A.22. *A family history X is equivalently described by the infinite sequence consisting first of the initial pair $(L, \nu(L))$ for the infected individual $\langle 0 \rangle$, followed by all the first-generation family histories X_i's of $\langle 0 \rangle$:*

$$X = \big\{L, \nu(L); X_1; X_2; \dots; X_i; \dots\big\}. \tag{62}$$

Proof. This equivalence results directly from the fact that all coordinates in the first-generation family histories $X_1, X_2, \dots$ comprise, *with no duplication*, all the coordinates in X, except for the original coordinates L and $\nu(L)$ of $\langle 0 \rangle$. □

Viewing a family history X as an infinite sequence of first-generation family histories leads to formalising the probability measure induced on Ω by X.

Lemma A.23 (Probability measure induced by X)**.** *The probability measure induced by a family history $X = \big\{L, \nu(L); X_1; X_2; \dots; X_i; \dots\big\}$, is given by:*

$$\mathbb{P}[X \in dx] = \mathbb{P}[L \in dl, \nu(L) = y] \bigotimes_{i=1}^{\infty} \mu_l(dx_i),$$

where μ_l is the probability measure induced by $X^{(l)}$, i.e. $\mu_l(dx_i) = \mathbb{P}[X^{(l)} \in dx_i]$. More explicitly in terms of finite-dimensional distributions, for any $i_1, \dots, i_k \in \mathbb{N}$ and $k \in \mathbb{N}$:

$$\mathbb{P}[L \in dl, \nu(L) = y, X_{i_1} \in dx_{i_1}, \dots, X_{i_k} \in dx_{i_k}] = q_l(y) g(l) dl \prod_{j=1}^{k} \mu_l(dx_{i_j}). \tag{63}$$

Finally Eq. (63) *can be refined further and the measure induced by X can be fully determined using only the life-length density and the density of the offspring process:*

$$\mathbb{P}[L \in dl, \nu(L) = y; X_{i_1} \in dx_{i_1}; \dots; X_{i_k} \in dx_{i_k}] = q_l(y) g(l) dl \prod_{j=1}^{k} q_{l+l_{i_j}}(y_{i_j}) g(l_{i_j}) dl_{i_j}. \tag{64}$$

Proof. Let us write as per Eq. (62), $X = \{L, \nu(L); X_1; X_2; \ldots; X_i; \ldots\}$. Then:

$$\begin{aligned}
\mathbb{P}[X \in dx] &= \mathbb{P}[L \in dl, \nu(L) = y, X_1 \in dx_1, X_2 \in dx_2, \ldots], \\
&= \mathbb{P}[L \in dl, \nu(L) = y] \bigotimes_{i=1}^{\infty} \mathbb{P}[X_i \in dx_i | L = l, \nu(L) = y], && \text{(Conditioning on the initial particle)} \\
&= \mathbb{P}[L \in dl, \nu(L) = y] \bigotimes_{i=1}^{\infty} \underbrace{\mathbb{P}[X_i \in dx_i | L = l]}_{\mu_l(dx_i)}, && (X_i \text{ indep. of } \nu(L)) \\
&= \underbrace{\mathbb{P}[L \in dl, \nu(L) = y]} \bigotimes_{i=1}^{\infty} \mu_l(dx_i), && \text{(By def. of } \mu_l) \\
&= \mathbb{P}[\nu(L) = y | L = l] \mathbb{P}[L \in dl] \bigotimes_{i=1}^{\infty} \mu_l(dx_i), && \text{(Conditioning on } L) \\
&= \mathbb{P}[\nu(l) = y] \mathbb{P}[L \in dl] \bigotimes_{i=1}^{\infty} \mu_l(dx_i), && (\nu(t) \text{ and } L \text{ indep.}) \\
&= q_l(y) g(l) dl \bigotimes_{i=1}^{\infty} \mu_l(dx_i). && \text{(By def. of } q_l(\cdot) \text{ and } g(\cdot))
\end{aligned}$$

We now prove Eq. (64). To simplify, we first consider a pair of first-generation family histories X_i and X_j and compute their joint probability with X :

$$\begin{aligned}
&\mathbb{P}[L \in dl, \nu(L) = y; X_i \in dx_i; X_j \in dx_j] \\
&= \mathbb{P}[L \in dl, \nu(L) = y; \underbrace{L_i \in dl_i, \nu_i(L + L_i) = y_i}_{X_i \in dx_i}; \underbrace{L_j \in dl_j, \nu_j(L + L_j) = y_j}_{X_j \in dx_j}], \\
&= \mathbb{P}[L_i \in dl_i, \nu_i(L + L_i) = y_i; L_j \in dl_j, \nu_j(L + L_j) = y_j | L = l, \nu(L) = y] \times \mathbb{P}[L \in dl, \nu(L) = y], \\
&= \mathbb{P}[L_i \in dl_i, \nu_i(L + L_i) = y_i; L_j \in dl_j, \nu_j(L + L_j) = y_j | L = l] \times \mathbb{P}[L \in dl, \nu(L) = y], \\
&= \mathbb{P}[L_i \in dl_i, \nu_i(L + L_i) = y_i | L = l] \mathbb{P}[L_j \in dl_j, \nu_j(L + L_j) = y_j | L = l] \times \mathbb{P}[L \in dl, \nu(L) = y], \\
&= \mathbb{P}[L_i \in dl_i, \nu_i(l + L_i) = y_i] \mathbb{P}[L_j \in dl_j, \nu_j(l + L_j) = y_j] \times \mathbb{P}[L \in dl, \nu(L) = y], \qquad (65)
\end{aligned}$$

where in the second equality we applied Bayes Law; in the third equality we used the fact that X_i and X_j are independent of $\nu(L)$ and in the fourth equality we used the independence of X_i from X_j. Next we simply observe that:

$$\begin{aligned}
\mathbb{P}[L_i \in dl_i, \nu_i(l + L_i) = y_i] &= \mathbb{P}[\nu_i(l + L_i) = y_i | L_i = l_i] \mathbb{P}[L_i \in dl_i], && \text{(By Def. of conditional prob.)} \\
&= \mathbb{P}[\nu_i(l + l_i) = y_i] \mathbb{P}[L_i \in dl_i], && \text{(By indep. of } L \text{ and } \nu(\cdot)) \\
&= q_{l+l_i}(y_i) g(l_i) dl_i. && \text{(By def. of } q_\cdot(\cdot) \text{ and } g(\cdot)).
\end{aligned}$$

Thus combining this last result with Eq. (65), we obtain:

$$\mathbb{P}[L \in dl, \nu(L) = y; X_i \in dx_i; X_j \in dx_j] = q_l(y) q_{l+l_i}(y_i) q_{l+l_j}(y_j) g(l) dl g(l_i) dl_i g(l_j) dl_j.$$

More generally for k first-generation family histories, we can generalise the formula above:

$$\begin{aligned}\mathbb{P}[L \in dl, \nu(L) = y; X_{i_1} \in dx_{i_1}; \ldots; X_{i_k} \in dx_{i_k}] & \\ &= q_l(y)g(l)dl \prod_{j=1}^{k} \mathbb{P}[X_{i_j} \in dx_{i_j}], \\ &= q_l(y)g(l)dl \prod_{j=1}^{k} q_{l+l_{i_j}}(y_{i_j})g(l_{i_j})dl_{i_j}.\end{aligned}$$

□

Remark A.24 (The measures $\mathbb{P}_i$ are not identical to $\mathbb{P}$)**.** In [20], the probability measures $\mathbb{P}_i(dx_i)$ on Ω_i are all identical between each other and also identical to the probability measure $\mathbb{P}(dx)$ on Ω. Here, however, in the context of time-varying offspring process, this is not the case anymore. The measures $\mathbb{P}_i(dx_i)$ on Ω_i are conditional measures of X_i on Ω given the initial life-length L :

$$\mathbb{P}_i(dx_i) = \mathbb{P}[X_i \in dx_i | L = l].$$

This is also why the l-shifted family histories need to be introduced within a context of a time-varying offspring distribution.

B Prevalence and incidence of the branching process

The main reason for constructing the probability spaces of infected individuals and of family histories was to build a measure-theoretic framework in order to give meaning to quantities such as the total number of infected individuals alive at time t in the family $F(X)$, which we will denote here $Z(t, X)$. In epidemiology, the expectation $\mathbb{E}[Z(t, X)]$ is commonly called *prevalence*. We show below how the build-up to the prevalence of the overall branching process starts at the level of one infected individual.

B.1 Prevalence

Before counting the overall number of infected individuals at any time point t in a branching process, we start by counting the number of infected individuals whose age is smaller than a given value a at a point in time t. We denote by $Z_\iota(a, t, X)$ the indicator function for the infected individual denoted $\langle\iota\rangle$ belonging to the family $F(X)$, which is equal to 1 if $\langle\iota\rangle$ is alive at time t and

if its age at time point t is lower than the value a. From the indicator $Z_\iota(a,t,X)$, it then becomes possible to count the number of infected individuals in the family $F(X)$ which were born between two time points $t-a$ and t. Indeed the total number of infected individuals in the family $F(X)$ who are alive at time t and whose age is smaller than a given value a at time t is defined as the sum of all the indicator functions $Z_\iota(a,t,X)$:

$$Z(a,t,X) := \sum_{\langle\iota\rangle\in F(X)} Z_\iota(a,t,X).$$

As $a\to\infty$, we denote by $Z_\iota(\infty,t,X) := Z_\iota(t,X)$ the function indicating simply if $\langle\iota\rangle$ is alive at time t. Finally, and this is where prevalence will emerge, the total number of infected individuals alive at time t in the family $F(X)$, denoted $Z(t,X) = Z(\infty,t,X)$, is defined by:

$$Z(t,X) := \sum_{\langle\iota\rangle\in F(X)} Z_\iota(\infty,t,X) = \sum_{\langle\iota\rangle\in F(X)} Z_\iota(t,X). \tag{66}$$

Remark B.1. A branching process can be defined directly through the stochastic process $Z(t,X)$. For a given family history X, the number of infected individuals alive at time t in X, i.e. $Z(t,X)$, is a stochastic process called a branching process, taking discrete values in $\mathbb{N}\cup\{0\}$:

$$\begin{aligned} Z : \mathbb{R}^+ \times X &\to \mathbb{N}\cup\{0\} \\ (t,X) \mapsto Z(t,X) &= \sum_{\langle\iota\rangle\in I(X)} Z_\iota(t,X). \end{aligned}$$

The formal definition of prevalence of a branching process follows.

