Mathematical and Statistical Methods in Epidemiology

Matthew Penn

A thesis submitted for the degree of
Doctor of Philosophy

University of Oxford
Department of Statistics

Supervisors:

Professor Christl Donnelly (University of Oxford)
Professor Samir Bhatt (University of Copenhagen)

Submission Date: Hilary Term 2024

Acknowledgements

There have been far too many people who have suffered through my ramblings about proofs, bugs in my code, trees, diseases, and (aren't they cool?) proofs throughout my studies. It would be genuinely impossible to mention you all (not least because I don't know the name of the anaesthetist who I monologued at while returning to consciousness), but below are the people who have been most important (and therefore most patient) to whom I would like to give my thanks.

To **Christl**, thank you for taking in a far-too-theoretical mathematician and helping them actually do some epidemiology rather than getting lost in a cave of ever-increasing abstraction and insanity. Thank you for every mind-numbingly dull lemma that you've read; for every set of in-depth corrections on all my poorly proof-read papers; for the close-knit community you forged in our group; and for your constant help in finding opportunities, both DPhil- and football-related. And, of course, thank you for all the hyphens.

To **Sam**, thank you for being the only person who wants to talk about Maths more than me. Thank you for always getting excited when I derived something new, even though I'd generally discover it was wrong an hour later; for all your enthusiasm in finding new problems in epidemiology, phylogenetics or football to look at; for all the stupid (but fun) ideas we've discussed and then discarded; and (perhaps most importantly) for always doing the coding when there's been something to do in R.

On that note, the programming language *R* gets an anti-acknowledgement. No one thinks that *c* stands for *vector*... but I do know that it's next to *v* on the keyboard. Suspicious.

To my amazing co-authors, thank you for bearing with all my grammar mistakes and bad writing to make our papers so much more appealing. In particular, **Neil**, thank you for all the hours you've spent turning my atrocious code (or worse, an unintelligible draft of my derivations) into wonderful, efficient packages. **Verity**, thank you for all the hundreds of times you put up with my code crashing (you should have worked with Neil instead...) and for the hellish month you spent fighting to upload all our files. **David**, thank you for actually knowing things about biology and for finding us so many useful examples to put in the papers.

Last, but not least of my co-authors, **Joe**, thank you for all the time you've spent poring over my

proofs to spot mistakes. Thank you for our stupid conversations about how to fix errors even when it meant believing that we could "differentiate with respect to the tree". Thank you for all the times you've tested my code; taught me the content that I needed to teach to undergrads (though, truth be told, I still don't really understand special relativity) and even marked problem sheets for me. It's a miracle you've had any time to do your own DPhil!

To the members of Team Donnelly, thank you for the joy and laughter that you have brought me over the last two-and-a-half years. **Ruth**, thank you for all your answers to my questions and for always knowing what's going on. **Nic**, thank you for all the wonderfully academically cynical conversations. **Sarah**, thank you for doing actually useful work to remind me and Nic that we're wrong. **Cathal**, thank you for all the excellent football chat to distract me from marking and **Tarek**, thank you for being the only other person to stay awake while we watched the football in Bologna. Finally, **Vik**[1], thank you for sharing the best moment of my DPhil with me at Oxford City's historic playoff final win (with thanks also to **Reece Fleet**, **Zac McEachran**, **Josh Ashby** and **Josh Parker** for making it happen). It has been a pleasure to get to know all of you, and I look forward to finding out where you all end up next.

To my family, thank you for supporting me throughout this process. Thank you to my **Mum** for your constant excitement about papers getting published; for convincing me that sometimes it's not appropriate to put a joke at the start of a paragraph; and for bearing with my slapdash schedule every time you've tried to arrange a call or a visit. Thank you to **Anna** for putting up with mine and Joe's stupid conversations even when you've been in the room too! And thank you to my **Dad** for passing on your joy of Maths (even if everyone else is maybe less thankful for that).

And finally, to the person who has suffered through enough Maths chat to last her a thousand lifetimes, who is probably the only person to have proof-read more equations than she's solved and who has borne with me to the level that she possibly now knows more about phylogenetic trees than real ones, a massive thank you to **Grace**. Thank you for all your patience on the evenings where my brain is full of Maths, for all the cold football matches you've watched (slept through); for all the much-needed distraction you've provided from work and all the love and affection you've shown me. You have made my DPhil experience infinitely better and I'm looking forward to our next steps together.

[1]Team Donnelly membership pending

Contents

Abstract

Epidemiology is underpinned by mathematical and statistical models which are used to answer key questions about the origin, spread, and control of diseases. With the rapidly increasing availability of data and computational resources, coupled with the vast number of deterministic and stochastic factors affecting the trajectory of epidemics, the complexity of these models continues to grow.

Many questions can be answered, at least in part, with these models by applying simple methods. For example, the variance of an epidemic under stochastic models can be estimated by running such a model many times. However, rigorous mathematical derivations allow these answers to be calculated more accurately, computed more efficiently, and, ultimately, understood in a deeper way.

This thesis seeks to provide mathematical insight into three key areas in epidemiology. First, the development of new methods for solving phylogenetic optimisation problems allows modern machine-learning and Bayesian techniques to be used to more accurately estimate the true evolutionary history of disease pathogens, as well as providing an explanation for the effectiveness of minimum evolution methods. Second, it considers the aleatoric uncertainty of epidemics, deriving explicit equations for the variance of an epidemic under a Crump-Mode-Jagers model. Finally, it considers the problem of optimal vaccination, deriving constraints and asymptotic limits on the optimal vaccination policy under a multi-group Susceptible-Infected-Recovered (SIR) model.

Abbreviations

Table 1: Abbreviations used in this thesis.

Abbreviation	**Meaning**
BME	Balanced Minimum Evolution
cdf	Cumulative Distribution Function
CTMC	Continuous-Time Markov Chain
GPU	Graphics Processing Unit
iid	Independent and Identically Distributed
KF	Kuhner-Felsenstein
LG	Le Gascuel
LHS	Left-Hand Side
LIFO	Last In First Out
MCMC	Markov Chain Monte Carlo
ME	Minimum Evolution
MLE	Maximum Likelihood Estimator
MRCA	Most Recent Common Ancestor
NB	Negative Binomial
NJ	Neighbour-Joining
NNI	Nearest-Neighbour Interchange
ODE	Ordinary Differential Equation
PDE	Partial Differential Equation
pdf	Probability Density Function
pgf	Probability Generating Function
pmf	Probability Mass Function
RF	Robinson-Foulds
RHS	Right-Hand Side
SARS	Severe Acute Respiratory Syndrome
SIR	Susceptible-Infected-Recovered
SPR	Subtree-Prune and Regraft
TMRCA	Time to Most Recent Common Ancestor
TPU	Tensor Processing Unit

Chapter 1: Introduction

Throughout history, infectious diseases have proven to be one of the most powerful untamed forces of nature, indiscriminately devastating even the most dominant civilisations [1]. Documented examples begin in the ancient world, with the 430BC plague of Athens being a key factor in their eventual defeat by Sparta in the Peloponnesian War [2]. As civilizations advanced and people became increasingly interconnected, the risk of epidemics grew. At the height of the Roman Empire, the 165AD Antonine Plague, sometimes considered the world's first pandemic [3], has been suggested to have had a substantial impact, with it and subsequent related epidemics perhaps even precipitating Rome's "third century crisis" [4]. More recently, the Black Death killed around 50% of Europe's population [5] and led to widespread social change, including the fall of medieval serfdom [6]; while the 1918-20 influenza epidemic, killing approximately 50 million people [7], may have affected the end of the First World War [8].

While modern medicine has reduced or even eliminated the risk from a wide range of diseases [9, 10, 11], the emergence of novel pathogens remains a substantial global threat, as illustrated by the worldwide impact of the COVID-19 pandemic [12, 13, 14]. With increasing availability and affordability of genetic data, phylogenetics - the study of evolutionary relationships - has become a key tool in identifying pathogens with pandemic potential [15, 16], while also providing insight into the evolution of variants of the disease in question [17, 18, 19].

However, despite advances in pandemic prevention, a range of environmental and societal factors are making pandemics increasingly likely [20, 21, 22]. Understanding the possible future behaviour of the prevalence of the disease is crucial in balancing the potential benefit of non-pharmaceutical interventions with their costs [23, 24, 25]. A complicating factor in achieving this goal is that pandemics have a high level of intrinsic randomness [26], so two outbreaks of identical pathogens could have vastly different results, and therefore having models that can not only provide mean forecasts but also accurately capture uncertainty is critical to ensure policy-makers are fully informed [27, 28].

Ultimately, where possible, many endemic diseases are controlled by vaccination [29, 30, 31]. However, when the supply or delivery rate of vaccines is limited, determining which people should be vaccinated

and the order in which this should be done becomes a matter of great importance [32, 33, 34]. Again, epidemiology can provide an answer to this question, as developing models to accurately forecast the outcomes of different strategies, as well as building algorithms to find the optimal strategy, ensures that the available resources will be used effectively [35, 36]. This effective distribution of vaccinations can greatly reduce the social [37] and economic [38] impacts of a pathogen, controlling its potential for spread within a population.

While vaccination can greatly reduce the impact of the disease, the risk can only be fully removed through complete eradication. For many diseases, such as COVID-19, such a goal is perhaps unattainable [39], but after the successful elimination of smallpox [40], it remains a realistic target for diseases such as polio [41]. Achieving eradication is a difficult and costly undertaking [42, 43], but combining previously-discussed methods of phylogenetic analysis [44], transmission modelling [45], and optimised controls such as vaccination [46] makes it possible, at least for certain diseases,, allowing countless future lives to be saved.

1.1 Thesis overview

This DPhil is motivated by the three key areas highlighted in the background: phylogenetics, epidemic uncertainty, and optimal vaccination. The primary objectives were

1) To make a meaningful contribution to our mathematical understanding of the models in question.

2) To develop a novel methodology to apply these models to real-world situations.

A summary of the six papers that seek to meet these goals is shown in Table 1.1.

1.1.1 Phylogenetics

Large amounts of genetic data are now available for a multitude of different species [52]. Although this should, in theory, increase the precision with which evolutionary histories can be inferred, many contemporary methods scale poorly and, therefore, it may be computationally infeasible to apply them to the entire dataset [53]. This is particularly true when considering the posterior distribution of evolutionary trees rather than simply trying to find the most likely tree [54]. In Papers I-III, we seek to address this issue by developing methods that could be used on large datasets.

Table 1.1: A summary of the six papers comprising this thesis.

Paper	**Title**	**Status**	**Reference**	**Topic**	**Objectives Fulfilled**
I	Phylo2Vec: a vector representation for binary trees	Accepted pending minor revisions at *Systematic Biology*	[47]	Phylogenetics	(2)
II	Leaping through tree space: continuous phylogenetic inference for rooted and unrooted trees	Published in *Genome Biology and Evolution*	[48]	Phylogenetics	(2)
III	Bayesian distance-based phylogenetics for the genomics era	Submitted to *Communications Biology*	-	Phylogenetics	(1) + (2)
IV	Intrinsic randomness in epidemic modelling beyond statistical uncertainty	Published in *Communications Physics*	[49]	Epidemic uncertainty	(1) + (2)
V	Optimality of maximal-effort vaccination	Published in *Bulletin of Mathematical Biology*	[50]	Vaccination	(1)
VI	Asymptotic analysis of optimal vaccination policies	Published in *Bulletin of Mathematical Biology*	[51]	Vaccination	(1) + (2)

Paper I

Paper I presents a novel method, Phylo2Vec, for representing trees, developing the concept of an integer representation that first appeared in [55]. Unlike previous integer representations, by motivating our construction from branching processes, Phylo2Vec is designed to be used for phylogenetic optimisation problems. It has a natural, efficient measure of tree distance and allows for the development of a simple optimisation algorithm with properties similar to subtree-prune and regraft methods, though also with the ability to make more radical moves. Alongside the paper, we also provide a Python package containing the algorithms discussed, allowing other researchers to use our representation to develop new methods.

Paper II

Building on the Phylo2Vec representation, Paper II describes a novel optimisation algorithm, GradME, which uses gradient-based methods to find the optimal tree with a balanced minimum evolution objective. Using information from all possible *ordered* trees - a natural subset of trees that arises from the Phylo2Vec construction - GradME is more effectively able to avoid getting stuck at suboptimal local minima. GradME outperforms the contemporary software FastME [56] on a number of datasets, albeit at an additional computational cost. We also develop mathematical theory to enable GradME to directly search for rooted trees and we provide an illustration of the effectiveness of this over the traditional midpoint-rooting method. For moderately-sized phylogenies we therefore believe that this paper offers a substantial development in phylogenetic optimisation.

Paper III

Paper III derives an approximate phylogenetic likelihood, the *entropic likelihood*, that can be calculated from a balanced minimum evolution objective function by adjusting the inter-taxa distance matrix. We show that, for feasible branch lengths, the entropic likelihood is very well-approximated by a linear function of the balanced minimum evolution objective function, while also being approximately linearly related to the traditional Felsenstein's likelihood. Therefore, this result provides mathematical justification for the closeness between tree inference from likelihood-based methods and balanced minimum evolution, unifying these two previously separate branches of research. Moreover, combining our entropic likelihood with the methods developed in Papers I and II, we detail an efficient algorithm for Bayesian analysis and show its utility by applying it to a large dataset of bird genomes.

1.1.2 Epidemic variance

As the availability of epidemiological data grows and the complexity of models increases, it becomes increasingly difficult to accurately and rigorously describe the behaviour of epidemic uncertainty under these models, with many contemporary works relying on simulation to estimate uncertainty [57, 27, 58]. However, while these estimates may be reasonably accurate, simulations can be computationally expensive, and it may be difficult to understand and compare the relative impact of different sources of uncertainty in the model.

Paper IV

To address these issues, Paper IV provides a detailed mathematical examination of a flexible branching process model of an epidemic. It derives renewal equations for the probability generating function, mean and variance of the prevalence under this model and characterises the source of the different terms in the variance equation, allowing their relative contributions to be compared. It also illustrates the importance of intrinsic uncertainty in epidemic modelling, proving that the prevalence is overdispersed (under very mild conditions) and illustrating its large size in real-world examples. Finally, it provides software to implement these methods, allowing researchers to efficiently assess uncertainty in a range of epidemic scenarios.

1.1.3 Optimal vaccination

When there is a limited supply of vaccinations during an ongoing epidemic, ensuring that they are optimally assigned within a population can substantially reduce the number of cases and deaths [35]. Solving the optimal vaccination problem must (in general) be done computationally [59] and the resultant solutions may be difficult to justify to policy-makers. Therefore, developing a set of rigorously-derived principles for the optimal policy can be a useful tool in improving our understanding of optimal vaccination.

Paper V

Paper V states and proves an intuitive theorem - that one should always vaccinate with *maximal effort* (giving out vaccines as early and as quickly as possible), providing the vaccination is effective and its effectiveness does not wane over time. It achieves this for a general heterogeneous Susceptible-Infected-Recovered (SIR) model for any number of groups and also considers the case where the cost of vaccination is included, in this scenario showing that the optimal policy involves maximal-effort vaccination up to some (possibly infinite) time. While this theorem may be conceptually obvious, it

provides a starting point for optimising vaccination policies, reducing the dimension of the feasible set by 1.

Paper VI

Building on the results proved in Paper V, Paper VI derives results on the leading-order optimal solution in two cases of asymptotic parameters. It shows that a small, vulnerable subgroup should always be vaccinated first (in the limit) and derives an explicit leading-order solution in the case of a small vaccination supply. These principles provide some insight into the structure of optimal vaccination policies and could be used to inform and explain results in more general settings.

1.2 Thesis structure

This thesis is structured as follows:

- Chapter 2 provides a background to the three areas of focus in this thesis and gives an overview of the relevant literature.
- Chapters 3-8 contain the papers which form the basis of this thesis.
- Chapter 9 provides a high-level discussion and summary of the findings of this thesis.
- Appendices A-F provide the appendices from the papers in this thesis.
- Appendix G summarises the other papers published during this DPhil which are not part of this thesis.

This is an integrated thesis, so each of the chapters contains its own introduction, literature review, and discussion. Thus, Chapters 2 and 9 are intended only as an overview of the relevant topics.

1.3 Other work

Papers I-VI comprise the majority of the work carried out during my DPhil period (October 2021 - April 2024). I have also been at least a joint-first author on six other papers and preprints, which are summarised in Appendix G. In general, these are out-of-scope of the topic of this thesis and have therefore not been included in the main work (the single exception, [60], has been excluded as I only contributed one of the theorems), with a number of them [61, 62, 63] complementing my work as a data analyst for a range of football clubs (Como 1907, Oxford United FC, Solihull Moors FC and Oxford City FC).

Chapter 2: Background and literature review

2.1 Phylogenetics

2.1.1 Introduction

Phylogenetics is the study of the evolutionary history of different taxa. The aim is to produce a phylogenetic tree that describes the evolutionary relatedness of the different taxa under consideration [64]. These trees are comprised of nodes (denoting the different taxa) and branches joining these nodes (representing the evolutionary distance between different divergence events). Note that the evolutionary distance may be different from the true time, as factors such as population size can affect the rate of evolution [65].

Understanding evolutionary dynamics is a crucial epidemiological endeavour [66, 67, 68]. For example, these dynamics drive the emergence of new diseases [69]. Most new human diseases, including Severe Acute Respiratory Syndrome (SARS) [70], H1N1 influenza [71], and Middle East Respiratory Syndrome (MERS) [72], are caused by zoonotic spillover, where viruses that were previously endemic in other animals undergo mutations that allow them to infect humans [73]. While often these infections occur in isolated clusters, as the pathogen is unable to sustain human-to-human transmission [74], there is heavy selective pressure towards mutations that allow for greater adaptation to human hosts [75], and therefore greater pandemic potential. With factors such as urbanisation [76] and climate change [77] increasing the rate of pathogen emergence, understanding this process allows effective mitigation strategies to be implemented and can therefore have substantial public health benefits [78].

The role of evolution in epidemics is not limited to the initial outbreak stage. As a recent example, the trajectory of the COVID-19 pandemic around the world was driven by rapidly emerging variants [79, 80, 81]. In addition to being more transmissible [82, 83] and, in some cases, a higher mortality rate [84], new variants can reduce the effectiveness of both natural and vaccine-derived immunity [85, 86]. This can have large implications for public health policy [87], and so modelling this evolution is a crucial aspect of epidemiology [88].

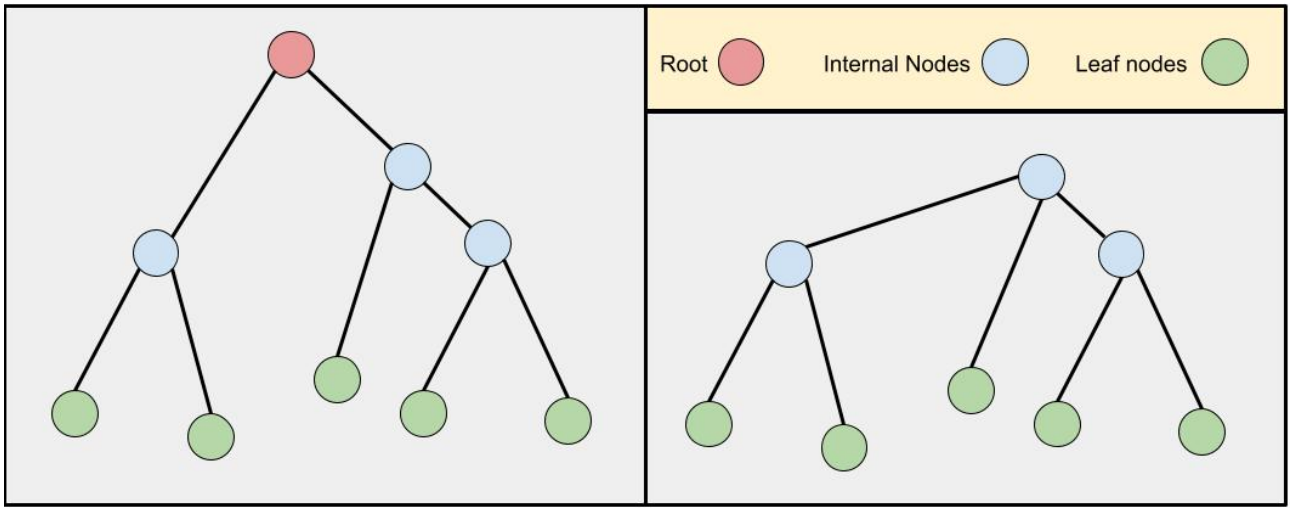


Figure 2.1: Examples of rooted (left panel) and unrooted (bottom right panel) phylogenetic trees.

Alongside understanding short-term pathogen evolution, the evolutionary history of both the pathogen and the host is a key determinant of the risk of the disease jumping from one species to another [89, 15]. Understanding the pathogen's evolutionary history can also provide insight into drug resistance [90] and factors that affect evolutionary pressure towards human-adapted diseases [91]. It is therefore crucial to have methods which can accurately and efficiently reconstruct these evolutionary histories [66].

2.1.2 Binary phylogenetic trees

While other frameworks exist, such as phylogenetic networks that allow horizontal gene transfer [92], phylogenetic trees are generally constrained to be binary [93, 94, 95]. In these binary trees, there are up to three types of nodes - *leaf nodes*, *internal nodes* and *the root*. Leaf nodes, which represent the taxa under consideration, have degree 1 and share an edge only with an internal node or the root. These internal nodes, which represent previously-evolved taxa that are ancestors of the leaf nodes, have degree 3, while, if a root is included, this has degree 2 and represents the most recent common ancestor of the taxa under consideration. A tree without a root is called *unrooted*, while a tree with a root is called *rooted*. Examples are shown in Figure 2.1.

These phylogenetic trees describe the evolutionary relatedness of different taxa [96]. Rooted trees have the simplest interpretation, as there is a sense of "direction" - that is, the path from the root to a leaf shows the evolutionary process from the most recent common ancestor to the taxon represented by that leaf. In rooted trees, one can define the *generation* of a node to be the number of edges on the shortest path between that node and the root (the generation of the root is therefore zero). One can

also define the *children* of a node x to be any nodes sharing an edge with node x that have a larger generation (necessarily, their generation will be one more than the generation of x). Similarly, unless node x is the root, the parent of node x is the node with which it shares an edge that has a smaller generation than x (again, this generation will be one less than the generation of x).

Unrooted trees have a more abstract interpretation, as they do not have this intrinsic sense of direction [97]. However, they still show the relatedness of taxa through the length of the paths between them. For example, in the unrooted tree in Figure 2.1, the leftmost taxon is most closely related to the second taxon from the left, as the path between them is shorter than the path between the leftmost taxon and any other taxon. Thus, both representations provide useful information about the evolutionary history of the taxa under consideration, though the rooted representation contains additional, biologically-motivated, constraints.

2.1.3 Phylogenetic distance

It is helpful when considering the space of phylogenetic trees to have a measure of distance between a pair of trees. Several different metrics have been developed, including a range of metrics based on the number of times a certain operation is performed, such as a subtree-prune and regraft (SPR) move (where a subtree is removed from the tree and attached to another section) [98]. However, while these distances are intuitive, they are generally difficult to calculate exactly for distant trees [99]. An alternative is the Robinson-Foulds distance [100]. This metric, which measures the number of leaf node partitions that are in one tree and not the other, can be efficiently calculated [101] and is therefore generally used in this thesis despite its relative lack of sensitivity [102].

2.1.4 Phylogenetic optimisation

Phylogenetic optimisation involves finding the best phylogenetic tree based on some objective function and an underlying dataset, such as genetic [103] or proteomic [104] information for each of the leaf taxa [105]. In both of these cases, the dataset is the values at different sites in each taxa (for example, in the genetic case, each site has value A, C, G or T). However, not every site will be available for each taxon (as, for example, it may not be possible to identify the equivalent section of the genome in distantly-related taxa) meaning that there may be gaps in the dataset [106].

In this setting, there are three main groups of objective functions [107]: likelihood, parsimony and distance-based objectives. Parsimony involves looking for the tree which would require the fewest

total mutations to generate the dataset [108]. This approach does not require any kind of mutation model [109], and is therefore simple to implement [110]. However, it is not guaranteed to be statistically consistent [111], as the *long branch attraction* phenomenon [112] can lead to incorrect inference. Because of this, parsimony is not used in any of the papers in this thesis.

Likelihood-based methods

Likelihood methods [113, 114, 115] involve specifying a model for the mutation of each site from one state to another. Under this model, by marginalising over the possible states at internal nodes, one can then calculate the probability that a given tree resulted in taxa with the observed data [116] - a calculation that is generally referred to Felsenstein's likelihood [117]. Under certain assumptions, calculating likelihoods in this manner is possible even when there are gaps in the data at the leaf nodes, as, again, the unknown states can be marginalised over [118]. Generally, likelihood methods are considered to produce the most reliable results of the three phylogenetic optimisation methods [119, 120, 121].

Given this likelihood function, there are two main possible methods of performing inference. Firstly, one can consider the problem of finding the maximum likelihood tree, and use this to provide a single representation of the evolutionary history [122]. A wide range of algorithms have been developed to find the maximum likelihood phylogenetic tree [113, 114, 123, 124], the vast majority of which use intuitive heuristics to decide which tree to evaluate next, such as nearest neighbour interchange (NNI) [125], where the position of two neighbouring subtrees is swapped; the previously mentioned SPR move; or tree bisection and reconnection (TBR) [126], where the tree is split into two parts, and the possible ways of reconnecting the resulting subtrees are explored. Alongside the topology, the branch lengths must also be optimized, although this is a far simpler task as it is simply the optimization of a, generally relatively small, number of continuous variables. Thus, it is possible to re-optimize at least some [127], if not all [128], of the branch lengths for each candidate tree, before performing a final optimization on the output maximum-likelihood tree [129].

However, while the simplicity of the results and (relatively) low computational costs of maximum likelihood estimation may be appealing, their focus on a single tree can fail to be a good representation of the potentially diverse feasible space of trees [130]. Additionally, and perhaps more concerningly, they can even fail to be reproducible [131], as different runs of the optimisation algorithm may converge to different trees.

Thus, it is often of interest to explore the posterior distribution of the set of phylogenetic trees under Felsenstein's likelihood through Bayesian analysis [132, 133]. In order to implement this approach, it is necessary to choose a prior on the space of topologies and branch lengths. The appropriateness of different topological priors depends on the type of tree under consideration, with unrooted trees often assigned uniform priors [134], while rooted trees may be given priors from biologically-motivated tree construction processes such as birth-death processes [135]. Branch lengths are commonly given exponential priors [136, 137, 138], motivated both by Markovian assumptions and empirical evaluation [139]. However, other alternatives, such as uniform priors [140] or Pareto priors [141] may be preferable depending on the exact biological scenario in question.

Again, many algorithms have been developed to achieve this goal [142, 143] which, unsurprisingly, are generally based on heuristic frameworks similar to their maximum likelihood equivalents. However, this analysis is often computationally expensive, particularly for large datasets, sometimes taking months [144, 145] or even years [144] of computation time. Therefore, the development of scalable sampling algorithms to cope with the ever-increasing availability of genetic data is an important problem in phylogenetics [54].

Distance-based methods

While both likelihood and parsimony use each unique site separately in each calculation of the objective function, distance-based approaches instead use them once to create a single, fixed, matrix describing the evolutionary distance between each pair of taxa [146]. The objective function then uses this matrix as its sole data source - a property that is particularly computationally advantageous when the amount of data is large [147], as the size of the distance matrix can be many orders of magnitude smaller than the number of sites in the dataset [148]. As discussed previously, the increasing availability of phylogenetic data makes this efficiency important [52].

Of the distance-based methods, balanced minimum evolution has the advantages of being computationally simple and statistically consistent [147, 149]. In balanced minimum evolution, one seeks to minimise the total length of the tree, where, for a given tree *topology* (that is, a tree where no lengths have been assigned to the branches), the inter-taxa distance matrix is used to provide an estimate of the branch lengths. The length L of a tree $\mathcal{T}$ can be calculated efficiently without the need to find

the individual branch lengths using the formula

$$L(\mathcal{T}) = 2\sum_{i<j} D_{ij}2^{-e_{ij}} \tag{2.1}$$

where D is the distance matrix and e_{ij} is the unweighted path length between taxon i and taxon j in the tree $\mathcal{T}$ [150]. Due to the simplicity of this formula, one can efficiently use the SPR heuristic to attempt to find the optimal tree [56]. However, this approach can fail to find the optimal tree, as we show in [48].

While the efficiency and statistical consistency of balanced minimum evolution is appealing, it tends to provide worse results in practice than likelihood methods [151]. However, it has been shown to achieve decent performance when applied to a variety of problems [152, 153, 154], and is a useful tool when the size of the genetic dataset is too large for likelihood methods to be computationally feasible [155]. Moreover, its relative mathematical tractability means that it is the focus of the novel methodology developed in Papers II and III.

2.2 Epidemic variance

2.2.1 Introduction

Randomness is fundamental to epidemic behaviour [156, 157, 158]. Differences in epidemic behaviour between otherwise equivalent deterministic and stochastic models can be substantial [159, 158, 160], particularly in the early and late stages of an epidemic when the number of cases is small [161], and so the effect of stochasticity deserves careful attention.

From a practical modelling standpoint, there are two sources of uncertainty to consider: aleatoric and epistemic uncertainty [162]. Epistemic uncertainty is uncertainty in the model parameters that are used to describe the epidemic, due to the finite amounts of available data [163]. This is present and can be taken into account in both deterministic and stochastic models [164].

Conversely, aleatoric uncertainty is the intrinsic randomness of the epidemic, and hence can only be captured by stochastic models [165]. Even if the model parameters were perfectly known, this uncertainty would still exist and have a meaningful effect on the distribution of the epidemic trajectory [158]. Thus, developing and understanding models that capture this aleatoric uncertainty is an

important topic of research.

The practical implications of epidemic stochasticity are far-reaching [26]. Under deterministic models such as the SIR model, epidemics cannot become extinct, although case numbers will tend towards zero [50]. However, this is possible in stochastic models (including the stochastic SIR model) and, under certain conditions (where the reproduction number is sufficiently larger than 1 [166]) leads to a bimodal distribution of the total number of cases [167], where it is highly likely that either the epidemic dies out while there are a small number of cases or after it has infected a substantial proportion of the population. This has important implications when considering the risk of emerging epidemics [168], while also highlighting the importance of early interventions [169].

Epidemic variance is also an important quantity to consider when modelling the impact on the public health system [170, 171]. Planning capacity around the epidemic trajectory of cases predicted by a deterministic model is insufficient, as the true number of cases may be substantially higher [172]. Even more dramatically, stochasticity can cause otherwise stable states for endemic diseases to become unstable [160], though a discussion of this stochastic resonance is beyond the scope of this thesis.

2.2.2 Stochastic models of epidemics

A number of frameworks have been developed to stochastically model epidemics [173], with stochastic SIR-style models being a popular choice for epidemiologists [174, 175, 176]. These models use a compartmental structure (for example, splitting the population into susceptible, infected, and recovered compartments), with individuals randomly moving from one compartment to another with rates dependent on the number of individuals in each compartment [177]. However, while the simplicity of these models is appealing, they fail to capture a number of important aspects that influence the variance of the epidemic. In particular, the number of cases caused by an infected individual can never be overdispersed [178], which means that one must modify the model to incorporate the effect of superspreading, that is, when a small number of individuals cause a large number of cases [179, 180, 181].

Another class of stochastic models used to simulate epidemics is agent-based models [182, 183]. By simulating at an individual level, these models allow for the consideration of a wide range of social, geographical, and epidemiological factors [184, 185, 186]. However, this complexity comes with a high computational cost [187], as well as the need for large numbers of empirically-estimated parameters [188].

Finally, branching process models, such as the model used in Paper IV, provide a naturally flexible method of modelling epidemics [189, 190, 191]. In branching process models, each individual, once infected, causes a certain random number of infections throughout, or at the end of, their infectious period [192]. Thus, one can represent the epidemic using a tree, with each node referring to an infection event.

In this thesis, we will limit our discussion to branching processes that have the *self-similarity property* [60]. This means that the number and timing of cases caused by an individual infected at some time t depends only on t (and the parameters of the epidemic) and is independent of the history of the epidemic. This is crucial in enabling the mathematical analysis carried out in [49].

A number of different models satisfying this property have been developed [193], including Galton-Watson processes [194], a discrete branching process where each individual causes a random number of new infections at the next timestep; Bellman-Harris processes [195], where each individual causes a single burst of new infections after a random period of time; and the more general Crump-Mode-Jagers process [196], which is the topic of [49].

2.2.3 The Crump-Mode-Jagers process

In a Crump-Mode-Jagers process [60], when individuals are infected, they remain infectious for a random period of time L. During this infectious period, they infect other individuals according to some counting process. These infections may occur in "bunches" - that is, the jumps in the counting process do not need to be of unit size - or individually. By choosing the distribution of this counting process and the random lifetime L, a wide range of epidemic scenarios can be modelled.

However, a key disadvantage of this model is that it cannot account for susceptible depletion [197] - the decrease in the number of susceptibles in the population over the course of an epidemic - and so the usefulness of the model is limited once the number of susceptibles begins to noticeably decrease [198]. In a true epidemic, the number of infections that an individual causes is dependent on previous infections, even if all individuals behave independently, as, for example, if the population is fully immune, then no further infections can occur. To address this issue, more complicated models are required, but this added complexity substantially reduces the mathematical tractability of the model [199].

2.3 Optimal vaccination

2.3.1 Introduction

Vaccination is one of the most effective and important methods of disease control [200, 201, 202, 203]. Unlike other methods such as lockdowns [204] and quarantine [205], the negative societal impact is limited to the small number of adverse reactions [206]. However, as was seen in the COVID-19 pandemic, vaccines can take a substantial amount of time to produce and distribute [207].

Because of this, it is crucial to consider the order in which people should be given vaccinations, as this can have a substantial impact on the number of cases and deaths from an epidemic [208, 209]. There are many examples of targeted vaccination programmes across the globe [210, 211]. Perhaps the most notable example is the COVID-19 pandemic, where, faced with a limited supply, governments were forced to decide which members of the population to vaccinate first [212, 213]. With diseases such as COVID-19, where the fatality rate by age was (in general) negatively correlated with the transmission rates [214, 215], there was a trade-off between reducing the infection rate in the population and protecting those who would become more severely ill if infected. In such scenarios, mathematical modelling is crucial to ensure that the best policy is chosen, particularly as there may be large differences in the optimal policy between different diseases [216].

Modelling population heterogeneity

Approximating the optimal vaccination policy relies on a good understanding of heterogeneity within a population. While heterogeneity exists at an individual level, as captured by agent-based models [183], for practical purposes, the population is generally divided into a set of groups based on characteristics such as household size [217], occupation [218], geography [219] and, most commonly, age [33, 220, 216].

In general, the easiest quantity to estimate across these different groups is the case fatality ratio [221], alongside related quantities such as hospitalisation rates [222] and the long-term individual-level impact of an infection [223]. However, the use of these quantities alone may lead to sub-optimal vaccination policies [224], as the most vulnerable group does not always cause the most infections, so it is also important to understand the transmission of the disease between and within these different groups.

A number of methods have been developed to model this, including stochastic models [225], network models [226], reaction-diffusion models [227] and discretised models [228], among many others

[187]. The increasing availability of data and the development of modern machine-learning technology means that the pace of development remains high [229, 230].

2.3.2 Multi-group SIR-type models

Despite these developments, classical deterministic compartmental models remain the most popular choices for epidemiologists [231]. In these models, each group within the population is divided into different compartments based on their infectious state. The simplest version of the model, the SIR model for a single group, is defined by equations

$$\frac{dS}{dt} = -\beta SI \tag{2.2}$$
$$\frac{dI}{dt} = \beta SI - \mu I \tag{2.3}$$
$$\frac{dR}{dt} = \mu I \tag{2.4}$$

where S, I and R are respectively the number (or proportion) of susceptible, infected and recovered (in the sense that they are no longer infectious, though this may have happened through death) members of the population. In this framework, β controls the transmission and μ controls the rate of recovery (or death) from the disease.

A number of studies use this formulation to carry out cost-benefit analyses of vaccination policies [232]. However, to explore the distribution of vaccinations across the population, it is necessary to use the multi-group model [233] where, for example, the number of susceptibles S_i in group i varies according to

$$\frac{dS_i}{dt} = -S_i \sum_j \beta_{ij} I_j \tag{2.5}$$

so that the members of group i can be infected (with different rates) by members of the other groups j.

Vaccination can then be added into the model, most commonly by adjusting (2.5) to be [234, 235, 236]

$$\frac{dS_i}{dt} = -S_i \sum_j \beta_{ij} I_j - U_i(t) S_i \tag{2.6}$$

where $U_i(t)$ is the vaccination rate in group i. These vaccinated members will move into another group in the population which, depending on the model assumptions, may still be able to be infected. However, under this construction, a group can never be fully vaccinated (provided the vaccination rate remains bounded by some $\mathcal{U}$) as (ignoring infection - which only reduces the number of susceptibles

who can be vaccinated)

$$\frac{dS_i}{dt} = -U_i(t)S_i \geq -\mathcal{U}S_i \tag{2.7}$$

which means

$$\frac{d}{dt}\Big(\log(S_i(t))\Big) \geq -\mathcal{U} \tag{2.8}$$

and hence

$$S_i(t) \geq S_i(0)e^{-\mathcal{U}t} > 0 \tag{2.9}$$

and so remains bounded above zero for all finite t. To solve this issue, in Papers V and VI, we consider a more general formulation where complete vaccination coverage is possible, although we show it has the formulation (2.6) as a special case.

A range of other compartments are used in the literature [237]. Most commonly, an *exposed* compartment, through which individuals pass between susceptible and infected, provides a more realistic model of the infectiousness profile [238, 239, 240], with other compartments including asymptomatic infections [241] and quarantine [242]. Other extensions are also possible, such as allowing the parameters to vary over time [243]. However, while these models are generally similar in structure, we restrict attention to the three basic compartments to reduce the complexity of the mathematical analysis in Papers V and VI.

2.3.3 The optimal vaccination problem

Different optimal vaccination problems have been formulated throughout the literature. One approach seeks to reduce the reproduction number [244] (a measure of how many secondary infections will be caused by an infected individual) as much as possible by vaccinating before infections arrive in a population [217, 245]. If this reproduction number can be reduced below one, then this will avoid an epidemic, but otherwise, it is crucial to consider the heterogeneity in mortality rates [216]. Furthermore, the assumption of instantaneous vaccination simplifies the problem, making it analytically solvable if there are only two groups [246], though a more realistic delivery timescale could cause a change in the optimal policy [247, 248].

In contrast, the optimisation problem used in Papers V and VI follows papers such as [249], [250] and [251] in considering minimising the overall deaths in a scenario in which vaccinations are delivered at a finite, limited rate and that vaccine supply is limited. This problem can be solved numerically using Pontryagin's Maximum Principle [252, 253, 254] and allows a wide range of scenarios to be con-

sidered [255].

Other factors can be included in the objective function such as the cost of vaccinations [256, 257]; different measures for epidemic impact such as years of life lost [258]; the use of vaccination alongside other control measures such as lockdowns [259]; vaccinations which require multiple doses [260]; vaccine hesitancy [261]; and uncertainty in vaccine effectiveness [262]. To simplify the analysis in Papers V and VI, we restrict attention to a simple objective function, though we do consider the effect of vaccination cost in Paper V.

Chapter 3: Paper I: Phylo2Vec: a vector representation for binary trees

Matthew J Penn et al. "Phylo2Vec: a vector representation for binary trees". In: arXiv preprint arXiv:2304.12693 (2023).

Status: This paper has been accepted pending minor revisions at *Systematic Biology*.

Abstract: Binary phylogenetic trees inferred from biological data are central to understanding the shared history among evolutionary units. However, inferring the placement of latent nodes in a tree is NP-hard and thus computationally expensive. State-of-the-art methods rely on carefully designed heuristics for tree search. These methods use different data structures for easy manipulation (e.g., classes in object-oriented programming languages) and readable representation of trees (e.g., Newick-format strings). Here, we present Phylo2Vec, a parsimonious encoding for phylogenetic trees that serves as a unified approach for both manipulating and representing phylogenetic trees. Phylo2Vec maps any binary tree with n leaves to a unique integer vector of length $n-1$. The advantages of Phylo2Vec are fourfold: i) fast tree sampling, (ii) compressed tree representation compared to a Newick string, iii) quick and unambiguous verification if two binary trees are identical topologically, and iv) systematic ability to traverse tree space in very large or small jumps. As a proof of concept, we use Phylo2Vec for maximum likelihood inference on five real-world datasets and show that a simple hill-climbing-based optimisation scheme can efficiently traverse the vastness of tree space from a random to an optimal tree.

Full Author List: Matthew J Penn, Neil Scheidwasser, Mark P Khurana, David A Duchêne, Christl A Donnelly and Samir Bhatt

Joint Authorship: M.J.P., N.S. and S.B. have joint authorship of this work.

Author contributions:[1] S.B., N.S. and M.J.P. conceived of the study. S.B. supervised. S.B.,

[1] As found in the published preprint.

N.S., M.J.P. designed the study. S.B., M.J.P. and N.S. performed optimisation runs. S.B., M.J.P., N.S. performed analysis. S.B., M.J.P. and N.S. drafted the first original draft. All authors contributed to editing the original draft. N.S. and D.A.D. contributed to revisions of the methodology. M.J.P., N.S. and S.B. drafted the appendix.

3.1 Introduction

Phylogenetic trees are a fundamental tool for depicting evolutionary processes, whether linguistic (evolution of different languages and language families) or biological (evolution of biological entities). Within the biological sciences, phylogenetic trees are integral to multiple research domains, including evolution [263], conservation [264], and epidemiology, where they allow us to better understand infectious disease transmission dynamics [265, 266].

A multitude of computer-readable formats have been proposed to store and represent (binary) phylogenetic trees. Although basic data structures such as arrays or linked lists can be used for this purpose, the Newick format, as outlined by [267] and [268], has emerged as the standard notation. This format characterises a tree through a string of nested parentheses. Each parenthesis encloses a pair of leaf nodes or subtrees, separated by a comma. Additional metadata such as branch lengths can be added after a colon which follows the leaf node or subtree. Although a compact and intuitive notation, several limitations exist. First, comparing (large) trees using the Newick format can be difficult for human readers, especially as isomorphic trees can be obtained by permuting nodes or subtrees within a set of parentheses. Second, verifying that two trees are identical from Newick strings requires additional steps, as two identical trees need not have the identical Newick string. Alternative, bijective representations do exist. For example, several methods have been proposed to assign unique integers to binary trees with unlabelled [269, 270], fully labelled [271], and partially labelled nodes [55] (only leaf nodes). More recently, [272] investigated several enumeration strategies of binary trees compatible with a perfect phylogeny. However, as mentioned by [55], using single-integer representations for downstream phylogenetic analyses is computationally difficult for large trees due to the factorial rate at which the size of binary tree space grows [273]. Conversely, several vector representations such as graph polynomials [274, 275] and the compact bijective ladderized vector [276] were introduced as a support for model selection and estimation of evolutionary or epidemiological parameters. Other vector representations of tree topology, such as pair matchings [277] and F matrices [278], focus on the polynomial-time computation of the distance between any two trees (to measure similarity). However, methods for systematically sampling random trees or changing tree topology with respect to an objective function by leveraging such vector representations have been understudied. In particular, creating sampling schemes (as done in Bayesian frameworks such as `BEAST` [143, 279] and `MrBayes` [142]) around standard tree arrangements is non-trivial, and, although inferring phylogenetic trees is a common task in evolutionary biology, tree search using any optimality criteria (including maximum likelihood) is NP-hard [280]. Another critical challenge is the size of the tree space: for a tree with n leaves, there

are $(2n-3)\cdot(2n-5)\cdot\ldots\cdot 5\cdot 3\cdot 1$ possible rooted binary trees [273]. Lastly, optimisation-based approaches often face a jagged "loss" landscape containing many trees with the same criterion score [281]. When considering inference, the choice of representation can be particularly relevant for application to real phylogenetic problems. For example, an application of the approach we introduce here can be used for continuous relaxation and gradient descent under the minimum evolution criterion [47]. For large phylogenies, the use of an efficient representation such as the compact bijective ladderized vector [276] has proven effective for deep learning-based, likelihood-free, inference [282] or diversification inference [283].

To overcome these limitations, we introduce Phylo2Vec, a new representation for any binary tree. In this framework, the topology of a binary tree can be completely described by a single integer vector $\boldsymbol{v}$ of dimension $n-1$, where n is the number of leaves in the tree. The vector's construction is intrinsically related to the branching pattern of the tree, and is defined by a simple constraint: $v_j \in \{0, 1, \ldots, 2(j-1)\}$ for all $j \in \{1, \ldots, n-1\}$. The approach we present here is most similar to that previously introduced by [55], but we focus on the integer representation and its mathematical properties, rather than counting or labelling trees. Additionally, this formulation naturally offers a new measure of distance between trees (e.g., by comparing two vectors using the Hamming distance) and yields a new mechanism to explore tree space which diverges from traditional heuristics such as subtree, prune and regraft (SPR). To further demonstrate its utility, as a proof of concept, we apply Phylo2Vec to several phylogenetic inference problems, where the task is to find an optimal tree given a set of genetic sequences using maximum likelihood estimation. While state-of-the-art frameworks for phylogenetic inference typically rely on search heuristics based on deterministic tree arrangements, Phylo2Vec provides the first steps to a more systematic criterion for optimisation.

3.2 Materials and Methods

The goal of this project was to develop a bijection (i.e., a one-to-one correspondence) between the set of binary rooted trees with n leaves to a constrained set of integer vectors of length $n-1$. We first describe an intuitive but incomplete (as not bijective) integer representation of trees as birth processes. Second, we define and characterise Phylo2Vec as a bijective generalisation of this first representation and formalise its properties. Third, we showcase the utility of Phylo2Vec by applying the representation for MLE-based phylogenetic inference on empirical datasets.

Our construction draws from an existing method of assigning integer counts to trees [55], although we focus on vector representations. It is distinct from [55] in labelling the tree edges, motivated by a simple and intuitive representation of birth processes. By applying this encoding to rooted binary trees,

we are able to move around tree space similarly to subtree-prune and regraft methods. Furthermore, we provide a rigorous proof of its bijectivity alongside a range of algorithms (all implemented into a Python package) which allows researchers to build on the phylogenetic optimisation algorithm we present here. Thus, we provide a significantly different method from those proposed previously [55], by focusing our efforts toward practical transitions in tree space.

3.2.1 An incomplete integer representation of trees as birth processes

Let $\mathcal{T}$ denote a rooted phylogenetic tree with n leaf nodes representing (biological) taxa, and $\mathcal{D}$ symbolise a key-value mapping (or dictionary) which associates a nonnegative integer (the keys) to each leaf node (the values).

Using this mapping, *for a subset of all trees*, we can summarise their topology using an integer vector $\boldsymbol{v}$ of size $n-1$ such that:

$$v_j \in \{0, 1, \ldots, j-1\} \quad \forall j \in \{1, \ldots, n-1\} \tag{3.1}$$

The construction of this vector is inspired by birth processes: assuming a two-leaf tree with leaves labelled 0 and 1, we process $\boldsymbol{v}$ from left to right. For each $j \in \{1, \ldots, n-1\}$, v_j (hereinafter noted as $\boldsymbol{v}[j]$) denotes the addition of leaf j such that, at iteration j, leaf j forms a cherry with leaf $\boldsymbol{v}[j]$. In other words, the branch leading to leaf $\boldsymbol{v}[j]$ "gives birth" to leaf j. Figure 3.1 illustrates algorithms to convert a tree to a vector and *vice versa*.

Although a simple representation of tree topology, it is easy to see from Equation 3.1 that this construction is incomplete. Indeed, there are j possible values for any $\boldsymbol{v}[j]$, and thus, for n leaves, there are $1 \cdot 2 \cdot \ldots \cdot (n-1) = (n-1)!$ possible vectors, which is less than the number of binary rooted trees, $(2n-3)!!$ (where !! denotes the semifactorial) [273, 284, 277]. This discrepancy stems from the assumptions of this construction, whereby a new leaf j has to form a cherry with a previously processed leaf $0, 1, \ldots, j-1$. For instance, leaf 2 has to form a cherry with either leaf 0 or 1, but cannot be an outgroup of the (0, 1) subtree. We thus denote trees that follow this incomplete construction of tree space as "ordered" trees, as they require a precise ordering of the leaf nodes.

3.2.2 Phylo2Vec

In this section, we define and formalise the properties of Phylo2Vec, an integer vector representation that extends the formulation presented above to be valid for any rooted binary tree. To ensure bijectivity to this space, we need the vector $\boldsymbol{v}$ to satisfy the following constraints:

$$v_j \in \{0, 1, \ldots, 2(j-1)\} \; \forall j \in \{1, \ldots, n-1\} \tag{3.2}$$

We say $\boldsymbol{v} \in \mathbb{V}$ if Equation 3.2 is satisfied. For this representation, there are $2j-1$ entries for any position j. Therefore, the number of possible vectors matches the number of possible binary rooted trees:

$$\prod_{j=1}^{n-1}(2j-1) = (2(n-1)-1)!! = (2n-3)!!$$

From this observation, we can prove the bijectivity of the mapping simply by showing injectivity - that is, that no two distinct vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ lead to the same tree. A proof is presented in the Appendix (Phylo2Vec details). Briefly, our proof relies on the fact that certain properties of pairs of nodes are preserved throughout the construction process - namely, that the most recent common ancestor (MRCA) of a pair of nodes is unchanged (once both nodes have been added to the tree) and that if one node is the ancestor of another at some stage of the construction process, then this remains true in the final tree. Then, if $\mathcal{T}$ and $\mathcal{T}'$ are the trees resulting from different vectors $\boldsymbol{v}$ and $\boldsymbol{v}$', respectively, we choose the smallest i such that $v_i \neq v_i'$. By considering the sets of leaf nodes descended from the edge to which i is added, we can show that the addition of node i causes a pair of nodes to either have a different MRCA or a different ancestral relationship. Therefore, since these properties are preserved throughout the construction process, we must have $\mathcal{T} \neq \mathcal{T}'$. This shows the injectivity of our mapping, with bijectivity following from the fact that the number of trees is the same as the number of possible vectors $\boldsymbol{v}$.

Recovering a tree from a Phylo2Vec vector

Building a binary tree from $\boldsymbol{v}$ follows closely the algorithm in Figure 3.1, but incorporates two additional requirements. First, we start from a two-leaf tree, whereby the leaves are labelled 0 and 1. The branches that lead to leaves 0, 1 are also labelled 0, 1, respectively. Second, we draw an additional node (called the "extra root") which is initially connected to the root by a branch labelled 2 (see second row in Figure 3.2).

The addition of a temporary root in the construction of $\boldsymbol{v}$ from a tree and vice versa ensures that there are $2j-1$ branches from which a leaf j can descend from. From these requirements, we can build a unique rooted tree $\mathcal{T}$ by processing $\boldsymbol{v}$ from left to right, where $\boldsymbol{v}[j]$ indicates the branch that will split and yield leaf j. Figure 3.2 shows a detailed example of this scheme, and other example representations for trees with $n = 4$ leaves are shown in Figure 3.3. We also describe (and prove its

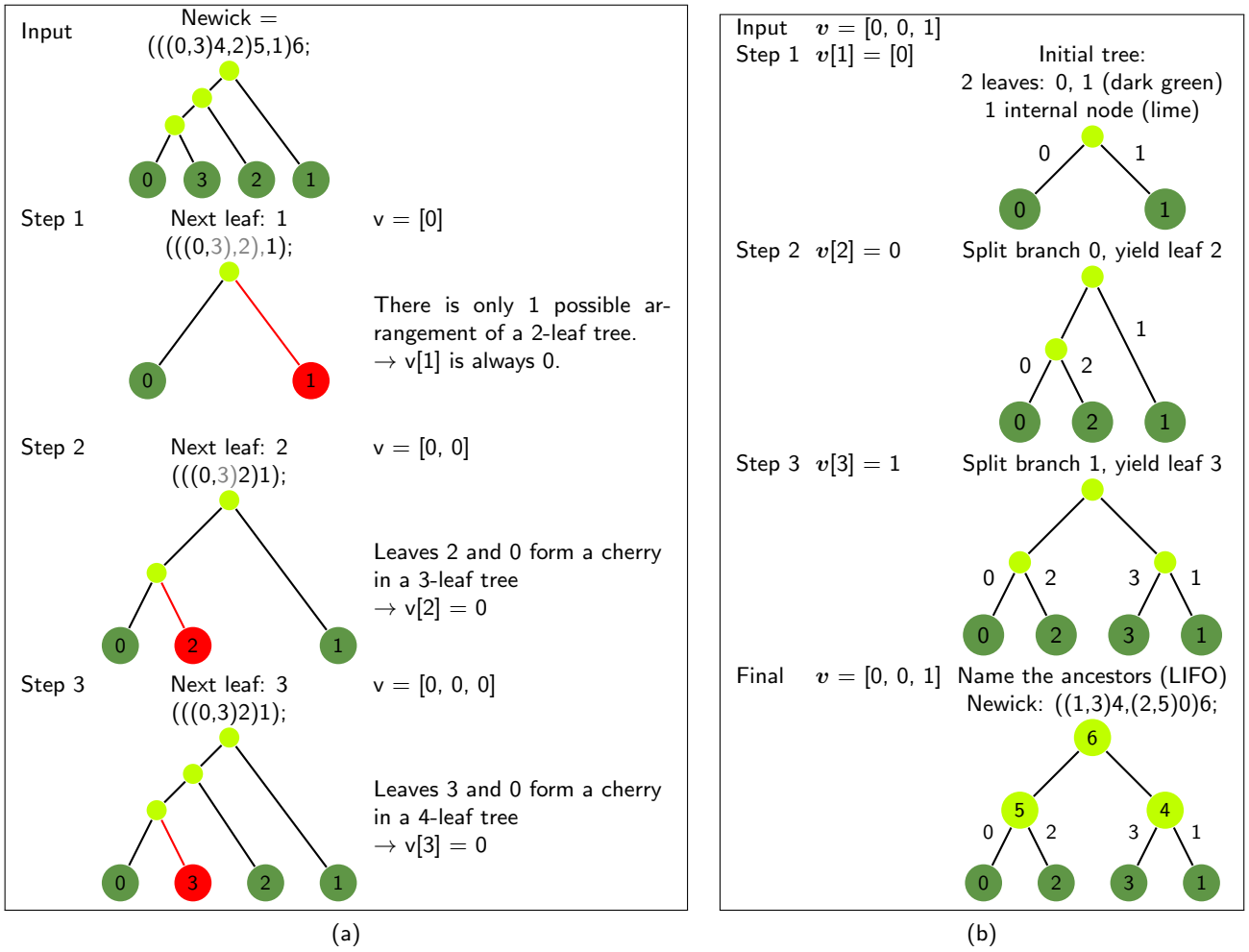


Figure 3.1: An incomplete integer representation of tree topology as birth processes. **(a)** Labelling a tree as an ordered vector: example for $\boldsymbol{v} = [0, 0, 0]$. We process leaves in ascending order. For each leaf j, we retrieve its sibling (or adjacent tip) in the Newick string, ignoring leaves $> j$. The adjacent tip corresponds to $\boldsymbol{v}[j]$. **(b)** Recovering a tree from an ordered vector: example for $\boldsymbol{v} = [0, 0, 1]$. We process $\boldsymbol{v}$ from left to right. Ancestors are named in last-in-first-out (LIFO) fashion: The ancestor of the last added leaf $L - 1$ (here, leaf 3) is named L (here, 4), the ancestor of the second-to-last added leaf $L - 2$ (here, leaf 2) is named $L + 1$ (here, 5) etc. In both cases, the lengths of the edges are arbitrary.

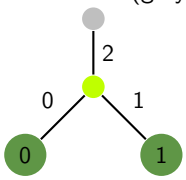
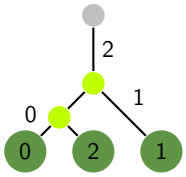
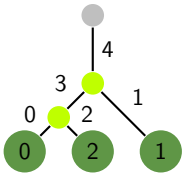
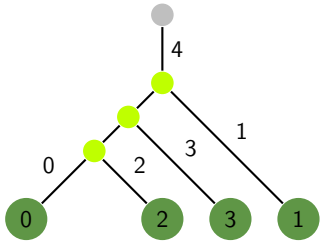
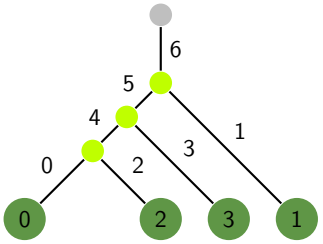
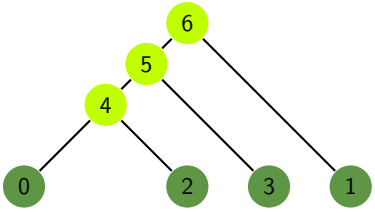
Input	$\boldsymbol{v} = [0, 0, 3]$		
Step 1	$\boldsymbol{v}[1] = [0]$	Initial tree: 2 leaves: 0, 1 (dark green) 1 internal node (lime) 1 extra root (grey) 	
Step 2	$\boldsymbol{v}[2] = 0$	Split branch 0, yield leaf 2 	Rename the branches 
Step 3	$\boldsymbol{v}[3] = 3$	Split branch 3, yield leaf 3 	Rename the branches 
Final	$\boldsymbol{v} = [0, 0, 3]$	Remove the extra root and name the ancestors 	

Figure 3.2: Recovering a tree from a Phylo2Vec vector: example for $\boldsymbol{v} = [0, 0, 3]$. We process $\boldsymbol{v}$ from left to right. The branch renaming step depends on the branch type. Leaf branches: For leaf branches, branches that end on leaves 0, ..., $L-1$ are labelled 0, ..., $L-1$, respectively. For internal branches, the next branch (L) to label is **i)** the deepest and **ii)** with the "highest" children (if there are ties for case 1.). We repeat the same process for internal branches $L+1, ..., 2(L-1)-1$, and label the last branch leading to the extra root $2(L-1)$. See Algorithm 8 and Figure A.2a for more details about implementation and complexity.

existence in the Appendix) an inverse algorithm to convert a tree represented in Newick format as a Phylo2Vec vector in Figure 3.4.

Complexity

The algorithm underlying Figure 3.2 and detailed in Algorithm 6 generally runs in linear time (see Fig. A.2a). A basic version using `NumPy` [285] runs in a few milliseconds for $n = 1000$ taxa on a modern CPU. The inverse algorithm (converting a Newick string to a Phylo2Vec $\boldsymbol{v}$), detailed in Algorithm 7, is of log linear complexity when internal nodes are already labelled (according to the scheme described in Fig. 3.2 Algorithm 8) and quadratic otherwise (see Fig. A.2b). Speedups are made through just-in-time compilation, e.g., using `Numba` [286] in Python.

Distances between trees

The formulation of Phylo2Vec as a one-to-one correspondence between binary trees and integer vectors constrained by Equation 3.2 naturally allows for a new measure of distance between trees. For any two Phylo2Vec trees $\boldsymbol{v}$ and $\boldsymbol{w}$, a Hamming distance can be defined as

$$\mu(\boldsymbol{v}, \boldsymbol{w}) = \sum_{i=1}^{n-1} \mathbb{I}_{v_i \neq w_i} \tag{3.3}$$

To compare this distance with other tree distance metrics, we consider a simple discrete random walk in the space of possible Phylo2Vec vectors $\mathbb{V}$. At each step, we create a new vector $\boldsymbol{w}$ from the previous vector $\boldsymbol{v}$ as follows. First, we choose a random subset of the indices $\mathcal{I} \subseteq \{2 \dots, n-2\}$. For each $i \notin \mathcal{I}$, we set $w_i = v_i$ and for each $i \in \mathcal{I}$, $w_i = \min(2(i-1), \max(v_i + J(i), 0))$ where the $J(i)$ are iid random variables uniform on the set $\{-1, 1\}$. Note that the values of $J(i)$ at different steps of the walk are also independent, and that the minimum and maximum in the definition of w_i ensure that it satisfies the constraint $0 \leq w_i \leq 2(i-1)$.

As $1, n-1 \notin \mathcal{I}$, we fix $w_1 = 0$ (by our constraints) and $w_{n-1} = 2(n-2)$ (to ensure that we move in the unrooted tree space for SPR distance calculations). Figure 3.5 compares μ to an approximate SPR distance [287], Robinson-Foulds (RF) distance [100] and Kuhner-Felsenstein (KF) distance [152]. We note that exact, rooted distance for SPR is NP-hard to compute [288] and therefore cannot be directly compared to our rooted Phylo2Vec formulation. For all distances, we see a nonlinear correspondence, especially for RF and KF distance. Small changes in $\boldsymbol{v}$ can lead to very large topological jumps, but equally, small jumps are also possible. Modifying several indexes in $\boldsymbol{v}$ results in significant jumps across tree space, leading to new trees that are very dissimilar. As a result, SPR, RF, or KF distances saturate as we increase the number of changes in $\boldsymbol{v}$ [289]. However, we note that small changes in $\boldsymbol{v}_i$

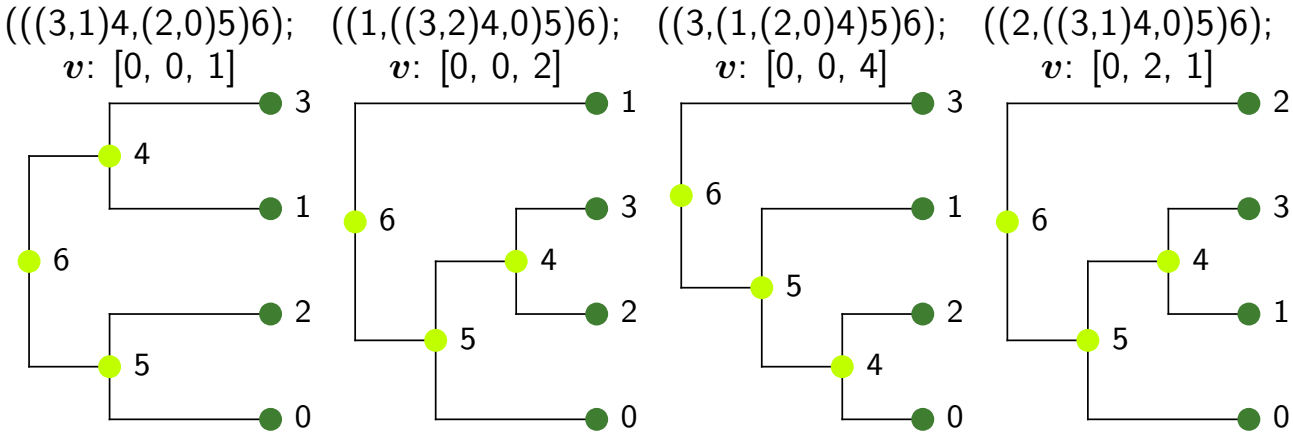


Figure 3.3: Example of trees with $n = 4$ leaves represented in both Newick and Phylo2Vec vector formats. Leaf and internal nodes are coloured in dark green and lime, respectively.

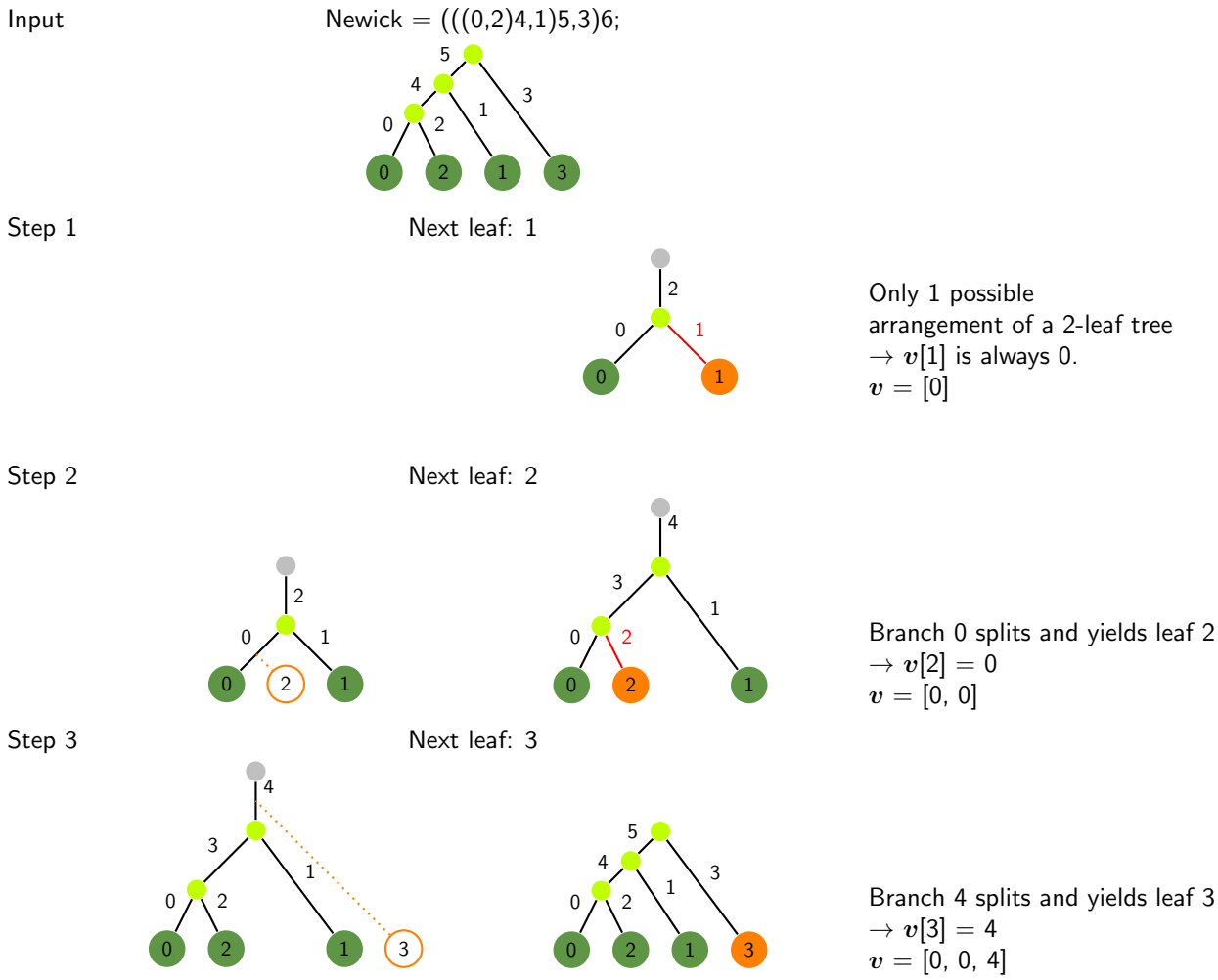


Figure 3.4: Labelling a tree as a Phylo2Vec vector $\boldsymbol{v}$: example for $\boldsymbol{v} = [0, 0, 4]$. We process leaves in ascending order. For each leaf j, we determine the branch that split and yielded leaf j, which corresponds to $\boldsymbol{v}[j]$. At each step, we re-label the branches with the same process as in Figure 3.2.

can also readily correspond to very minor topological changes.

In the exploration of tree space, the number of possible moves for both SPR and Phylo2Vec is of order $\mathcal{O}(n^2)$ (see Phylo2Vec details in the Appendix). Consequently, Phylo2Vec is expected to explore tree space in a similar manner than SPR, with proposals being less local than nearest neighbour interchange but also less global than those by tree bisection reconnection.

Whereas Figure 3.5 shows distances between unrooted trees, our framework is built on rooted phylogenies at its core. Knowing that all rootings produce the same likelihood due to the pulley principle and reversibility of nucleotide substitution models [268], we can, for any rooted phylogeny, switch to one that is rooted at a different outgroup and has exactly the same likelihood. Thus, an equivalence class $\mathcal{V}$ exists where, given a likelihood or parsimony score ℓ, any given Phylo2Vec vector $\boldsymbol{v} \in \mathcal{V}$ has the same $\ell(\boldsymbol{v})$, an SPR or RF distance of 0, but a Phylo2Vec distance of $\mu > 0$. In practice, μ is often very large between $\boldsymbol{v} \in \mathcal{V}$ (comparable to half the maximum SPR distance, see Fig. A.1), which makes switching a vector $\boldsymbol{v} \in \mathcal{V}$ to an equivalent $\boldsymbol{v}' \in \mathcal{V}$ an additional mechanism for tree space exploration.

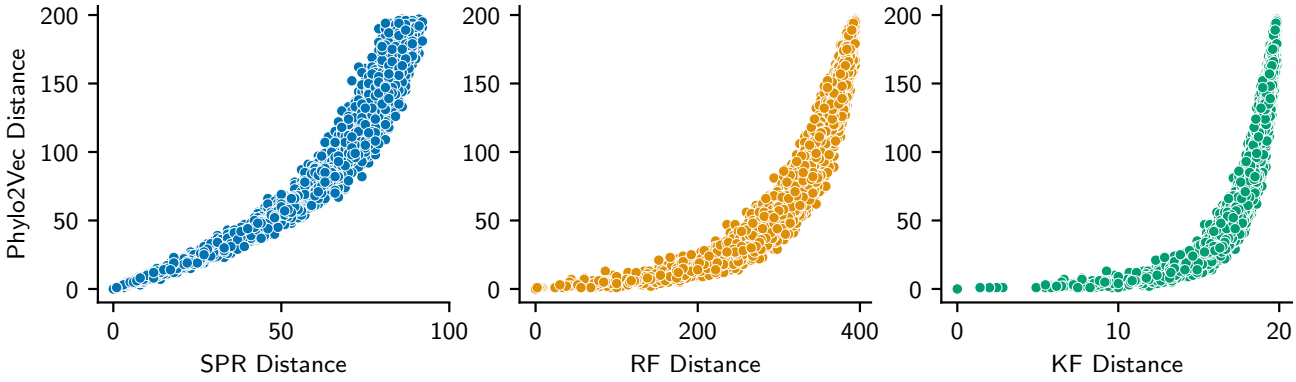


Figure 3.5: Comparison of Phylo2Vec moves with three popular tree distances: subtree-prune-and-regraft (SPR; left), Robinson-Foulds (RF; middle), and Kuhner-Felsenstein (KF; right). To generate the distances, a random walk of 5000 steps was performed from a random initial $\boldsymbol{v}$ with 200 taxa. At each step, each v_i can increment, decrement or remain unchanged.

Shuffling Indices

This distance between two trees is not symmetric with respect to the labelling of the trees, as discussed further in the Appendix (Phylo2Vec details). Depending on the choice of labelling, certain portions of the tree may be easier to optimise than others when performing phylogenetic inference. This is an undesirable quality and can be remedied by a simple reordering of indices within our algorithm. An example of a reordering algorithm is presented in Figure 3.6.

Consider a tree $\mathcal{T}$ where the leaves are labelled by a fixed set of indices $\{0, 1, \ldots, n-1\}$. Suppose that σ is a permutation of $\{0, 1, \ldots, n-1\}$, and consider a shuffled tree $\sigma(\mathcal{T})$ to be a tree with the

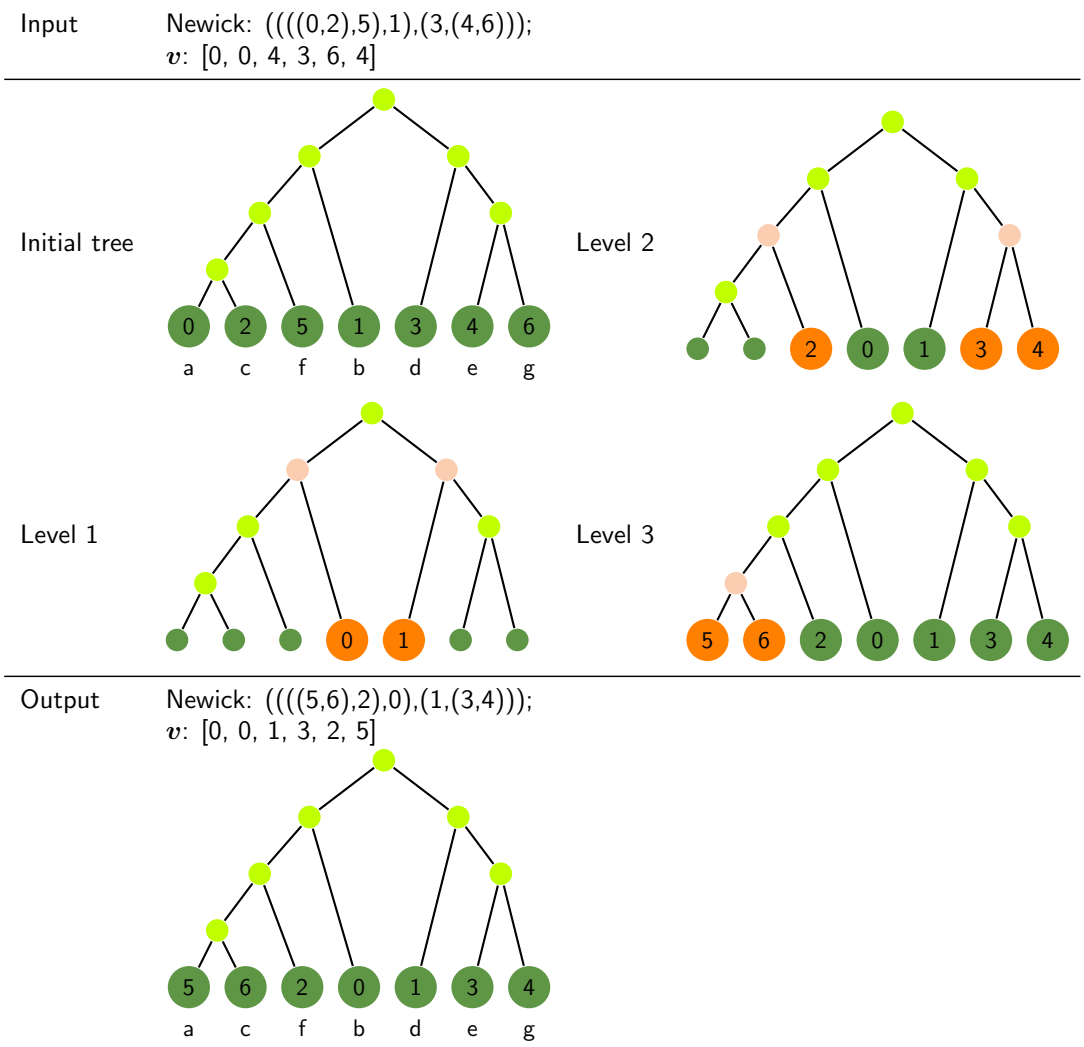


Figure 3.6: Example of a reordering scheme of $\boldsymbol{v}$ using level-order traversal. Starting from the root, for each level, we relabel the immediately descending leaf nodes with the smallest integers available (from 0 to $n-1$; shown in orange). The letters (a-g) indicate the taxa, showing that reordering the leaves does not affect tree topology but simply changes the integer-taxon mapping.

same topological structure as $\mathcal{T}$, but where, for each $j \in \{0, 1, \ldots, n-1\}$, the leaf with original label i now has label $\sigma(j)$.

Calculating the likelihood requires a tree and a set of genetic data $\mathcal{D} = (D_0, D_1, \ldots, D_{n-1})$, where D_j corresponds to the genotype of leaf j as well as tree $\mathcal{T}$. We can then write the likelihood as $L = L(\mathcal{T}, \mathcal{D})$. Moreover, defining the shuffled genetic data as $\sigma(D) = (D_{\sigma^{-1}(0)}, D_{\sigma^{-1}(21}, \ldots, D_{\sigma^{-1}(n-1)})$, we can then see that $L(\mathcal{T}, \mathcal{D}) = L(\sigma(\mathcal{T}), \sigma(\mathcal{D}))$. This occurs because when computing the likelihood, any calculation for $L(\mathcal{T}, \mathcal{D})$ that involves the node with original label i (and hence genetic data D_i) will now involve the node with label $\sigma(i)$ and hence genetic data $D_{\sigma^{-1}(\sigma(i))} = D_i$. Should the permutation only be applied to either the tree labels or the genetic data set, the resulting likelihood will likely be different from $L(\mathcal{T}, \mathcal{D})$. Thus, since the topological structure of $\mathcal{T}$ is the same as $D(\mathcal{T})$, the likelihood will remain unchanged. A more rigorous proof can be found in the Appendix (Phylo2Vec details).

One can also recover the vector $\boldsymbol{v}$ corresponding to the shuffled tree $\sigma(T)$. This is possible because of the bijective relationship between the space of $\boldsymbol{v}$'s and the space of trees. We provide an algorithm that inverts our map from $\boldsymbol{v}$ to M in the Appendix (Phylo2Vec details). Thus, one can equivalently define a shuffled vector $\sigma(\boldsymbol{v})$ (such that $\sigma(\boldsymbol{v})$ generates $\sigma(\mathcal{T})$) and consider the likelihood relationship as $L(\boldsymbol{v}, \mathcal{D}) = L(\sigma(\boldsymbol{v}), \sigma(\mathcal{D}))$. This allows for discrete optimisation steps to be taken with respect to the new shuffled $\boldsymbol{v}$, increasing the flexibility of the algorithm while removing the asymmetric effects of the initial labelling.

Branch lengths

In addition to tree topology, determining the branch lengths of a tree is an important facet in phylogenetic inference. When making small changes to the tree topology, a number of portions of the tree will remain identical and, therefore, it is likely that the optimal values of subtree branch lengths will not change. It is therefore helpful to represent branch lengths in a method that is robust to these changes to avoid carrying out the full optimisation process every time the topology is changed.

Within the Phylo2Vec framework, there are several approaches in which branch lengths can be integrated. First, given each $\boldsymbol{v}_j$ refers to the branch splitting and leading to leaf j, a simple solution would consist in adding a 2-column matrix specifying the position at which branch $\boldsymbol{v}_j$ splits and the length of the new branch yielding leaf j. Alternatively, it is possible to assign each leaf node a "position", calculate internal node positions as some weighted average of the positions of the nearby leaf nodes, and then calculate branch lengths based on the distance between a pair of nodes. This would have the advantage of branch lengths being independent of the choice of root, thus allowing to easily switch between the unrooted equivalence classes discussed previously.

For the examples in this paper, we used `RAxML-NG` [124] to optimise the branch lengths at each step of the algorithm without using information from previous branch lengths. This reduces the speed of the optimisation and is an area for improvement in future work.

3.2.3 Evaluation

Problem and data

To demonstrate the utility of Phylo2Vec, we apply our new representation for phylogenetic inference of five popular empirical molecular sequence datasets under the maximum likelihood (ML) criterion. This dataset corpus spans across different biological entities, taxa, and genetic sequence sizes.

Table 3.1: Evaluation datasets, sorted by number of taxa.

Name	Reference	Type	# taxa	# bases
Yeast	[290, 291]	Fungi	8	127,018
H3N2	[292]	Virus	19	1,407
M501	[293]	Animal	29	2,520
FluA	[294]	Virus	69	987
Zika	[292]	Virus	86	10,807

It has been proved that ML inference for phylogenetic trees is NP-hard [280] and therefore our key goal is to define a sensible heuristic that can explore the vastness of tree space. Moreover, the likelihood surface exhibits high curvature [281] and being trapped in a local optima is a persistent problem across all heuristic phylogenetic approaches.

Tree topology optimisation using hill-climbing

A simple way to explore the space of possible trees is to use hill climbing where we simply compute the difference in likelihood after a single element is changed. We define the neighbour matrix

$$\Delta\ell(\boldsymbol{v})_{ij} = \ell((v_1, \dots, v_{i-1}, j, v_{i+1}, \dots, v_{n-1})) - \ell(\boldsymbol{v}), \tag{3.4}$$

that is, the tree considered in the first likelihood has identical entries except for the i^{th} entry, which is changed to j. For (i, j) such that $v_i = j$ is infeasible, we set $\Delta\ell_{ij} = 0$. We have found that considering each row of the neighbour matrix yields good results, i.e., if $\max(\Delta\ell_i) > 0$, then we find $j = \text{argmax}_j(\Delta\ell_{ij})$ and change the value of v_i to j. This algorithm is guaranteed to converge to a point where $\max(\Delta\ell) \leq 0$ as no change in $\boldsymbol{v}$ results in a gradient that is greater than zero. Moreover, as there are only finitely many possible $\boldsymbol{v}$, and ℓ is strictly decreasing after each iteration of the while loop, the algorithm must converge in finite time. More complicated optimisation algorithms can be

readily created and is an especially useful aspect of our representation. An example is performing hill-climbing over paired changes in $\boldsymbol{v}$. Exploratory analysis suggests that paired changes are far more robust to being trapped in local minima, but at the cost of higher complexity. For challenging phylogenies, a simpler parsimony or minimum evolution score can be used to perform hill-climbing over pairs as an exploratory search.

However, as highlighted above, a fundamental asymmetry exists in Phylo2Vec which can make optimisation inefficient. A simple solution to mitigating this asymmetry is to reorder the integer-taxon mapping to obtain an ordered vector (and thus, an ordered tree), as described previously in An incomplete integer representation of trees as birth processes and Figure 3.6. The advantage of carrying out our hill climbing scheme on these ordered trees is that it removes the secondary effects of changing an element of $\boldsymbol{v}$ which can occur by the divergence in internal node labels. This prevents our algorithm from getting stuck in local minima, as it means that more parts of the tree can be easily edited.

The resulting algorithm is detailed in Algorithm 1. Our investigations have shown that all the possible trees that are one step from some ordered $\boldsymbol{v}$ are also one SPR move from the original tree (though the converse is not true - not all SPR moves will be one step from $\boldsymbol{v}$). This is proved in the Appendix (Phylo2Vec details). Thus, this application of our Phylo2Vec formulation falls within the SPR framework, and provides a mathematically convenient and principled way to explore tree space using well-tested SPR methodology.

We note that we could additionally explore rooted equivalence classes to further prevent being stuck in local minima. In particular, there is more freedom in the movements of nodes further down the tree, and re-rooting at the deepest node would allow all nodes to be easily moved to a variety of locations. However, for the experiments presented hereafter, we found this extra degree of freedom to be unnecessary.

Algorithm 1 Hill-climbing optimisation of a tree with n leaves

Input $\boldsymbol{v} \in \mathcal{T}_n$ ▷ Initialise with a random $\boldsymbol{v}$
$\ell_{\text{best}} \leftarrow \ell(\boldsymbol{v})$ ▷ Initial best likelihood value
repeat
  Reorder($\boldsymbol{v}$) ▷ Reorder the labels (see Figure 3.6)
  Sample $i \in \{1, \dots, n-1\}$ ▷ Sample an index of $\boldsymbol{v}$
  $\boldsymbol{G_i} \leftarrow \Delta\ell(\boldsymbol{v})_{ij}$ ▷ G_{ij} = likelihood difference from changing v_i to j
  $j \leftarrow \text{argmax}(\boldsymbol{G_i})$ ▷ Find the best change
  $v_i \rightarrow j$ ▷ Change v_i
  $\ell_{\text{best}} \leftarrow \ell(\boldsymbol{v})$
until $\max(G_{i=1,\dots,n-1}) = 0$ ▷ Continue iterating until local minimum

Additional properties of the Phylo2Vec vector

An additional advantage of having an integer vector representation for binary trees such as Phylo2vec is efficiency with respect to sampling, data storage, as well as assessing tree equality (with respect to topology). We highlight these properties in Figure 3.7 by performing several benchmarks against functions of shows the widely used R library `ape` [295]. Figure 3.7a shows how Phylo2Vec sampling of trees is several times faster than the function `rtree`, while also being simple in construction and implementation. Figure 3.7b verifies that the Phylo2Vec sampling distribution is indeed uniform. While we do explore other sampling schemes further, ordered trees present one avenue to perform constrained tree sampling. Figure 3.7c shows the storage costs in kB of Phylo2Vec as compared to a Newick string with *only* topological information. From these simple simulations we estimate a Phylo2Vec vector can be stored as an integer array or a string as much as a six times reduced storage cost. Finally, Figure 3.7d shows the time required to find a unique set of topologies from a set of trees. Phylo2Vec is several orders of magnitude faster than `unique.multiPhylo` in `ape`, and can be massively parallelised. This speed difference can be particularly useful in Bayesian settings.

3.2.4 Implementation

All Phylo2Vec algorithms and related optimisation methods presented in the main text were implemented in Python 3.10 using `NumPy` [285] and `numba` [286]. Tree manipulation scripts were written using `ete3` [296]. Dataset construction was based on `phangorn` [291] in R and `TreeTime` [292] in Python. Maximum likelihood estimation was performed using `RAxML-NG` [124]. An implementation is available at: https://github.com/Neclow/phylo2vec. Execution times were benchmarked using Python's `timeit` on a machine equipped with a 64-core CPU @ 7 GHz, with 256 GB of RAM.