Definition B.2 (Prevalence)**.** Given a branching process described by the stochastic process $Z(t,X)$, *prevalence* is defined as the expectation of the number of infected individuals alive at time t, i.e. $\mathbb{E}[Z(t,X)]$.

B.2 Incidence

In the study of infectious disease, two quantities of interest are commonly used in practice. We have just presented how prevalence, as the number of infected individuals alive at a time point t, builds up from individual-level indicator functions of whether an infected individual is alive and of a certain age. We now formally define incidence.

Definition B.3 (Incidence)**.** *Incidence*, denoted $I(t,X)$, is the point process defined as the number of new infected individuals *being born at a time point t in the family $F(X)$*.

This definition does not necessarily facilitate practical manipulation of the quantity in question in the study of infectious disease, hence we introduce an alternative definition of incidence based on the intensity of another point process, the cumulative incidence, which is itself easier to compute and manipulate. For this purpose we define:

Definition B.4 (Cumulative incidence)**.** *Cumulative incidence up to time* t, denoted $N_b(t, X,)$ is defined as the cumulative number of infected individuals that were born between time 0 and time t in the family $F(X)$.

Then, in practice and in continuous time settings, incidence is often described as the intensity of the cumulative incidence.

Definition B.5 (Incidence as intensity)**.** Incidence $I(t, X)$ is the intensity of the cumulative number of infected individuals $N_b(t, X)$ born up to time t.

Remark B.6. Our simulations are based on computing the quantity $\{\mathbb{E}[N_b(t, X)] - \mathbb{E}[N_b(t - dt, X)]\}/dt$, for small values of dt and approximating:

$$I(t, X) = \frac{\mathbb{E}[N_b(t, X)] - \mathbb{E}[N_b(t - dt, X)]}{dt}, \text{ for small } dt, \tag{67}$$

where we stress that for the moment this definition in Eq. (67) is purely formal as we we not know if this limit exists. And in fact it does not always,in which case we revert to distributional derivative.

Finally, we observe that we can also link cumulative incidence in expectation and prevalence, by introducing the cumulative number of deaths, $N_d(t, X)$ which is the number of infected individuals that died, up to time t, in the family $F(X)$. It follows immediately that the number of infected individuals alive at time t, namely $Z(t, X)$ can be described as:

$$Z(t, X) = N_b(t, X) - N_d(t, X).$$

And then:

$$\underbrace{\mathbb{E}[Z(t, X)]}_{\text{Prevalence}} = \mathbb{E}[N_b(t, X)] - \mathbb{E}[N_d(t, X)].$$

B.3 First-generation principle for prevalence

The result called *first-generation principle* is at the basis of most derivations obtained in the study of age-dependent branching processes. In particular, this principle is stated in Theorem 6.1, Ch.V

in [20]. To summarise it in words, we quote first [20]: *the branching process is the superposition of the processes initiated by the first-generation progeny of the initial infected individual.* Alternatively [23] describes the first generation principle as follows: *Each new offspring of the initial infected individual at time of transformation L can be treated as the ancestor of its own process X_i, which is itself a component of the process X. It then follows that the number of infected individuals alive in X at time t is equal to the sum of the numbers of infected individuals alive in all the sub-processes. This bookkeeping is correct for $t \geq L$, i.e. once the first infected individual has died.*

More formally, we assume that the initial infected individual $\langle 0 \rangle$ transforms at a given time point L into $\nu(L) = n > 0$ offspring. Then, counting the number of infected individuals $Z(t, X)$ alive in the family $F(X)$ at a time point $t \geq L$, is equivalent to counting, within the n first-generation families $(F(X_i))_{i=1}^n$, the number of infected individuals alive at a shifted time point $t - L$.

It is cumbersome but easy enough to show that Theorem 6.1, Ch. V in [20] still holds true in the context of a time-varying offspring process.

Theorem B.7 (First generation principle for prevalence)**.** *Let X be a family history. Assume the first infected individual $\langle 0 \rangle$ has a life length L, and that it transforms into at least one child, i.e. $\nu(L) := n > 0$. Let $(X_i)_{i=1}^n$ be the n first-generation family histories as defined in Def. A.14. Then at any point in time past the first transformation, $t \geq L$, the total number of infected individuals alive in the family history X is given by:*

$$Z(t, X) = \sum_{i=1}^{\nu(L)} Z(t - L, X_i), \tag{68}$$

where $X = (L, \nu(L); X_1, X_2, \ldots, X_i, \ldots)$.

In order to prove Eq. (68), we need the following lemma:

Lemma B.8. *For particle $\langle \iota \rangle = \langle i_1, \ldots, i_k \rangle$, the indicator function $Z_\iota(y, t, X) = 1$ if the three*

following conditions $C_1(X), C_2(X), C_3(X)$ *are satisfied simultaneously:*

$$Z_\iota(y,t,X)=1 \iff C_1(X): \begin{cases} i_1 & \leq \nu(L), \\ i_2 & \leq \nu_{i_1}(L+L_{i_1}), \\ \vdots & \\ i_k & \leq \nu_{i_1,i_2,\ldots,i_{k-1}}(L+L_{i_1}+L_{i_1,i_2}+\ldots+L_{i_1,\ldots,i_{k-1}}), \end{cases}$$

$$C_2(X): t-y \leq L+L_{i_1}+L_{i_1,i_2}+\ldots+L_{i_1,i_2,\ldots,i_{k-1}} \leq t,$$

$$C_3(X): L+L_{i_1}+L_{i_1,i_2}+\ldots+L_{i_1,i_2,\ldots,i_{k-1}}+L_{i_1,i_2,\ldots,i_k} > t,$$

where:

$$C_1(X) := \langle\iota\rangle \textit{ belongs to the } k\textit{-th generation of family } I(X)\ ,$$

$$C_2(X) := \langle\iota\rangle \textit{ was born between } t-y \textit{ and } t,$$

$$C_3(X) := \langle\iota\rangle \textit{ dies after time } t,$$

We can now prove Eq. (68).

Proof. Let X be a family history and we denote by L and $\nu(L)$ respectively the life length and the offspring number at time of death of the first particle $\langle 0\rangle$. First we recall the three conditions $C_1(X), C_2(X), C_3(X)$ for particle $\langle\iota\rangle = \langle i_1, i_2, \ldots, i_k\rangle$ in the family $I(X)$ to be alive at time t and of age smaller than y, where $C_1(X)$ means that $\langle\iota\rangle \in I_k(X)$, $C_2(X)$ means that $\langle\iota\rangle$ was born between $t-y$ and t and $C_3(X)$ means that $\langle\iota\rangle$ dies after time t (as per Lemma B.8):

$$C_1(X): \begin{cases} i_1 \leq \nu(L), \\ i_2 \leq \nu_{i_1}(L+L_{i_1}), \\ i_3 \leq \nu_{i_1,i_2}(L+L_{i_1}+L_{i_1,i_2}), \\ \vdots \\ i_k \leq \nu_{i_1,i_2,\ldots,i_{k-1}}(L+L_{i_1}+L_{i_1,i_2}+\ldots+L_{i_1,\ldots,i_{k-1}}). \end{cases}$$

$$C_2(X): t-y \leq L+L_{i_1}+L_{i_1,i_2}+\ldots+L_{i_1,i_2,\ldots,i_{k-1}} \leq t.$$

$$C_3(X): L+L_{i_1}+L_{i_1,i_2}+\ldots+L_{i_1,i_2,\ldots,i_{k-1}}+L_{i_1,i_2,\ldots,i_{k-1},i_k} > t.$$

We wish to show that for $0 \leq L \leq t$:

$$Z(t, X) = \sum_{i=1}^{\nu(L)} Z(t - L, X_i).$$

First for particle $\langle \iota \rangle = \langle i_1, i_2, \ldots, i_k \rangle$, and for a first-generation child $\langle i \rangle$, let us show that:

$$Z_{i,\iota}(t, X) = Z_\iota(t - L, X_i) \tag{69}$$

To show the equality in Eq. (69), we can assume that $\langle \iota \rangle = \langle i_1, i_2 \rangle$. Indeed the case where $\langle \iota \rangle = \langle 0 \rangle$ or the general case for $k \geq 2$ is dealt with in an exactly similar way as the case $k = 2$. So we want to show that for $\langle \iota \rangle = \langle i_1, i_2 \rangle$ when $0 \leq L \leq t$ and $1 \leq i \leq \nu(L)$:

$$Z_{i,\iota}(t, X) = Z_\iota(t - L, X_i),$$

i.e.

$$\underbrace{Z_{i,i_1,i_2}(t, X)}_{A_i(X)} = \underbrace{Z_{i_1,i_2}(t - L, X_i).}_{B_i(X_i)}$$

We recall that $Z_\iota(t, X)$ are indicator functions defined as:

$$Z_\iota(t, X) = 1 \iff \langle \iota \rangle \text{ is alive at } t \text{ in the family } I(X). \tag{70}$$

Thus applying Eq. (70) to $A_i(X)$:

$$A_i(X) = Z_{i,i_1,i_2}(t, X) = 1 \iff \langle i, i_1, i_2 \rangle \text{ is alive at } t \text{ in the family } I(X). \tag{71}$$

We now use the conditions $C_1(X), C_2(X), C_3(X)$ to get necessary and sufficient conditions for $\langle i, i_1, i_2 \rangle$ to be alive at time t. We will call these specific conditions $C_{1,A_i}(X), C_{2,A_i}(X), C_{3,A_i}(X)$:

$$C_{1,A_i}(X) : \begin{cases} i \leq \nu(L), \\ i_1 \leq \nu_i(L + L_i), \\ i_2 \leq \nu_{i,i_1}(L + L_i + L_{i,i_1}), \end{cases}$$

$$C_{2,A_i}(X) : 0 \leq L + L_i + L_{i,i_1} \leq t,$$

$$C_{3,A_i}(X) : L + L_i + L_{i,i_1} + L_{i,i_1,i_2} > t.$$

Now we turn to $B_i(X_i) = Z_\iota(t - L, X_i)$. This is another indicator function:

$$B_i(X_i) = Z_{i_1,i_2}(t - L, X_i) = 1 \iff \langle i_1, i_2 \rangle \text{ is alive at } t - L \text{ in the family } I(X_i).$$

Again we will denote by $C_{1,B_i}(X_i), C_{2,B_i}(X_i), C_{3,B_i}(X_i)$ the necessary and sufficient conditions for $\langle i_1, i_2\rangle$ to be alive at time $t-L$ in the family $I(X_i)$. (Recall we have assumed $t \geq L$.)

But here we introduce the following notations where the subscript $'$ means that we consider a descendent of child $\langle i\rangle$:

$$X' = X_i, \tag{72}$$

$$L'_\iota = L_{i,\iota}, \tag{73}$$

$$\tau'_\iota = \tau_{i,\iota}, \text{ i.e. } \tau'_\iota = L + L_i + L_{i,i_1} + L_{i,i_1,i_2}, \tag{74}$$

$$\nu'_\iota(\tau'_\iota) = \nu_{i,\iota}(\tau_{i,\iota}), \text{ i.e.: } \nu'_\iota(\tau'_\iota) = \nu_{i,\iota}(L + L_i + L_{i,i_1} + L_{i,i_1,i_2}). \tag{75}$$

We now recall the definition of the first-generation family history X_i as per Def. A.14:

$$X_i = (L_i, \nu_i(L+L_i); L_{i,i_1}, \nu_{i,i_1}(L + L_i + L_{i,i_1}); L_{i,i_1,i_2}, \nu_{i,i_1,i_2}(L + L_i + L_{i,i_1} + L_{i,i_1,i_2}); \ldots).$$

Replacing with the notations in Eqs. (72) to (75), we can re-write:

$$X' = (L', \nu'(L+L_i); L'_{i_1}, \nu'_{i_1}(L + L_i + L_{i,i_1}); L'_{i_1,i_2}, \nu'_{i_1,i_2}(L + L_i + L_{i,i_1} + L_{i,i_1,i_2})).$$

Now we can go back to the indicator function $B_i(X_i)$:

$$B_i(X_i) = 1 \iff Z_{i_1,i_2}(t-L, X_i) = 1$$
$$B_i(X') \iff Z_{i_1,i_2}(t-L, X') = 1$$

and we apply conditions $C_{1,B_i}(X_i), C_{2,B_i}(X_i), C_{3,B_i}(X_i)$ where t is replaced by $t-L$ and X_i is replaced by X' :

$$C_{1,B_i}(X') : \begin{cases} i_i \leq \nu'(L+L_i), \\ i_2 \leq \nu'_{i_1}(L + L_i + L_{i,i_1}), \end{cases}$$

$$C_{2,B_i}(X') : 0 \leq L_i + L_{i,i_1} \leq t - L,$$

$$C_{3,B_i}(X') : L_i + L_{i,i_1} + L_{i,i_1,i_2} > t - L.$$

To clarify, the conditions for $A_i(X)$ to be equal to 1 and for $B_i(X_i)$ to be equal to 1 are compared in the table below, where we recall by Eq. (73): $L' = L_i$ and by Eq. (75): $\nu'(L+L_i) = \nu_i(L+L_i)$

$Z_{i,i_1,i_2}(t,X)=1$	$Z_{i_1,i_2}(t-L,X_i)=1$
$i \leq \nu(L)$	
$i_1 \leq \nu_i(L+L_i)$	$i_1 \leq \nu'(L+L_i)$ i.e. $i_1 \leq \nu_i(L+L_i)$
$i_2 \leq \nu_{i,i_1}(L+L_i+L_{i,i_1})$	$i_2 \leq \nu'_{i_1}(L+L_i+L_{i,i_1})$ i.e. $i_2 \leq \nu_{i,i_1}(L+L_i+L_{i,i_1})$
$0 \leq L+L_i+L_{i,i_1} \leq t$	$0 \leq L'+L'_{i_1} \leq t-L$ i.e. $0 \leq L_i+L_{i,i_1} \leq t-L$
$L+L_i+L_{i,i_1}+L_{i,i_1,i_2} > t$	$L'+L'_{i_1}+L'_{i_1,i_2} > t-l$ i.e. $L_i+L_{i,i_1}+L_{i,i_1,i_2} \geq t-L$

But the condition $i \leq \nu(L)$ is already assumed to start with. So we have shown that for $0 \leq L \leq t, 1 \leq i \leq \nu(L)$:

$$Z_{i,\iota}(t,X) = Z_\iota(t-L,X_i). \tag{76}$$

We now sum both sides of Eq. (76) over all particles $\langle\iota\rangle \in I(X)$ and then over $i = \{1,\ldots,\nu(L)\}$:

$$\sum_{i=1}^{\nu(L)} \sum_\iota Z_{i,\iota}(t,X) = \sum_{i=1}^{\nu(L)} \sum_\iota Z_\iota(t-L,X_i). \tag{77}$$

On the left hand side of Eq. (77), this is equivalent to summing over all elements in $I(X)$ whose first coordinate is i, except for the element $\langle 0\rangle$. However $Z_0(t,X) = 0$ for $t \geq L$ because the first individual $\langle 0\rangle$ is not alive by construction after L. So on the left hand side we end up with the sum over all elements of the form $\langle i,\iota\rangle$ where $0 \leq i \leq \nu(L)$. This is exactly the sum over all the elements in the family $I(X)$. Thus the left hand side is equal to $Z(t,X)$.

Now for the right hand side of Eq. (77), we use Eq. (66):

$$\sum_{i=1}^{\nu(L)} \sum_\iota Z_\iota(t-L,X_i) = \sum_{i=1}^{\nu(L)} Z(t-L,X_i).$$

To summarise we have shown:

$$Z(t,X) = \sum_{i=1}^{\nu(L)} Z(t-L,X)$$

which is what we needed to prove. □

B.4 First generation principle for cumulative incidence

Although not stated specifically by [20] but used implicitly by [23], the first-generation principle also holds for the cumulative incidence up to time t, albeit with the difference that the initial infected individual $\langle 0\rangle$ needs to be accounted for in the cumulative sum.

Theorem B.9 (First generation principle for cumulative incidence)**.** *With similar assumptions as in theorem B.7, the cumulative incidence up to time $t \geq L$ in the family history X is given by:*

$$N_b(t, X) = 1 + \sum_{i=1}^{\nu(L)} N_b(t - L, X_i),$$

where $X = (L, \nu(L); X_1, X_2, \ldots, X_i, \ldots)$.

Proof. The proof is analogous to that of Theorem B.7. □

C Using probability generating functions to derive renewal equations

Generating functions have proven to be very convenient and minimal tools to summarise the entire probability distribution of a process. We introduce here various generating functions, namely $F(s,t)$ which corresponds to the branching process $Z(t, X)$, $F^{(l)}(s,t)$ for the process with shifted offspring distribution $Z(t, X^{(l)})$, $F_b(s,t)$ for the cumulative incidence $N_b(t, X)$ and finally $h(s,t)$ for the offspring process $\nu(t)$.

C.1 Generating functions related to the branching process

Here we give the formal definitions:

Definition C.1 (Generating function for $Z(t, X)$)**.** For a given family history X, the generating function associated to the probability distribution of the branching process $Z(t, X)$ is defined for $s \in \mathbb{C}$, $|s| < 1$ by:

$$\begin{aligned} F(s,t) &:= \sum_{r=0}^{\infty} \mathbb{P}[Z(t, X) = r] s^r \\ &= \mathbb{E}[s^{Z(t,X)}]. \end{aligned} \tag{78}$$

Whilst the branching process $Z(t, X)$ represents the number of infected individuals alive at time t in the family $F(X)$, we can also introduce the shifted branching process $Z(t, X^{(l)})$ which represents the number of infected individuals alive at time t in the family $F(X^{(l)})$. Similarly to Def. C.1, we introduce the generating function associated to the branching process with shifted offspring distribution:

Definition C.2 (Generating function for $Z(t, X^{(l)})$)**.** We associate a generating function $F^{(l)}(s,t)$ to the probability distribution of the shifted branching process $Z(t, X^{(l)})$:

$$\begin{aligned} F^{(l)}(s,t) &:= \sum_{r=0}^{\infty} \mathbb{P}[Z(t, X^{(l)}) = r] s^r \\ &= \mathbb{E}[s^{Z(t,X^{(l)})}]. \end{aligned}$$

Next, we introduce the generating function associated to the distribution of the cumulative incidence $N_b(t, X)$ up to time t.

Definition C.3 (Generating function for $N_b(t, X)$)**.** For a given family history X, the generating function associated to the probability distribution of the cumulative incidence $N_b(t, X)$ up to time t is defined for $s \in \mathbb{C}$, $|s| < 1$ by:

$$\begin{aligned} F_b(s,t) &:= \sum_{r=0}^{\infty} \mathbb{P}[N_b(t, X) = r] s^r \\ &= \mathbb{E}[s^{N_b(t,X)}]. \end{aligned}$$

Finally we introduce the generating function $h(s,t)$ for the offspring process $\nu(t)$ as it appears naturally in the derivation of the integral equations below.