3.3 Results

We test Phylo2Vec by performing inference on five popular empirical datasets described in Table 3.1. This dataset corpus spans across different biological entities, taxa, and genetic sequence sizes.

For each dataset, we use the optimisation procedure described in Evaluation, using `RAxML-NG` for branch length and substitution matrix optimisation. We report performance using the negative log version of the tree likelihood defined by [297].

Figure 3.8 shows the optimisation results for four of the datasets described in Table 3.1. We observe that from 10 random starting trees we always achieve the same minimal loss without being trapped in local optima. This is comparable to state-of-the-art software that also searches through topological

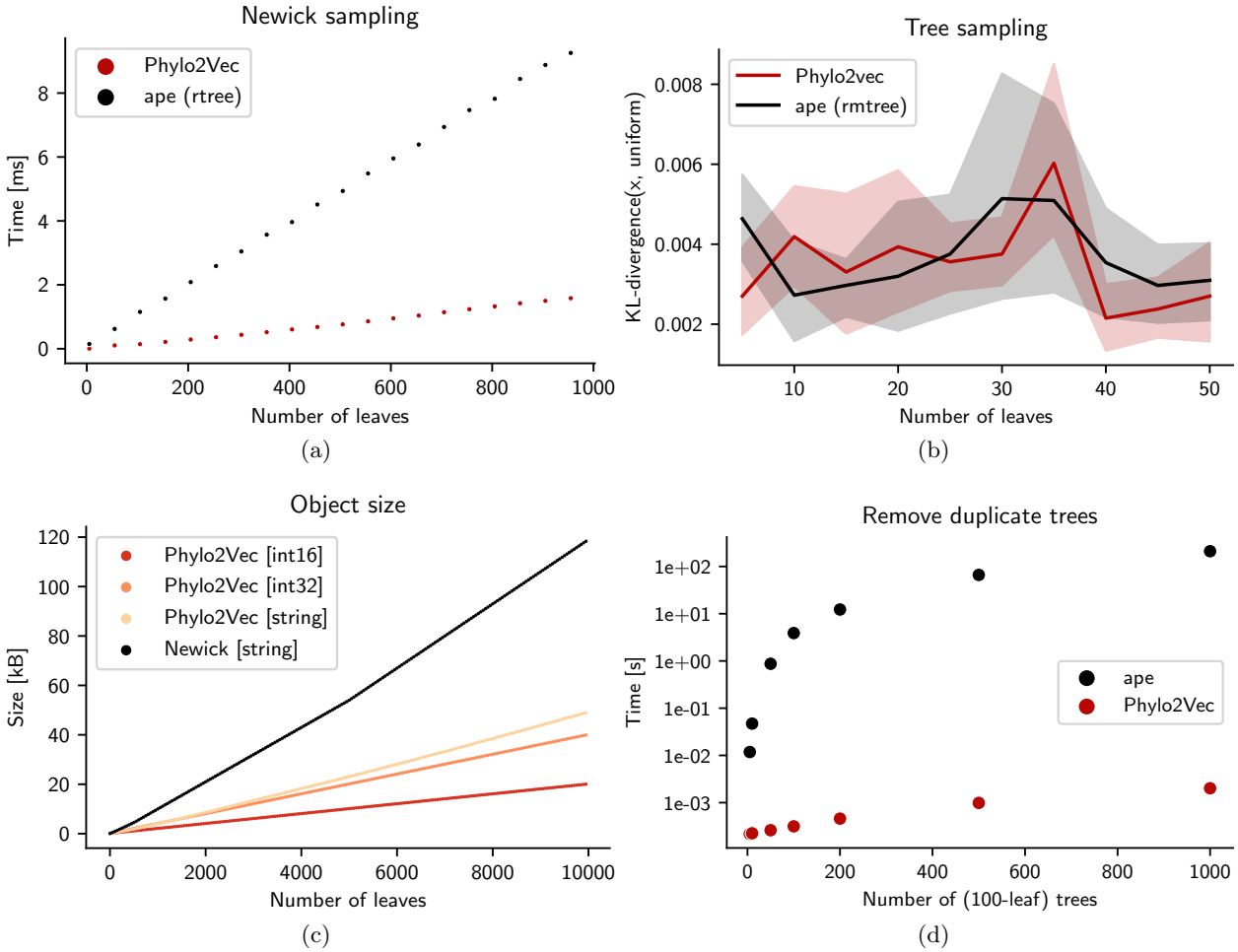


Figure 3.7: Phylo2Vec allows for fast and unbiased sampling, low memory or storage, and fast comparison of trees. **(a)** Average sampling time using `phylo2vec.utils.sample` and `rtree` from `ape`. Execution time was measured over 100 executions using Python's `timeit` and R's `microbenchmark`, respectively. **(b)** Sampling bias comparison. For each size and sampler, we sample 10000 trees and converted them first to their Phylo2Vec representation, and second to an integer using a method similar to that of [55]. We then compare the probability distributions of the integers generated by Phylo2Vec and `ape` sampling against the reference uniform distribution for each tree size using the Kullback-Leibler (KL) divergence. The lower the KL-divergence value, the more the reference distribution and the distribution of interest share similar information. **(c)** Object sizes for different tree sizes of Phylo2Vec vectors (stored as a 16- or 32-bit `numpy` integer array, or a string) compared against their Newick-format equivalents (without branch length information). **(d)** Average time for duplicate removal from a set of trees using Phylo2Vec (vectors) and the *unique.multiPhylo* function from `ape`. Execution time was measured over 30 executions using Python's `timeit` and R's `microbenchmark`, respectively.

space [113, 298]. For each dataset, only two epochs of changes (i.e., two passes through every index of $\boldsymbol{v}$) were generally needed to achieve a minimal negative log-likelihood. In addition, for M501 for example, only a total of around 10,000 likelihood optimisations for each run were needed to reach a minimum - a vanishingly small fraction of the total number of trees possible with 29 taxa ($\sim 8e^{36}$). The choice of the number of optimisations can be shortened depending on the optimisation stoppage criteria, but with the trade-off of being trapped in local minima. We also note that in the Zika virus example, two runs converged at a loss slightly (0.07%) greater than the minimum of the other eight

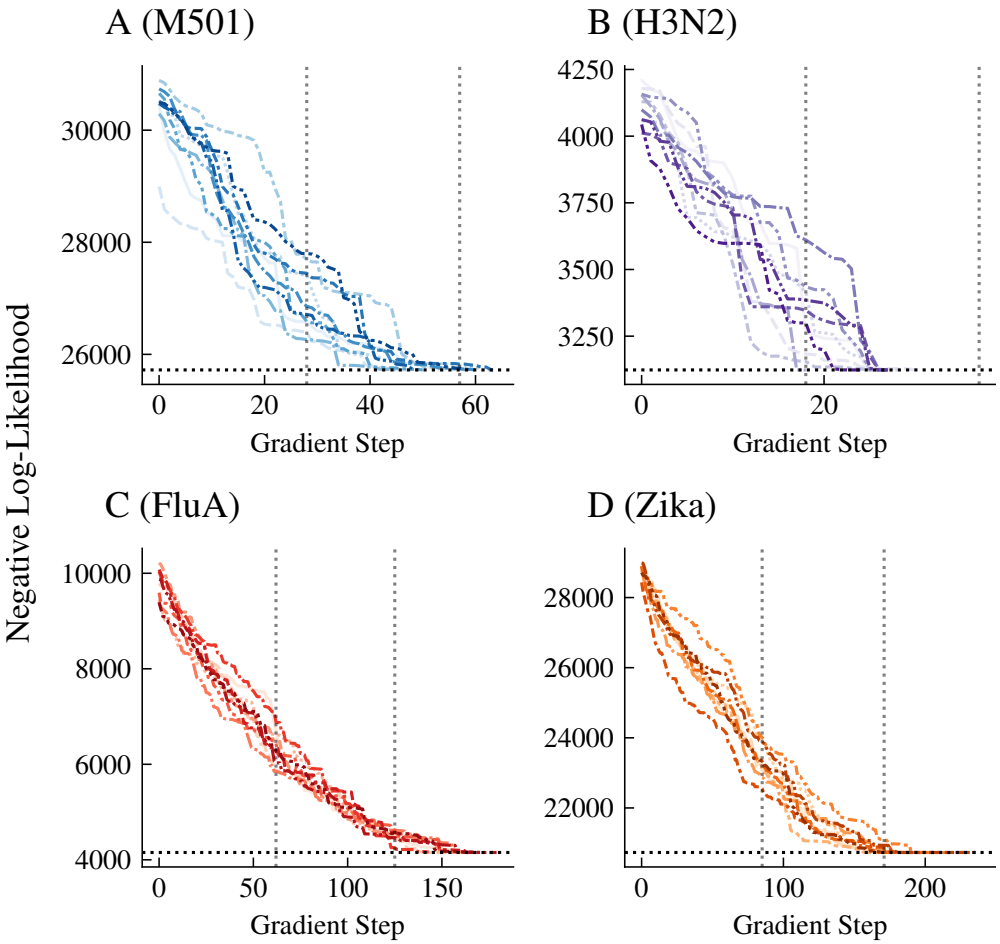


Figure 3.8: Phylo2Vec-based likelihood optimisation results for four datasets described in Table 3.1. The horizontal and vertical lines indicate local minima and epochs (i.e., one pass through every index of $\boldsymbol{v}$), respectively.

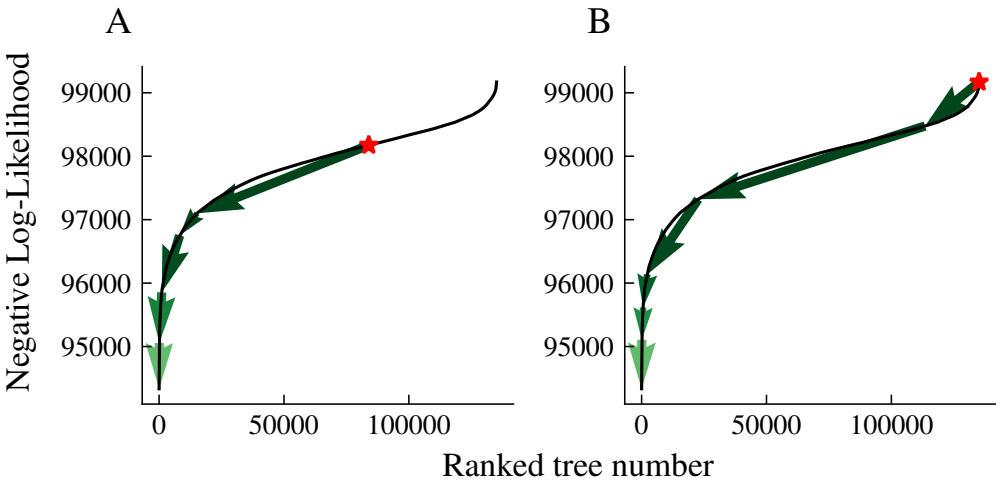


Figure 3.9: Negative log-likelihood path drawn from all possible trees of the Yeast dataset. A and B respectively show the path to the minimum from a random tree and the worst possible tree. The black line shows the sorted phylogenetic likelihoods for all trees. The arrows show the proposal moves for two searches, one from a random tree (A) and one from the worst possible tree (B).

runs. The resultant trees from these minima show that we get trapped in these suboptimal minima due to rooting issues, preventing single changes in $\boldsymbol{v}$ from finding a better optimum. This highlights once again that our algorithm is attempting to solve a more difficult problem than is strictly necessary by searching the space of rooted trees rather than unrooted trees. Due to the pulley principle [268], all rootings of an unrooted tree have the same negative log-likelihood and therefore no paths between rooted trees exist to aid our optimisation algorithm. In practice, especially for large phylogenies, it is common to begin optimisation from a sensible starting point [299] (e.g., a maximum parsimony or neighbour joining tree). In our experiments, we have chosen to start from a completely random tree to highlight the utility of simple algorithms based on Phylo2Vec to traverse tree space.

Subsequently, we apply the same optimisation procedure for the yeast dataset (8 taxa) initially presented in [290] and studied in [300]. Given the smaller number of taxa, we were able to exhaustively calculate the likelihood for every possible rooted tree. As shown in Figure 3.9, we notice a broad region of numerous trees with comparable likelihoods, in addition to a considerably smaller group of trees exhibiting increasing likelihood. Regardless of whether we start from a random tree or the worst possible tree, our algorithm quickly converges to the accurate tree reported in [290]. Across several runs, Algorithm 1 required 96 total likelihood evaluations - a very small fraction of the total number of trees.

3.4 Discussion

Phylo2Vec is a parsimonious representation for phylogenetic trees whose validity extends to any binary tree. This representation facilitates the calculation of distances between trees and allows the formation of a simple algorithm for phylogenetic optimisation. Following from trends in phylogenetics, Phylo2Vec could be integrated within state-of-the-art computing libraries (e.g., `libpll` [301] or `Beagle` [302]) to facilitate its use. We have not yet considered Bayesian inference, but this is likely a useful application of Phylo2Vec, where random walks can be trivially implemented (see Figure 3.5). Furthermore, Phylo2Vec can be useful in assessing topological convergence, for example, for a large phylogeny of 500 taxa and a million trees, extracting the unique set of topologies takes < 10 seconds on a single core in Python, and can be even faster with parallel computation. Although Phylo2Vec does allow for unrooted trees, it is primarily an algorithm for rooted trees. In the examples in this paper, we only consider reversible Markov models where rooting is irrelevant due to the pulley principle [268]. Irreversible Markov models are both mathematically and biologically more principled [303] but require rooted trees. Therefore, a useful application of Phylo2Vec could be in the inference of phylogenies with irreversible Markov models.

The use of empirical datasets served as a proof of concept that maximum likelihood estimation can be performed using Phylo2Vec vectors. We show that, using a simple hill-climbing scheme, we can recover the same topology optimum found by state-of-the-art MLE frameworks such as RAxML-NG [124]. It is important to note, however, that this approach is nowhere near as optimised as RAxML-NG. As it only performs topology changes at a single vector index at a time, its inherent greediness makes inference of large datasets difficult.

That being said, the simplicity of the Phylo2Vec formulation means that it can be used in other more efficient and complex optimisation schemes can also be developed. For instance, Phylo2Vec can also benefit from fast SPR changes [304] and other heuristic optimizations that are currently in RAxML(-NG). In addition, by construction, we have ensured that Phylo2Vec can be differentiable through transforming v into a matrix $W \in \mathbb{R}^{0,1}$ such that $W_{ij} = \mathbb{P}(\boldsymbol{v}_i = j)$. Via this transform, inference in a continuous tree space using gradient descent-based optimisation frameworks is theoretically possible, but its particulars remain to be developed. Similarly, we expect Phlyo2Vec-based representations to be applied in Monte Carlo tree search (MCTS) frameworks which may explore tree space more efficiently, or used as an embedding to regularly infer phylogenetic trees using well-established machine learning paradigms such as self-supervised learning from large existing tree libraries (e.g., TreeBase [305]).

3.5 Data and Code availability

All code relevant to reproduce the experiments is available online: https://github.com/Neclow/phylo2vec. Instructions to access the publicly available datasets are included in the `phylo2vec/datasets` folder of the repository.

Chapter 4: Paper II: Leaping through tree space: continuous phylogenetic inference for rooted and unrooted trees

Matthew J Penn et al. "Leaping through tree space: continuous phylogenetic inference for rooted and unrooted trees". In: Genome Biology and Evolution 15.12 (Dec. 2023), evad213.

Status: This paper has been published in *Genome Biology and Evolution*.

Abstract: Phylogenetics is now fundamental in life sciences, providing insights into the earliest branches of life and the origins and spread of epidemics. However, finding suitable phylogenies from the vast space of possible trees remains challenging. To address this problem, for the first time, we perform both tree exploration and inference in a continuous space where the computation of gradients is possible. This continuous relaxation allows for major leaps across tree space in both rooted and unrooted trees, and is less susceptible to convergence to local minima. Our approach outperforms the current best methods for inference on unrooted trees and, in simulation, accurately infers the tree and root in ultrametric cases. The approach is effective in cases of empirical data with negligible amounts of data, which we demonstrate on the phylogeny of jawed vertebrates. Indeed, only a few genes with an ultrametric signal were generally sufficient for resolving the major lineages of vertebrates. With cubic-time complexity and efficient optimisation via automatic differentiation, our method presents an effective way forwards for exploring the most difficult, data-deficient phylogenetic questions.

Full Author List: Matthew J Penn, Neil Scheidwasser, Joseph Penn, Christl A Donnelly, David A Duchêne and Samir Bhatt

Joint Authorship: M.J.P., N.S., D.A.D. and S.B. have joint authorship of this work.

Author contributions:[1] S.B. and M.J.P. conceived of the study. S.B. and C.A.D. supervised.

[1] As found in the published paper

S.B., N.S., M.J.P., and D.A.D. designed the study. S.B. and N.S. performed optimization runs. S.B., M.J.P., N.S., and D.A.D. performed analysis. All authors contributed to writing the original draft. M.J.P. and J.P. drafted the appendix.

4.1 Introduction

Phylogenetic inference, the task of reconstructing the evolutionary relationships across taxonomic units given observational data, has a wide range of theoretical and practical applications in biology, such as evolution [273, 268, 306], conservation [264] and epidemiology [307, 266], but also in comparative linguistics [308] and cultural anthropology [309, 310]. In particular, the COVID-19 pandemic has catalysed the development of efficient phylogenetic tools and methods to better understand the virus' origin, spread, and evolution [311, 312, 313, 314, 315, 276, 316]. For biological problems, tree inference is primarily informed by molecular sequence data (i.e., nucleotide or amino acid sequences), for which an extensive body of literature exists [317, 318, 120]. Other types of biological data such as morphology [319], fossils [320], and auditory communication in animals [321] can also be used as input.

Two key parameters considered when inferring a phylogenetic tree include the *topology*, the branching pattern that specifies the evolutionary relationships between operational taxonomic units, and *branch lengths*, the amount of evolutionary divergence that occurred between the branching events. A substantial amount of research has been conducted on how to parameterise branch lengths [322, 323], especially through the use of various molecular clocks [324]. Similarly, although to a lesser degree, progress has been made on methods for efficient exploration of the space of tree topologies[113], which is fundamentally challenging due to its combinatorial complexity. Indeed, for n taxa, there are $(2n-3)!!$ possible rooted tree arrangements, where $n!!$ denotes the semifactorial of n — even a small dataset of ten taxa can be enumerated by 34 million unique rooted trees. Moreover, finding the global optimal tree is NP-hard for all major optimality criteria (e.g., maximum parsimony [325], minimum evolution [326], maximum likelihood [280]). Methods such as linear programming [150] or branch and bound [327] can provide exact solutions, but are practically limited to problems with $\approx$ 15 or fewer taxa. To overcome these challenges, the overwhelming majority of state-of-the-art software (e.g., `MrBayes` [142], `PAUP` [328], `BEAST` [143], `PAML` [329], `RAxML(-NG)` [113, 124], `FastME` [56], `IQ-TREE` [114, 298]) rely on hand-engineered search heuristics to perform tree topology optimisation or Bayesian analysis. These are traditionally based on subtree pruning and regrafting (SPR) and tree bisection and reconnection (TBR) operations, which have empirically been shown to be the best available methods for exploring tree topology space [113, 330].

However, such methods still have limitations. First, hill climbing using heuristic approaches necessitates multiple evaluations of the objective function to pick the best move. While these heuristics are still polynomial, exhaustive exploration of single SPR operations is quadratic in complexity, and paired

operations (two sequential SPR changes) are quartic. Second, all the aforementioned tree arrangements are prone to being trapped in local optima and even if a global optimum is found, terraces of trees with identical quality exist [281]. This phenomenon is exacerbated when concatenating multiple genes in supermatrices [290, 331, 332] or when using genomic-scale datasets which require extensive computational resources.

To facilitate a better exploration of tree space, we propose **GradME**, a new direction for tree topology inference which expands the problem space using a continuous rather than discrete parameterisation of a phylogenetic tree. Generally, aside from considering metrics (e.g., distances in tree space) [289, 333, 334, 332], performing topological search in a continuous tree space has rarely been explored (for recent work in hyperbolic spaces, see [335, 336, 337, 338]). Furthermore, very few approaches have made use of gradient-based tree proposals [335, 339, 340, 333]. Although maximum likelihood and Bayesian inference criteria are more popular and generally considered state-of-the-art [142, 113, 143, 298, 328, 302], the GradME framework optimises tree topology under a balanced minimum evolution (BME) criterion [341, 147] using distance matrices as an input. This criterion is well-principled [342] but generally performs worse than likelihood-based [151, 120, 56]. However, the framing of the minimum evolution criterion [342, 343] has been proven to be statistically consistent [344, 268] and has repeatedly shown good (although not state-of-the-art) performance in various settings [152, 153, 154, 56].

To better explore the space of possible trees, we expand the space over which we need to search. Our novel vector representation of a phylogenetic tree, Phylo2Vec [47], has a natural continuous extension, allowing us to improve the ability to search parts of this space. Appealing to a common analogy that casts the optimal tree search problem as finding a needle in a haystack, our approach observes a much bigger haystack, but the hay is in very large bundles, many of whom have a needle, and for these bundles we have access to a (weak) magnet. Providing details to this analogy, the size of the usual phylogenetic haystack with n taxa is $(2n-3)!!$ [284], while we search a much larger haystack of size $(n!)^2$. There are $n!$ bundles in this larger haystack, each of which contains $n!$ trees, but for any tree, 2^{n-1} bundles will contain that tree. Although the proportion of bundles containing a needle shrinks exponentially, we propose a novel approach (Queue Shuffle) that chooses bundles that should be closer to one with a needle. For any given bundle, we also introduce a continuous objective function that can be efficiently traversed using gradient descent approaches (the weak magnet) developed for large-scale machine learning problems [345, 346, 347]. This continuous objective facilitates enormous changes to tree topology in a single step in a direction that improves the objective function. After searching any given bundle using the continuous objective, we use Queue Shuffle, which improves the

switch towards the next bundle to search. This counterintuitive approach offers a new addition to the existing heuristic methods used for topological inference, outperforming the current state-of-the-art but as currently stands with a larger overall complexity than neighbour-joining or other discrete approaches.

4.2 Methods

In the following, we describe GradME, a distance-based method for continuous phylogenetic inference of rooted and unrooted trees using gradient descent. The framework can be divided into three components: i) a continuous tree representation based on Phylo2Vec [47], a bijective integer representation of phylogenetic trees; ii) gradient-based optimisation using a continuous version of the balanced minimum evolution criterion [341, 147], iii) Queue Shuffle, a method to shuffle the integer-to-taxon mapping underlying Phylo2Vec for full tree space exploration. The overall approach works for both rooted and unrooted trees.

4.2.1 Balanced minimum evolution

Popular objective functions to infer the optimal tree from phylogenetic data include maximum parsimony [348], maximum likelihood [297] and minimum evolution [349]. Maximum likelihood and minimum evolution are provably statistically consistent [147, 268], whereas maximum parsimony can be inconsistent under certain conditions [111]. For small to moderate sized phylogenies, methods based on maximum likelihood (and Bayesian extensions) are generally considered state-of-the-art [113, 143, 298, 328]. However, approaches based on minimum evolution (ME) have also shown to yield adequate performance [152, 154, 56, 153, 280]. The first introductions of the minimum evolution (ME) paradigm [342, 343] sought to express evolutionary relationships through dissimilarity. They proved that, given unbiased estimates of the true evolutionary distances, the true phylogeny has an expected length shorter than any other possible phylogeny – thereby establishing the principled ME criterion. Currently, the best performing ME approach is that of balanced minimum evolution (BME) [341, 147], with `FastME` [56] being a popular software implementation. Its objective function can be written as:

$$\mathcal{L}(T) = \sum_{i,j} D_{ij} 2^{-e_{ij}} \tag{4.1}$$

where D_{ij} denotes a distance (e.g., based on molecular sequence data) between two taxa i and j and e_{ij} the number of branches in the path between taxa i and j (the path length [350]). This objective can be computed in a numerically stable fashion using the log-sum-exp trick (see Appendix B.11 for

an example snippet). A widely used approach to estimate the optimal tree greedily [344, 154] is the neighbour-joining (NJ) method [349]. When neighbour-joining is based on an additive distance measure, it reconstructs a unique tree, but still performs well with near-additive trees [351] and under small perturbations in the data [352]. However, despite these highly favourable properties, further heuristic optimisation on a NJ tree using SPR moves have proven to be even more accurate [56]. Once a tree topology is found, quadratic algorithms exist for estimating the branch lengths [353] as well as efficient approaches for molecular clock dating [354].

4.2.2 Balanced minimum evolution for rooted trees

Inference using BME is always restricted to unrooted trees [350, 149] with rooting chosen after inference through heuristics (e.g., midpoint rooting) or via a molecular clock (e.g., for serially sampled data). However, it is often of interest to find the optimal rooted tree for a set of taxa, as this provides extra biological context (e.g., to represent evolutionary paths).

In an unrooted tree, the BME objective function (Eq. 4.1) provides an efficient way of calculating the total length of a tree where the branch lengths are the BME estimators for approximating each D_{ij} with the distance from nodes i to j in the tree. However, this result does not hold in a rooted tree, as the addition of a root changes many of the path lengths. To remedy this, we consider adding a "root taxon" to the tree by joining it to the root node as taxon n. If the tree is roughly ultrametric, then we expect

$$D_{ni} \approx D^* \quad \forall i \neq n \tag{4.2}$$

where D^* is the (assumed constant) root-to-taxa distance. Of course, we do not know the sequence of the root but, as we will show, the value of D^* is unimportant — it is instead simply important that it is independent of i. Adding this root taxon as a leaf node transforms the tree from being rooted to being unrooted, where standard BME can be used. From this assumption, we prove two lemmas to ensure the framework's validity, showing that the optimal unrooted tree is obtained when the variation in the root-to-taxa distance is sufficiently small (Lemma B.1), and subsequently that, in all cases, the optimal rooting for an unrooted tree solves a biologically plausible optimisation problem (Lemma B.2).

First, Lemma B.1 shows that, if

$$|D_{ni} - D^*| < \delta \quad \forall i \neq n \tag{4.3}$$

then, using e^u_{ij} and e^r_{ij} to denote path lengths in the unrooted tree (u) containing taxon n and the

rooted tree (r) formed by removing taxon n,

$$\left| \sum_{i=0}^{n} \sum_{j=0}^{n} D_{ij} 2^{-e^u_{ij}} - \sum_{i=0}^{n-1} \sum_{j=1}^{n-1} D_{ij} 2^{-e^r_{ij}} - D^* \right| \leq \delta \tag{4.4}$$

where δ denotes a small number. Hence, the difference between the rooted and unrooted objective functions is approximately equal to the constant, D^*. Thus, for sufficiently small δ, by the discreteness of tree space, we can see that, if it is unique, the optimal unrooted tree under the *rooted* objective (using e^r_{ij}) will be the same as that under the unrooted objective (using e^u_{ij}), when the root taxon is used as an additional leaf.

Subsequently, Lemma B.2 shows that, for the correct unrooted tree, the BME-optimal rooting maximizes a simple heuristic (defined in Definition 2) for the root-to-tip distance. Equivalently, the optimal rooting ensures that the root is estimated to be the maximal possible distance back in time.

This is not an immediately biologically plausible objective for the root. Indeed, the cornerstone of BME is that we want the tree of *minimum* length, and it hence seems counter-intuitive to require the root that is the *maximum* distance backwards in time (though, by Lemma B.2, this does create the minimum length tree). However, the root of our tree must be the point that is furthest backwards in time. In particular, this means that the evolutionary direction needs to be away from the root. By setting our root such that the root-to-tip distance is maximised, we ensure that the root satisfies this constraint.

This method shares a similar motivation with midpoint rooting, which also seeks to maximise a heuristic for root-to-tip distance. However, the heuristic used in midpoint rooting uses only the two taxa which are furthest apart, while our rooting method uses distances from all taxa. Thus, we expect our method to be more robust, particularly as large inter-taxa distances are difficult to estimate, meaning that the additional information used in our heuristic should help to reduce errors. This is evidenced in Fig. B.2, where midpoint rooting leads to incorrect root placement.

Nonetheless, this property will not hold if the tree is not ultrametric. If taxa evolve at different rates at different times throughout the tree, then the root will be drawn towards taxa with high evolutionary rates. Thus, caution must be used when applying our rooted algorithm to such trees, although the unrooted algorithm will still give a correct unrooted tree. In this case, it may be best to find the optimal unrooted tree topology and then solve the rooting problem for this tree, rather than finding the optimal rooted tree, as the former will reduce the skewing effect of the heterogeneity in evolutionary rates. Because of this, the algorithm introduced in this paper has the flexibility to find the optimal unrooted or rooted tree.

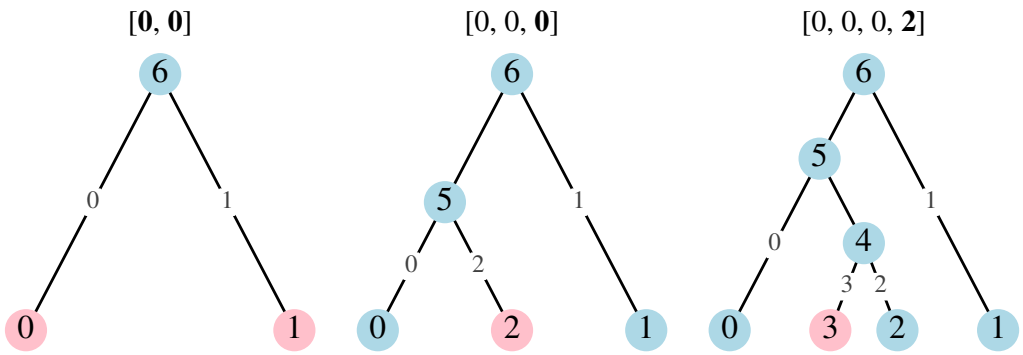


Figure 4.1: An example of the left-to-right construction of the ordered tree $v = [0, 0, 0, 2]$. We begin with two leaf nodes and two edges labelled 0 and 1, then append node 2 to its label edge 0, creating a new internal node and a pair of new edges. The new edge joining node 2 to the tree is labelled as edge 2. We then append node 3 to edge 2, again creating a new internal node and two new edges.

4.2.3 An ordered bijection to tree space

Previously, we introduced Phylo2Vec [47], a novel bijection between the space of phylogenetic trees and a space of integer vectors. In contrast to other bijections such as permutation matchings [277], changes in Phylo2Vec correspond to smooth changes in the tree space, e.g., single changes in a Phylo2Vec vector correspond to a limited set of SPR changes. On the other hand, Prüfer codes [355] form a bijection to the space of all m-ary trees, meaning that there is no guarantee to sample binary trees from random Prüfer sequences.

Here, we focus on the notion of *ordered trees* from [47], where it is possible to construct a tree from its vector in linear time. An ordered tree can be thought of as a birth process, such that when a birth occurs, the original node continues to live and retains its label, while the new node receives an incremented label. Accordingly, we introduce[2] an equivalent but more intuitive tree construction process for these ordered trees (see Fig. 4.1 for an example). We begin with two leaf nodes and two edges labelled 0 and 1. We then append nodes by joining them, *in order*, to edges connecting leaf nodes to the tree. This tree construction process can be summarised by a single vector $\boldsymbol{v}$, with[3] $v_0 = v_1 = 0$ and, for $m \geq 2$, v_m being the label of the edge to which node m is appended. Each index in v_m is subject to the simple constraint:

$$\begin{aligned} &v_0 = v_1 = 0 \quad \text{and} \\ &v_m \in \{0, 1, ..., m-1\} \quad \forall \quad m \geq 2 \end{aligned} \tag{4.5}$$

[2]When this paper was published, it referenced an older version of Paper I than the one found in this thesis. In this original version, we used a different, slower, and less intuitive tree construction method (though it resulted in the same topologies). After publishing Paper II and receiving feedback on Paper I, we then updated Paper I with a new construction method based on the one presented here in Paper II.

[3]Another minor difference between the Phylo2Vec representation here and in Paper I is that here we include a fixed $v_0 = 0$ entry, alongside $v_1 = 1$.

which is equivalent to the definition of ordered trees in [47]. Intuitively, a tree is ordered if, starting with a branch connecting the root to taxon 0, the taxa can be added in order of their label by appending a new branch to a *terminal* branch of the existing tree (i.e. a branch connecting a leaf node to the rest of the tree). Thus, the ordering of the taxa is in some sense "natural" for each possible ordered tree. It is proved in Lemma B.3 that for ordered $\boldsymbol{v}$, the algorithm presented above and the more general Phylo2Vec algorithm in [47] produce equivalent trees.

Note that, for a fixed integer-taxon labelling, the number of ordered trees is a subset of the number of possible trees. We discuss an efficient method to remedy this problem and explore all tree space in Appendix B.5.

4.2.4 A continuous representation of a tree

We introduce a continuous, probabilistic, representation of trees using a square matrix W which gives the distribution of a random ordered vector $\boldsymbol{v}$ with independent entries such that $W_{ij} = \mathbb{P}(v_i = j)$. Given Eq. 4.5, W is a lower-triangular, stochastic matrix (row sums to 1). Thus, W can probabilistically represent any *ordered* phylogenetic tree (a space of $(n-1)!$ trees). A simple approach to determining the most likely single tree from W is to take the column-wise argmax, yielding a single tree $\boldsymbol{v}$.

4.2.5 Gradient-based optimisation using the BME criterion

Using this continuous representation, we can find the optimal ordered tree. Defining $f(\boldsymbol{v})$ to be the BME objective function for the tree generated by an ordered vector $\boldsymbol{v}$, we then create a continuous objective, $F(W)$ by $F(W) = \mathbb{E}\big[f(\boldsymbol{v})\big]$.

The calculation of $F(W)$ follows our new method of constructing ordered trees from the vector $\boldsymbol{v}$. For a fixed (randomly chosen) tree, we define e^k_{ij} to be the path length between nodes i and j when nodes $0, 1, ..., k-1$ have been added to the tree (for $i, j < k$). Note that to find the rooted objective function, we initialise with $e^2_{10} = e^2_{01} = 2$, while to find the unrooted objective, we initialise with $e^2_{10} = e^2_{01} = 1$ (while the Phylo2Vec representation is an inherently rooted representation, this unrooted objective finds the tree length if the root were removed from the random rooted tree with distribution given by W). This is because, in a tree where the only leaf nodes are 0 and 1, these nodes are a path length of 2 apart if there is also a root (as the root is on this path) while otherwise, they are a path length of 1 apart.

If node k is appended to the edge joining either node i or j to the tree, then $e^{k+1}_{ij} = e^k_{ij} + 1$; otherwise, $e^{k+1}_{ij} = e^k_{ij}$. Similarly, if node k is appended to the edge joining node i to the tree, then

$e_{ik}^{k+1} = 2$ and otherwise, if node k is appended to the edge joining node x to the tree, then $e_{ik}^{k+1} = e_{ix}^{k} + 1$. Thus, using V_k to denote the random value of v_k, we can write

$$e_{ij}^{k+1} = e_{ij}^{k} + G_{ij}^{k}(V_k) \tag{4.6}$$

$$\Rightarrow e_{ij}^{n-1} = \sum_{k=2}^{n-2} G_{ij}^{k}(V_k) + e_{ij}^{2} \tag{4.7}$$

for some functions G_{ij}^{k} which are derived explicitly in Lemma B.4. Importantly, each term in the sum is independent, and hence, in Lemma B.4, a closed iterative system for the quantities $E_{ij}^{k} = \mathbb{E}(2^{-e_{ij}^{k}})$ can be calculated for $i < j$ as

$$E_{ij}^{k+1} = \begin{cases} E_{ij}^{k}\left[1 - \frac{1}{2}(W_{ki} + W_{kj})\right] & \text{if } i < j < k \\ \left[\frac{1}{2}\sum_{x \neq i} E_{ix}^{k} W_{kx}\right] + \frac{1}{4} W_{ki} & \text{if } i < k \end{cases} \tag{4.8}$$

with the remaining values for $i > j$ following by symmetry.

The objective function is a polynomial function of the entries of W and is linear in each fixed entry (that is, the diagonal entries of $\nabla^2 F$ are zero). Thus, by Lemma B.5, there is always a minimum at a "discrete tree" (that is, at a matrix W where for each row, one value is 1 and all the others are 0). Moreover, this simple form makes it easy to differentiate F analytically, numerically or automatically. Using state-of-the-art automatic differentiation [356], gradient descent can be used to efficiently minimise F and find the optimal *ordered* tree.

There may also be minima at non-discrete trees if multiple trees share the same, optimal, objective value. In our rooted optimisation, this is highly unlikely (as two topologically different trees having equal objectives places a dimension 1 condition on the distance matrix D, meaning the set of distance matrices for which this happens has measure 0 in the set of possible distance matrices), but when we use our unrooted algorithm, this will occur. This is because, as discussed in [47], there are $n - 1$ Phylo2Vec vectors which, when the root is removed, give the same unrooted tree. Thus, if multiple rooted vectors giving the same unrooted tree $\mathcal{U}$ are in the same space of ordered trees, then, if $\mathcal{U}$ is the optimal tree under this ordering, the algorithm may converge to a non-discrete W. In this case, taking the argmax safely recovers an optimal rooted tree as, by Lemma B.5, all possible trees according to W will have the optimal objective value.

The tree-space induced by our continuous objective function will have local minima whenever changing any single entry of the vector $\boldsymbol{v}$ causes the objective function to increase. As the Hamming distance between vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ is comparable to SPR distance [47], we therefore expect that the discrete subset of our continuous tree-space will be similar in structure to the space induced by SPR

moves. However, by starting from a uniformly distributed tree, where all possible ordered trees contribute to the objective, we expect that our algorithm is better able to pick up the "signal" from the true optimum, and avoid moving towards suboptimal local minima.

A Python-like algorithm to compute $\mathbb{E}(2^{-e})$ is shown in Algorithm 2. To find the E^m_{ij} terms, there are $\mathcal{O}(m)$ steps of $\mathcal{O}(m)$ (finding the $E^m_{m-1,j}$ terms) and $\mathcal{O}(m^2)$ steps of $\mathcal{O}(1)$ (finding the other E^m_{ij} terms). There are $\mathcal{O}(n)$ values of m that need to be considered, and hence this system can be solved in $\mathcal{O}(n^3)$ time.

Algorithm 2 Compute $E := \mathbb{E}(2^{-e})$

Require: n ▷ fixed number of taxa
Require: W ▷ stochastic input matrix[47]

$E = \text{zeros}((n, n))$
for $i \leftarrow 1$ to $n-1$ **do**
  $E^* = \text{zeros}((n, n))$
  for $k \leftarrow 0$ to $i-1$ **do**
    for $j \leftarrow 0$ to $k-1$ **do**
      $E^*_{[k,j]} = E_{[k,j]}(1 - \frac{1}{2}(W_{[i-1,j]} + W_{[i-1,k]}))$
    end for
    $E^*_{[i,k]} = \sum_{l=0}^{i-1} \frac{1}{2} E_{[l,k]} W_{[i-1,l]} + \frac{1}{4} W_{[i-1,k]}$
  end for
  $E = E^* + E^{*T}$
end for
Return E

4.2.6 Orderings

The continuous objective function is only defined for ordered trees, which, for a given labelling of the nodes, is a subset of the whole tree space. Thus, we define the concept of an *ordering* of the nodes, whereby changing the ordering allows the full space of trees to be explored.

Definition 1 ***Ordering:*** *Suppose that the nodes correspond to taxa with names $N_0, N_1, ...N_{n-1}$. We then define an ordering of the nodes to be a permutation, σ of the set $\{0, ..., n-1\}$ such that the node with name N_i is processed as node $\sigma(i)$ by the Phylo2Vec algorithm. It is necessary that the associated Phylo2Vec vector $\boldsymbol{v}(\sigma)$ is ordered.*

Note that we use the phrase "node x" to mean "the node with name N_x" and will always be explicit if we refer to a node by its label.

Given a tree, it is possible to generate a possible ordering of the nodes, σ, as well as the associated vector $\boldsymbol{v}(\sigma)$. We will do this by labelling the tree as follows, again distinguishing between the leaf node names N_i and their labels (which we will call $l(i)$).

Consider labelling the root node as 0. Then, label the children of this node as 0 and 1. One can continue this process inductively, choosing a node, labelled x, with unlabelled children and then labelling its children as x and y, where y is the smallest unused label. This process terminates when every node has been labelled. As discussed above, suppose that the label of the leaf node with name N_i is $l(i)$. Then, Lemmas B.6, B.7 and B.8 show that l is a possible ordering of the tree, and provide a method for calculating the associated ordered Phylo2Vec vector $\boldsymbol{v}(l)$.

Given a tree, one can use the previous labelling algorithm to show that there are at least 2^{n-1} possible orderings for the tree (as the children of each node can be labelled in either order). The exact number depends on the choice of which internal node is processed at each step (and so, a "balanced" tree where the leaves have a low generation has more possible orderings than a "ladder" tree where each leaf has a distinct generation). However, in general, it will be substantially larger than 2^{n-1} (we conjecture that for flat trees, it may be factorial in size) and hence there are numerous possibilities which will allow for a global minimum of the objective to be found.

4.2.7 Queue Shuffle: changing orderings to explore all the tree space

The number of ordered trees (different from ranked trees [125]), $(n-1)!$, is substantially smaller than the possible number of trees, $(2n-3)!!$ (albeit with a comparable growth pattern in n), and hence, while the optimal ordered tree will be closer to the true tree than a very large proportion of trees, it is very unlikely to be exactly equal to the true tree.

To fully explore tree space, one must shuffle the labels of the leaf nodes in the optimal ordered tree. Simply choosing a uniformly random permutation will lead to extremely inefficient optimisation, as each tree is only possible in approximately $1/2^n$ of the possible orderings. Instead, we use the topology of the optimal tree to inform our choice of permutation through a novel approach we call the *Queue Shuffle*. This ensures that the previous optimal tree can be written as an ordered tree in the new ordering, while also ensuring a smooth and efficient path through the space of orderings.

The Queue Shuffle is motivated by the labelling procedure discussed in the previous section, but ensures that the set of internal nodes with a given generation (that is, a given distance from the root) are processed consecutively. That is, we begin by processing all nodes with generation 0 (i.e., the root), then all internal nodes with generation 1, then all internal nodes with generation 2, and continue in this fashion until all internal nodes have been processed.

Algorithmically, this can be achieved by a "queue" of internal nodes to be processed. When an internal node is processed, any of its children that are also internal nodes are added to the back of

this queue. Thus, the queue is always in ascending order of generation, and it is simple to show that this ensures that nodes are processed in non-decreasing order of generation.

A crucial feature of this queue is that the child given the same label as its parent is placed *ahead* of the other child in the queue. This ensures that one can, in some way, control the order of processing by choosing the labelling of the children of each node. Moreover, it is vital for the theoretical result presented in the following section.

To add randomness into the labelling procedure, every time an internal node is processed, we randomly choose which child is given the label of their parent, and which child is given the next available label. This provides 2^{n-1} possible orderings for each tree. This stochasticity is helpful in ensuring that the algorithm does not get stuck – as discussed in the subsequent section, it ensures that a large class of similar trees will be considered after a few ordering proposals.

An algorithmic description of the Queue Shuffle is provided in Algorithm 3.

Algorithm 3 The Queue Shuffle

Require: $\mathcal{T}$ ▷ Current Tree
Require: $\mathcal{N} = \{\nu_0, \nu_1, ...\}$ ▷ set of all non-root nodes
$Q = [\nu_0, \nu_1]$ ▷ "queue" of nodes to process
$L = \{\nu_0 : 0, \nu_1 : 1\}$ ▷ node:label mapping
$l_{\text{next}} = 2$ ▷ next available label
$P = []$ ▷ processed nodes
while $Q \neq []$ **do**
    $\nu = Q[0]$ ▷ node to process
    Q = Q[1:] ▷ ν will be processed
    append(P, ν) ▷ ν will be processed
    if isLeaf(ν) **then**
        continue ▷ move to next node
    end if
    $a, b =$ randChildren(ν) ▷ get **randomly ordered** children of ν
    $L[a] = L[\nu]$ ▷ label a with ν's label
    $L[b] = l_{\text{next}}$ ▷ give b next available label
    $l_{\text{next}} = l_{\text{next}} + 1$
    append(Q, a) ▷ add a to the queue
    append(Q, b) ▷ add b to the queue
end while
Return L ▷ Ordering determined by values of L for leaf nodes

4.2.8 GradME

The Queue Shuffle completes our optimisation algorithm. We iteratively find the best ordered tree according to the current ordering and then use Queue Shuffle to change ordering, changing the space of explorable trees. The algorithm terminates when the optimal tree has not been improved upon for

a fixed number of iterations (note that, by construction, the previous optimal tree will always be in the new space of ordered trees). In the examples presented in this paper, only tens of iterations are needed from some random starting ordering, and less if a sensible starting ordering (such as from a neighbour-joining tree) is used.

We refer to the resulting system, combining the continuous tree representation, Queue Shuffle reordering, and the gradient-based optimisation framework using BME, as **GradME**.

4.2.9 Why does Queue Shuffle work?

A given tree is in the space of ordered trees for at least 2^{n-1} orderings. This means that we do not need to find a single optimal ordering, but have exponentially many which will return the true optimal tree. Very loosely considered, being able to explore $n!$ tree space reliably and efficiently with continuous optimisation, Queue Shuffle reduces the inferential task to one that is exponential.

However, while the number of optimal orderings grows exponentially, their proportion tends quickly to zero as n grows. It is therefore, perhaps, surprising that we are able to find an optimal ordering so quickly from merely tens of shuffles. The proportion of optimal orderings (approximately the ratio of ordered trees to total trees, $\frac{(n-1)!}{(2n-3)!!}$, ranges from 8×10^{-4} in our smallest dataset (14 taxa) to 6×10^{-29} in the largest (99 taxa; see Table 4.1).

This efficiency comes from the topology-dependence of the Queue Shuffle algorithm, which allows us to plot a relatively "smooth" path through the space of possible orderings. That is, the majority of trees in the new ordered space will have similar properties to the previous optimal tree and so, unless the previous tree was a local minimum of the objective, it is likely that one of these "close" trees will have a lower objective value.

Lemma B.9 shows that the expected distance from the root grows harmonically as the label increases. For large trees, the node with label $n-1$ has an expected distance from the root of approximately twice the expected distance from the root of the node label 0. This property is noticeable even for small trees – if $n = 10$, then the ratio of the expected distance to the root of node 9 and node 0 is approximately 1.65. Thus, nodes which are close to the root in the current optimum, will also be closer on average to the root in the new space of ordered trees. In essence, this means that "fewer slots are wasted" in the new ordered space – that is, there are fewer trees in the new space of ordered trees which are topologically far from the previous optimum (a tree that, after the first few iterations, is likely to be far closer to the true optimum than a randomly-chosen tree) and hence, more trees which are reasonable candidates for having lower objective values.

Lemma B.1 proves another example of the smoothness in transitions induced by the Queue Shuffle,

based on nearest neighbour interchange (NNI) moves. An NNI move considers the four subtrees attached to two non-root nodes that share an edge and swaps two of these subtrees. Lemma B.1 shows that, starting from a tree $\mathcal{T}$, any tree which is one NNI move away from $\mathcal{T}$ will be in the new space of ordered trees with probability at least $\frac{1}{4}$.

This ensures that this new space contains many sensible proposal trees. Perhaps the most surprising aspect of this result is that this probability is bounded below, independently of the topology. Thus, with high probability, the optimal tree will only remain the same for more than a few iterations if large sets of similar trees yield lower objective values than the current optimum.

That being said, the Queue Shuffle does not guarantee that the global minimum will be found, even if the gradient-based algorithm for optimising $F(W)$ always converges to the optimal tree. If a tree is "far" from the nearest tree with a better objective value, then it may take a very large number of shuffles (or, indeed, it may be impossible) to find a better tree. However, while the only theoretical guarantee is that Lemma B.1 shows it will quickly find better trees that can be formed by NNI, we expect that stronger conditions hold on its ability to "escape" from local minima.

Table 4.1: Evaluation datasets. rRNA/rDNA: ribosomal RNA/DNA, mtDNA: mitochondrial DNA. AA: amino acid. For the Jawed dataset, several subsets of the original dataset [357] were used (from 1,460 to 18,406 sites; cf. Fig. 4.2c).

Dataset	Reference	**# Sites**	**# Taxa**	**Type**	**Taxonomic rank**
DS1	[358]	1,949	27	rRNA (18S)	Tetrapods
DS2	[293]	2,520	29	rRNA (18S)	Acanthocephalans
DS3	[359]	1,812	36	mtDNA	Mammals; mainly Lemurs
DS4	[360]	1,137	41	rDNA (18S)	Fungi; mainly Ascomycota
DS5	[361, 362]	378	50	DNA	Lepidoptera
DS6	[363]	1,133	50	rDNA (28S)	Fungi; mainly Diaporthales
DS7	[364]	1,824	59	mtDNA	Mammals; mainly Lemurs
DS8	[365]	1,008	64	rDNA (28S)	Fungi; mainly Hypocreales
DS9	[366]	955	67	DNA	Poaceae (grasses)
DS10	[367]	1,098	67	DNA	Fungi; mainly Ascomycota
DS11	[368]	1,082	71	DNA	Lichen
Eutherian	[369]	1,338,678	37	DNA	Eutherian Mammals
Jawed	[357]	1,460-18,406	99	AA	Gnathostomata (jawed vertebrates)
Primates	[370, 299]	232	14	mtDNA	Mammals; mainly Primates

4.2.10 Computational complexity

The computational complexity for all distance-based algorithms requires an upfront computational cost of $\mathcal{O}(n^2)$ to compute the distance matrix. We will disregard this cost from subsequent comparisons. The standard neighbour-joining algorithm [349] has an overall computational complexity of $\mathcal{O}(n^3)$. FastME has a computational complexity of $\mathcal{O}(kn^2_{\text{Diam}}(T))$ (where $\text{Diam}(T)$ is the maximum path

length in a tree, which is generally much smaller than n) for k iterations where $k < n$ when n is large. When fully discrete, our algorithm also has the same complexity but with added mechanisms for escaping optima via Queue Shuffle. Therefore, a discrete setting is as computationally efficient as FastME (see Appendix B.9 for details).

Computing the expectation in Algorithm 2 has complexity $\mathcal{O}(n^3)$. A single gradient evaluation (that is, calculating $\frac{\partial F}{\partial W_{ij}}$ for some i and j) is also $\mathcal{O}(n^3)$ and therefore computing the full Jacobian is $\mathcal{O}(n^5)$. Our Queue Shuffle algorithm runs in $\mathcal{O}(n)$. Therefore, our optimisation for k steps and l shuffles yields a complexity of $\mathcal{O}(kln^5)$. The size of l is dependent on the choice of gradient optimiser, and the size of k varies if a sensible ordering is initialised.

Thus, the computational complexity of GradME is substantially higher than that of FastME and closer to that of FITCH [371]. This is due to the far greater mathematical complexity of the continuous objective function, $F(W)$. As it is an expectation over all possible ordered trees, the explicit formula for $F(W)$ is a polynomial in W with $(n-1)!$ different terms. Intuitively, the continuous space always considers a path between any two trees, something that becomes impossible with discrete settings. Thus, being able to compute it in polynomial time is a vast improvement on a naive approach, although it is still considerably less than the innovative FastME greedy approach. More savings should be possible, we hope to make further efficiency gains in future work.

4.2.11 Evaluation

We evaluate GradME on a diverse corpus of 14 empirical molecular sequence datasets (Table 4.1). The first 11 are commonly used to assess phylogenetic inference performance [372], whereas the last three were used to assess inference on rooted trees. For each dataset, we start from a random tree and optimise the W matrix to a tolerance of 1e-10 using gradient descent with Adafactor [347] optimisation. The distance matrix D is computed using the GTR+Γ substitution model for DNA and an LG model [373] for amino acids. Substitution model parameters for the GTR+Γ are also estimated using gradient descent with Adafactor using a pairwise maximum likelihood approach [318]. Jukes-Cantor [374], F81 [375] and TN93 [376] models were also tested for DNA, while stochastic gradient descent (SGD), RMSprop [377], and AdamW [345, 346] were also considered for optimisation (see Fig. B.3). To fairly assess the performance of GradME, we compare our framework to two well-established distance-based methods: `BioNJ` [378], based on the neighbour-joining algorithm [349], and `FastME` [56], based on balanced minimum evolution.

4.2.12 Implementation

Implementation of the BME criterion and the optimisation framework was written in Python using `Jax` [356] and `Optax` [379]. Optimisation was performed on a Xeon 2.30GHz (CPU; Intel Corporation) or on a single GeForce GTX 1080 (GPU; Nvidia Corporation). Evaluation of the `BioNJ` [378] and `FastME` [56] methods were performed via the R package `ape` [299] using `rpy2` [380]. Tree manipulation and visualisation scripts were written using `ete3` [296] and `NetworkX` [381]. An implementation is available at: https://github.com/Neclow/GradME.

4.3 Results

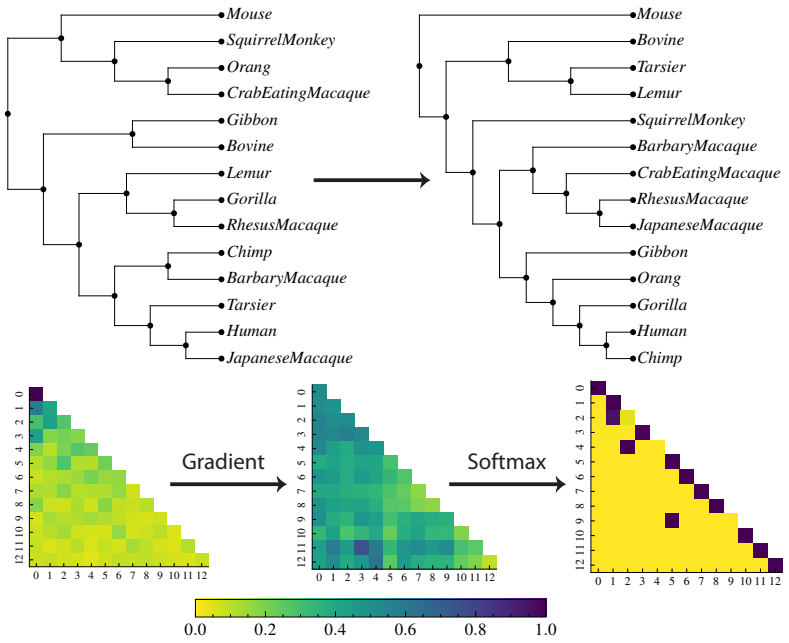
4.3.1 Tree traversal in continuous space

For any choice of label ordering, our approach admits a continuous gradient across $n!$ trees for n leaves. This gradient, which can be obtained readily via automatic differentiation, can rapidly traverse tree space to find trees with a close to optimal objective value. Fig. 4.2a shows a single gradient step for the small Primates dataset [299, 329]. Simply subtracting the gradient from a random initial tree, followed by softmax activation, results in an almost discrete W which corresponds to the best BME tree for a given substitution model. Note that if more gradient steps were taken, the W matrix would quickly become discrete (from Lemma B.5). The jump taken corresponds to six subtree-prune and regraft moves [299].

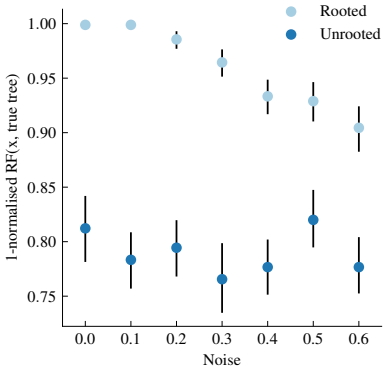
For larger alignments such as the popular Eutherian dataset [369], a single gradient step can result in 14 to 18 SPR moves. While the number of SPR moves achieved is large, this is achieved with a substantial increase in overall computational complexity when compared to FastME. We note that the gradient step size is dependent on the data and, as expected, greatly reduces as we approach an optimum.

4.3.2 A comparison to benchmark phylogenetic datasets

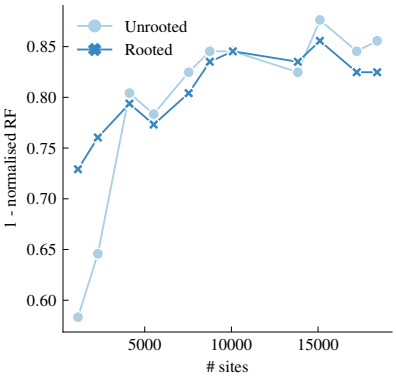
Table 4.2 presents a comparison of GradME with neighbour-joining (BioNJ) and FastME (subtree-prune and regraft version) over 11 popular phylogenetic benchmark datasets [372]. Both neighbour-joining and FastME are only able to infer a minimum length unrooted tree, and therefore we compare estimates only on unrooted trees. We always initialise our algorithm with a uniform, equiprobable tree, where the starting taxon labelling is random and optimised using Queue Shuffle. We estimate trees using distances from a GTR+Γ model estimated via maximum likelihood (see Appendix B.9 for details). As expected, FastME consistently outperforms BioNJ, with lower BME loss on all alignments. On the



(a) **Dataset: Primates**



(b) **Dataset: Simulated (20 taxa, 100k sites)**



(c) **Dataset: Jawed**

Figure 4.2: Results on empirical data **(a)** Starting from a random tree, represented by an $n \times n$ stochastic matrix, we compute the continuous gradient, apply softmax activation and increment the original matrix. In a single step, our gradient finds the correct tree at a distance of 6 subtree-prune and regraft moves from the random starting tree. **(b)** Simulating ultrametric trees of 20 taxa and 100,000 sites under an LG model of protein evolution. We add random uniform noise to all branch lengths to simulate departures from ultrametricity. Compared to the true tree via Robinson-Foulds distance, dark blue bars are midpoint rooting the best `FastME` tree and light blue bars are the inferred root from our approach. **(c)** Phylogenies for jawed vertebrates, where the number of genes (hence sites) are reduced to be more clocklike. Normalised Robinson-Foulds distance are shown between the best `ASTRAL` [382] tree, the best unrooted `FastME` tree which has been midpoint rooted (light blue) and our inferred rooting algorithm (dark blue). Performance for `FastME` reduces when the number of sites is small.

Table 4.2: Balanced minimum evolution loss scores for 11 phylogenetic benchmark datasets. Lower is better. Scores from BioNJ and FastME were obtained following the implementations in `ape` [299] using the same distance matrix as GradME. The distance matrix was estimated from a GTR+Γ model via maximum likelihood [318]. Our GradME approach always starts from a uniform tree distribution (every tree is equiprobable) with a random taxon ordering (optimised by Queue Shuffle). The best performing approaches for each dataset are denoted in bold. GradME either equalled or performed better than FastME. The topological accuracy, measured as one minus the Robinsons-Foulds distance is shown between GradME and FastME and GradME and a maximum likelihood gold standard from IQ-TREE also using a GTR+Γ model

Dataset	BioNJ	FastME	GradME	Topological accuracy between GradME and FastME	Topological accuracy between IQ-TREE and the best distance tree
DS1	0.3118613	**0.3101232**	**0.3101232**	1.00	0.54
DS2	3.725205	**3.7239944**	**3.7239944**	1.00	0.77
DS3	8.0115913	**8.0075588**	**8.0075588**	1.00	0.97
DS4	2.2528503	**2.2447615**	**2.2447615**	1.00	0.68
DS5	6.3077156	**6.2606057**	**6.2606057**	1.00	0.70
DS6	0.6249236	0.6228563	**0.6219367**	0.87	0.67
DS7	9.9174641	**9.882138**	**9.882138**	1.00	0.91
DS8	1.337924	**1.3252984**	**1.3252984**	1.00	0.82
DS9	0.3788481	**0.3788481**	**0.3788481**	1.00	0.66
DS10	1.1286037	**1.1247627**	**1.1247627**	1.00	0.78
DS11	1.313921	1.3096422	**1.3096415**	0.88	0.53

other hand, GradME always achieves a better or equal loss compared to FastME. We observe similar results when using different substitution models (e.g., F81). In the two examples where GradME does better than FastME, the topological accuracy, measured by one minus the Robinson-Foulds distance [100] is close to 0.9, suggesting FastME has converged to a similar tree. We note that FastME's performance is generally worse when using the nearest neighbour interchange heuristic (instead of the SPR-based heuristic). When compared to a maximum likelihood gold standard (IQ-TREE [298]), the best distance method does not recover the same tree as that from maximum likelihood, but in some cases, is very close (e.g., DS3 and DS7). Finally, we note that while GradME outperforms FastME, it is orders of magnitude slower and in most of the datasets FastME finds the same optimal tree as GradME.

4.3.3 Rooting ultrametric trees

Despite being applicable to the unrooted problem, our approach, at its core, works with rooted trees. As previously discussed, if we assume the existence of a distant outgroup, then the balanced minimum evolution objective can be used to optimise a rooted phylogenetic tree. In Appendix B.1, we show that, given an ultrametric unrooted tree, the optimal rooting maximises a heuristic for the root-to-tip distance in the tree. While this property only holds for ultrametric trees, our approach still works well

for near clock-like trees. As an experiment, we draw small (20 taxa) random ultrametric phylogenies with a total length of one, and simulate 100,000-residue protein sequences [299] down these trees under an LG [319] model of protein evolution, assuming random uniform amino acid base frequencies. In the ultrametric cases, all taxa are equidistant to the root, which corresponds to a strict molecular clock. We add uniform noise to all branch lengths to simulate departure from a strict clock. Fig. 4.2b shows the Robinson-Foulds [100] distance from the true tree to the *midpoint-rooted* best *unrooted* `FastME` tree (when SPR moves were used by `FastME`), and the distance to our inferred rooted tree. We see that when the tree is ultrametric, or close to ultrametric, our approach recovers the correct rooted tree. As expected, an increase in noise leads to a decrease in topological accuracy, although our approach still performs substantially better than midpoint rooting. We note that uniform noise is unlikely to be biologically realistic. Instead, deviations from a strict clock are more likely to be heterogeneous in certain clades or internal branches. However, for small departures, we believe our algorithm to reliably infer the correct tree and root simultaneously.

We implement our rooting algorithm on the popular mammal data from [369]. We infer a rooted tree via Queue Shuffle and also midpoint root the best FastME tree. Both trees, unrooted, have the same balanced minimum evolution loss, but our rooted loss is less than the FastME midpoint rooted loss. Our rooted tree correctly identifies *Gallus gallus* (red junglefowl) as the outgroup, while midpoint rooting pairs *Gallus gallus* with *Ornithorhynchus Anatinus* (platypus) (see Fig. B.2 for the rooted phylogenies).

4.3.4 Rooting the phylogeny of all jawed vertebrates

To perform a more detailed evaluation of our framework, we tested GradME's robustness for topological inference by finding the root of the large jawed vertebrates dataset from [357] with 99 taxa and 4593 genes. Given the reliance of our method on ultrametric data for inference of the root, we first made a fast measure of the ultrametricity of each gene-tree. To do this, we inferred the phylogeny of each gene using GradME, followed by midpoint rooting. The coefficient of variation in root-to-tip lengths was taken as a measure of ultrametricity. We then concatenated ranked genes into supermatrices including decreasing numbers of genes, and examined the performance of GradME with midpoint rooting against our method. All inferences were performed using the LG amino-acid substitution model to maintain simplicity. We placed special focus on our ability to use small portions of data for recovering the main groupings of vertebrates; these key groupings include the root separating cartilaginous (Chondrichthyes) versus boned vertebrates, ray-finned fishes (Actinopterygii), and the major groups of tetrapods and amniotes (amphibians, mammals, archosaurs, turtles, and lizards and relatives).

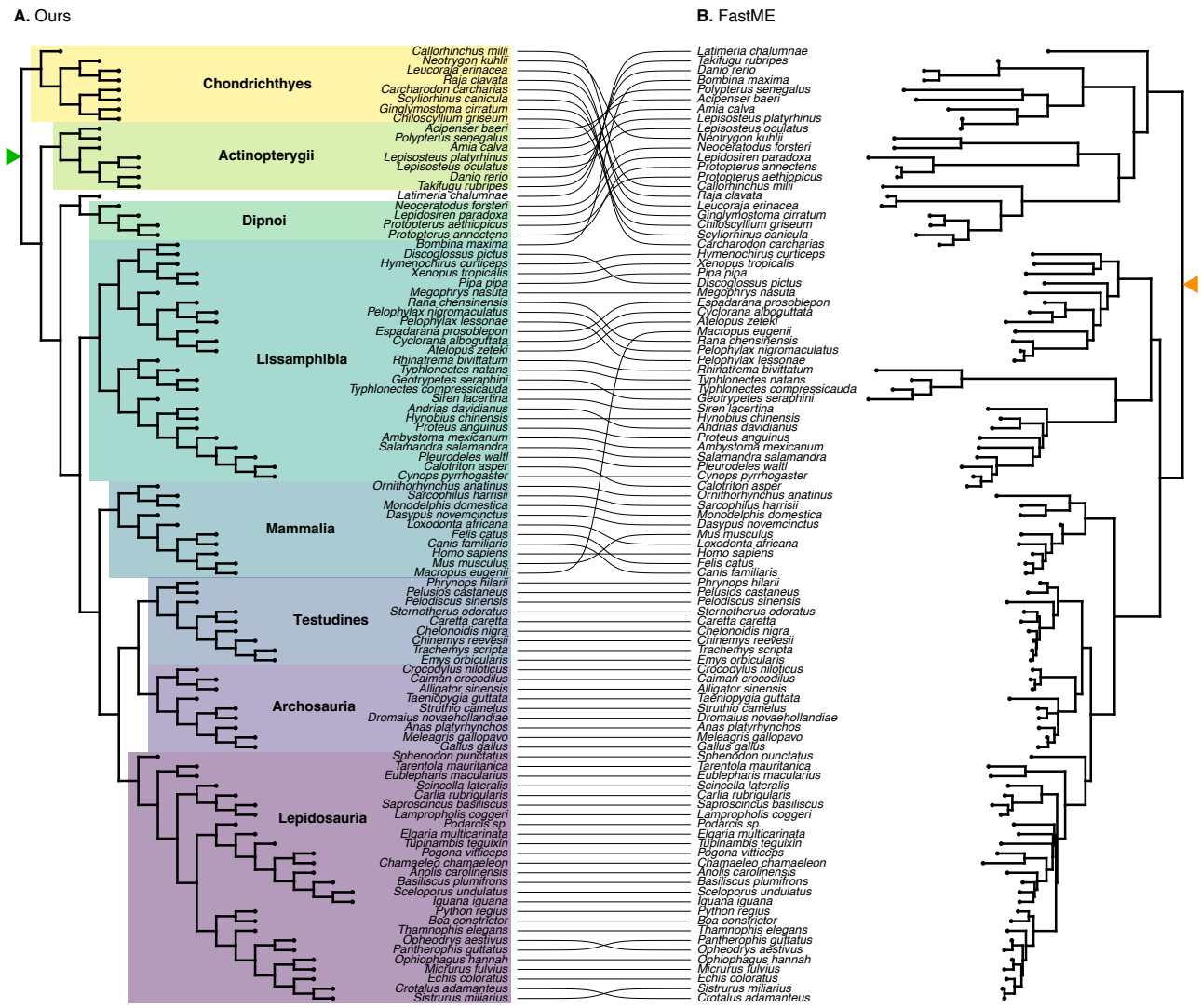


Figure 4.3: Phylogenetic inferences of the jawed vertebrates' phylogeny using the two most ultrametric loci from a data set of 99 taxa and 4593 genes [357]. **(a)** Inference using our approach leads to high accuracy in identifying the root and all major jawed vertebrate taxa. Note that, we do not estimate branch lengths, but only topology via balanced minimum evolution **(b)** inference using FastME and midpoint rooting leads to widespread error, primarily and critically near the root of the process.

A small number of genes with an ultrametric signal were generally sufficient for resolving many of the major lineages of vertebrates using both midpoint rooting and our approach (Fig. 4.2c). For larger numbers of genes, midpoint rooting and our approach are broadly similar. However, at the smallest numbers of genes (0.05%, 2 genes), midpoint rooting was unable to recover many of the early relationships among vertebrates, such as the root, monophyly of cartilaginous fishes, ray-finned fishes, Tetrapoda, or the mammals. Even small amounts of data (1460 amino-acids of 1,964,439; 0.07% of the original data) were sufficient for GradME to resolve the root as well as every major grouping of jawed vertebrates accurately (Fig. 4.3). The only exception was the controversial position of the Coelacanth, which was found to be sister of Dipnoi (lungfish) rather than the more widely accepted position as sister of Dipnoi plus Tetrapoda. While this remarkable performance under the simple LG model is in part attributed to the informative nature of highly ultrametric genes, our tree topology demonstrates the superiority of our approach in accuracy and efficiency over other fast methods in phylogenetics.

4.4 Discussion

We have introduced a new approach for exploring the vastness of tree space. Counterintuitively, our approach explores a much bigger space than the space of possible trees, but this larger space allows for new ways to find the best tree. The key to our method's success lies in transforming the phylogenetic tree search problem from a discrete to a continuous one, allowing us to achieve superior performance. To our knowledge, this is the first time a continuous, differentiable objective function for the inference of tree topology has been proposed, and it opens new possibilities for phylogenetic inference. Bayesian phylogenetics can be regarded as the most robust framework for inferring phylogenies, but has been to-date limited by the poor ability of random walk Metropolis-Hastings algorithms to explore tree space [383]. More efficient Hamiltonian Monte Carlo samplers have been proposed [333] to tackle this problem, and our framework presents a new avenue to jointly explore topology and branch lengths with efficient samplers. A remaining limitation of our approach is the need to shuffle labels to fully explore the space of all possible trees, and while the approach we use, Queue Shuffle, is mathematically and practically powerful, this step is still discrete. The possibility of permutation distributions such as the Gumbel-Sinkhorn distribution could allow for a fully differentiable algorithm. Finally, the complexity of our approach is $\mathcal{O}(n^5)$, which easily allows for large phylogenies up to a thousand, but not tens of thousands. However, computation on GPUs or TPUs in parallel can facilitate computational tractability.

A major benefit of our approach is that it naturally enables the estimation of the root node, which has been a long matter of interest in the biological sciences [384, 385, 386]. For genes where a strict

clock is a reasonable assumption, our method of traversing tree space in large steps reliably estimates both the correct tree topology and the root. Our approach will likely be useful in settings where genetic sequences are contemporaneous and time for measurable evolution is short, such as early epidemics or nosocomial settings. However, as we showed analytically, our approach will have reduced performance when considering rate heterogeneity and departures from a strict clock.

Tests on the relationships among jawed vertebrates demonstrate that even minimal amounts of data can be sufficient for our method to reach high accuracy in topology and root estimates. These results are consistent with previous work on large amounts of genome-scale data showing that clocklike loci to be the most suitable for phylogenetic inference [387]. Furthermore, our approach is effective with negligible amounts of data – where other methods are ineffective – making it a powerful addition to the existing toolkit for addressing recalcitrant questions of the tree of life.

Our approach is based on the minimum evolution principle, which has repeatedly shown to produce fast and accurate inference. Nonetheless, an interesting area for further study is to extend the continuous path length formulation to approximations of traditional phylogenetic likelihoods [297]. This would be particularly beneficial for implementation in Bayesian inference, since tree topology inference is a major obstacle to large hierarchical models [388, 389]. Our method is therefore a step towards more efficient sampling of the complex posterior distributions over tree topology.

Chapter 5: Paper III: Bayesian distance-based phylogenetics for the genomics era

Abstract: As whole genomes become widely available, traditional likelihood-based or Bayesian phylogenetic methods are demonstrating their limits in meeting the escalating computational demands. Conversely, distance-based phylogenetic methods are efficient and have a long history, but are rarely favoured due to their inferior performance. Here, we extend distance-based phylogenetics using an entropy-based likelihood of the evolution among pairs of taxa, allowing for fast Bayesian inference in large-scale datasets. We provide evidence of a link between the inference criteria used in distance methods and phylogenetic likelihoods, providing additional evidence for why distance-based approaches work well in practice. Using the entropic likelihood, we perform Bayesian inference on three phylogenetic benchmark datasets and find that estimates closely correspond with previous inferences and their bootstrap supports. We also apply this approach to a 60-million site alignment from 363 avian taxa, covering most avian families. We find the method to have outstanding performance relative to traditional methods, and reveal substantial uncertainty regarding the diversification events immediately after the K-Pg transition event. The entropic likelihood allows for efficient and accurate Bayesian inference of phylogeny and branch supports, enhancing the scalability of phylogenetic inference to accommodate the demands of the genomic era.

Full Author List: Matthew J Penn, Neil Scheidwasser, Joseph Penn, Mark P Khurana, Christl A Donnelly, David A Duchêne and Samir Bhatt

Author contributions: See end of chapter.

Joint Authorship: M.J.P., N.S., D.A.D. and S.B. have joint authorship of this work.

5.1 Introduction

The field of phylogenetics underpins a large portion of modern biological research, offering a powerful framework to describe branching processes across the tree of life. With the advent of vast amounts of genomic data, the field is now challenged by the fact that the space of possible phylogenetic trees scales double-factorially: for n samples (or leaves) there are $1 \cdot 3 \cdot 5 \ldots (2n-3)$ number of possible bifurcating rooted trees [273, 284]. This makes it impossible to consider every possible tree as a candidate, except in very small datasets (approximately ≤ 10 taxa). Whilst phylogenetic reconstruction under any existing method is NP-hard [280], methods based on Felsenstein's likelihood [375] are generally considered highly robust [120], where various heuristic approaches to tree search have shown excellent performance in both simulations [113, 114] and in comparisons with independent data from the fossil record [320]. This has led maximum likelihood to become the underlying criterion for many phylogenetic tree estimation frameworks [113, 298, 316]. However, the computational demand of maximum likelihood approaches is becoming impractical for many genome-scale and epidemiological datasets, raising questions about the feasibility of these analyses into the future of the genomics era [390, 391].

The primary caveat with modern phylogenetics is that the traditional Felsenstein's likelihood [375] is highly complex to compute for large datasets. Specifically, calculating the likelihood of a *single* tree is $\mathcal{O}(nNc^2)$ for n leaves, N unique site patterns and c character states. Even with the efficiency savings that occur when considering similar trees, the calculation remains highly complex, being of at least $\mathcal{O}(N)$. Thus, optimising by making the best move on the tree topology using subtree-pruning and regrafting (SPR) from the set of n^2 possible moves still gives a high complexity, of at least $\mathcal{O}(n^2 N k_L)$ for k_L optimisation steps. Conversely, distance-based approaches using the balanced minimum evolution (BME) criterion [392, 147] have a high initial pre-processing complexity, of $\mathcal{O}(Nn^2 + c^2 n^2)$, but far lower optimisation complexity that is independent of the number of unique site patterns N. Indeed, using a similar SPR scheme under this criterion has a worst-case complexity of $\mathcal{O}(k_B n^2 \text{Diam(T)})$ for k_B optimisation steps, where $\text{Diam}(T) \lesssim \mathcal{O}(n)$ is the maximum number of edges between each pair of taxa. In sum, the statistical benefits of using Felsenstein's likelihood can be outweighed by their high computational cost, providing a strong motivation for developing alternative fast approaches for phylogenetic inference.

The gap in complexity between likelihood and distance methods widens dramatically in Bayesian inference settings, where large numbers of likelihoods are evaluated. When analysing whole genomes with millions of sites, or thousands of taxa, likelihood methods become prohibitively slow and the only alternative is either a minimum evolution approach or maximum parsimony. While maximum

parsimony is consistent under certain conditions, BME has been proven to be more broadly statistically consistent under knowledge of the true model and when studying a single gene [344]. Although not as empirically accurate as maximum likelihood, the statistical guarantees of BME will ensure it is highly accurate in large-scale data regimes, particularly when considering branch supports and other forms of uncertainty (e.g., in branch lengths or date estimates) [56].

Although a point estimate of a single best tree can be useful, uncertainty quantification is another critical aspect to biological hypothesis testing, yet generally prohibitive in very large modern datasets. Existing solutions include bootstrapping [393, 394] and Bayesian posterior supports [395]. For all inferential approaches, bootstrapping is possible and has highly efficient implementations [396]. However, the interpretation of a bootstrap sample is non-probabilistic. It is a measure of re-sampling variance and of the robustness of the estimator over small changes in the data [397]. In contrast, Bayesian approaches yield the posterior probability for any given tree and for branches within a tree. Bayesian approaches, however, require a likelihood and are therefore unsuitable for inference under minimum evolution and maximum parsimony. Generalised Bayesian approaches [398] offer an alternative, but are as of yet underdeveloped for phylogenetics. Therefore, a key limitation of minimum evolution is that the objective criterion is not a likelihood and therefore cannot be used for hypothesis testing or, more importantly, for Bayesian analysis. Until now, uncertainty under this framework has only been quantified via bootstrapping.

To address the various shortcomings of phylogenetic methods for inference using large modern datasets, we develop a new likelihood function by considering the effect of using tree entropy as a prior distribution. Using simplifying approximations, we reduce this new likelihood to a tree length given an "entropic distance matrix" which we denote as d^S_{ij}. This entropic distance is the expected entropy of the evolution process across the path between a pair of taxa in the tree, given their genetic distance. The resulting entropic likelihood is highly correlated with the well-established likelihood of Felsenstein [297] that underpins all maximum likelihood and Bayesian phylogenetics (Figure 5.2), and we provide a mathematical justification for the closeness of this relationship. Through studying the similarities between the entropic distance d^S_{ij} and standard genetic distance d_{ij}, we find that, under a Markovian branching process, the entropic distance is extremely well-approximated by a linear function of the standard genetic distance (Figure 5.1). Specifically, given a continuous time Markov chain (CTMC) substitution model P (e.g., Jukes-Cantor model [374]), the entropy along a single branch is $S(t) = \sum_{a,b} \pi_a P_{ab}(t) \log(P_{ab}(t))$ (where π denotes stationary frequencies). Furthermore, the entropic tree length is approximately a linear function of the BME tree length. The error of this approximation has favourable mathematical properties (Figure 5.3), and this new formulation provides

additional insights into the robustness of BME as an optimality criterion for phylogenetic inference. Our work paves the way for Bayesian phylogenetic inference using distance-based methods, allowing for highly efficient probabilistic analysis of the massive datasets that define the genomic era.

5.2 Methods

5.2.1 Notation and preliminaries

Throughout this paper, we will use $\mathcal{T}$ to denote the true topology which, using the phylogenetic data, we are aiming to infer. $\mathcal{U}$ will be used for topology under consideration - which may or may not be the true tree.

The classical maximum likelihood problem in phylogenetics involves the construction of a weighted, binary tree with topology $\mathcal{U}$ and branch lengths $\boldsymbol{b}$, describing the evolutionary history of a set of n taxa, each of which corresponds to a leaf node of $\mathcal{U}$.

In general, a fixed number of sites of genomic or amino acid data is available for each taxon. Typically, the substitution process of a single site is modelled using a reversible continuous time Markov chain (CTMC) where the transition matrix/kernel is:

$$P_{ij}(t;\boldsymbol{\phi}) = \mathbb{P}(\text{A site changes from state } i \text{ to state } j \text{ in time } t) \tag{5.1}$$

where $\boldsymbol{\phi}$ represents the set of model parameters, which are generally estimated as part of the inference process. Estimating these parameters is a standard procedure, and so we will not refer to $\boldsymbol{\phi}$ in the methods.

We use Q to denote the Q-matrix of the substitution CTMC, and define π to be its stationary distribution. We assume that $|Q_{ij}| > 0$ (and hence $P_{ij}(t) > 0$) for all $t > 0$ and all i and j. This is not strictly necessary for the results presented in this paper, but simplifies the derivations.

Assuming that topology is independent of the genetic mutations (that is, for example, there are no mutations which make short branches more likely), one can simulate the sites of a tree (and to calculate likelihoods) as follows. Firstly must choose a node, r, denoted hereinafter as the *simulation root* - which corresponds to the node at which to begin the simulation. It will be assumed that the sites at node r are chosen according to the stationary distribution of the CTMC. It is shown in Lemma C.1 in the Appendix that the distribution of the sites of the tree is independent of which node is chosen to be the simulation root.

Thus, given a set of characters g for a given site on the internal *and* external nodes (totalling 2n-1), one can define a function $p(g,\mathcal{U},\boldsymbol{b})$ which gives the likelihood of these values being observed

for a given site on the topology $\mathcal{U}$ with branch lengths $\boldsymbol{b}$. This can be done by considering each edge separately, and will be derived later in the methods.

Felsenstein's likelihood

In general, only the characters on leaf nodes $l(g)$ are observed, rather than the full set of characters g.

Felsenstein's likelihood solves this problem by marginalising over the possible values of the sites of the internal nodes. For a site with characters $l(g)$ on the leaf nodes, the likelihood of that site (on the topology $\mathcal{U}$ with branch lengths $\boldsymbol{b}$) is then

$$\lambda_F(l(g),\mathcal{U},\boldsymbol{b}) = \sum_{h\in\mathcal{G}:l(h)=l(g)} p(h,\mathcal{U},\boldsymbol{b}) \tag{5.2}$$

where $\mathcal{G}$ denotes the set of possible site patterns over the whole tree.

Calculating the likelihood directly from (5.2) is inefficient, and in practice, λ_F can be computed through a post-order tree traversal, where one iteratively finds the likelihood of the subtree rooted at each node x, conditional on the value of the sites at x.

Under the assumption that different sites evolve independently, Felsenstein's likelihood L_F of the tree is then simply the product of the $\lambda_F(g,\mathcal{U},\boldsymbol{b})$ for each observed set of characters g. We use ℓ_F to denote the logarithm of Felsenstein's likelihood.

5.2.2 Motivation: a likelihood from balanced minimum evolution

In the balanced minimum evolution paradigm [392], the aim is to find the tree where, under a specific (balanced) method of branch length estimation, the length of this tree (that is, the sum of its branch lengths) is minimised. To produce a similarly-motivated likelihood for each possible tree (comprised of a topology $\mathcal{U}$ and branch lengths $\boldsymbol{b}$), we develop a measure, $E(\mathcal{U},\boldsymbol{b})$, of the entropy of this tree. This provides a similar measure of tree complexity - low expected likelihood (just like high length) means that a more complex evolutionary process would have occurred.

To create this entropy measure, we suppose that we observe the sites at both internal and external nodes of the tree. If, for a random simulated mutation process on $\mathcal{U}$, we observe sites G across all these nodes, then our entropy measure is simply the entropy (that is, the expected log-likelihood) of the random variable G.

To apply such principles to a likelihood setting, we consider a Bayesian approach to inferring the true tree. Suppose that for a tree with topology $\mathcal{U}$ and branch lengths $\boldsymbol{b}$, our prior distribution is

$\theta(\mathcal{U}, \boldsymbol{b})$. Suppose that the likelihood of observing the data, $\mathcal{D}$ is $P(\mathcal{D}|\mathcal{U}, \boldsymbol{b})$. Then, the posterior likelihood of a topology $\mathcal{U}$ is proportional to

$$L(\mathcal{U}) = \int_{\boldsymbol{B}} P(\mathcal{D}|\mathcal{U}, \boldsymbol{b})\theta(\mathcal{U}, \boldsymbol{b})d\boldsymbol{b} \tag{5.3}$$

where $\boldsymbol{B}$ denotes the set of possible values of $\boldsymbol{b}$ and $d\boldsymbol{b} = db_1 db_2 ... db_{2n-1}$.

In the classical Felsenstein's likelihood inference problem, θ is generally a function of $\boldsymbol{b}$ (for example, using the assumption that branch lengths are exponentially- or gamma-distributed) and is independent of the topology, $\mathcal{U}$.