Definition C.4 (Generating function for $\nu(t)$)**.** The generating function of the offspring process $\nu(t)$ is defined for $s \in \mathbb{C}$, $|s| < 1$ by:

$$\begin{aligned} h(s,t) &:= \sum_{k=0}^{\infty} \mathbb{P}[\nu(t) = k] s^k \\ &= \sum_{k=0}^{\infty} q_t(k) s^k. \end{aligned}$$

Remark C.5. The trivial cases $\mathbb{P}[\nu(t) = 1] = 1$ (each infected individual always transforms into one unique infected individual, i.e. itself) and $\mathbb{P}[\nu(t) = 0] = 1$ are always excluded.

In our context, we need the first and second moments of the processes to be finite and thus, given the classical connection (proven in Lemma C.13 below) between the partial derivatives of the generating functions evaluated at $s = 1$ and the moments of the processes, we make the same following assumptions for all processes involved:

Assumption C.6 (Finite first and second partial derivatives)**.**

$$\sum_{r=0}^{\infty} r\mathbb{P}[Z(t,X)=r] < \infty, \qquad \sum_{r=0}^{\infty} r^2\mathbb{P}[Z(t,X)=r] < \infty, \tag{79}$$

$$\sum_{r=0}^{\infty} r\mathbb{P}[Z(t,X^{(l)})=r] < \infty, \qquad \sum_{r=0}^{\infty} r^2\mathbb{P}[Z(t,X^{(l)})=r] < \infty, \tag{80}$$

$$\sum_{r=0}^{\infty} r\mathbb{P}[N_b(t,X)=r] < \infty, \qquad \sum_{r=0}^{\infty} r^2\mathbb{P}[N_b(t,X)=r] < \infty, \tag{81}$$

$$\sum_{k=0}^{\infty} kq_t(k) < \infty, \qquad \sum_{n=0}^{\infty} k^2 q_t(k) < \infty. \tag{82}$$

Equivalently:

Assumption C.7 (Finite first and second moments)**.**

$$\mathbb{E}[Z(t,X)] < \infty, \qquad \mathbb{E}[Z^2(t,X)] < \infty,$$

$$\mathbb{E}[Z(t,X^{(l)})] < \infty, \qquad \mathbb{E}[Z^2(t,X^{(l)})] < \infty,$$

$$\mathbb{E}[N_b(t,X)] < \infty, \qquad \mathbb{E}[N_b^2(t,X)] < \infty,$$

$$\mathbb{E}[\nu(t)] < \infty, \qquad \mathbb{E}[\nu^2(t)] < \infty.$$

Remark C.8. The expectation for $\nu(t)$ is taken with respect to the probability measure defined on $\mathcal{I}$.

It is then straightforward to prove that:

Lemma C.9. *The first and second partial derivatives of $F(s,t), F^{(l)}(s,t), F_b(s,t)$ and $h(s,t)$ with respect to s in the interior of the unit circle $|s|<1$ and for $s=1$ are continuous.*

Proof. This is a direct consequence of Eqs. (79) to (82). Indeed, for instance, the first part of Eq. (79) implies that $\partial_s F(s,t)$ is a convergent power series and as such, it is continuous. Similarly the second part of Eq. (79) implies that the derivative $\partial_s^2 F(s,t)$ is a convergent power series and thus continuous. This logic holds for the other generating functions introduced here as their first and second partial derivatives follow similar assumptions. □

Noticing that all quantities are related to the same family history X and that $X^{(l)}$ is also directly related to X (it includes the same life-length random variables as X but the offspring process is observed at a shifted time), we will now use the simplifying notations:

Notation C.10 (Simplified notations for $Z(t, X^{(l)}), N_b(t, X^{(l)})$, and $I(t, X^{(l)})$)**.**

$$\begin{aligned} Z(t,l) &:= Z(t, X^{(l)}), & Z(t,0) &:= Z(t, X), \\ N_b(t,l) &:= N_b(t, X^{(l)}), & N_b(t,0) &:= N_b(t, X), \\ I(t,l) &:= I(t, X^{(l)}), & I(t,0) &:= I(t, X). \end{aligned}$$

Now, given that $s \in \mathbb{C}$ is an auxiliary variable in the expression of a generating function, we are also simplifying all notations for the generating functions. These functional notations will greatly simplify the writing down of the derivations of the integral equations, although it should be noted that the downside of these functional notations is that the variable s does not explicitly appear anymore in the expression of the generating functions.

Notation C.11 (Simplified notations for generating functions)**.**

$$\begin{aligned} F(t,l) &:= F^{(l)}(s,t), & F(t,0) &: = F(s,t), \\ F_b(t,l) &:= F_b^{(l)}(s,t), & F_b(t,0) &:= F_b(s,t). \end{aligned}$$

Finally we will also introduced functional notations to simplify the derivation of the renewal equation of the prevalence and incidence.

Notation C.12 (Simplified notations for first partial derivatives of generating functions)**.**

$$\begin{aligned} f(t,l) &:= \partial_s F^{(l)}(s,t)_{|s=1} = \partial_s F(t,l)_{|s=1}, & f(t,0) &:= \partial_s F(s,t)_{|s=1} = \partial_s F(t,0)_{|s=1}, \\ f_b(t,l) &:= \partial_s F_b^{(l)}(s,t)_{|s=1} = \partial_s F_b(t,l)_{|s=1}, & f_b(t,0) &:= \partial_s F_b(s,t)_{|s=1} = \partial_s F_b(t,0)_{|s=1}. \end{aligned}$$

Next we can recall the general result for a probability distribution with finite first two moments, which says that the first and second partial derivatives of its generating functions evaluated at $s = 1$ are equal to the mean and variance of the associated processes.

Lemma C.13 (First two partial derivatives at $s = 1$)**.**

Mean:	*Variance:*	
$\partial_s F(t,0)\|_{s=1} = \mathbb{E}[Z(t,0)],$	$\partial_s^2 F(t,0)\|_{s=1} = \mathbb{E}[Z^2(t,0)] - \mathbb{E}^2[Z(t,0)],$	(83)
$\partial_s F(t,l)\|_{s=1} = \mathbb{E}[Z(t,l)],$	$\partial_s^2 F(t,l)\|_{s=1} = \mathbb{E}[Z^2(t,l)] - \mathbb{E}^2[Z(t,l)],$	(84)
$\partial_s F_b(t,0)\|_{s=1} = \mathbb{E}[N_b(t,0)],$	$\partial_s^2 F_b(t,0)\|_{s=1} = \mathbb{E}[(N_b(t,0))^2] - \mathbb{E}^2[N_b(t,0)],$	(85)
$\partial_s h(s,t)\|_{s=1} = \mathbb{E}[\nu(t)],$	$\partial_s^2 h(s,t)\|_{s=1} = \mathbb{E}[(\nu^2(t)] - \mathbb{E}^2[\nu(t))].$	(86)

Proof. Given the exact similitude between the generating functions, we only prove the result for the generating function of $Z(t,0)$:

$$F(t,0) = \sum_{r=0}^{\infty} p_Z(r,t)s^r = \sum_{r=0}^{\infty} P[Z(t,0)=r]s^r,$$

As a power series with non-negative coefficients, $F(t,0)$ is absolutely convergent and infinitely differentiable with a radius of convergence $|s| < 1$. In what follows, $s = 1$ should always be interpreted as $s = 1^-$. Using Eq. (78), we can differentiate $F(t,0)$ with respect to s :

$$\partial_s F(t,0) = \sum_{r=1}^{\infty} r p_r(t) s^{r-1},$$

and by Abel's theorem, we then take the limit at the boundary $s = 1$ of the domain of convergence to evaluate the partial derivative at $s = 1$:

$$\partial_s F(t,0)_{|s=1} = \sum_{r=1}^{\infty} r p_r(t), \tag{87}$$

$$= \mathbb{E}[Z(t,0)]. \tag{88}$$

Thus the first part of Eq. (83) is now proven. Similarly, as the power series defining $F(t,0)$ is infinitely differentiable and converges absolutely, we can differentiate a second time:

$$\partial_s^2 F(t,0) = \sum_{r=2}^{\infty} r(r-1) p_r(t) s^{r-2}.$$

Again using Abel's theorem and Eq. (79), we obtain:

$$\begin{aligned}
\partial_s^2 F(t,0)|_{s=1^-} &= \sum_{r=2}^{\infty} r(r-1)p_r(t), && \text{(Abel's theorem)} \\
&= \sum_{r=2}^{\infty} r^2 p_r(t) - \sum_{r=2}^{\infty} r p_r(t), && \text{(Each series is absolutely convergent)} \\
&= \sum_{r=1}^{\infty} r^2 p_r(t) - \sum_{r=1}^{\infty} r p_r(t), && \text{(The term } r=1 \text{ simplifies)} \\
&= \mathbb{E}[Z^2(t,0)] - \mathbb{E}[Z(t,0)],
\end{aligned}$$

where in the last line we have recognised respectively the first and the second moments of the branching process $Z(t,X)$. And this proves the second part of Eq. (83). All the other equations (from Eq. (84) to Eq. (86)) can be proven in exactly the same way. □

Because the mean of the offspring process has a specific name in the study of infectious disease, namely the *reproductive number*, we highlight here the previous result related to the generating

function of the offspring distribution:

$$\partial_s h(s,t)_{|s=1} = \underbrace{\mathbb{E}[\nu(t)]}_{\text{Reproductive number}}, \tag{89}$$

and we formally define:

Definition C.14 (Time-varying reproductive number $R(t)$)**.** The expectation of the offspring process is called the *reproductive number* $R(t)$ at time t.

$$R(t) := \mathbb{E}[\nu(t)]. \tag{90}$$

And it follows from Eq. (89) that the time-varying reproductive number is also equal to the partial derivative of the offspring generating function evaluated at $s = 1$:

$$R(t) = \partial_s h(s,t)_{|s=1}.$$

C.2 Renewal equations for the generating functions

Given that the mean of a stochastic process is equal to the first partial derivative of its generating function evaluated at $s = 1$, in order to derive integral equations for a quantity like prevalence, which is nothing else than $\mathbb{E}[Z(t,0)]$, we first derive an integral equation for its generating function $F(t,0)$. Then all that will remain doing will be to take the partial derivative of this integral equation and evaluate it at $s = 1$ to obtain the equation for prevalence. However given that the equation for $F(t,0)$ will prove to be a particular case of the integral equation for $F(t,l)$, we first derive the integral equation for the generating function $F(t,l)$ and then will set $l = 0$ to deduce the equation followed by $F(t,0)$.

We will also proceed similarly with the other quantities such as cumulative incidence $N_b(t,l)$ and incidence $I(t,l)$.

Lemma C.15. *For any $t \geq 0$, any shift $l \geq 0$, and any $s \in \mathbb{C}$ such that $|s| < 1$:*

$$F(t,l) = s[1 - G(t)] + \int_0^t h[F(t-u, l+u), l+u]dg(u). \tag{91}$$

Proof. Recall that in the family history $X^{(l)}$, the offspring process is observed at a future time, shifted by $l \geq 0$. But only the observation time points for the offspring process are shifted by a quantity l in $X^{(l)}$: the absolute time scale between X and $X^{(l)}$ stays unchanged. The main *trick*,

applied in the literature, is to condition the expectation and split it into two parts, one where calendar time t is still before first transformation time L, and a second part where transformation of the initial infected individual has already occurred i.e. where $t > L$. Thus by definition of the generating function $F(t,l)$:

$$F(t,l) = \mathbb{E}[s^{Z(t,l)}] = \underbrace{\mathbb{E}[s^{Z(t,l)}\mathbf{1}_{\{L>t\}}]}_{A_l} + \underbrace{\mathbb{E}[s^{Z(t,l)}\mathbf{1}_{\{L\leq t\}}]}_{B_l}. \tag{92}$$

We first derive a simpler expression for the term A_l. When $t < L$, the first transformation has not yet occurred and it is also the case when considering the family history $X^{(l)}$ for which calendar time stays unchanged. So when $t < L$ there is still only one particle in the process associated to $X^{(l)}$ and for any shift $l \geq 0$, $Z(t,l) = 1$. We can write:

$$\begin{aligned} A_l &= \mathbb{E}[s^{Z(t,l)}\mathbf{1}_{\{L>t\}}], \\ &= \mathbb{E}[s\mathbf{1}_{\{L>t\}}], \\ &= s\mathbb{P}[L > t], \\ &= s[1 - G(t)]. \end{aligned}$$

Now we consider the second term B_l where we use the fact that $\nu(\cdot)$ is observed at time $t+l$ in the history $X^{(l)}$ and to which we apply the Tower property of conditional expectations. Note that in the third line below, we apply the rule commonly called *take out what is known*, where, in addition to not being random, the element that is taken out of the conditional expectation also needs to be bounded.

$$\begin{aligned} B_l &= \mathbb{E}[s^{Z(t,l)}\mathbf{1}_{L<t}], \\ &= \mathbb{E}[\mathbb{E}[s^{Z(t,l)}\mathbf{1}_{\{L\leq t\}}|L = u, \nu(L+l) = k]], && \text{(Tower property)} \\ &= \mathbb{E}[\mathbb{E}[s^{Z(t,l)}|L = u, \nu(L+l) = k]\mathbf{1}_{\{L\leq t\}}], && \text{(Take out what is known)} \\ &= \int_0^t \sum_{k=0}^{\infty} \mathbb{E}[s^{Z(t,l)}|L = u, \nu(L+l) = k] \underbrace{\mathbb{P}[\nu(L+l) = k|L = u]}_{\nu \text{ and } L \text{ ind.}} \mathbb{P}[L \in du], && \text{(Law of total pb.)} \\ &= \int_0^t \sum_{k=0}^{\infty} \mathbb{E}[s^{Z(t,l)}|L = u, \nu(L+l) = k] \underbrace{\mathbb{P}[\nu(u+l) = k]}_{\text{Def of } q.(\cdot)} \underbrace{\mathbb{P}[L \in du]}_{\text{Def. of } g(\cdot)}, \\ &= \int_0^t \sum_{k=0}^{\infty} \underbrace{\mathbb{E}[s^{Z(t,l)}|L = u, \nu(L+l) = k]}_{C_l} q_{u+l}(k)dg(u). \end{aligned} \tag{93}$$

We now work out the expression for the inner conditional expectation C_l, where we denote:

$$Z_i(t,l) := Z(t, X_i^{(l)}),$$

and as a consequence where the first-generation principle can be re-written respectively for X and for $X^{(l)}$:

$$\text{For } X: Z(t,0) = \sum_{i=1}^{\nu(L)} Z_i(t-L,0), \qquad \text{for any } X^{(l)}: Z(t,l) = \sum_{i=1}^{\nu(L)} Z_i(t-L,l).$$

$$\begin{aligned}
C_l &= \mathbb{E}[s^{Z(t,l)}|L=u,\nu(L+l)=k], \\
&= \mathbb{E}[\prod_{i=1}^{\nu(L+l)} s^{Z_i(t-L,l)}|L=u,\nu(L+l)=k], && \text{(First-gen. principle applied to } X^{(l)}) \\
&= \mathbb{E}[\prod_{i=1}^{k} s^{Z_i(t-u,l)}|L=u], && (X_i^{(l)} \text{ does not depend on } \nu(L+l)) \\
&= \prod_{i=1}^{k} \mathbb{E}[s^{Z_i(t-u,l)}|L=u], && ((X_i^{(l)}|L=u)_i \text{ are independent}) \\
&= \prod_{i=1}^{k} \mathbb{E}[s^{Z(t-u,l+u)}], && (\text{By def. } (X_i^{(l)}|L=u)_i \sim X^{(l+u)}) \\
&= \Big(\mathbb{E}[s^{Z(t-u,l+u)}]\Big)^k, \\
&= \Big(F(t-u,l+u)\Big)^k. && (\text{By def. of } F(\cdot,\cdot)) \qquad (94)
\end{aligned}$$

Replacing the expression for C_l from Eq. (94) in Eq. (93), we obtain:

$$F(t,l) = s[1-G(t)] + \int_0^t \underbrace{\sum_{k=0}^{\infty} q_{l+u}(k)\Big(F(t-u,l+u)\Big)^k}_{\text{We recognise } h[F(t-u,l+u),l+u]} dg(u), \text{ i.e.:}$$

$$F(t,l) = s[1-G(t)] + \int_0^t h[F(t-u,l+u),l+u]dg(u). \qquad (95)$$

□

It is important to note that Eq. (91) is actually not a proper renewal equation for $F(t,l)$ such as the one derived in [20] when the offspring process is time invariant. Indeed Eq. (91) simply links the generating function $F(t,l)$ of the branching process $Z(t,l)$ at time t to the generating function $F(t-u,l+u)$ of the $(l+u)$-shifted process $Z(t-u,l+u)$ at time $t-u$. We will see in section E how to transform this equation into a renewal-like equation.

But nonetheless, we do end up with a system of equations, valid for any $l \geq 0$. The process of interest is of course the non-shifted one, i.e. $Z(t,0)$ associated to the family history X. Having

said this, in the context of a time-varying offspring distribution, it is the whole system of integral equations given by Eq. (91) which allows to solve a specific integral equation for $F(t,l)$ in a given case such as $l=0$. Here we re-write more specifically Eq. (91) for $l=0$, which again is the main generating function of interest corresponding to the original branching process $Z(t,0)$:

Lemma C.16. *For any $t \geq 0$, and $s \in \mathbb{C}$ such that $|s| < 1$:*

$$F(t,0) = s[1-G(t)] + \int_0^t h[F(t-u,0),u]dg(u). \tag{96}$$

Finally it can be shown in a very similar way that $F_b(t,l)$, the generating function associated to the process of cumulative incidence, is solution to the following integral equation:

Lemma C.17. *For any $t \geq 0$, and $s \in \mathbb{C}$ such that $|s| < 1$:*

$$F_b(t,l) = s[1-G(t)] + s\int_0^t h[F_b(t-u,l+u),l+u]g(u)du. \tag{97}$$

Proof. In order to obtain an integral equation for the generating function of the shifted cumulative incidence $N_b(t,l)$, we again condition on the time of first transformation L.

$$F_b(t,l) = \mathbb{E}[s^{N_b(t,l)}] = \underbrace{\mathbb{E}[s^{N_b(t,l)}\mathbf{1}_{\{L>t\}}]}_{A_b} + \underbrace{\mathbb{E}[s^{N_b(t,l)}\mathbf{1}_{\{L\leq t\}}]}_{B_b}.$$

We first derive an expression for the term A_b, where we notice that when $t < L$, the first transformation has not yet occurred and thus at $t < L$, the cumulative number of births $N_b(t,l)$, no matter which family history $X^{(l)}$ we are observing, is necessarily equal to 1 with probability 1. Thus the generating function $F_b(t,l)$ simplifies to $F_b(t,l) = s$ and we obtain:

$$\begin{aligned} A_b &= \mathbb{E}[s^{N_b(t,l)}\mathbf{1}_{\{L>t\}}], \\ &= \mathbb{E}[s\mathbf{1}_{\{L>t\}}], \\ &= s\mathbb{P}[L>t], \\ &= s[1-G(t)]. \end{aligned}$$