Under our minimal-complexity paradigm, we use the tree entropy as a prior as we expect simpler trees to be more likely. As entropy is fundamentally a log-likelihood quantity, we set (ignoring the normalisation constant),

$$\log(\theta(\mathcal{U}, \boldsymbol{b})) = E(\mathcal{U}, \boldsymbol{b}) \tag{5.4}$$

To simplify the calculation, we make the following assumptions. Firstly, we assume that we have enough data such that (using the consistency of the BME estimators), the likelihood P is closely concentrated around the BME branch lengths, $\boldsymbol{b}^*(\mathcal{D},\mathcal{U})$. That is,

$$P(\mathcal{D}|\mathcal{U}, \boldsymbol{b}) \approx P(\mathcal{D}|\mathcal{U})\delta(\boldsymbol{b} - \boldsymbol{b}^*(\mathcal{D},\mathcal{U})) \tag{5.5}$$

Secondly, we choose to ignore the additional information provided about the topology, $P(\mathcal{D}|\mathcal{U})$. Again, this mirrors the classical BME setup where the objective can be calculated from an unordered set of branch lengths. This reduces the total information used and, we conjecture, is the main reason why Felsenstein's likelihood generally outperforms balanced minimum evolution - even though both are asymptotically consistent, Felsenstein's likelihood utilises all the information from the data and will therefore perform better on finite datasets. However, the relationship between the topological posterior and the data is far more complex than the relationship between the branch length posterior and the data, so restricting ourselves to information about the branch lengths substantially reduces the complexity of the tree inference problem. Thus, ignoring normalisation, we suppose that

$$P(\mathcal{D}|\mathcal{U}) = 1 \tag{5.6}$$

Under these assumptions, we can evaluate the integral to see that L becomes

$$L(\mathcal{U}) = \theta(\mathcal{U}, \boldsymbol{b}^*(\mathcal{U}, \mathcal{D})) \tag{5.7}$$

and hence

$$\ell(\mathcal{U}) = E(\mathcal{U}, \boldsymbol{b}^*(\mathcal{U}, \mathcal{D})) \tag{5.8}$$

which is the expected likelihood of the tree with topology $\mathcal{U}$ and branch lengths given by the BME estimates.

In the remainder of this paper, we will derive an efficient way of calculating $L(\mathcal{U})$ and show that is closely approximated by a linear function of the standard BME objective. We will also show that this likelihood is closely-related to Felsenstein's likelihood, and can therefore be used to perform approximate inference in that paradigm.

5.2.3 Finding $\ell(\mathcal{U})$

Given the balanced minimum evolution estimates for branch lengths, $\boldsymbol{b}$, it is relatively straightforward to calculate the tree entropy E of a topology $\mathcal{U}$. We will now show how to calculate $p(g,\mathcal{U},\boldsymbol{b})$ and E.

Calculating $p(g,\mathcal{U},\boldsymbol{b})$

To begin, it is helpful to find an explicit formula for the function $p(g,\mathcal{U},\boldsymbol{b})$, which can be interepeted as Felsenstein's likelihood in the case that the sites at internal nodes are known.

Define the set of edges to be $\mathcal{E} = \{(e_i^1, e_i^2) \mid i = 0, 1, ..\}$, where e_i^1 and e_i^2 are the nodes which this edge connects, such that e_i^1 is the closest to the simulation root r (note that it is possible, and indeed necessary, that $e_i^k = e_j^l$ for some indices i, j, k, l). Suppose that z_i^j is the site value on node e_i^j and that b_i is the length of edge (e_i^1, e_i^2). Then, as the substitution CTMC on each edge is conditionally independent given the start and end values of the sites, we have

$$p(g, \mathcal{U}, \boldsymbol{b}) = \pi_{z_r} \prod_i P_{z_i^1, z_i^2}(b_i) \tag{5.9}$$

Here $P_{z_i^1, z_i^2}(b_i)$ is the probability of z_i^1 mutating into z_i^2 in time b_i, while z_r gives the value of the site at the simulation root. Note that, assuming time reversibility, Lemma C.1 shows that the value of $p(g, \mathcal{T})$ is independent of r.

Calculating E

We can now note that, as

$$\mathbb{E}\Big(\log(p(G,\mathcal{U},\boldsymbol{b}))\Big) = \mathbb{E}\Big(\log(\pi_{Z_r}) + \sum_i \log(P_{Z_i^1,Z_i^2}(b_i))\Big) = \mathbb{E}\Big(\log(\pi_{Z_r})\Big) + \sum_i \mathbb{E}\Big(\log(P_{Z_i^1,Z_i^2}(b_i))\Big) \tag{5.10}$$

We can ignore $\mathbb{E}(\log(\pi_{Z_r}))$, as it is independent of the topology and the data (this is simply the entropy of the starting genome at the simulation root). Moreover, using Lemma C.2, we know that as the length of the branch joining Z_i^1 and Z_i^2 is b_i,

$$(Z_i^1, Z_i^2) \overset{d}{=} (M_0, M_{b_i}) \tag{5.11}$$

where M is an independent copy of the mutation CTMC, and $\overset{d}{=}$ denotes equality in distribution. Thus, we have

$$\mathbb{E}\Big(\log(P_{z_i^1,z_i^2}(b_i))\Big) = -S(b_i) \tag{5.12}$$

where S is the entropy of our mutation CTMC run for time b_i (taken to be positive - hence the minus sign as log-likelihoods are always negative). Thus,

$$\ell(\mathcal{U}) = \mathbb{E}\Big(\log(p(G,\mathcal{U}))\Big) = \text{const.} - \sum_i S(b_i) \tag{5.13}$$

For trees with N sites, we can use the independence of different sites, meaning that the entropy E is simply N multiplied by $\mathbb{E}\big(\log(p(G,\mathcal{U}))\big)$.

5.2.4 The entropic likelihood

While (5.13) is computable, using the individual balanced estimates to the branch lengths b_i can lead to problems. For example, there is no guarantee that these estimates will be positive. Having the ability to use some prior on branch lengths is therefore helpful and, while we cannot efficiently do this on an individual level, we develop a method in this section for imposing priors on an inter-taxa scale.

To begin, note that the objective function (5.8) is -1 multiplied by the length of a tree where the branches have lengths $S(b_i)$ (rather than their original b_i lengths, which were the BME approximations). Thus, we seek to find an entropic distance matrix d^S such that d^S_{ij} is the distance between taxa i and j in our entropy-weighted tree (for a single site). In this case, the objective function is the

length of the BME tree with distance matrix d^S, meaning

$$\ell_S(\mathcal{U}) = -N \sum_{i \neq j} 2^{-e_{ij}} d_{ij}^S \tag{5.14}$$

where e_{ij} is the (unweighted) path length between i and j in $\mathcal{U}$. This sum over $i \neq j$ considers all ordered pairs i and j. We call the likelihood ℓ_S the *entropic likelihood*.

Models for approximating d_{ij}^S

If one supposes that the tree was generated according to some random process, then one can set

$$d_{ij}^S = \mathbb{E}\big(D_{ij}^S(d_{ij})\big) \tag{5.15}$$

where $D_{ij}^S(\tau)$ is a random variable with distribution equal to the entropic distance between two taxa on a tree generated according to this random process conditional on the distance between these taxa being equal to d_{ij}.

In this paper, we use a simple branching process where each branch has a length distributed according to some probability density function f, splits into two branches at the end of this length, and the process is stopped after some fixed time T (so that any branches still "active" at time T form a leaf node). We hope to examine more complicated models in the future, particularly those which allow for non-ultrametric trees (for example, by varying the mutation rate along the branches). However, as we will show in the subsequent sections, the approximate linearity of our entropic distance means that this model is still useful even in the absence of ultrametricity.

Lemma C.3 and Lemma C.4 provide formulae for the tree being generated by this branching process. In Lemma C.3, one has the renewal equation

$$\frac{1}{2}\mathbb{E}(D_{ij}^S(d_{ij})) := h(\tau) = (1 - F(\tau))S(\tau) + \int_0^\tau (h(\tau - t) + S(t))f(t)dt \tag{5.16}$$

where F is the cdf for the pdf f. For any parametric choice of branch length distribution, these renewal equations can be solved by Riemann sums. For example, in the case of an exponential distribution, Lemma C.4 shows an analytically tractable solution with rate θ. In this case,

$$\mathbb{E}(D_{ij}^S(d_{ij})) = 2\bigg(\int_0^{d_{ij}/2} \Big[e^{-\theta t}(S'(t) + \theta S(t)) + \int_0^t S(s)\theta^2 e^{-\theta s} ds \Big] dt \bigg) \tag{5.17}$$

and hence, the entropic likelihood is

$$\ell_S(\mathcal{T}) = -2N \sum_{i \neq j} \left\{ 2^{-e_{ij}} \left(\int_0^{d_{ij}/2} \left[e^{-\theta t}(S'(t) + \theta S(t)) + \int_0^t S(s)\theta^2 e^{-\theta s} ds \right] dt \right) \right\} \tag{5.18}$$

Comparison to Felsenstein's likelihood

The approximation (5.14) means we use the data differently to Felsenstein's likelihood. Rather than using the data to directly calculate likelihoods via $p(g,\mathcal{U})$, the data are now used to create an inter-taxa distance matrix, from which the times t_i are calculated.

This different application of the data means that the full dataset only needs to be used once in the optimisation process, as the distance matrix will remain the same throughout, independently of the topology under consideration. When the number of sites, N is much larger than the number of taxa n (which is often the case, particularly in phylogenomics) this results in a substantial saving in computational cost, as after a single pre-processing step, the dataset effectively reduces from size Nn to size n^2.

5.2.5 Connection to the classical balanced minimum evolution objective

Defining $K = \sum_{a \neq b} \pi_a Q_{ab}$ to be the expected instantaneous substitution rate, assuming that

$$\frac{K}{\theta} << 1 \tag{5.19}$$

and defining $\kappa := \min_{i,j} \left\{ \theta d_{ij} \right\}$, one can show that the approximation

$$\frac{1}{2}\mathbb{E}(D_{ij}^S(d_{ij})) \approx \frac{1}{2}\mathbb{E}\left\{ D_{ij}^S\left(\frac{\kappa}{\theta}\right) \right\} + \left(\frac{\theta d_{ij}}{2} - \kappa \right) \int_0^\infty \theta S(s) e^{-\theta s} ds \tag{5.20}$$

has a percentage error of approximate leading order $\mathcal{O}(\log(\frac{K}{\theta})^{-1})$. This is formalised in Theorem C.1 in the Appendix.

The assumption that $\frac{K}{\theta} \ll 1$ may seem contrived, but is in fact necessary for phylogenetic inference to be possible. Noting that branch lengths are of length $\mathcal{O}(\frac{1}{\theta})$, the expected number of substitutions per site on a branch is $\mathcal{O}(\frac{K}{\theta})$. If this number is not small, then the sequences at each leaf node will largely uncorrelated, regardless of the distance between them in the tree, and so it will not be possible to infer a shared genetic history. Note that $\frac{K}{\theta}$ is a non-dimensional quantity and independent of any reparametrisation of time.

Using this result, we can note that (5.20) can be written as

$$\mathbb{E}(D_{ij}^{S}(d_{ij})) \approx a + bd_{ij} \tag{5.21}$$

for some constants a and b. Then, the entropic likelihood (5.14) becomes (using the Kraft Equality)

$$\ell_S(\mathcal{T}) = -b\sum_{i\neq j} 2^{-e_{ij}} d_{ij} + a\sum_{i\neq j} 2^{-e_{ij}} \tag{5.22}$$

$$= -b\sum_{i\neq j} 2^{-e_{ij}} d_{ij} + a \tag{5.23}$$

which is a linear function of the classical BME objective and hence, in particular, the optimal tree under this objective will be the optimal BME tree. A visualisation of how linear this relationship is for three datasets (see Table 5.1) is shown in Figure 5.1.

Table 5.1: Evaluation datasets

Dataset	Reference	**# Sites**	**# Taxa**	**Type**	**Taxonomic rank**
DS1	[358]	1,949	27	rRNA (18S)	Tetrapods
DS2	[293]	2,520	29	rRNA (18S)	Acanthocephalans
DS3	[359]	1,812	36	mtDNA	Mammals; mainly Lemurs

Note that the value of b can be written as

$$b = \Big(d_{ij}\theta\Big)\Big(\int_0^\infty \theta S(s) e^{-\theta s} ds\Big) = \Big(\mathbb{E}(\text{branches between } i \text{ and } j)\Big)\Big(\mathbb{E}(\text{entropy on a branch})\Big) \tag{5.24}$$

and hence, the new distances form a simple approximation to the entropy between each pair of taxa.

5.2.6 Analytical and empirical comparison to Felsenstein's likelihood

As we show in this section, our entropic likelihood is similar, though not exactly equivalent, to the classical Felsenstein's likelihood. In particular, we demonstrate that the two are highly linearly correlated, meaning that, using an empirically-calculated scaling factor, our entropic likelihood can be used to approximate sampling under Felsenstein's likelihood. We do this by deriving our likelihood directly from the expected Felsenstein's likelihood, and analyse the impact of each required approximation.

Setup

Consider the expected Felsenstein's log-likelihood of a single-site tree $\mathcal{U}$ with branch lengths $\boldsymbol{b}^*$ given by balanced minimum evolution. This is equal to the entropy of the leaf nodes when the mutation

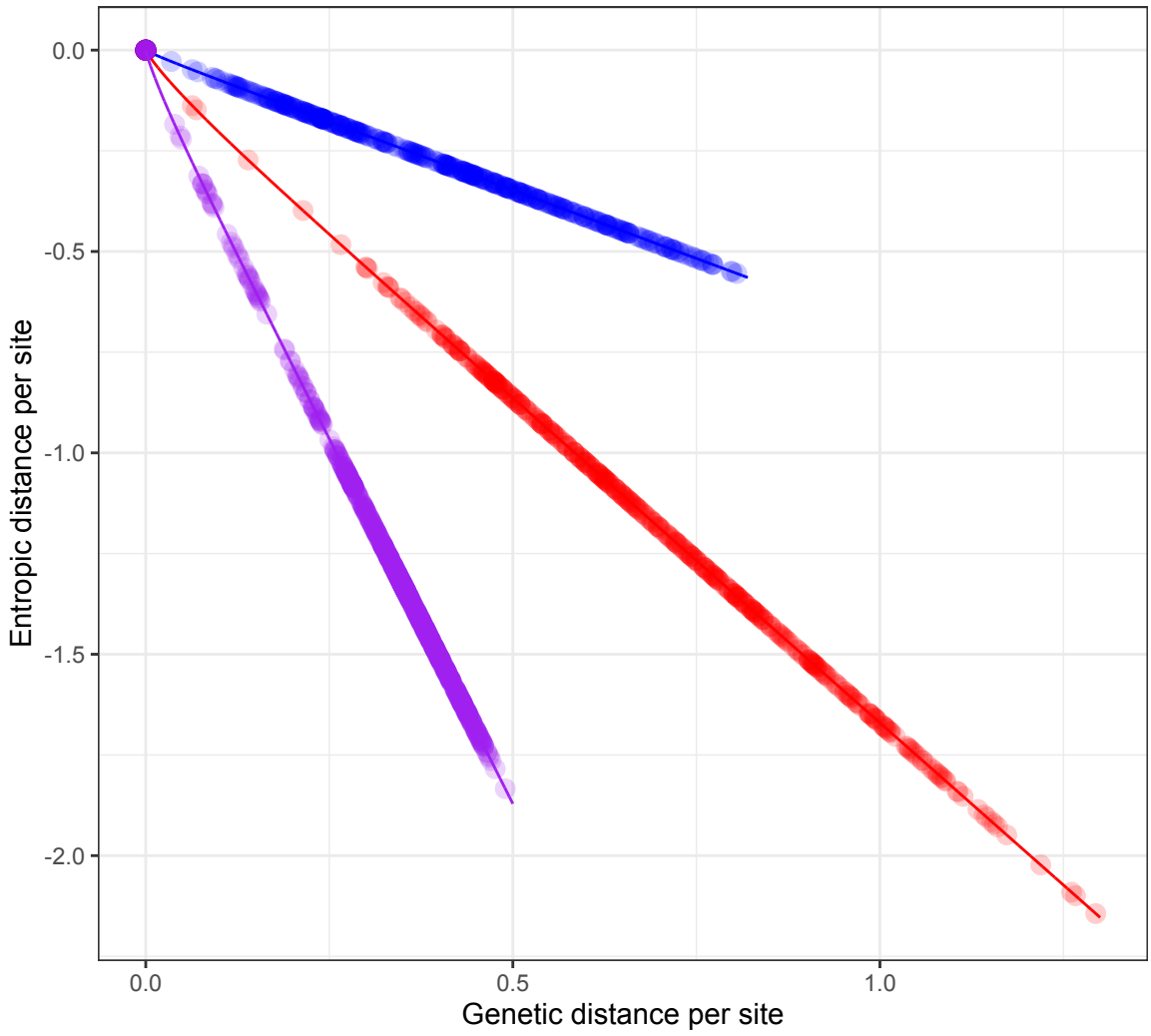


Figure 5.1: **Genetic distances d_{ij} against corresponding entropic distances $\mathbb{E}(D_{ij}^S(d_{ij}))$ for three empirical datasets (see Table 5.1).** For all datasets we see a strong linear relationship and correlation near equal to 1. Note that non-linearity tends to exist close to zero, where there are fewer data.

process is realised on the true tree $\mathcal{T}$ with the true branch lengths $\boldsymbol{\beta}$ and then the resulting leaf sites are mapped to their equivalents in $\mathcal{U}$.

Using X to denote the set of leaf nodes and, later, Y to denote the set of internal nodes, we define $H(X|\mathcal{T}, \boldsymbol{\beta})$ to be the expected Felsenstein's log-likelihood. In the large sites limit, this will be equal to leading order to the observed Felsenstein's log-likelihood, and so we shall assume that the two are equal in the subsequent derivation.

Using all nodes

To reduce notation, we omit the conditional dependence of H on $\boldsymbol{\beta}$ and $\mathcal{T}$ in this section.

As we vary $\mathcal{U}$ (and therefore also $\boldsymbol{b}^*$), we expect Felsenstein's log-likelihood $H(X)$ and the log-likelihood on all nodes, $H(X, Y)$ to be closely related to each other. Noting that

$$H(X, Y) = H(X) + H(Y|X) \tag{5.25}$$

the difference between them will be determined by the amount of information that the leaf nodes provide about the internal nodes.

In trees with a very low number of expected mutations per site (over the whole length of the tree), as are commonly found in epidemiological applications [399], we expect that $H(X, Y) \approx H(X)$, as one can (with high probability) infer the states at the internal nodes from the leaf nodes - essentially using the fact that sites are extremely unlikely to mutate twice or more.

Conversely, in trees with a very high number of expected mutations per site, the sites at each node are approximately independent of each other (and therefore simply independent realisations of the stationary distribution of the mutation CTMC). In this case, as there are approximately the same number of leaf and internal nodes, we expect $\frac{1}{2}H(X, Y) \approx H(X)$.

In between these two extremes where there are intermediate numbers of expected mutations per site, when we consider trees of similar topologies and branch lengths (and therefore $H(Y|X)$ varies slowly) it therefore seems reasonable to approximate

$$H(X) \approx \kappa_1 H(X, Y) \tag{5.26}$$

for some fixed $\kappa_1 \in [\frac{1}{2}, 1]$. That is, we expect the fully-observed likelihood to be approximately

proportional to Felsenstein's likelihood.

Varying the sampling distribution

To recover our original new likelihood $\ell(\mathcal{U})$ from (5.13), we must make the approximation

$$H(X, Y|\mathcal{T}, \boldsymbol{\beta}) \approx \kappa_2 H(X, Y|\mathcal{U}, \boldsymbol{b}^*) = \kappa_2 \ell(\mathcal{U}) \tag{5.27}$$

The effect of this approximation is to change the assumed distribution of the sites when calculating the entropy. $H(X, Y|\mathcal{T}, \boldsymbol{\beta})$ assumes the true distribution but, in general, this is unknown. When considering the tree $(\mathcal{U}, \boldsymbol{b}^*)$, we hence instead perform our calculation *as if this tree were the correct tree.*

In Lemma C.6, we justify this approximation, and show that we expect $\kappa_2 \in [1, 2]$. In essence, the primary reason for this is that as we consider trees $\mathcal{U}$ which are further from the true tree $\mathcal{T}$, their entropy will be higher and therefore there is a double effect on $H(X, Y|\mathcal{T}, \boldsymbol{\beta})$ - it increases both because $\mathcal{U}$ is a higher entropy tree, and because the true sequence distribution is diverging from that given by $\mathcal{U}$.

Unlike the previous approximation (5.26), where we essentially just change the likelihood scale, the approximation (5.27) does add meaningful error and means that the correlation between the entropic and Felsenstein's likelihood is not as strong as, for example, the correlation between the entropic likelihood and the BME objective - which are virtually the same. However, we have found that these errors are not prohibitive, as we show in the subsequent section, where we make simple empirical comparisons between the entropic and Felsenstein's likelihood.

The entropic likelihood

The final approximation comes from the use of our entropic distance matrix to approximate the value of $\ell(\mathcal{U})$. The validity of this approximation depends on how well our heuristic tree branching process model describes the true process of tree creation, and can change the gradient of the linear relationship between the entropic likelihood and Felsenstein's likelihood.

To examine the effect of this, suppose that we have two branching models, A and B which determine the construction of a tree. In model A, we suppose that, for a uniformly chosen branch in the tree of length T_i, we have $\mathbb{E}(S(T_i)) = s_A$ and $\mathbb{E}(T_i) = \tau_A$, with similar definitions for model B.

Then, for taxa which are a sufficiently large distance Δ apart, we expect entropic distances $S.$ to be approximately equal to the average entropy on a branch, $s.$, multiplied by the number of branches,

approximately $\frac{\Delta}{\tau_\cdot}$. That is,

$$S_A = \frac{s_A}{\tau_A}\Delta \quad \text{and} \quad S_B = \frac{s_B}{\tau_B}\Delta \tag{5.28}$$

Thus, S_A and S_B are linearly related, but the ratio between them (and therefore the ratio between the entropic likelihoods given by the two tree construction models) will not necessarily be 1. While it is readily possible to re-estimate model parameters for each tree, this will not resolve all the previous sources of error in our approximation. We therefore propose calibrating our entropic likelihood against. Felsenstein's likelihood via a linear scaling with gradient m.

Analytical summary

Thus, we see that our entropic likelihood and Felsenstein's likelihood should be approximately linearly related. This is an important result, not only for justifying the validity of the entropic likelihood, but also in providing an explanation for the utility of BME - we can use the approximate linear relationship between BME and the entropic likelihood to show that we also expect BME to be approximately linearly related to Felsenstein's likelihood. This theoretical finding opens the door for approximate Bayesian inference using a distance matrix and BME, simply by performing a calibration to estimate the linear scaling coefficient m. It is critical to note that, even after scaling, the entropic likelihood will not exactly match Felsentstein's likelihood. However, given linearity, the two likelihoods will be of similar magnitudes for a given tree.

Illustrative empirical comparisons to Felsenstein's likelihood

First, we illustrate through a simple empirical example that a linear approximation is reasonable. As highlighted in the motivation, we do not expect this likelihood to perform as well as Felsenstein's likelihood because we ignore information provided about the topology $p(\mathcal{D}|\mathcal{U})$. In contrast, Felsenstein's likelihood incorporates this information through a post-order tree traversal and marginalisation. However, using our entropic likelihood is a reasonable choice when computational tractability prohibits the use of Felsenstein's likelihood.

To show the limitations of our entropic likelihood, we simulate a biologically realistic alignment from a known tree. Our tree simulation follows a birth death process with 50 species and with $\lambda = 0.5$, $\mu = 0.1$, $\rho = 1$ and time since origin (TMRCA) of 65 (reflecting many major radiations since the K-Pg transition event). On this birth death process, we simulate white noise with a mean rate and rate variation of 0.05 according to an exponentiated white noise process, resulting in samples that are distributed log normally. A birth death tree with noise is considered the true tree $\mathcal{T}$ and is created using the `TreeSim` [400] and `NELSI` [401] packages in R. To simulate an alignment down this tree we

use the *seqSim* function in `phangorn` [402] in R, four nucleotides and no gaps or unknown bases. We sample the six possible transition rates as well as the four possible base frequencies from a Dirichlet distribution with $\alpha_i = 5 \; \forall i$. We assume a Gamma shape of 1 with 4 categories and sample rates using the *discrete.gamma* function in `phangorn`. Finally, we assume a sequence length of $5,000$ base pairs. An example of a resultant tree is shown in Figure 5.2. For all inferences, we assume a Jukes-Cantor substitution model and therefore are performing inference under model misspecification. We compare our likelihood to Felsenstein's likelihood for alignments simulated as described above. Our likelihood is:

$$\ell_S(\mathcal{T}) = -N \sum_{i,j} \mathbb{E}(D_{ij}^S(d_{ij}))2^{-\epsilon_{ij}} \tag{5.29}$$

Assuming exponential branch lengths ($f(t) \sim \text{Exponential}(\theta)$) we use the distance matrix found by solving the renewal equation

$$\frac{1}{2}\mathbb{E}(D_{ij}^S(d_{ij})) := h(\tau) = (1 - F(\tau))S(\tau) + \int_0^\tau (h(\tau - t) + S(t))f(t)dt \tag{5.30}$$

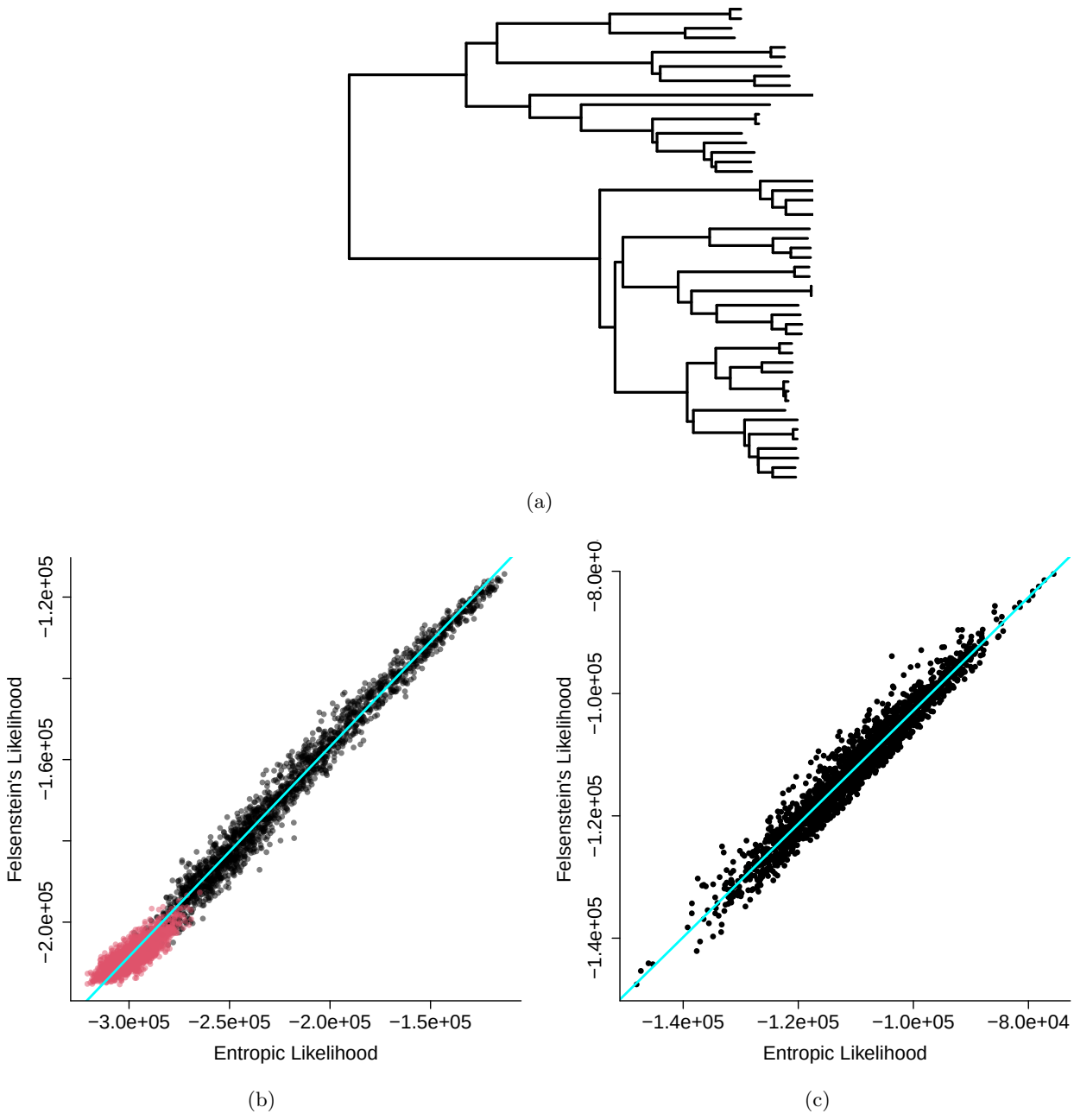
The rate of the exponential distribution was specified as $\hat{\theta} = \frac{n-2}{L}$ where L is the tree length (this formula is derived in Lemma C.5). Using Felsenstein's likelihood to estimate an optimal tree was performed using `optim.pml` implemented in the `phangorn` library [402] in R.

Simulating 2000 random alignments and optimising our likelihood and Felsenstein's likelihood and examining one minus the normalised Robinson-Foulds distance [100] to the known true tree results in a median topological accuracy of 0.9787 $[0.9361 - 1]$ for both our entropic likelihood approach and Felsenstein's likelihood approach. Examining the likelihoods for these 2000 random alignments, we see a strong linear relationship between the optimal likelihoods (see Figure 5.2 Right) with $m = 0.92$. Therefore, for the optimal tree, the scaling is close to one. Picking one single fixed tree and alignment, we can explore the relationship between our entropic likelihood and Felsenstein's likelihood for suboptimal trees given the alignment. Generating 2000 trees by subtree-prune and regraft (SPR) operations on the optimal tree, where the number of operations is randomly drawn uniformly from integers between 1 and 50 (i.e. maximum 50 sequential SPR moves) also results in a strong linear relationship (see Figure 5.2 Middle black points). Sampling 2000 entirely random trees and evaluating both likelihoods (see Figure 5.2 Middle red points) further interpolates the linear trend. We note that when examining suboptimal trees $m = 0.508$, and is therefore not a close scaling. However, the linear relationship that our theory suggests holds for trees close to the optimal tree, all the way to entirely random trees.

Next we show how the gradient, m, between the Felsenstein and entropic likelihoods varies between

0 and 1 across alignments with different rates. As we have noted from our discussion of the κ_1 scaling parameter in (5.26), our theory suggests with low rates we expect a gradient, m, of around one when compared to Felsenstein's likelihood, and with high rates we expect m to decrease. This decrease is also caused by the entropic likelihood approximation, as higher rates lead to entropy being less closely linear. Simulating trees using the same procedure as above with 50 taxa and 5,000 sites, we explored how rate variation affects our approximation via m. Simulated trees were rescaled following the procedure outlined in [403]. Briefly, varying evolutionary rates were set as the root to tip divergences. We rescaled root-to-tip distances to vary between a very wide range of $\{0.005, 2.5\}$ [403], and for each rate simulated 500 trees to estimate uncertainty and for each tree we explore the distribution around the optimal tree via SPR changes close to the optimal tree, where the number of sequential random SPR changes is drawn from $\sim \text{Poisson}(1) + 1$. To mitigate the effects of a misspecified tree construction model, we estimate θ for each new SPR tree which, for the exponential case, is available in closed form as $\hat{\theta} = \frac{n-2}{L}$ where L is the length of the tree found as the BME objective. This helps reduce the error for trees far from the true tree (which through classical BME theory, will have larger values of L), but does not reduce the impact of the overall model misspecification (an impact, which as discussed, will increasingly affect the gradient as the substitution rate grows). As shown in Figure 5.3 (left figure), re-estimating θ does result in coefficient m that is close to 1 when rates of substitution are low. However, as rates increase and multiple substitutions happen at multiple sites, the estimates of m become progressively worse. The topological accuracy of the best tree however remains excellent and comparable to Felsenstein's likelihood for all rates Figure 5.3 (middle figure). Similarly, the mean absolute percentage error between Felsenstein's likelihood and a linear model with the entropic likelihood does grow with rate, but not substantially, and considering the log scale, the linear approximation is good with up to half a percentage point of error.

In summary, the linearity of our entropic likelihood with Felsenstein's likelihood is a very useful property, and justifies the previous use of BME in likelihood proposals e.g. [404, 113]. However, because there is no guarantee that the gradient is 1, and indeed we expect it not to be for real data [403], Bayesian distance-based MCMC inference cannot be performed without a linear calibration. A solution we utilise is to simply perform a calibration of the entropic likelihood against Felsenstein's likelihood to estimate the gradient correction. This calibration only requires a few hundred trees and can be performed as an average across subsets of taxa for large numbers of sequences, or as a subset of the number of sites for a very large number of sites (e.g. a whole genome). The resultant posterior is of course with respect to our approximate likelihood and will not converge to Felsenstein's likelihood, but a judicious choice of data and model to ensure a good distance approximation will make



(a)

(b)

(c)

Figure 5.2: **Comparison of entropic and Felsenstein's likelihoods assuming exponential branch lengths.** **(a)** Example single true tree from simulation. **(b)** A comparison of the likelihoods of suboptimal trees inferred from data simulated through the true tree. Black dots are suboptimal trees generated by performing random SPR moves from the best estimated tree, and red dots are entirely random trees. **(c)** A comparison of the likelihoods for the best tree across 2000 simulated alignments.

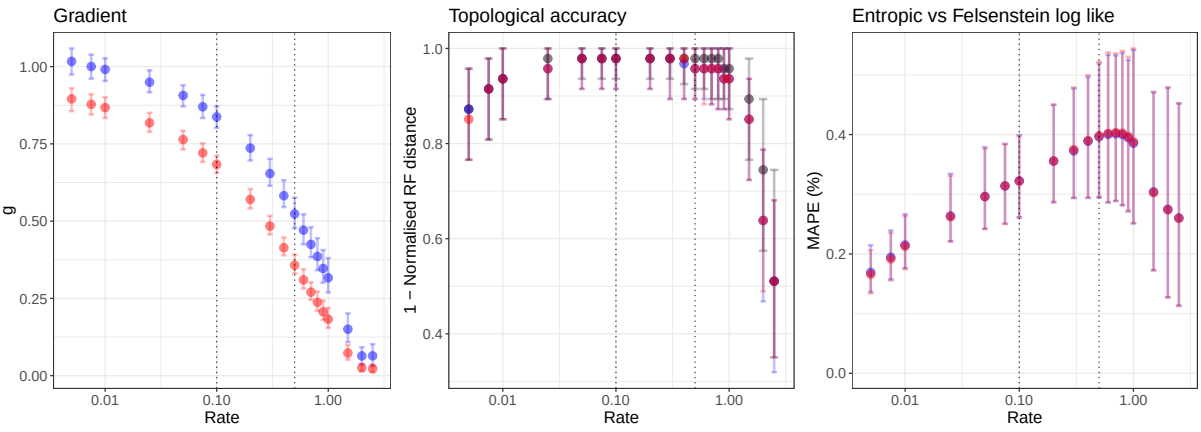


Figure 5.3: **Performance of the entropic likelihood across evolutionary rates.** Left: The variation of the scaling between the entropic and Felsenstein's likelihood as a function of rate, where red indicates the entropic distance estimated using equation 5.17 assuming exponentially distributed branch lengths. Blue points use the same procedure, but the branching rate θ is re-estimated for each tree and a new entropic distance found. Middle: Comparison of topological accuracy for Felsenstein's likelihood in black, entropic distance in red, and entropic distances when re-estimating θ in blue. Right: The mean absolute percentage error between Felsenstein's likelihood and a linear model of entropic distance. The percentage error is on a log likelihood scale. Black dotted lines show where most empirical data exist [403].

it close. We note that our theory and these illustrative simulations show that the BME objective can be treated as a likelihood that needs to be linearly calibrated against a choice of phylogenetic likelihood. For example, for the example datasets in Table 5.1, m using entropic likelihoods are estimated as $m = \{0.9119, 0.7882, 0.7635\}$. One can also calculate the gradients between standard BME and Felsenstein's likelihood, which are $\{8264, 4731, 4730\}$.

5.2.7 Implementation procedures

To estimate the entropic distance matrix, we first require a sequence alignment, from which we estimate the standard distance matrix D. This can be easily achieved for common reversible Markov substitution models. Next, we need to specify a branch length distribution $p(\beta)$. This can be chosen *a priori* or taken from a reasonable estimate from a tree. In our examples, we first find an (approximately) optimal BME tree, calculate the least squares solution to the branch lengths, and then use these lengths to inform our distribution. Assuming an exponential distribution, $p(\beta) \sim \text{Exp}(\lambda)$ where λ is the reciprocal of the mean across all branch lengths in the tree (the maximum likelihood solution). Other distributions can readily be used. Given a parametric choice for $p(\beta)$, the density and cumulative distribution functions are defined and the renewal equation 5.16 can be solved. Solving this equation for values of D results in a new entropic distance matrix D^s. Given D^s, we can choose a sampling of N trees $\{\mathcal{U}_1, \ldots, \mathcal{U}_N\}$ and regress our likelihood against another phylogenetic likelihood such as Felsenstein's likelihood to obtain an estimate of the scaling m. We note, that due to the

theory we introduce showing the linearity between the entropic distance and the genetic distance, this calibration can be done directly on the standard BME objective (taken from D) rather than the entropic likelihood (taken from D^s) with only small additional error. However, the resultant likelihood will be orders of magnitude different when performing the regression from standard BME, which can introduce numerical sensitivity.

5.3 Results and discussion

5.3.1 Bayesian distance-based inference on standard benchmark datasets

To compare current phylogenetic inference approaches with the entropic likelihood, we started by using three classical benchmark datasets in phylogenetics. As detailed in Table 5.1 included data on tetrapods [358] (27 taxa, 1949 sites), acanthocephalans [293] (29 taxa, 2520 sites), and mammals [359] (36 taxa, 1812 sites; hereafter DS1-3). For each of the three datasets, we performed phylogenetic inference using (i) maximum likelihood implemented in RAxML-NG [124] optimisation with 100 different starts and 1000 bootstraps with the `autMRE` function (extended majority-rule consensus tree criteria with a cut-off of 0.03). From the 100 different starts, a subset of unique local minima was created. (ii) BME using FastME [56] with 1000 bootstraps. We also included (iii) the new continuous vector-based BME inference using GradME [48] to search for an optimal tree. We implement the entropic likelihood by running a Random Walk Metropolis Hastings Markov Chain Monte Carlo for 20 million iterations, with 500,000 burn-in discarded and across 20 chains, thinning every 10 samples. Convergence was assessed by calculating the R-hat statistic on the chain log-likelihoods, as well as visual inspection of the log-likelihood trace plots. To guarantee linearity between likelihoods (see Methods), we calibrated the entropic likelihood to Felsenstein's likelihood using 1000 trees, perturbed from the best balanced minimum evolution tree with SPR(x), where x is the number of sequential SPR moves and $x \sim \text{Poisson}(1) + 1$ - that is mostly small changes of one SPR from the best tree, but occasionally large changes. The results are visualised in Figure 5.4. To maintain consistency with previous studies [372], all analyses were performed on a Jukes-Cantor substitution model [374].

To compare all optimal, bootstrap, and posterior trees across analyses, we calculated pairwise Robinson-Foulds distances across trees for the datasets enumerated in Table 5.1. To visualise these distances, we follow [405] and plot the first two components of a multidimensional scaling reduction of the distance matrix. For DS1 (see Figure 5.4a), a very large number of trees were sampled ($\sim$ 30,000) and we see the distance tree is relatively far from the maximum likelihood tree (around 0.5 normalised RF distance from the RAxML-NG modes). The bootstrap distributions from maximum

likelihood and distance estimation overlap, as does the Markov chain Monte Carlo (MCMC) posterior. The MCMC entropic posterior is more concentrated but with considerable variation around the BME optimum. This narrower sampling of the posterior set of unique trees is expected [397].

In DS2, the continuous tree search [48] found a better tree than the optimal FastME tree, which was the most similar to the RAxML-NG tree. In this dataset, only one RAxML-NG mode was found across all 100 runs, and the bootstrap (blue circles) is tighter than in DS1 (Figure 5.4b). The entropic MCMC posterior ranges across the BME bootstrap samples and includes trees very close to the best maximum likelihood tree. In DS2, FastME selected a tree which was suboptimal, while GradME found a tree with a smaller length that was closer to the maximum likelihood best tree. Reassuringly, the entropic MCMC samples around both of these modes, and a third mode that is close to the maximum likelihood tree. For DS3, we once again see that the entropic MCMC explores several modes that correspond closely to the BME bootstrap distribution, and samples trees close to one of the best maximum likelihood trees. Both the FastME and RAxML-NG bootstraps overlap considerably, and the distance between optimal trees is small (normalised RF distance ~ 0.1).

Overall, these analyses demonstrate that the entropic likelihood facilitates MCMC sampling, with the distribution of samples overlapping with a distance-based bootstrap. The number of unique topologies varies depending on how well-resolved the tree is based on a distance matrix; for DS1, over 30,000 unique topologies were found, for DS2, only 413 were found, while for DS3, only 382 were found. The correspondence between the distance-based posterior and the distance-based bootstrap validates the theoretical results above, suggesting that this likelihood can be viewed as BME with a suitable scaling. This means that by using the entropic distance matrix as opposed to a standard distance matrix (e.g., Jukes-Cantor distances), it is possible to perform Bayesian model-based inference using all the theory and methods already developed in BME. BME has a quadratic complexity and can scale to thousands of taxa, opening the possibility of Bayesian inference on huge datasets with thousands of taxa and millions of sites. We note that we have only used a single distance matrix to estimate the posterior distributions, propagating uncertainty from the distance matrix will create a larger posterior distribution of unique trees.

5.3.2 Bayesian inference on 363 genomes from the Bird 10,000 Genomes (B10K) project

To showcase the results that can be obtained from analyses using the entropic likelihood, we used data from the Bird 10,000 Genomes (B10K) project [406], an initiative that aims to generate representative draft genome sequences from all extant bird species. Here we analyse the release of 363 genomes

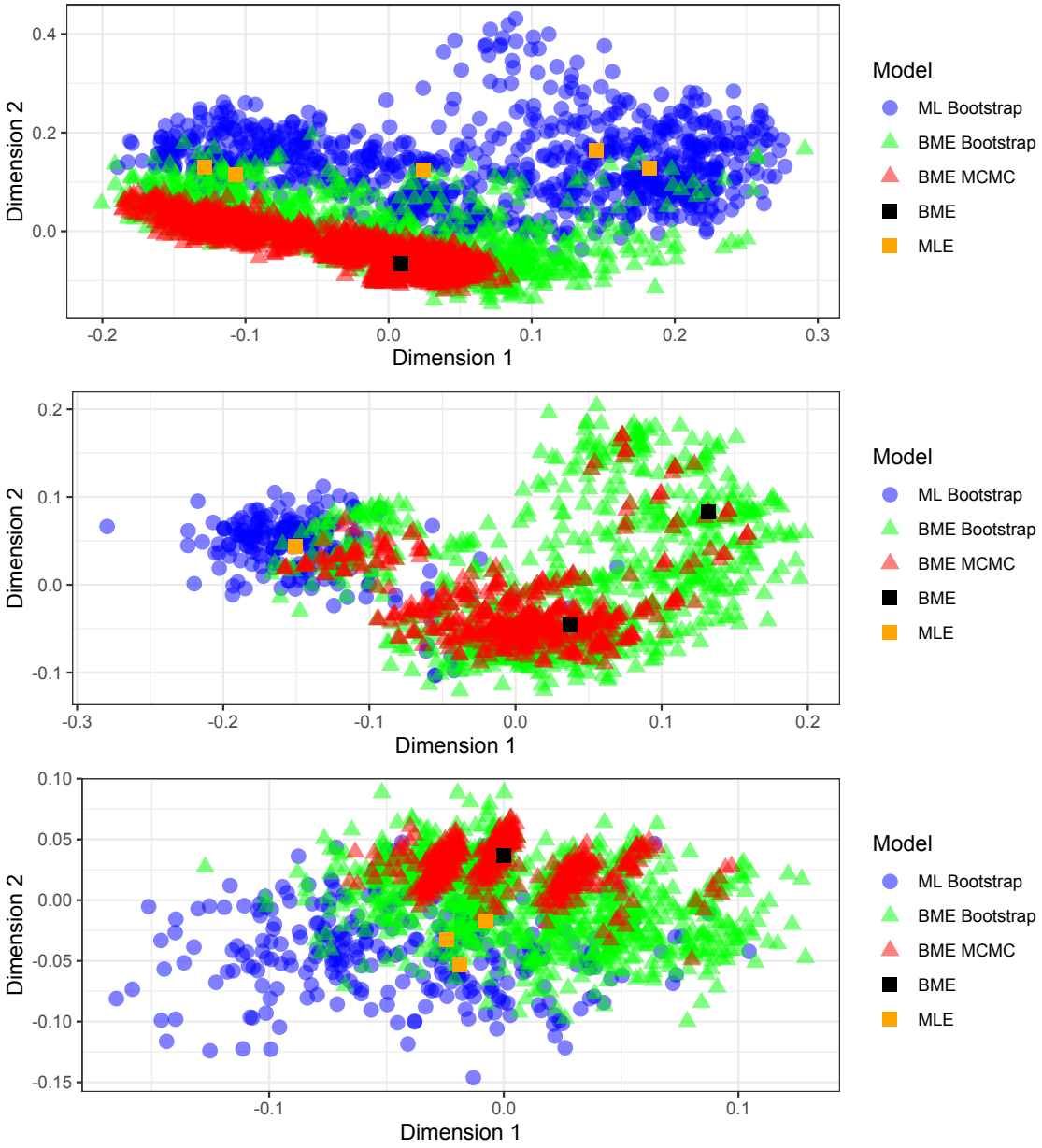


Figure 5.4: For the set of unique trees, the generalised Robinson-Foulds distance is calculated, and the distance matrix reduced by multidimensional scaling. A Jukes-Cantor model was used. ML Bootstrap: RAxML-NG [124] bootstrap, BME bootstrap: FastME [56] bootstrap, BME MCMC is the method presented in this paper, and BME and MLE are the point estimates. To facilitate visualisation, only a random sample of 5000 trees is shown.

representing 92% of all bird families. We estimated pairwise genetic distance under a GTR+Γ substitution model (general time-reversible model with gamma distributed rate variation among sites) using the approach outlined in [318]. Given free parameters for rates, $S \in \mathbb{R}^6_{\geq 0}$, frequencies, $\pi \in \mathbb{R}^4_{\geq 0}$ and $\sum \pi_i = 1$, the time between two sequences t_{ij}, and genetic sequence data $\mathcal{G}$, the log-likelihood for the transitions between a pair of taxa i and j is

$$\mathcal{L}_{ij}(\mathcal{G}|S, \pi, t_{ij}) = \sum_a \sum_b \kappa_{ab}^{ij} \log(P_{ab}(t_{ij}; S, \pi)) \tag{5.31}$$

where κ_{ab}^{ij} is the number of $a \to b$ transitions from taxon i to taxon j. We approximate the optimal parameters by minimizing the total negative log-likelihood (that is, the sum over i and j of $\mathcal{L}_{ij}$) using gradient descent in `Jax` [356].

We considered two alignments, a smaller alignment, A_1, of 100,000 sites that was originally used to estimate a maximum likelihood tree in RAxML [113, 406] and a much larger alignment, A_2, of 63.4 million sites from intergenic regions across the genomes for which maximum likelihood tree estimation is highly impractical. For both alignments A_1 and A_2, assuming a GTR+Γ substitution model as above, distance matrices D_1 and D_2 could be estimated. Comparing these distance matrices to the RAxML tree created from A_1, we find the distance matrix from the smaller alignment, D_1, yields a topological accuracy (one minus the normalised Robinson-Foulds distance) to the RAxML tree of 91% - that is, 91% of all splits are shared between the trees. Using the distance matrix from the larger alignment D_2, the topological accuracy to the RAxML tree increases substantially to 98% - as expected, the performance gap between our method and Felsenstein's narrows as the amount of data increases. Critically, both distance matrices are highly linearly related (intercept 0.004911 and gradient 1.051129), and therefore we can perform a likelihood calibration, applying a calibration on A_1 to the more reliable distance matrix D_2. The calibration of the entropic likelihood to Felsenstein's likelihood was done on A_1 using 500 trees, perturbed from the best balanced minimum evolution tree with SPR(x), where x is the number of sequential SPR moves and $x \sim \text{Poisson}(1) + 1$.

Our avian family-level Bayesian phylogenetic estimate is highly congruent with inferences from previous genome-scale studies. It supports the major groupings of birds being Palaeognathae, Gallo-anseres and Neoaves [406]. The analysis supports Neoaves as being split into 9 of the 10 major monophyletic lineages described previously [407], and largely with maximal posterior supports (Figure 5.5). The only exception is Afroaves, which is split into two groups as also suggested in several previous studies [407]. Importantly, this split is one of the few showing lower posterior support. Here, the lowest posterior support was found in nodes that are widely accepted to have occurred soon after

the Cretaceous-Palaeogene (K-Pg) boundary and which pinpoint the uncertainty in the relationships among the 10 major groups of Neoaves. We also found low uncertainties primarily in the relative placements among Strisores, Aequornithes, Phaethontimorphae, Opisthocomiformes, and Cursorimorphae, which have long been recognised as difficult to place in the avian tree of life [408, 409].

These results are consistent with the genome-wide disagreement in the relationships early after the post-K-Pg transition. Critically, the entropic likelihood Bayesian approach provides evidence that those nodes with low support cannot be placed confidently using nucleotide data alone. For illustration, fast molecular dating using penalised likelihood was performed as implemented in the `ape` [299] R package (function *chronos*) using the avian maximum clade credibility tree and allowing for substantial variation in rates across lineages ($\Lambda = 0.001$). We included 29 fossil constraints and a constraint on crown Neoaves to fall between 60 and 70 million years ago. As expected under these calibrations, the resulting dates support the so-called 'big-bang' scenario of rapid radiation scenario in most avian lineages occurring after the K-Pg mass extinction event [410], and are overall in line with prior expectations.

5.3.3 Conclusions

In this paper, we have showcased the outstanding potential for distance-based approaches to perform highly efficient and accurate phylogenetic inference on large datasets while considering uncertainty in inferences, allowing for model selection, and incorporating complex models of evolution. Distance-based phylogenetics has traditionally been considered less effective than likelihood-based approaches due to the lack of marginalisation across internal nodes, the reduction of large amounts of sequence data into a single distance matrix, and the lack of model-based customisation. Yet, we demonstrate that these methods are robust to many of these perceived weaknesses. An inherent limitation of distance-based phylogenetics will be the inability to account for unknown states within a given tree to a greater extent than Felsenstein's likelihood-based approaches. Conversely, the expanding volume of phylogenetic data, in both site numbers and taxa, is placing substantial strain on maximum likelihood approaches, such that distance-based methods will play a critical role in genome-scale phylogenetic inference. This is already evident in major phylogenomics initiatives where computation times can extend into months, and where Bayesian solutions are infeasible [406]. Consequently, when considering large numbers of taxa (i.e., in the thousands) or site patterns (i.e., in the tens of millions), distance-based approaches are among the only viable solutions along with parsimony-based approaches.

Considering the inherent uncertainty in these distance-based phylogenetic inferences, the only viable approach until now has been bootstrapping, which provides limited meaningful information

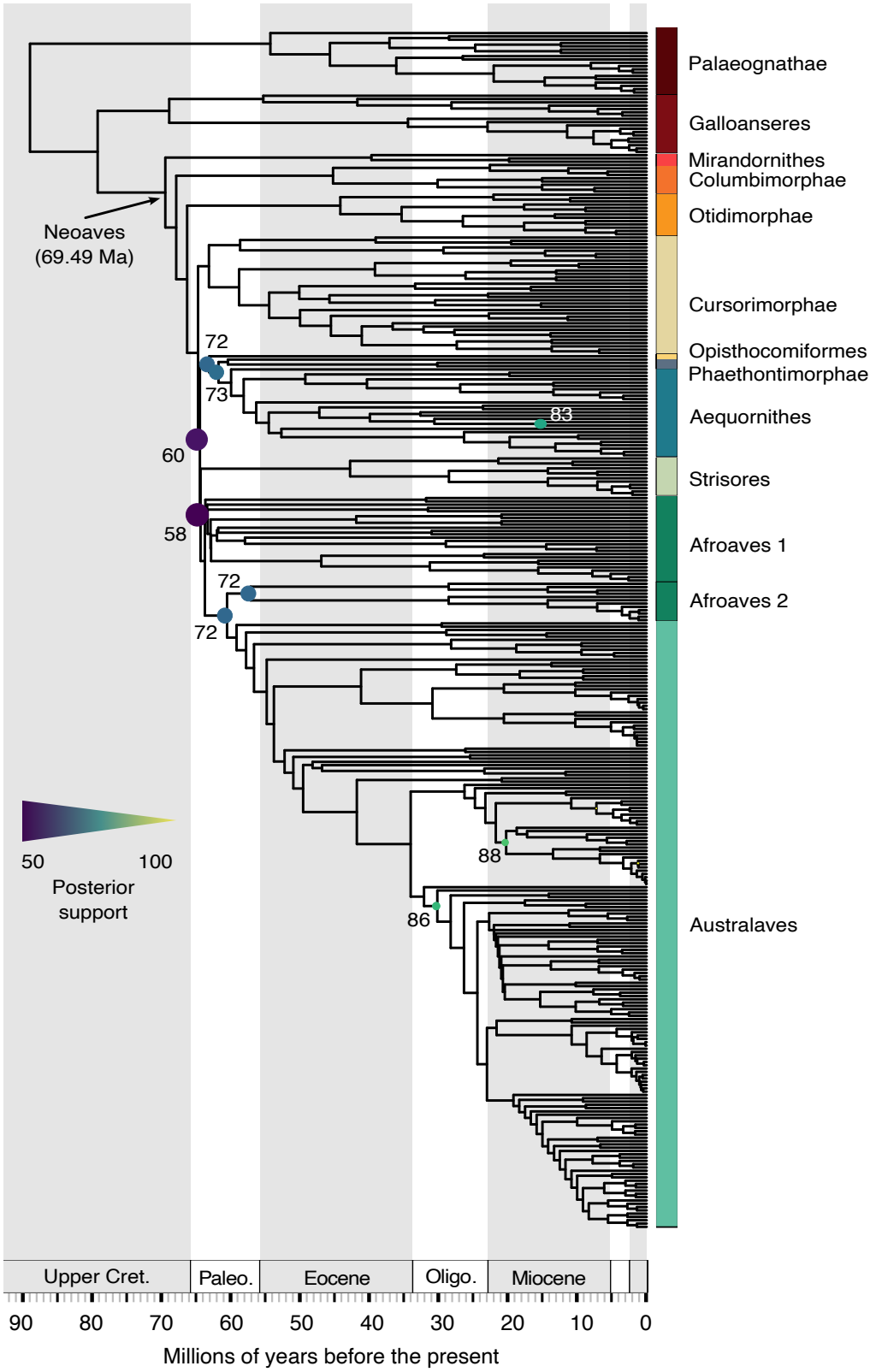


Figure 5.5: Maximum clade credibility tree with posterior node support for 363 bird taxa built from a distance matrix of ~ 60 million sites. The tree has a normalised Robinson-Foulds distance of 0.02 (8 splits) from the maximum likelihood tree, and the posterior uncertainty in tree topology concentrates on the K-Pg boundary, in line with prior expectations [406]. Molecular dating was performed for visual purposes on a subset of calibrations from the original study.

when the sample size of sites is very large. Our proposed entropic likelihood approach provides an intuitive alternative via Bayesian model-based inference as applied to distance phylogenetics. We have shown how distance-based posteriors produce sensible distributions of trees with a close correspondence to the bootstrap. When factoring the quadratic complexity of distance-based approaches, the entropic likelihood also presents new avenues for Bayesian inference of large phylogenetic datasets, as we have shown on genome-scale data on avian taxa.

The theory we introduce arrives at an approximation of Felsenstein's likelihood, but in which only inter-taxa pairs are considered. This entropic likelihood is very closely related to BME - by far the best performing of all distance-based approaches. It has previously been shown that BME works as a special case of the weighted least-squares approach to phylogenetic inference [344]. To show this, however, a fundamental assumption is that the variance is $\propto 2^{\tau}$, where τ is the topological distance taken, which is empirically justified but nevertheless a strong assumption. In contrast, we arrive at what is essentially BME with a distance matrix scaled by a constant, and show this is an approximation of the standard phylogenetic likelihood. Our theory therefore provides a different and robust justification for the use of BME and suggests that the standard BME objective function is approximately proportional to a likelihood. We prove that the error from this approximation is small, casting new confidence on fast methods for making accurate model-based inferences from genome-scale data.

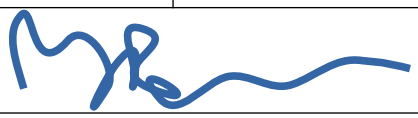
Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	**Bayesian distance-based phylogenetics for the genomics era**
Publication Status	**Submitted to Nature Communications Biology**
Publication Details	-

Student Confirmation

Student Name:	**Matthew Penn**		
Contribution to the Paper	**Conceived of and developed the mathematical framework for the key ideas in this paper.** **Carried out some initial experiments to verify the methods.** **Contributed to writing the paper jointly with the other authors. Wrote the appendix.**		
Signature		Date	**20/03/24**

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: **Dr Samir Bhatt**		
Supervisor comments		
Signature 	Date	**20/03/24**

This completed form should be included in the thesis, at the end of the relevant chapter.

Chapter 6: Paper IV: Intrinsic randomness in epidemic modelling beyond statistical uncertainty

Matthew J Penn et al. "Intrinsic randomness in epidemic modelling beyond statistical uncertainty". In: Communications Physics 6.1 (2023), p. 146.

Status: This paper has been published in *Communications Physics*

Abstract: Uncertainty can be classified as either aleatoric (intrinsic randomness) or epistemic (imperfect knowledge of parameters). The majority of frameworks assessing infectious disease risk consider only epistemic uncertainty. We only ever observe a single epidemic, and therefore cannot empirically determine aleatoric uncertainty. Here, we characterise both epistemic and aleatoric uncertainty using a time-varying general branching process. Our framework explicitly decomposes aleatoric variance into mechanistic components, quantifying the contribution to uncertainty produced by each factor in the epidemic process, and how these contributions vary over time. The aleatoric variance of an outbreak is itself a renewal equation where past variance affects future variance. We find that superspreading is not necessary for substantial uncertainty, and profound variation in outbreak size can occur even without overdispersion in the offspring distribution (i.e. the distribution of the number of secondary infections an infected person produces). Aleatoric forecasting uncertainty grows dynamically and rapidly, and so forecasting using only epistemic uncertainty is a significant underestimate. Therefore, failure to account for aleatoric uncertainty will ensure that policy-makers are misled about the substantially higher true extent of potential risk. We demonstrate our method, and the extent to which potential risk is underestimated, using two historical examples.

Full Author List: Matthew J Penn, Daniel J Laydon, Joseph Penn, Charles Whittaker, Christian Morgenstern, Oliver Ratmann, Swapnil Mishra, Mikko S Pakkanen, Christl A Donnelly and Samir Bhatt

Joint Authorship M.J.P. and S.B. have joint authorship of this work.

Author contributions:[1] S.B. and M.J.P. conceived and designed the study. S.B. performed analysis with assistance from M.J.P.. M.J.P., D.J.L. and S.B. drafted the original manuscript. M.J.P. drafted the Supplementary Information with assistance from J.P.. M.J.P., D.J.L., J.P., C.W., C.M., O.R., S.M., M.S.P., C.A.D., and S.B. revised the manuscript and contributed to its scientific interpretation. S.B., M.J.P. and C.A.D. supervised the work.

[1] As found in the published paper.

6.1 Introduction

Infectious diseases remain a major cause of human mortality. Understanding their dynamics is essential for forecasting cases, hospitalisations, and deaths, and to estimate the impact of interventions. The sequence of infection events defines a particular epidemic trajectory – the outbreak – from which we infer aggregate, population-level quantities. The mathematical link between individual events and aggregate population behaviour is key to inference and forecasting. The two most common analytical frameworks for modelling aggregate data are susceptible-infected-recovered (SIR) models [411] or renewal equation models [412, 60]. Under certain specific assumptions, these frameworks are deterministic and equivalent to each other [413]. Several general stochastic analytical frameworks exist [414, 60], and to ensure analytical tractability make strong simplifying assumptions (e.g. Markov or Gaussian) regarding the probabilities of individual events that lead to emergent aggregate behaviour.

We can classify uncertainty as either aleatoric (due to randomness) or epistemic (imprecise knowledge of parameters) [415]. The study of uncertainty in infectious disease modelling has a rich history in a range of disciplines, with many different facets [416, 417, 418]. These frameworks commonly propose two general mechanisms to drive the infectious process. The first is the infectiousness, which is a probability distribution for how likely an infected individual is to infect someone else. The second is the infectious period, i.e. how long a person remains infectious. The infectious period can also be used to represent isolation, where a person might still be infectious but no longer infects others and therefore is considered to have shortened their infectious period. Consider fitting a renewal equation to observed incidence data [60], where infectiousness is known but the rate of infection events $\rho(\cdot)$ must be fitted. The secondary infections produced by an infected individual will occur randomly over their infectious period g, depending on their infectiousness ν. The population mean rate of infection events is given by $\rho(t)$, and we assume that this mean does not differ between individuals (although each individual has a different random draw of their number of secondary infections). In Bayesian settings, inference yields multiple posterior estimates for ρ, and therefore multiple incidence values. This is epistemic uncertainty: any given value of ρ corresponds to a single realisation of incidence. However, each posterior estimate of ρ is in fact only the mean of an underlying offspring distribution (i.e. the distribution of the number of secondary infections an infected person produces). If an epidemic governed by identical parameters were to happen again, but with different random draws of infection events, each realisation would be different, thus giving aleatoric uncertainty.

When performing inference, infectious disease models tend to consider epistemic uncertainty only due to the difficulties in performing inference with aleatoric uncertainty (e.g. individual-based mod-

els) or analytical tractability. There are many exceptions such as the susceptible-infected-recovered model, which has stochastic variants that are capable of determining aleatoric uncertainty [414] and have been used in extensive applications (e.g. [419]). However, we will show that this model can underestimate uncertainty under certain conditions. An empirical alternative is to characterise aleatoric uncertainty by the final epidemic size from multiple historical outbreaks [420, 421] but these are confounded by temporal, cultural, epidemiological, and biological context, and therefore parameters vary between each outbreak. Here, following previous approaches [414], we analyse aleatoric uncertainty by studying an epidemiologically-motivated stochastic process, serving as a proxy for repeated realisations of an epidemic. Within our framework, we find that using epistemic uncertainty alone is a vast underestimate, and accounting for aleatoric uncertainty shows potential risk to be much higher. We demonstrate our method using two historical examples: firstly the 2003 severe acute respiratory syndrome (SARS) outbreak in Hong Kong, and secondly the early 2020 UK COVID-19 epidemic.

6.2 Results

6.2.1 An analytical framework for aleatoric uncertainty

A time-varying general branching processes proceeds as follows: first, an individual is infected, and their infectious period is distributed with probability density function g (with corresponding cumulative distribution function G). Second, while infectious, individuals randomly infect others (via a counting process with independent increments), driven by their infectiousness ν and a rate of infection events ρ. That is, an individual infected at time l, will, at some later time while still infectious t, generate secondary infections at a rate $\rho(t)\nu(t-l)$. $\rho(t)$ is a population-level parameter closely related to the time-varying reproduction number $R(t)$ (see Methods and [60] for further details), while $\nu(t-l)$ captures the individual's current infectiousness (note that $t-l$ is the time since infection). We allow multiple infection events to occur simultaneously, and assume individuals behave independently once infected, thus allowing mathematical tractability [422]. Briefly, we model an individual's secondary infections using a stochastic counting process, which gives rise to secondary infections (i.e. offspring) that are either Poisson or Negative Binomial distributed in their number (see Appendix D). We study the aggregate of these events (prevalence or incidence) through closed-form probability generating functions and probability mass functions. Our approach models epidemic evolution through intuitive individual-level characteristics while retaining analytical tractability. Importantly, the mean of our process follows a renewal equation [423, 60, 424]. Our formulation unifies mechanistic and individual-based modelling within a single analytical framework based on branching processes. Figure 6.1 shows

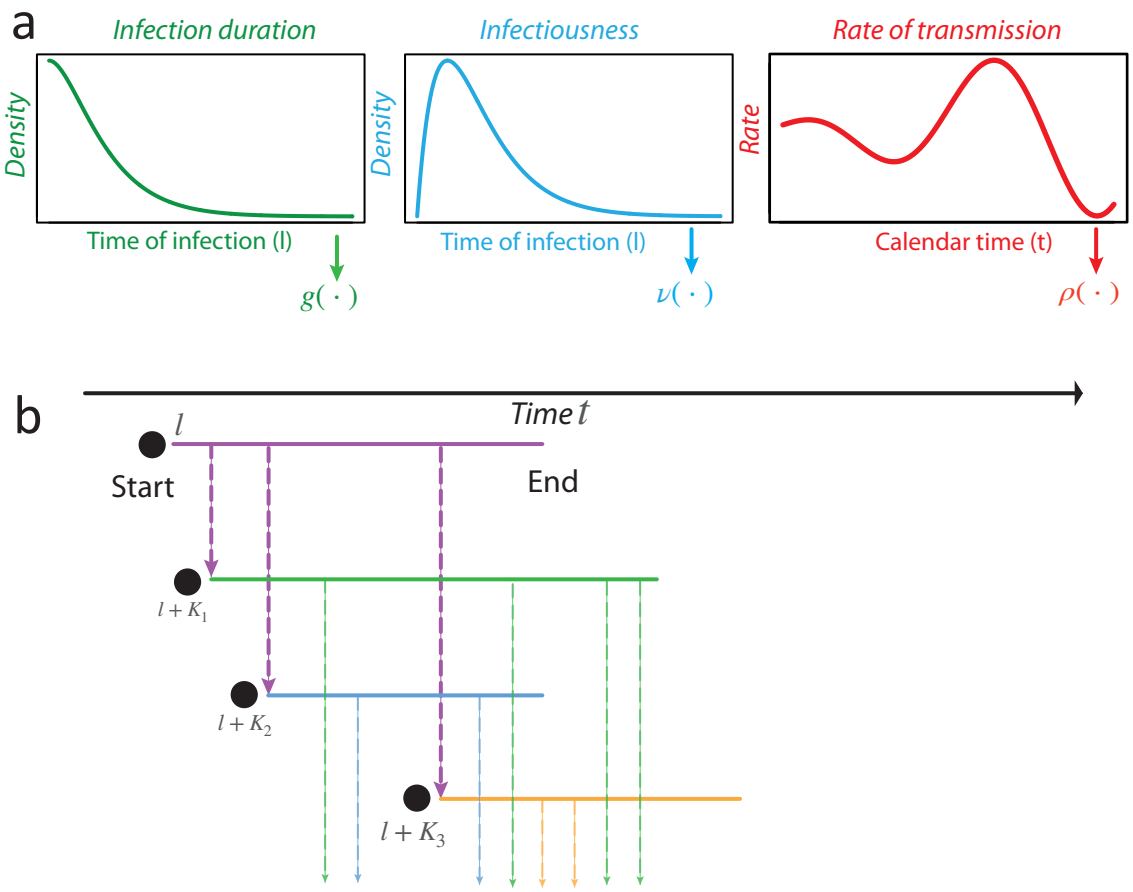


Figure 6.1: Schematic of a time-varying general branching process. (a) shows schematics for the infectious period, an individual's time-varying infectiousness (both functions of time post infection t^*), and the population-level mean rate of infection events. The infectious period is given by probability density function g. For each individual their (time-varying) infectiousness and rate of infection events are given by ν and ρ respectively. In an example (b), an individual is infected at time l, and infects three people (random variables K, purple dashed lines) at times $l + K_1$, $l + K_2$ and $l + K_3$. The times of these infections are given by a random variable with probability density function $\sim \frac{\rho(t)\nu(t-l)}{\int_l^t \rho(u)\nu(u-l)du}$. Each new infection then has its own infectious period and secondary infections (thinner coloured lines).

a schematic of this process. Formal derivation is in Appendix D.

Randomness occurs at individual level, and there is a distribution of possible realisations of the epidemic given identical parameters. Simulating our general branching process would be cumbersome using the standard approach of Poisson thinning [425], and inference from simulation is more challenging still. Using probability generating functions, we analytically derive important quantities from the distribution of the number of infections, including the (central) moments and marginal probabilities given ρ, g and ν (with or without epistemic uncertainty). We additionally use the probability generating function to prove general, closed-form, analytical results such as the decomposition of variance into mechanistic components, and the conditions under which overdispersion exists (i.e. where variance is greater than the mean). Finally, we derive a general probability mass function (likelihood function) for incidence.

If infection event $k = 0, \ldots, n$ occurred at time τ_k and produced y_k infections, let x_{kj} denote the end time of the infectious period of the j^{th} infection at event k. Note that $\tau_0 = l$ is the time of the first infection event and $y_0 = 1$. Then the likelihood $L_{\text{InfPeriod}}$ of each infected person's infectious period is a product over all infections given by

$$L_{\text{InfPeriod}} = \prod_{k=0}^{n} \prod_{j=1}^{y_k} g(x_{kj} - \tau_k, \tau_k). \tag{6.1}$$

The likelihood of there being y_k infections at time τ_k is given by

$$L_{\text{InfTime}} = \prod_{k=1}^{n} \left(\sum_{i=0}^{k-1} \sum_{j=1}^{y_i} \mathbb{1}_{\{x_{ij} < \tau_k\}} p_{y_k}(\tau_k, \tau_i) \right), \tag{6.2}$$

where $p_{y_k}(\tau_k, \tau_i)$ is the (infinitesimal) rate at which an individual infected at τ_i causes y_k infections at time τ_k, provided it is still infectious. Finally, the probability that no other infections occurred between the infection events at times $(\tau_k)_{k=0}^{n}$ is given by

$$L_{\text{Only}} = \exp\left(- \sum_{i=0}^{n} \sum_{j=1}^{y_i} \int_{\tau_i}^{\min(t, x_{ij})} r(u, \tau_i) du \right), \tag{6.3}$$

where r is the infection event rate and t is the current time. Note the term $\exp(-x)$ comes from our assumption that infection events occur according to inhomogeneous Poisson Processes. Our full likelihood L_{Full} is then

$$L_{\text{Full}} = L_{\text{InfPeriod}} \times L_{\text{InfTime}} \times L_{\text{Only}}. \tag{6.4}$$

Full derivations of these quantities are provided in Appendix D. If discrete time is assumed, equation 6.4 simplifies to a likelihood commonly used for inference [426]. Markov Chain Monte Carlo can be used on equation 6.4 to sample aleatoric incidence realisations, but it is often simpler to solve the probability generating function with complex integration. The probability generating function, equations for the variance, and derivations of the probability mass function are found in Appendix D, and a summary of the main analytical results is found in the Methods.

6.2.2 The dynamics of uncertainty

We derive the mean and variance of our branching process. The general variance Equation 6.9 (see Methods) captures uncertainty in prevalence over time, where individual-level parameters govern each infection event. This equation comprises three terms: the timing of secondary infections from the infectious period (Equation 6.9a); the offspring distribution (Equation 6.9b); and propagation of un-

certainty through the descendants of the initial individual (Equation 6.9c). Importantly, this last term depends on past variance, showing that the infection process itself contributes to aleatoric variance, and does not arise only from uncertainty in individual-level events. In short, unlike common Gaussian stochastic processes, the general variance in disease prevalence is described through a renewal equation. Therefore, future uncertainty depends on past uncertainty, and so the uncertainty around subsequent epidemic waves has memory. Additionally, uncertainty is driven by a complex interplay of time-varying factors, and not simply proportional to the mean. For example, a large first wave of infection can increase the variance of the second wave. As such, the general variance equation 6.9 disentangles and quantifies the causes of uncertainty, which remain obscured in brute-force simulation experiments [414].

Consider a toy simulated epidemic with $\rho(t) = 1.4 + \sin(0.15t)$, where the offspring distribution is Poisson in both timing and number of secondary infections, and where infectiousness ν is given by the probability density function $\nu \sim \text{Gamma}(3, 1)$, and, similarly, the infectious period $g \sim \text{Gamma}(5,1)$. Here the parameters of the Gamma distribution are the shape and scale respectively. The resulting variance is counterintuitive. We prove analytically that overdispersion emerges despite a non-overdispersed Poisson offspring distribution. The second wave has a lower mean but a higher variance than the first wave (Figure 6.2), because uncertainty is propagated. If the variance were Poisson, i.e. equal to the mean, the second wave would instead have a smaller variance due to fewer infections. Initially, uncertainty from individuals is largest, but as the epidemic progresses, compounding uncertainty propagated from the past dominates [Figure 6.2, bottom right]. Note that in this example with zero epistemic uncertainty (we know the parameters perfectly), aleatoric uncertainty is large.

In Equation 6.9, the first two terms account for uncertainty in the infectious periods of all infected individuals. The third term denotes the uncertainty from the offspring distribution. By construction, the timing of infections is an inhomogenous Poisson process, where at each infection time the number of infections is random. The third term (Equation 6.9b) contains the second moment of the offspring distribution, which is the variability around its mean (i.e. $\rho(t)$). The second moment quantifies the extent of possible superspreading. In contrast to other studies [427, 428], we find that individual-level overdispersion in the offspring distribution is less important than explosive epidemics. Under a null Poisson model, with no overdispersion (see Poisson case in Figure 6.2), substantial aleatoric uncertainty arises from a Poisson offspring distribution combined with variance propagation. We rigorously prove via the Cauchy-Schwarz inequality that, under a mild condition on the possible spread of the epidemic, the variance of number of infections at a given time is always greater than the mean, and hence is overdispersed. Overdispersion in the offspring infection distribution is therefore not necessary for high

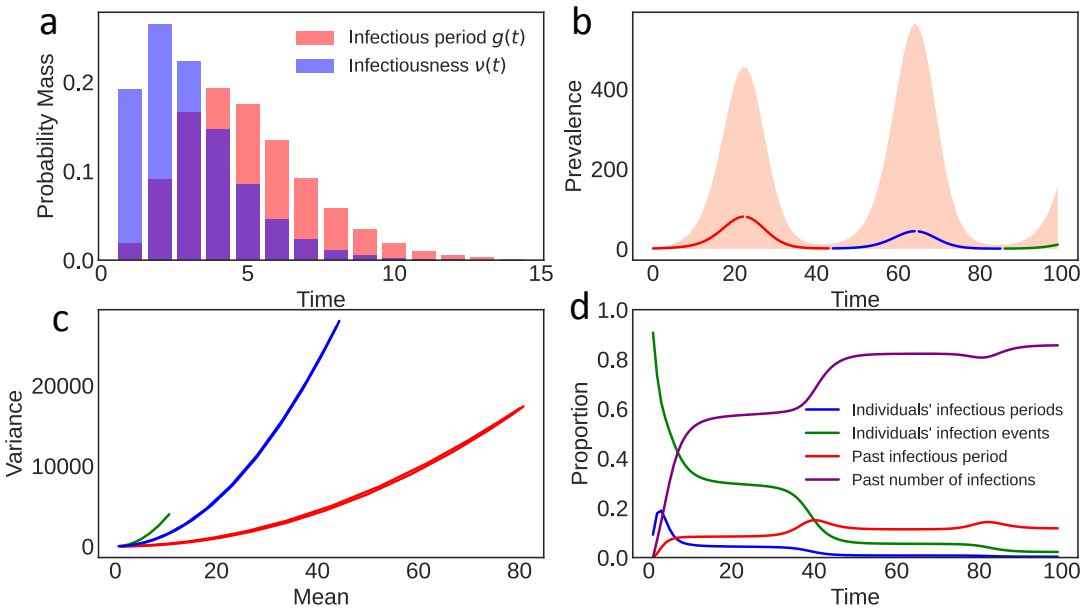


Figure 6.2: Aleatoric uncertainty without overdispersed offspring distribution. Plots show simulated epidemic where $\rho(t) = 1.4+\sin(0.15t)$, with a Poisson offspring distribution. We use infectiousness $\nu \sim$ Gamma$(3, 1)$, and infectious period $g \sim$ Gamma(5,1). (a) Overlap between g and the infectiousness ν, where g controls when the infection ends e.g. by isolation. (b) Predicted mean and 95% aleatoric uncertainty intervals for prevalence. Note there is no epistemic uncertainty as the parameters are known exactly. (c) Phase plane plot showing the mean plotting against the variance. (d) Proportional contribution to the variance from the individual terms in Equation 6.9. Compounding uncertainty from past events is the dominant contributor to overall uncertainty.

aleatoric uncertainty, although it still increases variance at both individual-level and population-level.

We derive the conditional variance, with known past events but unknown future events. Conditional variance grows proportionally to the square of the mean, with additional terms containing the previous variance. Therefore, aleatoric uncertainty grows and forecasting exercises based only on epistemic uncertainty greatly underestimate the risk of very large epidemics, and this underestimation becomes more severe as the forecast horizon expands or as the epidemic grows.

6.2.3 Aleatoric uncertainty over the SARS 2003 epidemic

To demonstrate the importance of aleatoric uncertainty, we analyse daily incidence of symptom onset in Hong Kong during the 2003 severe acute respiratory syndrome (SARS) outbreak [429, 426, 430]. The epidemic struck Hong Kong in March-May 2003, with a case fatality ratio of 15%. We fit a Bayesian renewal equation assuming a random walk prior distribution for the rate of infection events ρ [60], using Equation 6.4 for inference. We ignore g and assume that the distribution of generation times mirrors the distribution of infectiousness, i.e. that the infectiousness ν equals the generation time [429]. Note these parameter choices are illustrative and do not affect our main conclusions. The fitted $\rho(t)$ in Figure 6.3 (top left) shows two major peaks, consistent with the major transmission events in the epidemic

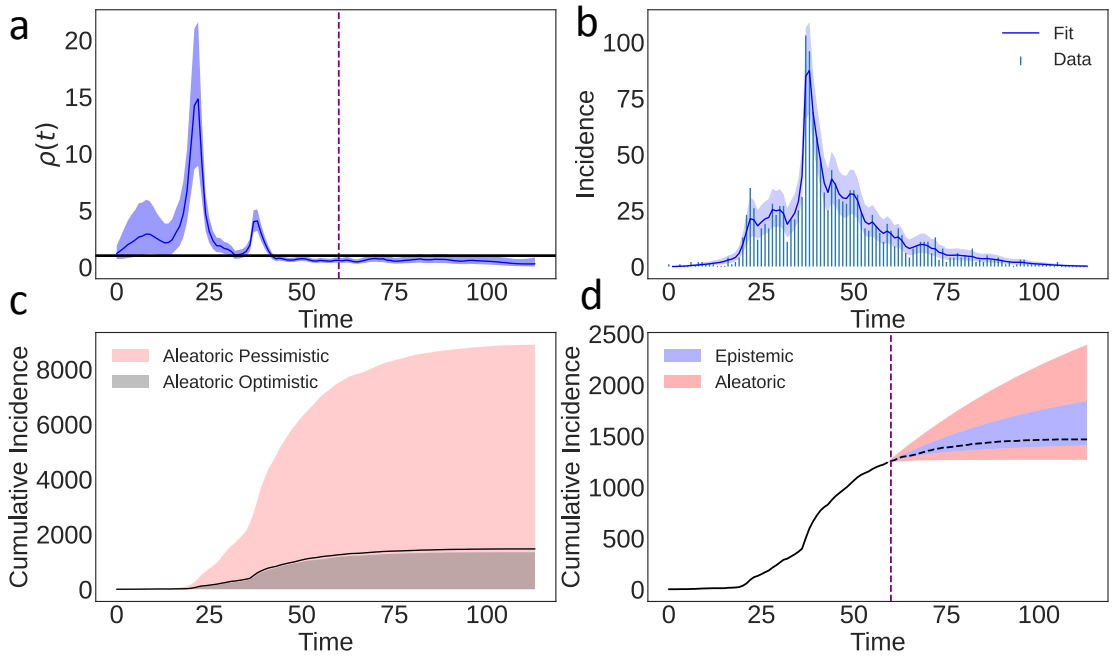


Figure 6.3: The 2003 SARS epidemic in Hong Kong [429, 426]. (a) $\rho(t)$ with 95% epistemic uncertainty. (b) Fitted incidence mean, 95% epistemic uncertainty with observational noise from using Equation 6.4. Data is daily incidence of symptom onset. (c) Aleatoric uncertainty from the start of the epidemic under an optimistic and pessimistic $\rho(t)$. (d) Epistemic (blue) and epistemic and aleatoric uncertainty (red) while keeping ρ constant at the forecast data (dotted line). Forecasting is from day 60.

[430]. Figure 6.3 (top right) shows the mean epistemic fit, with epistemic (posterior) uncertainty tightly distributed around the data. Figure 6.3 (bottom left) shows the aleatoric uncertainty under optimistic and pessimistic scenarios (i.e. the upper and lower bounds of $\rho(t)$ in Figure 6.3 (top right)). The pessimistic scenario includes the possibility of extinction, but also an epidemic that could have been more than six times larger than that observed. The optimistic scenario suggests we would observe an epidemic of at worst comparable size to that observed. Finally, Figure 6.3 (bottom right) shows epistemic and aleatoric forecasts at day 60 of the epidemic, fixing $\rho(t)$ using the 95% epistemic uncertainty interval to be constant at either $\rho(t \geq 60) = 0.38$ or $\rho(t \geq 60) = 0.83$ and simulating forwards. While the epistemic forecast does contain the true unobserved outcome of the epidemic, it underestimates true forecast uncertainty, which is 1.3 times larger. The range of the constant ρ for forecast is below 1, and yet we still see substantial aleatoric uncertainty. If ρ were above 1 for a sustained period, aleatoric uncertainty would play a smaller role [431], but this is rare with real epidemics, where susceptible depletion, behavioural changes or interventions keep ρ around 1. Our results therefore highlight that epistemic uncertainty drastically underestimates potential epidemic risk.

6.2.4 Aleatoric risk assessment in the early 2020 COVID-19 pandemic in the UK

To demonstrate the practical application of our model, we retrospectively examine the early stage of the COVID-19 pandemic in the UK, using only information available at the time. While the date of the first locally transmitted case in the UK remains unknown (likely mid-January 2020 [432]), COVID-19 community transmission was confirmed in the UK by late January 2020, and we therefore start our simulated epidemic on January 31st 2020. We consider uncertainty in the predicted number of deaths on March 16th 2020 [25], during which time decisions regarding non-pharmaceutical interventions were made. Testing was extremely limited during this period, and COVID-19 death data were unreliable. For this illustration, we assume that we did not know the true number of COVID-19 deaths, as was the case for many countries in early 2020. Policymakers then needed estimates of the potential death toll, given limited knowledge of COVID-19 epidemiology and unreliable national surveillance.

We simulated an epidemic from a time-varying general branching process with a Negative Binomial offspring distribution, using parameters that were largely known by March 16th 2020 (Table 6.1). The infection fatality ratio, infection-to-onset distribution and onset-to-death distribution were convoluted with incidence [60] to estimate numbers of deaths. Estimated COVID-19 deaths and uncertainty estimates between January 31st and March 16th 2020 are shown in Figure 6.4 (Top). While the epistemic uncertainty contains the true number of deaths, it is still an underestimate, and including aleatoric uncertainty, we find that the epidemic could have had more than four times as many deaths. Consider a hypothetical intervention on March 17th 2020 (Figure 6.4 (bottom)) that completely stops transmission. Deaths would still occur from those already infected but no new infections would arise. In this hypothetical case, the aleatoric uncertainty would still be 2.5 times the actual deaths that occurred (when in fact transmission was never zero or close to it). This hypothetical scenario highlights the scale of aleatoric uncertainty, and demonstrates that our method can be useful in assessing risk in the absence of data by giving a reasonable worst case. Further, we observe that using only epistemic uncertainty provides a reasonably good fit in a relatively short time-horizon (Figure 6.4, Top), but soon afterwards greatly underestimates uncertainty (Figure 6.4, Bottom). The fits using aleatoric uncertainty provide a more reasonable assessment of uncertainty. While we concentrate on the upper bound, the lower bound on the worst-case scenario still exceeds zero, and therefore the epidemic going extinct by March 16th in the worst-case with no external seeding would have been very unlikely. Aleatoric uncertainty highlights a more informative reasonable worst-case estimate than epistemic uncertainty alone, and could be a useful metric for a policymaker in real time, with low-quality data, without requiring simulations from costly, individual-based models.