Note that $s[1-G(t)]$ is the probability for the initial particle of surviving after time t. Next we

consider the second term B_b and follow the same procedure as for B_l in the previous proof:

$$
\begin{aligned}
B_b &= \mathbb{E}[s^{N_b(t,l)}\mathbf{1}_{\{L\le t\}}], \\
&= \mathbb{E}[\mathbb{E}[s^{N_b(t,l)}\mathbf{1}_{\{L\le t\}}|L=u,\nu(L+l)=k]], && \text{(Tower property)} \\
&= \mathbb{E}[\mathbb{E}[s^{N_b(t,l)}|L=u,\nu(L+l)=k]\mathbf{1}_{\{L\le t\}}], && \text{(Take out what is known)} \\
&= \int_0^t \sum_{k=0}^{\infty} \mathbb{E}[[s^{N_b(t,l)}|L=u,\nu(u+l)=k]\mathbb{P}[\nu(L+l)=k|L=u]\mathbb{P}[L\in du], && \text{(By def of } \mathbb{E} \text{ and law of total pb)} \\
&= \int_0^t \sum_{k=0}^{\infty} \mathbb{E}[[s^{N_b(t,l)}|L=u,\nu(L+l)=k]\mathbb{P}[\nu(u+l)=k]\mathbb{P}[L\in du], && (L \text{ and } \nu(\cdot) \text{ are independent}) \\
&= \int_0^t \sum_{k=0}^{\infty} \underbrace{\mathbb{E}[s^{N_b(t,l)}|L=u,\nu(L+l)=k]}_{C_b} q_{u+l}(k)g(du). && (98)
\end{aligned}
$$

We now work out the expression for the inner conditional expectation C_b, where we denote:

$$N_{b,i}(t,l) := N_b(t, X_i^{(l)}),$$

and as a consequence, where the first-generation principle for cumulative incidence can be re-written:

$$
\begin{aligned}
\text{For X: } N_b(t,0) &= \sum_{i=1}^{\nu(L)} N_{b,i}(t-L,0), \\
\text{For } X^{(l)} : N_b(t,l) &= \sum_{i=1}^{\nu(L)} N_{b,i}(t-L,l).
\end{aligned}
$$

$$
\begin{aligned}
C_b &= \mathbb{E}[s^{N_b(t,l)}|L=u,\nu(L+l)=k], \\
&= \mathbb{E}[\prod_{i=1}^{\nu(L+l)} s^{1+N_{b,i}(t-L,l)}|L=u,\nu(L+l)=k], && \text{(By first gen. principle for } N_b(t,l)) \\
&= s\mathbb{E}[\prod_{i=1}^{k} s^{N_{b,i}(t-u,l)}|L=u], && (X_i^{(l)} \text{ does not depend on } \nu(L)) \\
&= s\prod_{i=1}^{k} \mathbb{E}[s^{N_{b,i}(t-u,l)}|L=u], && ((X_i^{(l)}|L=u)_i \text{ are independent}) \\
&= s\prod_{i=1}^{k} \mathbb{E}[s^{N_b(t-u,l+u)}] && (X_i^{(l)}|L=u) \sim X^{(l+u)}), \\
&= s\Big(\mathbb{E}[s^{N_b(t-u,l+u)}]\Big)^k, \\
&= s\Big([F_b(t-u,l+u)]\Big)^k && \text{(By def. of } F_b(\cdot,\cdot)). \quad (99)
\end{aligned}
$$

Finally we combine Eq. (99) above with Eq. (98):

$$\begin{aligned} B_b &= \int_0^t \sum_{k=0}^{\infty} C_b q_{u+l}(k) g(du), \\ &= s \int_0^t \underbrace{\sum_{k=0}^{\infty} q_{u+l}(k)[F_b(t-u, u+l)]^k}_{\text{We recognise } h[f_b(t-u,u+l),u+l]} g(du), \\ &= s \int_0^t h[F_b(t-u, u+l), u+l] dg(u). \end{aligned}$$

We thus obtain:

$$F_b(t,l) = s[1-G(t)] + s \int_0^t h[F_b(t-u, u+l), u+l] dg(u). \tag{100}$$

□

We can now use this general equation and apply it to the original branching process with a 0-shift offspring by setting $l = 0$ in Eq. (97):

Lemma C.18. *For any $t \geq 0$, and $s \in \mathbb{C}$ such that $|s| < 1$:*

$$F_b(t,0) = s[1-G(t)] + s \int_0^t h[F_b(t-u, u), u] dg(u).$$

D Renewal equations for prevalence, cumulative incidence and incidence of the branching process

D.1 Renewal equation for prevalence

To derive an integral equation for the prevalence of the original branching process, i.e. for $\mathbb{E}[Z(t,0)]$, we rely on the relationships between the first moment of $Z(t,0)$ and the partial derivative of $F(t,0)$ evaluated at $s = 1$ given in the first part of Eq. (84), which we recall here:

$$\mathbb{E}[Z(t,0)] = \partial_s F(t,0)|_{s=1},$$

and using functional notation, this can be re-written:

$$\mathbb{E}[Z(t,0)] = f(t,0).$$

As proceeded with the derivation of the renewal-type equations for the generating functions in the previous section, here we also first prove the more general result related to the shifted process and

we will then obtain a renewal equation for the prevalence of the original (non-shifted) process by simply setting $l = 0$ in the general integral equation.

Lemma D.1. *Prevalence of the l-shifted process* $\mathbb{E}[Z(t,l)]$*, also denoted* $f(t,l)$*, is solution to the renewal-type equation given by:*

$$f(t,l) = 1 - G(t) + \int_0^t R(l+u) f(t-u, l+u) g(u) du. \tag{101}$$

Proof. We recall the integral equation for the associated generating function $F(t,l)$ as per Eq. (91):

$$F(t,l) = s[1 - G(t)] + \int_0^t h[F(t-u, l+u), l+u] dg(u).$$

Being a generating function, $F(t,l)$ is a complex power series in s with non-negative coefficients, and as such, is infinitely differentiable within its radius of convergence, i.e. within the interior of the unit circle for $|s| < 1$. It is also absolutely convergent and consequently its first two partial derivatives $\partial_s F(t,l)$ and $\partial_s^2 F(t,l))$ are continuous for $|s| \leq 1$. We can take the partial derivative on both sides of Eq. (91):

$$\partial_s F(t,l) = 1 - G(t) + \int_0^t \partial_s F(t-u, l+u)\, \partial_s h[F(t-u, l+u), l+u] g(u) du. \tag{102}$$

To obtain an integral equation for $f(t,l)$, it suffices to evaluate Eq. (102) at $s = 1^-$ (denoted here $s = 1$). By continuity at $s = 1^-$ of all the functions involved in Eq. (102), the limit can be passed inside the integral :

$$\partial_s F(t,l)_{|s=1} = 1 - G(t) + \int_0^t \partial_s F(t-u, l+u)_{|s=1} \partial_s h[F(t-u, l+u), l+u]_{|s=1} g(u) du. \tag{103}$$

Now using our functional notations at time point $t - u$, we write:

$$\partial_s F(t-u, l+u)_{|s=1} = f(t-u, l+u). \tag{104}$$

Additionally, by Abel's theorem:

$$\lim_{s \to 1^-} F(t-u, l+u) = \lim_{s \to 1^-} \sum_{r=0}^{\infty} \mathbb{P}[Z(t-u, l+u) = r] s^r = \sum_{r=0}^{\infty} \mathbb{P}[Z(t-u, l+u) = r] = 1. \tag{105}$$

And by continuity of $\partial_s h(s,t)$ as $s \to 1^-$:

$$\partial_s h[F(t-u, l+u), l+u]_{|s=1} = \partial_s h(s, l+u)_{|s=1}. \tag{106}$$

Finally using the general fact that $\partial_s h[s, \cdot]_{|s=1} = R(\cdot)$, we obtain:

$$\partial_s h[F(t-u, l+u), l+u]_{|s=1} = R(l+u). \tag{107}$$

Now going back to Eq. (103), we replace using Eq. (104) and Eq. (107) to obtain:

$$\partial_s F(t,l)_{|s=1} = 1 - G(t) + \int_0^t R(l+u) f(t-u, l+u) g(u) du,$$

i.e.

$$f(t,l) = 1 - G(t) + \int_0^t R(l+u) f(t-u, l+u) g(u) du,$$

where the term $1-G(t)$ is the survival probability of the first infected individual, sometimes called *index infection.* □

It now suffices to set $l = 0$ in the above equation to derive the renewal-type equation for the prevalence of the original branching process.

Lemma D.2 (Renewal equation for prevalence)**.** *The prevalence of the branching process,* $\mathbb{E}[Z(t,0)] = f(t,0)$*, is solution to the following renewal equation:*

$$f(t,0) = 1 - G(t) + \int_0^t R(u) f(t-u, u) g(u) du. \tag{108}$$

D.2 Renewal equation for cumulative incidence

We recall that incidence $I(t,0)$ at time t can be defined in Eq. (67) as the point process intensity of the cumulative incidence:

$$I(t,0) = \lim_{\delta t \to 0} \frac{\mathbb{E}[N_b(t,0)] - \mathbb{E}[N_b(t-\delta t, 0)]}{\delta t}. \tag{109}$$

Thus in order to obtain an integral equation for $I(t,0)$, we first could derive an integral equation for the expectation of the cumulative incidence $\mathbb{E}[N_b(t,0)]$, and then study the limit above.

Here again, we will first derive the more general renewal equation for $f_b(t,l)$ and then set $l = 0$ to obtain the renewal equation for the expectation of the cumulative incidence of the original branching process. Similarly to above, we rely on the relationship between the first moment of $N_b(t,l)$ and the first partial derivative of its generating function evaluated at $s = 1$.

$$f_b(t,l) = \mathbb{E}[N_b(t,l)] = \partial_s F_b(t,l)|_{s=1}.$$

We now have all the elements to show that the expectation of the cumulative incidence with shifted offspring process follows an integral equation.

Lemma D.3 (Integral equation for the expectation of the cumulative incidence with shifted offspring)**.** *$f_b(t,l) = \mathbb{E}[N_b(t,l)]$ is a solution to the integral equation given by:*

$$f_b(t,l) = 1 + \int_0^t R(l+u) f_b(t-u, l+u) g(u) du. \tag{110}$$

Proof. Our starting point is Eq. (97) for $F_b(t,l)$:

$$F_b(t,l) = s[1-G(t)] + s\int_0^t h[F_b(t-u, l+u), l+u] g(u) du. \tag{111}$$

$F_b(t,l)$ is a power series with non-negative coefficients and as such is infinitely differentiable within its radius of convergence, i.e. within the interior of the unit circle for $|s| < 1$. It is also absolutely convergent, so its the first two partial derivatives $\partial_s F_b(t,l)$ and $\partial_s^2 F_b(t,l)$ are continuous for $|s| \leq 1$. Consequently we can take the first partial derivative on both sides of Eq. (97) and write:

$$\begin{aligned}\partial_s F_b(t,l) = 1 - G(t) &+ \int_0^t h[F_b(t-u, l+u), l+u] g(u) du \\ &+ \int_0^t \partial_s F_b(t-u, l+u)\, \partial_s h[F_b(t-u, l+u), l+u] g(u) du.\end{aligned}$$

To obtain an integral equation for $f_b(t,l)$, it suffices to evaluate the equation above at $s = 1^-$ (denoted here $s = 1$). By continuity at $s = 1^-$ of all the functions involved:

$$\begin{aligned}\partial_s F_b(t,l)_{|s=1} = 1 - G(t) &+ \int_0^t h[F_b(t-u, l+u), l+u]_{|s=1} g(u) du \\ &+ \int_0^t \partial_s F_b(t-u, l+u)_{|s=1} \partial_s h[F_b(t-u, l+u), l+u)]_{|s=1} g(u) du,\end{aligned} \tag{112}$$

Now, using our functional notation at time point $t-u$, we can write:

$$\partial_s F_b(t-u, l+u)_{|s=1} = f_b(t-u, l+u). \tag{113}$$

Additionally, by Abel's theorem:

$$\lim_{s\to 1^-} F_b(t-u, l+u) = \lim_{s\to 1^-} \sum_{r=0}^{\infty} \mathbb{P}[N_b(t-u, l+u) = r] s^r = \sum_{r=0}^{\infty} \mathbb{P}[N_b(t-u, l+u) = r] = 1.$$

And by continuity of $\partial_s h(s,t)$ as $s \to 1^-$:

$$\partial_s h[F_b(t-u, l+u), l+u]_{|s=1} = \partial_s h(s, l+u)_{|s=1}.$$

Finally using the fact that $\partial_s h(s,\cdot) = R(\cdot)$ (Eq. (89)), we obtain:

$$\partial_s h[F_b(t-u, l+u), l+u]_{|s=1} = R(l+u).$$

Now going back to Eq. 112, and noticing again by Abel's theorem applied to $h(s,t)$ that $h(s,u)_{|s=1} = 1$, we obtain:

$$\begin{aligned}
\partial_s F_b(t,l)_{|s=1} &= 1 - G(t) + \int_0^t \underbrace{h[F_b(t-u,l+u)_{|s=1}, l+u]}_{=h(1,l+u)=1} g(u)du \\
&\quad + \int_0^t \underbrace{\partial_s F_b(t-u,l+u)_{|s=1}}_{f_b(t-u,l+u)} \underbrace{\partial_s h[F_b(t-u,l+u),y)_{|s=1}}_{R(l+u)} g(u)du, \\
&= 1 - G(t) + \underbrace{\int_0^t g(u)du}_{G(t)} + \int_0^t R(l+u) f_b(t-u,l+u) g(u)du, \\
&= 1 + \int_0^t R(l+u) f_b(t-u,l+u) g(u)du,
\end{aligned}$$

i.e.:

$$f_b(t,l) = 1 + \int_0^t R(l+u) f_b(t-u,l+u) g(u)du.$$

□

Finally it now suffices again to set $l = 0$ in the above equation to derive the renewal-type equation for the expectation of the cumulative incidence of the original branching process.

Lemma D.4 (Integral equation for the expectation of the cumulative incidence)**.** $f_b(t,0) = \mathbb{E}[N_b(t,0)]$ *is solution to the integral equation given by:*

$$f_b(t,0) = 1 + \int_0^t R(u) f_b(t-u,u) g(u)du. \tag{114}$$

D.3 Renewal equation for incidence

Having derived a renewal-type equation for the cumulative incidence, we now are in a position to derive a renewal-type equation for the incidence point process, using the fact that incidence can be seen as the intensity of the expectation of the cumulative incidence. We first derive the general equation for $I(t,l)$ as follows.

Lemma D.5 (Renewal equation for incidence with shifted offspring)**.** *The incidence $I(t,l)$ for an l-shifted age-dependent branching process with time-varying reproduction number $R(t)$ is solution to the integral equation given by:*

$$I(t,l) = R(t+l)g(t) + \int_0^t R(l+u) I(t-u,l+u) g(u)du. \tag{115}$$

Proof. Our starting point is the integral equation for $\mathbb{E}[N_b(t,l)] = f_b(t,l)$, in Eq. (110):

$$f_b(t,l) = 1 + \int_0^t R(l+u) f_b(t-u, l+u) g(u) du, \tag{116}$$

which we can re-write at time $t - \delta t$:

$$f_b(t-\delta t, l) = 1 + \int_0^{t-\delta t} R(l+u) f_b(t-\delta t-u, l+u) g(u) du. \tag{117}$$

Next we take the difference between Eq. (116) and Eq. (117):

$$\begin{aligned}
f_b(t,l) - f_b(t-\delta t, l) &= \int_0^t R(l+u) f_b(t-u, l+u) g(u) du \\
&\quad - \int_0^{t-\delta t} R(l+u) f_b(t-\delta t-u, l+u) g(u) du, \\
&= \int_0^{t-\delta t} R(l+u) f_b(t-u, l+u) g(u) du + \int_{t-\delta t}^t R(l+u) f_b(t-u, l+u) g(u) du \\
&\quad - \int_0^{t-\delta t} R(l+u) f_b(t-\delta t-u, l+u) g(u) du, \\
&= \int_0^{t-\delta t} R(l+u) [f_b(t-u, l+u) - f_b(t-\delta t-u, l+u)] g(u) du \\
&\quad + \int_{t-\delta t}^t R(l+u) f_b(t-u, l+u) g(u) du.
\end{aligned}$$

In order to obtain an integral equation for $I(t,l)$, we now divide by δt,

$$\begin{aligned}
\frac{f_b(t,l) - f_b(t-\delta t, l)}{\delta t} &= \int_0^{t-\delta t} R(l+u) \frac{f_b(t-u, l+u) - f_b(t-\delta t-u, l+u)}{\delta t} g(u) du \\
&\quad + \frac{1}{\delta t} \int_{t-\delta t}^t R(l+u) f_b(t-u, l+u) g(u) du.
\end{aligned} \tag{118}$$

(noting here that we assumed we could divide by δt *inside the integral*). Again we assume that both terms on the right hand side of Eq. (118) converge absolutely so we take the limit of each term separately as $\delta t \to 0$:

$$\begin{aligned}
&\lim_{\delta t \to 0} \frac{f_b(t,l) - f_b(t-\delta t, l)}{\delta t} \\
&\qquad = \underbrace{\lim_{\delta t \to 0} \int_0^{t-\delta t} R(l+u) \frac{f_b(t-u, l+u) - f_b(t-\delta t-u, l+u)}{\delta t} g(u) du}_{A_1} \\
&\qquad + \underbrace{\lim_{\delta t \to 0} \frac{1}{\delta t} \int_{t-\delta t}^t R(l+u) f_b(t-u, l+u) g(u) du}_{A_2}.
\end{aligned}$$

The term A_1 can be re-written, passing the limit inside the integral:

$$\begin{aligned}
A_1 &= \int_0^t R(l+u) \lim_{\delta t \to 0} \frac{[f_b(t-u, l+u) - f_b(t-\delta t-u, l+u)]}{\delta t} g(u) du, \\
&= \int_0^t R(l+u) I(t-u, l+u) g(u) du.
\end{aligned}$$

Finally the term A_2 gives:

$$A_2 = \lim_{\delta t \to 0} \frac{1}{\delta t} \int_{t-\delta t}^{t} R(l+u) f_b(t-u, l+u) g(u) du = \lim_{\delta t \to 0} \frac{1}{\delta t} \int_{t-\delta t}^{t} a(u) du,$$

where $a(u) = R(l+u) f_b(t-u, l+u) g(u)$. So:

$$\begin{aligned} A_2 &= \lim_{\delta t \to 0} \frac{1}{\delta t} \int_{t-\delta t}^{t} a(u) du = a(t), \\ &= R(l+t) f_b(0, l+t) g(t), \\ &= R(l+t) g(t), \end{aligned}$$

because $f_b(0, t+l) = \mathbb{E}[N_b(0, t+l)] = 1$. We thus obtain:

$$I(t,l) = R(l+t)g(t) + \int_0^t R(l+u) I(t-u, l+u) g(u) du.$$

□

Again, it is now straight-forward to set $l = 0$ above and derive the renewal equation followed by the incidence process $I(t, 0)$:

Lemma D.6 (Renewal equation for incidence)**.** *The incidence $I(t, 0)$ of an age-dependent branching process with time-varying reproduction number $R(t)$ is a solution to the integral equation given by:*

$$I(t,0) = R(t)g(t) + \int_0^t R(u) I(t-u, u) g(u) du.$$

E Discretisation procedure to simulate renewal-type equations

In order to simulate and solve numerically the renewal-type equations obtained above for prevalence, cumulative incidence and incidence, it has proven necessary to put in place a discretisation procedure allowing to approximate the quantities in question using the integral equations.