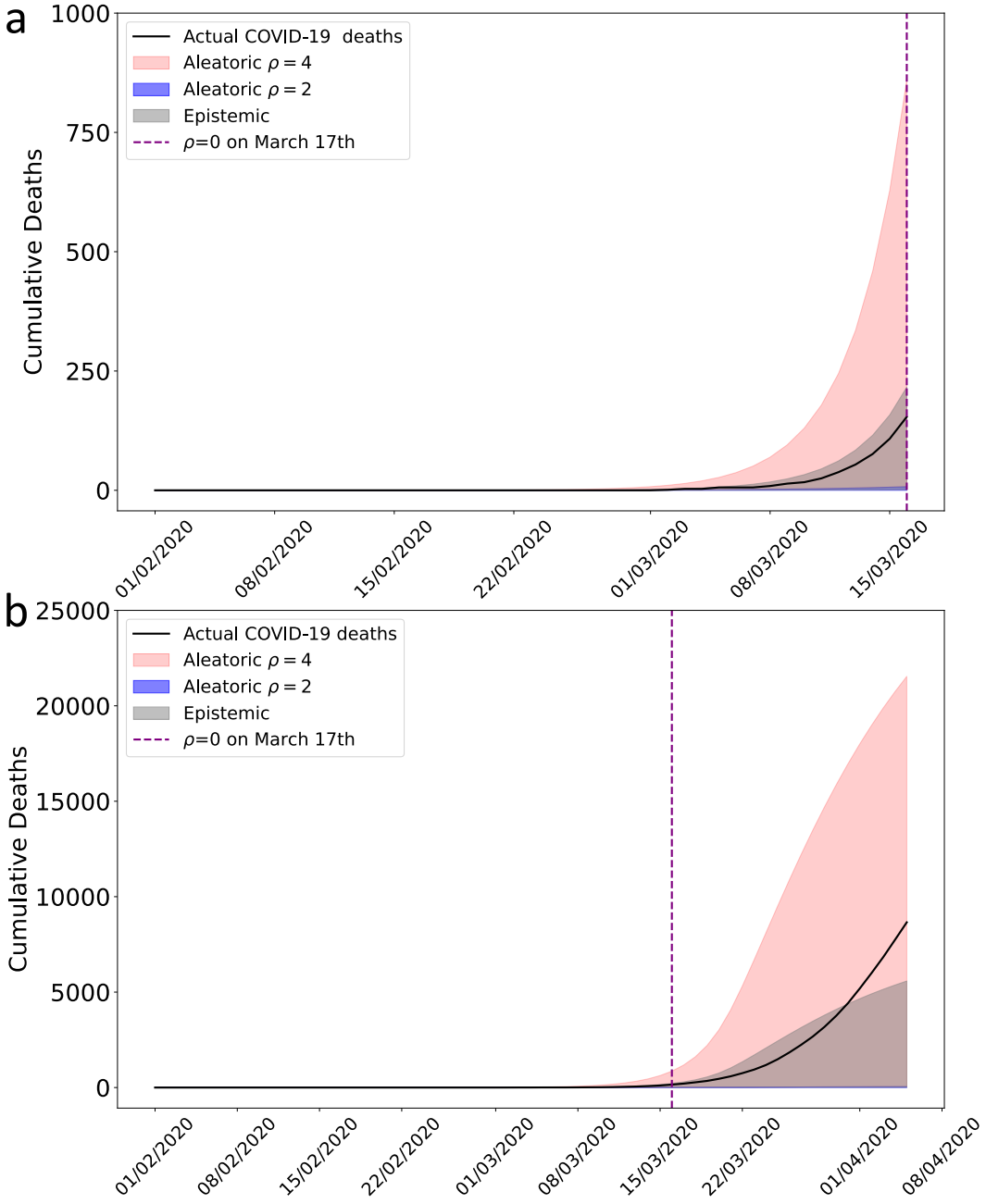


Figure 6.4: Early 2020 COVID-19 pandemic in the UK. (a) shows a simulated epidemic using parameters available on March 16th 2020 (Table 6.1), for a plausible range of $\rho = R_0$ between 2 and 4. The black lines indicate actual COVID-19 deaths, which we assume no knowledge of. The purple line is March 17th 2020, we set transmission to zero i.e. $\rho = 0$, to simulate an intervention that stops transmission completely. The grey envelope is the epistemic uncertainty and the red envelope the aleatoric uncertainty. (b) is the same as the top plot, except time is extended past March 17th with transmission being zero.

Epidemiological Parameter	Value or Distribution	Citation
Infection Fatality Ratio	0.9%	[433, 434]
Basic Reproduction Number	$2-4$	[25, 435]
Serial Interval Distribution	$\sim$ Gamma$(7.82, \frac{1}{0.62})$	[433, 24, 436]
Onset-to-Death Distribution	$\sim$ Gamma$(1.45, 10.43)$	[433, 191]
Infection-to-Onset Distribution	$\sim$ Gamma$(35.16, 6.9)$	[24, 433]
Overdispersion Coefficient	0.53	[437]

Table 6.1: Epidemiological parameters available on March 16th 2020 used in branching process simulation

6.3 Discussion

Stochastic models more realistically model natural phenomena than deterministic equations [438], and particularly so with infection processes [439]. Accordingly, individual-based models have found much success [440, 441] in capturing the complex dynamics that emerge from infectious disease outbreaks, and have been highly influential in policy [25]. However, despite a plethora of alternatives, many analytical frameworks still tend to be deterministic [426, 24, 266], and only consider statistical, epistemic parameter uncertainty. Frameworks that expand deterministic, mechanistic equations to include stochasticity use a Gaussian noise process [414], or restrict the process to be Markovian. Markovian branching processes require the infection period or generation time to be exponentially distributed - a fundamentally unrealistic choice for most infectious diseases. Further, a Gaussian noise process is unlikely to be realistic [421].

Our results show that individual-level uncertainty is overshadowed by uncertainty in the infection process itself. Profound overdispersion in infectious disease epidemics is not simply a result of overdispersion in the offspring distribution, but is fundamental and inherent to the branching process. We rigorously prove that even with a Poisson offspring distribution (not characterised by overdispersion), overdispersion in resulting prevalence or incidence is still virtually always guaranteed. We show that forecast uncertainty increases rapidly, and therefore common forecasting methods almost certainly underestimate true uncertainty. Similar to other existing frameworks, our approach provides a different methodological tool to evaluate uncertainty in the presence of little to no data, assess uncertainty in forecasting, and retrospectively assess an epidemic. Other approaches, such as agent based models, could also be readily used. However, the framework we present permits the unpicking of dynamics analytically and from first principles without a black box simulator. Equally, this is also a limitation, since new and flexible mechanisms cannot be easily integrated or considered.

We have considered only a small number of mechanisms that generate uncertainty. Cultural, behavioural and socioeconomic factors could introduce even greater randomness. Therefore, our frame-

work may underestimate true uncertainty in infectious disease epidemics. The converse is also likely, contact network patterns and spatial heterogeneity also limit the routes of transmission, such that the variability in anything but a fully connected network will be lower. Furthermore, our assumption of homogeneous mixing and spatial independence overestimates uncertainty. A sensible next step for future research to study the dynamics of these branching processes over complex networks. Finally, at the core of all branching frameworks, is an assumption of independence, which is unlikely to be completely valid (people mimic other people in their behaviour) but is necessary for analytical tractability. Studying the effect of this assumption compared to agent-based models would also be a useful area of future research.

We provide one approach to determining aleatoric uncertainty. Other approaches based on stochastic differential equations, Markov processes, reaction kinetics, or Hawkes processes all have their respective advantages and disadvantages. The differences in model-specific aleatoric uncertainty and how close the models come to capturing the true, unknown, aleatoric uncertainty is a fundamental question moving forwards. In this paper we have provided yet another approach to characterise aleatoric uncertainty, where this approach is most useful and how it can be reconciled with existing approaches will be an interesting area of study.

6.4 Methods

Detailed derivations of the methods can be found in Appendix D.

A time-varying general branching process proceeds as follows: first, a single individual is infected at some time l, and their infectious period L is distributed with probability density function g (and cumulative distribution function G). Second, during their infectious period, they randomly infect other individuals, affected by their infectiousness $\nu(t-l)$, and their mean number of secondary infections, which is assumed to be equal to the population-level rate of infection events $\rho(t)$. $\rho(t)$ is closely related to the time-varying reproduction number $R(t)$ (see [60] for details). The infectious period g accounts for variation in individual behaviour. If people take preventative action to reduce onward infections, their reduced infection period can stop transmission despite remaining infectious. Where infectious individuals do not change their behaviour, g can be ignored and individual-level transmission is controlled by infectiousness ν only. Each newly infected individual then proceeds independently by the same mechanism as above. Specifics can be found in Appendix D.

Formally, if an individual is infected at time s, their number of secondary infections is given by a stochastic counting process $\{N(t,s)\}_{t\geq s}$, which is independent of other individuals and has independent increments. We assume here that the epidemic occurs in continuous time, and hence that

$N(t,s)$ is continuous in probability, although we consider discrete-time epidemics in Appendix D. To aid calculation, we suppose $N(t,s)$ can be defined from a Lévy Process $\mathcal{N}(t)$ - that is, a process with both independent and identically distributed increments - via $N(t,s) = \mathcal{N}\left(\int_s^t r(k,s)dk\right)$ for some non-negative rate function r. It is assumed that each counting process $\{N(t,s)\}_{t\geq s}$ is defined from an independent copy of $\mathcal{N}(t)$. This formulation has two advantages: first, the dependence of $N(t,s)$ on s is restricted to the rate function r; and second, if $J_{\mathcal{N}}(t)$ counts the number of infection events in $\mathcal{N}(t)$ (where here infection events refer to an increase, of any size, in $N(t,s)$), then $J_{\mathcal{N}}(t)$ is a Poisson process with some rate κ [442]. We can then define $J(t,s)$ to be the counting process of infection events in $N(t,s)$, and $Y(v)$ to be size of the infection event (i.e. the number of secondary infections that occur) at time v. We assume that Y is independent of s, although such a dependence would curtail superspreading to depend on infectiousness, and could be incorporated into the framework. Therefore, $J(t,s)$ is an inhomogeneous Poisson Process (and so $N(t,s)$ has been characterised as an inhomogeneous compound Poisson Process). We consider the cases where $N(t,s)$ is itself an inhomogeneous Poisson process, and where $N(t,s)$ is a Negative Binomial process. This allows us to examine effects of overdispersion in the number of secondary infections, although our framework allows for more complicated distributions.

Here, $r(t,l) = \rho(t)\nu(t-l)$ where $\rho(t)$ models the population-level rate of infection events, and $\nu(t-l)$ models the infectiousness of an individual infected at time l. If $\nu(t-l)$ is sufficiently well characterised by the generation time (i.e. where the timing of secondary infections mirrors tracks their infectiousness), and the infectious period can be ignored, then the integral $\int_l^t r(s,l)ds$ has the same scale as the commonly used reproduction number $R(t)$ [60]. The branching process yields a series of birth and death times for each individual (i.e. the time of infection and the end of the infectious period respectively), from which prevalence (the number of infections at any given time) or cumulative incidence (the total number of infections up to any time) can be defined.

6.4.1 Probability generating function

We derive the probability generating function for a time-varying age-dependent branching process, allowing derivation of the mean and higher-order moments (full derivations can be found in Appendix D). We consider two special cases for the number of new infections $Y(v)$ at each infection event: firstly, $Y(v) = 1$, meaning the secondary infections form a Poisson Process, and, secondly, a logarithmic (log series) distribution yielding a Negative Binomial process of secondary infections. In both cases, we assume that the distribution of $Y(v)$ is equal for all values of v. In the Poisson case, the number of new infections at each infection time is, by definition, one. Therefore, the number of infections

an individual creates is Poisson distributed, and closely clustered around the mean rate of infection events. Conversely, the Negative Binomial case, more realistically allows multiple infections to occur at each infection time, and so the number of infections an individual causes is overdispersed. The pgf (probability generating function), $F(t,l;s) = E(s^{Z(t,l)})$ of some function $Z(t,l)$ such as the prevalence or cumulative incidence, can be derived by conditioning on the lifetime, L, of the first individual. That is,

$$E\left(s^{Z(t,l)}\right) = \left(1 - G(t-l,l)\right) E\left(s^{Z(t,l)} \middle| L \geq t-l\right) + \int_0^{t-l} E\left(s^{Z(t,l)} \middle| L = u\right) g(u,l)du \tag{6.5}$$

Note that if the individuals directly infected by the initial individual are infected at times $l+t_1, ..., l+t_n$, then

$$Z(t,l) = 1 + \sum_{i=1}^{n} Z(t, l+t_i) \tag{6.6}$$

This observation allows us to write the generating function $F(t,l)$ as a function of $F(t,u)$ for $u \in (t,l)$. As $F(t,t) = s$, this allows us to iteratively find the value of $F(t,l)$. Explicitly, we have

$$\underbrace{F(t,l;s)}_{\text{pgf}} = \underbrace{(1 - G(t-l,l))}_{P(L>t-l)} q_1 \underbrace{\left(\int_0^{t-l} \overbrace{f(\underbrace{F(t,l+k;s)}_{\text{pgf of process started at } l+k})}^{\text{pgf of } Y} \overbrace{\rho(l+k)\nu(k)dk}^{\text{infection rate at time } l+k} \right)}_{\text{pgf of } J(t,l)}$$

$$+ \int_0^{t-l} \underbrace{q_2 \left(\int_0^{u} \overbrace{f(\underbrace{F(t,l+k;s)}_{\text{pgf of process started at } l+k})}^{\text{pgf of } Y} \overbrace{\rho(l+k)\nu(k)}^{\text{infection rate at time } l+k} dk \right)}_{\text{pgf of } J(l+u,l)} \underbrace{g(u,l)du}_{P(L=l+u)}, \tag{6.7}$$

where $q_1(z;s) = se^z$, and where $q_2(z) = e^z$ in the case where $Z(t,l)$ refers to prevalence, whereas $q_2(z;s) = se^z$ in the case where $Z(t,l)$ refers to cumulative incidence. Note also that $f(z) = z - 1$ in the Poisson case and $f(z) = -\phi\left[\log\left(1 - (1 - \frac{\phi}{1+\phi}z)\right) - \log\left(\frac{\phi}{1+\phi}\right)\right]$ in the log-series case and that the constant κ is absorbed into ρ.

The key intuition in understanding Equation 6.7 is that for an integer random variable X and iid (independent and identically distributed) random variables Y_i, $E(s^{\sum_{i=1}^{X} Y_i}) = G_X(G_Y(s))$, where G_X and G_Y are the generating functions of X and Y_i respectively. Thus, we expect the pgfs of the various parts of our model to combine via composition, as occurs in the equation above.

Mean incidence can recovered from both prevalence (via back calculation [60]) and cumulative incidence. In Equation 6.7 for the Negative Binomial case, ϕ is the degree of overdispersion. Equation 6.7 is solvable using via quadrature and the fast Fourier transform via a result from complex analysis

[443] and scales easily to populations with millions of infected individuals, and the probability mass function can be computed to machine precision (a full derivation is available in Appendix D).

6.4.2 Variance decomposition

For simplicity, we only summarise the decomposition for prevalence, but an analogous and highly similar derivation for cumulative incidence can be found in Appendix D. We can derive an analytical equation for the mean and variance of the entire branching process (full derivations can be found in Appendix D as well as the mathematical properties of the variance equations). The mean prevalence $M(t,l)$ is given by

$$M(t,l) = (1 - G(t-l,l)) + \int_0^{t-l} M(t,l+u)\rho(l+u)\nu(u)\mathbb{E}(Y)(1-G(u,l))du. \tag{6.8}$$

Note, ρ can be scaled to absorb the $\mathbb{E}(Y)$ and κ constants. Equation 6.8 is consistent with that previously derived in [60]. The second moment, $W(t,l) := \mathbb{E}(Z(t,l)(Z(t,l)-1))$ allows us to determine the variance, $V(t,l)$ as $V(t,l) = W(t,l) + M(t,l) - M(t,l)^2$. The variance can be decomposed into three mechanistic components.

$$\begin{aligned} V(t,l) = & \underbrace{\int_0^{t-l} \left[\int_0^u M(t,l+k)\rho(l+k)\nu(k)dk\right]^2 g(u,l)du - M(t,l)^2}_{\text{(9a): uncertainty from the infectious period}} \\ & + \underbrace{(1-G(t-l,l))\left[1 + 2\int_0^{t-l} M(t,l+u)\rho(l+u)\nu(u)du + \left(\int_0^{t-l} M(t,l+u)\rho(l+u)\nu(u)du\right)^2\right]}_{\text{(9a continued): uncertainty from the infectious period}} \\ & + \underbrace{\int_0^{t-l} M(t,l+u)^2\mathbb{E}(Y^2)\rho(l+u)\nu(u)(1-G(u,l))du}_{\text{(9b): uncertainty from the offspring distribution}} \\ & + \underbrace{\int_0^{t-l} V(t,l+u)\rho(l+u)\nu(u)(1-G(u,l))du}_{\text{(9c): uncertainty propagated from the past}}. \end{aligned} \tag{6.9}$$

The general variance equation 6.9 captures the evolution of uncertainty in population-level disease prevalence over time, where fixed individual-level disease transmission parameters govern each infection event. Unlike the simple Galton-Watson process, we find that previously unknown factors also determine aleatoric variation in disease prevalence. Specifically, the general variance equation 6.9 comprises three terms, one for the infectious period (Equation 6.9a), one for the number and timing of secondary infections (Equation 6.9b), and a term that propagates uncertainty through descendants of the initial individual (Equation 6.9c). Importantly, the last term (Equation 6.9c) depends on past variance, showing that the infection process itself contributes to aleatoric variance, and this is distinct

from the uncertainty in individual infection events. In short, and unlike Gaussian stochastic processes, the general variance in disease prevalence is described through a renewal equation. Intuitively then, uncertainty in an epidemic's future trajectory is contingent on past infections, and the uncertainty around consecutive epidemic waves are connected. As such, the general variance equation 6.9 allows us to disentangle important aspects of infection dynamics that remain obscured in brute-force simulations [414].

6.4.3 Overdispersion

We define an epidemic to be expanded if at time t there is a non-zero probability that the prevalence, not counting the initial individual or its secondary infections, is non-zero.

Note that this is a very mild condition on an epidemic - in a realistic setting, the only way for an epidemic to not be expanded is if it is definitely extinct by time t, or if t is small enough that tertiary infections have not yet occurred.

Large aleatoric variance intrinsic to our branching process implies that the prevalence of new infections (that is, prevalence excluding the deterministic initial case) is always strictly overdispersed at time t, providing the epidemic is expanded at time t. A full proof is given in Appendix D, but we provide here a simpler justification in the special case that $G(t-l,l) = 1$.

In this case, prevalence of new infections is equal to standard prevalence, and the equations for $M(t,l)$ and $V(t,l)$ simplify significantly. Switching the order of integration in the equation for $M(t,l)$ gives

$$M(t,l) = \int_0^{t-l} M(t,l+u)\rho(l+u)\nu(u)\mathbb{E}(Y)(1-G(u,l))du \tag{6.10}$$

$$= \int_0^{t-l} \left[\int_0^u M(t,l+k)\rho(l+k)\mathbb{E}(Y)\nu(k)dk\right] g(u,l)du \tag{6.11}$$

and hence, the Cauchy-Schwarz Inequality shows that

$$M(t,l)^2 \leq \left(\int_0^{t-l} \left[\int_0^u M(t,l+k)\mathbb{E}(Y)\rho(l+k)\nu(k)dk\right]^2 g(u,l)du\right) \tag{6.12}$$

as $\int_0^{t-l} g(u,l)du = 1$ (note that this follows from splitting the integrand into $\int_0^u M(t,l+k)\rho(l+k)\mathbb{E}(Y)\nu(k)dkg(u,l)^{0.5}$ and $g(u,l)^{0.5}$). Thus, noting $1-G(t-l,l) = 0$, the first term, (6.9a), in the variance equation is non-negative.

The remaining terms can be dealt with as follows. The sum of (6.9b) and (6.9c) is (using $Y(l+u,l)^2 \geq Y(l+u,l)$) bounded below by $\int_0^{t-l} \mathbb{E}(Z(t,l+u)^2)\mathbb{E}(Y)\rho(l+u)\nu(u)(1-G(u,l))du$. Finally, noting that $Z(t,l+u)^2 \geq Z(t,l+u)$, this is bounded below by $\int_0^{t-l} M(t,l+u)\mathbb{E}(Y)\rho(l+u)\nu(u)(1-G(u,l))du =$

$M(t,l)$. Hence, $V(t,l) \geq M(t,l)$ holds.

To show strict overdispersion, note that for $V(t,l) = M(t,l)$ to hold, it is necessary that

$$\int_0^{t-l} \mathbb{E}(Z(t,l+u)^2)\mathbb{E}(Y)\rho(l+u)\nu(u)(1-G(u,l))du = \int_0^{t-l} M(t,l+u)\mathbb{E}(Y)\rho(l+u)\nu(u)(1-G(u,l))du \tag{6.13}$$

and hence, for each u (as $\mathbb{E}(Y) > 0$)

$$\mathbb{E}\Big[Z(t,l+u)(Z(t,l+u)-1)\Big] = 0 \quad \text{or} \quad \rho(l+u)\nu(u)(1-G(u,l)) = 0 \tag{6.14}$$

If new infections can be caused, then more than one new infection can be caused by the continuity in probability of the secondary infection process. Thus, if an individual infected at $l+u$ has $\mathbb{E}\Big[Z(t,l+u)(Z(t,l+u)-1)\Big] = 0$, this individual cannot cause new infections whose infection trees have non-zero prevalence at time $l+u$. Hence, the condition 6.14 is equivalent to the epidemic being non-expanded at time t, as at each time $l+u$, either no infections are possible from the initial individual (the second condition being equal to zero), or any individuals that are infected at time $l+u$ contribute zero prevalence at time t from the new infections they cause.

Hence, $Z(t,l)$ is strictly overdispersed for expanded epidemics. This means that Gaussian approximations are unlikely to be useful.

6.4.4 Variance midway through an epidemic

It is important to calculate uncertainty starting midway through an epidemic, conditional on previous events. This derivation is significantly more algebraically involved than the other work in this paper. For simplicity, we assume that $N(t,l)$ is an inhomogeneous Poisson Process, and that $L = \infty$ for each individual.

Suppose that prevalence (here equivalent to cumulative incidence) $Z(t,l) = n+1$. We create a strictly increasing sequence $l = B_0 < B_1 < \cdots < B_n$ of $n+1$ infection times, which has probability density function

$$\underbrace{f_{\boldsymbol{B}}(\boldsymbol{b})}_{\text{Joint pdf}} = \underbrace{\frac{1}{P(Z(t,l)=n+1)}}_{\text{normalising constant}} \underbrace{\prod_{i=1}^{n}\left(\rho(b_i)\sum_{j=0}^{i-1}\nu(b_i-b_j)\right)}_{\text{infection rates at each } b_i} \underbrace{\exp\left[-\sum_{i=0}^{n}\int_0^{t-b_i}\rho(s+l)\nu(s)ds\right]}_{\text{probability of no other infections}}, \tag{6.15}$$

Then, the variance at time $t+s$ is given by

$$\underbrace{\text{var}(Z(t+s,l))}_{\text{variance}} = \underbrace{\int_{b=0}^{t} \sum_{i=0}^{n} V^*(t+s,b) f_{B_i}(b) db}_{\text{variance from subsequent cases}} \ldots$$
$$\ldots + \underbrace{\int_{b=0}^{t}\int_{c=0}^{t} \sum_{i=0}^{n}\sum_{j=0}^{n} M^*(t+s,b)M^*(t+s,c)(f_{B_i,B_j}(b,c) - f_{B_i}(b)f_{B_j}(c)) dbdc}_{\text{variance from unknown infection times}}, \tag{6.16}$$

where $M^*(t+s,b)$ and $V^*(t+s,b)$ are the mean and variance of the size of the infection tree (i.e. prevalence or cumulative incidence) at time $t+s$, caused by an individual infected at time b, ignoring all individuals they infected before time t. These quantities are calculated from M and V. Note also that f_{B_i} and f_{B_i,B_j} are the one-and-two-dimensional marginal distributions from $f_{\boldsymbol{B}}$.

6.4.5 Bayesian inference and for SARS epidemic in Hong Kong

The data for the SARS epidemic in Hong Kong consist of 114 daily measurements of incidence (positive integers), and an estimate of the generation time [444] obtained via the R package EpiEstim [426]. We ignore the infectious period g and set the infectiousness ν to the generation interval. The inferential task is then to estimate a time varying function ρ from these data using Equation 6.4. As we note in Equation 6.4 and in Appendix D, discretisation simplifies this task considerably. Our prior distributions are as follows

$$\phi \sim \text{Normal}^{+}(0,1)$$
$$\sigma \sim \text{Exponential}(100)$$
$$\epsilon \sim \text{Normal}(0,\sigma)$$
$$\rho(t) = \rho(t-1) + \epsilon_t$$

where ρ is modelled as a discrete random walk process. The renewal likelihood in Equation 6.4 is vectorised using the approach described in [60]. Fitting was performed in the probabilistic programming language Numpyro, using Hamiltonian Monte Carlo[445] with 1000 warm-up steps and 6000 sampling steps across two chains. The target acceptance probability was set at 0.99 with a tree depth of 15. Convergence was evaluated using the RHat statistic[446].

Forecasts were implemented through sampling using MCMC from Equation 6.4. In order to use Hamiltonian Markov Chain Monte Carlo, we relax the discrete constraint on incidence and allow it to be continuous with a diffuse prior. We ran a basic sensitivity analysis using a Random Walk

Metropolis with a discrete prior to ensure this relaxation was suitable. In a forecast setting, incidence up to a time point ($T = 60$) is known exactly and given as $y^{t \leq T}$ and we have access to an estimate for $\rho(t > T)$ in the future. In our case we fix $\rho(t > T) = \rho(T)$.

Our code is available at https://github.com/MLGlobalHealth/uncertainity_infectious_diseases.git.

6.4.6 Numerically calculating the probability mass function via the probability generating function

Following [447] and [448] (originally from [443]), the probability mass function p can be recovered through a pgf F's derivatives at $s = 0$. i.e. $\mathbb{P}(n) = \frac{1}{n!} \left(\frac{d}{ds}\right)^n F(s; t, \tau)|_{s=0}$ This is generally computationally intractable. A well-known result from complex analysis [443] holds that $f^{(n)}(a) = \frac{n!}{2\pi i} \oint \frac{f(z)}{(z-a)^{n+1}} dz$ and therefore $\mathbb{P}(n) = \frac{1}{2\pi i} \oint \frac{F(z;t,\tau)}{z^{n+1}} dz$ This integral can be very well approximated via trapezoidal sums as $\mathbb{P}(n) = \frac{1}{Mr^n} \sum_{m=0}^{M-1} F(re^{2\pi im/M}; t, \tau) e^{-2\pi inm/M}$ where $r = 1$[448]. The probability mass function for any time and n can be determined numerically. One needs $M \geq n$, which requires solving n renewal equations for the generating function and performing a fast Fourier transform. This is computationally fast, but may become slightly burdensome for epidemics with very large numbers of infected individuals (millions). A derivation of this approximation is provided in the Appendix D.

Chapter 7: Paper V: Optimality of maximal-effort vaccination

Matthew J Penn and Christl A Donnelly. "Optimality of maximal-effort vaccination". In: Bulletin of Mathematical Biology 85.8 (2023).

Status: This paper has been published in the *Bulletin of Mathematical Biology*.

Abstract: It is widely acknowledged that vaccinating at maximal effort in the face of an ongoing epidemic is the best strategy to minimise infections and deaths from the disease. Despite this, no one has proved that this is guaranteed to be true if the disease follows multi-group SIR (Susceptible–Infected–Recovered) dynamics. This paper provides a novel proof of this principle for the existing SIR framework, showing that the total number of deaths or infections from an epidemic is decreasing in vaccination effort. Furthermore, it presents a novel model for vaccination which assumes that vaccines assigned to a subgroup are distributed randomly to the unvaccinated population of that subgroup. However, as the novel model provides a strictly larger set of possible vaccination policies, the results presented in this paper hold for both models.

Author contributions: Attached at the end of the chapter.

7.1 Introduction

The COVID-19 pandemic has illustrated the importance of quickly implementing vaccination policies which target particular groups within a population [216]. The difference in final infections between targeted policies and uniform distribution to the entire population can be significant [209, 208] and so it is important that the models underlying these decisions provide realistic predictions of the outcomes of different policies.

One of the most commonly used models to project epidemics is the multi-group SIR (Susceptible-Infected-Recovered) model [449, 233, 450]. This model divides the population into different groups based on characteristics such as age or occupation. Each group is then further sub-divided into categories of susceptible, infected and recovered. Where vaccination does not give perfect immunity, further sub-categorisation based on vaccination status can also be used [451], as will be done in this paper.

While many other approaches have been developed either by adding compartments to the SIR framework [35] or using completely different models such as networks [452] or stochastic simulations [225], the multi-group SIR model remains popular because of its comparatively small number of parameters and its relatively simple construction and solution. In this paper, attention will thus be restricted to the multi-group SIR model, although it would be beneficial for future work to consider a wider range of disease models.

There are two general frameworks that are used to model optimal vaccination policies in a resource-limited setting. The first, used in papers such as [245] and [217], seeks to reduce the reproduction number, R_0 of the epidemic as much as possible by vaccinating before infections arrive in a population. It is simple to show that in this case, one should use all of the vaccinations available, and so this problem will not be considered further in this paper.

The second framework, used in papers such as [233] and [234] aims to minimise the total cost of an epidemic. This is the framework that will be discussed in this paper. The "cost" of an epidemic is, in general, defined to be the number of deaths or infections, with many papers also considering the cost of vaccination alongside the cost of other control measures, such as isolation, lockdown or treatment [453].

One important principle which underlies all of these vaccination policies is the assumption that giving people their first dose of vaccine as soon as possible reduces the number of infections. Of course, this only holds when the timescale considered is sufficiently short for effects such as waning immunity to be negligible, and a more complicated framework would be needed to model these effects. However,

the acceptance of at least short-term optimality of maximal vaccination effort has been highlighted in the COVID-19 pandemic response, as countries began their vaccination programmes as soon as vaccines became available [454].

To the best of the authors' knowledge, no one has provided a mathematical proof that in a general, multi-group SIR model with imperfect vaccination, it is always best to vaccinate people as early as possible. Of course, it is not difficult to create a conceptually sound justification - vaccinating more people means that fewer people will catch the disease which will reduce the total number of infections. However, the SIR model is an approximation of the process of a disease spreading, and so it is important that it obeys this principle for all physical parameter values and vaccination policies.

Some special cases of the theorem presented in this paper have been previously proved in the literature. In particular, a significant number of papers have considered the optimal vaccination policy for a homogeneous population, with [455] first proving that, in this case, it is optimal to vaccinate at maximal effort (if one ignores the cost of vaccination). This proof held for vaccination policies that were finite sums of point mass "impulse" vaccinations, and has been generalised by papers such as [234, 235, 456, 457] to a much wider class of vaccination policies, although the proof was still restricted to a single group and to perfect vaccination. Moreover, [234] notes that the case of imperfect vaccination (where vaccinated individuals can still get infected, although at a lower rate) remained a topic of open investigation, and so it can not easily be solved using the same methods presented in these papers. A slight extension is made in [458] where it is shown that maximal effort is optimal in the case of perfect vaccination of any number of disconnected groups, but the full problem is still far from understood.

The general method of proof in the literature relies on Pontryagin's Maximum Principle, which is difficult to apply to multi-group models due to the more complex structure of the equations. It is simple to characterise the solution in terms of the adjoint variables, as is done in [249] and [459] for a two-group model with imperfect vaccination, in [460] for a general n-group model with perfect vaccination and in [461] for a six-group model with imperfect vaccination. However, determining whether this solution corresponds to the maximal effort solution in the case of zero vaccination cost requires the analysis of the adjoint ODE system, which is often just as complicated as the original disease model. In particular, the fact that vaccinated people need to be no more infectious, no more susceptible and be infected for no more time than unvaccinated people means that any analysis of the adjoint system would be complicated, as the properties of all the constituent parameters would need to be used.

Thus, in this paper, a novel approach is developed. Rather than attempting to use the general optimal control theory methodology, the specific structure of the SIR equation system is exploited.

Using this, an inequality is derived which shows that if a given vaccination policy, $\tilde{\boldsymbol{U}}$ vaccinates each individual at least as early as another vaccination policy, $\boldsymbol{U}$, then the latter policy will lead to at least as many deaths (or equivalently, infections) as the former. As well as providing a constraint on the optimal solution, this theorem also highlights important structural properties of the model, as it shows that the number of deaths is everywhere non-increasing in the vaccination rates, rather than this just holding near the optimal solutions.

Also introduced in this paper is a more general model of vaccination than the one normally used in the literature. The one that is typically used (in almost all papers cited in this work such as [234, 235, 236]) models decreasing vaccination uptake by assuming that the total rate of people being vaccinated is the product of a vaccination rate and the proportion of susceptible people in the population. The model introduced in this paper allows for more flexibility in modelling the demand. However, the standard vaccination model is a special case of the general model introduced here, and so the theoretical results proved in this paper can be used by those following the standard model.

Alongside proving that the final infected, recovered and dead populations are non-increasing with increased and earlier vaccination effort, some cautionary contradictions to perhaps intuitive conjectures are also provided which show the importance of mathematical proof instead of simply intuition. In particular, it is shown that increased vaccination (under this model) can lead to, at a fixed finite time of the simulation, higher infection rates or a higher death count, despite the longer-term better performance of this policy - as the infection curve is "flattened" [462]. Indeed, it is results similar to these which make the proof of the optimality of maximal effort difficult, as it means that one must be very careful when constructing the inequalities that do hold for all models.

7.2 Modelling

7.2.1 Disease transmission and vaccination model

Suppose that the population is divided into n subgroups, such that population of people in group i is N_i and define

$$N := \sum_{i=1}^{n} N_i. \tag{7.1}$$

Define the compartments of people as follows, for $i = 1, ..., n$:

$$S_i := \text{Number of people that are in group } i, \text{ are susceptible, and are unvaccinated,} \tag{7.2}$$

$$I_i := \text{Number of people that are in group } i, \text{ are currently infected, and}$$
$$\text{were infected while unvaccinated ,} \tag{7.3}$$

$$R_i := \text{Number of people that are in group } i, \text{ are recovered (or dead), and} \tag{7.4}$$
$$\text{were infected while unvaccinated,}$$

$$S_i^V := \text{Number of people that are in group } i, \text{ are susceptible and are vaccinated,} \tag{7.5}$$

$$I_i^V := \text{Number of people that are in group } i, \text{ are infected} \tag{7.6}$$
$$\text{and were infected after being vaccinated,}$$

$$R_i^V := \text{Number of people that are in group } i, \text{ are recovered (or dead)} \tag{7.7}$$
$$\text{and were infected after being vaccinated.} \tag{7.8}$$

This paper introduces a more general and flexible framework for vaccination, which is motivated as follows. It is assumed that there is a record of people who have received a vaccination and that protection from vaccination does not decay over time, so that no one is vaccinated more than once. Thus, if a total number, $U_i(t)dt$, of people in group i are given vaccines in a small time interval $(t, t + dt)$, and these vaccines are distributed randomly to the unvaccinated population in group i, the total population of susceptibles given vaccines in group i is

$$SU_i(t)dt \times \mathrm{P}\Big(\text{A person in group } i \text{ is in } S_i \mid \text{A person is in group } i \text{ is unvaccinated}\Big) \tag{7.9}$$

which is equal to

$$\frac{U_i(t)dtS_i(t)}{N_i - \int_0^t U_i(s)ds}, \tag{7.10}$$

as $\int_0^t U_i(s)ds$ is the total population that are in group i and have been vaccinated before time t. For the remainder of this section, this vaccination model will be referred to as the "general" model

This results in the following model, based on SIR principles

$$\frac{dS_i}{dt} = -\sum_{j=1}^{n}(\beta_{ij}^1 I_j + \beta_{ij}^2 I_j^V)S_i - \frac{U_i(t)S_i}{N_i - W_i(t)}, \tag{7.11}$$

$$\frac{dI_i}{dt} = \sum_{j=1}^{n}(\beta_{ij}^1 I_j + \beta_{ij}^2 I_j^V)S_i - \mu_i^1 I_i, \tag{7.12}$$

$$\frac{dR_i}{dt} = \mu_i^1 I_i, \tag{7.13}$$

$$\frac{dS_i^V}{dt} = -\sum_{j=1}^{n}(\beta_{ij}^3 I_j + \beta_{ij}^4 I_j^V)S_i^V + \frac{U_i(t)S_i}{N_i - W_i(t)}, \tag{7.14}$$

$$\frac{dI_i^V}{dt} = \sum_{j=1}^{n}(\beta_{ij}^3 I_j + \beta_{ij}^4 I_j^V)S_i^V - \mu_i^2 I_i^V, \tag{7.15}$$

$$\frac{dR_i^V}{dt} = \mu_i^2 I_i^V, \tag{7.16}$$

where

$$W_i(t) := \int_0^t U_i(s)ds. \tag{7.17}$$

Here, β_{ij}^1 represents transmission from the unvaccinated members of group j to the unvaccinated members of group i, β_{ij}^2 represents transmission from vaccinated members to unvaccinated members, β_{ij}^3 represents transmission from unvaccinated members to vaccinated members and β_{ij}^4 represents transmission from vaccinated members to vaccinated members. Additionally, μ_i^1 represents the infectious period of unvaccinated infected members in group i while μ_i^2 represents the infectious period of vaccinated members. Note that the superscript denotes different parameter values, so that β_{ij}^2 is not necessarily the square of β_{ij}^1.

To ensure that vaccination is "locally effective" (that is, a vaccinated individual is no more likely to transmit or be infected by the disease, and is infectious for no longer than an unvaccinated individual in the same subgroup), and that the parameters are epidemiologically feasible, the following constraints are imposed:

$$\beta_{ij}^1 \geq \beta_{ij}^2, \beta_{ij}^3 \geq \beta_{ij}^4 \geq 0 \quad \text{and} \quad \mu_i^2 \geq \mu_i^1 > 0 \tag{7.18}$$

Note that there is no constraint on the ordering of β_{ij}^2 and β_{ij}^3. It is assumed for convenience that all variables except the S_i and I_i are initially zero. Finally, we assume that all initial conditions are non-negative.

Ultimately, the objective of the vaccination programme will be to minimise a weighted sum of the

total infections in each group - that is

$$\sum_{i=1}^{n} p_i(R_i(\infty) + \kappa_i R_i^V(\infty)). \tag{7.19}$$

Here p_i is the weighting of a member of group i who is infected before being vaccinated, while $p_i\kappa_i$ is the weighting of a member of group i who is infected after being vaccinated. These parameters could be chosen to capture one of a range of objectives, such as minimizing deaths, minimizing hospitalisations, or minimizing total infections. Again assuming "local effectiveness" of the vaccination, it is imposed that $\kappa_i \leq 1$, as vaccination should not increase the severity of the infection.

The equations (7.11) - (7.16) sum to zero on the right-hand side, and so for each i,

$$S_i(t) + I_i(t) + R_i(t) + S_i^V(t) + I_i^V(t) + R_i^V(t) = N_i \quad \forall t \geq 0. \tag{7.20}$$

It will be assumed that the populations and parameters have been scaled such that $N = 1$, Finally, it is assumed that

$$W_i(t) \leq N_i \quad \forall t \geq 0 \quad \text{and} \quad W_i(t) = N_i \Rightarrow \frac{U_i(t)S_i}{N_i - W_i(t)} = 0. \tag{7.21}$$

to ensure feasibility of the vaccination policies.

7.2.2 Comparison to the standard vaccination model

A more common model of vaccination in the literature is the "standard" vaccination model ([234], [235] and [236]), where equation (7.11) becomes

$$\frac{dS_i}{dt} = -\sum_{j=1}^{n}(\beta_{ij}^1 I_j + \beta_{ij}^2 I_j^V)S_i - U_i^*(t)S_i, \tag{7.22}$$

Here, $U_i^*(t)$ is the vaccination rate in this model. In general, $U_i^*(t)$ is constrained such that $U_i^*(t) \leq \mathcal{U}_i(t)$ for some function $\mathcal{U}_i(t)$

The $U_i^*(t)S_i$ term seeks to capture the fact that vaccination uptake will decrease even if the vaccination "effort" (or, equivalently, the doses available) remains constant. However, the rate at which uptake decreases is fixed by the model. For example, if the vaccination effort $U_i^*(t)$ was equal to a constant $\mathcal{U}_i$ and was much quicker than the rate of infection, then the leading order equation is

$$\frac{dS_i}{dt} = -\mathcal{U}_i S_i \Rightarrow S_i(t) = S_i(0)e^{-\mathcal{U}_i t} \tag{7.23}$$

and hence

$$\frac{dS_i}{dt} = -\mathcal{U}_i S_i(0) e^{-\mathcal{U}_i t} \tag{7.24}$$

which means that vaccination uptake decreases exponentially. For some pandemics, such as COVID-19, where demand remained high until a large proportion of the population had been vaccinated, as shown in [463], such a model may be inappropriate.

The general vaccination model provides substantially more flexibility. For example, it is possible for a group to be completely vaccinated in the general case, whereas this is impossible in the standard case (while one may never be able to fully vaccinate a human population, it would be possible, for example, in a group of animals on a farm). Moreover, by bounding the vaccination rate $U_i(t)$ above by some function of vaccination demand $K_i(W(t), t)$, decreasing vaccination uptake can still be modelled.

7.2.3 Recovery of the standard model

The standard model is a special case of the general model, meaning that the results of this paper are applicable to both frameworks. To show this, one can solve the equation

$$\frac{U_i(t)}{N_i - W_i(t)} = U_i^*(t) \Rightarrow \frac{d}{dt}\Big(\log(N_i - W_i(t)) + W_i^*(t) \Big) = 0, \tag{7.25}$$

where

$$W_i^*(t) := \int_0^t U_i^*(s) ds. \tag{7.26}$$

Thus, by integrating (7.25), and noting that $W_i^*(0) = W_i(0) = 0$

$$\log(N_i - W_i(t)) + W_i^*(t) = \log(N_i) \tag{7.27}$$

and so

$$W_i(t) = N_i(1 - e^{-W_i^*(t)}). \tag{7.28}$$

The constraint $U_i^*(t) \leq \mathcal{U}_i$ is equivalent to $U_i(t) \leq (N_i - W_i(t))\mathcal{U}_i$ and so this can also be represented in the general model. Thus, given any standard vaccination policy $\boldsymbol{U}^*$, it can be replaced by a general policy $\boldsymbol{U}$ (although the converse does not hold as $W_i(t) = N_i$ requires $W_i^*(t) = \infty$).

Moreover, note that $W_i^*(t)$ is increasing in $W_i(t)$. Thus, if a pair of general policies $\boldsymbol{U}$ and $\tilde{\boldsymbol{U}}$ satisfy $W_i(t) \leq \tilde{W}_i(t)$ then this inequality is preserved by the corresponding standard policies as $W_i^*(t) \leq \tilde{W}_i^*(t)$. This property means that the theorems proved in this paper will hold for both models (as they will be proved using the general model).

7.3 Optimisation Problem

Now that the model has been formulated, it is possible to set up the optimisation problem that will be considered in the remainder of this paper.

7.3.1 Constraints on $U_i(t)$

In order to assist the proof of the theorems, it is necessary to make some (unrestrictive) assumptions on the vaccination rates, $U_i(t)$.

Firstly, there are the physical constraints that for each $i \in \{1, ..., n\}$

$$U_i(t) \geq 0 \quad \text{and} \quad \int_0^t U_i(s)ds \leq N_i \quad \forall t \geq 0. \tag{7.29}$$

It is also necessary that $U_i(t)$ is within the class of functions such that solutions to the model equations exist and are unique. Discussion of the exact conditions necessary for this to hold is outside the scope of this paper. However, from the Picard-Lindelöf Theorem [464], a sufficient condition for this is that $U_i(t)$ is a piecewise Lipschitz continuous function. While this is not a necessary condition, this illustrates that this assumption will hold for a large class of functions. However, it will be helpful throughout the course of the proof to explicitly assume two conditions on $U_i(t)$ - namely, that it is bounded and that it is Lebesgue integrable on $\Re$ for each i.

For the remainder of this paper, define the set of feasible vaccination policies, C, to be the set of functions $\boldsymbol{U}$ satisfying (7.29) such that unique solutions to the model equations exist with these functions as the vaccination policy and such that each $U_i(t)$ is bounded and Lebesgue integrable on $\Re$.

7.3.2 Optimisation problem

The aim is to choose the vaccination policy $\boldsymbol{U} \in C$ such that the total number of deaths (or any linear function of the infections in each subgroup) is minimised while meeting additional constraints on vaccine supply and vaccination rate. It is assumed that the maximal rate of vaccination at time t is $A(t)$ and that there is a total (non-decreasing) supply of $B(t)$ vaccinations that has arrived by time t. Thus, for each i, $U_i(t)$ is constrained to satisfy

$$\sum_{i=1}^{n} U_i(t) \leq A(t) \quad \text{and} \quad \sum_{i=1}^{n} W_i(t) \leq B(t). \tag{7.30}$$

As previously discussed, it is assumed that each infection of unvaccinated people in group i is weighted by some p_i and that the infection is no more serious for those that have been vaccinated, so that the weighting of an infection of a vaccinated person in group i is $p_i\kappa_i$, where $\kappa_i \leq 1$. Thus, the objective function is

$$H(\boldsymbol{U}) := \sum_{i=1}^{n} p_i \Big(R_i(\infty) + \kappa_i R_i^V(\infty) \Big) \tag{7.31}$$

where, for example

$$R_i(\infty) = \lim_{t\to\infty} (R_i(t)). \tag{7.32}$$

Note these limits exist as R_i is non-decreasing and bounded by Lemma E.17. Hence, the optimisation problem is

$$\min \Bigg\{ \sum_{i=1}^{n} p_i \Big(R_i(\infty) + \kappa_i R_i^V(\infty) \Big) : \sum_{i=1}^{n} U_i(t) \leq A(t), \quad \sum_{i=1}^{n} W_i(t) \leq B(t) \quad \forall i, t... \tag{7.33}$$

$$\text{and} \quad \boldsymbol{U} \in C \Bigg\}. \tag{7.34}$$

7.4 Main results

The main results of this paper are as follows. Firstly, it is shown that the objective function is non-increasing in vaccination effort.

Theorem 7.1 *Suppose that $\boldsymbol{U}, \tilde{\boldsymbol{U}} \in C$. Suppose further that for each $i \in \{1, ..., n\}$ and $t \geq 0$*

$$\int_0^t U_i(s)ds \leq \int_0^t \tilde{U}_i(s)ds \tag{7.35}$$

Then

$$H(\boldsymbol{U}) \geq H(\tilde{\boldsymbol{U}}). \tag{7.36}$$

Then, it is shown that if an optimal solution exists, there is an optimal maximal effort solution.

Theorem 7.2 *Suppose that B is differentiable, and that there is an optimal solution $\boldsymbol{U}$ to (7.34). Then, define the function*

$$\chi(t) := \begin{cases} A(t) & \text{if} \quad \int_0^t \chi(s)ds < B(t) \\ \min(A(t), B'(t)) & \text{if} \quad \int_0^t \chi(s)ds \geq B(t) \end{cases} \tag{7.37}$$

and suppose that $\chi(t)$ exists and is bounded. Then, there exists an optimal solution $\tilde{\boldsymbol{U}}$ to the problem

(7.34) such that

$$\sum_{i=1}^{n} \tilde{W}_i(t) = \min\left(\int_0^t \chi(s)ds, 1 \right). \tag{7.38}$$

Moreover, if $\chi(t)$ is continuous almost everywhere, there exists an optimal solution $\tilde{\boldsymbol{U}}$ such that

$$\sum_{i=1}^{n} \tilde{U}_i(t) = \begin{cases} \chi(t) & \text{if } \int_0^t \chi(s)ds < 1 \\ 0 & \text{otherwise} \end{cases} \tag{7.39}$$

It is perhaps concerning to the reader that the existence of χ is left as an assumption in this theorem. However, while the exact conditions on its existence are beyond the scope of this paper, it certainly exists for a wide class of functions $A(t)$ and $B(t)$, as proved in Lemma E.14.

Finally, it is shown that this principle still holds if the cost of vaccination is considered.

Theorem 7.3 *Under the assumptions of Theorem 7.2, consider a modified objective function $\mathcal{H}$ given by*

$$\mathcal{H}(\boldsymbol{U}) = H(\boldsymbol{U}) + F(\boldsymbol{W}(\infty)) \tag{7.40}$$

for any non-decreasing (in each argument) function F. Then, with χ defined to be the maximal vaccination effort as in Theorem 7.2, there exists an optimal solution $\tilde{\boldsymbol{U}}$ such that, for some $\tau \geq 0$

$$\sum_{i=1}^{n} \tilde{W}_i(t) = \begin{cases} \int_0^t \chi(s)ds & \text{if } t \leq \tau \\ W_i(\infty) & \text{otherwise} \end{cases}. \tag{7.41}$$

Moreover, if χ is continuous almost everywhere, then there is an optimal solution $\tilde{\boldsymbol{U}}$ such that

$$\sum_{i=1}^{n} U_i(t) = \begin{cases} \chi(t) & \text{if } t \leq \tau \\ 0 & \text{otherwise} \end{cases}. \tag{7.42}$$

7.5 Sketch proof

The full proofs of Theorems 7.1, 7.2 and 7.3 can be found in Appendix E. However, this section provides a high-level sketch of the main arguments.

7.5.1 Bounds on the inter-group infectious forces

Define

$$K_{ij}(t) = \frac{\beta_{ij}^1}{\mu_j^1} R_j(t) + \frac{\beta_{ij}^2}{\mu_j^2} R_j^V(t) \tag{7.43}$$

and

$$L_{ij}(t) := \frac{\beta_{ij}^3}{\mu_i^1} R_j(t) + \frac{\beta_{ij}^4}{\mu_i^2} R_j^V(t). \tag{7.44}$$

$K_{ij}(t)$ can be interpreted as the total infectious force up to time t from the members of group j on the unvaccinated members of group i as

$$K_{ij}(t) = \int_0^t (\beta_{ij}^1 I_j(\tilde{t}) + \beta_{ij}^2 I_j^V(\tilde{t}))d\tilde{t}. \tag{7.45}$$

Similarly, $L_{ij}(t)$ can be interpreted as the total infectious force up to time t from the members of group j on the vaccinated members of group i.

The first part of the proof shows that increasing the vaccination effort will decrease these infectious forces. To facilitate the proof, some extra assumptions are made on the parameters (which will be removed in subsequent propositions).

Proposition 7.1 *Suppose that $U_i(t)$ and $\tilde{U}_i(t)$ are right-continuous step functions. Moreover, suppose that*

$$\beta_{ij}^1 > \beta_{ij}^3 > 0 \quad \forall i, j \in \{1, ..., n\}, \tag{7.46}$$

$$S_i(0)I_i(0) > 0 \quad \forall i \in \{1, ...n\}. \tag{7.47}$$

and that

$$W_i(t) < N_i \quad \forall t \geq 0 \quad \textit{and} \quad \forall i \in \{1, ..., n\} \tag{7.48}$$

Then,

$$K_{ij}(t) \geq \tilde{K}_{ij}(t) \quad \textit{and} \quad L_{ij}(t) \geq \tilde{L}_{ij}(t) \quad \forall t \geq 0. \tag{7.49}$$

This proposition is proved by contradiction in two parts. Firstly, a time T is introduced, which is the infimum of the times where at least one of $K_{ij}(t) < \tilde{K}_{ij}(t)$ or $L_{ij}(t) < \tilde{L}_{ij}(t)$ for some i and j. As the infectious forces do not satisfy this condition in $[0, T]$, one can show that, necessarily, they must all have been equal in $[0, T]$, which means that one must have $W_i(t) = \tilde{W}_i(t)$ for all $t \in [0, T]$.

From here, the proof can proceed by a short-time linearisation, considering the small interval $[T, T + \delta]$. The condition on U_i and $\tilde{U}_i$ being step functions allows for them to be considered constant in this interval. It can then be shown that (7.49) must hold in $[T, T + \delta]$, which contradicts the definition of T and completes the proof.

7.5.2 A proof for a restricted parameter and policy set

Proposition 1 can be extended to prove the result of Theorem 7.1 under the more restrictive set of conditions it introduced.

Proposition 7.2 *Under the conditions of Proposition 1, for any* $t \geq 0$ *and* $i \in \{1, ..., n\}$

$$I_i(t) + I_i^V(t) + R_i(t) + R_i^V(t) \geq \tilde{I}_i(t) + \tilde{I}_i^V(t) + \tilde{R}_i(t) + \tilde{R}_i^V(t) \tag{7.50}$$

and

$$R_i(t) \geq \tilde{R}_i(t). \tag{7.51}$$

Moreover, for any $\lambda \in [0, 1]$

$$R_i(\infty) + \lambda R_i^V(\infty) \geq \tilde{R}_i(\infty) + \lambda \tilde{R}_i^V(\infty) \tag{7.52}$$

and hence, the objective function is lower for $\tilde{U}$*, provided the conditions of Proposition 1 are met.*

This comes from finding $S_i + S_i^V$ in terms of K_{ij}, L_{ij} and W, and showing that $S_i + S_i^V \leq \tilde{S}_i + \tilde{S}_i^V$ - that is, that more people were infected in the U_i case. Taking limits, and using a similar approach to consider the number of unvaccinated infections then shows the required result.

7.5.3 Generalisation

This result can be generalised to the original set of parameters and vaccination policies by using the continuous dependence of the number of infections on the parameters and the vaccination policy.

From here, it is simple to weaken the inequalities on the parameters introduced in Proposition 1. The treatment of the vaccination policies requires more care, as it is not necessarily true that a Lebesgue integrable $\boldsymbol{U}$ can be approximated by step functions. However, its integral, $\boldsymbol{W}$, can be approximated by the integral of step functions, and this allows the result of Proposition 2 to be generalised to Theorem 7.1.

7.5.4 Theorem 7.2

Theorem 7.2 is proved as follows. Firstly, one can show that, for any vaccination policy $\boldsymbol{U}$ and $t \geq 0$,

$$\min\left(\int_0^t \chi(s)ds, 1\right) \geq \int_0^t \sum_{i=1}^n U_i(s)ds, \tag{7.53}$$

using the definition of χ in terms of the constraints on $\boldsymbol{U}$. This means that the total rate of vaccination given by $\tilde{\boldsymbol{U}}$ is at least as high as that given by $\boldsymbol{U}$.

One can then show that $\chi(t) \leq A(t)$

$$\int_0^t \chi(s)ds \leq B(t) \tag{7.54}$$

which means that $\tilde{\boldsymbol{U}}$ satisfies the vaccination constraints.

From here, one can transform any optimal vaccination policy $\boldsymbol{U}$ into suitable $\tilde{\boldsymbol{U}}$. Initially, the quantities $\tilde{W}_i(t)$ are constructed. The details of this are left to the appendix but the general principle is that the policy $\boldsymbol{U}$ is compressed in time so that the total number of vaccinations given out matches $\min\left(\int_0^t \chi(s)ds, 1\right)$. It may also be necessary to add additional vaccinations if the overall total differs - these can be assigned in proportion to the number of unvaccinated people in each group.

This construction ensures that the feasibility constraints $\tilde{W}_i \leq N_i$ are satisfied. Moreover, one can show that $\tilde{W}_i$ is Lipschitz continuous, which allows for the construction of a derivative $\tilde{U}_i$ which integrates to $\tilde{W}_i$. Finally, one can show that $\tilde{W}_i(t) \geq W_i(t)$, meaning that, by Theorem 7.1, $\tilde{\boldsymbol{U}}$ must also be an optimal vaccination policy.

7.5.5 Theorem 7.3

The proof of Theorem 7.3 then follows from a similar construction to Theorem 7.2 - the only difference is that no additional vaccinations are assigned by $\tilde{\boldsymbol{U}}$ compared to $\boldsymbol{U}$.

7.6 Limitations of Theorem 7.1

It is helpful to consider the limitations of Theorem 7.1, as it does not prove that every conceivable cost function is non-increasing in vaccination effort. This will be illustrated through some examples based on theoretical COVID-19 outbreaks in the United Kingdom.

Using the work of [214], one can split the UK into 16 age-groups (comprising five-year intervals from 0 to 75 and a group for those aged 75+) which mix heterogeneously. The contact matrices estimated in [214] allow for the construction of a matrix $\boldsymbol{\beta}^*$, which will be proportional to each of the matrices $\boldsymbol{\beta}^\alpha$ in the model.

As illustrated in [435], estimation of the basic reproduction number R_0 for COVID-19 is complicated, and a wide range of estimates have been produced. For the examples in this paper, a reproduction number of 4 will be used, meaning that β^1 will be scaled so that the largest eigenvalue

of the matrix given by

$$M_{ij} = \frac{\beta^1_{ij} N_i}{\mu^1_i} \tag{7.55}$$

is equal to 4. Note that the population of each group $\boldsymbol{N}$ - normalised to have total sum 1 - is taken from [465]. Moreover, based on the estimates in [449], the value of μ^1_i and, in the first example, μ^2_i will be set equal to $\frac{1}{14}$.

To model the effectiveness of vaccination, the estimates of [466] will be used so that $\beta^2 = 0.77\beta^1$, (modelling the reduction in infectiousness), $\beta^3 = 0.3\beta^1$ (modelling the reduction in susceptibility) and $\beta^4 = 0.77 \times 0.3 \times \beta^1$ (assuming these effects are independent). Finally, the initial conditions used are $S_i(0) = (1 - 10^{-4})N_i$ and $I_i(0) = (10^{-4})N_i$ for each i, modelling a case where 0.01% of the population is initially infected. It should be emphasised however, that this model has purely been made for illustrative purposes and substantially more detailed fitting analysis would be required to use it for forecasting COVID-19 in the UK.

In both the subsequent examples, it will be assumed that 0.5% of the population is vaccinated homogeneously each day in the vaccination case. This will be compared to a case with no vaccination.

7.6.1 Infections are not decreasing for all time

While the overall number of infections will decrease as vaccination effort increases, the infections at a particular point in time will not. Figure 7.1 shows that the effect of vaccination is both to reduce, but also delay the peak of the infections. This is an important consideration when deciding vaccination policy, as increasing infections at a time in the year when hospitals are under more pressure could have negative consequences, and so it is important not to simply assume that vaccination will reduce all infections at all times.

7.6.2 Deaths are not decreasing for all time

Perhaps most surprisingly, the total deaths in the epidemic may at some finite times (although not at $t = \infty$) be higher when vaccination occurs, at least under the assumptions of the SIR model. This is a rarer phenomenon, but is possible if vaccination increases the recovery rate as well as decreasing infectiousness.

For illustrative purposes, suppose that vaccination doubles the recovery rate (so that $\mu^2_i = \frac{1}{7}$) and has no effect on mortality rates. Then, using [221] to get age-dependent mortality rates for COVID-19, Figure 7.2 shows that initially, the number of deaths is higher in the case of vaccination. This occurs because the higher value of μ^2 means that vaccinated people move more quickly to the R^V

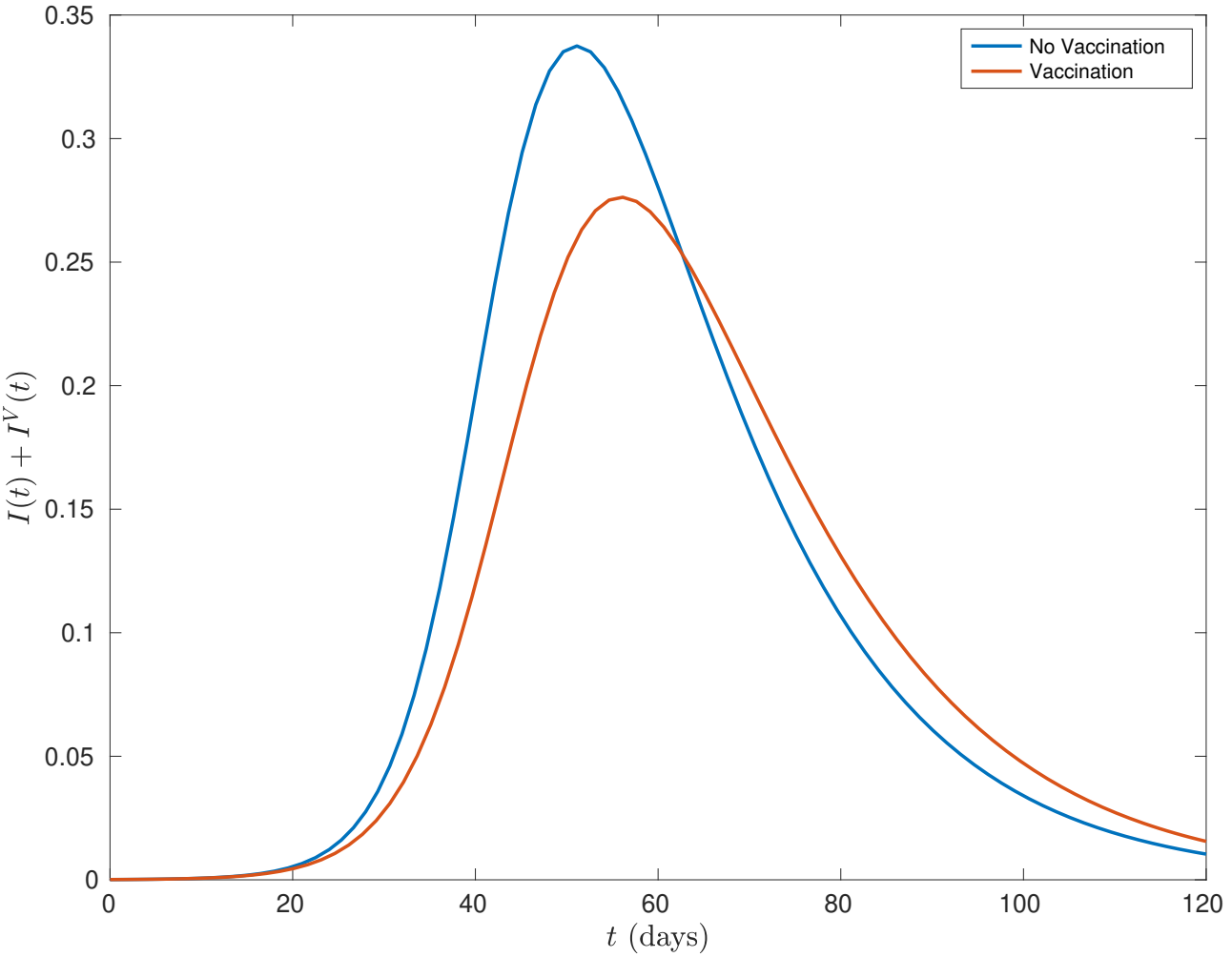


Figure 7.1: A comparison of the total infections over time for a simulated COVID-19 epidemic in the UK, depending on whether a uniform vaccination strategy of constant rate is used.

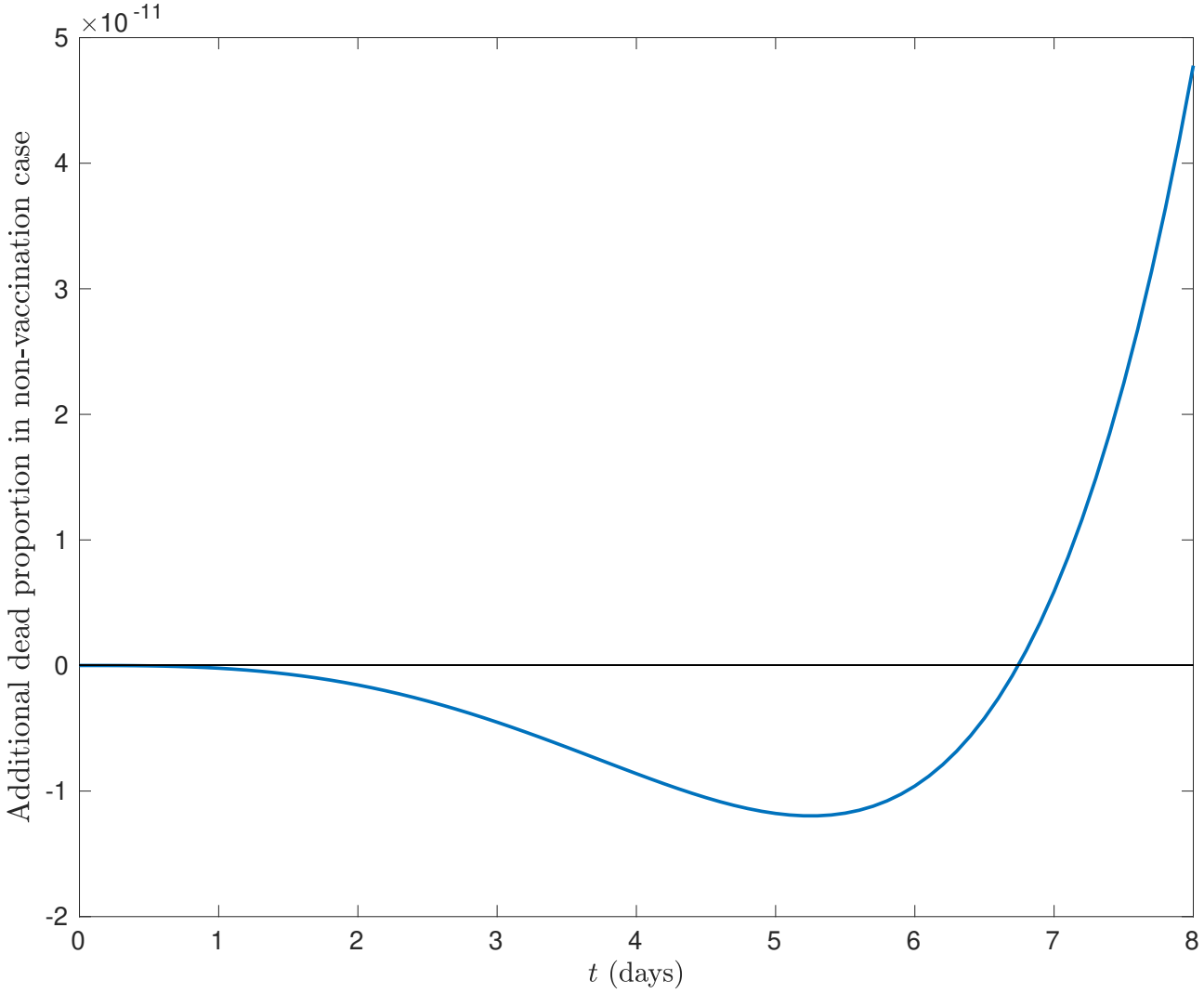


Figure 7.2: The difference between proportion of the population that has died by each time t in the case of vaccination and non-vaccination. Positive values indicate that the deaths are higher in the non-vaccination case.

compartment than their unvaccinated counterparts and so, while they will infect fewer people, when the number of infections is comparable in the early epidemic, this means that more people will die. Indeed, this property can still hold if vaccination reduced mortality rates (although this reduces the already small difference between the two further - in this example, one needs $\kappa_i \gtrsim 0.9999$ for deaths to ever be lower in the non-vaccinated case).

Of course, this is not a realistic reflection of the course of an epidemic - the reason for μ^2 being higher is that vaccinated people are likely to get *less* ill rather than dying more quickly - but it illustrates a potential limitation of the SIR framework. One possible way to avoid this problem would be to split the recovered compartment up into the truly recovered and dead subsections. Then, vaccination could increase the speed at which infected members of the population moved to the recovered compartment, but not the speed at which they moved to the dead compartment. This would remove the possibility of seeing the counter-intuitive behaviour of Figure 7.2.

7.7 Discussion

It is comforting that the multi-group SIR model does indeed satisfy the condition that the final numbers of infections and deaths are non-increasing in vaccination effort. This shows the importance of ensuring that vaccinations are available as early as possible in a disease outbreak. To achieve this, it is important that good plans for vaccine roll-out and supply chains are available in advance of them being needed to ensure that maximum benefit from the vaccination programme is obtained.

For $n > 1$, there are, of course, many possible maximal-effort vaccination policies. The results of this paper, in effect, reduce the dimension of the space of possible vaccination policies from n to $n-1$, as one can assume that an optimal policy satisfies the condition (7.38) in Theorem 7.2. However, choosing the correct groups to prioritise is still of crucial importance and can have a substantial impact on the effectiveness of the vaccination campaign [216]. Applying similarly rigorous techniques to finding the optimal vaccination policy is beyond the scope of this paper, although we extended the results of this paper to apply asymptotic techniques to understand the behaviour of the optimal solution under certain special cases in [51].

However, there are limitations to these results. Indeed, while the final numbers of infections and deaths are guaranteed to decrease, this is not necessarily true at a given finite time. In particular, vaccination can move the peak of the epidemic, and so it is important to consider the consequences of this, particularly if only a small number of lives are saved by vaccination.

Moreover, while this has not been discussed in this paper, it is also important to emphasise that these results only apply if vaccine efficacy does not decay over time. Indeed, if vaccination efficacy does decay significantly, then vaccinating the most vulnerable groups in a population very early may be worse than vaccinating them later, unless booster jabs are available. If the main epidemic occurs long after the vulnerable have been vaccinated, their immunity may have worn off significantly by the time that the majority of disease exposure occurs. Thus, in this case a more detailed analysis would be needed to determine the optimal vaccination rate.

The authors believe that future models for optimal vaccination should consider using the more general vaccination model introduced in this paper. This allows for greater flexibility in modelling the effect of decreasing demand. Of course, this modified model is slightly more complicated, and care needs to be taken to avoid numerical instabilities arising from the removable singularity in the $\frac{U_i S_i}{N_i - W_i}$ term when $W_i \to N_i$. However, it has been shown that many of the standard properties of SIR models, and indeed the results of this paper, still hold for this model, and so these extra technical difficulties appear to be a small price to pay for the significantly increased accuracy and potentially

large difference between the optimal solutions for the two models.

The results of this paper could be extended to cover a wider range of disease models that are currently being used in the literature. In particular, the next step could be to prove the results for SEIR (Susceptible-Exposed-Infected-Recovered) models, and indeed models with multiple exposed compartments for each subgroup. This would help to build a general mathematical theory of maximal-effort vaccination that would provide evidence for the reliability of contemporary epidemiological modelling.

7.8 Conclusion

The results of this paper are summarised below:

- Vaccinating at maximal effort is optimal for a multi-group SIR model with non-decaying vaccination efficacy.
- The general vaccination model introduced in this paper provides greater flexibility in modelling the effect of decreasing vaccination uptake.
- While vaccinating at maximal effort gives optimality, there can be finite times at which, according to the SIR model, infections or deaths are higher if vaccination has occurred.

Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	**Optimality of maximal-effort vaccination**
Publication Status	**Published in the Bulletin of Mathematical Biology**
Publication Details	**Matthew J Penn and Christl A Donnelly. "Optimality of maximal-effort vaccination". In: Bulletin of Mathematical Biology 85.8 (2023).**

Student Confirmation

Student Name:	**Matthew Penn**		
Contribution to the Paper	**Conceived of and developed the mathematical results in this paper.** **Carried out the numerical experiments.** **Wrote the first draft of the paper.**		
Signature		Date	**20/03/24**

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: **Professor Christl Donnelly**		
Supervisor comments **As described above, this is Matt's initiative and follow-through. I provided supervision and suggestions for editing.**		
Signature 	Date	20 March 2024

This completed form should be included in the thesis, at the end of the relevant chapter.

Chapter 8: Paper VI: Asymptotic analysis of optimal vaccination policies

Matthew J Penn and Christl A Donnelly. "Asymptotic analysis of optimal vaccination policies". In: Bulletin of Mathematical Biology 85.3 (2023)

Status: This paper has been published in the *Bulletin of Mathematical Biology*.

Abstract: Targeted vaccination policies can have a significant impact on the number of infections and deaths in an epidemic. However, optimising such policies is complicated, and the resultant solution may be difficult to explain to policy-makers and to the public. The key novelty of this paper is a derivation of the leading-order optimal vaccination policy under multi-group susceptible–infected–recovered dynamics in two different cases. Firstly, it considers the case of a small vulnerable subgroup in a population and shows that (in the asymptotic limit) it is optimal to vaccinate this group first, regardless of the properties of the other groups. Then, it considers the case of a small vaccine supply and transforms the optimal vaccination problem into a simple knapsack problem by linearising the final size equations. Both of these cases are then explored further through numerical examples, which show that these solutions are also directly useful for realistic parameter values. Moreover, the findings of this paper give some general principles for optimal vaccination policies which will help policy-makers and the public to understand the reasoning behind optimal vaccination programmes in more generic cases.

Author contributions: See end of chapter

Notes:

1. The version presented in this thesis contains a correction to the proof of Theorem 2 - this has been submitted to the journal and we are currently awaiting their response.
2. In the published manuscript, the appendix section "Supplementary Lemmas" restates a number of results from Paper V. Since Paper V is also part of this thesis, we have omitted these restated results and instead referenced the relevant lemmas from Paper V directly.

8.1 Introduction

The trajectory of an epidemic can be dramatically changed by the implementation of a vaccination program, as has been shown in the case of COVID-19 [467]. These vaccination programs are most effective when they target specific groups in a population [216], although the optimal targeting strategy is dependent on the properties of the disease and vaccine [35]. Thus, it is important to have robust methods to determine the optimal strategy whenever a new epidemic emerges.

In recent years, the epidemiological literature has grown rapidly, and a wide range of models have been developed and analysed. These include branching-process models [60]; network-based models [468] and machine-learning-based models [469], among many others [470].

However, despite these innovations, compartmental models, where the population is split into a number of subgroups and disease transmission is modelled by a system of differential equations [471], remain a popular choice for epidemiologists, and have been widely used for modelling the COVID-19 pandemic [472]. As discussed in [472], a number of different compartment structures have been used, while many authors have also sought to model the effect of government interventions and quarantining procedures [473, 474, 475].

One such compartmental model that is widely used [449, 233, 450] is the multi-group SIR (Susceptible-Infected-Recovered) model. This is an extension of the classical SIR model [411] and has been used to model a range of diseases such as measles [476], influenza [477] and COVID-19 [478]. It provides a general framework with which to assess the effectiveness of different vaccination policies, while also remaining mathematically tractable, allowing theorems about its behaviour to be rigorously proved [50]. It splits a population up into a number of inter-connected subgroups (such as age groups [479]) and captures the different transmission dynamics between each group. This construction highlights the dual benefit that vaccination can have - vaccines that are infection-reducing directly protect the individuals that are vaccinated while transmission-reducing vaccines can also indirectly protect un-vaccinated individuals [480].

This dual benefit can significantly complicate the optimal vaccination problem when there is a negative correlation between the infectiousness of a group and the vulnerability of its members to the disease. Examples of this occur when the population is divided by age for diseases such as COVID-19 [481] and seasonal influenza [482]. In such cases, the optimal strategy may not be obvious and could be highly dependent on uncertain parameters [483], while the seemingly intuitive solution may be significantly sub-optimal [484]. Moreover, the complicated methods used to find the optimal solution, involving solving the adjoint equations derived via Pontryagin's Maximum Principle [460, 461] means

that the optimal solution may be difficult to understand or qualitatively justify to policy-makers.

When attempting to understand a complicated problem such as finding the optimal vaccination policy, it is often helpful to look at cases with extreme parameter values via asymptotic analysis, which helps the problem to be analytically solvable (at least to leading order). This can help form general principles for optimal vaccination policies. These principles can then be used both to form heuristics for finding the true optimal policy in a more general setting and also to explain the resultant optimal solution, as it is often comprised of a mixture of policies resulting from these principles.

There have been a number of recent papers that have used asymptotic analysis to derive general principles. [485] discusses a model with high reproduction numbers and shows that in this case, it is often optimal to vaccinate the less infectious groups in a population. Moreover, [246], building on the work of [486], linearises the model equations and derives a simple knapsack problem, although the solution to this problem is only optimal when considering the short-term evolution of the epidemic. Other special cases are investigated in [458] (which looks at a population with disconnected subgroups) and [487] (which examines the critical vaccination fraction for a population with separable mixing).

Two cases will be considered in this paper which both provide novel contributions to the literature. Firstly, the case of a population with a small vulnerable subgroup will be analysed, and it will be shown that, in the asymptotic limit (as the size of this population group tends to zero and its relative vulnerability tends to infinity), any vaccination policy is eventually outperformed by one where this group is vaccinated first. Of course, the concept that vaccinating vulnerable groups is important has been raised in many previous papers, such as [35] and [488], but the mathematically rigorous asymptotics presented here provide new evidence for the importance of this principle.

The second case to be discussed is that of a small total vaccination supply. The key novel result that will be shown is that (to leading order) the optimal vaccination problem reduces to a linear knapsack problem which can be easily solved. This knapsack problem differs from the one in [246] because, by linearising the final size equations rather than the model ODEs (ordinary differential equations), the optimal solutions and predictions of their behaviour are valid for the full evolution of the epidemic, rather than just in the short-term. Again, the case of a small vaccine supply has been examined in many papers such as [489, 490, 491], but these papers have simply analysed the optimisation problem in the standard way, without deriving the explicit leading order solution as is done in this paper.

In order to prove these results, it is necessary to build on previous literature. A number of results from [50] (found in Appendix E) are used in the course of the proof alongside some well-established results, such as the final size of an epidemic in SIR-type models [492]. However, the theorems presented

in the main text are completely novel, with their proofs requiring a significant extension of the current literature. In particular, the various propositions in the proofs (found in Appendix F) are, to the best of the authors' knowledge, new to the literature. Some of these results, such as, for example, the proof that epidemic final size is continuously dependent on initial conditions and the vaccination policy found in Proposition F.2 may also be helpful to those seeking to prove similar results.

The main analytic results will be further investigated through examples and, in particular, the small supply case will be used to show that it is not always optimal to vaccinate the most infectious group, even when all groups are equally vulnerable. The UK population's age structure will be used to relate these results to a realistic example, and optimal small-supply vaccination policies will be approximated for diseases with different age-dependent case fatality ratios.

The paper is structured as follows. Firstly, the multi-group SIR model will be introduced. Then, analytic results will be presented in the case of a small vulnerable subgroup, which will be explored through numerical examples. Finally, analytic results related to a small vaccination supply will be presented and again, examples will be used to illustrate the findings.

8.2 Modelling

8.2.1 Disease transmission and vaccination model

The model used in this paper is identical to the model presented in [50] and this section is simply a summary of the modelling section in [50]. The population is divided into n subgroups and each subgroup i is further divided into six compartments:

$$S_i := \text{Number of people that are in group } i\text{, are susceptible, and are unvaccinated} \tag{8.1}$$

$$I_i := \text{Number of people that are in group } i\text{, are currently infected, and were infected while unvaccinated} \tag{8.2}$$

$$R_i := \text{Number of people that are in group } i\text{, are recovered, and were infected while unvaccinated} \tag{8.3}$$

$$S_i^V := \text{Number of people that are in group } i\text{, are susceptible and are vaccinated} \tag{8.4}$$

$$I_i^V := \text{Number of people that are in group } i\text{, are infected and were infected after being vaccinated} \tag{8.5}$$

$$R_i^V := \text{Number of people that are in group } i\text{, are recovered and were infected after being vaccinated.} \tag{8.6}$$

$$\tag{8.7}$$

Using SIR principles, the model becomes

$$\frac{dS_i}{dt} = -\sum_{j=1}^{n}(\beta_{ij}^1 I_j + \beta_{ij}^2 I_j^V)S_i - \frac{U_i(t)S_i}{N_i - W_i(t)} \tag{8.8}$$

$$\frac{dI_i}{dt} = \sum_{j=1}^{n}(\beta_{ij}^1 I_j + \beta_{ij}^2 I_j^V)S_i - \mu_i^1 I_i \tag{8.9}$$

$$\frac{dR_i}{dt} = \mu_i^1 I_i \tag{8.10}$$

$$\frac{dS_i^V}{dt} = -\sum_{j=1}^{n}(\beta_{ij}^3 I_j + \beta_{ij}^4 I_j^V)S_i^V + \frac{U_i(t)S_i}{N_i - W_i(t)} \tag{8.11}$$

$$\frac{dI_i^V}{dt} = \sum_{j=1}^{n}(\beta_{ij}^3 I_j + \beta_{ij}^4 I_j^V)S_i^V - \mu_i^2 I_i^V \tag{8.12}$$

$$\frac{dR_i^V}{dt} = \mu_i^2 I_i^V \tag{8.13}$$

where

$$W_i(t) := \int_0^t U_i(s)ds, \tag{8.14}$$

and

$$N_i = S_i(t) + I_i(t) + R_i(t) + S_i^V(t) + I_i^V(t) + R_i^V(t) \tag{8.15}$$

is the size of group i. Moreover, the β_{ij}^α terms represent transmission from group j to group i and the μ_i^α terms give the infectious period of the relevant individuals in group i.

Here, $U_i(t)dt$ gives the number of individuals in group i that are vaccinated in the small time interval $[t, t+dt]$ and hence, $W_i(t)$ is the number of individuals that have been vaccinated in group i in $[0, t]$. It is assumed that these vaccinations are assigned randomly to the unvaccinated members of group i, so that each vaccine is given to a susceptible member of group i with probability

$$\frac{\text{number of susceptible members}}{\text{number of unvaccinated members}} = \frac{S_i}{N_i - W_i(t)} \tag{8.16}$$

Thus, the total rate of susceptibles being vaccinated is $\frac{U_i(t)S_i}{N_i - W_i(t)}$.

Note that there is a slight difference between this model and the one commonly found in the literature (in [234], [235] and [236] among many others) which set the vaccination term equal to $S_i U_i(t)$ instead of $\frac{U_i(t)S_i}{N_i - W_i(t)}$. As discussed in [50], this corresponds to vaccines that are randomly distributed to the whole population, which can be seen by rewriting the vaccination term as

$$S_i U_i(t)dt = \frac{S_i}{N_i} \times N_i U_i(t)dt \tag{8.17}$$

The first term on the right-hand side is then the probability of a randomly chosen member of group i being susceptible, while the second term is the total number of vaccines assigned in a small time interval $[t, t+dt]$, noting that here the dimension of $U_i(t)$ is 1/time (compared to the model used in this paper where the dimension of $U_i(t)$ is population/time) and hence it is necessary to scale by $N_i dt$ to convert $U_i(t)$ into a number of vaccines.

This is in contrast to the model in this paper which corresponds to vaccines that are randomly distributed only to the unvaccinated population. [50] provides justification for the use of this "unvaccinated-only model", which is therefore the one that will be used in this paper. However, they are structurally very similar, and so it would be possible to apply the results in this paper to the more commonly found model.

To deal with the (removable) singularity that can occur when $W_i = N_i$, it is assumed that

$$W_i(t) \leq N_i \quad \forall t \geq 0 \quad \text{and} \quad W_i(t) = N_i \Rightarrow \frac{U_i(t) S_i}{N_i - W_i(t)} = 0 \tag{8.18}$$

To capture the benefits of vaccination, there are additional constraints put on the β_{ij}^{α} and μ_j^{α} terms which are

$$\beta_{ij}^1 \geq \beta_{ij}^2, \beta_{ij}^3 \geq \beta_{ij}^4 \quad \text{and} \quad \mu_i^1 \leq \mu_i^2. \tag{8.19}$$

Finally, it will be assumed throughout the remainder of this paper that the population sizes are normalised so that

$$\sum_{i=1}^{n} N_i = 1 \tag{8.20}$$

Further details are given in [50].

8.2.2 Optimisation problem

The optimal vaccination problem considered in this paper aims to find the vaccination policy, $\boldsymbol{U}$, which minimises a weighted sum of the total number of infections in each group. Thus, the problem is

$$\min \Bigg\{ \sum_{i=1}^{n} p_i \Big(R_i(\infty) + \kappa_i R_i^V(\infty) \Big) : \sum_{i=1}^{n} U_i(t) \leq A(t), \quad \sum_{i=1}^{n} W_i(t) \leq B(t), \\ U_i(t) \geq 0, \quad W_i(t) \leq N_i \quad \forall t \geq 0 \Bigg\}. \tag{8.21}$$

Here, $A(t)$ represents the maximal vaccination rate, $B(t)$ represents the maximal vaccine supply and $R_i(\infty)$ and $R_i^V(\infty)$ are the limiting values of $R_i(t)$ and $R_i^V(t)$ as $t \to \infty$. The weights p_i and $p_i \kappa_i$ could

be interpreted in a number of ways, depending on the quantity of interest. For example, $p_i = \kappa_i = 1$ if one wanted to minimise infections, or p_i and $p_i\kappa_i$ could be the case fatality ratio of unvaccinated and vaccinated members of group i respectively if one wanted to minimise deaths. However, it is important to note that $\kappa_i \leq 1$ for each i as vaccinated members of the population should be no more vulnerable to the disease that their unvaccinated counterparts.

It is helpful to define $H(\boldsymbol{U})$ to be the objective function - that is

$$H(\boldsymbol{U}) = \sum_{i=1}^{n} p_i \Big(R_i(\infty) + \kappa_i R_i^V(\infty) \Big), \tag{8.22}$$

where R_i and R_i^V are found from solving the model equations with vaccination policy given by $\boldsymbol{U}$.