So first here we observe that the generic form of the integral equations which we need to solve numerically is:

$$f(t,l) = h(t,l) + \int_0^t R(l+u) f(t-u, l+u) g(u) du, \tag{119}$$

where the generic function $h(t,l)$ translated respectively into $h(t,l) = 1 - G(t)$ for prevalence, then into $h(t,l) = 1$ for cumulative incidence and finally into $h(t,l) = R(t+l)g(t)$ for incidence. The first step is to transform this generic integral equation into a renewal-like equation. For this purpose we write:

$$l := c - t,$$

where $c \in \mathbb{R}$ is a newly introduced parameter which will highlight the renewal form of this equation. Indeed, we introduce the function $f_c(t)$ such as:

$$f_c(t) := f(t, c-t),$$

so we can re-write Eq. (119):

$$f_c(t) = \underbrace{h(t,c-t)}_{h_c(t)} + \int_0^t R(c-(t-u))f(t-u, c-(t-u))g(u)du, \text{ i.e.:}$$

$$f_c(t) = h_c(t) + \int_0^\infty R(c-(t-u))f_c(t-u)g(u)du. \tag{120}$$

Eq (120) highlights the fact that the integral equations derived earlier in a context of time varying offspring distribution can still be viewed as renewal-like equations. Next, in order to discretise Eq (120), we want to solve approximately for $f_c(i\Delta)$, for any $i = 0, 1, \ldots, n$ and for a step size $\Delta > 0$. We call the approximate numerical solution $\hat{f}_c(i\Delta)$. To start with, clearly we should set $\hat{f}_c(0) := h_c(0)$. Then for $i \geq 1$ and $0 \leq j \leq i$, we define recursively:

$$\hat{f}_c(i\Delta) := h_c(i\Delta) + \sum_{j=1}^{i} R(c-(i-j)\Delta)\hat{f}_c((i-j)\Delta)g((j-1)\Delta)\Delta. \tag{121}$$

The objective is to vectorise these equations in order to decrease computation time. So next, letting n be the number of time steps, and choosing n values for c, namely, $c_1, \ldots, c_n \in \mathbb{R}$, for $0 \leq i \leq n$, we define the three n-dimensional vectors:

$$\hat{\boldsymbol{f}}(i\Delta) := \begin{bmatrix} \hat{f}_{c_1}(i\Delta) \\ \vdots \\ \hat{f}_{c_n}(i\Delta) \end{bmatrix} \in \mathbb{R}^n, \ \boldsymbol{h}_c(i\Delta) := \begin{bmatrix} h_{c_1}(i\Delta) \\ \vdots \\ h_{c_n}(i\Delta) \end{bmatrix} \in \mathbb{R}^n, \qquad \boldsymbol{R}(i\Delta) := \begin{bmatrix} R(c_1 - i\Delta) \\ \vdots \\ R(c_n - i\Delta) \end{bmatrix} \in \mathbb{R}^n.$$

We can then rewrite Equ. (121) using this vectorisation:

$$\hat{\boldsymbol{f}}(i\Delta) := \boldsymbol{h}_c(i\Delta) + \sum_{j=1}^{i} g[(j-1)\Delta]\boldsymbol{R}[(i-j)\Delta] \odot \hat{\boldsymbol{f}}[(i-j)\Delta]\Delta, \tag{122}$$

where $\odot$ stands for element-wise vectors multiplication (and below matrices). Vectorising one dimension further, we then define the $n \times (i+1)$ matrices for $i = 1, \ldots, n$:

$$\begin{aligned}
\boldsymbol{A}_i &:= \Big[\boldsymbol{R}(0) \quad \boldsymbol{R}(\Delta) \ldots \quad \boldsymbol{R}(i\Delta)\Big] \in \mathbb{R}^{n\times(i+1)}, \\
\boldsymbol{B}_i &:= \Big[\hat{\boldsymbol{f}}(0) \quad \hat{\boldsymbol{f}}(\Delta) \quad \ldots \quad \hat{\boldsymbol{f}}(i\Delta)\Big] \in \mathbb{R}^{n\times(i+1)}, \\
\boldsymbol{G}_i &:= \Delta \begin{bmatrix} g(i\Delta) & \ldots & g(\Delta) & g(0) \\ \vdots & & & \\ g(i\Delta) & \ldots & g(\Delta) & g(0) \end{bmatrix} \in \mathbb{R}^{n\times(i+1)},
\end{aligned}$$

such that Eq.(122) can be re-written as a product of matrices for $i = 1, \ldots, n$:

$$\hat{\boldsymbol{f}}(i\Delta) = \boldsymbol{h}_c(i\Delta) + \text{RowSum}(\boldsymbol{G}_{i-1} \odot \boldsymbol{A}_{i-1} \odot \boldsymbol{B}_{i-1}).$$

The matrix $\boldsymbol{B}_i$ is built up recursively using:

$$\boldsymbol{B}_i = \text{Concatenate}\Big(\boldsymbol{B}_{i-1}, \boldsymbol{h}_c(i\Delta) + \text{RowSum}(\boldsymbol{G}_{i-1} \odot \boldsymbol{A}_{i-1} \odot \boldsymbol{B}_{i-1})\Big) \in \mathbb{R}^{n\times(i+1)}.$$

Finally it suffices to notice that by setting $c_i := i\Delta$, and extracting for $i = 1, \ldots, n$ the terms $\hat{f}_{i\Delta}(i\Delta)$, from the diagonal of $\boldsymbol{B}_n$, we obtain the approximate solution of the renewal-type equation we were looking for:

$$\hat{f}_{i\Delta}(i\Delta) \approx f_{i\Delta}(i\Delta) = f(i\Delta, i\Delta - i\Delta) = f(i\Delta, 0).$$

F Consistency of renewal equations with incidence/prevalence formula

We wish to show that the renewal equations derived earlier for prevalence $f(t,l)$, incidence $I(t,l)$ and cumulative incidence $N_b(t,l)$ are consistent with the *back calculating formula* commonly used in epidemiology applications[12], i.e.:

$$f(t,l) = [(1-G) * I(\cdot, l)](t). \tag{123}$$

To check for consistency in Eq. (123), we will need the following intermediary result which we briefly prove:

Lemma F.1. *Let $f_1(\cdot,\cdot), f_2(\cdot)$, and $f_3(\cdot)$ be non-negative, Lebesgue measurable functions. Then:*

$$\int_0^t \Big(\int_0^u f_1(u-s, s) f_2(s) ds \Big) f_3(t-u) du = \int_0^t (f_1(\cdot, s) * f_3)(t-s) f_2(s) ds. \tag{124}$$

Proof. We prove Eq. (124):

$$\begin{aligned}\int_0^t \Big(\int_0^u f_1(u-s,s) f_2(s) ds \Big) f_3(t-u) du &= \int_0^t \int_0^u f_1(u-s,s) f_2(s) f_3(t-u) ds du, \\ &= \int_0^t \int_0^t f_1(u-s,s) f_3(t-u) f_2(s) \mathbf{1}_{s \le u} ds du, \\ &= \int_0^t \int_0^t f_1(u-s,s) f_3(t-u) \mathbf{1}_{s \le u} du f_2(s) ds.\end{aligned}$$

Then we substitute $v := u - s$ and note that the constraint $s \le u$ translates into $v \ge 0$, while $t \ge u$ translated into $t - s \ge v$. Thus:

$$\begin{aligned}\int_0^t \int_0^t f_1(u-s,s) f_3(t-u) \mathbf{1}_{s \le u} du f_2(s) ds. &= \int_0^t \int_0^{t-s} f_1(v,s) f_3(t-s-v) dv f_2(s) ds \\ &= \int_0^t [f_1(\cdot, s) * f_3](t-s) f_2(s) ds.\end{aligned}$$

□

Now we check that Eq. (123) hold by first recalling the integral equation derived for incidence:

$$I(t,l) = R(l+t)g(t) + \int_0^t R(l+u) I(t-u, l+u) g(u) du.$$

We observe that this equation should be more precisely re-written:

$$I(t,l) = R(l+t)g(t) + \int_0^{t^-} R(l+u) I(t-u, l+u) g(u) du. \tag{125}$$

Next, using the Delta function, we note that the term $R(l+t)g(t)$ can be expressed as an integral over the atom $\{t\}$:

$$\begin{aligned}R(l+t)g(t) &= \int_{\{t\}} R(l+u)\delta_0(t-u) g(u) du, && \text{(By def. of } \delta_0(\cdot)) \\ &= \int_{\{t\}} R(l+u) I(t-u, l+u) g(u) du. && \text{(Because for any } l : I(0, l+u) = \delta_0(0)\end{aligned}$$

Thus the term $R(l+t)g(t)$ can be subsumed inside the integral leaving out a Dirac delta term. Taking this observation into account, we convolve both sides of Eq. (125) with $1 - G$ on the left to obtain:

$$((1-G) * I(\cdot, l))(t) = (1-G) * \delta_0(t) + \left((1-G) * \int_0^{\cdot} I(\cdot - s, l+s) R(l+s) g(s) ds \right)(t). \tag{126}$$

Using Lemma F.1, we set $f_1(t,s) := I(t, l+s)$, $f_2(t) := R(l+t)g(t)$ and $f_3(t) := 1 - G(t)$, and

obtain:

$$\left((1-G) * \int_0^{\cdot} I(\cdot - s, l+s)R(l+s)g(s)ds\right)(t) = \int_0^t \Big(\int_0^u I(u-s, l+s)R(l+s)g(s)ds\Big)(1-G)(t-u)du,$$
$$= \int_0^t (I(\cdot, s) * (1-G))(t-s)R(l+s)g(s)ds,$$

where in the first equality, we simply applied the definition of convolution. Now regrouping this result in Eq. (126), we finally obtain:

$$\underbrace{((1-G) * I(\cdot, l))(t)}_{w(t,l)} = (1-G)(t) + \int_0^t \underbrace{((1-G) * I(\cdot, s))(t-s)}_{w(t-s,l)} R(l+s)g(s)ds.$$

We observe that the function w(t,l) follows exactly the same integral equation as the prevalence $f(t,l)$ in Eq. (101). By unicity of the solution proven in [23], we deduce that $((1-G) * I(\cdot, l))(t)$ is equal to the prevalence.