It will be assumed throughout this paper that all "feasible" $\boldsymbol{U}$ are sufficiently smooth for all the quoted theorems to hold. In general, this does not significantly restrict $\boldsymbol{U}$ - for example, the results in [50] simply require that each $U_i(t)$ is bounded and Lebesgue integrable, while Theorems 8.1 and 8.2 require only that $\boldsymbol{U}$ has finite support. Moreover, it is assumed that $B(t)$ is non-decreasing (as total supply should not decrease over time) and piecewise differentiable.

8.3 Results

8.3.1 A small, vulnerable subgroup

Consider the case where one of the groups in the population (which, without loss of generality, will be assumed to be group 1) is very small and vulnerable. That is, the population N_1 satisfies

$$N_1(\epsilon) = \epsilon << 1 \tag{8.23}$$

while the weights satisfy

$$p_1(\epsilon) = p_1 \quad \text{and} \quad p_i(\epsilon) = p_i^* \epsilon \quad \forall i \neq 1 \tag{8.24}$$

for some constants p_1 and p_i^*. It will be assumed that all κ_i are constant. In this setting, group 1 contains a very small proportion of the population, but each member of group 1 is much more vulnerable than the rest of the population.

Thus, this case is practically valid when there is a small subsection of the population that carries the majority of the vulnerability to a disease. As will be discussed further in Section 3.2.3, this has applicability to diseases such as COVID-19, where the majority of the deaths occur significantly older people, while it could also apply to diseases where there are rare conditions that cause a minority of

people to be much more vulnerable.

It is mathematically convenient to rescale the parameters p_i so that only p_1 depends on ϵ. This can be done by multiplying all the p_i terms by $\frac{1}{p_1\epsilon}$ so that

$$\tilde{p}_1(\epsilon) = \frac{1}{\epsilon} \quad \text{and} \quad \tilde{p}_i(\epsilon) = \frac{p_i^*}{p_1} := \tilde{p}_i \quad \forall i \neq 1. \tag{8.25}$$

This leads to an equivalent optimisation problem in the sense that the optimal vaccination policy will be the same. This occurs because the only change to the objective function is a scalar multiplication of $\frac{1}{p_1\epsilon}$ to each of the terms. Note that while this multiplicative factor tends to infinity as ϵ tends to 0, this system is only analysed for non-zero values of ϵ, and hence this rescaling is valid.

Analytic results

The first result presented in this section shows that, in the limit of a group with small size and large vulnerability (with the total cost of the whole group being infected, $N_1\tilde{p}_1$, remaining constant) any fixed vaccination policy where the vulnerable group is not vaccinated first will eventually (that is, for sufficiently small ϵ) be outperformed by a similar policy where the vulnerable group is vaccinated first.

Group 1 will be given a population size $N_1 = \epsilon$ and an infection cost $\tilde{p}_1 = \frac{1}{\epsilon}$ (recall that the $\tilde{p}_i$ represent the rescaled values of p_i, and so it is acceptable that $\tilde{p}_1 > 1$ for small ϵ). It will be assumed that the initial conditions in the group are proportional to ϵ, so that there exists some $\sigma \in (0, 1]$ such that the initial susceptible population is $\sigma\epsilon$ and the initial infected population is $(1 - \sigma)\epsilon$.

Before stating the full theorem, it is helpful to explain the various constraints and variables that will be introduced. Define, for each value of $\epsilon \geq 0$, $\boldsymbol{U}(t; \epsilon)$ to be the "fixed" vaccination policy where group 1 is not vaccinated first. Of course, the vaccination policy cannot be completely fixed, as the size, ϵ, of group 1 is decreasing, and so it will simply be assumed that the vaccines given out to each group satisfies

$$|W_i(t; \epsilon) - W_i(t; 0)| < \epsilon \quad \forall t \geq 0 \quad \text{and} \quad \forall i \in \{1, ..., n\} \tag{8.26}$$

Note that all groups are allowed to have small changes in the number of vaccinations they receive - this allows, for example, for vaccinations that would have been given to group 1 being reassigned as group 1's population shrinks.

Moreover, to reduce the lengths of the proofs, it will be assumed that $\boldsymbol{U}$ has uniformly bounded finite support - that is, there is some constant t_U such that for each $i \in \{1, ..., n\}$,

$$t > t_U \Rightarrow U_i(t; \epsilon) = 0 \quad \forall t, \epsilon \geq 0 \tag{8.27}$$

In order for group 1 to not be vaccinated first in the limit as $\epsilon \to 0$, there must be some time τ at which some fixed proportion w of the other groups have been vaccinated, while at least some fixed proportion $(1-\alpha)$ of group 1 has not been vaccinated. That is,

$$W_1(\tau;\epsilon) < \alpha\epsilon \quad \text{and} \quad \sum_{i=1}^{n} W_i(\tau;\epsilon) > w. \tag{8.28}$$

One can also define a vaccination policy $\tilde{\boldsymbol{U}}(t;\epsilon)$ where group 1 is vaccinated first. This will be done by re-directing all vaccinations from the $\boldsymbol{U}(t;\epsilon)$ policy to group 1 until it is fully vaccinated, and keeping the same vaccination policy after group 1 is fully vaccinated (ignoring any vaccines that $\boldsymbol{U}(t;\epsilon)$ assigns to group 1 after this time).

To ensure convergence of the model at $\epsilon = 0$, given $\Pi(\epsilon)$ defined by

$$\Pi(\epsilon) := \left\{ i : \exists t \geq 0 \quad \text{s.t.} \quad I_i(t;\epsilon) > 0 \right\}, \tag{8.29}$$

it will be assumed that $\Pi(\epsilon) = \{1, ..., n\}$ for all $\epsilon > 0$ (as any groups which never suffer any infections can be ignored) and that $\Pi(0) = \{2, ..., n\}$. While this second condition may not be strictly necessary for the theorem to hold, it is unrestrictive, and ensures convergence - if this were not the case, then it would be possible that infection in some set of groups were seeded only by group 1. Thus, when $\epsilon = 0$, these groups would suffer no infections, while for any $\epsilon > 0$, they would have an epidemic of size independent (at leading order) of ϵ.

The final condition on the model is that the people in group 1 can be infected by other groups, and that vaccinated members of group 1 gain protection from this infection. That is, there is some $i \in \{1, ..., n\}$ such that

$$\beta_{1i}^1 > \beta_{1i}^3 \geq 0. \tag{8.30}$$

This is an important condition, as if people in group 1 could only be infected by other members of group 1 then the total number of infections in group 1 would decay as $\epsilon \to 0$, meaning that it would no longer necessarily be optimal to vaccinate group 1 first (as most people in group 1 would not catch the disease anyway for small ϵ).

With these considerations, Theorem 8.1 can now be stated.

Theorem 8.1 *Suppose that for all $\epsilon > 0$,*

$$N_1(\epsilon) = \epsilon, \quad S_1(0;\epsilon) = \epsilon\sigma, \quad I_1(0;\epsilon) = (1-\sigma)\epsilon \quad \textit{and} \quad \tilde{p}_1(\epsilon) = \frac{1}{\epsilon} \tag{8.31}$$

for some $\sigma \in (0,1)$ and that all other parameter values and initial conditions are independent of ϵ.

Consider any vaccination policy with uniformly bounded finite support given by $\boldsymbol{U}(t;\epsilon)$ *and suppose that there exists fixed* $\alpha, \tau, w > 0$ *such that*

$$W_1(\tau;\epsilon) < \alpha\epsilon \quad \textit{and} \quad \sum_{i=1}^{n} W_i(\tau;\epsilon) > w \quad \forall \epsilon > 0. \tag{8.32}$$

Define a new policy, $\tilde{\boldsymbol{U}}(t;\epsilon)$*, given by*

$$\tilde{U}_1(t;\epsilon) = \begin{cases} \sum_{i=1}^{n} U_i(t) & \textit{if } \sum_{i=1}^{n} W_i(t;\epsilon) \leq \epsilon \\ 0 & \textit{otherwise} \end{cases} \tag{8.33}$$

and, for $i \neq 1$*,*

$$\tilde{U}_i(t;\epsilon) = \begin{cases} 0 & \textit{if } \sum_{i=1}^{n} W_i(t;\epsilon) \leq \epsilon \\ U_i(t;\epsilon) & \textit{otherwise} \end{cases}. \tag{8.34}$$

Suppose that for each $i \in \{1, ..., n\}$ *and* $t \geq 0$*,*

$$|W_i(t;0) - W_i(t;\epsilon)| < \epsilon. \tag{8.35}$$

Define

$$\Pi(\epsilon) := \{i : \exists t \geq 0 \quad s.t. \quad I_i(t;\epsilon) > 0\} \tag{8.36}$$

and suppose that $\Pi(\epsilon) = \{1, ..., n\}$ *for any* $\epsilon > 0$ *and that* $\Pi(0) = \{2, ..., n\}$*. Finally, suppose that there exists an* $i \in \{2, ..., n\}$ *such that*

$$\beta_{1i}^{1} > \beta_{1i}^{3} \geq 0. \tag{8.37}$$

Then, the policy $\tilde{\boldsymbol{U}}$ *is feasible and for sufficiently small* ϵ*,*

$$H(\boldsymbol{U}(t;\epsilon)) > H(\tilde{\boldsymbol{U}}(t;\epsilon)). \tag{8.38}$$

For the second theorem, it is helpful to note that, using the results in [50], if one defines

$$\chi(t) := \begin{cases} A(t) & \text{if} \quad \int_0^t A(s)ds < B(t) \\ \min(A(t), B'(t)) & \text{if} \quad \int_0^t A(s)ds \geq B(t) \end{cases}, \tag{8.39}$$

then (assuming that there is an optimal solution, and under mild smoothness conditions on $\boldsymbol{U}$, A and

B) there must be an optimal solution satisfying

$$\sum_{i=1}^{n} W_i(t) = \max\left(\int_0^t \chi(s)ds, 1\right). \tag{8.40}$$

The following theorem then proves that the limiting optimal vaccination policy vaccinates the vulnerable group as quickly as possible. To reduce the length of the proof, it will be assumed that $\sigma = 1$, so that (in the small ϵ limit) all members of group 1 can be vaccinated before being infected.

Theorem 8.2 *With the definitions of Theorem 8.1, suppose additionally that*

$$\sum_{j=2}^{n}(\beta_{1j}^1 - \beta_{1j}^3)I_j(0;\epsilon) > 0. \tag{8.41}$$

That is, the initial difference between the infective force on vaccinated and unvaccinated members of the population is positive. Suppose further that

$$\sigma = 1. \tag{8.42}$$

Suppose an optimal vaccination policy for each ϵ is given by $\overline{\boldsymbol{U}}(t;\epsilon)$ and suppose that $\overline{\boldsymbol{U}}(t;\epsilon)$ has uniformly bounded finite support. Then, there exists an η depending only on α, τ, w and the model parameters such that, for any $\boldsymbol{U}$ satisfying the condition (8.32) as defined in Theorem 8.1.

$$\epsilon \in (0,\eta) \Rightarrow H(\boldsymbol{U}) > H(\overline{\boldsymbol{U}}). \tag{8.43}$$

Moreover, there is a sequence of optimal vaccination policies, $\overline{\boldsymbol{U}}(t;\epsilon)$, which satisfies

$$\lim_{\epsilon\to 0}\left(\frac{\overline{W}_1(t;\epsilon)}{\epsilon}\right) = 1 \quad \forall t \quad s.t. \quad \int_0^t \chi(s)ds > 0. \tag{8.44}$$

Note that the existence of an optimal vaccination policy has been assumed in the statement of this theorem. The authors believe that an optimal policy should exist, as Proposition F.2 in the appendices can be used to show that $H(\boldsymbol{U})$ is continuous. However, more care would need to be taken with the smoothness assumptions on $\boldsymbol{U}$ to create a rigorous proof of this.

Theorems 8.1 and 8.2 are proved in the appendices.

Examples

To illustrate these analytic results, consider a simple two-group example. Suppose that group 1 is small, vulnerable, and non-infectious, while group 2 is large, invulnerable and infectious. These groups

could be interpreted as "old" and "young" respectively, although there is no specific physical situation being modelled here.

Suppose the transmission matrices are given by

$$\beta^1 = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}, \quad \beta^2 = \chi\beta^1 \quad \beta^3 = \rho\beta^1 \quad \text{and} \quad \beta^4 = \chi\rho\beta^1 \tag{8.45}$$

for some parameters χ and ρ which will be varied. This corresponds to the case of vaccination having (independently) an effectiveness χ at stopping people being infected and ρ at stopping infected people transmitting the disease. Moreover, suppose that

$$\mu_i^\alpha = 1 \quad \forall i, \alpha \tag{8.46}$$

and

$$N_1 = \epsilon, \quad \tilde{p}_1 = \frac{1}{\epsilon}, \quad \kappa_1 = 1 \quad N_2 = 1, \quad \tilde{p}_2 = p^* \quad \text{and} \quad \kappa_2 = 1, \tag{8.47}$$

for some parameter p^* that will be varied. Finally, suppose that the initial conditions are

$$S_1(0;\epsilon) = \epsilon, \quad I_1(0;\epsilon), = 0 \quad S_2(0;\epsilon) = 1 - I^* \quad \text{and} \quad I_2(0;\epsilon) = I^*, \tag{8.48}$$

for some parameter I^* that will be varied, and that the vaccination constraints are given by

$$A(t) = 1 \quad \text{and} \quad B(t) = \max(t, 1). \tag{8.49}$$

Consider therefore a vaccination policy where group 2, the infectious group, is vaccinated first (and hence, as $B(\infty) = N_2$, it is the only group that is vaccinated). That is,

$$U_1(t;\epsilon) = 0 \quad \text{and} \quad U_2(t;\epsilon) = \begin{cases} 1 & \text{if } t \leq 1 \\ 0 & \text{otherwise} \end{cases}. \tag{8.50}$$

Hence, with $\tilde{\boldsymbol{U}}$ defined as in Theorem 8.1, one has

$$\tilde{U}_1(t;\epsilon) = \begin{cases} 1 & \text{if } t \leq \min(1,\epsilon) \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad \tilde{U}_2(t;\epsilon) = \begin{cases} 1 & \text{if } t \in (\epsilon, 1] \\ 0 & \text{otherwise} \end{cases}. \tag{8.51}$$

Fig. 8.1 shows a comparison of the objective values $H(\boldsymbol{U}(t;\epsilon))$ and $H(\tilde{\boldsymbol{U}}(t;\epsilon))$ for different values of ϵ. As expected, when $\epsilon = 1$, vaccinating the more infectious group first is optimal (as they have the same

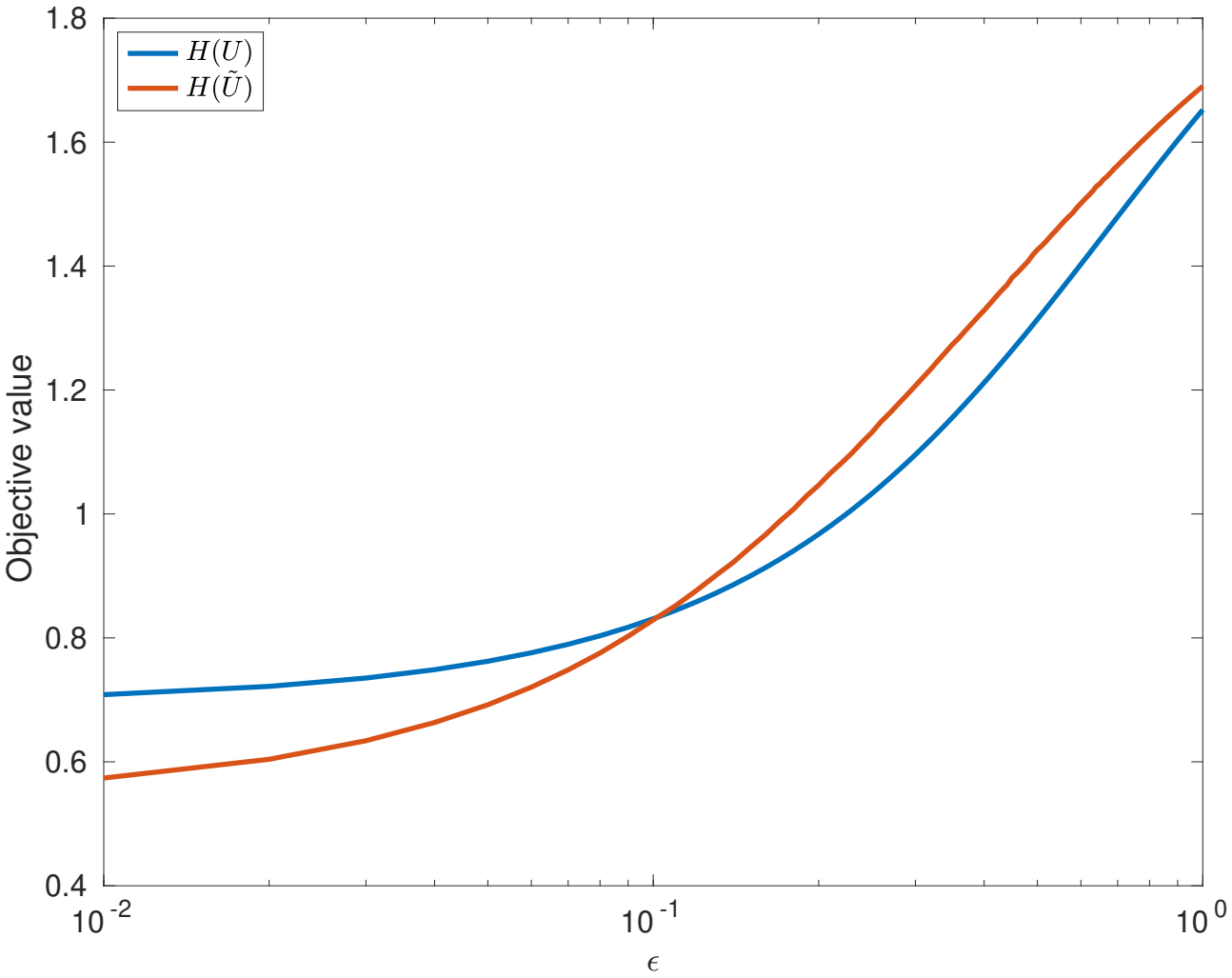


Figure 8.1: A comparison of the two vaccination policies, $\boldsymbol{U}(t;\epsilon)$ (where the infectious group is vaccinated first) and $\tilde{\boldsymbol{U}}(t;\epsilon)$ (where the vulnerable group is vaccinated first) for different values of ϵ. Note that here, $I^* = 0.01$, $\chi = \rho = 0.5$ and $p^* = 1$.

vulnerability in this case), while for ϵ smaller than around 0.1, it becomes more effective to vaccinate the vulnerable group first, illustrating the results of Theorem 8.1.

It is useful to consider the approximate smallness of ϵ required in Theorem 8.1. That is, how small ϵ needs to be in order for $\tilde{\boldsymbol{U}}(t;\epsilon)$ to be the better vaccination policy. To explore this, define, for each value of I^* and p^*,

$$\epsilon^*(I^*, p^*) := \inf\left(\left\{\epsilon : H(\tilde{\boldsymbol{U}}(t;\epsilon)) > H(\boldsymbol{U}(t;\epsilon))\right\} \cup \{1\}\right). \tag{8.52}$$

That is, $\epsilon^*(I^*, p^*)$ is the smallest value of ϵ such that vaccinating group 2 first is better that the $\tilde{\boldsymbol{U}}$ policy, with a cut-off value at 1 (as it is possible that for some parameter values, the $\tilde{\boldsymbol{U}}$ policy is always better).

Fig. 8.2 shows the behaviour of $\epsilon^*(I^*, p^*)$. As expected, ϵ^* is decreasing in I^* - this is because when there are fewer initial infectious people, there is more time to vaccinate the infectious group before

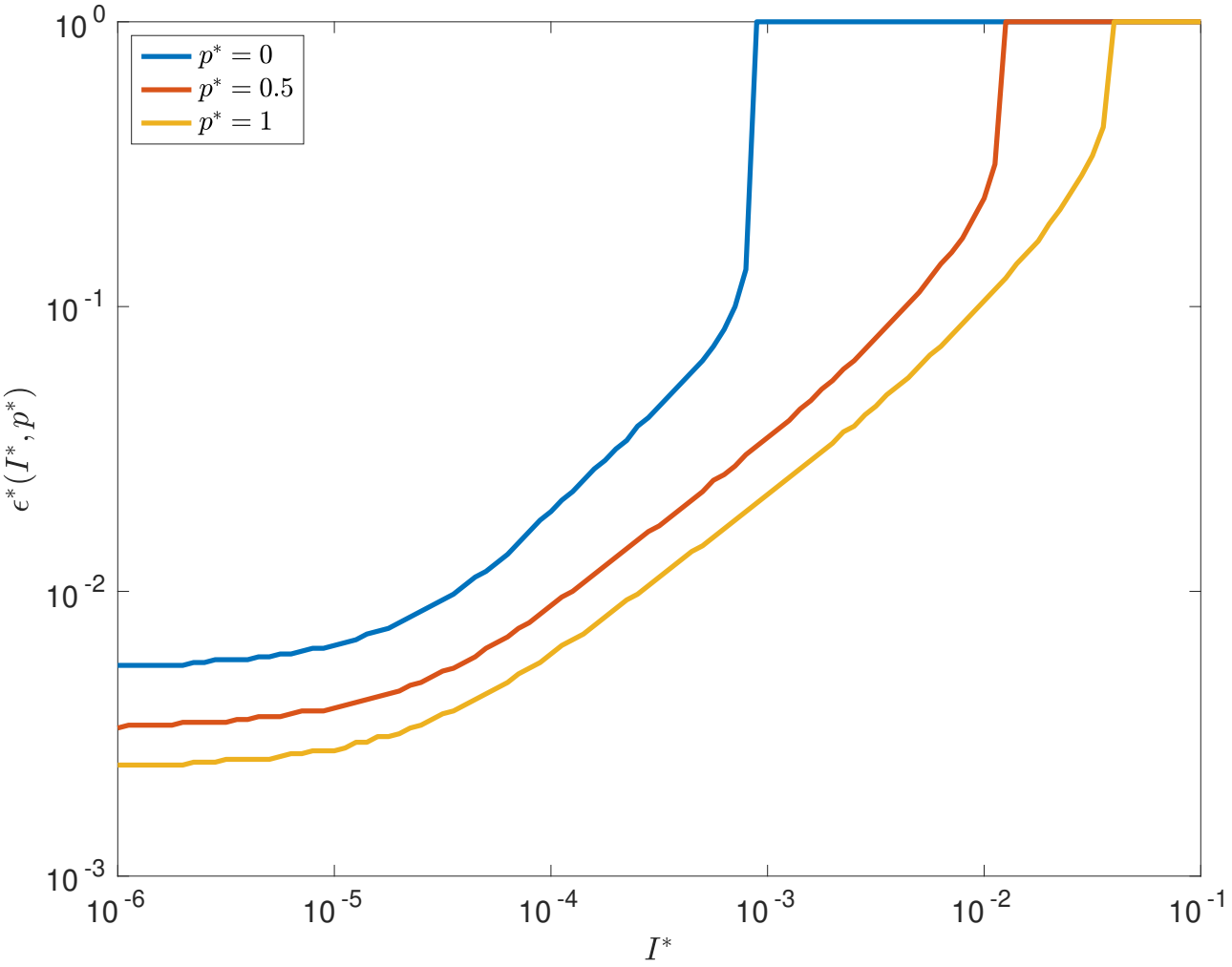


Figure 8.2: A plot of $\epsilon^*(I^*, p^*)$, the highest value of ϵ for which $\boldsymbol{U}$ is a better vaccination policy that $\tilde{\boldsymbol{U}}$. Note that ϵ^* is capped at 1, so that a value of 1 indicates that there were no values found of ϵ^* such that $\boldsymbol{U}$ was the better policy. Note that here, $\chi = \rho = 0.5$.

the epidemic has a chance to grow, reducing the peak of the epidemic. Moreover, ϵ^* is decreasing in p^*, as higher values of p^* mean that the number of infections in group 2 is valued higher.

Moreover, Fig. 8.2 suggests that, for each fixed p^*, ϵ^* is uniformly bounded below for all I^*. Indeed, this is expected as when I^* is very small, there are negligible infections within the interval $t \in [0, 1]$ and so the vaccination policies $\boldsymbol{U}$ and $\tilde{\boldsymbol{U}}$ are in effect being carried out in a completely uninfected population. As the R_0 (that is, the initial growth rate of the disease) number of a fully vaccinated population (in this case) is greater than 1, $I(t; \epsilon)$ will reach an $O(1)$ value regardless of the vaccination policy. Thus, while decreasing I^* will increase the time to reach this $O(1)$ value, it will not significantly change the final infections in the epidemic, and hence ϵ^* should converge to a fixed value for small I^*.

When the fully vaccinated population has an R_0 lower than 1, the difference between $\boldsymbol{U}$ and $\tilde{\boldsymbol{U}}$ is more distinct. Indeed, provided I^* is small enough for vaccination to be completed before many infections have occurred, one would expect $O(I^*)$ infections in group 2 in either of the two vaccination

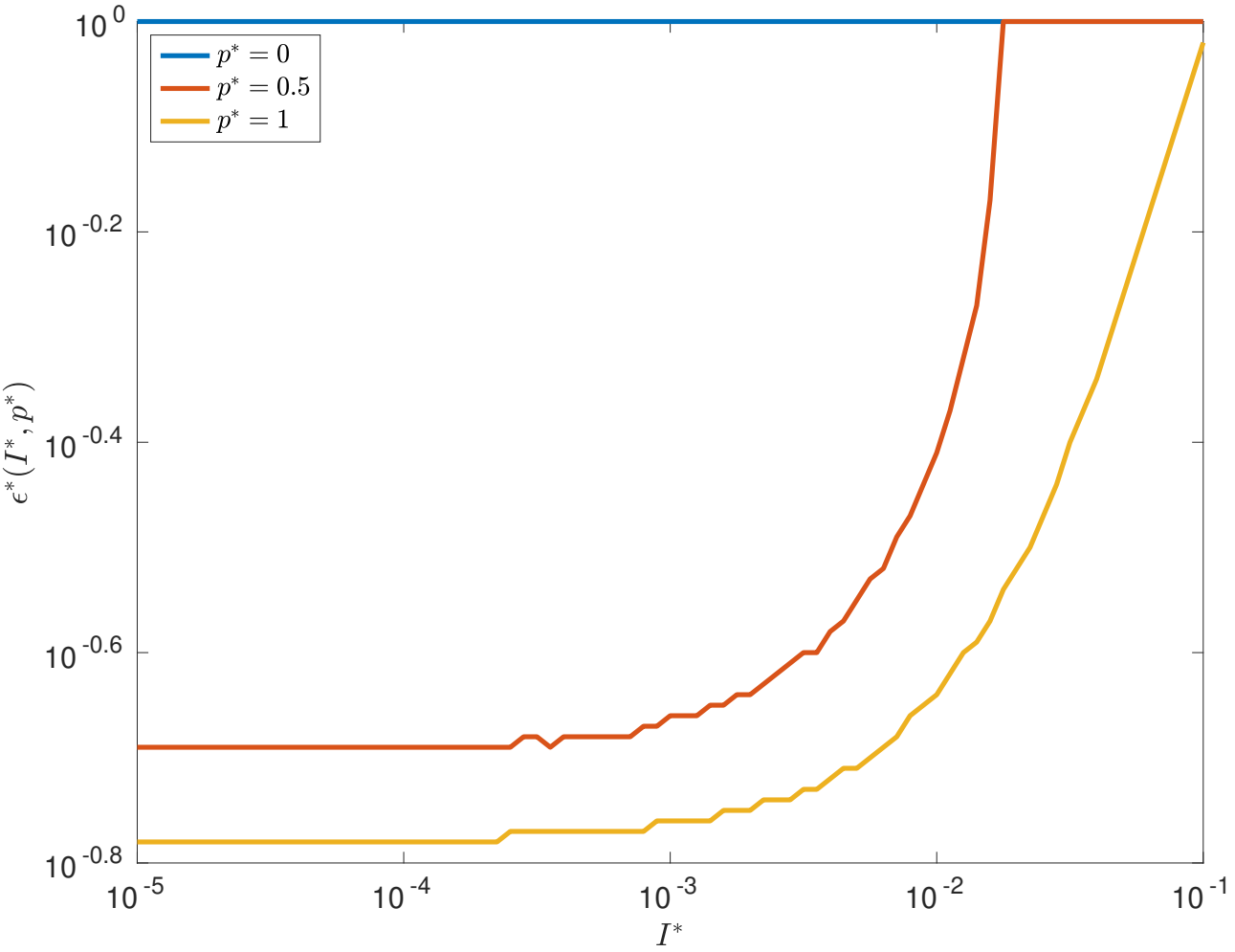


Figure 8.3: A plot of $\epsilon^*(I^*, p^*)$, the highest value of ϵ for which $\boldsymbol{U}$ is a better vaccination policy that $\tilde{\boldsymbol{U}}$, in the case of complete vaccination effectiveness (so $\chi = \rho = 0$). Note that, because the values of the objective function are $O(I^*)$, there is some numerical instability which has caused some non-smoothness of the plot.

policies (for sufficiently small ϵ), as in both policies, the size of the infected compartment will be decreasing after the vaccination has been completed. However, in the $\boldsymbol{U}$ case, one would expect $O(I^*\epsilon)$ infections in total in group 1 (as there is an $O(I^*)$ infection force on a group of size $O(\epsilon)$ for $O(1)$ time), while in the $\tilde{\boldsymbol{U}}$ case, one would expect $O(I^*\epsilon^2)$ infections in total in group 1, as the population of this group is only of size $O(\epsilon)$ for $O(\epsilon)$ time. This behaviour is illustrated in Fig. 8.3, which shows that ϵ^* converges to significantly higher values than in Fig. 8.2 - indeed, in the case that $p^* = 0$, it appears that $\boldsymbol{U}$ is never optimal for any $\epsilon \leq 1$.

8.3.2 A small vaccination supply

In this section, the case of a small, immediately available vaccine supply will be considered. In this case, it will be possible to analytically derive the optimal vaccination policy (in the limit of small supply).

This case may be particularly relevant if there was an outbreak of a disease where a vaccine already existed (so that some vaccinations are available immediately), but where supplies were limited, and scaling production would take a significant amount of time. An example of this can be found in the recent mpox outbreak [493] where the United Kingdom initially purchased 20 000 smallpox vaccines. This small figure - not even enough to vaccinate 0.1% of the UK population [465] - would certainly fall within the small vaccination supply case.

Moreover, one can use the results in this section regardless of the time at which vaccinations become available (that is, they are not only relevant at the start of an epidemic). This would be of practical use whenever vaccine production is slow, or when the disease is sufficiently mild (or vaccine production is sufficiently expensive) that a large-scale vaccination program is not deemed economically feasible.

Analytic results

To state the analytic result from this section, it is helpful to define

$$\beta'_{ij} = \begin{cases} \beta^1_{ij} & \text{if } i,j \leq n \\ \beta^2_{i(n-j)} & \text{if } i \leq n < j \leq 2n \\ \beta^3_{(n-i)j} & \text{if } j \leq n < i \leq 2n \\ \beta^4_{(n-i)(n-j)} & \text{if } n < i,j \leq 2n \end{cases}, . \tag{8.53}$$

This large transmission matrix captures the dynamics of all $2n$ susceptible and infectious groups in the model (both vaccinated and unvaccinated). Indeed, after vaccination has been completed, there is no movement from S_i to S_i^V so β' allows for the model to be considered as a 2n-group SIR model without vaccination. Thus, in particular, one can derive a simple final size relation for the total number of infections in the epidemic. Similarly, define

$$\mu'_i = \begin{cases} \mu^1_i & \text{if } i \leq n \\ \mu^2_{(i-n)} & \text{if } n < i \leq 2n \end{cases} \tag{8.54}$$

and

$$p'_i = \begin{cases} p_i & \text{if } i \leq n \\ \kappa_{(i-n)} p_{(i-n)} & \text{if } n < i \leq 2n \end{cases}. \tag{8.55}$$

In this case of small supply, it is possible to effectively differentiate the final size of the epidemic with respect to the vaccination policy, and use the resultant linear approximation to form a simple knapsack

problem for the optimal vaccination policy. This will involve writing the objective in the form

$$H(\boldsymbol{U}(t;\epsilon)) = H(\boldsymbol{0}) + \boldsymbol{y}^T \boldsymbol{W}(\tau(\epsilon);\epsilon) + o(\epsilon) \tag{8.56}$$

where $\boldsymbol{W}$ is the final vaccination amounts in each group. To define the gradient, $\boldsymbol{y}$, it is necessary to use the inverse of a matrix $\boldsymbol{Q}$ given by

$$Q_{ij} = \frac{1}{1 - e^{-\sum_{j=1}^{2n} \frac{\beta'_{ij}}{\mu'_j} R_j(\infty;0)}} \left[\delta_{ij} + \frac{S_i(0;0) e^{-\sum_{j=1}^{2n} \frac{\beta'_{ij}}{\mu'_j} R_j(\infty;0)} \beta'_{ij}}{\mu'_j} \right], \tag{8.57}$$

where as before, the variables $f_i(t;\eta)$ indicate the value of the relevant model variable at time t, given that the parameter ϵ is equal to η, and δ_{ij} is the Kronecker delta. Then, $\boldsymbol{y}$ is defined by

$$\boldsymbol{x} = \boldsymbol{Q}^{-T} \boldsymbol{p}' \quad \text{and} \quad y_i = \frac{S_i(0;0)}{N_i}(x_{i+n} - x_i) \quad \forall i \in \{1, ..., n\}. \tag{8.58}$$

These definitions allow for the theorem to be stated.

Theorem 8.3 *Suppose that, for all* $\epsilon > 0$

$$B(t;\epsilon) = \epsilon \quad \forall t \geq 0. \tag{8.59}$$

and that all other parameter values and initial conditions are independent of ϵ*. Suppose that* $A(t)$ *is a continuous function with*

$$A(0) > 0 \tag{8.60}$$

and that the matrix M *is invertible. For sufficiently small* ϵ*, define*

$$\tau(\epsilon) := \inf \left\{ t : \int_0^t A(s) ds = \epsilon \right\}. \tag{8.61}$$

Suppose that $\boldsymbol{U}$ *satisfies the condition*

$$\sum_{i=1}^{n} U_i(s) = \min \left(\int_0^t \chi(s) ds, 1 \right), \tag{8.62}$$

where χ *is defined in (8.40). Then, for sufficiently small* ϵ*, the objective function is given by*

$$H(\boldsymbol{U}(t;\epsilon)) = H(\boldsymbol{0}) + \boldsymbol{y}^T \boldsymbol{W}(\tau(\epsilon);\epsilon) + o(\epsilon). \tag{8.63}$$

Moreover, if there is a unique element of $\boldsymbol{y}$ *equal to the minimum of* $\boldsymbol{y}$ *then the optimal vaccination policy (to leading order in* ϵ*) is uniquely given by*

$$U_i(t;\epsilon) = \begin{cases} A(t) & \text{if } i = \min\{y_i : i \in \{1,...,n\}\} \quad \text{and } \int_0^t A(s)ds < \epsilon \\ 0 & \text{otherwise} \end{cases}. \tag{8.64}$$

The second part of the theorem assumes a unique minimal element of $\boldsymbol{y}$. This is not guaranteed to happen, and if there were multiple groups with equal values of $\boldsymbol{y}$, this could mean that the effectiveness of vaccinating these groups would be equal to $O(\epsilon)$. However, any sets of parameters satisfying this condition would be unstable to small perturbations (as a trivial example, consider perturbing the initial susceptible populations $S_i(0,0)$ of the groups with a minimal values of y_i). Thus, in any practical scenario, the probability that the best estimates of the parameters give multiple minimal values of y_i is very small.

Theorem 8.3 is proved in the appendices.

Vaccinating a homogeneous population

To illustrate the effectiveness of this approximation, consider first an example of a homogeneous population (so $n = 1$). Consider the case where $\beta^1 = \beta$, $\beta^2 = \beta^3 = 0.5\beta$ and $\beta^4 = 0.25\beta$ for some parameter β that will be varied. Suppose moreover that

$$N_1 = \mu_1^1 = \mu_1^2 = p_1 = \kappa_1 = A(t) = 1, \quad S_1(0) = 1 - 10^{-4} \quad \text{and} \quad I_1(0) = 10^{-4}. \tag{8.65}$$

Finally, suppose $B(t) = \epsilon$ where ϵ will be varied.

Fig. 8.4 shows a comparison of the predicted and actual change in number of deaths, ρ_1 for two values of ϵ. It illustrates that, even when $\epsilon = 0.1$, a relatively large value, $\boldsymbol{y}$ gives a good approximation of the true value (found by simulation). Moreover, when $\epsilon = 0.01$, the two lines are almost indistinguishable. This is useful validation for the approximation, as the correction term was simply proved to be $o(\epsilon)$ rather than, for example, $O(\epsilon^2)$, and so it is encouraging that the predictions are so close.

An interesting property of Fig. 8.4 is that the value of β for which vaccination is most effective appears to be very close to $S(0)\beta = 1$ (as $S(0) \sim 1$). Note that here, as $\mu = 1$, this is equal to the initial reproduction number of the disease. This has the perhaps surprising consequence that if one has a set of disconnected, equally vulnerable subgroups, a small vaccination supply should be assigned to a group with initial reproduction number close to 1, rather than giving it to the group with the

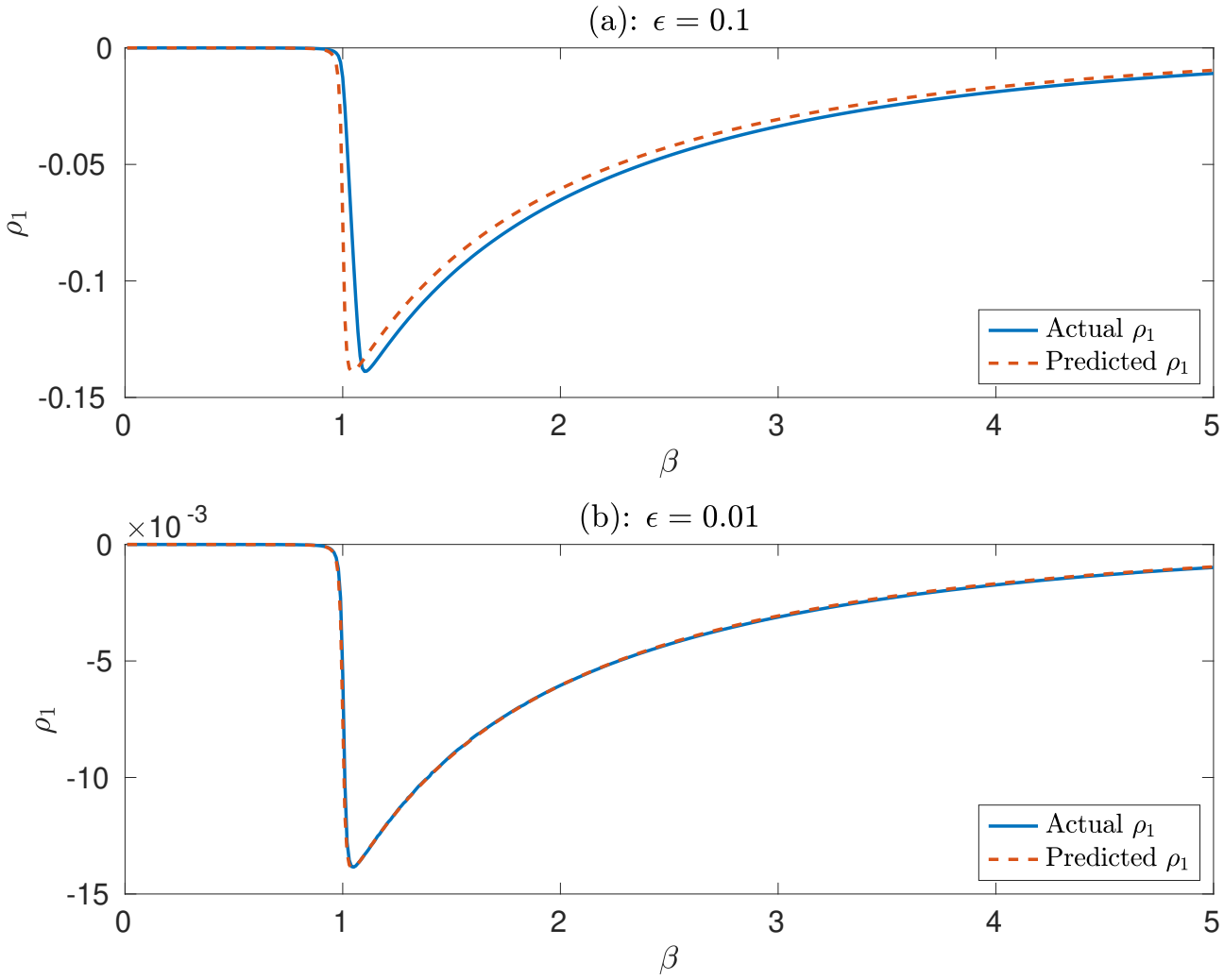


Figure 8.4: A comparison of the predicted and actual values of the change in deaths, ρ_1, in the case of a homogeneous population for two different values of vaccination supply, ϵ and for different values of infectivity, β. Note the different scales on the two y axes.

highest value of β (that is, the most group with the most infectious individuals). This result is in line with the findings of [485] which showed that vaccinating less infectious groups can be more effective, and is an important consideration for vaccination policy planning.

Application to age-structured populations

Consider assigning a small quantity of vaccinations to an age-structured population, using the example of the UK. The disease model has been estimated using the inter-age-group contact matrices $\mathbf{\Lambda}$ from [214], alongside population estimates $\boldsymbol{N}$ from [465]. As in [214], this gives a transmission matrix of

$$\beta_{ij}^1 = \beta \frac{\Lambda_{ij}}{N_j} \tag{8.66}$$

for some scalar parameter β. As in the previous section, it will be assumed that

$$\mu_i^\alpha = 1 \quad \forall i, \alpha \tag{8.67}$$

and

$$\beta^2 = 0.5\beta^1, \quad \beta^3 = 0.5\beta^1 \quad \text{and} \quad \beta^4 = 0.25\beta^1. \tag{8.68}$$

It will also be assumed that the initial infected population is small, so that, for each i

$$S_i(0;\epsilon) = (1 - 10^{-4})N_i \quad \text{and} \quad I_i(0;\epsilon) = 10^{-4}N_i. \tag{8.69}$$

In the following examples, β will be chosen so that the disease-free next generation matrix of a completely unvaccinated population, given by

$$R_{ij} = \frac{N_i \beta_{ij}^1}{\mu_j^1} = \beta_{ij}^1 \tag{8.70}$$

has a spectral radius (that is, largest eigenvalue) equal to 4. This sets the R_0 number in the overall population to be 4. To illustrate the population structure, Fig. 8.5 shows a heatmap of the matrix R_{ij}. This highlights the strongly assortative nature of the contacts (that is, members of a subgroup are most likely to be contacts with members of their own subgroup), while also showing that contacts are lower for older age groups.

Now, two different age-dependent case-fatality ratios will be considered - uniform case-fatality and approximate COVID-19 case fatality, taken from [80]. In both cases, it will be assumed that vaccination reduces the case fatality ratio by 90% (following the results of [80] for the COVID-19

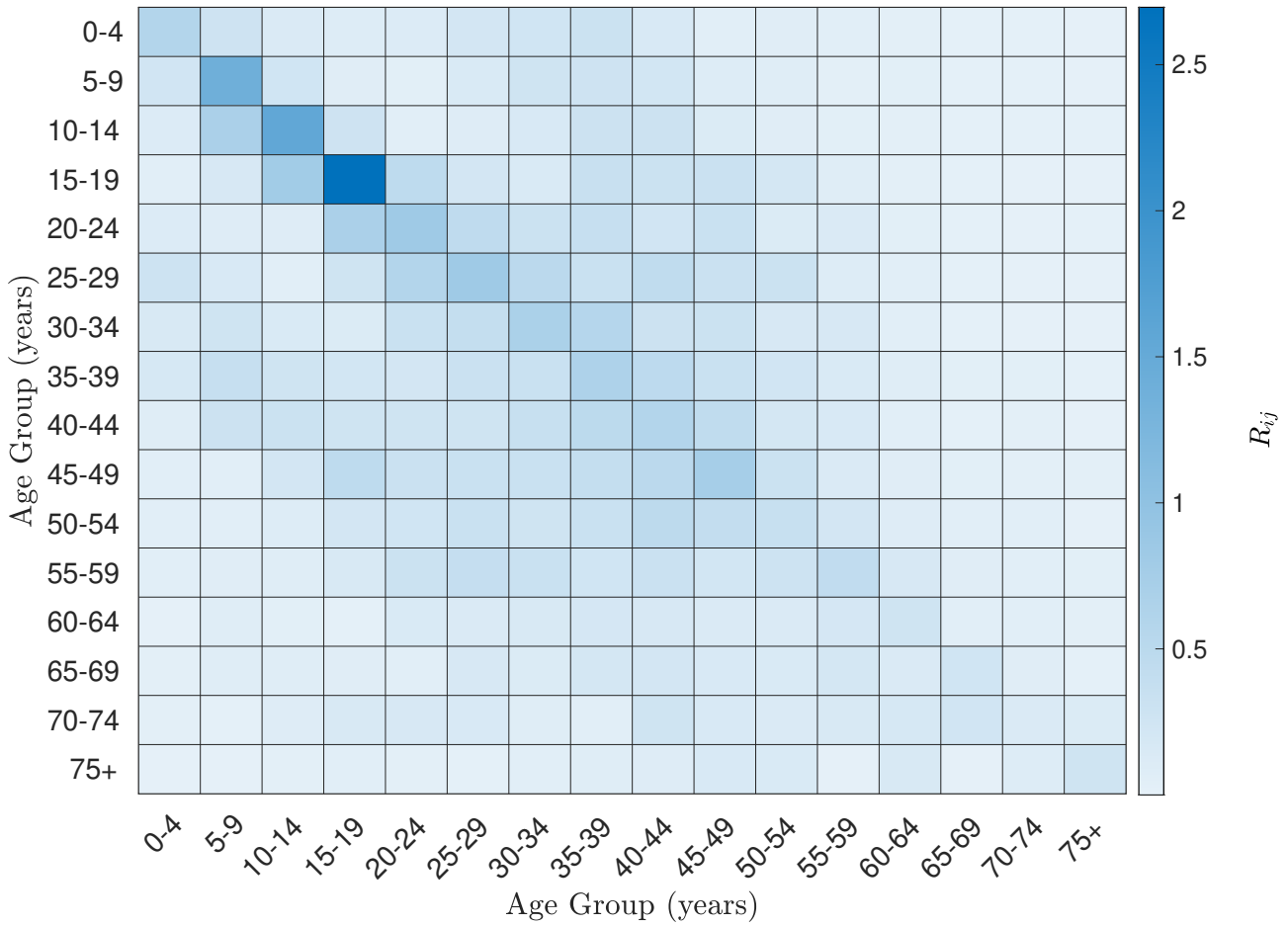


Figure 8.5: A heatmap of the next generation matrix for the age-structured UK population.

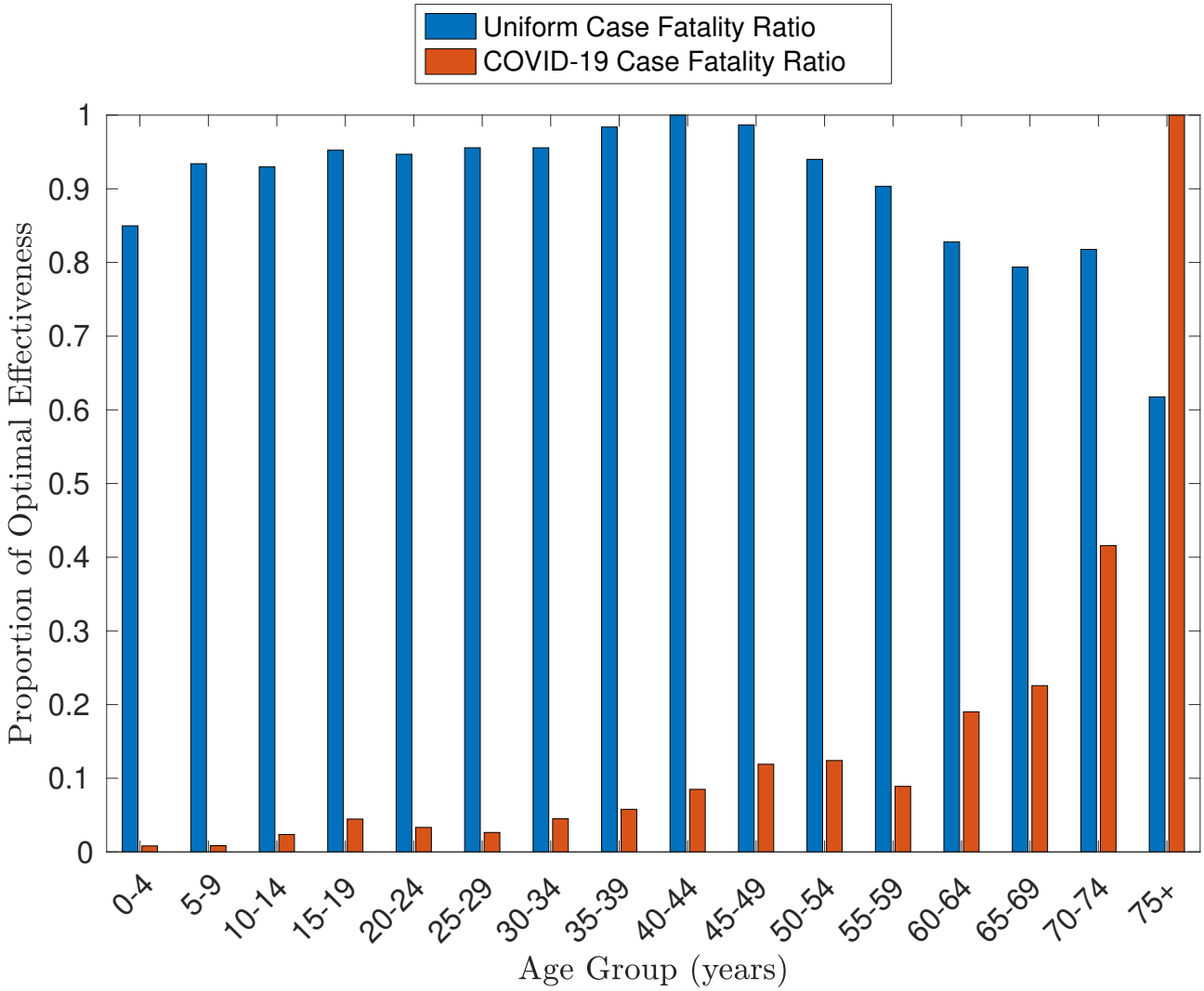


Figure 8.6: The effectiveness of assigning a small quantity of vaccines to each age group as a proportion of the optimal effectiveness.

vaccines) so that $\kappa_i = 0.1$ for all i. However, it is worth emphasising that this model is simply based on real-world data, and does not seek to accurately model the COVID-19 pandemic.

Fig. 8.6 shows the effectiveness of vaccinating each age group in the two different cases, as a proportion of the optimal effectiveness. Note that here the proportion of effectiveness of assigning vaccine to group i is given by $\frac{y_i}{\min_j(y_j)}$, as each y_j is non-positive. It highlights that the significantly higher mortality rates for COVID-19 for the older age groups means that vaccinating them is much more effective than vaccinating the other age groups. This is an example of Theorems 8.1 and 8.2, as the oldest age group makes up a relatively small percentage (around 9%) of the population, but, if one scales p such that it has median value 1, the $p_i N_i$ value for the oldest age group is approximately 20, and so is definitely $O(1)$ rather than $O(\epsilon)$.

A perhaps surprising exception to the general correlation between effectiveness and mortality is the relatively low effectiveness of vaccinating the 55-59 year old age group, which is lower than the 45-49

year old and 50-54 year old groups. This illustrates the non-intuitive nature that optimal vaccination policies can take, and the importance of investigating their behaviour fully. The main reason for this low effectiveness is that, while the 55-59 year old age group is more vulnerable to COVID-19 than the younger groups, according to [214], they have much less contact with the 75+ year old age group and thus vaccinating this group provides significantly less secondary protection to most vulnerable members of the population. The authors speculate that this could be due to a significant number of the parents of the 55-59 year old age group having died (particularly in comparison to the younger groups), reducing their links with the 75+ year old age group. Moreover, those in the 55-59 year old age group may also not be old enough to have many 75+ year olds in their social circles (in comparison to members of older groups). However, further investigation would be needed to justify this claim.

In the case of uniform mortality, the vaccination policy becomes even less intuitive, as Figure 8.6 shows that the optimal age group to vaccinate is the 40-44 year olds. Indeed, from Fig. 8.5, it may seem that the 15-19 year old group would be the best group to vaccinate, as they have the highest overall transmission - that is, the maximum value of

$$\text{Total infectious force of group } j := \sum_{i=1}^{16} R_{ij}. \tag{8.71}$$

However, if instead, one considers

$$\text{Total external infectious force of group } j := \sum_{i=1, i\neq j}^{16} R_{ij}, \tag{8.72}$$

then it is the 35-39 and the 40-44 age groups which have the highest values. This can be considered in conjunction with the results of the previous subsection, which showed that vaccinating groups with R_0 numbers close to 1 is optimal for disconnected populations. Indeed, the "secondary effect" of vaccinations (that is, the number of people who are not vaccinated, but are protected from the disease because of vaccines given to others) can be higher for groups with lower internal infectious force, particularly when their external infectious force is higher.

Finally, it is useful to again explore the range of values for ϵ for which $\mathbf{y}$ gives a good approximation of the true number of infections. As the minimum (scaled so that the total population size is 1) value of N_i is 0.0498 in this case, ϵ will be tested at 0.0498. The results of this are shown in Fig. 8.7, which again illustrates the effectiveness of this approximation. Indeed, the largest error across either case is of order 10^{-4}, which in turn is of order $\epsilon^2\boldsymbol{y}$. This suggests that the $o(\epsilon)$ correction term in Theorem 8.3 is significantly smaller than ϵ, which increases the usefulness of this approximation. However, further

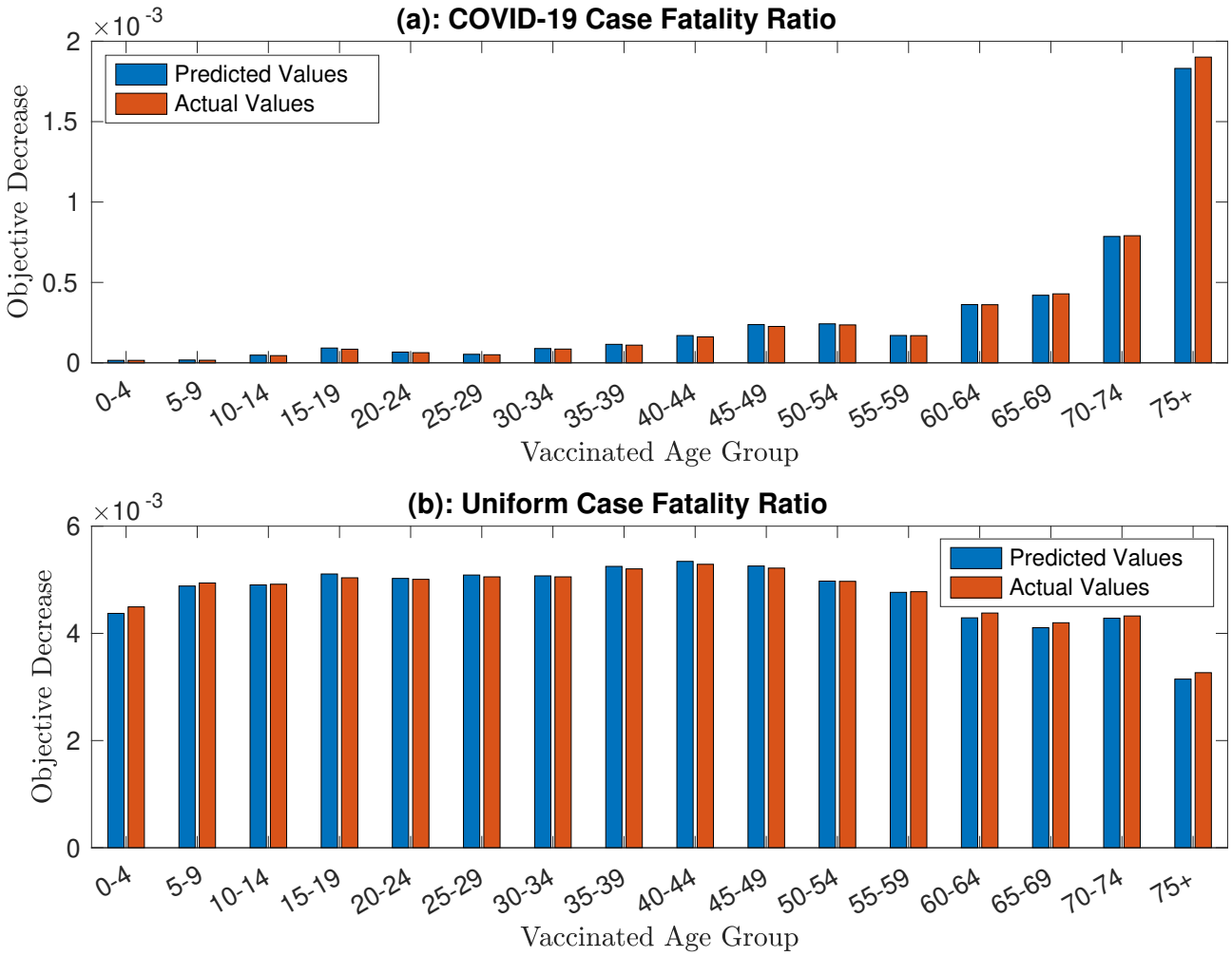


Figure 8.7: A comparison of the predicted and simulated change in the objective function when vaccinating each individual group at $\epsilon = 0.0498$. Both the cases of a COVID-19 case fatality ratio (in (a)) and a uniform case fatality ration (in (b)) are presented.

investigation is needed to determine whether this correction is of $O(\epsilon^2 \boldsymbol{y})$ for all parameter values.

8.4 Discussion

This paper has shown two general principles for optimal vaccination policies by looking at the asymptotic behaviour of the optimal policy in the case of extreme parameters. Firstly, it has shown that small, vulnerable groups should in general be vaccinated first, regardless of the overall timetable of vaccination. This is an important result as it requires very little data on the population - merely the case fatality ratios and populations of the different subgroups - and in particular needs no forecasting of future transmission trends or vaccine supply.

The analytically derived results (in the limiting case) also show that the effect of vaccinating this small group far outweighs the effect of vaccinating any of the other groups. Indeed, if the size of the vulnerable group is $O(\epsilon)$ and the case fatality ratio of the other groups is $O(\epsilon)$, then Theorem 8.1

shows that vaccinating the vulnerable group will lead to an $O(\epsilon)$ decrease in the number of fatalities, while vaccinating the same number of people from another group will only decrease this by $O(\epsilon^2)$. This result is of practical importance for diseases such as COVID-19, where the majority of the fatalities would be from certain age groups within the population. In particular, it provides strong evidence for the importance of sharing vaccines on a global scale, as this is the only way to ensure that vaccinations can be given to all people who are most vulnerable to the disease.

However, this result should be used with caution, as it certainly does not imply that a population should always be vaccinated in order of decreasing vulnerability to the disease. The optimal vaccination policy is, in general, a balance between directly protecting the vulnerable by vaccinating them, and by indirectly protecting them by vaccinating those groups with the highest infectiousness. This is shown in Figure 8.6 by the fact that, when a COVID-19 case fatality ratio is used, the relative effectiveness of vaccinating each age group does not decrease everywhere with age. The results of Theorems 8.1 and 8.2 simply provide a principle that in the asymptotic limit, the optimal strategy is to vaccinate small, vulnerable groups first. In the absence of data on vaccination effectiveness (which is crucial in determining whether indirectly protecting the more vulnerable population may be better), this provides a mathematically sound justification for beginning with the most vulnerable members of a population while gathering data to determine the rest of the vaccination policy.

The second principle derived in this paper was a linear approximation to the change in number of fatalities from a disease, which allows for the estimation of the optimal vaccination policy in the case of a small total supply. Again, this principle is flexible, applying for any set of parameters, and provides a computationally cheap way of the approximating the optimal solution, even for large numbers of groups, as it merely requires the solution of a linear system involving the same number of variables as the number of groups.

A useful feature of this approximation is that it appears to have high accuracy even for reasonably large values of the total supply, such as when 10% of the population can be vaccinated. Figs. 8.4 and 8.7 show that there is very little deviation between the predicted and actual values of the objective function and so suggest that this is a flexible and widely applicable method of approximation, even when the population contains a large number of subgroups. However, it would be helpful to strengthen the results of Theorem 8.3 to get a stronger bound on the error for small ϵ to ensure that this similarity holds for all models.

The results presented in Fig. 8.4 are also informative for vaccination policy. As shown in Fig. 8.4, in a completely homogeneous population, vaccination has the most effect when the reproduction number ($\frac{\beta}{\mu}$ in this case) is slightly bigger than 1, with a steep decline in effectiveness for reproduction

numbers below 1, and a more gradual decline for large reproduction numbers. This result allows one to consider the "vaccination leverage" of a population - that is, the effectiveness that a small quantity of vaccination can have - and shows that, even in the case of homogeneous case fatality ratios, vaccinating in order of infectiousness may be far from optimal, as it is much more difficult to reduce infections in a highly infectious population.

Indeed, a similar idea was shown to apply when the UK age structure was considered. In the case of uniform case fatality, the optimal group to assign a small amount of vaccinations to was the 40-44 age group which, as shown by Fig. 8.5, is not the most infectious group. This perhaps counter-intuitive results highlights the importance of mathematically justifying the principles one uses to decide on optimal vaccination policies, as "common-sense" arguments may in fact give false conclusions. Communicating such principles to governments and policy-makers will be crucial in future pandemics, particularly ones with more homogeneous case fatality ratios where the optimal policy is not as intuitive as for diseases like COVID-19.

An important limitation of Theorem 8.3 is that the optimal policies for small vaccination supplies do not necessarily generalise to give the beginning of the optimal vaccination policy in the case of a much larger vaccination supply. Indeed, it is possible to have bifurcations in the optimal vaccination policy as the supply increases - for example, it can become possible to completely avoid an epidemic by vaccinating a large quantity of an infectious group. Thus, while the linear approximation can be a useful starting point when attempting to estimate the optimal strategy, it is important to consider alternatives when a large proportion of the population can be vaccinated.

The results of this paper are only applicable if the trajectory of the disease in question can be well-approximated by multi-group SIR dynamics. In particular, this requires there to be reasonably high levels of the disease in a population (otherwise stochastic dynamics change the epidemic behaviour [494]), and for population subgroups to be sufficiently large (again to prevent stochasticity dominating). Moreover, the model assumptions would not hold if individuals could be re-infected, or if the effect of vaccination was not eternal (though if the timescale of the epidemic was sufficiently shorter than the timescale of immunity decay, then the model would still provide a good approximation).

A final barrier to using the results in this paper is that that estimation error in the model parameters could lead to the optimal solutions being incorrectly calculated. Estimating the β_{ij}^{α} parameters is particularly complicated, especially in a multi-group setting where it is difficult to establish the chain of transmission between different groups. Because of this, building models based on contact rates between groups (estimated using surveys [214]) or proxies such as commuting patterns [495]) may be the best method, at least to provide priors on the parameters. Theorems 8.1 and 8.2 are significantly

less susceptible to errors in parameters, as they do not require any of the β_{ij}^{α} or μ_j^{α} parameters to be known, although the level of "smallness" of ϵ would vary depending on the disease in question. Theorem 3 is significantly more susceptible to error, as all the model parameters are needed. However, while there may be bifurcations in the optimal strategy, the optimal value of the objective function should depend continuously on the parameters (a fact which could be proved by extending the results of Proposition F.2), limiting the effect of small estimation errors.

Despite this, the authors expect that similar results to those presented in Theorem 3 will hold for a very wide class of deterministic models. Essentially, the only necessary characteristic of the model that is required by Theorem 3 is that the objective function, $H(\boldsymbol{U})$, is a continuously differentiable function of the vaccination policy $\boldsymbol{U}$ in some neighbourhood of $\mathbf{0}$. Indeed, $\boldsymbol{y}$ in Theorem 3 can be replaced by $\nabla \boldsymbol{H}(\mathbf{0})$ in a general setting. Certainly, it should be conceptually simple (though perhaps algebraically complicated) to generalise this result to other compartmental models such as SEIR (Susceptible-Exposed-Infected-Recovered) and even those modelling vector-transmitted diseases.

The authors also expect that Theorem 8.1 will hold for general models where the effect of vaccination is eternal. The essential points in the proof of Theorem 8.1 are that vaccinating the small group does not affect the overall vaccination program (to leading order) and that it does have an O(1) effect on the objective function. Both of these should still hold in a wide range of models, although it may be difficult to define the meaning of "very small group" and "very vulnerable group" - particularly in more complicated settings such as individual-based models.

This work could be extended by deriving more principles for extreme parameter values and investigating whether they generalise to realistic model parameters. By combining the existing results in this paper and others such as [485] with potential new ones, one could create an algorithm that creates good heuristics of optimal vaccination policies that could be used as starting points for accurately approximating the optimal policy for a general parameter set. This could have significant implications for the design of vaccination policies, as it would enable the optimisation problem to be estimated for very complex models, as the time taken to converge to an optimal solution would significantly decrease given good initial heuristics.

8.5 Conclusion

The results of this paper are summarised below:

- If a sufficiently vulnerable, sufficiently small population exists in this multi-group SIR model, it is optimal to vaccinate this group first.

- For small overall vaccination supplies, the optimal vaccination problem can be well approximated by a simple knapsack problem.
- This linearisation appears to be a good approximation even for relatively large vaccination supplies (such as 10% of the population).
- This linearisation shows that, in the case of uniform case fatality, it is not necessarily optimal to vaccinate the most infectious group.

Statement of Authorship for joint/multi-authored papers for PGR thesis

To appear at the end of each thesis chapter submitted as an article/paper

The statement shall describe the candidate's and co-authors' independent research contributions in the thesis publications. For each publication there should exist a complete statement that is to be filled out and signed by the candidate and supervisor **(only required where there isn't already a statement of contribution within the paper itself).**

Title of Paper	**Asymptotic analysis of optimal vaccination policies**
Publication Status	**Published in the Bulletin of Mathematical Biology**
Publication Details	**Matthew J Penn and Christl A Donnelly. "Asymptotic analysis of optimal vaccination policies". In: Bulletin of Mathematical Biology 85.3 (2023)**

Student Confirmation

Student Name:	**Matthew Penn**		
Contribution to the Paper	**Conceived of and developed the mathematical results in this paper.** **Carried out the numerical experiments.** **Wrote the first draft of the paper.**		
Signature		Date	20/03/24

Supervisor Confirmation

By signing the Statement of Authorship, you are certifying that the candidate made a substantial contribution to the publication, and that the description described above is accurate.

Supervisor name and title: **Professor Christl Donnelly**		
Supervisor comments **As described above, this is Matt's initiative and follow-through. I provided supervision and suggestions for editing.**		
Signature 	Date	20 March 2024

This completed form should be included in the thesis, at the end of the relevant chapter.

Chapter 9: Summary and conclusions

9.1 Summary of main findings

9.1.1 Phylogenetics

Papers I-III detailed a range of novel phylogenetic methodology which has the potential to be of high value to future research. The Phylo2Vec representation developed in Paper I has many advantages compared to its contemporaries, such as the Newick format [496], including its natural ability to efficiently sample trees and its easily-computable tree distance metric. It is a representation specifically designed for optimisation and enables a range of optimisation procedures to be simply implemented, including a hill-climbing algorithm that showed comparable performance to optimisation via SPR moves. While this hill-climbing algorithm was substantially slower than contemporary algorithms, it provides an important proof of concept, illustrating that the Phylo2Vec representation is a sensible parameterisation of tree space that can allow for the development of more effective optimisation algorithms. With the help of the open-source software package that was made available with this paper, we hope that future researchers can exploit this representation to develop such methods.

A complicating factor in the Phylo2Vec construction is its label-asymmetry, where the labels assigned to the different taxa can have a substantial effect on the optimisation procedure. In particular, since changing v_i has the most direct effect on node i, the condition $v_i \in \{0, ..., 2(i-1)\}$ means that there are fewer ways of directly affecting the position of i when i is small. Thus, implementing the most simple optimisation procedure, where the labels assigned to each taxon remain fixed, may lead to ineffective optimisation if nodes with low labels are initially in the wrong position in the tree. We address this limitation in Paper I by using a "shuffling" algorithm which changes the labels assigned to each node, allowing our optimisation algorithm to converge to the correct tree. However, it is important to consider this property when designing new algorithms.

Building on Phylo2Vec, GradME, developed in Paper II, represents a novel method of tree optimisation. By recasting the balanced minimum evolution optimisation problem into a continuous space (for each subset of ordered trees), it is able to make use of the plethora of tools that have been developed to optimise differentiable objective functions. It showed good performance in comparison to FastME,

finding a better tree in two of the eleven main test cases and recovering a substantially more accurate tree on the jawed vertebrate dataset. Moreover, its ability to directly consider rooting trees, using our novel theory of rooted balanced minimum evolution, showed promise, outperforming midpoint rooting on a Eutherian Mammal phylogeny.

Despite its success, there are some limitations to GradME. While the majority of the optimisation process is continuous, GradME currently can only perform calculations within an ordered tree space. This means that there is a necessarily discrete element to GradME, the Queue Shuffle, to change the ordering of the taxa labels. However, while this makes the optimisation procedure more complicated, we have found the Queue Shuffle to be very effective, supported by our theoretical results, and so do not see this as a major limitation.

More importantly, GradME suffers from a high computational complexity in comparison to other algorithms such as FastME. Thus, it could not be easily applied to large optimisation problems where it is likely to have the most theoretical benefit over FastME. However, by applying it to sections of a large tree in turn, or by developing an algorithm which more efficiently calculates only a part of the gradient, we hope that it could be a useful tool for large trees. Moreover, the fact that it was able to infer two small trees more accurately than FastME shows that there are cases where it can improve the accuracy of optimisation while still terminating in a reasonable timeframe.

A final limitation of GradME is the fact that it only functions for a balanced minimum evolution objective rather than for Felsenstein's likelihood. The more complicated mathematical structure of this likelihood means that we have not been able to develop a similar algorithm (and, even if it were possible, it is probable that the high computational cost of GradME would be even more pronounced in this case). However, the work of Paper III in linking balanced minimum evolution to an approximate phylogenetic likelihood means that GradME could still be used for Bayesian inference through, for example, Hamiltonian Markov Chain Monte Carlo (MCMC).

This link between these two different optimisation paradigms, developed in Paper III, has important consequences for phylogenetics. The derivation of the entropic likelihood, together with the explanation of its correlation with Felsenstein's likelihood and the proof of its asymptotic equivalence with the balanced minimum evolution objective function, is a substantial theoretical breakthrough. This result demonstrates mathematically why these two objective functions have been shown to lead

to similar trees so frequently when applied to large phylogenetic datasets.

Most importantly, the methodology presented in Paper III allows for a Bayesian application of distance-based methodology. Comparing the results of MCMC using our entropic likelihood to FastME's bootstrap provides the expected results, as the entropic likelihood's region of posterior support lies within that of the bootstrap. Applying this likelihood to a dataset of 60 million sites and 363 taxa, something which would be computationally infeasible with traditional likelihood methods, provided a sensible tree reconstruction, closely aligning with previous studies of this dataset. With comparatively small computational overhead costs, we believe that our methods could allow for a new era of phylogenomic analysis, allowing the posteriors of a range of genome-scale datasets to be rapidly explored using our single, simple-to-implement method.

Despite these advantages, the approximations inherent to the entropic likelihood mean that its performance is not as high on datasets where Felsenstein's likelihood can also be used. These errors are particularly pronounced if long branches are present in the dataset as their lengths are difficult to estimate when calculating the inter-taxa distance matrix. Moreover, distance-based methods lack the ability to marginalise over missing sites in the data, meaning that the pairwise distances may not all be calculated from the entire genome, with these gaps particularly prevalent if the taxa in question have high evolutionary distances between them. Thus, the entropic likelihood is not a replacement for Felsenstein's likelihood but instead a tool that can be used on much larger datasets to gain fast and informative, though not completely accurate, insights.

Overall, the work in Papers I-III satisfied both the aims of this thesis. By showing that balanced minimum evolution can be applied to rooted trees and, more importantly, deriving a link between it and Felsenstein's likelihood, they have advanced our theoretical understanding of this phylogenetic paradigm. The algorithms developed in Papers II and III, supported by the new tree representation in Paper I, provide exciting new directions for phylogenetics, presenting novel approaches for analysing a range of datasets.

9.1.2 Epidemic variance

Paper IV provided a robust mathematical foundation for understanding the behaviour of a range of commonly-used branching processes in epidemiology. Many of these models are special cases of the Crump-Mode-Jagers process considered in this paper and so equivalent results can be easily found

by substituting the appropriate functions into the formulae presented in this work. While the main text of the paper focused on the variance of epidemics in continuous time, the supplementary material provided a range of additional results, including a variety on discrete epidemics, which broadens the potential utility of this paper.

Starting from a renewal equation for the probability generating function, which could be used to derive a wide range of other results, Paper IV explored the resultant renewal equation for the variance of prevalence. It was possible to decompose this equation to identify the contribution of different parts of the model to uncertainty, something which would be extremely difficult to do accurately through simulation alone. This decomposition allows for an assessment of the impact of different modelling choices and perhaps even the impact of interventions on the stochastic behaviour of the epidemic as well as simply its mean. Moreover, the linearity of this renewal equation means that it can be efficiently solved and so the uncertainty can be quickly evaluated for a wide range of scenarios. The paper also included an equation for the prevalence variance when forecasts are made midway through an epidemic, providing another useful tool for epidemiologists.

A key theoretical result from this paper was that, in all realistic scenarios, the prevalence was overdispersed, even without overdispersion in the number of infections from a given individual. This highlights the importance of understanding and including this intrinsic variance in the behaviour of epidemics, particularly since this dominated the epistemic uncertainty from the estimation of the model parameters in the presented examples.

However, while these results are powerful, they are limited by the assumptions inherent to the Crump-Mode-Jagers process. In particular, once a non-negligible proportion of the population has been infected with a disease, the reproduction number will begin to naturally decrease because of susceptible depletion (as a proportion of each infected person's contacts will be immune to the disease). As well as its obvious impact on the mean prevalence, this phenomenon will also have an effect on the uncertainty (as a trivial example, the number of people who can be infected is bounded above by the population size - a condition which cannot be included within the Crump-Mode-Jagers framework). Thus, care must be taken when applying these models during large epidemics, particularly when making long-term forecasts where this dependence between epidemic trajectory and parameter behaviour is especially important.

Overall, this section of the thesis clearly satisfied the aim of contributing to our mathematical understanding of branching process models, providing formulae describing key quantities and categorizing their constituent parts to aid in their interpretation. These results also allow for efficient calculations of variance in real-world situations, and we hope that the code we provided alongside Paper IV will help future researchers make use of this theory.

9.1.3 Optimal vaccination

Papers V and VI derived a range of results describing the behaviour of the optimal vaccination policy at different values of the model parameters. The main result of Paper V (broadly, that vaccinating more people or vaccinating people earlier will not increase the number of infections or deaths from a disease) is deeply intuitive. Indeed, this result has been assumed, either explicitly or implicitly, both throughout a range of papers in the literature and by policy-makers around the world. However, while this means that there are no immediate real-world implications from Paper V, it is important that assumptions such as this are rigorously tested, particularly when the same models are used to recommend vaccination policies. Were this result not to hold under such a model, then this could either have serious implications for the model's reliability in providing sensible vaccination policies or, more seriously, our understanding of epidemic behaviour under vaccination. As was shown by considering the counter-intuitive behaviour of deaths at finite times, models do not always behave in the way one might expect them to, and so establishing this result for such a broadly-used type of model is important, even though Paper V simply confirmed our intuition.

A limitation of Paper V is that this result is currently restricted to multi-group SIR models. As many contemporary modellers add in different compartments, with an exposed compartment being a particularly common choice, this paper will not apply directly to their models. However, this paper provides an important first step towards a more general result that could apply to a much wider range of models, as will be discussed in the "Future Work" section.

Paper VI examined the behaviour of the optimal vaccination policy under two different cases of "extreme" parameters. Firstly, it considered the case of a small, vulnerable subgroup within a population. Such groups of clinically extremely vulnerable people have often been treated separately from those of the same age group as was shown, for example, during the COVID-19 pandemic [497]. Paper VI provided justification for this approach, showing that a sufficiently small, sufficiently vulnerable subgroup within a population will be fully vaccinated at any (fixed) non-zero time in the optimal vac-

cination policy. Although the threshold for the optimality of vaccinating vulnerable subgroups first depends on the exact disease in question, this general theoretical result, explored further in the paper through simulation, means that the most clinically vulnerable people should be strongly considered for early vaccination.

The second section of Paper VI considered the case of a small vaccination supply. In this scenario, it derived an approximate solution to the optimal vaccination problem, which, for each group, approximated the effect of assigning the small vaccine supply to them. This novel result allows the rapid calculation of approximately optimal policies even when a reasonable proportion of the population (such as the approximately 5% tested in the paper) can be vaccinated. Moreover, it could be used as a computationally cheap heuristic in the creation of optimal policies when larger supplies of vaccine are available.

However, while these both of these results provide theoretical insights with useful practical implications, there is currently no measure of the error of these approximations for a given set of parameter values, with the theorems simply holding in their asymptotic limit. Thus, care must be used in applying them to real-world scenarios, and it should be particularly emphasised that the range of parameters for which they are good approximations will be affected by the properties of the disease under consideration. However, they nevertheless provide useful starting points to approximate and explain optimal vaccination policies.

The papers in this section make a clear novel contribution to our theoretical understanding of the optimal vaccination problem. Alongside the main theorems, there is also a wide range of smaller lemmas proved in the supplementary material which may be of use to future researchers looking to develop similar results. The methodological applications of these papers are more limited than the other papers in this thesis, with many of the results simply confirming pre-existing assumptions about disease behaviour, but the small-supply solution in Paper VI has the potential to be used practically in future algorithms.

9.2 Future work

Many of the papers in this thesis provide a "first step" in a particular research direction, and therefore there is much that could be done to extend and refine their results. Therefore, this section details some future avenues of research, again separated into the three broad areas discussed in this thesis.

Phylogenetics

A key limitation of balanced minimum evolution, the theory on which the results presented in this section ultimately rely, is the reduction in the amount of information used to calculate the objective, condensing millions or even billions of data points into a distance matrix often containing merely thousands of entries. Of course, this is also the source of the efficiency of the method presented in Paper III, but it seems likely that a sensible middle ground between balanced minimum evolution and exact likelihood methods could be found. This could be achieved by increasing the size of the reduced dataset, perhaps to $\mathcal{O}(n^3)$, where n is the number of taxa, rather than $\mathcal{O}(n^2)$, while maintaining its independence from the genome length. For example, it appears likely to the authors that somehow recording the uncertainty in the distance estimates could be helpfully integrated into an objective function. This may be particularly useful in datasets with large numbers of gaps to avoid overweighting small, but commonly aligned, segments of the genome.

Another important area for future research would be attempting to understand the error distribution when approximating Felsenstein's likelihood with the entropic likelihood. By treating the entropic likelihood as a noisy approximation of Felsenstein's, one could then improve the accuracy of the resultant samples. Accounting for this source of uncertainty would provide more realistic estimates of the posterior distribution rather than the overconfident clustering around the balanced minimum evolution optimum which can occur, particularly on smaller datasets.

Finally, increasing the efficiency with which the gradient of the continuous objective introduced in Paper II can be calculated would allow it to be used on datasets with larger numbers of taxa. Following the FastME method, a discrete "gradient" can be calculated in at worst $\mathcal{O}(n^3)$ time, while our continuous gradient requires $\mathcal{O}(n^5)$ operations. Narrowing this gap is crucial in ensuring that GradME, shown to provide more accurate results than FastME, can be used on these larger datasets where it can provide meaningful biological insights.

Epidemic variance

Understanding the effect of susceptible depletion and other more complex phenomena such as waning immunity on the variance of epidemics is crucial when presenting information and recommendations to policy-makers after a substantial proportion of the population has already been infected. Achieving this will require a different approach to that used in Paper IV as the self-similarity property will no longer hold. However, we believe that, at least in some simple cases, it should be possible to provide an

analytic quantification of the uncertainty. A good starting point would be to consider the asymptotic behaviour, both in short and long timescales, and, building on this, to potentially consider the final size of an epidemic.

There are also connections between the work in this chapter and the work that has been carried out on phylogenetic trees. In particular, one could combine the Crump-Mode-Jagers model of disease transmission with our tree construction methods to attempt to estimate the connections between different recorded cases. Developing an objective function which depends both on the disease transmission model and genetic data could be a powerful tool to supplement traditional contact-tracing methods in estimating true transmission within a population. Combining the method of Papers III and IV, it could also be possible to calculate credible intervals for this transmission, even when the number of cases is reasonably large.

Optimal vaccination

As discussed in more detail in Paper V and Paper VI, it appears that most of the theorems presented in these works should hold for a much wider range of disease transmission models. Developing results that held for a much more general class of models would ensure that they were more widely applicable without the need to repeat derivations for the multitude of different model variations that are present in the literature. By distilling the proofs of Papers V and VI into their key steps and recasting these in a more general setting, the author believes that generalising these results would be possible, though there would undoubtedly be complications that would require further mathematical insight.

Another area for development would be to consider the case of waning immunity. In this case, many of the results proved in this thesis, such as the optimality of maximal-effort vaccination, would cease to be true (as, for example, vaccinating a population long before the epidemic arrives would mean that their immunity would have almost completely disappeared when infections begin to rise). However, understanding the optimal policy in these scenarios is still vital, as basing decisions on a model without waning immunity may lead to unintended negative long-term consequences.

9.3 Conclusions

Like all branches of science, epidemiology is constantly evolving and, as illustrated by the previous section, progress often brings with it a multitude of new questions. However, the techniques and results developed in this thesis have provided some useful insights into a range of epidemiological

areas, as summarised in the bullet points below.

- Phylo2Vec, a simple integer-valued vector representation of phylogenetic tree space, imparts a sensible topology on the space of trees and provides the flexibility for developing a range of methods for optimising tree objective functions (Papers I-III).
- The balanced minimum evolution objective function may also be applied to rooted trees, which can allow for more accurate inference (Paper II).
- By transforming the balanced minimum evolution problem to an expectation maximisation problem over a continuous parameter space, gradient-based methodologies can outperform contemporary optimisation algorithms (Paper II).
- Balanced minimum evolution can be interpreted as proportional to an approximate likelihood, closely related to Felsenstein's likelihood, providing justification for its usefulness on a range of phylogenetic datasets (Paper III).
- Using distance-based methodologies to perform Bayesian inference is efficient and effective, allowing insights to be derived on genome-scale data without a prohibitive computational cost (Paper III).
- Aleaotoric variance plays a crucial role in epidemic behaviour and can dominate the effects of epistemic variance (Paper IV).
- In a general Crump-Mode-Jagers branching process model, an explicit renewal equation for the aleatoric prevalence variance can be derived (Paper IV).
- In all realistic scenarios, the prevalence is always overdispersed, even without overdispersion of the infections caused by an individual member of the population (Paper IV).
- It is always optimal to vaccinate people as early as possible and as much as possible under the considered multi-group SIR model of disease transmission (Paper V).
- A sufficiently small, sufficiently vulnerable subgroup in a population should (in the asymptotic limit) be vaccinated as quickly as possible to reduce the cost of an epidemic (Paper VI).
- When the vaccine supply is small, one can derive a linear leading-order approximation to the optimal vaccination objective function that performs well in a realistic scenario with 5% vaccination coverage (Paper VI).

Bibliography

[1] Jocelyne Piret and Guy Boivin. "Pandemics throughout history". In: *Frontiers in Microbiology* 11 (2021), p. 631736.

[2] JFD Shrewsbury. "The plague of Athens". In: *Bulletin of the History of Medicine* 24.1 (1950), pp. 1–25.

[3] Kyle Harper. "Pandemics and passages to late antiquity: rethinking the plague of c. 249–270 described by Cyprian". In: *Journal of Roman Archaeology* 28 (2015), pp. 223–260.

[4] Christer Bruun. "The Antonine Plague in Rome and Ostia". In: *Journal of Roman Archaeology* 16 (2003), pp. 426–434.

[5] Ole Jørgen Benedictow. *The Complete History of the Black Death*. Boydell & Brewer, 2021.

[6] Remi Jedwab, Noel D Johnson, and Mark Koyama. "The economic impact of the Black Death". In: *Journal of Economic Literature* 60.1 (2022), pp. 132–178.

[7] Niall PAS Johnson and Juergen Mueller. "Updating the accounts: global mortality of the 1918-1920 "Spanish" influenza pandemic". In: *Bulletin of the History of Medicine* (2002), pp. 105–115.

[8] Andrew T Price-Smith. *Contagion and chaos: disease, ecology, and national security in the era of globalization*. MIT press, 2008.

[9] Irena Ilic and Milena Ilic. "Historical review: Towards the 50th anniversary of the last major smallpox outbreak (Yugoslavia, 1972)". In: *Travel Medicine and Infectious Disease* 48 (2022), p. 102327.

[10] Marco A Biamonte, Jutta Wanner, and Karine G Le Roch. "Recent advances in malaria drug discovery". In: *Bioorganic & Medicinal Chemistry Letters* 23.10 (2013), pp. 2829–2843.

[11] Sarentha Chetty et al. "Recent advancements in the development of anti-tuberculosis drugs". In: *Bioorganic & Medicinal Chemistry Letters* 27.3 (2017), pp. 370–386.

[12] Gulfaraz Khan et al. "Novel coronavirus pandemic: A global health threat". In: *Turkish Journal of Emergency Medicine* 20.2 (2020), p. 55.

[13] Osama M Al-Quteimat and Amer Mustafa Amer. "The impact of the COVID-19 pandemic on cancer patients". In: *American Journal of Clinical Oncology* (2020).

[14] Jaspreet Singh and Jagandeep Singh. "COVID-19 and its impact on society". In: *Electronic Research Journal of Social Sciences and Humanities* 2 (2020).

[15] Amanpreet Behl et al. “Threat, challenges, and preparedness for future pandemics: A descriptive review of phylogenetic analysis based predictions”. In: *Infection, Genetics and Evolution* 98 (2022), p. 105217.

[16] Delphine Destoumieux-Garzón et al. “Getting out of crises: Environmental, social-ecological and evolutionary research is needed to avoid future risks of pandemics”. In: *Environment International* 158 (2022), p. 106915.

[17] Mahmoud Kandeel et al. “Omicron variant genome evolution and phylogenetics”. In: *Journal of Medical Virology* 94.4 (2022), pp. 1627–1632.

[18] S-W Chan et al. “Analysis of a new hepatitis C virus type and its phylogenetic relationship to existing variants”. In: *Journal of General Virology* 73.5 (1992), pp. 1131–1141.

[19] Susmita Shrivastava et al. “Whole genome sequencing, variant analysis, phylogenetics, and deep sequencing of Zika virus strains”. In: *Scientific reports* 8.1 (2018), p. 15843.

[20] Nita Madhav et al. *Pandemics: risks, impacts, and mitigation*. 2018.

[21] Jalal Poorolajal. *The global pandemics are getting more frequent and severe*. 2021.

[22] Katherine F Smith et al. “Global rise in human infectious disease outbreaks”. In: *Journal of the Royal Society Interface* 11.101 (2014), p. 20140950.

[23] Muzna Alvi and Manavi Gupta. “Learning in times of lockdown: how Covid-19 is affecting education and food security in India”. In: *Food security* 12.4 (2020), pp. 793–796.

[24] Seth Flaxman et al. “Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe”. In: *Nature* (June 2020).

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

[26] Tom Britton and David Lindenstrand. “Epidemic modelling: aspects where stochasticity matters”. In: *Mathematical Biosciences* 222.2 (2009), pp. 109–116.

[27] Michael Barnett, Greg Buchak, and Constantine Yannelis. “Epidemic responses under uncertainty”. In: *Proceedings of the National Academy of Sciences* 120.2 (2023), e2208111120.

[28] Ruth McCabe et al. “Communicating uncertainty in epidemic models”. In: *Epidemics* 37 (2021), p. 100520.

[29] Johan Christiaan Bester. “Measles and measles vaccination: a review”. In: *JAMA Pediatrics* 170.12 (2016), pp. 1209–1215.

[30] Zhou Xing, Mangalakumari Jeyanathan, and Fiona Smaill. “New approaches to TB vaccination”. In: *Chest* 146.3 (2014), pp. 804–812.

[31] Edward T Ryan and Stephen B Calderwood. “Cholera vaccines”. In: *Clinical Infectious Diseases* 31.2 (2000), pp. 561–565.

[32] PAUL E M FINE and Jacqueline A Clarkson. “Individual versus public priorities in the determination of optimal vaccination policies”. In: *American Journal of Epidemiology* 124.6 (1986), pp. 1012–1020.

[33] Johannes Müller. “Optimal vaccination patterns in age-structured populations”. In: *SIAM Journal on Applied Mathematics* 59.1 (1998), pp. 222–241.

[34] Alberto Olivares and Ernesto Staffetti. “Optimal control-based vaccination and testing strategies for COVID-19”. In: *Computer Methods and Programs in Biomedicine* 211 (2021), p. 106411.

[35] Sam Moore et al. “Modelling optimal vaccination strategy for SARS-CoV-2 in the UK”. In: *PLoS Computational Biology* 17.5 (2021), e1008849.

[36] Mario Coccia. “Optimal levels of vaccination to reduce COVID-19 infected individuals and deaths: A global analysis”. In: *Environmental research* 204 (2022), p. 112314.

[37] Jeroen Luyten and Philippe Beutels. “The social value of vaccination programs: beyond cost-effectiveness”. In: *Health Affairs* 35.2 (2016), pp. 212–218.

[38] Sachiko Ozawa et al. “Estimated economic impact of vaccinations in 73 low-and middle-income countries, 2001–2020”. In: *Bulletin of the World Health Organization* 95.9 (2017), p. 629.

[39] Alicia Blair et al. “The end of the elimination strategy: decisive factors towards sustainable management of COVID-19 in New Zealand”. In: *Epidemiologia* 3.1 (2022), pp. 135–147.

[40] Patrick Berche. “Life and death of smallpox”. In: *La Presse Médicale* 51.3 (2022), p. 104117.

[41] Alejandra Bellatin et al. “Overcoming vaccine deployment challenges among the hardest to reach: lessons from polio elimination in India”. In: *BMJ Global Health* 6.4 (2021), e005125.

[42] John Furesz. “Difficulties in polio eradication”. In: *The Lancet* 357.9254 (2001), pp. 477–478.

[43] Radboud J Duintjer Tebbens et al. “Economic analysis of the global polio eradication initiative”. In: *Vaccine* 29.2 (2010), pp. 334–343.

[44] David Jorgensen et al. “The role of genetic sequencing and analysis in the polio eradication programme”. In: *Virus Evolution* 6.2 (2020), veaa040.

[45] Kathryn Glass and Belinda Barnes. “Eliminating infectious diseases of livestock: a metapopulation model of infection control”. In: *Theoretical Population Biology* 85 (2013), pp. 63–72.

[46] Shuo Jiang et al. “Mathematical models for devising the optimal Ebola virus disease eradication”. In: *Journal of Translational Medicine* 15.1 (2017), pp. 1–10.

[47] Matthew J Penn et al. *Phylo2Vec: a vector representation for binary trees*. 2023. arXiv: 2304.12693.

[48] Matthew J Penn et al. “Leaping through Tree Space: Continuous Phylogenetic Inference for Rooted and Unrooted Trees”. In: *Genome Biology and Evolution* 15.12 (Dec. 2023), evad213.

[49] Matthew J Penn et al. “Intrinsic randomness in epidemic modelling beyond statistical uncertainty”. In: *Communications Physics* 6.1 (2023), p. 146.

[50] Matthew J Penn and Christl A Donnelly. “Optimality of maximal-effort vaccination”. In: *Bulletin of Mathematical Biology* 85.8 (2023).

[51] Matthew J Penn and Christl A Donnelly. “Asymptotic analysis of optimal vaccination policies”. In: *Bulletin of Mathematical Biology* 85.3 (2023), p. 15.

[52] Hervé Philippe et al. “Phylogenomics”. In: *Annual Review of Ecology, Evolution, and Systematics* 36 (2005), pp. 541–562.

[53] Cheong Xin Chan and Mark A Ragan. “Next-generation phylogenomics”. In: *Biology direct* 8.1 (2013), pp. 1–6.

[54] Jacob L Steenwyk et al. “Incongruence in the phylogenomics era”. In: *Nature Reviews Genetics* 24.12 (2023), pp. 834–850.

[55] F James Rohlf. “Numbering binary trees with labeled terminal vertices”. In: *Bulleting of Mathematical Biology* 45.1 (1983), pp. 33–40.

[56] Vincent Lefort, Richard Desper, and Olivier Gascuel. “FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program”. In: *Molecular Biology and Evolution* 32.10 (2015), pp. 2798–2800.

[57] Michael Dunne et al. “Complex model calibration through emulation, a worked example for a stochastic epidemic model”. In: *Epidemics* 39 (2022), p. 100574.

[58] Tapiwa Ganyani, Christel Faes, and Niel Hens. “Simulation and analysis methods for stochastic compartmental epidemic models”. In: *Annual Review of Statistics and Its Application* 8 (2021), pp. 69–88.

[59] Toru Kitagawa and Guanyi Wang. “Who should get vaccinated? Individualized allocation of vaccines over SIR network”. In: *Journal of Econometrics* 232.1 (2023), pp. 109–131.

[60] Mikko S Pakkanen et al. “Unifying incidence and prevalence under a time-varying general branching process”. In: *Journal of Mathematical Biology* 87.2 (2023), p. 35.

[61] Matthew J Penn, Christl A Donnelly, and Samir Bhatt. “Continuous football player tracking from discrete broadcast data”. In: *arXiv preprint arXiv:2311.14642* (2023).

[62] Matthew J. Penn et al. “Sherlock — A flexible, low-resource tool for processing camera-trapping images”. In: *Methods in Ecology and Evolution* 15.1 (2024), pp. 91–102.

[63] Matthew J Penn and Christl A Donnelly. “Analysis of a double Poisson model for predicting football results in Euro 2020”. In: *PLoS One* 17.5 (2022), e0268511.

[64] Charles Semple, Mike Steel, et al. *Phylogenetics*. Vol. 24. Oxford University Press on Demand, 2003.

[65] Robert Lanfear, Hanna Kokko, and Adam Eyre-Walker. “Population size and the rate of evolution”. In: *Trends in Ecology & Evolution* 29.1 (2014), pp. 33–41.

[66] Massimo Ciccozzi et al. “The phylogenetic approach for viral infectious disease evolution and epidemiology: An updating review”. In: *Journal of Medical Virology* 91.10 (2019), pp. 1707–1724.

[67] Michael W Gaunt et al. “Phylogenetic relationships of flaviviruses correlate with their epidemiology, disease association and biogeography”. In: *Journal of General Virology* 82.8 (2001), pp. 1867–1876.

[68] James I Brooks and Paul A Sandstrom. “The power and pitfalls of HIV phylogenetics in public health”. In: *Canadian Journal of Public Health* 104 (2013), e348–e350.

[69] Nathan D Wolfe, Claire Panosian Dunavan, and Jared Diamond. “Origins of major human infectious diseases”. In: *Nature* 447.7142 (2007), pp. 279–283.

[70] Rui-Heng Xu et al. “Epidemiologic clues to SARS origin in China”. In: *Emerging Infectious Diseases* 10.6 (2004), p. 1030.

[71] Gabriele Neumann, Takeshi Noda, and Yoshihiro Kawaoka. “Emergence and pandemic potential of swine-origin H1N1 influenza virus”. In: *Nature* 459.7249 (2009), pp. 931–939.

[72] Ahmad Sharif-Yakan and Souha S Kanj. “Emergence of MERS-CoV in the Middle East: origins, transmission, treatment, and perspectives”. In: *PLoS Pathogens* 10.12 (2014), e1004457.

[73] Mark Woolhouse and Eleanor Gaunt. “Ecological origins of novel human pathogens”. In: *Critical reviews in microbiology* 33.4 (2007), pp. 231–242.

[74] BB Chomel. “Zoonoses”. In: *Reference Module in Biomedical Sciences* (2014).

[75] Elisa Visher et al. “The three Ts of virulence evolution during zoonotic emergence”. In: *Proceedings of the Royal Society B* 288.1956 (2021), p. 20210900.

[76] James M Hassell et al. “Urbanization and disease emergence: dynamics at the wildlife–livestock–human interface”. In: *Trends in Ecology & Evolution* 32.1 (2017), pp. 55–67.

[77] Amr El-Sayed and Mohamed Kamel. “Climatic changes and their role in emergence and re-emergence of diseases”. In: *Environmental Science and Pollution Research* 27 (2020), pp. 22336–22352.

[78] Nazmun Nahar et al. “A controlled trial to reduce the risk of human Nipah virus exposure in Bangladesh”. In: *Ecohealth* 14 (2017), pp. 501–517.

[79] Tony Kirby. “New variant of SARS-CoV-2 in UK causes surge of COVID-19”. In: *The Lancet Respiratory Medicine* 9.2 (2021), e20–e21.

[80] Owen Dyer. *Covid-19: South Africa’s surge in cases deepens alarm over omicron variant*. 2021.

[81] Ewen Callaway et al. “Delta coronavirus variant: scientists brace for impact”. In: *Nature* 595.7865 (2021), pp. 17–18.

[82] Yusha Araf et al. “Omicron variant of SARS-CoV-2: genomics, transmissibility, and responses to current COVID-19 vaccines”. In: *Journal of Medical Virology* 94.5 (2022), pp. 1825–1832.

[83] Erik Volz. “Fitness, growth and transmissibility of SARS-CoV-2 genetic variants”. In: *Nature Reviews Genetics* 24.10 (2023), pp. 724–734.

[84] Elizabeth Bast et al. “Increased risk of hospitalisation and death with the delta variant in the USA”. In: *The Lancet Infectious Diseases* 21.12 (2021), pp. 1629–1630.

[85] Deepa Vasireddy et al. “Review of COVID-19 variants and COVID-19 vaccine efficacy: what the clinician should know?” In: *Journal of Clinical Medicine Research* 13.6 (2021), p. 317.

[86] Salim S Abdool Karim and Quarraisha Abdool Karim. “Omicron SARS-CoV-2 variant: a new chapter in the COVID-19 pandemic”. In: *The Lancet* 398.10317 (2021), pp. 2126–2128.

[87] John P Moore. “Approaches for optimal use of different COVID-19 vaccines: issues of viral variants and vaccine efficacy”. In: *JAMA* 325.13 (2021), pp. 1251–1252.

[88] James Kyle Miller, Kimberly Elenberg, and Artur Dubrawski. “Forecasting emergence of COVID-19 variants of concern”. In: *PLoS One* 17.2 (2022), e0264198.

[89] Alaina C Pfenning-Butterworth, T Jonathan Davies, and Clayton E Cressler. “Identifying co-phylogenetic hotspots for zoonotic disease”. In: *Philosophical Transactions of the Royal Society B* 376.1837 (2021), p. 20200363.

[90] Nancy A Chow et al. “Tracing the evolutionary history and global expansion of Candida auris using population genomic analyses”. In: *MBio* 11.2 (2020), pp. 10–1128.

[91] Felix M Key et al. “Emergence of human-adapted Salmonella enterica is linked to the Neolithization process”. In: *Nature Ecology & Evolution* 4.3 (2020), pp. 324–333.

[92] L Jetten. *Characterising tree-based phylogenetic networks*. 2015.

[93] Damian Bogdanowicz and Krzysztof Giaro. “Matching split distance for unrooted binary phylogenetic trees”. In: *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 9.1 (2011), pp. 150–160.

[94] Weronika Buczynska and Jaroslaw A Wisniewski. “On the geometry of binary symmetric models of phylogenetic trees”. In: *Journal of the European Mathematical Society* 9.3 (2007), pp. 609–635.

[95] Yun S Song. “On the combinatorics of rooted binary phylogenetic trees”. In: *Annals of Combinatorics* 7.3 (2003), pp. 365–379.

[96] Pravech Ajawatanawong. “Molecular phylogenetics: Concepts for a newcomer”. In: *Network Biology* (2017), pp. 185–196.

[97] Mark Wilkinson et al. “Of clades and clans: terms for phylogenetic relationships in unrooted trees”. In: *Trends in Ecology & Evolution* 22.3 (2007), pp. 114–115.

[98] Chris Whidden and Frederick A Matsen. “Calculating the unrooted subtree prune-and-regraft distance”. In: *IEEE/ACM transactions on Computational Biology and bioinformatics* 16.3 (2018), pp. 898–911.

[99] Lavanya Kannan, Hua Li, and Arcady Mushegian. “A polynomial-time algorithm computing lower and upper bounds of the rooted subtree prune and regraft distance”. In: *Journal of Computational Biology* 18.5 (2011), pp. 743–757.

[100] David F Robinson and Leslie R Foulds. “Comparison of phylogenetic trees”. In: *Mathemetical Biosciences* 53.1-2 (1981), pp. 131–147.

[101] Nicholas D Pattengale, Eric J Gottlieb, and Bernard ME Moret. “Efficiently computing the Robinson-Foulds metric”. In: *Journal of Computational Biology* 14.6 (2007), pp. 724–735.

[102] Jucheol Moon and Oliver Eulenstein. “Cluster matching distance for rooted phylogenetic trees”. In: *Bioinformatics Research and Applications: 14th International Symposium, ISBRA 2018, Beijing, China, June 8-11, 2018, Proceedings 14*. Springer. 2018, pp. 321–332.

[103] Marine Murtskhvaladze et al. “Phylogeny of caucasian rock lizards (Darevskia) and other true lizards based on mitogenome analysis: Optimisation of the algorithms and gene selection”. In: *PLoS One* 15.6 (2020), e0233680.

[104] Mun Hua Tan et al. “MitoPhAST, a new automated mitogenomic phylogeny tool in the post-genomic era with a case study of 89 decapod mitogenomes including eight new freshwater crayfish mitogenomes”. In: *Molecular Phylogenetics and Evolution* 85 (2015), pp. 180–188.

[105] David Penny. “Criteria for optimising phylogenetic trees and the problem of determining the root of a tree”. In: *Journal of Molecular Evolution* 8 (1976), pp. 95–116.

[106] AJ Dissanayake et al. “Applied aspects of methods to infer phylogenetic relationships amongst fungi”. In: *Mycosphere* 11.1 (2020), pp. 2652–2676.

[107] Paschalia Kapli, Ziheng Yang, and Maximilian J Telford. “Phylogenetic tree building in the genomic age”. In: *Nature Reviews Genetics* 21.7 (2020), pp. 428–444.

[108] David L Swofford. *Phylogenetic analysis using parsimony*. 1998.

[109] Mike Steel and David Penny. “Parsimony, likelihood, and the role of models in molecular phylogenetics”. In: *Molecular Biology and Evolution* 17.6 (2000), pp. 839–850.

[110] Caro-Beth Stewart. “The powers and pitfalls of parsimony”. In: *Nature* 361.6413 (1993), pp. 603–607.

[111] Joseph Felsenstein. “Cases in which Parsimony or Compatibility Methods will be Positively Misleading”. In: *Systematic Biology* 27.4 (1978), pp. 401–410.

[112] Bastien Boussau et al. “Strepsiptera, phylogenomics and the long branch attraction problem”. In: *PLoS One* 9.10 (2014), e107709.

[113] Alexandros Stamatakis. “RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies”. In: *Bioinformatics* 30.9 (2014), pp. 1312–1313.

[114] Lam-Tung Nguyen et al. “IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies”. In: *Molecular Biology and Evolution* 32.1 (2015), pp. 268–274.

[115] Shamita Malik, Dolly Sharma, and Sunil Kumar Khatri. “Parallel implementation of D-Phylo algorithm for maximum likelihood clusters”. In: *IET Nanobiotechnology* 11.2 (2017), pp. 134–142.

[116] David Posada and Keith A Crandall. “Felsenstein phylogenetic likelihood”. In: *Journal of Molecular Evolution* 89.3 (2021), pp. 134–145.

[117] Joseph Felsenstein. “Maximum-likelihood estimation of evolutionary trees from continuous characters.” In: *American Journal of Human Genetics* 25.5 (1973), p. 471.

[118] Elena Rivas and Sean R Eddy. “Probabilistic phylogenetic inference with insertions and deletions”. In: *PLoS Computational Biology* 4.9 (2008), e1000172.

[119] Xiaofan Zhou et al. “Evaluating fast maximum likelihood-based phylogenetic programs using empirical phylogenomic data sets”. In: *Molecular Biology and Evolution* 35.2 (2018), pp. 486–503.

[120] Ziheng Yang and Bruce Rannala. “Molecular phylogenetics: principles and practice”. In: *Nature Reviews Genetics* 13.5 (2012), pp. 303–314.

[121] Simon Whelan and David A Morrison. “Inferring trees”. In: *Bioinformatics: Volume I: Data, Sequence Analysis, and Evolution* (2017), pp. 349–377.

[122] Jack Sullivan. “Maximum-likelihood methods for phylogeny estimation”. In: *Methods in enzymology*. Vol. 395. Elsevier, 2005, pp. 757–779.

[123] Stéphane Guindon et al. “Estimating maximum likelihood phylogenies with PhyML”. In: *Bioinformatics for DNA sequence analysis* (2009), pp. 113–137.

[124] Alexey M Kozlov et al. “RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference”. In: *Bioinformatics* 35.21 (2019), pp. 4453–4455.

[125] Lena Collienne and Alex Gavryushkin. “Computing nearest neighbour interchange distances between ranked phylogenetic trees”. In: *Journal of Mathematical Biology* 82.1-2 (Jan. 2021), p. 8.

[126] Steven Kelk and Simone Linz. “A tight kernel for computing the tree bisection and reconnection distance between two phylogenetic trees”. In: *SIAM Journal on Discrete Mathematics* 33.3 (2019), pp. 1556–1574.

[127] Alexandros Stamatakis, Thomas Ludwig, and Harald Meier. “RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees”. In: *Bioinformatics* 21.4 (2005), pp. 456–463.

[128] Craig A Stewart et al. “Parallel implementation and performance of fastDNAml: a program for maximum likelihood phylogenetic inference”. In: *Proceedings of the 2001 ACM/IEEE conference on Supercomputing*. 2001, pp. 20–20.

[129] Alexey M Kozlov et al. “RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference”. In: *Bioinformatics* 35.21 (2019), pp. 4453–4455.

[130] Mark P Simmons and John Kessenich. “Divergence and support among slightly suboptimal likelihood gene trees”. In: *Cladistics* 36.3 (2020), pp. 322–340.

[131] Xing-Xing Shen et al. “An investigation of irreproducibility in maximum likelihood phylogenetic inference”. In: *Nature communications* 11.1 (2020), p. 6096.

[132] Nicholas M Fountain-Jones et al. “Emerging phylogenetic structure of the SARS-CoV-2 pandemic”. In: *Virus Evolution* 6.2 (2020), veaa082.

[133] Oliver WM Rauhut and Diego Pol. “Probable basal allosauroid from the early Middle Jurassic Cañadón Asfalto Formation of Argentina highlights phylogenetic uncertainty in tetanuran theropod dinosaurs”. In: *Scientific Reports* 9.1 (2019), p. 18826.

[134] Fabrícia F Nascimento, Mario dos Reis, and Ziheng Yang. “A biologist’s guide to Bayesian phylogenetic analysis”. In: *Nature ecology & evolution* 1.10 (2017), pp. 1446–1454.

[135] Mark P Khurana et al. “The Limits of the Constant-rate Birth–Death Prior for Phylogenetic Tree Topology Inference”. In: *Systematic Biology* 73.1 (2024), pp. 235–246.

[136] Sebastian Höhna et al. “RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language”. In: *Systematic biology* 65.4 (2016), pp. 726–736.

[137] John P Huelsenbeck and Fredrik Ronquist. “Bayesian analysis of molecular evolution using MrBayes”. In: *Statistical methods in molecular evolution* (2005), pp. 183–226.

[138] Liangliang Wang, Alexandre Bouchard-Côté, and Arnaud Doucet. “Bayesian phylogenetic inference using a combinatorial sequential Monte Carlo method”. In: *Journal of the American Statistical Association* 110.512 (2015), pp. 1362–1374.

[139] Chris Venditti, Andrew Meade, and Mark Pagel. “Phylogenies reveal new interpretation of speciation and the Red Queen”. In: *Nature* 463.7279 (2010), pp. 349–352.

[140] Bryan Kolaczkowski and Joseph W Thornton. “Effects of branch length uncertainty on Bayesian posterior probabilities for phylogenetic hypotheses”. In: *Molecular biology and evolution* 24.9 (2007), pp. 2108–2118.

[141] Stefan Ekman and Rakel Blaalid. “The devil in the details: interactions between the branch-length prior and likelihood model affect node support and branch lengths in the phylogeny of the Psoraceae”. In: *Systematic Biology* 60.4 (2011), pp. 541–561.

[142] John P Huelsenbeck and Fredrik Ronquist. “MRBAYES: Bayesian inference of phylogenetic trees”. In: *Bioinformatics* 17.8 (2001), pp. 754–755.

[143] Alexei J Drummond and Andrew Rambaut. “BEAST: Bayesian evolutionary analysis by sampling trees”. In: *BMC Evolutionary Biology* 7 (2007), p. 214.

[144] Sandra Álvarez-Carretero and Mario dos Reis. “Bayesian phylogenomic dating”. In: *The Molecular Evolutionary Clock: Theory and Practice* (2020), pp. 221–249.

[145] Luiza Guimarães Fabreti and Sebastian Höhna. “Nucleotide substitution model selection is not necessary for Bayesian inference of phylogeny with well-behaved priors”. In: *Systematic Biology* (2023), syad041.

[146] Xuhua Xia and Xuhua Xia. “Distance-based phylogenetic methods”. In: *Bioinformatics and the Cell: Modern Computational Approaches in Genomics, Proteomics and Transcriptomics* (2018), pp. 343–379.

[147] Richard Desper and Olivier Gascuel. “Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle”. In: *Algorithms in Bioinformatics: Second International Workshop, WABI 2002 Rome, Italy, September 17–21, 2002 Proceedings 2*. Springer. 2002, pp. 357–374.

[148] Frédéric Delsuc, Henner Brinkmann, and Hervé Philippe. “Phylogenomics and the reconstruction of the tree of life”. In: *Nature Reviews Genetics* 6.5 (2005), pp. 361–375.

[149] Daniele Catanzaro et al. “A tutorial on the balanced minimum evolution problem”. In: *European Journal of Operational Research* 300.1 (2022), pp. 1–19.

[150] Daniele Catanzaro et al. “The Balanced Minimum Evolution Problem”. In: *INFORMS Journal of Computing* 24.2 (2012), pp. 276–294.

[151] Morgan N Price, Paramvir S Dehal, and Adam P Arkin. “FastTree: computing large minimum evolution trees with profiles instead of a distance matrix”. In: *Molecular Biology and Evolution* 26.7 (2009), pp. 1641–1650.

[152] Mary K Kuhner and Joseph Felsenstein. “A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates.” In: *Molecular Biology and Evolution* 11.3 (1994), pp. 459–468.

[153] Sudhir Kumar and Sudhindra R Gadagkar. “Efficiency of the neighbor-joining method in reconstructing deep and shallow evolutionary relationships in large phylogenies”. In: *Journal of Molecular Evolution* 51.6 (2000), p. 544.

[154] Olivier Gascuel and Mike Steel. “Neighbor-joining revealed”. In: *Molecular Biology and Evolution* 23.11 (2006), pp. 1997–2000.

[155] Tom van der Valk et al. “Million-year-old DNA sheds light on the genomic history of mammoths”. In: *Nature* 591.7849 (2021), pp. 265–269.

[156] Mick Roberts et al. “Nine challenges for deterministic epidemic models”. In: *Epidemics* 10 (2015), pp. 49–53.

[157] Valerie Isham. “Assessing the variability of stochastic epidemics”. In: *Mathematical Biosciences* 107.2 (1991), pp. 209–224.

[158] Linda JS Allen. “An introduction to stochastic epidemic models”. In: *Mathematical Epidemiology*. Springer, 2008, pp. 81–130.

[159] George H Weiss and Menachem Dishon. “On the asymptotic behavior of the stochastic and deterministic models of an epidemic”. In: *Mathematical Biosciences* 11.3-4 (1971), pp. 261–265.

[160] David Alonso, Alan J McKane, and Mercedes Pascual. “Stochastic amplification in epidemics”. In: *Journal of the Royal Society Interface* 4.14 (2007), pp. 575–582.

[161] Peter Czuppon et al. “The stochastic dynamics of early epidemics: probability of establishment, initial growth rate, and infection cluster size at first detection”. In: *Journal of the Royal Society Interface* 18.184 (2021), p. 20210575.

[162] Matias Valdenegro-Toro and Daniel Saromo Mori. “A deeper look into aleatoric and epistemic uncertainty disentanglement”. In: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE. 2022, pp. 1508–1516.

[163] Robert Evans. “SAGE advice and political decision-making:‘Following the science’in times of epistemic uncertainty”. In: *Social Studies of Science* 52.1 (2022), pp. 53–78.

[164] Paulo JS Silva et al. “Optimized delay of the second COVID-19 vaccine dose reduces ICU admissions”. In: *Proceedings of the National Academy of Sciences* 118.35 (2021), e2104640118.

[165] David J Spiegelhalter. *Understanding uncertainty*. 2008.

[166] Luis F Gordillo et al. “Bimodal epidemic size distributions for near-critical SIR with vaccination”. In: *Bulletin of Mathematical Biology* 70 (2008), pp. 589–602.

[167] Norman TJ Bailey. “The total size of a general stochastic epidemic”. In: *Biometrika* (1953), pp. 177–185.

[168] Robin N Thompson, Christopher A Gilligan, and Nik J Cunniffe. “Will an outbreak exceed available resources for control? Estimating the risk from invading pathogens using practical definitions of a severe epidemic”. In: *Journal of the Royal Society Interface* 17.172 (2020), p. 20200690.

[169] Jonathan Dushoff and Sang Woo Park. “Speed and strength of an epidemic intervention”. In: *Proceedings of the Royal Society B* 288.1947 (2021), p. 20201556.

[170] Tamer Oraby et al. “Modeling the effect of lockdown timing as a COVID-19 control measure in countries with differing social contacts”. In: *Scientific reports* 11.1 (2021), p. 3354.

[171] Nicholas G Davies et al. “Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study”. In: *The Lancet Public Health* 5.7 (2020), e375–e385.

[172] Anna Maria Gambaro et al. “ICU capacity expansion under uncertainty in the early stages of a pandemic”. In: *Production and Operations Management* (2023).

[173] Meksianis Z Ndii and Asep K Supriatna. “Stochastic mathematical models in epidemiology”. In: *Information* 20 (2017), pp. 6185–6196.

[174] Chunyan Ji and Daqing Jiang. “Threshold behaviour of a stochastic SIR model”. In: *Applied Mathematical Modelling* 38.21-22 (2014), pp. 5067–5079.

[175] Elisabetta Tornatore, Stefania Maria Buccellato, and Pasquale Vetro. “Stability of a stochastic SIR system”. In: *Physica A: Statistical Mechanics and its Applications* 354 (2005), pp. 111–126.

[176] Xianghua Zhang and Ke Wang. “Stochastic SIR model with jumps”. In: *Applied Mathematics Letters* 26.8 (2013), pp. 867–874.

[177] Linda JS Allen. *An introduction to stochastic processes with applications to biology*. CRC press, 2010.

[178] Jonathan A Chávez Casillas. “A Stochastic Model for the Early Stages of Highly Contagious Epidemics by using a State-Dependent Point Process”. In: *arXiv preprint arXiv:2209.08612* (2022).

[179] Andreas Eilersen and Kim Sneppen. “SARS-CoV-2 superspreading in cities vs the countryside”. In: *Apmis* 129.7 (2021), pp. 401–407.

[180] Masao Fukui and Chishio Furukawa. “Power laws in superspreading events: Evidence from coronavirus outbreaks and implications for SIR models”. In: *MedRxiv* (2020), pp. 2020–06.

[181] Yu Liu et al. “Characterizing super-spreading in microblog: An epidemic-based information propagation model”. In: *Physica A: Statistical Mechanics and its Applications* 463 (2016), pp. 202–218.

[182] Elizabeth Hunter, Brian Mac Namee, and John D Kelleher. "A taxonomy for agent-based models in human infectious disease epidemiology". In: *Journal of Artificial Societies and Social Simulation* 20.3 (2017).

[183] Steven F Railsback and Volker Grimm. *Agent-based and individual-based modeling: a practical introduction*. Princeton University Press, 2019.

[184] Cliff C Kerr et al. "Covasim: an agent-based model of COVID-19 dynamics and interventions". In: *PLoS Computational Biology* 17.7 (2021), e1009149.

[185] Navid Mahdizadeh Gharakhanlou and Navid Hooshangi. "Spatio-temporal simulation of the novel coronavirus (COVID-19) outbreak using the agent-based modeling approach (case study: Urmia, Iran)". In: *Informatics in Medicine Unlocked* 20 (2020), p. 100403.

[186] Erik Cuevas. "An agent-based model to evaluate the COVID-19 transmission risks in facilities". In: *Computers in Biology and Medicine* 121 (2020), p. 103827.

[187] Aniruddha Adiga et al. "Mathematical models for covid-19 pandemic: a comparative analysis". In: *Journal of the Indian Institute of Science* 100.4 (2020), pp. 793–807.

[188] Nicolas Hoertel et al. "A stochastic agent-based model of the SARS-CoV-2 epidemic in France". In: *Nature Medicine* 26.9 (2020), pp. 1417–1421.

[189] Frank Ball and Peter Donnelly. "Strong approximations for epidemic models". In: *Stochastic processes and their applications* 55.1 (1995), pp. 1–21.

[190] Alastair Jamieson-Lane and Bernd Blasius. "Calculation of epidemic arrival time distributions using branching processes". In: *Physical Review E* 102.4 (2020), p. 042301.

[191] Swapnil Mishra et al. "On the derivation of the renewal equation from an age-dependent branching process: an epidemic modelling perspective". In: (June 2020). arXiv: 2006.16487 [q-bio.PE].

[192] James C Frauenthal. *Mathematical modeling in epidemiology*. Springer Science & Business Media, 2012.

[193] Maroussia Slavtchova-Bojkova. "Branching processes modelling for coronavirus (COVID'19) pandemic". In: *13th International Conference on Information Systems and Grid Technologies, ISGT*. Vol. 2020. 2020, p. 2656.

[194] Theodore Edward Harris et al. *The theory of branching processes*. Vol. 6. Springer Berlin, 1963.

[195] M González, R Martínez, and M Slavtchova-Bojkova. "Stochastic monotonicity and continuity properties of the extinction time of Bellman-Harris branching processes: an application to epidemic modelling". In: *Journal of Applied Probability* 47.1 (2010), pp. 58–71.

[196] Giacomo Plazzotta and Caroline Colijn. "Crump-Mode-Jagers branching processes in disease outbreaks". In: *9th European Conference on Mathematical and Theoretical Biology*. 2014.

[197] Marc Lipsitch et al. "Depletion-of-susceptibles bias in influenza vaccine waning studies: how to ensure robust results". In: *Epidemiology & Infection* 147 (2019), e306.

[198] Jérôme Levesque, David W Maybury, and RHA David Shaw. "A model of COVID-19 propagation based on a gamma subordinated negative binomial branching process". In: *Journal of Theoretical Biology* 512 (2021), p. 110536.

[199] Christine Jacob. "Branching processes: their role in epidemiology". In: *International journal of environmental research and public health* 7.3 (2010), pp. 1186–1204.

[200] Chaitanya Shivaramaiah et al. "Coccidiosis: recent advancements in the immunobiology of Eimeria species, preventive measures, and the importance of vaccination as a control tool against these Apicomplexan parasites". In: *Veterinary Medicine: Research and Reports* (2014), pp. 23–34.

[201] Gordon Ada et al. "The importance of vaccination". In: *Frontiers in Biosciences* 12 (2007), pp. 1278–90.

[202] James A Roth. "Veterinary vaccines and their importance to animal health and public health". In: *Procedia in Vaccinology* 5 (2011), pp. 127–136.

[203] Helen Bedford. "Measles and the importance of maintaining vaccination levels". In: *Epidemiology* 14.1 (2004), pp. 153–168.

[204] Xavier Bonal and Sheila González. "The impact of lockdown on the learning gap: family and school divisions in times of crisis". In: *International Review of Education* 66.5-6 (2020), pp. 635–655.

[205] Nnamdi Nkire et al. "COVID-19 pandemic: demographic predictors of self-isolation or self-quarantine and impact of isolation and quarantine on perceived stress, anxiety, and depression". In: *Frontiers in Psychiatry* 12 (2021), p. 553468.

[206] Ariana Remmel et al. "COVID vaccines and safety: what the research says". In: *Nature* 590.7847 (2021), pp. 538–540.

[207] Ingrid Torjesen. *Covid-19 vaccine shortages: what is the cause and what are the implications?* 2021.

[208] Carlo Delfin S Estadilla et al. "Impact of vaccine supplies and delays on optimal control of the COVID-19 pandemic: mapping interventions for the Philippines". In: *Infectious Diseases of Poverty* 10.04 (2021), pp. 46–59.

[209] Marcia C Castro and Burton Singer. "Prioritizing COVID-19 vaccination by age". In: *Proceedings of the National Academy of Sciences* 118.15 (2021).

[210] Chandrakant Lahariya. "A brief history of vaccines & vaccination in India". In: *The Indian Journal of Medical Research* 139.4 (2014), p. 491.

[211] Saki Takahashi et al. "The geography of measles vaccination in the African Great Lakes region". In: *Nature Communications* 8.1 (2017), p. 15585.

[212] Kate M Bubar et al. "Model-informed COVID-19 vaccine prioritization strategies by age and serostatus". In: *Science* 371.6352 (2021), pp. 916–921.

[213] Chris Baraniuk. "Covid-19: How the UK vaccine rollout delivered success, so far". In: *British Medical Journal* 372 (2021).

[214] Kiesha Prem, Alex R Cook, and Mark Jit. "Projecting social contact matrices in 152 countries using contact surveys and demographic data". In: *PLoS Computational Biology* 13.9 (2017), e1005697.

[215] Fabrizio Natale et al. "COVID-19 cases and case fatality rate by age". In: *European Commission* 52.2 (2020), pp. 154–164.

[216] Meagan C Fitzpatrick and Alison P Galvani. "Optimizing age-specific vaccination". In: *Science* 371.6532 (2021), pp. 890–891.

[217] Niels G Becker and Dianna N Starczak. "Optimal vaccination strategies for a community of households". In: *Mathematical Biosciences* 139.2 (1997), pp. 117–132.

[218] Ana Babus, Sanmay Das, and SangMok Lee. "The optimal allocation of COVID-19 vaccines". In: *Economics Letters* 224 (2023), p. 111008.

[219] Somayeh Momenyan and Mahmoud Torabi. "Modeling the spatio-temporal spread of COVID-19 cases, recoveries and deaths and effects of partial and full vaccination coverage in Canada". In: *Scientific Reports* 12.1 (2022), p. 17817.

[220] Ze-yu Zhao et al. "The optimal vaccination strategy to control COVID-19: a modeling study in Wuhan City, China". In: *Infectious Diseases of Poverty* 10.06 (2021), pp. 48–73.

[221] Clara Bonanad et al. “The effect of age on mortality in patients with COVID-19: a meta-analysis with 611,583 subjects”. In: *Journal of the American Medical Directors Association* 21.7 (2020), pp. 915–918.

[222] Liza Coyer et al. “Hospitalisation rates differed by city district and ethnicity during the first wave of COVID-19 in Amsterdam, the Netherlands”. In: *BMC Public Health* 21 (2021), pp. 1–9.

[223] Diane Godeau et al. “Return-to-work, disabilities and occupational health in the age of COVID-19”. In: *Scandinavian journal of work, environment & health* 47.5 (2021), p. 408.

[224] Julia R Gog et al. “Vaccine escape in a heterogeneous population: insights for SARS-CoV-2 from a simple model”. In: *Royal Society Open Science* 8.7 (2021), p. 210530.

[225] Frank G Ball and Owen D Lyne. “Optimal vaccination policies for stochastic epidemics among a population of households”. In: *Mathematical Biosciences* 177 (2002), pp. 333–354.

[226] Victor M Preciado et al. “Optimal vaccine allocation to control epidemic outbreaks in arbitrary networks”. In: *52nd IEEE conference on decision and control*. IEEE. 2013, pp. 7486–7491.

[227] Fudong Ge and YangQuan Chen. “Optimal vaccination and treatment policies for regional approximate controllability of the time-fractional reaction–diffusion SIR epidemic systems”. In: *ISA Transactions* 115 (2021), pp. 143–152.

[228] Omar Zakary, Mostafa Rachik, and Ilias Elmouki. “On the analysis of a multi-regions discrete SIR epidemic model: an optimal control approach”. In: *International Journal of Dynamics and Control* 5 (2017), pp. 917–930.

[229] Farrukh Saleem et al. “Machine learning, deep learning, and mathematical models to analyze forecasting and epidemiology of COVID-19: A systematic literature review”. In: *International journal of environmental research and public health* 19.9 (2022), p. 5099.

[230] Timothy L Wiemken and Robert R Kelley. “Machine learning in epidemiology and health outcomes research”. In: *Annual Review of Public Health* 41.1 (2020), pp. 21–36.

[231] Youssoufa Mohamadou, Aminou Halidou, and Pascalin Tiam Kapen. “A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of COVID-19”. In: *Applied Intelligence* 50.11 (2020), pp. 3913–3925.

[232] Laetitia Laguzet and Gabriel Turinici. “Global optimal vaccination in the SIR model: properties of the value function and application to cost-effectiveness analysis”. In: *Mathematical Biosciences* 263 (2015), pp. 180–197.

[233] Daron Acemoglu et al. “Optimal targeted lockdowns in a multigroup SIR model”. In: *American Economic Review: Insights* 3.4 (2021), pp. 487–502.

[234] Elsa Hansen and Troy Day. “Optimal control of epidemics with limited resources”. In: *Journal of Mathematical Biology* 62.3 (2011), pp. 423–451.

[235] Gul Zaman, Yong Han Kang, and Il Hyo Jung. “Stability analysis and optimal vaccination of an SIR epidemic model”. In: *BioSystems* 93.3 (2008), pp. 240–249.

[236] Tapan Kumar Kar and Ashim Batabyal. “Stability analysis and optimal control of an SIR epidemic model with vaccination”. In: *Biosystems* 104.2-3 (2011), pp. 127–135.

[237] Lu Tang et al. “A review of multi-compartment infectious disease models”. In: *International Statistical Review* 88.2 (2020), pp. 462–513.

[238] Shaobo He, Yuexi Peng, and Kehui Sun. “SEIR modeling of the COVID-19 and its dynamics”. In: *Nonlinear Dynamics* 101 (2020), pp. 1667–1680.

[239] Suwardi Annas et al. “Stability analysis and numerical simulation of SEIR model for pandemic COVID-19 spread in Indonesia”. In: *Chaos, solitons & fractals* 139 (2020), p. 110072.

[240] Md Haider Ali Biswas, Luís Tiago Paiva, M Do Rosario de Pinho, et al. “A SEIR model for control of infectious diseases with constraints”. In: *Mathematical Biosciences and Engineering* 11.4 (2014), pp. 761–784.

[241] Maurice Görtz and Joachim Krug. “Nonlinear dynamics of an epidemic compartment model with asymptomatic infections and mitigation”. In: *Journal of Physics A: Mathematical and Theoretical* 55.41 (2022), p. 414005.

[242] M Soledad Aronna, Roberto Guglielmi, and Lucas Machado Moschen. “A model for COVID-19 with isolation, quarantine and testing as control measures”. In: *Epidemics* 34 (2021), p. 100437.

[243] Xinwei Wang et al. “Optimal vaccination strategy of a constrained time-varying SEIR epidemic model”. In: *Communications in Nonlinear Science and Numerical Simulation* 67 (2019), pp. 37–48.

[244] Paul Van den Driessche and James Watmough. “Further notes on the basic reproduction number”. In: *Mathematical Epidemiology* (2008), pp. 159–178.

[245] Andrew N Hill and Ira M Longini Jr. “The critical vaccination fraction for heterogeneous epidemic models”. In: *Mathematical Biosciences* 181.1 (2003), pp. 85–106.

[246] Isabelle J Rao and Margaret L Brandeau. "Optimal allocation of limited vaccine to control an infectious disease: Simple analytical conditions". In: *Mathematical Biosciences* 337 (2021), p. 108621.

[247] Wen Cao et al. "Optimizing Spatio-Temporal Allocation of the COVID-19 Vaccine Under Different Epidemiological Landscapes". In: *Frontiers in Public Health* 10 (2022), p. 921855.

[248] Mostak Ahmed, Md Abdullah Bin Masud, and Md Manirul Alam Sarker. "Bifurcation analysis and optimal control of discrete SIR model for COVID-19". In: *Chaos, Solitons & Fractals* 174 (2023), p. 113899.

[249] Heting Zhang et al. "Optimal control strategies for a two-group epidemic model with vaccination-resource constraints". In: *Applied Mathematics and Computation* 371 (2020), p. 124956.

[250] Kuan Yang et al. "Optimal vaccination policy and cost analysis for epidemic control in resource-limited settings". In: *Kybernetes* 44.3 (2015), pp. 475–486.

[251] Luca Bolzoni et al. "Optimal control of epidemic size and duration with limited resources". In: *Mathematical Biosciences* 315 (2019), p. 108232.

[252] Lukman Hakim. "A Pontryagin principle and optimal control of spreading COVID-19 with vaccination and quarantine subtype". In: *Commun. Math. Biol. Neurosci.* 2023 (2023), Article–ID.

[253] Petter Ögren and Clyde F Martin. "Vaccination strategies for epidemics in highly mobile populations". In: *Applied Mathematics and Computation* 127.2-3 (2002), pp. 261–276.

[254] Richard E Kopp. "Pontryagin maximum principle". In: *Mathematics in Science and Engineering*. Vol. 5. Elsevier, 1962, pp. 255–279.

[255] Kaihui Liu and Yijun Lou. "Optimizing COVID-19 vaccination programs during vaccine shortages". In: *Infectious Disease Modelling* 7.1 (2022), pp. 286–298.

[256] Phuc Le and Michael B Rothberg. "Determining the optimal vaccination schedule for herpes zoster: a cost-effectiveness analysis". In: *Journal of General Internal Medicine* 32 (2017), pp. 159–167.

[257] Bernard S Bloom et al. "A reappraisal of hepatitis B virus vaccination strategies using cost-effectiveness analysis". In: *Annals of Internal Medicine* 118.4 (1993), pp. 298–306.

[258] Mark A Miller et al. "Prioritization of influenza pandemic vaccination to minimize years of life lost". In: *The Journal of infectious diseases* 198.3 (2008), pp. 305–311.

[259] Carlos Garriga, Rody Manuelli, and Siddhartha Sanghi. "Optimal management of an epidemic: Lockdown, vaccine and value of life". In: *Journal of Economic Dynamics and Control* 140 (2022), p. 104351.

[260] Francesco Parino et al. "A model predictive control approach to optimally devise a two-dose vaccination rollout: A case study on COVID-19 in Italy". In: *International journal of robust and nonlinear control* 33.9 (2023), pp. 4808–4823.

[261] Shenmin Zhang. "An Optimal Vaccine Allocation Model Considering Vaccine Hesitancy and Efficacy Rates Among Populations". In: *IEEE Access* 11 (2023), pp. 27693–27701.

[262] Satyaki Roy, Ronojoy Dutta, and Preetam Ghosh. "Optimal time-varying vaccine allocation amid pandemics with uncertain immunity ratios". In: *IEEE Access* 9 (2021), pp. 15110–15121.

[263] Hélène Morlon, Matthew D. Potts, and Joshua B. Plotkin. "Inferring the dynamics of diversification: A coalescent approach". In: *PLoS Biology* 8.9 (2010).

[264] Jonathan Rolland et al. "Using phylogenies in conservation: New perspectives". In: *Biology Letters* 8.5 (2011), pp. 692–694.

[265] Rolf J Ypma, W Marijn van Ballegooijen, and Jacco Wallinga. "Relating phylogenetic trees to transmission trees of infectious disease outbreaks". In: *Genetics* 195.3 (2013), pp. 1055–1062.

[266] Nuno R Faria et al. "Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil". In: *Science* (Apr. 2021).

[267] Gary Olsen. *Gary Olsen's interpretation of the "Newick's 8: 45" tree format standard*. https://phylipweb.github.io/phylip/newick_doc.html. 1990.

[268] Joseph Felsenstein. *Inferring phylogenies*. Vol. 2. Sunderland, MA: Sinauer Associates, 2004.

[269] Doron Rotem and Yaakov L Varol. "Generation of binary trees from ballot sequences". In: *Journal of the ACM (JACM)* 25.3 (1978), pp. 396–404.

[270] Andrzej Proskurowski. "On the generation of binary trees". In: *Journal of the ACM (JACM)* 27.1 (1980), pp. 1–2.

[271] Donald Knuth. *The Art of Computer Programming*. Vol. 1. Reading, MA: Addison-Wesley, 1973.

[272] Julia A Palacios et al. "Enumeration of binary trees compatible with a perfect phylogeny". In: *Journal of Mathematical Biology* 84.6 (2022), p. 54.

[273] Luigi L Cavalli-Sforza and Anthony WF Edwards. "Phylogenetic analysis. Models and estimation procedures". In: *American Journal of Human Genetics* 19.3 Pt 1 (1967), p. 233.

[274] Pengyu Liu. “A tree distinguishing polynomial”. In: *Discrete Applied Mathematics* 288 (2021), pp. 1–8.

[275] Pengyu Liu et al. “Analyzing Phylogenetic Trees with a Tree Lattice Coordinate System and a Graph Polynomial”. In: *Systematic Biology* 71.6 (2022), pp. 1378–1390.

[276] Jakub Voznica et al. “Deep learning from phylogenies to uncover the epidemiological dynamics of outbreaks”. In: *Nature Communications* 13.1 (2022), pp. 1–14.

[277] Persi W Diaconis and Susan P Holmes. “Matchings and phylogenetic trees”. In: *PNAS* 95.25 (1998), pp. 14600–14602.

[278] Jaehee Kim, Noah A Rosenberg, and Julia A Palacios. “Distance metrics for ranked evolutionary trees”. In: *PNAS* 117.46 (2020), pp. 28876–28886.

[279] Remco Bouckaert et al. “BEAST 2: a software platform for Bayesian evolutionary analysis”. In: *PLoS Computational Biology* 10.4 (2014), e1003537.

[280] Sebastien Roch. “A short proof that phylogenetic tree reconstruction by maximum likelihood is hard”. In: *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 3.1 (2006), pp. 92–94.

[281] Michael J Sanderson, Michelle M McMahon, and Mike Steel. “Terraces in phylogenetic tree space”. In: *Science* 333.6041 (2011), pp. 448–450.

[282] Ammon Thompson et al. “Deep learning and likelihood approaches for viral phylogeography converge on the same answers whether the inference model is right or wrong”. In: *Systematic Biolpgy* (Jan. 2024).

[283] Sophia Lambert, Jakub Voznica, and Hélène Morlon. “Deep Learning from Phylogenies for Diversification Analyses”. In: *Systematic Biology* (2023).

[284] Joseph Felsenstein. “The Number of Evolutionary Trees”. In: *Systematic Biology* 27.1 (1978), pp. 27–33.

[285] Charles R. Harris et al. “Array programming with NumPy”. In: *Nature* 585.7825 (Sept. 2020), pp. 357–362.

[286] Siu Kwan Lam, Antoine Pitrou, and Stanley Seibert. “Numba: A LLVM-based Python JIT compiler”. In: *Proc. LLVM-HPC*. 2015, pp. 1–6.

[287] Leonardo De Oliveira Martins, Diego Mallo, and David Posada. “A Bayesian supertree model for genome-wide species tree reconstruction”. In: *Systematic Biology* 65.3 (2016), pp. 397–416.

[288] Magnus Bordewich and Charles Semple. “On the Computational Complexity of the Rooted Subtree Prune and Regraft Distance”. In: *Annals of Combinatorics* 8.4 (2005), pp. 409–423.

[289] Katherine St. John. “The Shape of Phylogenetic Treespace”. In: *Systematic Biology* 66.1 (2017), e83–e94.

[290] Antonis Rokas et al. “Genome-scale approaches to resolving incongruence in molecular phylogenies”. In: *Nature* 425.6960 (2003), pp. 798–804.

[291] Klaus Peter Schliep. “phangorn: phylogenetic analysis in R”. In: *Bioinformatics* 27.4 (2011), pp. 592–593.

[292] Pavel Sagulenko, Vadim Puller, and Richard A Neher. “TreeTime: Maximum-likelihood phylodynamic analysis”. In: *Virus Evolution* 4.1 (2018), vex042.

[293] James R Garey et al. “Molecular evidence for Acanthocephala as a subtaxon of Rotifera”. In: *Journal of Molecular Evolution* 43 (1996), pp. 287–292.

[294] Mathieu Fourment and Aaron E Darling. “Evaluating probabilistic programming and fast variational Bayesian inference in phylogenetics”. In: *PeerJ* 7 (2019), e8272.

[295] Emmanuel Paradis and Klaus Schliep. “ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R”. In: *Bioinformatics* 35 (2019), pp. 526–528.

[296] Jaime Huerta-Cepas, François Serra, and Peer Bork. “ETE 3: reconstruction, analysis, and visualization of phylogenomic data”. In: *Molecular Biology and Evolution* 33.6 (2016), pp. 1635–1638.

[297] Joseph Felsenstein. “Statistical inference of phylogenies”. In: *Journal of the Royal Statistical Society Series A* 146.3 (1983), pp. 246–262.

[298] Bui Quang Minh et al. “IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era”. In: *Molecular Biology and Evolution* 37.5 (2020), pp. 1530–1534.

[299] Emmanuel Paradis, Julien Claude, and Korbinian Strimmer. “APE: analyses of phylogenetics and evolution in R language”. In: *Bioinformatics* 20.2 (2004), pp. 289–290.

[300] Daniel Money and Simon Whelan. “Characterizing the phylogenetic tree-search problem”. In: *Systematic Biology* 61.2 (2012), pp. 228–239.

[301] Tomas Flouri et al. “The phylogenetic likelihood library”. In: *Systematic Biology* 64.2 (2015), pp. 356–362.

[302] Daniel L Ayres et al. “BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics”. In: *Systematic Biology* 61.1 (2012), pp. 170–173.

[303] Jeremy G Sumner et al. “Is the general time-reversible model bad for molecular phylogenetics?” In: *Systematic Biology* 61.6 (2012), pp. 1069–1074.

[304] Stéphane Guindon et al. “New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0”. In: *Systematic Biology* 59.3 (2010), pp. 307–321.

[305] William H Piel et al. “TreeBASE v. 2: A Database of Phylogenetic Knowledge”. In: *e-BioSphere*. 2009.

[306] Brian C O’Meara. “Evolutionary inferences from phylogenies: a review of methods”. In: *Annual Reviews in Ecology, Evolution and Systematics* 43 (2012), pp. 267–285.

[307] Bryan T Grenfell et al. “Unifying the epidemiological and evolutionary dynamics of pathogens”. In: *Science* 303.5656 (2004), pp. 327–332.

[308] Ruth Mace and Clare J Holden. “A phylogenetic approach to cultural evolution”. In: *Trends Ecol. Evol.* 20.3 (2005), pp. 116–121.

[309] Mark Collard, Stephen J Shennan, and Jamshid J Tehrani. “Branching, blending, and the evolution of cultural similarities and differences among human populations”. In: *Evolution and Human Behaviour* 27.3 (2006), pp. 169–184.

[310] David A Morrison. “Are phylogenetic patterns the same in anthropology and biology?” In: *bioRxiv* (2014).

[311] Frédéric Lemoine et al. “COVID-Align: Accurate online alignment of hCoV-19 genomes using a profile HMM”. In: *Bioinformatics* 37.12 (2021), pp. 1761–1762.

[312] Áine O’Toole et al. “Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool”. In: *Virus Evolution* 7.2 (2021), veab064.

[313] Stephen W. Attwood et al. “Phylogenetic and phylodynamic approaches to understanding and combating the early SARS-CoV-pandemic”. In: *Nature Reviews Genetics* 23.9 (2022), pp. 547–562. DOI: 10.1038/s41576-022-00483-8.

[314] Theo Sanderson. “Taxonium, a web-based tool for exploring large phylogenetic trees”. In: *eLife* 11 (2022).

[315] Yatish Turakhia et al. “Pandemic-scale phylogenomics reveals the SARS-CoV-2 recombination landscape”. In: *Nature* 609.7929 (2022), pp. 994–997.

[316] Nicola De Maio et al. “Maximum likelihood pandemic-scale phylogenetics”. In: *Nature Genetics* (2023), pp. 1–7.

[317] Michael J Sanderson and H Bradley Shaffer. “Troubleshooting molecular phylogenetic analyses”. In: *Annu. Rev. Ecol. Syst.* 33.1 (2002), pp. 49–72.

[318] Ziheng Yang. *Computational Molecular Evolution*. OUP Oxford, 2006.

[319] Michael SY Lee and Alessandro Palci. “Morphological phylogenetics in the genomic age”. In: *Current Biology* 25.19 (2015), R922–R929.

[320] Hélène Morlon, Todd L Parsons, and Joshua B Plotkin. “Reconciling molecular phylogenies with the fossil record”. In: *Proceedings of the National Academy of Sciences U.S.A.* 108.39 (2011), pp. 16327–16332.

[321] Jozsef Arato and W Tecumseh Fitch. “Phylogenetic signal in the vocalizations of vocal learning and vocal non-learning birds”. In: *Philosophical Transactions of the Royal Society B* 376.1836 (2021), p. 20200241.

[322] Lindell Bromham and David Penny. “The modern molecular clock”. In: *Nature Reviews Genetics* 4.3 (2003), pp. 216–224.

[323] Mario Dos Reis, Philip CJ Donoghue, and Ziheng Yang. “Bayesian molecular clock dating of species divergences in the genomics era”. In: *Nature Reviews Genetics* 17.2 (2016), pp. 71–80.

[324] E Zuckerkandl. “Molecular disease, evolution, and genetic heterogeneity”. In: *Horiz. Biochem. Biophys.* (1962), pp. 189–225.

[325] Les R Foulds and Ronald L Graham. “The Steiner problem in phylogeny is NP-complete”. In: *Advances in Applied Mathematics* 3.1 (1982), pp. 43–49.

[326] William HE Day. “Computational complexity of inferring phylogenies from dissimilarity matrices”. In: *Bulletin of Mathematical Biology* 49.4 (1987), pp. 461–467.

[327] M. D. Hendy and David Penny. “Branch and bound algorithms to determine minimal evolutionary trees”. In: *Mathematical Biosciences* 59.2 (1982), pp. 277–290.

[328] James C Wilgenbusch and David Swofford. “Inferring evolutionary trees with PAUP”. In: *Current Protocols in Bioinformatics* 1 (2003), pp. 6.4.1–6.4.28.

[329] Ziheng Yang. “PAML 4: phylogenetic analysis by maximum likelihood”. In: *Molecular Biology and Evolution* 24.8 (2007), pp. 1586–1591.

[330] Hyun Jung Park, Seung-Jin Sul, and Tiffani L Williams. "Large-scale analysis of phylogenetic search behavior". In: *Advances in Experimental Medicine and Biology* 680 (2010), pp. 35–42.

[331] Alan de Queiroz and John Gatesy. "The supermatrix approach to systematics". In: *Trends in Ecology and Evolution* 22.1 (2007), pp. 34–41.

[332] Olga Chernomor, Arndt von Haeseler, and Bui Quang Minh. "Terrace Aware Data Structure for Phylogenomic Inference from Supermatrices". In: *Systematic Biology* 65.6 (Nov. 2016), pp. 997–1008.

[333] Vu Dinh et al. "Probabilistic Path Hamiltonian Monte Carlo". In: *ICML*. Vol. 70. PMLR, 2017, pp. 1009–1018.

[334] Louis J Billera, Susan P Holmes, and Karen Vogtmann. "Geometry of the Space of Phylogenetic Trees". In: *Advances in Applied Mathematics* 27.4 (2001), pp. 733–767.

[335] Hirotaka Matsumoto, Takahiro Mimori, and Tsukasa Fukunaga. "Novel metric for hyperbolic phylogenetic tree embeddings". In: *Biology Methods and Protocols* 6.1 (2021), bpab006.

[336] Benjamin Wilson. *Learning phylogenetic trees as hyperbolic point configurations*. 2021. arXiv: 2104.11430 [cs.LG].

[337] Matthew Macaulay, Aaron Darling, and Mathieu Fourment. "Fidelity of hyperbolic space for Bayesian phylogenetic inference". In: *PLoS Computational Biology* 19.4 (2023), e1011084.

[338] Takahiro Mimori and Michiaki Hamada. *GeoPhy: Differentiable Phylogenetic Inference via Geometric Gradients of Tree Topologies*. 2023. arXiv: 2307.03675 [cs.LG].

[339] Cheng Zhang and Frederick A Matsen IV. "Variational Bayesian phylogenetic inference". In: *ICLR*. 2018.

[340] Luca Nesterenko, Bastien Boussau, and Laurent Jacob. "Phyloformer: towards fast and accurate phylogeny estimation with self-attention networks". In: *bioRxiv* (2022), pp. 2022–06.

[341] Y Pauplin. "Direct calculation of a tree length using a distance matrix". In: *Journal of Molecular Evolution* 51.1 (2000), pp. 41–47.

[342] K K Kidd and L A Sgaramella-Zonta. "Phylogenetic analysis: concepts and methods". In: *American Journal of Human Genetics* 23.3 (1971), pp. 235–252.

[343] Andrey Rzhetsky and Masatoshi Nei. "A simple method for estimating and testing minimum-evolution trees". In: *Molecular Biology and Evolution* 9.5 (1992), pp. 945–967.

[344] Richard Desper and Olivier Gascuel. “Theoretical foundation of the balanced minimum evolution method of phylogenetic inference and its relationship to weighted least-squares tree fitting”. In: *Molecular Biology and Evolution* 21.3 (2004), pp. 587–598.

[345] Diederik P. Kingma and Jimmy Ba. “Adam: A Method for Stochastic Optimization”. In: *ICLR*. 2015.

[346] Ilya Loshchilov and Frank Hutter. “Decoupled Weight Decay Regularization”. In: *ICLR*. 2019.

[347] Noam Shazeer and Mitchell Stern. “Adafactor: Adaptive learning rates with sublinear memory cost”. In: *ICML*. PMLR. 2018, pp. 4596–4604.

[348] W M Fitch and E Margoliash. “Construction of phylogenetic trees”. In: *Science* 155.3760 (1967), pp. 279–284.

[349] Naruya Saitou and Masatoshi Nei. “The neighbor-joining method: a new method for reconstructing phylogenetic trees.” In: *Molecular Biology and Evolution* 4.4 (1987), pp. 406–425.

[350] Charles Semple and Mike Steel. “Cyclic permutations and evolutionary trees”. In: *Advances in Applied Mathematics* 32.4 (2004), pp. 669–680.

[351] K Atteson. “The Performance of Neighbor-Joining Methods of Phylogenetic Reconstruction”. In: *Algorithmica* 25.2 (1999), pp. 251–278.

[352] Radu Mihaescu, Dan Levy, and Lior Pachter. “Why neighbor-joining works”. In: *Algorithmica* 54.1 (2009), pp. 1–24.

[353] Radu Mihaescu and Lior Pachter. “Combinatorics of least-squares trees”. In: *Proceedings of the National Academy of Sciences U.S.A.* 105.36 (2008), pp. 13206–13211.

[354] Thu-Hien To et al. “Fast Dating Using Least-Squares Criteria and Algorithms”. In: *Systematic Biology* 65.1 (2016), pp. 82–97.

[355] H.-C. Chen and Y.-L. Wang. “An efficient algorithm for generating Prüfer codes from labelled trees”. In: *Theory of Computing Systems* 33 (2000), pp. 97–105.

[356] James Bradbury et al. *JAX: composable transformations of Python+NumPy programs*. Version 0.3.13. 2018. URL: http://github.com/google/jax.

[357] Iker Irisarri et al. “Phylotranscriptomic consolidation of the jawed vertebrate timetree”. In: *Nature Ecology and Evolution* 1.9 (2017), pp. 1370–1378.

[358] S Blair Hedges, Kirk D Moberg, and Linda R Maxson. “Tetrapod phylogeny inferred from 18S and 28S ribosomal RNA sequences and a review of the evidence for amniote relationships.” In: *Molecular Biology and Evolution* 7.6 (1990), pp. 607–633.

[359] Ziheng Yang and Anne D Yoder. "Comparison of likelihood and Bayesian methods for estimating divergence times using multiple gene loci and calibration points, with application to a radiation of cute-looking mouse lemur species". In: *Systematic Biology* 52.5 (2003), pp. 705–716.

[360] Daniel A Henk, Alex Weir, and Meredith Blackwell. "Laboulbeniopsis termitarius, an ectoparasite of termites newly recognized as a member of the Laboulbeniomycetes". In: *Mycologia* 95.4 (2003), pp. 561–564.

[361] Andrew VZ Brower. "Phylogenetic relationships among the Nymphalidae (Lepidoptera) inferred from partial sequences of the wingless gene". In: *Proceedings of the Royal Society Series B* 267.1449 (2000), pp. 1201–1211.

[362] Clemens Lakner et al. "Efficiency of Markov chain Monte Carlo tree proposals in Bayesian phylogenetics". In: *Systematic Biology* 57.1 (2008), pp. 86–103.

[363] Ning Zhang and Meredith Blackwell. "Molecular phylogeny of dogwood anthracnose fungus (Discula destructiva) and the Diaporthales". In: *Mycologia* 93.2 (2001), pp. 355–365.

[364] Anne D Yoder and Ziheng Yang. "Divergence dates for Malagasy lemurs estimated from multiple gene loci: geological and evolutionary context". In: *Molecular Ecology* 13.4 (2004), pp. 757–773.

[365] Amy Y Rossman et al. "Molecular studies of the Bionectriaceae using large subunit rDNA sequences". In: *Mycologia* 93.1 (2001), pp. 100–110.

[366] Amanda L Ingram and Jeff J Doyle. "Is Eragrostis (Poaceae) monophyletic? Insights from nuclear and plastid sequence data". In: *Systematic Botany* 29.3 (2004), pp. 545–552.

[367] Sung-Oui Suh and Meredith Blackwell. "Molecular phylogeny of the cleistothecial fungi placed in Cephalothecaceae and Pseudeurotiaceae". In: *Mycologia* 91.5 (1999), pp. 836–848.

[368] Scott Kroken and John W Taylor. "Phylogenetic species, reproductive mode, and specificity of the green alga Trebouxia forming lichens with the fungal genus Letharia". In: *Bryologist* (2000), pp. 645–660.

[369] Sen Song et al. "Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model". In: *Proceedings of the National Academy of Sciences U.S.A.* 109.37 (2012), pp. 14942–14947.

[370] Masami Hasegawa and Hirohisa Kishino. "Confidence limits of the maximum-likelihood estimate of the hominoid three from mitochondrial-DNA sequences". In: *Evolution* 43.3 (1989), pp. 672–677.

[371] Joseph Felsenstein. "An alternating least squares approach to inferring phylogenies from pairwise distances". In: *Systematic Biology* 46.1 (1997), pp. 101–111.

[372] Chris Whidden and Frederick A Matsen 4th. "Quantifying MCMC exploration of phylogenetic tree space". In: *Systematic Biology* 64.3 (2015), pp. 472–491.

[373] Si Quang Le and Olivier Gascuel. "An improved general amino acid replacement matrix". In: *Molecular Biology and Evolution* 25.7 (2008), pp. 1307–1320.

[374] Thomas H Jukes, Charles R Cantor, et al. "Evolution of protein molecules". In: *Mammalian Protein Metabolism* 3 (1969), pp. 21–132.

[375] Joseph Felsenstein. "Evolutionary trees from DNA sequences: a maximum likelihood approach". In: *Journal of Molecular Evolution* 17 (1981), pp. 368–376.

[376] Koichiro Tamura and Masatoshi Nei. "Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees." In: *Molecular Biology and Evolution* 10.3 (1993), pp. 512–526.

[377] Tijmen Tieleman and Geoffrey Hinton. *Lecture 6e - rmsprop: Divide the gradient by a running average of its recent magnitude*. Coursera: Neural networks for machine learning. Slides. 2012. URL: https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf.

[378] Olivier Gascuel. "BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data." In: *Molecular Biology and Evolution* 14.7 (1997), pp. 685–695.

[379] Igor Babuschkin et al. *The DeepMind JAX Ecosystem*. 2020. URL: http://github.com/deepmind.

[380] Laurent Gautier, Michał Krassowski, et al. *rpy2: Python interface to the R language*. 2021. URL: https://github.com/rpy2/rpy2.

[381] Aric A. Hagberg, Daniel A. Schult, and Pieter J. Swart. "Proc. SciPy". In: 2008, pp. 11–15.

[382] Chao Zhang et al. "ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees". In: *BMC Bioinformatics* 19.6 (2018), pp. 15–30.

[383] Michael Betancourt. "A conceptual introduction to Hamiltonian Monte Carlo". In: *arXiv preprint arXiv:1701.02434* (2017).

[384] John P Huelsenbeck, Jonathan P Bollback, and Amy M Levine. “Inferring the root of a phylogenetic tree”. In: *Systematic Biology* 51.1 (2002), pp. 32–43.

[385] Fernando Domingues Kümmel Tria, Giddy Landan, and Tal Dagan. “Phylogenetic rooting using minimal ancestor deviation”. In: *Nature Ecology and Evolution* 1 (2017), p. 193.

[386] Suha Naser-Khdour, Bui Quang Minh, and Robert Lanfear. “Assessing Confidence in Root Placement on Phylogenies: An Empirical Study Using Nonreversible Models for Mammals”. In: *Systematic Biology* 71.4 (2022), pp. 959–972.

[387] Mezzalina Vankan, Simon Y W Ho, and David A Duchêne. “Evolutionary Rate Variation among Lineages in Gene Trees has a Negative Impact on Species-Tree Inference”. In: *Systematic Biology* 71.2 (2022), pp. 490–500.

[388] Marc A Suchard et al. “Hierarchical phylogenetic models for analyzing multipartite sequence data”. In: *Systematic Biology* 52.5 (2003), pp. 649–664.

[389] Sebastián Duchêne et al. “Cross-validation to select Bayesian hierarchical models in phylogenetics”. In: *BMC Evolutionary Biology* 16.1 (2016), p. 115.

[390] Sudhir Kumar. “Embracing Green Computing in Molecular Phylogenetics”. In: *Molecular Biology and Evolution* 39.3 (2022), msac043.

[391] Paul Simion, Frédéric Delsuc, and Herve Philippe. “To What Extent Current Limits of Phylogenomics Can Be Overcome?” In: *Phylogenetics in the Genomic Era*. Ed. by Celine Scornavacca, Frédéric Delsuc, and Nicolas Galtier. No commercial publisher — Authors open access book, 2020, 2.1:1–2.1:34. URL: https://hal.science/hal-02535366.

[392] Yves Pauplin. “Direct calculation of a tree length using a distance matrix.” In: *Journal of Molecular Evolution* 51.1 (2000).

[393] Joseph Felsenstein. “Confidence limits on phylogenies: an approach using the bootstrap”. In: *Evolution* 39.4 (1985), pp. 783–791.

[394] B Efron, E Halloran, and S Holmes. “Bootstrap confidence levels for phylogenetic trees”. In: *Proceedings of the National Academy of Sciences U. S. A.* 93.23 (1996), pp. 13429–13434.

[395] B Rannala and Z Yang. “Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference”. In: *Journal of Molecular Evolution* 43.3 (1996), pp. 304–311.

[396] Diep Thi Hoang et al. “UFBoot2: Improving the Ultrafast Bootstrap Approximation”. In: *Molecular Biology and Evolution* 35.2 (2018), pp. 518–522.

[397] Susan Holmes. “Bootstrapping Phylogenetic Trees: Theory and Methods”. In: *Statistical Science* 18.2 (2003), pp. 241–255.

[398] P G Bissiri, C C Holmes, and S G Walker. “A general framework for updating belief distributions”. In: *Journal of the Royal Statistical Society Series B Statistical Methodology* 78.5 (2016), pp. 1103–1130.

[399] Nicola De Maio et al. “Mutation rates and selection on synonymous mutations in SARS-CoV-2”. In: *Genome biology and evolution* 13.5 (2021), evab087.

[400] Tanja Stadler. “Simulating trees with a fixed number of extant species”. In: *Systematic Biology* 60.5 (Oct. 2011), pp. 676–684.

[401] Simon Y W Ho, Sebastián Duchêne, and David Duchêne. “Simulating and detecting autocorrelation of molecular evolutionary rates among lineages”. In: *Molecular Ecology Resources* 15.4 (July 2015), pp. 688–696.

[402] Klaus Peter Schliep. “phangorn: phylogenetic analysis in R”. In: *Bioinformatics* 27.4 (2011), pp. 592–593.

[403] Seraina Klopfstein, Tim Massingham, and Nick Goldman. “More on the Best Evolutionary Rate for Phylogenetic Analysis”. In: *Systematic Biology* 66.5 (Sept. 2017), pp. 769–785.

[404] Stéphane Guindon et al. “New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0”. In: *Systematic Biology* 59.3 (2010), pp. 307–321.

[405] Chris Whidden et al. “Systematic Exploration of the High Likelihood Set of Phylogenetic Tree Topologies”. In: *Systematic Biology* 69.2 (2020), pp. 280–293.

[406] Shaohong Feng et al. “Dense sampling of bird diversity increases power of comparative genomics”. In: *Nature* 587.7833 (2020), pp. 252–257.

[407] Erich D Jarvis et al. “Whole-genome analyses resolve early branches in the tree of life of modern birds”. In: *Science* 346.6215 (Dec. 2014), pp. 1320–1331.

[408] Sushma Reddy et al. “Why do phylogenomic data sets yield conflicting trees? Data type influences the avian tree of life more than taxon sampling”. In: *Systematic Biology* 66.5 (2017), pp. 857–879.

[409] Peter Houde et al. “Phylogenetic Signal of Indels and the Neoavian Radiation”. In: *Diversity* 11.7 (2019).

[410] Stephen L Brusatte, Jingmai K O'Connor, and Erich D Jarvis. "The Origin and Diversification of Birds". In: *Current Biology* 25.19 (Oct. 2015), R888–98.

[411] W O Kermack and A G McKendrick. "A contribution to the mathematical theory of epidemics". In: *Proceedings of the Royal Society of London A Mathematical and Physical Sciences* 115.772 (Aug. 1927), pp. 700–721.

[412] Christophe Fraser. "Estimating individual and household reproduction numbers in an emerging epidemic". In: *PLoS One* (2007). ISSN: 19326203. DOI: 10.1371/journal.pone.0000758.

[413] David Champredon et al. "Two approaches to forecast Ebola synthetic epidemics". In: *Epidemics* 22 (Mar. 2018), pp. 36–42.

[414] Linda J S Allen. "A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis". In: *Infectious Disease Modelling* 2.2 (May 2017), pp. 128–142.

[415] Armen Der Kiureghian and Ove Ditlevsen. "Aleatory or epistemic? Does it matter?" In: *Structural Safety* 31.2 (Mar. 2009), pp. 105–112.

[416] Mario Castro et al. *The turning point and end of an expanding epidemic cannot be precisely forecast*. 2020.

[417] I Neri and L Gammaitoni. "Role of fluctuations in epidemic resurgence after a lockdown". In: *Scientific Reports* 11.1 (Mar. 2021), p. 6452.

[418] Samuel V Scarpino and Giovanni Petri. "On the predictability of infectious disease outbreaks". In: *Nature Communications* 10.1 (2019), p. 898.

[419] Giulia Pullano et al. "Underdetection of cases of COVID-19 in France threatens epidemic control". In: *Nature* 590.7844 (Feb. 2021), pp. 134–139.

[420] Felix Wong and James J Collins. *Evidence that coronavirus superspreading is fat-tailed*. 2020.

[421] Pasquale Cirillo and Nassim Nicholas Taleb. "Tail risk of contagious diseases". In: *Nature Physics* 16.6 (May 2020), pp. 606–613.

[422] Theodore Edward Harris et al. *The theory of branching processes*. Vol. 6. Springer Berlin, 1963.

[423] Kris V Parag and Christl A Donnelly. "Using information theory to optimise epidemic models for real-time prediction and estimation". In: *PLoS Computational Biology* 16.7 (July 2020), e1007990.

[424] Sam Abbott et al. *EpiNow2: Estimate Real-Time Case Counts and Time-Varying Epidemiological Parameters*. 2020. DOI: 10.5281/zenodo.3957489.

[425] Y Ogata. "On Lewis' simulation method for point processes". In: *IEEE Transactions on Information Theory* 27.1 (Jan. 1981), pp. 23–31.

[426] Anne Cori et al. "A new framework and software to estimate time-varying reproduction numbers during epidemics". In: *American Journal of Epidemiology* 178.9 (Nov. 2013), pp. 1505–1512.

[427] M E Woolhouse et al. "Heterogeneities in the transmission of infectious agents: implications for the design of control programs". In: *Proceedings of the National Academy of Sciences U. S. A.* 94.1 (Jan. 1997), pp. 338–342.

[428] J O Lloyd-Smith et al. "Superspreading and the effect of individual variation on disease emergence". In: *Nature* 438.7066 (Nov. 2005), pp. 355–359.

[429] Marc Lipsitch et al. "Transmission dynamics and control of severe acute respiratory syndrome". In: *Science* 300.5627 (June 2003), pp. 1966–1970.

[430] Lee Shiu Hung. "The SARS epidemic in Hong Kong: what lessons have we learned?" In: *Journal of the Royal Society Medicine* 96.8 (Aug. 2003), pp. 374–378.

[431] Andrew Barbour and Gesine Reinert. "Approximating the epidemic curve". In: *Electronic Journal of Probability* 18 (Jan. 2013), pp. 1–30.

[432] O Pybus, A Rambaut, COG-UK-Consortium, et al. "Preliminary analysis of SARS-CoV-2 importation & establishment of UK transmission lineages". In: *Virological. org* (2020).

[433] S Flaxman et al. "Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries". In: *Statistics and Computing* in press.1 (Sept. 2020), pp. 111–119.

[434] Nicholas F Brazeau et al. "Estimating the COVID-19 infection fatality ratio accounting for seroreversion using statistical modelling". In: *Communications in Medicine* 2 (May 2022), p. 54.

[435] Ying Liu et al. "The reproductive number of COVID-19 is higher compared to SARS coronavirus". In: *Journal of Travel Medicine* (2020). ISSN: 17088305. DOI: 10.1093/jtm/taaa021.

[436] Mrinank Sharma et al. "Understanding the effectiveness of government interventions against the resurgence of COVID-19 in Europe". In: *Nature Communications* 12.1 (Oct. 2021), p. 5820.

[437] Adam J Kucharski et al. "Early dynamics of transmission and control of {COVID}-19: a mathematical modelling study". In: *Lancet Infect Dis* 3099.20 (2020), p. 2020.01.31.20019901.

[438] Mumford. "The dawning of the age of stochasticity". In: *Mathematics: Frontiers and Perspectives* (2000).

[439] P W Anderson. "More is different". In: *Science* 177.4047 (Aug. 1972), pp. 393–396.

[440] Lander Willem et al. “Lessons from a decade of individual-based models for infectious disease transmission: a systematic review (2006-2015)”. In: *BMC Infectious Diseases* 17.1 (Sept. 2017), p. 612.

[441] Neil M Ferguson et al. “Strategies for containing an emerging influenza pandemic in Southeast Asia”. In: *Nature* 437.7056 (Sept. 2005), pp. 209–214.

[442] David Applebaum. *Lévy Processes and Stochastic Calculus*. Cambridge University Press, Apr. 2009.

[443] J N Lyness. “Numerical algorithms based on the theory of complex variable”. In: *Proceedings of the 1967 22nd national conference*. ACM ’67. Association for Computing Machinery, Jan. 1967, pp. 125–133.

[444] Ake Svensson. “A note on generation times in epidemic models”. In: *Mathematical Biosciences* 208.1 (July 2007), pp. 300–311.

[445] Matthew D. Hoffman et al. “Stochastic variational inference”. In: *Journal of Machine Learning Research* (2013). ISSN: 15324435.

[446] Andrew Gelman et al. *Bayesian Data Analysis*. Chapman & Hall/CRC, July 2003.

[447] Joel C Miller. “A primer on the use of probability generating functions in infectious disease modeling”. In: *Infectious Disease Modelling* 3 (2018), pp. 192–248.

[448] Folkmar Bornemann. “Accuracy and stability of computing high-order derivatives of analytic functions by Cauchy integrals”. In: *Foundations of Computational Mathematics* 11.1 (Feb. 2011), pp. 1–63.

[449] Vishaal Ram and Laura P Schaposnik. “A modified age-structured SIR model for COVID-19 type viruses”. In: *Scientific Reports* 11.1 (2021), pp. 1–15.

[450] Toshikazu Kuniya. “Global behavior of a multi-group SIR epidemic model with age structure and an application to the chlamydia epidemic in Japan”. In: *SIAM Journal on Applied Mathematics* 79.1 (2019), pp. 321–340.

[451] Kazuki Kuga and Jun Tanimoto. “Impact of imperfect vaccination and defense against contagion on vaccination behavior in complex networks”. In: *Journal of Statistical Mechanics: Theory and Experiment* 2018.11 (2018), p. 113402.

[452] Lijuan Chen and Jitao Sun. “Optimal vaccination and treatment of an epidemic network model”. In: *Physics Letters A* 378.41 (2014), pp. 3028–3036.

[453] Yuting Fu et al. "Optimal lockdown policy for vaccination during COVID-19 pandemic". In: *Finance Research Letters* 45 (2022), p. 102123.

[454] Edouard Mathieu et al. "A global database of COVID-19 vaccinations". In: *Nature Human Behaviour* 5.7 (2021), pp. 947–953.

[455] Andris Abakuks. "Some optimal isolation and immunisation policies for epidemics". PhD thesis. University of Sussex, 1972.

[456] Richard Morton and Kenneth H Wickwire. "On the optimal control of a deterministic epidemic". In: *Advances in Applied Probability* 6.4 (1974), pp. 622–635.

[457] Yinggao Zhou et al. "Optimal vaccination policies for an SIR model with limited resources". In: *Acta Biotheoretica* 62.2 (2014), pp. 171–181.

[458] Lotty E Duijzer et al. "Dose-optimal vaccine allocation over multiple populations". In: *Production and Operations Management* 27.1 (2018), pp. 143–159.

[459] Eleni Zavrakli et al. "Optimal age-specific vaccination control for COVID-19". In: *arXiv preprint arXiv:2104.15088* (2021).

[460] Hamza Boutayeb et al. "Automated optimal vaccination and travel-restriction controls with a discrete multi-region SIR epidemic model". In: *Communications in Mathematical Biology and Neuroscience* 2021.Article ID 22 (2021), Article–ID.

[461] Sunmi Lee, Michael Golinski, and Gerardo Chowell. "Modeling optimal age-specific vaccination strategies against pandemic influenza". In: *Bulletin of Mathematical Biology* 74.4 (2012), pp. 958–980.

[462] Chris Kenyon. "Flattening-the-curve associated with reduced COVID-19 case fatality rates-an ecological analysis of 65 countries". In: *Journal of Infection* 81.1 (2020), e98–e99.

[463] Hannah Ritchie et al. "Coronavirus pandemic (COVID-19)". In: *Our world in data* (2020). URL: https://ourworldindata.org/coronavirus.

[464] Peter J Collins. *Differential and integral equations*. OUP Oxford, 2006.

[465] UN. *World population prospects - population division*. Aug. 2019. URL: https://population.un.org/wpp/Download/Standard/Population/.

[466] Natalie E Dean and M Elizabeth Halloran. "Protecting the herd with vaccination". In: *Science* 375.6585 (2022), pp. 1088–1089.

[467] David E Bloom, Daniel Cadarette, and Maddalena Ferranna. "The societal value of vaccination in the age of COVID-19". In: *American Journal of Public Health* 111.6 (2021), pp. 1049–1054.

[468] Jamie Bedson et al. “A review and agenda for integrated disease models including social and behavioural factors”. In: *Nature Human Behaviour* 5.7 (2021), pp. 834–846.

[469] LJ Muhammad et al. “Supervised machine learning models for prediction of COVID-19 infection using epidemiology dataset”. In: *SN Computer Science* 2.1 (2021), pp. 1–13.

[470] Fred Brauer, Carlos Castillo-Chavez, and Zhilan Feng. *Mathematical models in epidemiology*. Vol. 32. Springer, 2019.

[471] Anas Abou-Ismail. “Compartmental models of the COVID-19 pandemic for physicians and physician-scientists”. In: *SN Comprehensive Clinical Medicine* 2.7 (2020), pp. 852–858.

[472] Lingcai Kong et al. “Compartmental structures used in modeling COVID-19: a scoping review”. In: *Infectious Diseases of Poverty* 11.1 (2022), pp. 1–9.

[473] Raffaele Vardavas, Pedro Nascimento de Lima, and Lawrence Baker. “Modeling COVID-19 nonpharmaceutical interventions: exploring periodic NPI strategies”. In: *medRxiv* (2021).

[474] Tomas de-Camino-Beck. “A modified SEIR Model with Confinement and Lockdown of COVID-19 for Costa Rica”. In: *medRxiv* (2020).

[475] R Adhikari et al. “Inference, prediction and optimization of non-pharmaceutical interventions using compartment models: the PyRoss library”. In: *arXiv preprint arXiv:2005.09625* (2020).

[476] Lisa Sattenspiel and Klaus Dietz. “A structured epidemic model incorporating geographic mobility among regions”. In: *Mathematical Biosciences* 128.1-2 (1995), pp. 71–91.

[477] Fred Brauer. “Epidemic models with heterogeneous mixing and treatment”. In: *Bulletin of Mathematical Biology* 70.7 (2008), pp. 1869–1885.

[478] Glenn Ellison. *Implications of heterogeneous SIR models for analyses of COVID-19*. Tech. rep. National Bureau of Economic Research, 2020.

[479] Ira M Longini Jr, Eugene Ackerman, and Lila R Elveback. “An optimization model for influenza A epidemics”. In: *Mathematical Biosciences* 38.1-2 (1978), pp. 141–157.

[480] Martin Eichner et al. “Direct and indirect effects of influenza vaccination”. In: *BMC Infectious Diseases* 17.1 (2017), pp. 1–8.

[481] Fuminari Miura et al. “Optimal vaccine allocation for COVID-19 in the Netherlands: A data-driven prioritization”. In: *PLoS Computational Biology* 17.12 (2021), e1009697.

[482] Noelle-Angelique M Molinari et al. “The annual impact of seasonal influenza in the US: measuring disease burden and costs”. In: *Vaccine* 25.27 (2007), pp. 5086–5096.

[483] Nuru Saadi et al. “Models of COVID-19 vaccine prioritisation: a systematic literature search and narrative review”. In: *BMC Medicine* 19.1 (2021), pp. 1–11.

[484] Jean-François Delmas, Dylan Dronnier, and Pierre-André Zitt. “Optimal vaccination: various (counter) intuitive examples”. In: *Journal of Mathematical Biology* 86.2 (2023), p. 26.

[485] Nir Gavish and Guy Katriel. “Optimal vaccination at high reproductive numbers: sharp transitions and counterintuitive allocations”. In: *Proceedings of the Royal Society B* 289.1983 (2022), p. 20221525.

[486] Gregory S Zaric and Margaret L Brandeau. “Resource allocation for epidemic control over short time horizons”. In: *Mathematical Biosciences* 171.1 (2001), pp. 33–58.

[487] Evelot Duijzer et al. “The most efficient critical vaccination coverage and its equivalence with maximizing the herd effect”. In: *Mathematical Biosciences* 282 (2016), pp. 68–81.

[488] Jonathan Dushoff et al. “Vaccinating to protect a vulnerable subpopulation”. In: *PLoS medicine* 4.5 (2007), e174.

[489] Eunha Shim. “Optimal allocation of the limited COVID-19 vaccine supply in South Korea”. In: *Journal of clinical medicine* 10.4 (2021), p. 591.

[490] Eunha Shim. “Prioritization of delayed vaccination for pandemic influenza”. In: *Mathematical Biosciences and Engineering* 8.1 (2011), p. 95.

[491] Jan Medlock and Lauren Ancel Meyers. “Optimizing allocation for a delayed influenza vaccination campaign”. In: *PLoS Currents* 1 (2009).

[492] Roy M Anderson and Robert M May. *Infectious diseases of humans: dynamics and control*. Oxford University Press, 1992.

[493] Elisabeth Mahase. “Monkeypox: Healthcare workers will be offered smallpox vaccine as UK buys 20 000 doses”. In: *British Medical Journal* 377 (2022), o1379.

[494] Frank Ball and Peter Neal. “A general model for stochastic SIR epidemics with two levels of mixing”. In: *Mathematical Biosciences* 180.1-2 (2002), pp. 73–102.

[495] Matt J Keeling and Peter J White. “Targeting vaccination against novel infections: risk, age and spatial structure for pandemic influenza in Great Britain”. In: *Journal of the Royal Society Interface* 8.58 (2011), pp. 661–670.

[496] Emmanuel Paradis. *Definition of formats for coding phylogenetic trees in R*. 2006.

[497] TM Cook and JV Roberts. “Impact of vaccination by priority group on UK deaths, hospital admissions and intensive care admissions from COVID-19”. In: *Anaesthesia* 76.5 (2021), pp. 608–616.

[498] Willy Feller. “On the Integral Equation of Renewal Theory”. In: *The Annals of Mathematical Statistics* (1941). ISSN: 0003-4851. DOI: 10.1214/aoms/1177731708.

[499] David Champredon, Jonathan Dushoff, and David J.D. Earn. “Equivalence of the Erlang-distributed SEIR epidemic model and the renewal equation”. In: *SIAM Journal on Applied Mathematics* (2018). ISSN: 00361399. DOI: 10.1137/18M1186411.

[500] Kenny S Crump and Charles J Mode. “A general age-dependent branching process. I”. In: *Journal of Mathematical Analysis and Applications* 24.3 (1968), pp. 494–508.

[501] Kenny Crump and Charles J Mode. “A general age-dependent branching process. II”. In: *Journal of Mathematical Analysis and Applications* 25.1 (Jan. 1969), pp. 8–17.

[502] Marek Kimmel. “The point-process approach to age- and time-dependent branching processes”. In: *Advances in Applied Probability* 15.1 (1983), pp. 1–20.

[503] Ole Barndorff-Nielsen and G F Yeo. “Negative binomial processes”. In: *Journal of Applied Probability* 6.3 (Dec. 1969), pp. 633–647.

[504] Diómedes Bárcenas. “The fundamental theorem of calculus for Lebesgue integral.” In: *Divulgaciones Matemáticas* 8.1 (2000), pp. 75–85.

[505] Jagdish Chandra and Paul W Davis. “Linear generalizations of Gronwall’s inequality”. In: *Proceedings of the American Mathematical Society* 60.1 (1976), pp. 157–160.

[506] Binyamin Schwarz. “Totally positive differential systems”. In: *Pacific Journal of Mathematics* 32.1 (1970), pp. 203–229.

[507] Abraham Berman and Robert J Plemmons. *Nonnegative matrices in the mathematical sciences*. SIAM, 1994.

[508] Thomas Scott Blyth and Edmund F Robertson. *Basic linear algebra*. Springer Science & Business Media, 2002.

[509] Matthew J Penn and Matthew G Hennessy. “Optimal loading of hydrogel-based drug-delivery systems”. In: *Applied Mathematical Modelling* 112 (2022), pp. 649–668.

[510] Joanna Marks and Matthew J Penn. “Barely a passing resemblance: Why women’s football stands out from the crowd”. In: *Significance* 20.6 (2023), pp. 22–25.

Appendix - Paper I

In this appendix, we present a rigorous mathematical framework summarising the results in this paper. We also detail further algorithms that may be useful in the implementation of Phylo2Vec.

A.1 Notations and definitions

A.1.1 Notations

We shall use the notation $\mathcal{T}$ to refer to a tree, $\boldsymbol{b}$ for the branch lengths of $\mathcal{T}$ and n as the number of leaves (or taxa).

A.1.2 Node labels

In a tree with n leaves, we set the convention that the leaf nodes are *labelled* from 0 to $n-1$, the internal nodes are *labelled* from n to $2n-3$ and the root (if it exists) is labelled as $2n-2$.

A.1.3 Generation

By definition, there is a unique path connecting each pair of nodes in a tree. Suppose that the path from node i to the root contains the nodes $i, a_1, \ldots, a_{g_i}$ in that order (and the root is hence a_{g_i}). We call g_i the *generation* of node i. The path from any a_x to the root must be $a_x, a_{x+1}, \ldots, a_{g_i}$ and hence $g_{a_x} = g_i - x$.

A.1.4 Unrooting

Suppose that, in a rooted tree, the two children of the root are x and y. To "unroot" the tree, we remove the edges connecting x and y to the root and add in a new edge connecting x and y. This edge is given length $b_x + b_y$ where, for example, b_x is the length of the original edge joining x to the root.

A.2 Phylo2Vec details

A.2.1 Vector representation

As described in the main text, Phylo2Vec is a way to represent binary trees with n leaves using a single integer vector, $\boldsymbol{v}$ of dimension $n-1$ that is simply constrained by

$$\boldsymbol{v}[j] \in \{0, 1, \dots, 2(j-1)\} \; \forall j \in \{1, \dots, n-1\}$$

The construction of the tree from this representation and the bijectivity between $\boldsymbol{v}$ and the space of trees are covered in the main text.

A.2.2 Bijectivity of v to the space of all possible trees

We can show that our mapping from $\mathbb{V}$ to the space of trees is a bijection.

Lemma A.1 *The mapping between the set of vectors $\boldsymbol{v} \in \mathbb{V}$ and the set of (topologically equivalent, labelled) trees is a bijection.*

Proof: As, by construction, the number of possible vectors $\boldsymbol{v}$ and the number of trees are the same ($(2n-3)!!$), it is simply necessary to show that this mapping is injective.

For any two nodes a and b, we define $M(a, b)$ to be their most recent common ancestor (MRCA). Moreover, for any nodes a and b, we use the notation $a \prec b$ if a is an ancestor of b.

We can use the fact that the MRCA of any pair of leaf nodes in the tree is unchanged during the Phylo2Vec construction process (once both nodes have been added to the tree). Note that the path from a node x to the root changes through the addition of an extra node if and only if a new node is appended to an edge on this path. Thus, if $M(a, b) = x$, x will remain on the paths from a and b to the root. Moreover, nodes of higher generation than x can only be added to one of the paths (as otherwise there would have to be an edge connecting at least one node of higher generation than x on both the original paths, contradicting x being the MRCA). Thus, the MRCA of a and b will be unchanged. Similarly, we can see that $a \prec b$ for two a and b at a given stage of the algorithm, this relationship will be unchanged throughout the construction process.

Suppose that two vectors $\boldsymbol{v}$ and $\boldsymbol{v}'$ result in trees $\mathcal{T}$ and $\mathcal{T}'$, and that $\boldsymbol{v} \neq \boldsymbol{v}'$. Define $i := \min\{j : v_j \neq v'_j\}$.

Define X to be the set of leaf nodes that, just before node i is added, are descended from the edge to which node i is appended when constructing the tree according to $\boldsymbol{v}$, and X' to be the equivalent set for $\boldsymbol{v}'$. We have $X \neq X'$, as each edge has a unique set of nodes descended from it.

If both $X \setminus X'$ and $X' \setminus X$ are non-empty, suppose that $a \in X \setminus X'$ and $b \in X' \setminus X$. Then, in $\mathcal{T}$, it must be the case that $M(b,i) \prec M(a,i)$ (both at this stage in the algorithm, and in the final tree, by the previous argument) once node i has been added (as the $M(a,i)$ will be the new internal node and $M(b,i)$ cannot this new internal node as $b \notin X$). A similar argument shows that $M(a,i) \prec M(b,i)$ in $\mathcal{T}'$ and hence $\mathcal{T} \neq \mathcal{T}'$.

If only one of $X \setminus X'$ and $X' \setminus X$ is non-empty, suppose without loss of generality that $X \setminus X'$ is non-empty and choose $a \in X \setminus X'$. We can choose a distinct node $b \in X' \cap X$ (as otherwise, if $X' \cap X$ and $X' \setminus X$ are both empty, we must have one of X and X' being empty which is a contradiction as every edge has at least one leaf node descended from it). Then, $M(a,i)$ is the newly-added internal node in the construction of $\mathcal{T}'$, but not in the construction of $\mathcal{T}$ (as $b \notin X$). However, in both cases, $M(b,i)$ is the newly-added node. Hence, in $\mathcal{T}$, $M(a,i) = M(b,i)$, but in $\mathcal{T}'$, they are distinct. Thus, $\mathcal{T}' \neq \mathcal{T}$.

Hence, topological non-equivalence holds in both cases, and the map from $\boldsymbol{v}$ to the set of trees is therefore injective and therefore bijective.

A.2.3 Label-asymmetry of v-induced distance

As discussed in the main text, $\boldsymbol{v}$ induces a natural distance function between trees - namely, that the distance between $\boldsymbol{v}$ and $\boldsymbol{w}$ is equal to $\sum_{i=0}^{k} \mathbb{I}_{v_i \neq w_i}$.

However, this distance function is dependent on the labels assigned to each leaf (and is hence label-asymmetric). A simple example of this can be found in the case of four leaves.

Consider the tree given by $\boldsymbol{v}_1 = (0,1,2)$. This is a ladder tree (that is, each internal node and the root is parent to at least one leaf node) with nodes in order $0, 1, \{2,3\}$ (where the $\{2,3\}$ is used to denote the fact that nodes 2 and 3 are from the same generation and so could be read in either order). This is distance 1 away from $\boldsymbol{v}_2 = (0,1,4)$, which is again a ladder tree with nodes now ordered as $3, 0, \{1,2\}$. Thus, if distance were symmetric, then any ladder tree with ordered nodes $a, b, \{c,d\}$ would be distance 1 away from the ladder trees with ordered nodes $d, a, \{b,c\}$ and $c, a, \{b,d\}$. However, the ladder tree with ordered nodes $2, 0, \{1,3\}$ is given by $\boldsymbol{v}_3 = (0,2,1)$, which is distance 2 away from $\boldsymbol{v}_1$. Thus, the distance function is not label-symmetric.

A.2.4 Unrooted tree equivalence classes

Noting from the main text that there are $(2k-3)!! = \prod_{i=0}^{k-2}(2i+1)$ unrooted trees with $k+1$ leaves, we can partition the space of trees with $k+1$ leaves into $(2k-3)!!$ equivalence classes $\{\mathcal{E}_i : i = 1, 2, \ldots, (2k-3)!!\}$ such that the removing the root from each of the trees in a given class $\mathcal{E}_i$ (according to the procedure described in Notations and definitions in the Appendix) results in the

same unrooted tree. These equivalence classes all contain $(2k-1)$ trees (as, reversing the unrooting procedure, a root can be added onto each of the $(2k-1)$ edges of an unrooted tree).

From the previous work, Felsenstein's likelihood is constant in each equivalence class. This has two consequences when attempting to maximise the likelihood. Firstly, a global maximum in the space of rooted trees will not be unique, as any of the other trees in the same equivalence class will also give a maximal likelihood. Secondly, if $\boldsymbol{v}$ is a local maximum (that is, all $\boldsymbol{w}$ with one entry different to $\boldsymbol{v}$ give a lower likelihood value), equivalence classes provide a way to change to a new vector $\boldsymbol{u}$ without having to decrease the likelihood.

Similarly to the case of reordering, the set of equivalence classes of vectors that are distance 1 from $\boldsymbol{v}$ will not in general be the same as the set with distance 1 from $\boldsymbol{u}$. One can either change equivalence class randomly or (unlike in the case of reordering), as there are only $2k-1$ members of each class, systematically iterate through them. Finding the vectors $\boldsymbol{w}$ corresponding to the trees in an equivalence class can again be done by constructing the tree matrix M, making the appropriate changes and then using the inversion algorithm to create a new vector representation.

Our experiments have shown that changing $\boldsymbol{v}$ is effective in increasing algorithmic performance. However, we found that reordering provided better increase in performance, and therefore did not present this other switching method in the algorithm in the main text.

A.2.5 Number of moves

Consider a vector $\boldsymbol{v}$ of length $n-1$ representing a tree with n leaves. For each entry j, there are $2j-1-1 = 2(j-1)$ possible moves. Summing for all $j \in 1, \ldots, n-1$ gives:

$$\sum_{j=1}^{n-1} 2(j-1) = (n-1)(n-2) \approx n^2$$

Thus, the number of possible moves for Phylo2Vec is of the order $\mathcal{O}(n^2)$.

A.3 Algorithms

The algorithms provided here are presented in rough descriptive pseudocode; they are written in human-readable fashion to facilitate comprehension, but are not necessarily optimised. Please refer to the GitHub repository (https://github.com/Neclow/phylo2vec) for detailed implementation.

Algorithm 4 Labelling a rooted tree as an ordered $\boldsymbol{v}$

Input rooted tree $\mathcal{T}$ with n leaves
$\boldsymbol{v} \leftarrow [0]$ ▷ Initial integer vector
for all leaves $j = 2, ..., n-1$ **do**
  $\mathcal{T}_j \leftarrow$ subtree with leaves $0, ..., j$
  $s \leftarrow$ leaf in $\mathcal{T}_j$ such that j and s form a cherry
  $\boldsymbol{v}$.append(s)
end for
return $\boldsymbol{v}$

Algorithm 5 Recovering an ordered rooted tree from an ordered $\boldsymbol{v}$

Input $\boldsymbol{v}$ of size $n-1$ satisfying Eq. 3.1 ▷ Ordered integer vector
$\mathcal{T} \leftarrow$ a rooted tree with two leaves 0 and 1 ▷ Initial tree
for all $j = 2, ..., n-1$ **do**
  Add a new leaf j to $\mathcal{T}$ such that it forms a cherry with $\boldsymbol{v}[j]$
  $a(j, \boldsymbol{v}[j]) \leftarrow 2(n-1) - j + 1$ ▷ Ancestor of j and $\boldsymbol{v}[j]$
end for
return $\mathcal{T}$

Algorithm 6 Recovering a rooted tree from a Phylo2Vec $\boldsymbol{v}$

Input $\boldsymbol{v}$ of dimension $n-1$ satisfying Eq. 3.2 ▷ Phylo2Vec vector
pairs $\leftarrow [(0, 1)]$ ▷ The objective of the algorithm is find the nodes to merge and their order. The first pair is always (0,1) for a 2-leaf tree.
for all leaves $j = 2, ..., n-1$ **do**
  if $\boldsymbol{v}[j] \leq j$ **then** ▷ This is like an ordered vector. The branch leading to $\boldsymbol{v}$[j] gives birth birth to leaf j
    next pair $\leftarrow (\boldsymbol{v}[j], j)$
    position $\leftarrow 1$ ▷ This pair corresponds to a cherry at height 0, so it should be drawn first.
  else ▷ The branch to be split is internal
    position $\leftarrow \boldsymbol{v}[j]$ - len(pairs) + 1 ▷ The index at which the next pair will be inserted
    descendant $\leftarrow$ pairs[position - 1][1] ▷ A node that we processed beforehand that is deeper than the branch $\boldsymbol{v}[j]$
    next pair $\leftarrow$ (pairs[descendant, j])
  end if
  pairs.insert(position, next pair)
end for
return pairs ▷ The pairs indicate how to build the tree (and form a Newick string)

Algorithm 7 Labelling a rooted tree as a Phylo2Vec vector

Input rooted tree r with n leaves ▷ We assume a Newick string with integer nodes and labelled internal nodes.
Ex: (((2,3)6,1)7,(0,4)5)8;

$M \leftarrow \text{reduce}(r)$ ▷ "Reduce" the Newick string to a $(n-1) \times 3$ matrix.
Columns 1-2 = children nodes
Column 3 = ancestor
Ex:
$$\begin{matrix} 0 & 4 & 5 \\ 2 & 3 & 6 \\ 1 & 6 & 7 \\ 5 & 7 & 8 \end{matrix}$$

$C \leftarrow \text{extract\_cherries}(M)$ ▷ Starting from the smallest internal node, replace the internal nodes by their smallest child (and discard the third column). This is equivalent to the pairs in Algorithm 6
Ex:
$$\begin{matrix} 0 & 4 \\ 2 & 3 \\ 1 & 2 \\ 0 & 1 \end{matrix}$$

$\boldsymbol{v} \leftarrow \text{build\_vector}(C)$ ▷ Use the same logic as Algorithm 6 to retrieve each $\boldsymbol{v}_j$ from the position of the containing leaf j

return $\boldsymbol{v}$

Algorithm 8 Labelling edges in the Phylo2Vec framework

Input rooted tree $\mathcal{T}$ with n leaves ▷ Leaves labelled $1, \ldots, n-1$
$\boldsymbol{e} \leftarrow [\,]$ ▷ Edges
for all leaves $j = 1, ..., n-1$ **do**
  $\boldsymbol{e}(j, a(j)) \leftarrow j$ ▷ The edge from which leaf j descends from is labelled j
end for
$j \leftarrow n$ ▷ The label of the next edge
for all heights $h = 1, ..., n-1$ **do** ▷ height = number of edges on the path from a given node to the furthest leaf node
  $a_j \leftarrow$ ancestor of height h with the greatest child
  $\boldsymbol{e}(a_j, a(a_j)) \leftarrow j$ ▷ The edge connecting a_j and its ancestor is labelled j
  $j \leftarrow j + 1$ ▷ Increment j
end for
return $\boldsymbol{e}$

A.4 Supplementary Figures

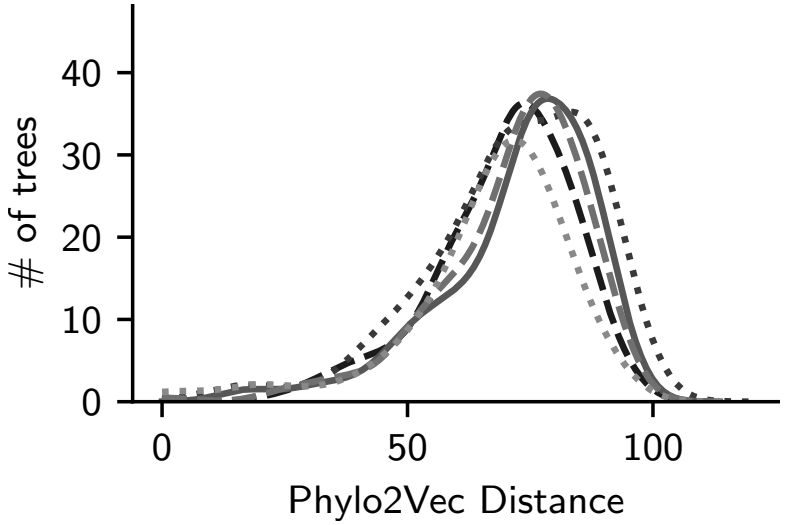


Figure A.1: Density plots of Phylo2Vec distance μ for an equivalence set of rooted trees with 200 taxa with a fixed Phylo2Vec vector $\boldsymbol{v}$. 5 random $\boldsymbol{v}$ are shown.

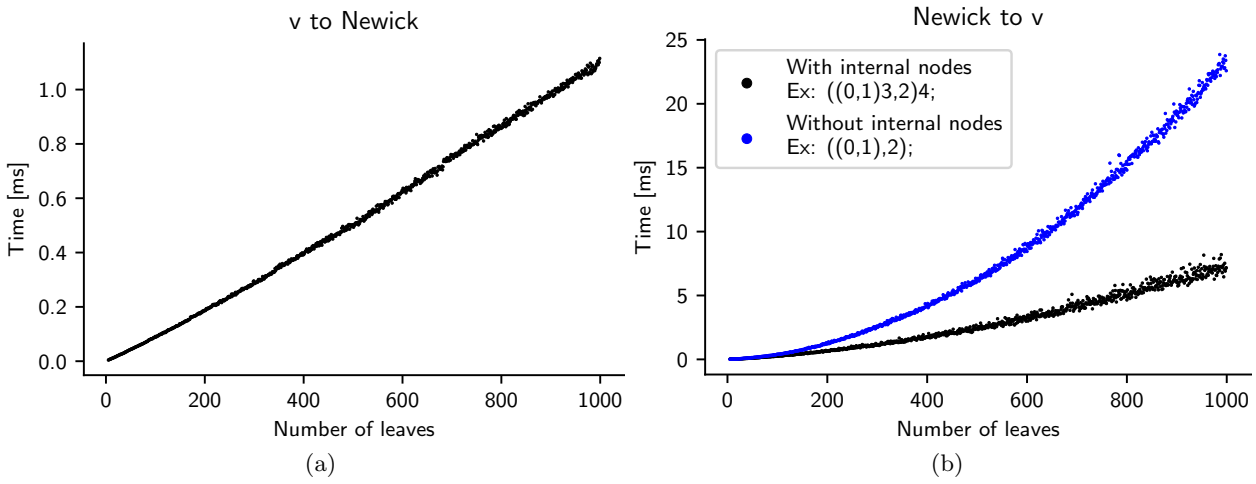


Figure A.2: Execution times for different tree sizes of **(a)** Algorithm 6 $\big(\mathcal{O}(n)\big)$ to recover a rooted tree from a Phylo2Vec $\boldsymbol{v}$ and **(b)** Algorithm 7 $\big($with internal nodes: $\mathcal{O}(n \log n)$; without internal nodes: $\mathcal{O}(n^2)\big)$ to label a rooted tree as a Phylo2Vec $\boldsymbol{v}$. For each size, we evaluated the execution time with a fixed configuration of 100 executions in a loop and 7 repeats using `timeit` in Python. When the number of taxa is large, manipulations on long strings to produce or process the Newick string can increase memory and thus increase execution time.

Appendix - Paper II

B.1 BME for rooted trees

B.1.1 Comparing the rooted and unrooted objectives

Lemma B.1 *Consider adding an extra taxon, n, to the set of taxa such that, for some D^* and δ.*

$$|D_{ni} - D^*| < \delta \quad \forall i \neq n \tag{B.1}$$

By joining node n to the root, this creates an unrooted tree $\mathcal{T}^u$, and we can create a rooted tree $\mathcal{T}^r$ by removing node n. If e^u_{ij} denotes inter-taxa distance in $\mathcal{T}^u$ and e^r_{ij} denotes inter-taxa distance in $\mathcal{T}^r$, then

$$\left| \sum_{i=0}^{n}\sum_{j=0}^{n} D_{ij}2^{-e^u_{ij}} - \sum_{i=0}^{n-1}\sum_{j=0}^{n-1} D_{ij}2^{-e^r_{ij}} - D^* \right| \leq \delta \tag{B.2}$$

Proof: Firstly, note that for $i, j < n$,

$$e^r_{ij} = e^u_{ij} \tag{B.3}$$

as the path between i and j will not contain any leaf nodes other than i and j, and therefore will not contain node n. Then, using the fact that $D_{nn} = 0$,

$$\sum_{i=0}^{n}\sum_{j=0}^{n} D_{ij}2^{-e^u_{ij}} = \sum_{j=0}^{n-1} D_{nj}2^{-e^u_{nj}} + \sum_{i=0}^{n-1} D_{in}2^{-e^u_{in}} + \sum_{i=0}^{n-1}\sum_{j=0}^{n-1} D_{ij}2^{-e^u_{ij}} \tag{B.4}$$

$$\leq \sum_{j=0}^{n-1}(D^* + \delta)2^{-e_{nj}} + \sum_{i=0}^{n-1}(D^* + \delta)2^{-e^u_{in}} + \sum_{i=0}^{n-1}\sum_{j=0}^{n-1} D_{ij}2^{-e^r_{ij}} \tag{B.5}$$

We can make progress with this sum by noting the Kraft Equality, found throughout the literature [150]

$$\sum_{j} 2^{-e^u_{nj}} = \frac{1}{2} \tag{B.6}$$

This means that

$$\sum_{i=0}^{n}\sum_{j=0}^{n} D_{ij}2^{-e^u_{ij}} \leq \sum_{i=0}^{n-1}\sum_{j=0}^{n-1} D_{ij}2^{-e^r_{ij}} + D^* + \delta \tag{B.7}$$

Similarly, one can show

$$\sum_{i=0}^{n}\sum_{j=0}^{n} D_{ij}2^{-e^u_{ij}} \geq \sum_{i=0}^{n-1}\sum_{j=0}^{n-1} D_{ij}2^{-e^r_{ij}} + D^* - \delta \tag{B.8}$$

and hence the result follows.

B.1.2 Understanding the BME rooting

Definition 2 ***Distance to root heuristic:*** *Under the assumption that the tree is ultrametric, we estimate the distance between the root and the leaf taxa using the following algorithm:*

1) *Find two leaf nodes that share a parent that is not the root. If no such pair exists, then move to step 3.*

2) *Let d_1 and d_2 be the distance between each leaf node and its parent. Suppose that d_3 is the distance between the parent and its parent (i.e. the grandparent of the leaves). Remove the leaves from the tree, update the distance between the parent and its parent to be $d_3 + \frac{d_1+d_2}{2}$ and return to step 1*

3) *If d_1 and d_2 are the distances between the two remaining children and the root, then the distance between the root and the taxa is $\frac{d_1+d_2}{2}$.*

Lemma B.2 *Consider the true unrooted tree $\mathcal{T}$ and suppose that the distance matrix D gives the true (time) distances between each pair of taxa in $\mathcal{T}$.*

Then, the optimal rooting — that is, the edge in the tree such that placing the root on this edge minimizes the BME objective — maximises the distance to root heuristic defined in Definition 2.

Proof: Choose an edge, b on which to place the root. We define an indicator function X_b such that $X_b(A, B)$ is 1 if b is on the path between nodes A and B and 0 otherwise. Adding a root to edge b changes the objective function by halving the weight assigned to each D_{AB} such that $X_b(A, B) = 1$ as the path length between these nodes will increase by 1. Thus, using f_r to denote the rooted objective and f_u to denote the unrooted objective,

$$f_r = f_u - \frac{1}{2} \sum_{X_b(A,B)=1} D_{AB} 2^{-e_{AB}} \tag{B.9}$$

where the e_{AB} are the path lengths in the original, unrooted tree and both A and B are allowed to vary in the sum. Hence, for a fixed unrooted tree topology, the optimal rooting will solve

$$\max_b \left\{ \sum_{X_b(A,B)=1} D_{AB} 2^{-e_{AB}} \right\} \tag{B.10}$$

By Lemma B.11,

$$\sum_{X_b(A,B)=1} D_{AB} 2^{-e_{AB}} = 2 \sum_A 2^{-g_A} D_{Ar} \tag{B.11}$$

while, by Lemma B.12, the root-to-tip heuristic, $\mathcal{D}$, satisfies

$$\mathcal{D} = \sum_{A} 2^{-g_A} D_{Ar} \tag{B.12}$$

and hence, the optimal rooting solves

$$\max_{b} \left\{ \mathcal{D} \right\} \tag{B.13}$$

as required.

B.2 Ordered Trees

Definition 3 ***Left-to-right construction algorithm*** *An ordered tree can be constructed as follows:*

1) *Begin with nodes* 0 *and* 1*, each joined to a root. Label the edge joining node* 0 *to the root as edge* 0*, and the edge joining node* 1 *to the root as edge* 1

2) *Process the nodes in order 2, 3, 4... When processing node k, join it to edge v_k (creating a new internal node — this is well-defined as $v_k < k$). Label the edge joining node k to the tree as edge k.*

Lemma B.3 *The left-to-right algorithm in Definition 3 and the standard Phylo2Vec algorithm defined in [47] produce the same tree, provided $\boldsymbol{v}$ is ordered.*

Note: This lemma refers to an old version of Paper I where a different tree construction algorithm was presented. While I have included it for completeness, I suggest the reader ignore this lemma as the construction algorithm presented in the most recent version of Paper 1 in this thesis is equivalent to the left-to-right construction algorithm described above.

Proof: To show that the left-to-right algorithm gives an equivalent tree, define $\mathcal{T}$ to be the tree resulting from the standard Phylo2Vec algorithm and $\mathcal{T}'$ to be the tree resulting from this new left-to-right algorithm. We proceed by induction on the number of nodes, n, noting that the case $n = 2$ is trivial.

Suppose now that the algorithms are equivalent for $n = m$. Choose some ordered $\boldsymbol{v}$ of length $m+1$ (so that this corresponds to $n = m + 1$) and consider processing the first node using the Phylo2Vec algorithm, so that (using the fact that $\boldsymbol{v}$ is ordered so that no nodes are skipped), node m merges with node v_m.

From this step, the Phylo2Vec algorithm proceeds as if there were $n-1$ nodes and the vector was $\tilde{\boldsymbol{v}} = (v_0, v_1, ..., v_{m-1})$. Define $\tilde{\mathcal{T}}$ to be the tree given by $\tilde{\boldsymbol{v}}$. Hence, $\tilde{\mathcal{T}}$ can be created from $\mathcal{T}$ by removing nodes v_m and m (and the edges connecting them to the tree) and relabelling their parent as v_m. Equivalently (by reversing this process), $\mathcal{T}$ can be created from $\tilde{\mathcal{T}}$ by adding a node to the edge joining leaf node v_m to the tree, and connecting node m to this edge.

Moreover, from the inductive hypothesis, $\tilde{\mathcal{T}}$ can be constructed by using the left-to-right algorithm on $\tilde{\boldsymbol{v}}$. As this algorithm processes v_m last, this means that after m nodes have been added to the tree by the left-to-right algorithm applied to $\boldsymbol{v}$, the current tree is given by $\tilde{\mathcal{T}}$.

The final step of the left-to-right algorithm applied to $\boldsymbol{v}$ is to add a node to the edge joining leaf node v_m to the tree, and to connect node m to this edge. As previously discussed, this creates the tree $\mathcal{T}$ and hence, $\mathcal{T} = \mathcal{T}'$ as required.

B.3 A continuous objective function

B.3.1 Construction

Lemma B.4 *For a randomly chosen tree with distribution W, define, for $i, j, < k$, e_{ij}^k to be the path length between taxa i and j when k nodes have been added to the tree (using the left-to-right construction algorithm). Define $E_{ij}^k := \mathbb{E}(2^{-e_{ij}^k})$. Then, for $i < j$*

$$E_{ij}^{k+1} = \begin{cases} E_{ij}^k \left[1 - \frac{1}{2}(W_{ki} + W_{kj})\right] & \text{if } i < j < k \\ \left[\frac{1}{2}\sum_{x \neq i} E_{ix}^k W_{kx}\right] + \frac{1}{4} W_{ki} & \text{if } i < k \end{cases} \tag{B.14}$$

with the remaining values following by symmetry.

Proof: Adding node k to the tree increases the path length between nodes i and j by 1 if and only if $V_k = i$ or $V_k = j$. As this condition is independent of other values of $\boldsymbol{V}$, using e_{ij}^k to be the path length after the k nodes $\{0, ..., k-1\}$ have been added, one can write

$$2^{-e_{ij}^{k+1}} = 2^{-(e_{ij}^k + \mathbb{I}\{V_k \in \{i,j\}\})} = 2^{-e_{ij}^k} \times 2^{-\mathbb{I}\{V_k \in \{i,j\}\}} \tag{B.15}$$

This is a product of independent random variables and so, defining $E_{ij}^k := \mathbb{E}(2^{-e_{ij}^k})$ and noting that $\mathbb{I}\{V_k \in \{i,j\}\}$ is a Bernoulli random variable with probability $W_{ki} + W_{kj}$, this equation becomes

$$E_{ij}^{k+1} = E_{ij}^k \left[(1 - (W_{ki} + W_{kj}) + \frac{1}{2}(W_{ki} + W_{kj})\right] = E_{ij}^k \left[1 - \frac{1}{2}(W_{ki} + W_{kj})\right] \tag{B.16}$$

To close this iterative system, note that when node j is added to the tree, it will be a path length 2 from the leaf node V_j connecting the edge it is joined to. Moreover, the distance between node j and any other nodes $x \neq V_j$ in the tree will be equal to one plus the distance between node x and node V_j. That is,

$$2^{-e_{ij}^{j+1}} = \begin{cases} 2^{-(1+e_{i,V_j}^{j})} & \text{if } V_j \neq i \\ \frac{1}{4} & \text{if } V_j = i \end{cases} \tag{B.17}$$

Thus, conditioning on the value of V_j,

$$E_{ij}^{j+1} = \left[\frac{1}{2}\sum_{x \neq i} E_{ix}^{j} W_{jx}\right] + \frac{1}{4} W_{ji} \tag{B.18}$$

Finally, by symmetry, $E_{ji}^{j+1} = E_{ij}^{j+1}$. Noting that E_{ij}^{k} is undefined (and unnecessary) for $k < \max(i,j)+1$, as nodes i and j have not both been added to the tree, (B.16) and (B.18) hence form a closed system. This can be solved inductively, finding all E_{ij}^{m} terms for $m = 2, 3, ..., n$.

B.3.2 Discrete minima

Lemma B.5 *Define $f(\boldsymbol{v})$ to be the objective function for the discrete tree given by the Phylo2Vec vector $\boldsymbol{v}$. Then, if V^* is the set of vectors $\boldsymbol{v}$ which minimize f, any optimal W satisfies*

$$\mathbb{P}(\boldsymbol{V} \in V^* | W) = 1 \tag{B.19}$$

Moreover if $|V^*| = 1$ then there is a unique $\boldsymbol{v}$ minimizing f, then there is a unique W minimizing F, which is the matrix such that

$$W_{m,j} = \mathbb{I}\{v_m = j\} \quad \forall m, j \tag{B.20}$$

Proof: Define V to be the set of ordered tree vectors. Then,

$$F(W) = \sum_{\boldsymbol{u} \in V} \mathbb{P}(\boldsymbol{V} = \boldsymbol{u}|W) f(\boldsymbol{u}) = \sum_{\boldsymbol{u} \in V} \prod_{m=1}^{n-1} W_{m,u_m} f(\boldsymbol{u}) \tag{B.21}$$

As

$$\sum_{\boldsymbol{u} \in V} \mathbb{P}(\boldsymbol{V} = \boldsymbol{u}|W) = 1 \tag{B.22}$$

we see that $F(W)$ is a weighted average of the values of $f(\boldsymbol{u})$ for $\boldsymbol{u} \in V$. Thus, for any $\boldsymbol{v} \in V^*$

$$F(W) \geq f(\boldsymbol{v}) \tag{B.23}$$

and

$$F(W) = f(\boldsymbol{v}) \quad \Rightarrow \quad \mathbb{P}(\boldsymbol{V} \in V^* | W) = 1 \tag{B.24}$$

as required. If $V = \{\boldsymbol{v}\}$, this minimum requires

$$\mathbb{P}(\boldsymbol{V} = \boldsymbol{v} | W) = 1 \tag{B.25}$$

and therefore, using (B.21) W is uniquely defined by

$$W_{m,j} = \mathbb{I}\{v_m = j\} \quad \forall m, j \tag{B.26}$$

as required.

B.4 Orderings

This section considers the labelling algorithm introduced in the main text.

Lemma B.6 *Two nodes have the same label only if they share an ancestor with that label.*

Proof: Suppose that this is false, and that nodes x and y have the same label, L, but do not share an ancestor with that label. Define $a(x)$ and $a(y)$ to be the nodes of lowest generation (that is, the nodes closest to the root) with label L such that they are ancestors of x and y respectively. Note that, by assumption, $a(x) \neq a(y)$ and also, neither can be the root (as the root is the ancestor of all nodes). Moreover, they cannot share a parent, as the children of a parent are labelled differently. Thus, without loss of generality, one can assume that $a(x)$ was labelled first. By definition, the parent of $a(x)$ does not have label L and hence, L must have been the smallest unused label when node $a(x)$ was labelled. A similar argument for $a(y)$ shows that L must have been the smallest unused label when node $a(y)$ was labelled. However, when node $a(y)$ was labelled, L had been used to label $a(x)$, giving the required contradiction.

Lemma B.7 *The set of leaf node labels is a permutation of the set* $\{0, 1, ..., n-1\}$.

Proof: Firstly, note that as there are $n-1$ internal nodes, and a single new label is introduced every time the children of an internal node are labelled, the set of labels used (across all nodes) must be $\{0, 1, ..., n-1\}$.

Suppose that leaf nodes x and y have the same label L. By Lemma B.6, they must share an ancestor a with that label. Define b to be the shared ancestor with the highest generation (that is,

furthest from the root) such that b has label L. Then, either nodes x and y are the children of b (in which case, they have distinct labels) or they are descendants of distinct children, c and d, of b. In the second case, one can impose without loss of generality that c does not have label L and, as label L has already been used to label b, we know that all descendants of c do not have label L. Hence, in both cases, one of x and y does not have label L as required.

Lemma B.8 *For each $i \in \{0, 1, ..., n-1\}$, define $x(i)$ to be the node of the highest generation with label i. Define (for $i > 0$), $y(i)$ to be the label of the parent of $x(i)$. Then, with ordering l, the tree is given by*

$$v_0 = 0 \quad \textit{and} \quad v_i = y(i) \quad \forall i > 0 \tag{B.27}$$

Note: The published version of this proof relies on the old construction algorithm from Paper I (see the text after Lemma B.3 for more details). To make this proof more understandable in the context of this thesis, we include a slightly edited version here (though the overall structure is the same).

Proof: Firstly, note that $\boldsymbol{v}$ is ordered, as the label of a child is greater than or equal to that of its parent. Hence, as the parent of $x(i)$ does not have label i, it must have a label strictly less than i and so $v_i < i$ as required.

One can then proceed by induction on the number of nodes. The case $n = 2$ is trivial, and so suppose it holds for $n = m$ and consider a tree with $n = m + 1$. We suppose that our Phylo2Vec vector is $\boldsymbol{v}$ and that $\boldsymbol{u}$ is a vector containing all but the last entry of $\boldsymbol{v}$ (so that $\boldsymbol{u}$ is a Phylo2Vec vector for a tree with m nodes).

Consider the node with label m in the tree given by $\boldsymbol{v}$. No other node can have label m (as, otherwise, one of its children would have label greater than m, contradicting Lemma B.7). When node m was added to the tree, a node with label $y(m)$ - the parent of node m - was appended to the edge joining node v_m (as this determined how node m was appended to the tree) to its parent. Note that node $y(m)$ must have label v_m as it cannot have the same label as its other child, node m. Thus, $v_m = y(m)$ ensures that the node with label m was added to the tree corrected.

Reversing the last step in our construction algorithm (and leaving node labels unchanged), we now have a tree given by $\boldsymbol{u}$. The values of $y(i)$ for this tree are unchanged (in particular, $y(y(m))$ depends on the label of an ancestor of the parent of node m) and hence, by induction, $\boldsymbol{u}$ correctly generates

the rest of the tree. Thus, the correct tree is generated by $\boldsymbol{v}$ as required.

B.5 Queue Shuffle: generating principled ordering proposals

B.5.1 Asymmetry of ordered tree spaces

Lemma B.9 *Define g_m^k to be the expected distance from the root of the node labelled m in a tree with k nodes chosen uniformly from the space of ordered trees. Define the harmonic sum function*

$$H(m) = \begin{cases} \sum_{j=1}^{m} \frac{1}{j} & \text{if } m \geq 1 \\ 1 & \text{if } m = 0 \end{cases} \tag{B.28}$$

Then,

$$g_m^k = H(k-1) + H(m) - 1 \tag{B.29}$$

Proof: From our left-to-right construction algorithm, conditioning on the value of v_{k-1},

$$g_{k-1}^k = \sum_{m=0}^{k-2} \frac{1}{k-1}(1 + g_m^{k-1}) \tag{B.30}$$

as adding node $k-1$ according to $v_{k-1} = m$ means that the path length between node $k-1$ and the root will be one more than the path length of between m and the root. If $v_{k-1} = m$, it will also increase the path length between node m and the root by 1 and so

$$g_m^k = \frac{k-2}{k-1} g_m^{k-1} + \frac{1}{k-1}(g_m^{k-1} + 1) = g_m^{k-1} + \frac{1}{k-1} \tag{B.31}$$

The initial conditions of this system are that

$$g_0^2 = g_1^2 = 1 \tag{B.32}$$

as in a two-node rooted tree, the leaves are distance 1 from the root.

We claim by induction on k that the solution to this system is (B.29).

Note that this holds for $k = 2$ as $H(1) = H(0) = 1$. Moreover, under the inductive hypothesis that it holds for a tree with $k-1$ nodes, for $m < k-1$,

$$g_m^k = g_m^{k-1} + \frac{1}{k-1} = H(k-2) + H(m) - 1 + \frac{1}{k-1} = H(k-1) + H(m) - 1 \tag{B.33}$$

and

$$g_{k-1}^{k} = \sum_{m=0}^{k-2} \frac{1}{k-1}(1 + g_m^{k-1}) \tag{B.34}$$

$$= \sum_{m=0}^{k-2} \frac{1}{k-1}(1 + H(k-2) + H(m) - 1) \tag{B.35}$$

$$= \sum_{m=0}^{k-2} \frac{1}{k-1}(H(k-2) + H(m)) \tag{B.36}$$

$$= H(k-2) + \frac{1}{k-1} \sum_{m=0}^{k-2} H(m) \tag{B.37}$$

$$= H(k-1) + \frac{1}{k-1} \sum_{m=1}^{k-2} \sum_{j=1}^{m} \frac{1}{j} \tag{B.38}$$

Now,

$$\sum_{m=1}^{k-2} \sum_{j=1}^{m} \frac{1}{m} = \sum_{j=1}^{k-2} \sum_{m=j}^{k-2} \frac{1}{j} = \sum_{j=1}^{k-2} \frac{k-1-j}{j} = (k-1)H(k-2) - (k-2) \tag{B.39}$$

and hence

$$g_{k-1}^{k} = H(k-1) + H(k-2) - \frac{(k-2)}{(k-1)} = H(k-1) + H(k-1) - 1 \tag{B.40}$$

as required.

B.5.2 Nearest Neighbour Interchange

Definition 4 *Define $\tau(\sigma)$ to be the space of possible trees given an ordering σ.*

Definition 5 *For a tree $\mathcal{T}$, define $Q(\mathcal{T})$ to be the random ordering generated by Queue Shuffle.*

Lemma B.10 *Consider a tree $\mathcal{T}$ and suppose that another tree, $\mathcal{T}'$ is one NNI move away from $\mathcal{T}'$. Then*

$$\mathbb{P}\left[\mathcal{T}' \in \tau\left(Q(\mathcal{T})\right)\right] \geq \frac{1}{4} \tag{B.41}$$

To facilitate the proof, we first note that for any permutation σ, by Lemma B.13, that

$$\mathbb{P}\left[Q(\mathcal{T}) = \sigma\right] \in \left\{0, \frac{1}{2^{n-1}}\right\} \tag{B.42}$$

Now, we consider labelling the tree as shown in the left panel of Fig. B.1. Suppose that the edge to which the four subtrees are joined has nodes i and j, with node i being a higher generation than node j. Suppose that the subtrees rooted at the child of j are labelled as $\mathcal{S}_j$ and $\mathcal{S}_c$. Suppose that the other child of i (that is, the child not equal to j) is the root of a subtree $\mathcal{S}_i$. Finally, suppose that

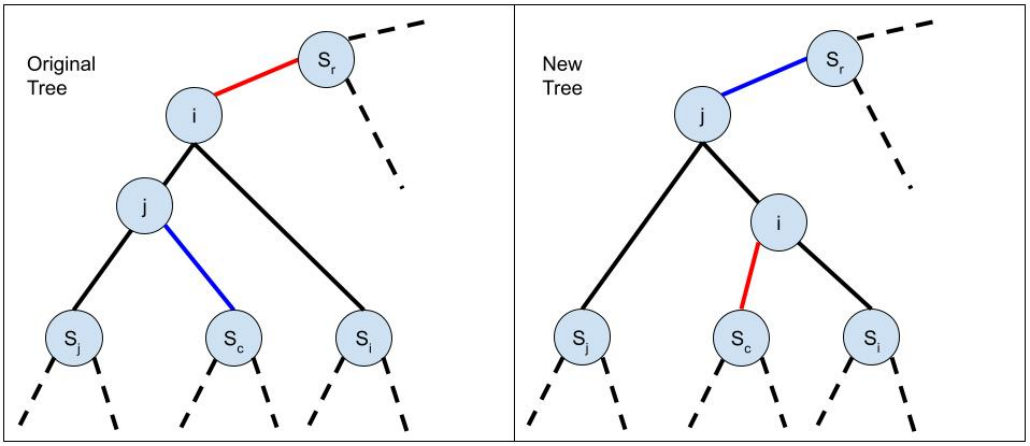


Figure B.1: An illustration of the labelling used in the proof of Lemma B.10. The new tree in the right hand panel is made after swapping subtrees $\mathcal{S}_c$ and $\mathcal{S}_r$.

the fourth subtree is labelled $\mathcal{S}_r$ (this is the subtree containing the root). A single NNI move in this context involves swapping a pair of subtrees from the set $\{\mathcal{S}_i, \mathcal{S}_j, \mathcal{S}_r, \mathcal{S}_c\}$.

By symmetry, swapping the pair $\mathcal{S}_j$ and $\mathcal{S}_c$ or the pair $\mathcal{S}_r$ and $\mathcal{S}_i$ has no impact, as the new tree will be topologically equivalent to $\mathcal{T}$ (which is always in the new shuffled space). Thus, one needs only to prove the result for the tree formed by swapping $\mathcal{S}_c$ and $\mathcal{S}_r$. This is topologically equivalent to the tree in the right panel of Fig. B.1. We assume that this tree is $\mathcal{T}'$

For a given permutation σ, define $l_i(\sigma)$ to be the label assigned to node i, $l_j(\sigma)$ to be the label assigned to node j, $l_i^{\mathcal{S}}(\sigma)$ to be the label assigned to the root of $\mathcal{S}_i$, $l_j^{\mathcal{S}}(\sigma)$ to be the label assigned to the root of $\mathcal{S}_j$ and $l_c^{\mathcal{S}}(\sigma)$ to be the label assigned to the root of $\mathcal{S}_c$

Then, by Lemma B.14,

$$\mathbb{P}\left[l_j^{\mathcal{S}}(Q(\mathcal{T})) = l_i(Q(\mathcal{T}))\right] = \frac{1}{4} \tag{B.43}$$

Now, suppose for a permutation σ that $l_j^{\mathcal{S}}(\sigma) = l_i(\sigma)$. Following the proof of Lemma B.14 (as $r_B = 0$), $l_c^{\mathcal{S}}(\sigma)$ was the smallest available label when processing node j, which must be bigger than the smallest available label when node i was processed (as j is the child of i and hence processed later). Thus, $l_c^{\mathcal{S}}(\sigma) > l_i^{\mathcal{S}}(\sigma)$.

A further important result proved in Lemma B.14 is that node j must have been placed ahead of the root of $\mathcal{S}_i$ in the queue. Thus, all nodes in $\mathcal{S}_i$ are labelled either as $l_i^{\mathcal{S}}(\sigma)$ or with a label that is greater than $l_c^{\mathcal{S}}(\sigma)$, as all internal nodes in $\mathcal{S}_i$ are processed after node j. Moreover, all nodes in $\mathcal{S}_j$ are labelled as either $l_j^{\mathcal{S}}(\sigma)$ or a label greater than $l_c(\sigma)$ as they were processed after node j.

Define $\boldsymbol{v}$ to be the vector giving $\mathcal{T}$ under σ. Then, define a new vector $\boldsymbol{v}'$ by

$$v'_m = \begin{cases} l_i^{\mathcal{S}}(\sigma) & \text{if } m = l_c^{\mathcal{S}}(\sigma) \\ v_m & \text{otherwise} \end{cases} \tag{B.44}$$

Then, $\boldsymbol{v}'$ is ordered as $l_c^{\mathcal{S}}(\sigma) > l_i^{\mathcal{S}}(\sigma)$. We now show that $\boldsymbol{v}'$ generates $\mathcal{T}'$ by using the left-to-right construction algorithm. Note that up to the point that the node labelled $l_c^{\mathcal{S}}(\sigma)$ is processed by this algorithm, no nodes have appended to the edges connecting either the root of $\mathcal{S}_i$ or $\mathcal{S}_j$ to the tree. This holds because all other nodes in $\mathcal{S}_i$ and $\mathcal{S}_j$ have labels greater than $l_c^{\mathcal{S}}(\sigma)$. Changing from $\boldsymbol{v}$ to $\boldsymbol{v}'$ means that the leaf node labelled $l_c^{\mathcal{S}}(\sigma)$ is joined to the edge connecting the root of $\mathcal{S}_i$ rather than the root of $\mathcal{S}_j$. The rest of the tree is then constructed identically as the vectors $\boldsymbol{v}$ and $\boldsymbol{v}'$ are the same.

Thus, changing from $\boldsymbol{v}$ to $\boldsymbol{v}'$ generates the tree in the right panel of Fig. B.1, and therefore the tree formed by the NNI move swapping the trees $\mathcal{S}_c$ and $\mathcal{S}_r$.

This completes the proof as it shows that

$$\mathbb{P}\left[\mathcal{T}' \in \tau\left(Q(\mathcal{T})\right) \middle| l_j^{\mathcal{S}}(\sigma) = l_i(\sigma)\right] = 1 \tag{B.45}$$

and hence, by Lemma B.14

$$\mathbb{P}\left[\mathcal{T}' \in \tau\left(Q(\mathcal{T})\right)\right] \geq \mathbb{P}\left[\mathcal{T}' \in \tau\left(Q(\mathcal{T})\right) \middle| l_j^{\mathcal{S}}(\sigma) = l_i(\sigma)\right] \mathbb{P}\left[l_j^{\mathcal{S}}(\sigma) = l_i(\sigma)\right] = \frac{1}{4} \tag{B.46}$$

B.6 Eutherian mammal phylogeny [369]

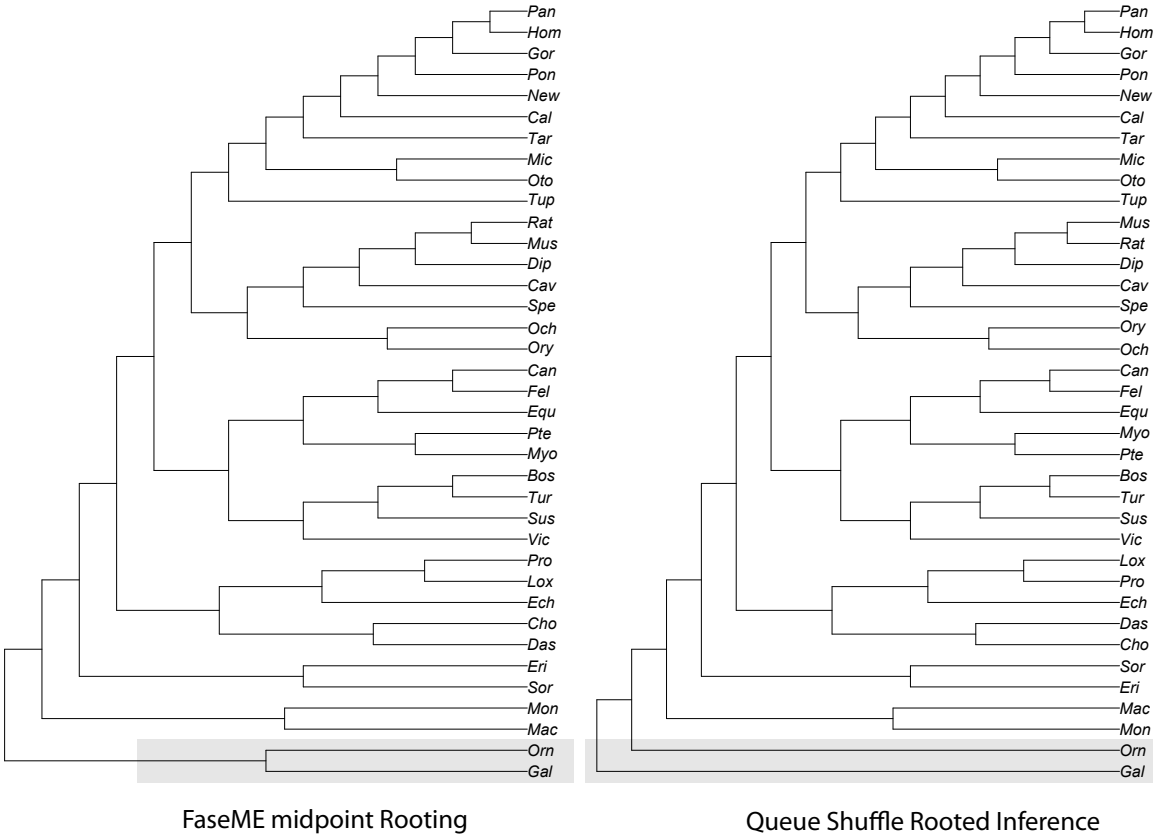


Figure B.2: Comparison of the best unrooted FastME tree that has been midpoint rooted to an optimised rooted tree via Queue Shuffle. Queue Shuffle correctly places *Gallus gallus* as the outgroup of mammals. Branch lengths are ignored and trees are displayed as ultrametric.

B.7 Convergence analysis

Figure B.3 compares the performance of different optimisers (Adafactor [347], AdamW [346], RMSprop [377], and SGD) under different DNA substitutions models (JC69 [374], F81 [375], TN93 [376]) on a single optimisation step (no subsequent reordering with Queue Shuffle) for a maximum of 5000 steps. Four learning rates were considered, logarithmically spaced from 0.001 to 1.0. Whereas convergence speed appears to be independent of the chosen DNA substitution model, the results varied widely with respect to the optimisation algorithm. In particular, Adafactor optimisation produced the best performance, with increasing convergence speed as learning rates were higher.

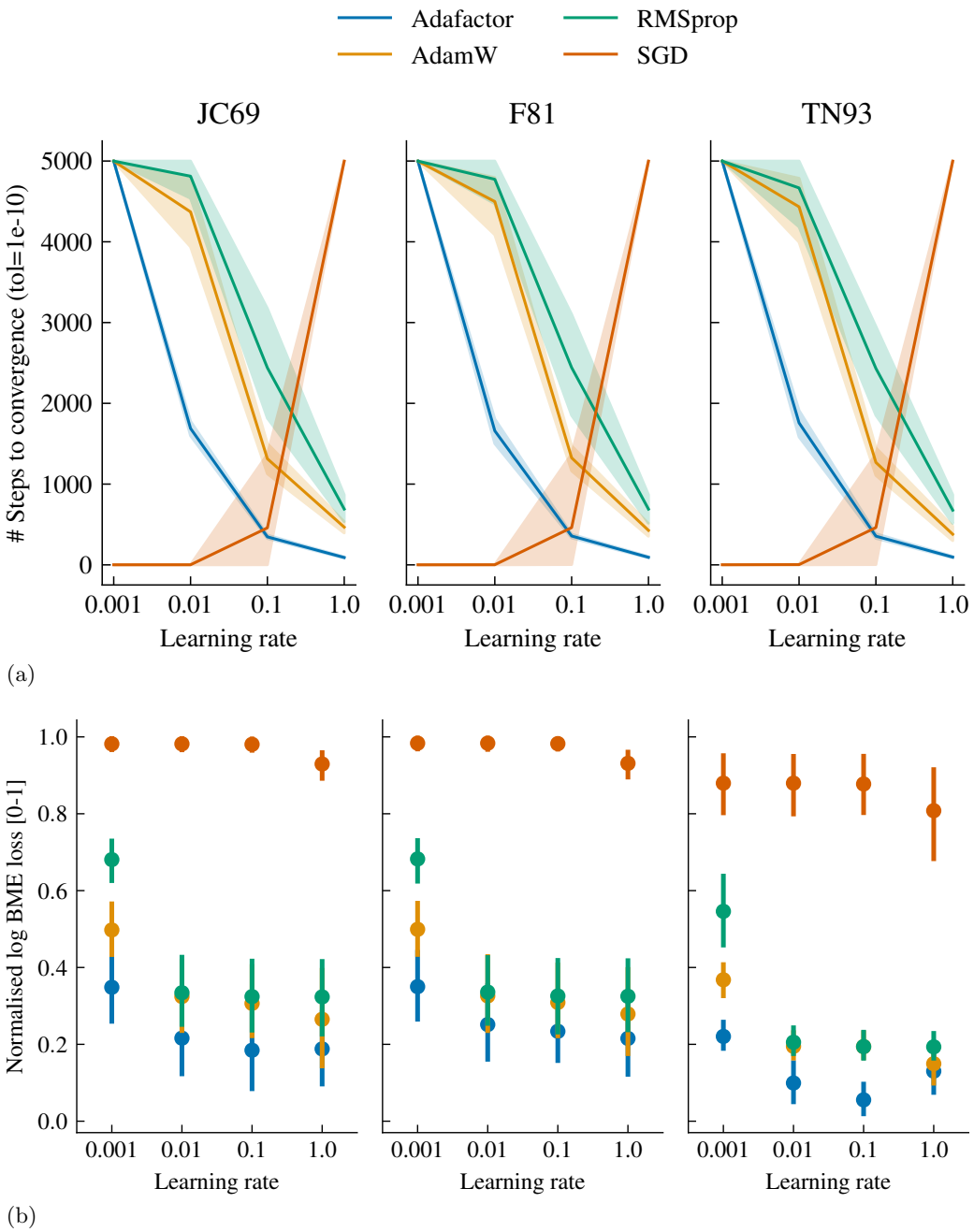


Figure B.3: Convergence analysis of different optimisers and DNA substitution models for the datasets DS1-DS11 (see Table 5.1). **(a)** Number of steps needed to reach convergence (tolerance: 1e-10). **(b)** Loss reached at convergence. For each dataset, the log BME losses were scaled using min-max normalisation. Error bars denote 95% confidence intervals computed with 1000 bootstraps.

B.8 Miscellaneous lemmas

B.8.1 Supplementary lemmas for Lemma B.2

Lemma B.11 *Using the notation of Lemma B.2, define D_{Ar} to be the distance between node A and the root, and g_A to be the path length (i.e. the generation of node A). Then,*

$$\sum_{X_b(A,B)=1} D_{AB}2^{-e_{AB}} = 2\sum_{A} 2^{-g_A} D_{Ar} \tag{B.47}$$

where the e_{AB} terms refer to the path lengths in the original unrooted tree.

Proof: We proceed by induction on the number of leaf nodes, n. For $n = 2$, the root must be placed on the edge connecting nodes 0 and 1 and $e_{01} = 1$. Then

$$\sum_{X_b(A,B)=1} D_{AB}2^{-e_{AB}} = \frac{1}{2}(D_{01} + D_{10}) = \left(2^{1-1}D_{0r} + 2^{1-1}D_{1r}\right) \tag{B.48}$$

as required, noting that $D_{0r} + D_{1r} = D_{01}$. Suppose that our claim holds for $n = k$ and consider a tree with $n = k + 1$. Choose a pair of sibling leaf nodes (i.e. leaf nodes joined to the same internal node) y and z such that the root is not on an edge joining one of these nodes to the tree (this must exist as, for $n > 2$, there are at least one pairs of sibling leaf nodes). Suppose that y and z are connected to the internal node w. Now, note that the contribution to the objective of node y is

$$2\sum_{B:X_b(y,B)=1} D_{yB}2^{-e_{yB}} = 2\sum_{B:X_b(w,B)=1} (D_{wB} + D_{wy})2^{-e_{yB}} \tag{B.49}$$

$$= 2\sum_{B:X_b(w,B)=1} D_{wB}2^{-e_{yB}} + 2D_{wy}\sum_{B:X_b(w,B)=1} 2^{-e_{yB}} \tag{B.50}$$

$$= \sum_{B:X_b(w,B)=1} D_{wB}2^{-e_{wB}} + 2D_{wy}\sum_{B:X_b(w,B)=1} 2^{-e_{yB}} \tag{B.51}$$

where the factor of 2 comes from the fact that both (y, B) and (B, y) must be considered. The final term in this equation can be simplified. Consider the subtree whose leaves are the root and all nodes such that $X_b(w, B) = 1$. Then, by the Kraft Equality on this subtree

$$\sum_{B:X_b(w,B)=1} 2^{-g_B} = \frac{1}{2} \tag{B.52}$$

as g_B gives the distance between the root and node B. Moreover, as $e_{yB} = g_y + g_B - 1$ (where the -1

accounts for the fact that $\mathcal{T}$ is unrooted)

$$\sum_{B:X_b(w,B)=1} 2^{-e_{yB}} = 2^{1-g_y} \sum_{B:X_b(w,B)=1} 2^{-g_B} = 2^{-g_y} \tag{B.53}$$

Thus, (B.51) becomes

$$2 \sum_{X_b(y,B)=1} D_{yB} 2^{-e_{yB}} = \sum_{B:X_b(w,B)=1} D_{wB} 2^{-e_{wB}} + 2D_{wy} 2^{-g_y} \tag{B.54}$$

Thus, the change to $\sum_{X_b(A,B)=1} D_{AB} 2^{-e_{AB}}$ caused by adding nodes y and z to the tree is to subtract

$$2 \sum_{X_b(w,B)=1} D_{wB} 2^{-e_{wB}} \tag{B.55}$$

(as node w is no longer a leaf node) and to add the contributions from nodes y and z

$$2\Bigg(\sum_{B:X_b(y,B)=1} D_{yB} 2^{-e_{yB}} + \sum_{B:X_b(z,B)=1} D_{zB} 2^{-e_{zB}} \Bigg) \tag{B.56}$$

$$= \sum_{X_b(w,B)=1} D_{wB} 2^{-e_{wB}} + 2D_{wy} 2^{-g_y} + \sum_{X_b(w,B)=1} D_{wB} 2^{-e_{wB}} + 2D_{wz} 2^{-g_z} \tag{B.57}$$

There is no (y, z) term to consider as $X_b(y, z) = 0$. Note that the $\sum_{X_b(w,B)=1} D_{wB} 2^{-e_{wB}}$ terms cancel. Define $\mathcal{T}'$ to be the tree when node y and z are removed so, as it now has k leaf nodes, the inductive hypothesis can be used to show

$$\sum_{A,B\in\mathcal{T}:X_b(A,B)=1} D_{AB} 2^{-e_{AB}} = \sum_{A',B'\in\mathcal{T}':X'_b(A',B')=1} D_{A'B'} 2^{-e'_{A'B'}} + 2D_{wy} 2^{-g_y} + 2D_{wz} 2^{-g_z} \tag{B.58}$$

$$= 2 \sum_{A'\in\mathcal{T}'} 2^{-g'_{A'}} D_{A'r'} + 2D_{wy} 2^{-g_y} + 2D_{wz} 2^{-g_z} \tag{B.59}$$

$$= 2 \sum_{A'\in\mathcal{T}'/\{w\}} 2^{-g'_{A'}} D_{A'r'} + 2D_{wr} 2^{-g_w} + 2D_{wy} 2^{-g_y} + 2D_{wz} 2^{-e_{zr}} \tag{B.60}$$

where dashes are used to denote quantities in $\mathcal{T}'$. Note that

$$\sum_{A'\in\mathcal{T}'/\{w\}} 2^{-g'_{A'}} D_{A'r'} = \sum_{A\in\mathcal{T}/\{y,z\}} 2^{-g_A} D_{Ar} \tag{B.61}$$

as only the nodes w, y and z are affected by changing from $\mathcal{T}$ to $\mathcal{T}'$. Moreover, note that

$$2^{-g_w} = 2^{-g_y} + 2^{-g_z} \tag{B.62}$$

as $g_w = g_y - 1 = g_z - 1$, and, as the distances are additive,

$$D_{wy} + D_{wr} = D_{yr} \quad \text{and} \quad D_{wz} + D_{wr} = D_{zr} \tag{B.63}$$

Hence

$$\begin{aligned} 2D_{wy}2^{-g_y} + 2D_{wz}2^{-g_z} + 2D_{wr}2^{-g_w} &= 2D_{wy}2^{-g_y} + 2D_{wz}2^{-g_z} + 2D_{wr}(2^{-g_y} + 2^{-g_z}) && \text{(B.64)} \\ &= 2(D_{wy} + D_{wr})2^{-g_y} + 2(D_{wz} + D_{wr})2^{-g_z} && \text{(B.65)} \\ &= 2D_{yr}2^{-g_y} + 2D_{zr}2^{-g_z} && \text{(B.66)} \end{aligned}$$

Thus,

$$\sum_{X_b(A,B)=1} D_{AB}2^{-e_{AB}} = \sum_A 2^{-g_A} D_{Ar} \tag{B.67}$$

and hence the claim holds by induction as required.

Lemma B.12 *Using the notation of Lemma B.2,*

$$\mathcal{D} = \sum_A 2^{-g_A} D_{Ar} \tag{B.68}$$

Proof: This can be proved through induction on the number of leaf nodes. It clearly holds if there are only two nodes in the tree. Assume it holds whenever there are k leaf nodes, and consider a tree with $k+1$ leaf nodes. We process the first pair of leaf nodes, y and z, and assume that their associated distances to their parent, w are D_{wy} and D_{wz} and that the new tree created is $\mathcal{T}'$ with distances D'. Then, we have, by our inductive hypothesis

$$\begin{aligned} \mathcal{D} &= \sum_{A' \in \mathcal{T}'} 2^{-g'_{A'}} D'_{A'r'} && \text{(B.69)} \\ &= \sum_{A' \in \mathcal{T}'/\{w\}} 2^{-g'_{A'}} D'_{A'r'} + 2^{-g'_w} D'_{wr'} && \text{(B.70)} \\ &= \sum_{A \in \mathcal{T}/\{y,z\}} 2^{-g_A} D_{Ar} + 2^{-g_w}\left(D_{wr} + \frac{D_{wy} + D_{wz}}{2}\right) && \text{(B.71)} \\ &= \sum_{A \in \mathcal{T}/\{y,z\}} 2^{-g_A} D_{Ar} + 2^{-g_w}\left(\frac{D_{yr} + D_{zr}}{2}\right) && \text{(B.72)} \\ &= \sum_{A \in \mathcal{T}} 2^{-g_A} D_{Ar} && \text{(B.73)} \end{aligned}$$

as required, where we have used the fact that $D_{wr} + D_{wy} = D_{yr}$, $D_{wr} + D_{wz} = D_{zr}$ and that $2^{-g_w} = 2 \times 2^{-g_y} = 2 \times 2^{-g_z}$.

B.8.2 Supplementary lemmas for Lemma B.10

Lemma B.13 *Using the notation of Lemma B.10, for a given permutation σ,*

$$\mathbb{P}\left[Q(\mathcal{T}) = \sigma\right] \in \left\{0, \frac{1}{2^{n-1}}\right\} \tag{B.74}$$

Proof Suppose that every (internal and external) node in $\mathcal{T}$ is assigned a fixed unique position, such that one can consistently define a "leftmost" and "rightmost" node from a pair of nodes. Then, the randomness in the Queue Shuffle algorithm can be represented by a uniform random vector $\boldsymbol{r} \in \{0,1\}^{n-1}$ such that when the m^{th} internal node is processed, the leftmost child is given the same label as its parent if and only if $r_{m+1} = 1$. Each distinct $\boldsymbol{r}$ results in a distinct ordering (as, given the leaf labels, one can generate the unique labelling of the tree as a parent must have a label equal to the minimum of the labels of its two children). Hence, each possible labelling is equally likely, giving (B.74) and completing the proof.

Lemma B.14 *Using the notation of Lemma B.10*

$$\mathbb{P}\left[l_j^{\mathcal{S}}(Q(\mathcal{T})) = l_i(Q(\mathcal{T}))\right] = \frac{1}{4} \tag{B.75}$$

Moreover, $l_j^{\mathcal{S}}(Q(\mathcal{T})) = l_i(Q(\mathcal{T}))$ if and only if node j was placed ahead of the root of $\mathcal{S}_i$ in the queue

Proof Using the $\boldsymbol{r}$ defined in the proof of Lemma B.13, define the random indices A and B such that r_A corresponds to processing node i and r_B corresponds to processing node j. Note that r_A and r_B are independent (though B will in general depend on r_A).

Now, suppose without loss of generality that j is the leftmost child of i and that the root of $\mathcal{S}_j$ is the leftmost child of j. As children have labels greater than or equal to their parents,

$$\mathbb{P}\left[l_j^{\mathcal{S}}(Q(\mathcal{T})) = l_i(Q(\mathcal{T}))\right] = \mathbb{P}\left[l_j^{\mathcal{S}}(Q(\mathcal{T})) = l_j(Q(\mathcal{T})) = l_i(Q(\mathcal{T}))\right] \tag{B.76}$$

and hence

$$\mathbb{P}\left[l_j^{\mathcal{S}}(Q(\mathcal{T})) = l_i(Q(\mathcal{T}))\right] = \mathbb{P}(r_A = r_B = 1) = \frac{1}{4} \tag{B.77}$$

which is the first result. The second result follows immediately from the fact that $r_A = 1$.

B.9 Estimation of GTR+Γ distances

To estimate distances under a GTR+Γ substitution model we use the approach outlined in [318]. Assuming a general time reversible rate symmetric matrix Q, the transition probability matrix over time is found via the matrix exponential $P(t) = e^{Qt}$. This matrix exponential can be readily computed via eigendecomposition.

Including variable rates among sites for a Gamma distribution g, $P(t) = \int e^{Qu} g(u) du$, which can again be estimated via eigendecomposition. Given parameters for rates, S, frequencies, π, the time between two sequences t_{ij}, and genetic sequence data $\mathcal{G}$, the log-likelihood for the transitions between a pair of taxa i and j is

$$\mathcal{L}_{ij}(\mathcal{G}|S, \pi, t_{ij}) = \sum_a \sum_b \kappa_{ab}^{ij} \log(P_{ab}(t_{ij}; S, \pi)) \tag{B.78}$$

where κ_{ab}^{ij} is the number of $a \to b$ transitions from taxon i to taxon j. We approximate the optimal parameters by maximizing the total log-likelihood (that is, the sum over i and j of $\mathcal{L}_{ij}$) using gradient descent in Jax.

B.10 Fast discrete hill-climbing with Phylo2Vec

The computational complexity of GradME is substantially higher than that of FastME due to the continuous nature of the algorithm. Thus, particularly for large numbers of taxa, using a similarly fast algorithm, at least to get close to the optimal tree, may be preferable.

Because of this, we have developed an alternative, discrete algorithm which has the same computational complexity as FastME. At each step, our algorithm outputs a matrix Δ such that Δ_{ij} is the change in the objective function if the value of v_i were changed to be equal to j. Δ allows us to perform a hill-climbing optimisation, as the change corresponding to the minimum value of Δ is made.

Analogously to FastME, Δ can be calculated in $\mathcal{O}(n^2 \text{diam}(\mathcal{T}))$ time, where $\text{diam}(\mathcal{T})$ is the maximum inter-taxa path length of the tree (this is generally substantially smaller than n). This algorithm works exclusively for unrooted trees, though a similar algorithm could be developed using our rooted objective.

Motivated by the method of [147], our algorithm begins by calculating directed edge weight vectors $\boldsymbol{w}_e^{\pm}$ for the edges in $\mathcal{T}$. Each edge e naturally partitions $\mathcal{T}$ into two disjoint subtrees, $\mathcal{S}_1^e$ and $\mathcal{S}_2^e$, with each one being rooted at a node connected to e. We suppose that $\mathcal{S}_1^e$ is the tree not containing some

fixed node X. Then, we assign $\boldsymbol{w}_e^+$ to be the balanced distance between the root of $\mathcal{S}_1^e$ and each of the individual leaf nodes in $\mathcal{S}_2^e$, and $\boldsymbol{w}_e^-$ to be the balanced distance between the root of $\mathcal{S}_2^e$ and each of the individual leaf nodes in $\mathcal{S}_1^e$. These edge weights can be calculated efficiently by using an iterative scheme, using the fact that the weight on a given edge from an internal node x to another node y is the mean of the weights of the two edges from nodes z and a (both distinct from y) to x (recalling that these weights are directional). This iterative scheme is closed by the fact that the weights on edges from leaf nodes to the tree are given by the appropriate columns of the distance matrix D.

This weighted tree can be used to calculate the distance between any pair of subtrees. For each fixed subtree, $\mathcal{S}$, one can iteratively calculate the distance $d_{\mathcal{SR}}$ between it and (disjoint) subtrees $\mathcal{R}$, using the fact that the precalculated $\boldsymbol{w}$ terms give the distance between each subtree and each leaf, and that for any pair of subtrees $\mathcal{R}_1$ and $\mathcal{R}_2$ which share a parent node and can therefore be combined in a subtree $\mathcal{R}_3$,

$$d_{\mathcal{SR}_3} = \frac{1}{2}\left(d_{\mathcal{SR}_1} + d_{\mathcal{SR}_2}\right) \tag{B.79}$$

From [147], there exists a simple formula for the difference in objective function caused by pruning a subtree and regrafting it to an adjacent edge. If an internal node x is joined to nodes which are the roots of disjoint subtrees A, B and C, then the difference between attaching a subtree K to the edge joining x and C and attaching the subtree K to the edge joining x and B is

$$\frac{1}{4}\left(\delta_{AB} + \delta_{KC} - \delta_{AC} - \delta_{KB}\right) \tag{B.80}$$

where here, δ denotes inter-subtree distance. These δ terms are not equivalent to the d terms, as removing K from the original tree creates a different set of possible subtree pairs. However, they can be calculated in $\mathcal{O}(n^2\mathrm{diam}(\mathcal{T}))$ time from the distances d following the methods of [147].

Explicitly, our algorithm moves outwards from the parent of the subtree which we are removing. We take B to be the node we are currently processing, x to be the previously processed node on this path, C to the node processed two iterations ago, and A to be the other node connected to x. Then, one has simply

$$\delta_{AB} = d_{AB} \quad \text{and} \quad \delta_{KB} = d_{KB} \tag{B.81}$$

Calculating δ_{AC} is slightly more complicated as the subtree K has been removed from C. If the original root of K was a path length l from C, and if that root shares a parent node (in C) with a subtree F, then

$$\delta_{AC} = d_{AC} + 2^{-l}(d_{AF} - d_{AK}) \tag{B.82}$$

as the subtree F now has double the weight in C. Finally, to calculate d_{CK}, one must calculate the distance between K and all subtrees connected to, but disjoint from the path taken from the tree root to C. If these subtrees are $\mathcal{S}_1, ..., \mathcal{S}_m$ respectively and are distance $1, ..., m$ from C, then

$$\delta_{KC} = \sum_{q=1}^{m} 2^{-q} d_{C\mathcal{S}_q} \tag{B.83}$$

It is this step which pushes the complexity of our algorithm above n^2

Thus, one can calculate the objective differences caused by each subtree-prune and regraft move by repeatedly applying this formula for each pruned subtree, essentially "moving" this subtree around the remaining tree.

The final step of our algorithm is to find the corresponding subtree-prune and regraft move for each possible change of $\boldsymbol{v}$. When v_i is changed to be equal to j, the subtree that is moved is the largest subtree of $\mathcal{T}$ containing i such that the leaves all have labels greater than or equal to i (in many cases, this will simply be the node i). For each value of i, one can find the edges which were connected to each leaf node when node i was attached in the left-to-right construction algorithm. The edge connected to node j at this stage gives the location of the regraft position of the subtree.

B.11 Code: continuous BME objective function

```python
import jax.numpy as np

from jax import jit, lax
from jax.scipy.special import logsumexp

@jit
def get_edges_exp_log(W, rooted):
    """Calculate the log-expectation of the objective value of a tree drawn with distribution W.
    We calculate and update E_ij throughout the left-to-right construction procedure

    Args:
        W (jax.numpy.array): Tree distribution
        rooted (bool): True is the tree is rooted, otherwise False

    Returns:
        E (jax.numpy.array): Log of the expected objective value of a tree drawn with W
    """
    # Add jnp.finfo(float).eps to W.tmp to avoid floating point errors with float32
    W_tmp = (
        jnp.pad(W, (0, 1), constant_values=jnp.finfo(float).eps) + jnp.finfo(float).eps
    )

    n_leaves = len(W) + 1

    E = jnp.zeros((n_leaves, n_leaves))
    E = E.at[1, 0].set(
        0.5 * E[0, 0] * W_tmp[0, 0] + jnp.log(0.25 * (2 - rooted) * W_tmp[0, 0])
    )
    E = E + E.T

    trindx_x, trindx_y = jnp.tril_indices(n_leaves - 1, -1)

    def body(carry, _):
        E, i = carry

        E_new = jnp.zeros((n_leaves, n_leaves))

        trindx_x_i = jnp.where(trindx_x < i, trindx_x, 1)
        trindx_y_i = jnp.where(trindx_x < i, trindx_y, 0)

        indx = (trindx_x_i, trindx_y_i)

        E_new = E_new.at[indx].set(
            E[indx] + jnp.log(
                1 + jnp.finfo(float).eps - 0.5 * (W_tmp[i - 1, indx[1]] + W_tmp[i - 1, indx[0]])
            )
        )

        # exp array
        mask_Ei = jnp.where(jnp.arange(n_leaves) >= i, 0, 1)
        exp_array = E * jnp.where(jnp.arange(n_leaves) >= i, 0, mask_Ei.T)

        # coef array
        mask_Wi = jnp.where(jnp.arange(n_leaves) >= i, 0, 0.5 * W_tmp[i - 1])
        coef_array = (jnp.zeros_like(W_tmp) + mask_Wi).at[:, i].set(0.25 * W_tmp[i - 1])
        coef_array = coef_array * (1 - jnp.eye(W_tmp.shape[0]))
```

```
        # logsumexp
        tmp = logsumexp(exp_array, b=coef_array, axis=-1) * mask_Ei

        E_new = E_new.at[i, :].set(tmp)

        # Update E
        E = E_new + E_new.T

        return (E, i + 1), None

    (E, _), _ = lax.scan(body, (E, 2), None, length=n_leaves - 2)

    return E

@jit
def bme_loss_log(W, D, rooted):
    """Log version of the BME loss function

    Args:
        W (jax.numpy.array): Tree distribution
        D (jax.numpy.array): Distance matrix
        rooted (bool): True is the tree is rooted, otherwise False

    Returns:
        loss (float): BME loss
    """
    E = get_edges_exp_log(W, rooted)
    loss = logsumexp(E, b=D)
    return loss
```

Appendix - Paper III

C.1 Connection of entropic likelihood to BME

In this section, we show that the percentage error in the BME approximation is small in the $\frac{K}{\theta} << 1$ limiting case. We do this with an exponential tree model, using the results of Lemma C.4.

Theorem C.1 *Define*

$$K = \sum_{a \neq b} \pi_a Q_{ab} \tag{C.1}$$

Consider taking $Kt \to 0$ while keeping the stationary distribution π and the ratios $\frac{Q_{ab}}{Q_{cd}}$ constant (i.e. one can change t or scale the matrix Q by a constant multiple).

For some fixed k_1 and variable $k_2 > k_1$ define a linear BME approximation to the entropic likelihood $f(k_2)$ to be

$$f(k_2) = \mathbb{E}\left\{ D_{ij}^S\left(\frac{k_1}{\theta}\right)\right\} + \frac{k_2 - k_1}{\theta}\left(\int_0^\infty S(s)\theta^2 e^{-\theta s} ds\right) \tag{C.2}$$

Then,

$$\frac{\mathbb{E}\left\{ D_{ij}^S\left(\frac{k_2}{\theta}\right)\right\} - \mathbb{E}\left\{ D_{ij}^S\left(\frac{k_1}{\theta}\right)\right\}}{f(k_2) - \mathbb{E}\left\{ D_{ij}^S\left(\frac{k_1}{\theta}\right)\right\}} = 1 + \mathcal{O}\left(\frac{1}{\log\left(\frac{K}{\theta}\right)}\right) \tag{C.3}$$

and, moreover, the percentage error is small

$$\frac{f(k_2) - \mathbb{E}\left\{ D_{ij}^S\left(\frac{k_2}{\theta}\right)\right\}}{f(k_2)} = \mathcal{O}\left(\frac{1}{\log\left(\frac{K}{\theta}\right)}\right) \tag{C.4}$$

Proof: We begin with the true entropy difference between $\frac{k_1}{\theta}$ and $\frac{k_2}{\theta}$, which is

$$\mathbb{E}\left\{ D_{ij}^S\left(\frac{k_2}{\theta}\right)\right\} - \mathbb{E}\left\{ D_{ij}^S\left(\frac{k_1}{\theta}\right)\right\} = \int_{\frac{k_1}{\theta}}^{\frac{k_2}{\theta}} \left\{ e^{-\theta t}(S'(t) + \theta S(t)) + \int_0^t S(s)\theta^2 e^{-\theta s} ds \right\} dt \tag{C.5}$$

This can be rewritten as

$$\mathbb{E}\left\{D_{ij}^S\left(\frac{k_2}{\theta}\right)\right\} - \mathbb{E}\left\{D_{ij}^S\left(\frac{k_1}{\theta}\right)\right\} \tag{C.6}$$

$$= \frac{k_2-k_1}{\theta}\left(\int_0^\infty S(s)\theta^2 e^{-\theta s}ds\right) + \int_{\frac{k_1}{\theta}}^{\frac{k_2}{\theta}}\left\{e^{-\theta t}(S'(t)+\theta S(t)) - \int_t^\infty S(s)\theta^2 e^{-\theta s}ds\right\}dt \tag{C.7}$$

which, after twice integrating $\int_t^\infty S(s)\theta^2 e^{-\theta s}ds$ by parts (integrating the exponential and differentiating the entropy), is

$$\mathbb{E}\left\{D_{ij}^S\left(\frac{k_2}{\theta}\right)\right\} - \mathbb{E}\left\{D_{ij}^S\left(\frac{k_1}{\theta}\right)\right\} = \frac{k_2-k_1}{\theta}\left(\int_0^\infty S(s)\theta^2 e^{-\theta s}ds\right) - \int_{\frac{k_1}{\theta}}^{\frac{k_2}{\theta}}\int_t^\infty S''(s)e^{-\theta s}dsdt \tag{C.8}$$

$$= f(k_2) - \mathbb{E}\left\{D_{ij}^S\left(\frac{k_1}{\theta}\right)\right\} - \int_{\frac{k_1}{\theta}}^{\frac{k_2}{\theta}}\int_t^\infty S''(s)e^{-\theta s}dsdt \tag{C.9}$$

Thus,

$$\frac{\mathbb{E}\left\{D_{ij}^S\left(\frac{k_2}{\theta}\right)\right\} - \mathbb{E}\left\{D_{ij}^S\left(\frac{k_1}{\theta}\right)\right\}}{f(k_2) - \mathbb{E}\left\{D_{ij}^S\left(\frac{k_1}{\theta}\right)\right\}} = 1 - \frac{\int_{\frac{k_1}{\theta}}^{\frac{k_2}{\theta}}\int_t^\infty S''(s)e^{-\theta s}dsdt}{\frac{k_2-k_1}{\theta}\times\theta^2\int_0^\infty S(s)e^{-\theta s}ds} \tag{C.10}$$

and so we simply need to justify that

$$\left|\frac{\int_{\frac{k_1}{\theta}}^{\frac{k_2}{\theta}}\int_t^\infty S''(s)e^{-\theta s}dsdt}{\frac{k_2-k_1}{\theta}\times\theta^2\int_0^\infty S(s)e^{-\theta s}ds}\right| = \mathcal{O}\left(\frac{1}{\log\left(\frac{K}{\theta}\right)}\right) \tag{C.11}$$

To begin, we consider the integral in the denominator. Note that

$$\int_0^\infty \theta S(s)e^{-\theta s}ds = \int_0^\infty S\left(\frac{\tau}{\theta}\right)e^{-\tau}d\tau \tag{C.12}$$

Now, choose some C, and note that (as $S(t)$ must be bounded between 0 and $-\log(N_s)$, where N_s is the number of states)

$$\left|\int_C^\infty S\left(\frac{\tau}{\theta}\right)e^{-\tau}d\tau\right| \leq -\log(N_s)e^{-C} \tag{C.13}$$

and so, provided $C >> 1$, this is exponentially small.

Moreover, in the region $\frac{KC}{\theta} << 1$, we can use Lemma C.8 to approximate

$$\int_0^C S\left(\frac{\tau}{\theta}\right)e^{-\tau}d\tau = \int_0^C\left\{\frac{K\tau}{\theta}\log\left(\frac{K\tau}{\theta}\right) + \mathcal{O}\left(\frac{K\tau}{\theta}\right)\right\}e^{-\tau}d\tau \tag{C.14}$$

The two conditions on C can be simultaneously satisfied by, for example, setting $C = \left(\frac{K}{\theta}\right)^{-0.5}$.

Now, note that

$$\int_0^C \mathcal{O}\left(\frac{K\tau}{\theta}\right) e^{-\tau} d\tau = \mathcal{O}\left(\frac{K}{\theta}\right) \int_0^C \tau e^{-\tau} d\tau = \mathcal{O}\left(\frac{K}{\theta}\right) \tag{C.15}$$

and hence, the leading order term in C.14 is

$$\int_0^\infty \theta S(s) e^{-\theta s} ds \sim \frac{K\tau}{\theta} \log\left(\frac{K\tau}{\theta}\right) \int_0^C e^{-\tau} d\tau \sim \frac{K\tau}{\theta} \log\left(\frac{K\tau}{\theta}\right) \tag{C.16}$$

again using the fact that e^{-C} is exponentially small.

Now, to simplify the numerator of C.11, we begin by considering the inner integral evaluated at some possible value of t, which we set to be $\frac{k}{\theta}$ for $k \in (k_1, k_2)$ (and therefore $k = \mathcal{O}(1)$). Now,

$$\int_{\frac{k}{\theta}}^\infty \theta S''(s) e^{-\theta s} ds = \int_k^\infty S''\left(\frac{\tau}{\theta}\right) e^{-\tau} d\tau \tag{C.17}$$

Again, we can solve this problem by splitting the integral. Note that, using Lemma C.7, $S''(s) \to 0$ as $s \to \infty$, so it must be bounded by some $\mathcal{S}$ in the region $[k, \infty)$. Defining

$$X = \left(\frac{K}{\theta}\right)^{-0.5} \tag{C.18}$$

we therefore see that

$$\left| \int_X^\infty S''\left(\frac{\tau}{\theta}\right) e^{-\tau} d\tau \right| \leq \mathcal{S} e^{-X} \tag{C.19}$$

which is exponentially small as $X \to 0$. Moreover, in the region $\tau < X$, we have $\frac{K\tau}{\theta} << 1$ and so, using Lemma C.8, $S''(t) \sim \frac{K}{t}$. Thus,

$$\left| \int_k^X S''\left(\frac{\tau}{\theta}\right) e^{-\tau} d\tau \right| \sim \left| \int_k^X \frac{K\theta}{\tau} e^{-\tau} d\tau \right| \leq \left| \frac{K\theta}{k_1} \int_0^\infty e^{-\tau} d\tau \right| = \frac{K\theta}{k_1} \tag{C.20}$$

Therefore, using these results in C.11, and noting that we have multiplied each of the integrals by θ in their consideration (which of course cancels)

$$\left| \frac{\int_{\frac{k_1}{\theta}}^{\frac{k_2}{\theta}} \int_t^\infty S''(s) e^{-\theta s} ds dt}{\frac{k_2 - k_1}{\theta} \times \theta^2 \int_0^\infty S(s) e^{-\theta s} ds} \right| \lesssim \left| \frac{\int_{\frac{k_1}{\theta}}^{\frac{k_2}{\theta}} \frac{K\theta}{k_1} dt}{\frac{k_2 - k_1}{\theta} \times \theta^2 \frac{K}{\theta} \log\left(\frac{K}{\theta}\right)} \right| = \frac{1}{k_1 \log\left(\frac{K}{\theta}\right)} \tag{C.21}$$

as required for equation C.3. It is worth noting that, while k_1 is fixed throughout this theorem, the result would not hold in a $k_1 \to 0$ limit as the approximation breaks down for extremely small branch lengths.

The final result follows as rearranging C.3 shows that

$$\frac{f(k_2) - \mathbb{E}\left\{D_{ij}^S\left(\frac{k_2}{\theta}\right)\right\}}{f(k_2)} = \left(\frac{\mathbb{E}\left\{D_{ij}^S\left(\frac{k_1}{\theta}\right)\right\}}{f(k_2)} - 1\right)\mathcal{O}\left(\frac{1}{\log\left(\frac{K}{\theta}\right)}\right) \tag{C.22}$$

and so, using the fact that the linear approximation is monotonic, $|f(k_2)| > \left|\mathbb{E}\left\{D_{ij}^S\left(\frac{k_1}{\theta}\right)\right\}\right|$ and the result follows.

C.2 Supporting Lemmas

C.2.1 Independence of simulation root

We claim that the distribution of sites on the tree is independent of the choice of simulation root. Define $p(g, r)$ to be the probability of the site pattern g given that the simulation root is r.

Lemma C.1 *If $\mathcal{N}$ is the set of nodes in $\mathcal{T}$,*

$$p(g, r) = p(g, s) \quad \forall r, s \in \mathcal{N} \tag{C.23}$$

Proof: As before, define the set of edges to be $\mathcal{E} = \{(e_i^1, e_i^2)|i = 0, 1, ..\}$, where e_i^1 and e_i^2 are the nodes which this edge connects, such that e_i^1 is the closer to the simulation root. Suppose that z_i^j is the site value on node e_i^j and that b_i is the length of edge (e_i^1, e_i^2).

Suppose first that r and s are adjacent, so that they share an edge. Suppose that when the root is r, this edge is labelled as e_1 in E. Then, note that, using detailed balance,

$$\pi_{z_r} P_{z_1^1, z_1^2}(t_i) = \pi_{z_r} P_{z_r, z_s}(t_i) = \pi_{z_s} P_{z_s, z_r}(t_i) \tag{C.24}$$

When the root is s, it is also possible to label this edge as e_1, but now $z_1^1 = s$ and $z_1^2 = r$. Thus,

$$p(g, s) = \pi_{z_s} P_{z_1^2, z_1^1}(t_i) \prod_{i \geq 2} P_{z_i^1, z_i^2}(t_i) = \pi_{z_r} P_{z_r, z_s}(t_i) \prod_{i \geq 2} P_{z_i^1, z_i^2}(t_i) = p(g, r) \tag{C.25}$$

and so the likelihood is unchanged.

For any pair of non-adjacent nodes, r and s, one can find the path of nodes $y_1, .., y_m$ between them. Then, $p(g, r) = p(g, s)$ follows from the fact that $p(g, r) = p(g, y_1) = p(g, y_2)... = p(g, s)$.

C.2.2 Pairwise distributions

Lemma C.2 *Let M_t be a stationary reversible substitution CTMC. For a given pair of taxa, x and y, which are distance t apart in a (known) tree, $\mathcal{T}$, define V_x and V_y to be the values of a particular state in those two taxa. Then,*

$$(M_t, M_0) \stackrel{d}{=} (V_x, V_y) \tag{C.26}$$

Proof: Considering r as a root, one can define the *most recent common ancestor*, m, of x and y to be the node furthest from r that is on both the path between x and r and the path between y and r. (Note that it is possible that $m = r$).

Define $\mathcal{V}$ to be the set of possible states for a node, and define π_s to be the stationary distribution. Then, using V_r and V_m to be the state at r and m respectively,

$$\mathbb{P}(V_x = a, V_y = b) = \sum_{c \in \mathcal{V}} \mathbb{P}(V_x = a, V_y = b | V_r = c) \mathbb{P}(V_r = c) \tag{C.27}$$

$$= \sum_{c \in \mathcal{V}} \mathbb{P}(V_x = a, V_y = b | V_r = c) \pi_c \tag{C.28}$$

$$= \sum_{c,d \in \mathcal{V}} \mathbb{P}(V_x = a, V_y = b | V_m = d, V_r = c) \mathbb{P}(V_m = d | V_r = c) \pi_c \tag{C.29}$$

$$= \sum_{c,d \in \mathcal{V}} \mathbb{P}(V_x = a, V_y = b | V_m = d) \mathbb{P}(V_m = d | V_r = c) \pi_c \tag{C.30}$$

Now, note that (defining t_r to be the distance between r and m), using the detailed balance equations gives

$$\sum_{c \in \mathcal{V}} \mathbb{P}(V_m = d | V_r = c) \pi_c = \sum_{c \in \mathcal{V}} P_{cd}(t_r) \pi_c = \sum_{c \in \mathcal{V}} P_{dc}(t_r) \pi_d = \pi_d \tag{C.31}$$

and so,

$$\mathbb{P}(V_x = a, V_y = b) = \sum_{d \in \mathcal{V}} \mathbb{P}(V_x = a, V_y = b | V_m = d) \pi_d \tag{C.32}$$

Now, the substitution process on the path from m to a is independent of the substitution process on the path from m to b. Thus, if the distance from m to a is t_a and the distance from m to b is t_b,

$$\mathbb{P}(V_x = a, V_y = b) = \sum_{d \in \mathcal{V}} \pi_d P_{da}(t_a) P_{db}(t_b) \tag{C.33}$$

Using detailed balance shows

$$\mathbb{P}(V_x = a, V_y = b) = \sum_{d \in \mathcal{V}} \pi_a P_{ad}(t_a) P_{db}(t_b) \tag{C.34}$$

Finally, note that

$$\pi_a \sum_{d \in \mathcal{V}} P_{ad}(t_a) P_{db}(t_b) = \mathbb{P}(M_0 = a) \sum_{d \in \mathcal{V}} \mathbb{P}(M_{t_a} = d | M_0 = a) \mathbb{P}(M_{t_a + t_b} = b | M_{t_a} = d) \tag{C.35}$$

$$= \mathbb{P}(M_0 = a) \mathbb{P}(M_{t_a + t_b} = b | M_0 = a) \tag{C.36}$$

$$= \mathbb{P}(M_0 = a, M_{t_a + t_b} = b) \tag{C.37}$$

and the result follows from the fact that $t_a + t_b = t$.

C.2.3 Entropic distance for a general branching process

Suppose that a tree $\mathcal{T}$ is generated as follows. Begin by choosing a total tree time T, and start the tree at a root node ρ (which can ultimately be removed from the tree to make it unrooted). We assume that two iid trees are connected by ρ. For each of these trees, we simulate a branch of random length T_0 (indepedently for each tree) according to some pdf f, with cumulative distribution function (cdf) F. If $T < T_0$, then we create a leaf node at a distance T from the root and terminate the generation of this tree. Otherwise, if $T_0 < T$, then we create an internal node at a distance T_0 from the root. We then generate two independent trees rooted at this internal node, but with a total time of $T - T_0$. The construction process stops when there are no more trees that need to be generated.

Lemma C.3 *Given two taxa i and j which are a distance of 2τ apart on $\mathcal{T}$, the expected entropic distance, $\mathbb{E}(D_{ij}^S) := 2h(\tau)$ satisfies the renewal equation*

$$\mathbb{E}(D_{ij}^S) = 2h(\tau) = 2(1 - F(\tau))S(\tau) + 2\int_0^\tau (h(\tau - t) + S(t)) f(t) dt \tag{C.38}$$

Proof: Define a to be the most recent common ancestor of i and j (that is, the node furthest from ρ which is on the path from both i to ρ and j to ρ). As the total distance from each leaf node to the root is fixed, it is necessary that both the distance from a to i and the distance from a to j is τ.

The two subtrees rooted at a are therefore iid, and so $\mathbb{E}(D_{ij}^S) = \frac{\mathbb{E}(D_{ia}^S)}{2}$. We can calculate $\mathbb{E}(D_{ia}^S)$ using the self-similarity property of the tree-generation process. Conditioning on the first branch length, T_0, of this tree

$$h(\tau) := \mathbb{E}(D_{ia}^S) = \int_0^\infty \mathbb{E}(D_{ia}^S | T_0 = t) f(t) dt \tag{C.39}$$

If $T_0 > \tau$, then this subtree has only one branch, and therefore $D^S_{ia} = S(\tau)$. Otherwise, we get an entropy of $S(T_0)$, and then get the entropy of a subtree of length $(\tau - T_0)$. Thus,

$$h(\tau) = (1 - F(\tau))S(\tau) + \int_0^\tau (h(\tau - t) + S(t))f(t)dt \tag{C.40}$$

as required.

C.2.4 Entropic distance for a Markovian branching process

Lemma C.4 *With the setup of Lemma C.4, if the branch lengths are exponentially distributed with mean $\frac{1}{\theta}$, then*

$$\mathbb{E}(D^S_{ij}) = 2\left(\int_0^\tau \left[e^{-\theta t}(S'(t) + \theta S(t)) + \int_0^t S(s)\theta^2 e^{-\theta s} ds \right] dt \right) \tag{C.41}$$

Proof: In this case, (C.40) can be rewritten as

$$h(\tau) = e^{-\theta\tau}S(\tau) + \int_0^\tau (h(\tau - t) + S(t))\theta e^{-\theta t} dt \tag{C.42}$$

$$= e^{-\theta\tau}S(\tau) + \int_0^\tau h(t)\theta e^{-\theta(\tau - t)} dt + \int_0^\tau S(t)\theta e^{-\theta t} dt \tag{C.43}$$

and hence

$$h(\tau)e^{\theta\tau} = S(\tau) + \int_0^\tau h(t)\theta e^{\theta(t)} dt + e^{\theta\tau} \int_0^\tau S(t)\theta e^{-\theta t} dt \tag{C.44}$$

Differentiating for $\tau > 0$ (to ensure that S is differentiable) gives

$$e^{\theta\tau}(h'(\tau) + \theta h(\tau)) = S'(\tau) + h(\tau)\theta e^{\theta\tau} + \theta S(\tau) + \theta e^{\theta\tau} \int_0^\tau S(t)\theta e^{-\theta t} dt \tag{C.45}$$

Hence,

$$h'(\tau) = e^{-\theta\tau}(S'(\tau) + \theta S(\tau)) + \int_0^\tau \theta^2 S(t) e^{-\theta t} dt \tag{C.46}$$

and so, using Lemma C.8 to show that the singularity of $S'(t)$ at $t = 0$ is logarithmic, and therefore integrable

$$h(\tau) = \int_0^\tau \left[e^{-\theta t}(S'(t) + \theta S(t)) + \int_0^t S(s)\theta^2 e^{-\theta s} ds \right] dt \tag{C.47}$$

C.2.5 Maximum likelihood estimation of θ for a Markovian branching process

Lemma C.5 *Given the branching process in Lemma C.4 and a total tree length of $\mathcal{L}$ the maximum likelihood estimator $\hat{\theta}$ of θ is*

$$\hat{\theta} = \frac{n - 2}{\mathcal{L}} \tag{C.48}$$

Proof: We can construct our tree using a "time-to-next-event" construction. That is, we consider the events $\mathcal{E}_1, \mathcal{E}_2, ...$ where a branch terminates, or when the time reaches T_0 and the process ends, in increasing order of the time at which they terminate. We use l_k to denote the time at which $\mathcal{E}_k$ occurs.

Using the memoryless property of the exponential distribution, immediately after $\mathcal{E}_k$, we have $k+2$ active branches and hence

$$l_{k+1} \sim l_k + \min\Big(T_0 - l_k, \text{Exp}((k+2)\theta)\Big) \tag{C.49}$$

We know that, as we have n leaf nodes, there must have been $n-2$ splitting events up to time T_0. For a given set of times l_k at which these events happened, we have a probability density function (pdf), $f(\boldsymbol{l})$

$$f(\boldsymbol{l}) = \Big[\prod_{i=1}^{n-2}(i+1)\theta e^{-(i+1)\theta(l_i - l_{i-1})}\Big] e^{-n\theta(T_0 - l_{n-2})} \tag{C.50}$$

This is proportional to

$$g(\boldsymbol{l}) = \theta^{n-2}\exp\Big[-\sum_{i=1}^{n-2}(i+1)\theta(l_i - l_{i-1}) - n\theta(T_0 - l_{n-2})\Big] \tag{C.51}$$

Now, we know that the total length of the tree is

$$\mathcal{L} = \sum_{i=1}^{n-2}(i+1)(l_i - l_{i-1}) + (T_0 - l_{n-2})n \tag{C.52}$$

by summing the $(i+1)$ branch lengths between events $\mathcal{E}_i$ and $\mathcal{E}_{i-1}$. Thus,

$$g(\boldsymbol{l}) = \theta^{n-2}\exp\Big[-\theta\mathcal{L}\Big] \tag{C.53}$$

We can find the maximiser of this by noting

$$\log(g) = (n-2)\log(\theta) - \mathcal{L}\theta \tag{C.54}$$

and hence, after differentiating

$$\hat{\theta} = \frac{n-2}{\mathcal{L}} \tag{C.55}$$

as required.

C.2.6 Varying the sampling distribution

Lemma C.6

$$H(X,Y|\mathcal{T},\boldsymbol{\beta}) \approx \kappa_2 H(X,Y|\mathcal{U},\boldsymbol{b}^*) \tag{C.56}$$

for some $\kappa_2 \in [1,2]$

"Proof": We do not prove this rigorously, but instead provide a rough justification as to why we expect this result to hold.

We define p_i to be the probability of sequence i in the true tree $\mathcal{T}$ and q_i to be the probability in $\mathcal{U}$

Rewriting the equation stated in the lemma in these terms, we seek to show

$$\sum_i p_i \log(q_i) \approx \kappa_2 \sum_i q_i \log(q_i) \tag{C.57}$$

Now, consider the solution to

$$\max_{\boldsymbol{q}} \{ \sum_i p_i \log(q_i) \Big| \sum_i q_i \log(q_i) = C, \sum_i q_i = 1 \} \tag{C.58}$$

When $\mathcal{U}$ is close to $\mathcal{T}$, we know that finding the balanced minimum evolution tree lengths will make q_i approximate p_i, and should therefore approximately maximise $\sum_i p_i \log(q_i)$, given that the entropy of $\sum_i q_i \log(q_i)$ has increased to some constant C. This approximation is not perfect, and so $\boldsymbol{q}$ will not exactly solve this optimization problem, but it is a good estimate with which we can examine the effect on our entropies. To solve this problem we consider a Lagrange multiplier to get

$$\mathcal{L}(\boldsymbol{q},\lambda,\mu) = \sum_i (p_i - \lambda q_i)\log(q_i) - \mu q_i \tag{C.59}$$

Differentiating and setting to zero gives

$$\frac{p_i}{q_i} - \lambda(1 + \log(q_i)) = \mu \tag{C.60}$$

Multiplying by q_i

$$p_i - \lambda q_i - \lambda q_i \log(q_i) - \mu q_i = 0 \tag{C.61}$$

and summing over i gives an equation

$$1 - \lambda - \lambda C - \mu = 0 \tag{C.62}$$

and so

$$\mu = 1 - (C+1)\lambda \tag{C.63}$$

and hence

$$p_i - \lambda q_i - \lambda q_i \log(q_i) - (1 - (C+1)\lambda)q_i = 0 \tag{C.64}$$

so

$$p_i = \lambda q_i \log(q_i) + (1 - C\lambda)q_i = q_i(\lambda \log(q_i) + (1 - C\lambda)) \tag{C.65}$$

From here, we work in the case where $|\mathcal{G}|$ is large and suppose that $C = \phi \log(\frac{1}{|\mathcal{G}|})$ for $\phi \in (0,1)$ and $\phi = \mathcal{O}(1)$ (noting that the maximal entropy is $\log(\frac{1}{|\mathcal{G}|})$). Thus, $|C| >> 1$ also.

We know that for each i, as $p_i \geq 0$

$$q_i \geq \exp\left[C - \frac{1}{\lambda}\right] \tag{C.66}$$

If this held to equality, summing over i would yield

$$\exp\left[C - \frac{1}{\lambda}\right] = \frac{1}{|\mathcal{G}|} \tag{C.67}$$

We therefore instead suppose that there exists some $\gamma = \mathcal{O}(1)$ such that

$$\exp\left[C - \frac{1}{\lambda}\right] = \left(\frac{1}{|\mathcal{G}|}\right)^{\gamma} \tag{C.68}$$

which means

$$\lambda = \frac{1}{C - \gamma \log(\frac{1}{|\mathcal{G}|})} = \frac{1}{C(1 - \gamma\phi)} \tag{C.69}$$

Now, multiplying our equation by $\frac{p_i}{q_i}$, and summing over i, we have

$$\sum_i \frac{p_i^2}{q_i} = \lambda \sum_i p_i \log(q_i) + (1 - C\lambda) \tag{C.70}$$

We suppose that our trees are close so that $\sum_i \frac{p_i^2}{q_i} = \mathcal{O}(1)$ (which follows as $\frac{p_i}{q_i} \approx 1$ and hence we are left with approximately $\sum_i p_i = 1$). Ignoring this term then yields

$$\sum_i p_i \log(q_i) \approx C + \frac{1}{\lambda} = C(2 - \gamma\phi) \tag{C.71}$$

Note that the case $\gamma = \phi = 1$ means that C is the maximal entropy for this distribution and therefore all the q_i are equal and so, as expected, we would get $\sum_i p_i \log(q_i) = C$. However, in general, we

expect $\gamma\phi \in (0,1)$ (noting that, necessarily $|\sum_i p_i \log(q_i)| \geq |\sum_i p_i \log(p_i)|$).

This is the required result, as the left-hand-side (under this approximate construction) is equal to $H(X,Y|\mathcal{U},\boldsymbol{b}^*)$ and the right-hand-side equal to $H(X,Y|\mathcal{T},\boldsymbol{\beta})$. Hence, we expect that

$$\kappa_2 = C(2-\gamma\phi) \in (1,2) \tag{C.72}$$

Of course, γ and ϕ are functions of C, but provided we only make small changes to C, we expect them to vary slowly, and so the required linear formula will approximately hold for close trees.

C.2.7 Large time behaviour of $S''(t)$

Lemma C.7

$$\lim_{t\to\infty}(S''(t)) = 0 \tag{C.73}$$

Proof: First, note that, as the substitution CTMC is reversible,

$$\pi_i Q_{ij} = \pi_j Q_{ji} \Rightarrow \sqrt{\pi_i} Q_{ij} \frac{1}{\sqrt{\pi_j}} = \frac{1}{\sqrt{\pi_i}} Q_{ji} \sqrt{\pi_j} \tag{C.74}$$

Hence, the matrix $\tilde{Q}$ given by

$$\tilde{Q}_{ij} = \sqrt{\pi_i} Q_{ij} \frac{1}{\sqrt{\pi_j}} \tag{C.75}$$

is symmetric and therefore diagonalisable by Spectral Theorem. Noting that

$$\tilde{Q} = \Pi^{\frac{1}{2}} Q \Pi^{-\frac{1}{2}} \tag{C.76}$$

where Π is a diagonal matrix with entries π_i, $\tilde{Q}$ is similar to Q, the matrix Q is therefore diagonalisable. Now, applying this to the forward equations gives

$$P(t) = \exp(Qt) = M^{-1} e^{Dt} M P(0) \tag{C.77}$$

for some matrix M and a diagonal matrix D. Hence, $P(t)$ is a weighted sum of exponentials, and its asymptotic behaviour is controlled by the leading order term. As, by the Ergodic theorem, P has a finite limit,

$$\lim_{t\to\infty} P_{ab}(t) = \pi_b, \tag{C.78}$$

the dominant exponential terms must have non-positive exponent, meaning

$$P_{ij}^{(k)}(t) \to 0 \quad \forall k \geq 1 \tag{C.79}$$

where $P_{ij}^{(k)}(t)$ denotes the k^{th} derivative of $P_{ij}(t)$.

Recall that

$$-S(t) = \sum_{a,b} \pi_a P_{ab}(t) \log(P_{ab}(t)) \tag{C.80}$$

One could show the result of the lemma by computing the second derivative manually - instead, we use a briefer, though less detailed argument. Note that $S''(t)$ will be a linear function of each P''_{ab} and a quadratic function of each P'_{ab}. Every term in the resultant sum will be multiplied by at least one of these derivatives. The only terms which singularities will be $\log(P_{ab})$, P_{ab}^{-1} and P_{ab}^{-2}, and as each P_{ab} is bounded away from 0 for large t, under the assumption that $\pi_i > 0$ for all i, these converge to finite limits. Thus, each term in the resultant sum will converge to 0 as required.

C.2.8 Behaviour of S near $t = 0$

Lemma C.8 *Define*

$$K = \sum_{a \neq b} \pi_a Q_{ab} \tag{C.81}$$

Consider taking $Kt \to 0$ while keeping the stationary distribution π and the ratios $\frac{Q_{ab}}{Q_{cd}}$ constant (i.e. one can change t or scale the matrix Q by a constant multiple). Then,

$$S(t) \sim Kt \log(Kt) + \mathcal{O}(Kt) \quad \textit{as } Kt \to 0 \tag{C.82}$$

$$S'(t) \sim K \log(Kt) + o(K \log(Kt)) \quad \textit{as } Kt \to 0 \tag{C.83}$$

and

$$S''(t) \sim \frac{K}{t} + o(\frac{K}{t}) \quad \textit{as } Kt \to 0 \tag{C.84}$$

Proof: All terms in sum defining K have the same sign, and so $Kt \to 0$ implies that $K\pi_a Q_{ab} \to 0$ for all $a \neq b$. As π is constant, this means that $KQ_{ab} \to 0$ for all $a \neq b$ and hence, summing these shows that $KQ_{aa} \to 0$ for all a.

The forward equations for a CTMC state that

$$P'(t) = PQ \quad \text{and} \quad P(0) = I \tag{C.85}$$

Thus,

$$P_{ab}(t) \sim \begin{cases} Q_{ab}t + \mathcal{O}(Q_{ab}^2 t^2) & \text{if } a \neq b \\ (1 - Q_{aa}t) + \mathcal{O}(Q_{aa}^2 t^2) & \text{if } a = b \end{cases} \quad \text{as } Kt \to 0 \tag{C.86}$$

Now, as $Kt \to 0$

$$S(t) = \sum_{a,b} \pi_a P_{ab}(t) \log(P_{ab}(t)) \tag{C.87}$$

$$\sim \sum_{a \neq b} \pi_a (Q_{ab}t + \mathcal{O}(Q_{ab}^2 t^2))(\log(Q_{ab}t + \mathcal{O}(Q_{ab}^2 t^2))) + \sum_a \pi_a (1 - Q_{ab}t) \log(1 - Q_{ab}t) \tag{C.88}$$

$$\sim \sum_{a \neq b} \pi_a (Q_{ab}t + \mathcal{O}(K^2 t^2))(\log(Q_{ab}t + \mathcal{O}(K^2 t^2))) + \mathcal{O}(Kt) \tag{C.89}$$

$$\sim \sum_{a \neq b} \pi_a Q_{ab} t \log(Q_{ab}t) + \mathcal{O}(Kt) \tag{C.90}$$

Now, note that the value of $\frac{\pi_a Q_{ab}}{K}$ is constant (and hence $\mathcal{O}(1)$) by assumption. Thus,

$$S(t) \sim \sum_{a \neq b} \pi_a Q_{ab} t \log(Kt) + \mathcal{O}(Kt) = Kt \log(Kt) + \mathcal{O}(Kt) \tag{C.91}$$

which is the required result.

The derivatives follow by noting that the ignored terms are all of the form $(Kt)^n$ and $(Kt)^m \log(Kt)$ for $n \geq 1$ and $m \geq 2$. As these terms all have larger powers of Kt and $Kt << 1$, the derivatives of these terms will be much smaller than the derivative of the leading order term, and hence we can simply differentiate the leading order term to find the leading order derivatives.

Appendix - Paper IV

D.1 Summary

This supplement provides full derivations of the results from the main text. The results are, as in the main text, presented for an epidemic occurring in continuous time, although some additional results on discrete epidemics are given in the final note of this supplement. The supplement is structured as follows.

- The first note, "Modelling", provides a precise definition of the branching process model used throughout the paper.
- The second note, "Probability generating functions" derives probability generating functions (pgfs) for prevalence and cumulative incidence. It also discusses their efficient solution, including some special cases in which one can speed up the solution process
- The third note, "Properties of the prevalence variance", derives the equation for the variance (via the previously derived equations for the pgf) and explores its properties, providing explanations for the various terms and proving that the prevalence of new infections is (under a mild condition on the possible spread of the epidemic) overdispersed.
- The fourth note, "Likelihood functions" contains the derivations of the pgf of the infection event times and the likelihood function presented in the main text.
- The fifth note, "Assessing future variance during an epidemic" derives the equation for variance of future cases when the cumulative incidence is known at some point in time.
- Finally, the sixth note, "Discrete epidemics" provides a range of similar results in the discrete setting, and shows the convergence of the pgf to its continuous equivalent as the step-size tends to zero.

D.2 Background literature on renewal equations

A common approach to modelling infectious diseases is to use the renewal equation. The early theory on the properties of the renewal equation can be found here [498]. Epidemiologically derived descriptions can be found here [412, 426] where the renewal equation is framed in an epidemiological

framework with reference to infection processes. The link between the renewal equation and the popular susceptible-infected-recovered models can be found here [499]. The basics of branching processes can be found here [422]. In what follows, we will arrive at a renewal equation from first principles by first starting with the probability generating function of a general branching process.

D.3 Modelling

D.3.1 Branching process framework

We present a general time-varying age-dependent branching process that is most similar to the general branching process initially proposed by Crump, Mode and Jagers [500, 501]. Following [60], in our process, we begin with a single individual infected at some time l whose infectious period is a random variable distributed by cumulative distribution function $G(\cdot, l)$, admitting a probability density $g(\cdot, l)$. During this individual's life length, the individual gives rise to an integer-valued random number of secondary infections according to a counting processes $\{N(t,l)\}_{t\geq l}$ ($\{N(t,l)\}$ is the number of secondary infections) where t is a global "calendar" time. The amount of time for which the individual has been infected before time t is therefore $t - l$.

For each infection event time - that is, for each v such that

$$v \in \left\{ u \leq t : \lim_{s \to u_-} (N(s,l)) \neq \lim_{s \to u_+} (N(s,l)) \right\} \tag{D.1}$$

we then define a random variable

$$Y(v,l) := \lim_{s \to v_+} (N(s,l)) - \lim_{s \to v_-} (N(s,l)) \tag{D.2}$$

to be the size of the infection event at time v; that is, this is the number of individuals that are infected (by the initial individual) at time v. Throughout this paper, it will be assumed that $Y = Y(v)$, so that Y does not depend on the length of time for which an individual has been infected. However, this assumption could be removed from the model if desired.

Each newly infected individual then proceeds, independently, in the same way as the initial individual. The only change is that the time at which they are infected will be different (but, for example, the infection tree rooted at an individual infected at time $s > l$ is equal in distribution to the full infection tree if one started an epidemic with $l = s$). This self-similarity property underpins the

derivations in the subsequent notes, as it allows an epidemic to be characterised purely by the "first generation" of infected individuals (and hence, the equations are derived using the "first generation principle").

D.3.2 The counting process, $N(t,l)$

Our framework relies on the assumption that the counting processes $N(t,l)$ have independent increments and are continuous in probability:

$$\lim_{\delta\to 0}\left[\mathbb{P}\Big(N(t+\delta,l)-N(t,l)\Big)\right]=0 \quad \forall t\geq l\geq 0 \tag{D.3}$$

This condition excludes any discrete formulations of the epidemic process. It will be shown later in the supplement that discrete epidemics (which are not continuous in probability), are structurally different as extra terms appear in the equations for the pgf. However, the equations in the continuous case are recovered as the step-size of the discrete process tends to zero.

A further assumption on $N(t,l)$ is that it can be constructed from a Lévy Process - that is, there is some non-negative rate function $r(t,l)$ and some Lévy Process $\mathcal{N}(t)$ such that

$$N(t,l)=\mathcal{N}\Big(\int_l^t r(s,l)ds\Big) \tag{D.4}$$

Note that the counting processes relating to different individuals are independent, and hence will come from different independent copies of the base process $\mathcal{N}$.

This assumption is important because it means that the counting process of "infection events" (that is, points in time such that the value of $N(t,l)$ changes) is an inhomogeneous Poisson Process, which can be shown as follows. Consider a counting process, $J_{\mathcal{N}}(t,l)$ that counts the increases in $\mathcal{N}$. That is,

$$J_{\mathcal{N}}(t):=\left|\left\{u\leq t:\lim_{s\to u_-}(\mathcal{N}(s))\neq\lim_{s\to u_+}(\mathcal{N}(s))\right\}\right| \tag{D.5}$$

where here $|\cdot|$ denotes the number of elements in a set. Then, as $\mathcal{N}$ is a Lévy Process, $J_{\mathcal{N}}(t)$ has iid (independent and identically distributed) increments and is non-decreasing in t with jumps of size 1 and thus follows a Poisson Process with some rate κ [442]. Thus, if $J(t,l)$ is the counting process of infection events in $\mathcal{N}(t,l)$, then

$$J(t,l)=J_{\mathcal{N}}\Big(\int_l^t r(s,l)ds\Big) \tag{D.6}$$

and hence, $J(t,l)$ is an inhomogeneous Poisson Process with rate $\kappa r(t,l)$ as required. In particular, defining

$$\lambda(t,l) := \int_l^t r(s,l)ds, \tag{D.7}$$

$J(t,l)$ has a generating function of

$$\mathcal{J}_{(t,l)}(s) = e^{\kappa\lambda(t,l)(s-1)} \tag{D.8}$$

D.3.3 The rate function, $r(t,l)$

Throughout the examples in this paper, the rate function $r(t,l)$ will be given as

$$r(t,l) = \rho(t)\nu(t-l) \tag{D.9}$$

Here, $\rho(t)$ is a population-level infection event rate. Note that, because the number of infections caused at each infection rate may be greater than 1 (that is one may have $J(t,l) < N(t,l)$), $\rho(t)$ cannot necessarily be interpreted in direct analogue to the reproduction number. $\nu(t-l)$ gives the infectiousness of an individual after it has been infected for time $(t-l)$. It will be assumed that $\int_0^\infty \nu(s)ds = 1$ so that ρ can be interpreted as the infection event rate.

D.3.4 Smoothness assumptions

Note that, throughout the derivations of this paper, the smoothness of ρ, ν and g will not be explicitly considered when taking limits - it will be assumed that they are sufficiently smooth for "natural" results to hold. The authors believe that the results of this paper will hold for any piecewise continuous choices for these functions, although more detailed analysis would be needed to provide a rigorous proof of this. It is possible that they hold for much wider classes of functions, but this seems to the authors to be outside the realm of epidemiological interest, as it appears implausible that any of these functions would not be piecewise continuous in a realistic setting.

Moreover, it will be assumed that unique solutions to the equations for the pgf, mean and variance exist. Again, a proof of this property is beyond the scope of this work, although the classes of equations presented in this paper are common across the literature, and it is likely that interested readers with a pure mathematical background could find applicable results to address this issue.

D.3.5 Special cases for $N(t,l)$

Throughout this paper, two special cases for $N(t,l)$ are considered - the case where $N(t,l)$ is itself an inhomogeneous Poisson Process, and the case where $N(t,l)$ is a Negative Binomial process. These were used to construct the figures in the paper and explanations as to how they can be used will be presented throughout this supplement.

D.4 Probability generating functions

D.4.1 General case

Define $F(t,l;s) := E\left(s^{Z(t,l)}\right)$ to be the generating function of $Z(t,l)$. For simplicity of notation the dependence of F on s will be suppressed.

To derive the generating function $F(t,l)$, we condition on the infection period (lifetime) of the initial case, L.

$$E\left(s^{Z(t,l)}\right) = \int_0^\infty E\left(s^{Z(t,l)}\middle| L=u\right)g(u,l)du \tag{D.10}$$

$$= \int_{t-l}^\infty E\left(s^{Z(t,l)}\middle| L=u\right)g(u,l)du + \int_0^{t-l} E\left(s^{Z(t,l)}\middle| L=u\right)g(u,l)du \tag{D.11}$$

The counting process of the first individual, $N(t,l)$ is independent of this first individual's infection period L. If $L > t-l$ then this individual is still infectious and able to infect others at time t. Therefore, conditional on $L > t-l$, the number of people they have infected before time t is independent of L (as all infections from $N(s,l)_{l\leq s\leq t}$ are counted, irrespectively of the value of L). That is (the first term in D.11)

$$\int_{t-l}^\infty E\left(s^{Z(t,l)}\middle| L=u\right)g(u,l)du = \int_{t-l}^\infty E\left(s^{Z(t,l)}\middle| L\geq t-l\right)g(u,l)du \tag{D.12}$$

and hence, the first integral in Supplementary Equation D.11 can be simplified to give

$$E\left(s^{Z(t,l)}\right) = \left(1-G(t-l,l)\right)E\left(s^{Z(t,l)}\middle| L\geq t-l\right) + \int_0^{t-l} E\left(s^{Z(t,l)}\middle| L=u\right)g(u,l)du \tag{D.13}$$

Let us consider the second part of Supplementary Equation D.11. Suppose first that $L=u$ for some $u < t-l$ so that the index case is no longer alive at time t. Thus, the number of infection events caused by the index case is given by $J(l+u,l)$.

Define the set of times at which these infected events occurred to be $\{K_1,...,K_{J(l+u,l)}\}$ where here, importantly, the K_i are labelled in a random order (so it is not necessarily the case that $K_1 < ... <$

$K_{J(l+u,l)}$). As J is an inhomogeneous Poisson Process and $N(t,l)$ is continuous in probability, the K_i are therefore iid with pdf (probability density function)

$$f_K(k) = \frac{r(l+k,l)}{\int_0^u r(l+s,l)ds} \tag{D.14}$$

It is perhaps helpful to note that this is the step which relies on N being continuous in probability. If this were not the case and $N(t,l)$ had non-zero probability of increasing at some time s, then the knowledge that $K_1 = s$ would give some information about K_2, as the fact that $K_2 \neq s$ would change its probability distribution, meaning K_1 and K_2 would not be independent. Conversely, in the continuous case, $K_1 = s$ removes an event of zero measure from the probability space of K_2, and hence K_1 and K_2 are still independent.

Now, by the self-similarity property ([422, 502]) we have

$$Z(t,l) = \sum_{i=1}^{J(l+u,l)} \sum_{j=1}^{Y(l+K_i(l+u,l))} Z_{ij}(t, l+K_i(l+u,l)) \tag{D.15}$$

where each Z_{ij} is an independent copy of Z that is equal in distribution. Z_{ij} denotes the jth individual corresponding to infection event time i. The two summations, from all previous infections, sum over all the infection events and their sizes. This summation is valid as each individual behaves independently once it has been infected.

Recall that if X_i are iid random variables (with a generating function, $G_X(s)$) and if Y is a non-negative integer-valued random variable (again with a generating function, $G_Y(s)$), then,

$$E\left(s^{\sum_{i=1}^{Y} X_i}\right) = G_Y(G_X(s)) \tag{D.16}$$

By defining $\mathcal{J}_{(t,l)}$ to be the generating function of $J(t,l)$, this relationship allows us to write $\mathbb{E}(s^{Z(t,l)}|L = u)$ as

$$\mathbb{E}(s^{Z(t,l)}|L=u) = \mathcal{J}_{(l+u,l)}\left(E\left[s^{\sum_{j=1}^{Y(l+K(l+u,l))} Z_j(t,l+K(l+u,l))}\right]\right) \tag{D.17}$$

where here, K is equal in distribution to the K_i. Conditioning on the value of K,

$$E\left[s^{\sum_{j=1}^{Y(l+K)} Z_j(t,l+K)}\right] = \int_0^u E\left[s^{\sum_{j=1}^{Y(l+k)} Z_j(t,l+k)}\right] \frac{r(l+k,l)}{\lambda(l+u,l)} dk \tag{D.18}$$

Thus, defining $\mathcal{Y}_{(l+k)}$ to be the generating function of $Y(l+k)$

$$E\left[s^{\sum_{j=1}^{Y(l+K)} Z_j(t,l+K)}\right] = \int_0^u \mathcal{Y}_{(l+k)}(F(t,l+k))\frac{r(l+k,l)}{\lambda(l+u,l)}dk \tag{D.19}$$

We can equivalently write this as an exponential, using the fact that $J(t,l)$ is Poisson distributed:

$$\mathbb{E}(s^{Z(t,l)}|L=u) = \mathcal{J}_{(l+u,l)}\left(\int_0^u \mathcal{Y}_{(l+k)}(F(t,l+k))\frac{r(l+k,l)}{\lambda(l+u,l)}dk\right) \tag{D.20}$$

$$= \exp\left[\kappa\lambda(l+u,l)\left(\int_0^u \mathcal{Y}_{(l+k)}(F(t,l+k))\frac{r(l+k,l)}{\lambda(l+u,l)}dk - 1\right)\right] \tag{D.21}$$

An identical derivation can be performed on the first integral in Supplementary Equation D.11 (swapping $t-l$ for u and multiplying by s to account for the initial case, which is counted in the prevalence at t when $L > t-l$), resulting in

$$\mathbb{E}(s^{Z(t,l)}|L \geq t-l) = s\mathcal{J}_{(t,l)}\left(\int_0^{t-l} \mathcal{Y}_{(l+k)}(F(t,l+k))\frac{r(l+k,l)}{\lambda(t,l)}dk\right) \tag{D.22}$$

$$= s\exp\left[\kappa\lambda(t,l)\left(\int_0^{t-l} \mathcal{Y}_{(l+k)}(F(t,l+k))\frac{r(l+k,l)}{\lambda(t,l)}dk - 1\right)\right] \tag{D.23}$$

and therefore, this yields an overall pgf

$$F(t,l) = s\Big(1-G(t-l,l)\Big)\mathcal{J}_{(t,l)}\left(\int_0^{t-l} \mathcal{Y}_{(l+u)}(F(t,l+u))\frac{r(l+u,l)}{\lambda(t,l)}du\right)... \\ ... + \int_0^{t-l} \mathcal{J}_{(l+u,l)}\left(\int_0^u \mathcal{Y}_{(l+k)}(F(t,l+k))\frac{r(l+k,l)}{\lambda(l+u,l)}dk\right)g(u,l)du \tag{D.24}$$

or, equivalently

$$F(t,l) = s\Big(1-G(t-l,l)\Big)\exp\left[\kappa\lambda(t,l)\left(\int_0^{t-l} \mathcal{Y}_{(l+k)}(F(t,l+k))\frac{r(l+k,l)}{\lambda(t,l)}dk - 1\right)\right]... \\ ... + \int_0^{t-l} \exp\left[\kappa\lambda(l+u,l)\left(\int_0^u \mathcal{Y}_{(l+k)}(F(t,l+k))\frac{r(l+k,l)}{\lambda(l+u,l)}dk - 1\right)\right]g(u,l)du \tag{D.25}$$

Note that by absorbing κ into the rate function $r(l+k,l)$, it can be assumed that $\kappa = 1$. Intuitively this is simply scaling the probability density by the number of points.

D.4.2 Solving the pgf equation

Practically, one will always set $l = 0$ for an epidemic, and so only the values $F(t,0)$ are directly relevant. However, it is still necessary to solve for $F(t,l)$ for $0 \leq l \leq t$. In the language of PDEs (partial differential equations) and, specifically, the Cauchy problem, this can be explained by the fact

that the "data curve" is the line $t = l$ (as the values of $F(t,t)$ are known to be equal to s) and the "characteristics" of the system are the lines $t =$ constant. Thus, to calculate the value of $F(t,0)$, it is necessary to follow the characteristic from (t,t) to $(t,0)$ and hence calculate $F(t,l)$ for $0 \leq l \leq t$.

Hence, following [60], solving Supplementary Equation D.25 can be greatly facilitated by defining an auxiliary equation $F_c(t) = F(c, c-t)$ and allows us to write Supplementary Equation D.25 as an equation in one variable. This is

$$F_c(t) = s\Big(1 - G(t,l)\Big)\mathcal{J}_{(c,c-t)}\bigg(\int_0^t \mathcal{Y}_{(c-t+u)}(F_c(t-u))\frac{r(c-t+u,c-t)}{\lambda(c,c-t)}du\bigg)... \\ ... + \int_0^t \mathcal{J}_{(c-t+u,c-t)}\bigg(\int_0^u \mathcal{Y}_{(c-t+k)}(F_c(t-k))\frac{r(c-t+k,c-t)}{\lambda(u,c-t)}dk\bigg)g(u,l)du \tag{D.26}$$

or, equivalently

$$F_c(t) = s\Big(1 - G(t,l)\Big)\exp\bigg[\lambda(c,c-t)\kappa\bigg(\int_0^t \mathcal{Y}_{(c-t+u)}(F_c(t-u))\frac{r(c-t+u,c-t)}{\lambda(c,c-t)}du - 1\bigg)\bigg]... \\ ... + \int_0^t \exp\bigg[\lambda(u,c-t)\kappa\bigg(\int_0^u \mathcal{Y}_{(c-t+k)}(F_c(t-k))\frac{r(c-t+k,c-t)}{\lambda(u,c-t)}dk - 1\bigg)\bigg]g(u,l)du \tag{D.27}$$

D.4.3 Poisson case

If $N(t,l)$ is an inhomogeneous Poisson Process, then, as the infection event size for a Poisson Process is always 1 [442], one has $\mathcal{Y}_{(t)}(s) = s$. To aid understanding below in the Negative Binomial case, it is helpful to note that the Lévy Process, $\mathcal{N}$, can hence be characterised by

$$\mathbb{P}(\mathcal{N}(t+dt) - \mathcal{N}(t) = 0) = 1 - \kappa dt$$
$$\mathbb{P}(\mathcal{N}(t+dt) - \mathcal{N}(t) = 1) = \kappa dt$$
$$\mathbb{P}(\mathcal{N}(t+dt) - \mathcal{N}(t) > 1) = o(dt)$$

Setting $\kappa = 1$ as discussed above, the generating function equation becomes

$$F(t,l) = s\Big(1 - G(t-l,l)\Big)\exp\bigg[\bigg(\int_0^{t-l} F(t,l+k)\rho(l+k)\nu(k)dk - \lambda(t,l)\bigg)\bigg]... \tag{D.28}$$
$$... + \int_0^{t-l}\exp\bigg[\bigg(\int_0^u F(t,l+k)\rho(l+k)\nu(k)dk - \lambda(l+u,l)\bigg)\bigg]g(u,l)du \tag{D.29}$$

This equation can be further simplified by recalling that

$$\lambda(t,l) := \int_l^t r(u,l)du = \int_0^{t-l} r(u+l,l)du = \int_0^{t-l}\rho(u+l)\nu(u)du \tag{D.30}$$

therefore

$$F(t,l) = s\Big(1 - G(t-l,l)\Big) \exp\left[\left(\int_0^{t-l} F(t,l+k)\rho(l+k)\nu(k)dk - \int_0^{t-l} \rho(l+k)\nu(k)dk\right)\right]... \\ ... + \int_0^{t-l} \exp\left[\left(\int_0^{u} F(t,l+k)\rho(l+k)\nu(k)dk - \int_0^{u} \rho(l+k)\nu(k)du\right)\right] g(u,l)du$$

$$= s\Big(1 - G(t-l,l)\Big) \exp\left[\left(\int_0^{t-l} \rho(l+k)\nu(k)\left(F(t,l+k)-1\right)dk\right)\right]... \\ ... + \int_0^{t-l} \exp\left[\left(\int_0^{u} \rho(l+k)\nu(k)dk\left(F(t,l+k)-1\right)\right)\right] g(u,l)du \tag{D.31}$$

For computational ease the auxiliary function equation is then

$$F_c(t) = s\Big(1 - G(t,l)\Big) \exp\left[\left(\int_0^{t} \left(F_c(t-u)-1\right)\rho(c-t+u)\nu(u)du\right)\right]... \\ ... + \int_0^{t} \exp\left[\left(\int_0^{u} \left(F_c(t-k)-1\right)\rho(c-t+k)\nu(k)dk\right)\right] g(u,l)du \tag{D.32}$$

D.4.4 Inhomogeneous Negative Binomial case

Our derivation follows from the well-known relationship that the Negative Binomial distribution arises from a compound Poisson distribution. For $p \in (0,1)$ and $\phi \in \mathbb{R}^+$, if

$$X = \sum_{i=1}^{N} Y_i \tag{D.33}$$

where

$$N \sim \text{Poisson}(-\phi \ln(p)) \tag{D.34}$$

and each Y_i is independent of N, iid, and follows a logarithmic series distribution

$$Y_i \sim \text{Logarithmic}(1-p) \tag{D.35}$$

then the random variable X is Negative Binomial distributed. This can easily be proven using pgfs. Therefore, we have $\kappa = -\ln(p)\phi$ and can calculate the pgf for Y as $\mathcal{Y}(s) = \frac{\ln(1-(1-p)s)}{\ln(p)}$. These can then be substituted into our general Supplementary Equation D.25.

For clarity we re-derive this relationship explicitly. We have

$$\mathcal{N}(t) \sim \text{NB}(\phi t, p) \tag{D.36}$$

As $M(t)$ has iid increments,

$$\mathbb{P}\Big(\mathcal{N}(t+dt)-\mathcal{N}(t)=k\Big)=\mathbb{P}\Big(\mathcal{N}(dt)=k\Big)=\frac{(k+\phi dt-1)(k+\phi dt-2)...\phi dt}{k!}(1-p)^k p^{\phi dt} \tag{D.37}$$

Thus, to leading order, for $k>0$, one has

$$\mathbb{P}\Big(\mathcal{N}(t+dt)-\mathcal{N}(t)=k\Big)=\frac{(1-p)^k\phi dt}{k}+o(dt) \tag{D.38}$$

while if $k=0$,

$$\mathbb{P}\Big(\mathcal{N}(t+dt)-\mathcal{N}(t)=0\Big)=p^{\phi dt}=1+\ln(p)\phi dt+o(dt) \tag{D.39}$$

(noting that $\ln(p)<0$). This means that the infection event process $J_{\mathcal{N}}$ satisfies

$$\mathbb{P}\Big(J_{\mathcal{N}}(t+dt)-J_{\mathcal{N}}(t)=0\Big)=1+\ln(p)\phi dt+o(dt) \tag{D.40}$$

and

$$\mathbb{P}\Big(J_{\mathcal{N}}(t+dt)-J_{\mathcal{N}}(t)=1\Big)=\sum_{k=1}^{\infty}\frac{(1-p)^k\phi dt}{k}+o(dt) \tag{D.41}$$

$$=-\ln(p)\phi dt+o(dt) \tag{D.42}$$

and hence, $J_{\mathcal{N}}$ is a Poisson Process of rate $-\ln(p)\phi$ [503]. Thus, one has

$$\kappa=-\ln(p)\phi \tag{D.43}$$

as expected. Moreover, the pmf (probability mass function) of an infection event size, Y is given by

$$\mathbb{P}(Y=k)=\frac{(1-p)^k}{-k\ln(p)} \tag{D.44}$$

One can hence find the generating function as

$$\mathcal{Y}(s)=\sum_{k=1}^{\infty}\frac{((1-p)s)^k\phi}{-k\ln(p)} \tag{D.45}$$

Noting that

$$\sum_{k=1}^{\infty}\frac{(1-p)^k}{-k\ln(p)}=1 \tag{D.46}$$

one has

$$\mathcal{Y}(s) = \frac{\ln(1-(1-p)s)}{\ln(p)} \sum_{k=1}^{\infty} \frac{(1-(1-(1-p)s))^k}{-k\ln(1-(1-p)s)} = \frac{\ln(1-(1-p)s)}{\ln(p)} \tag{D.47}$$

These results can be substituted into the general formula to give

$$\begin{aligned} F(t,l) =& s\Big(1-G(t-l,l)\Big)\exp\left[-\phi\Big(\int_0^{t-l}\ln(1-(1-p)F(t,l+u))\rho(u+l)\nu(u)du + \ln(p)\lambda(t,l)\Big)\right]... \\ &... + \int_0^{t-l}\exp\left[-\phi\Big(\int_0^{u}\ln(1-(1-p)F(t,l+k))\rho(k+l)\nu(k)dk + \ln(p)\lambda(u,l)\Big)\right]g(u,l)du \end{aligned} \tag{D.48}$$

As in the Poisson case, this equation can be simplified by factoring λ

$$\begin{aligned} F(t,l) =& s\Big(1-G(t-l,l)\Big)\exp\left[-\phi\Big(\int_0^{t-l}\left(\ln(1-(1-p)F(t,l+u))-\ln(p)\right)\rho(u+l)\nu(u)du\Big)\right]... \\ &... + \int_0^{t-l}\exp\left[-\phi\Big(\int_0^{u}\left(\ln(1-(1-p)F(t,l+k))-\ln(p)\right)\rho(k+l)\nu(k)dk\Big)\right]g(u,l)du \end{aligned} \tag{D.49}$$

The easier-to-solve auxiliary function is given by

$$\begin{aligned} F_c(t) =& s\Big(1-G(t-l,l)\Big)\exp\left[-\phi\Big(\int_0^{t}\left(\ln(1-(1-p)F_c(t-u))-\ln(p)\right)\rho(c-t+u)\nu(u)du\Big)\right]... \\ &... + \int_0^{t}\exp\left[-\phi\Big(\int_0^{u}\left(\ln(1-(1-p)F_c(t-k))-\ln(p)\right)\rho(c-t+k)\nu(k)dk\Big)\right]g(u,l)du \end{aligned} \tag{D.50}$$

If $p = \frac{\phi}{1+\phi}$, then the Poisson case (with $\kappa = 1$) is recovered in the $\phi \to \infty$ limit.

Note that $\mathbb{E}[N(t,l)] = \frac{\phi\lambda(t,l)(1-p)}{p}$ while in our case, we impose that $\mathbb{E}[N(t,l)] = \lambda(t,l)$. Solving for p we can see $p = \frac{\phi}{1+\phi}$ and this relation can be substituted into Supplementary Equation D.50. Note that this agrees with the definition of p in the Poisson limit.

D.4.5 Cumulative incidence

Similar to prevalence, cumulative incidence can be calculated by counting all previous infections as well as current ones. Following an identical derivation to prevalence the pgf for cumulative incidence simply requires multiplying the second integral by s as the initial infection is counted in the cumulative

incidence regardless of the value of L.

$$F(t,l) = s\Big(1 - G(t-l,l)\Big)\mathcal{J}_{(t,l)}\bigg(\int_0^{t-l} \mathcal{Y}_{(l+u)}(F(t,l+u))\frac{r(l+u,l)}{\lambda(t,l)}du\bigg)... \\ ... + s\int_0^{t-l} \mathcal{J}_{(l+u,l)}\bigg(\int_0^{u} \mathcal{Y}_{(l+k)}(F(t,l+k))\frac{r(l+k,l)}{\lambda(t,l)}dk\bigg)g(u,l)du \tag{D.51}$$

D.4.6 A simplified pgf ignoring g

By assuming $L = \infty$ and therefore $G(u,l) = 0 \ \ \forall u$, the pgf for prevalence (or, in this case, equivalently, cumulative incidence) simplifies to

$$F(t,l) = s\mathcal{J}_{(t,l)}\bigg(\int_0^{t-l} \mathcal{Y}_{(l+u,l)}(F(t,l+u))\frac{r(l+u,l)}{\lambda(t,l)}du\bigg)$$

Additional computational savings can be gained in our case $r(t,l) = \rho(t)\nu(t-l)$ if the infectiousness ν decays to zero quickly. This means that the auxiliary equation used for computation can be truncated to some time $\min(t,T)$. For example, in the Poisson case this becomes,

$$F_c(t) = \exp\left[\bigg(\int_0^{\min(t,T)} \left(F_c(t-u) - 1\right)\rho(c-t+u)\nu(u)du\bigg)\right]s \tag{D.52}$$

and in the Negative Binomial case this becomes,

$$F_c(t) = \exp\left[-\phi\bigg(\int_0^{\min(t,T)} \left(\ln(1-(1-p)F_c(t-u)) - \ln(p)\right)\rho(c-t+u)\nu(u)du\bigg)\right]s \tag{D.53}$$

These computational savings allow computation of the pgf for millions of iterations in minutes.

D.4.7 Calculating the probability mass function via the pgf

Following [447] and [448] (originally from [443]), by the properties of pgfs, the probability mass function p can be recovered through a pgf F's derivatives at $s = 0$

$$\mathbb{P}(n) = \frac{1}{n!}\left(\frac{d}{ds}\right)^n F(s;t,\tau)|_{s=0}$$

This is generally computationally intractable. A well-known result from complex analysis [443] holds that

$$f^{(n)}(a) = \frac{n!}{2\pi i}\oint \frac{f(z)}{(z-a)^{n+1}}\,dz. \tag{D.54}$$

Therefore,

$$\mathbb{P}(n) = \frac{1}{2\pi i} \oint \frac{F(z;t,\tau)}{z^{n+1}} dz \tag{D.55}$$

This integral can be done on a closed circle around the origin such that $z = re^{i\theta}$ and $dz = izd\theta$ - i.e.

$$\mathbb{P}(n) = \frac{1}{2\pi} \int_0^{2\pi} \frac{F(re^{i\theta};t,\tau)}{(re^{i\theta})^n} d\theta \tag{D.56}$$

Finally through substitution $\theta = 2\pi u$ such that $d\theta = 2\pi du$, where $u \in [0,1]$ we find

$$\mathbb{P}(n) = \int_0^1 \frac{F(re^{2\pi iu};t,\tau)}{r^n e^{2\pi iun}} du \tag{D.57}$$

Since trapezoidal sums are known to converge geometrically for periodic analytic functions (Davis 1959) a simple approximation becomes

$$\mathbb{P}(n) = \frac{1}{Mr^n} \sum_{m=0}^{M-1} F(re^{2\pi im/M};t,\tau) e^{-2\pi inm/M} \tag{D.58}$$

Bornemann[448] suggest using $r = 1$.

The probability mass function for any time and n can be determined numerically. One needs $M \geq n$, which requires solving n renewal equations for the generating function and performing a fast Fourier transform. This is generally computationally fast, but may become slightly burdensome for epidemics with very large numbers of infected individuals.

D.5 Properties of the prevalence variance

D.5.1 Derivation of equation for mean prevalence

Before deriving the equation for the prevalence variance, it is important to derive the equation governing the mean prevalence. This has been previously derived in [60], although here, we re-derive it from our new pgfs. First note that

$$\begin{aligned}
&\frac{\partial}{\partial s}\left(\mathcal{J}_{(t,l)}\left(\int_0^{t-l} \mathcal{Y}_{(l+u,l)}(F(t,l+u))\frac{r(l+u,l)}{\lambda(t,l)}du\right)\right)... \\
&= \left[\int_0^{t-l} F_s(t,l+u)\frac{r(l+u,l)}{\lambda(t,l)}\mathcal{Y}'_{(l+u,l)}(F(t,l+u))du\right]\left[\mathcal{J}'_{(t,l)}\left(\int_0^{t-l} \mathcal{Y}_{(l+u,l)}(F(t,l+u))\frac{r(l+u,l)}{\lambda(t,l)}du\right)\right]
\end{aligned} \tag{D.59}$$

Now, setting $s = 1$ so that $F(\cdot,\cdot) = 1$ and $F_s(\cdot,\cdot) = M(\cdot,\cdot)$, one has

$$\Big[\int_0^{t-l} M(t,l+u)\frac{r(l+u,l)}{\lambda(t,l)}\mathcal{Y}'_{(l+u,l)}(1)du\Big]\Big[\mathcal{J}'_{(t,l)}\Big(\int_0^{t-l}\mathcal{Y}_{(l+u,l)}(1)\frac{r(l+u,l)}{\lambda(t,l)}du\Big)\Big)\Big] \tag{D.60}$$

Now, define $B(t) = \mathbb{E}(Y(t))$ so that $\mathcal{Y}'_{(l+u)}(1) = B(l+u)$. Moreover, $\mathcal{Y}_{(l+u)}(1) = 1$ so the equation becomes

$$\Big[\int_0^{t-l} M(t,l+u)\frac{r(l+u,l)}{\lambda(t,l)}B(l+u)du\Big]\Big[\mathcal{J}'_{(t,l)}\Big(\int_0^{t-l}\frac{r(l+u,l)}{\lambda(t,l)}du\Big)\Big)\Big] \tag{D.61}$$

Now, necessarily

$$\int_0^{t-l}\frac{r(l+u,l)}{\lambda(t,l)}du = 1 \Rightarrow \mathcal{J}'_{(t,l)}\Big(\int_0^{t-l}\frac{r(l+u,l)}{\lambda(t,l)}du\Big) = \mathbb{E}(J(t,l)) = \kappa\lambda(t,l) \tag{D.62}$$

and so, this results in

$$\int_0^{t-l} M(t,l+u)\frac{r(l+u,l)}{\lambda(t,l)}B(l+u)\kappa\lambda(t,l)du \tag{D.63}$$

Moreover, evaluating

$$\mathcal{J}_{(t,l)}\Big(\int_0^{t-l}\mathcal{Y}_{(l+u)}(F(t,l+u))\frac{r(l+u,l)}{\lambda(t,l)}du\Big) \tag{D.64}$$

at $s = 1$ (necessary for the term where, by the product rule, we differentiate the multiple of s) gives

$$\mathcal{J}_{(t,l)}\Big(\int_0^{t-l}\mathcal{Y}_{(l+u)}(1)\frac{r(l+u,l)}{\lambda(t,l)}du\Big) = \mathcal{J}_{(t,l)}\Big(\int_0^{t-l}1\times\frac{r(l+u,l)}{\lambda(t,l)}du\Big) \tag{D.65}$$

$$= \mathcal{J}_{(t,l)}(1) \tag{D.66}$$

$$= 1 \tag{D.67}$$

Thus, the derivative of the full generating function equation gives

$$M(t,l) = (1-G(t-l,l))\Big[1+\int_0^{t-l}M(t,l+u)\frac{r(l+u,l)}{\lambda(t,l)}B(l+u)\kappa\lambda(t,l)du\Big]... \tag{D.68}$$

$$...+\int_0^{t-l}\int_0^{u}M(t,l+k)\frac{r(l+k,l)}{\lambda(l+u,l)}B(l+k)\kappa\lambda(l+u,l)g(u,l)dkdu \tag{D.69}$$

This can be simplified significantly. Note that,

$$\int_0^{t-l}\int_0^{u}M(t,l+k)\frac{r(l+k,l)}{\lambda(l+u,l)}B(l+k)\kappa\lambda(l+u,l)g(u,l)dkdu =$$
$$\int_0^{t-l}\int_0^{u}M(t,l+k)r(l+k,l)B(l+k)\kappa g(u,l)dkdu \tag{D.70}$$

Moreover, one can change the order of integration to get

$$\int_0^{t-l}\int_k^{t-l} M(t,l+k)r(l+k,l)B(l+k)\kappa g(u,l)dudk = \int_0^{t-l} M(t,l+k)r(l+k,l)B(l+k)\kappa(G(t-l,l)-G(k,l)) \tag{D.71}$$

and hence, one can write the equation for $M(t,l)$ as

$$M(t,l) = (1-G(t-l,l)) + \int_0^{t-l} M(t,l+u)r(l+u,l)B(l+u)\kappa(1-G(u,l))du \tag{D.72}$$

Note that, for the Poisson special case, $B(l+u,l)=1$ and for the Negative Binomial special case, $B(l+u,l)=\frac{p-1}{p\ln(p)}=-\frac{1}{\ln(p)\phi}$. In both cases, it may improve the epidemiological interpretation of ρ to absorb the $B(l+u,l)\kappa$ term into ρ (so that ρ becomes a measure of the rate of new infections). This gives the simpler equation

$$M(t,l) = (1-G(t-l,l)) + \int_0^{t-l} M(t,l+u)\rho(l+u)\nu(u)(1-G(u,l))du \tag{D.73}$$

which agrees with [60].

D.5.2 Derivation of equation for prevalence variance

The equation for variance can now be found by taking the second derivative of the pgf. Define $W(t,l) := \mathbb{E}(Z(t,l)(Z(t,l)-1))$. Note that this then gives the variance, $V(t,l)$ as $V(t,l) = W(t,l) + M(t,l) - M(t,l)^2$.

Consider first the term

$$s\bigg(1-G(t-l,l)\bigg)\mathcal{J}_{(t,l)}\bigg(\int_0^{t-l}\mathcal{Y}_{(l+u)}(F(t,l+u))\frac{r(l+u,l)}{\lambda(t,l)}du\bigg) \tag{D.74}$$

The first derivative of this term is equal to

$$\begin{aligned}&\bar{G}(t-l,l)\mathcal{J}_{(t,l)}\bigg(\int_0^{t-l}\mathcal{Y}_{(l+u)}(F(t,l+u))\frac{r(l+u,l)}{\lambda(t,l)}du\bigg)+...\\ &s\bar{G}(t-l,l)\bigg[\int_0^{t-l}F_s(t,l+u)\mathcal{Y}'_{(l+u)}(F(t,l+u))\frac{r(l+u,l)}{\lambda(t,l)}du\bigg]\mathcal{J}'_{(t,l)}\bigg(\int_0^{t-l}\mathcal{Y}_{(l+u)}(F(t,l+u))\frac{r(l+u,l)}{\lambda(t,l)}du\bigg)\end{aligned} \tag{D.75}$$

Then, the second derivative is equal to

$$2\bar{G}(t-l,l)\left[\int_0^{t-l} F_s(t,l+u)\mathcal{Y}'_{(l+u)}(F(t,l+u))\frac{r(l+u,l)}{\lambda(t,l)}du\right]\mathcal{J}'_{(t,l)}\left(\int_0^{t-l}\mathcal{Y}_{(l+u)}(F(t,l+u))\frac{r(l+u,l)}{\lambda(t,l)}du\right)+$$
$$+s\bar{G}(t-l,l)\left[\int_0^{t-l} F_s(t,l+u)\mathcal{Y}'_{(l+u)}(F(t,l+u))\frac{r(l+u,l)}{\lambda(t,l)}du\right]^2\mathcal{J}''_{(t,l)}\left(\int_0^{t-l}\mathcal{Y}_{(l+u)}(F(t,l+u))\frac{r(l+u,l)}{\lambda(t,l)}du\right)$$
$$+s\bar{G}(t-l,l)\left[\int_0^{t-l} F_{ss}(t,l+u)\mathcal{Y}'_{(l+u)}(F(t,l+u))\frac{r(l+u,l)}{\lambda(t,l)}du\right]\mathcal{J}'_{(t,l)}\left(\int_0^{t-l}\mathcal{Y}_{(l+u)}(F(t,l+u))\frac{r(l+u,l)}{\lambda(t,l)}du\right)$$
$$+s\bar{G}(t-l,l)\left[\int_0^{t-l} F_s^2(t,l+u)\mathcal{Y}''_{(l+u)}(F(t,l+u))\frac{r(l+u,l)}{\lambda(t,l)}du\right]\mathcal{J}'_{(t,l)}\left(\int_0^{t-l}\mathcal{Y}_{(l+u)}(F(t,l+u))\frac{r(l+u,l)}{\lambda(t,l)}du\right) \tag{D.76}$$

Now, one can evaluate this as $s=1$. Note that

$$\begin{aligned}\int_0^{t-l}\mathcal{Y}_{(l+u)}(F(t,l+u))\frac{r(l+u,l)}{\lambda(t,l)}du &= \int_0^{t-l}\mathcal{Y}_{(l+u)}(1)\frac{r(l+u,l)}{\lambda(t,l)}du\\ &= \int_0^{t-l}1\times\frac{r(l+u,l)}{\lambda(t,l)}du\\ &= 1\end{aligned} \tag{D.77}$$

Moreover, define $B^W(t) := \mathbb{E}(Y(t)(Y(t)-1))$ and $C^W(t,l) := \mathbb{E}(J(t,l)(J(t,l)-1))$. Note also $\mathbb{E}(J(t,l)) = \lambda(t,l)$. Thus, the second derivative evaluated at $s=1$ is

$$\begin{aligned}&2\bar{G}(t-l,l)\left[\int_0^{t-l}M(t,l+u)B(l+u)\frac{r(l+u,l)}{\lambda(t,l)}du\right]\kappa\lambda(t,l)\\ &+\bar{G}(t-l,l)\left[\int_0^{t-l}M(t,l+u)B(l+u)\frac{r(l+u,l)}{\lambda(t,l)}du\right]^2C^W(t,l)\\ &+\bar{G}(t-l,l)\left[\int_0^{t-l}W(t,l+u)B(l+u)\frac{r(l+u,l)}{\lambda(t,l)}du\right]\kappa\lambda(t,l)\\ &+\bar{G}(t-l,l)\left[\int_0^{t-l}M(t,l+u)^2B^W(l+u)\frac{r(l+u,l)}{\lambda(t,l)}du\right]\kappa\lambda(t,l)\end{aligned} \tag{D.78}$$

Noting that $J(t,l)$ is Poisson, one has

$$C^W(t,l)+\mathbb{E}(J(t,l))-\mathbb{E}(J(t,l)^2) = \mathrm{var}(J(t,l)) = \mathbb{E}(J(t,l)) \tag{D.79}$$

and hence

$$C^W(t,l) = \mathbb{E}(J(t,l))^2 \tag{D.80}$$

Define

$$\chi(t,l,k) := \kappa\Big[W(t,l+k)B(l+k)r(l+k,l)+M(t,l+k)^2B^W(l+k)r(l+k,l)\Big] \tag{D.81}$$

Then, the same process can be carried out for the second part of the equation to give

$$\begin{aligned}W(t,l) =&2\bar{G}(t-l,l)\bigg[\int_0^{t-l} M(t,l+u)B(l+u)r(l+u,l)du\bigg] + \bar{G}(t-l,l)\int_0^{t-l}\chi(t,l,k)dk...\\ &...+\bar{G}(t-l,l)\bigg[\int_0^{t-l} M(t,l+u)B(l+u)\kappa r(l+u,l)du\bigg]^2...\\ &...+\int_0^{t-l}\bigg[\int_0^{u} M(t,l+u)B(l+u)\kappa r(l+u,l)du\bigg]^2 g(u,l)du...\\ &...+\int_0^{t-l}\int_0^{u}\chi(t,l,k)dkg(u,l)du \end{aligned} \tag{D.82}$$

For ease of notation, define

$$S(t,l,u) := \bigg[\int_0^{u} M(t,l+k)B(l+k)\kappa r(l+k,l)dk\bigg]^2 \tag{D.83}$$

so that

$$\begin{aligned}W(t,l) =&2\bar{G}(t-l,l)\bigg[\int_0^{t-l} M(t,l+u)B(l+u)r(l+u,l)du\bigg] + \bar{G}(t-l,l)\int_0^{t-l}\chi(t,l,k)dk...\\ &...+\bar{G}(t-l,l)S(t,l,t-l)+\int_0^{t-l} S(t,l,u)g(u,l)du+\int_0^{t-l}\int_0^{u}\chi(t,l,k)dkg(u,l)du \end{aligned} \tag{D.84}$$

Now, changing the order of integration in the final term (as was done in the derivation of the mean prevalence), this can be rewritten as

$$\begin{aligned}W(t,l) =& 2\bar{G}(t-l,l)\bigg[\int_0^{t-l}\kappa M(t,l+u)B(l+u)r(l+u,l)du\bigg] + \bar{G}(t-l,l)S(t,l,t-l)...\\ &...+\int_0^{t-l} S(t,l,u)g(u,l)du+\int_0^{t-l}\chi(t,l,k)\bar{G}(k,l)dk \end{aligned} \tag{D.85}$$

From this, we can create an equation for $\mathbb{E}(Z(t,l)^2) := X(t,l) = W(t,l) + M(t,l)$ by defining

$$\chi^X(t,l,k) = \kappa\bigg[X(t,l+k)B(l+k,l)r(l+k,l) + M(t,l+k)^2 B^W(l+k,l)r(l+k,l)\bigg] \tag{D.86}$$

and then simply adding the equation for M to give

$$\begin{aligned}X(t,l) =& \bar{G}(t-l,l) + 2\bar{G}(t-l,l)\bigg[\int_0^{t-l}\kappa M(t,l+u)B(l+u)r(l+u,l)du\bigg] + \bar{G}(t-l,l)S(t,l,t-l)...\\ &...+\int_0^{t-l} S(t,l,u)g(u,l)du+\int_0^{t-l}\chi^X(t,l,k)\bar{G}(k,l)dk \end{aligned} \tag{D.87}$$

Finally, to form the equation for the variance $V(t,l) = X(t,l) - M(t,l)^2$, note that

$$\chi^X(t,l,k) = \kappa\Big[X(t,l+k)B(l+k)r(l+k,l) + M(t,l+k)^2(\mathbb{E}(Y(l+k)^2) - B(l+k))r(l+k,l)\Big] \tag{D.88}$$

$$= \kappa\Big[V(t,l+k)B(l+k)r(l+k,l) + M(t,l+k)^2\mathbb{E}(Y(l+k)^2)r(l+k,l)\Big] \tag{D.89}$$

$$:= \chi^V(t,l,k) \tag{D.90}$$

and hence, subtracting $M(t,l)^2$ from both sides of the equation for $X(t,l)$ gives

$$V(t,l) = \bar{G}(t-l,l) + 2\bar{G}(t-l,l)\Big[\int_0^{t-l} \kappa M(t,l+u)B(l+u)r(l+u,l)du\Big] + \bar{G}(t-l,l)S(t,l,t-l)... \\ ... + \int_0^{t-l} S(t,l,u)g(u,l)du + \int_0^{t-l} \chi^V(t,l,k)\bar{G}(k,l)dk - M(t,l)^2 \tag{D.91}$$

D.5.3 An explanation of the variance equation

There are two main sources of uncertainty in the infection process - the infectious period of an individual, and the number and timing of infections that occur during this infectious period. One can show that the variance splits into three terms - one for each of these two sources of uncertainty from the initial individual, and one which propagates the uncertainty through the descendants of the initial individual.

Each term will be derived by assuming that all other parts of the model are deterministic. To begin, suppose that the infectious period of the initial individual is random but all other parts of the model are deterministic, so that, given that the initial individual is infectious at time $l+u$, it will infect $B(l+u)r(l+u,l)dt$ people in the interval $[u, u+dt]$ (note that this is an abstraction to illustrate the source of this variance, as it is impossible for non-integer numbers of infections to occur). Moreover, it is assumed that each of these individuals have given rise to exactly $M(t,l+u)$ infections at time t.

Then, note that

$$\begin{aligned}
\text{var}(Z(t,l)) &= \mathbb{E}(Z(t,l)^2) - \mathbb{E}(Z(t,l))^2 && \text{(D.92)}\\
&= \int_0^\infty \mathbb{E}(Z(t,l)^2|L=u)g(u,l)du - M(t,l)^2 && \text{(D.93)}\\
&= \int_0^{t-l}\left[\int_0^u M(t,l+k)B(l+k)r(l+k,l)dk\right]^2 g(u,l)du... && \text{(D.94)}\\
&...+\bar{G}(t-l,l)\left(1+\int_0^{t-l} M(t,l+k)B(l+k)r(l+k,l)dk\right)^2 - M(t,l)^2\\
&= \bar{G}(t-l,l) + 2\bar{G}(t-l,l)\int_0^{t-l} M(t,l+k)B(l+k)r(l+k,l)dk + ... && \text{(D.95)}\\
&...+\bar{G}(t-l,l)S(t,l,t-l) + \int_0^{t-l} S(t,l,u)g(u,l)du - M(t,l)^2
\end{aligned}$$

which recovers all the terms of the variance equation except for $\int_0^{t-l}\chi^V(t,l,k)\bar{G}(k,l)dk$.

Now, suppose that the infectious period of the initial individual is deterministic in the sense that they infect others at a rate of $r(l+k,l)\bar{G}(k,l)$, i.e. the expected rate at time $l+k$. Thus, the number of infection events in the interval $[l+(k-1)dt, l+kdt]$ is (to leading order in dt) a Poisson variable, A_k, with mean $r(l+k,l)\bar{G}(k,l)dt$ and hence the number of infections is that Poisson variable multiplied by $Y(l+k,l)$. Finally, note that, as before, any individuals born at time $l+k$ will be assumed to deterministically cause $M(t,l+k)$ active infections at time t. Thus,

$$\begin{aligned}
\text{var}(Z(t,l)) &= \int_{k=0}^{k=t-l} \text{var}(M(t,l+k)Y(l+k)A_k) && \text{(D.96)}\\
&= \int_{k=0}^{k=t-l} \mathbb{E}((M(t,l+k)Y(l+k)A_k)^2) - \int_{k=0}^{k=t-l} \mathbb{E}((M(t,l+k)Y(l+k)A_k))^2 && \text{(D.97)}\\
&= \int_{k=0}^{k=t-l} M(t,l+k)^2\mathbb{E}(Y(l+k)^2)\mathbb{E}(A_k^2) - \int_{k=0}^{k=t-l} B(l+k)^2 r(l+k,l)^2\bar{G}(k,l)^2dt^2M(t,l+k)^2 && \text{(D.98)}
\end{aligned}$$

Ignoring the dt^2 term as it has zero measure, and noting that Y and A_k are independent

$$\text{var}(Z(t,l)) = \int_0^{t-l} M(t,l+k)^2\mathbb{E}(Y(l+k,l)^2)r(l+k,l)\bar{G}(k,l)dt \quad \text{(D.99)}$$

which is again a term from the variance equation.

The final term, $\int_0^{t-l} V(t,l+k)B(l+k)\bar{G}(k,l)r(l+k,l)dk$ denotes the propagation of uncertainty through future generations. Indeed, if the infection process of the initial individual (and its infectious

period) are assumed to be fully deterministic, then one simply has

$$\text{var}(Z(t,l)) = \int_0^{t-l} \text{var}(Z(t,l+k))\mathbb{E}(\text{number of individuals born at } l+k) \tag{D.100}$$

which can easily be seen to give the correct term.

D.5.4 Overdispersion

For the purposes of this note, it is helpful to create the following definition

Expanded: An epidemic is called "expanded" at time t, if there is a non-zero probability that the prevalence, not counting the initial individual or its secondary infections, is non-zero.

In this note, it will be shown that, if $\tilde{Z}(t,l)$ is the prevalence of *new* infections (that is, the prevalence without counting the initial case) then if the epidemic is expanded at time t, $\tilde{Z}(t,l)$ is strictly overdispersed. That is

$$\text{var}(\tilde{Z}(t,l)) > \mathbb{E}(\tilde{Z}(t,l)) \quad \text{or} \quad \mathbb{E}(\tilde{Z}(t,l+k))\rho(l+k,l)\nu(k)\bar{G}(k,l) = 0 \quad \forall k \in (0,t-l) \tag{D.101}$$

The second condition ensures that, at each k, either the likelihood of a new infection being caused at time $l+k$, or the probability of an individual who was infected at time $l+k$ causing subsequent infections whose infection tree has non-zero prevalence at time t, is zero. Hence, it is equivalent to the epidemic not being expanded at time t.

It is crucial to use $\tilde{Z}(t,l)$ rather than $Z(t,l)$, as otherwise the deterministic initial case means that, for early times, the prevalence is underdispersed (as, for example $\mathbb{E}(Z(l,l)) = 1$ and $\text{var}(Z(l,l)) = 0$). Moreover, the condition on the tertiary infections is necessary as, otherwise, if $N(t,l)$ is Poissonian, then $\tilde{Z}(t,l)$ is also Poissonian (and therefore not strictly overdispersed).

It is helpful to derive equations for the quantities for the mean $\tilde{M}(t,l)$ and the variance $\tilde{V}(t,l)$ of the new infection prevalence. This can be done by following the methods of the previous note. The derivations are mostly identical, and so will not be covered in detail. However, the key point is to note that the equation for the pgf, $\tilde{F}$, becomes

$$\begin{aligned}\tilde{F}(t,l) = \Big(1-G(t-l,l)\Big)\mathcal{J}_{(t,l)}\bigg(\int_0^{t-l}\mathcal{Y}_{(l+u)}(\tilde{F}(t,l+u))\frac{r(l+u,l)}{\lambda(t,l)}du\bigg)... \\ ... + \int_0^{t-l}\mathcal{J}_{(l+u,l)}\bigg(\int_0^{u}\mathcal{Y}_{(l+k)}(\tilde{F}(t,l+k))\frac{r(l+k,l)}{\lambda(l+u,l)}dk\bigg)g(u,l)du\end{aligned} \tag{D.102}$$

as the factor of s in the first term is discarded. This equation can then be differentiated as before to

show that

$$\tilde{M}(t,l) = (1-G(t-l,l))\Big[\int_0^{t-l} \tilde{M}(t,l+u)\frac{r(l+u,l)}{\lambda(t,l)}B(l+u)\kappa\lambda(t,l)du\Big]... \\ ... + \int_0^{t-l}\int_0^{u} \tilde{M}(t,l+k)\frac{r(l+k,l)}{\lambda(l+u,l)}B(l+k)\kappa\lambda(l+u,l)g(u,l)dkdu \tag{D.103}$$

and then rearranged to

$$\tilde{M}(t,l) = \int_0^{t-l} \tilde{M}(t,l+u)r(l+u,l)B(l+u)\kappa(1-G(u,l))du \tag{D.104}$$

Defining $\tilde{S}$ as the analogue to S, by

$$\tilde{S}(t,l,u) = \Big[\int_0^{u} \tilde{M}(t,l+k)B(l+k)\kappa r(l+k,l)dk\Big]^2 \tag{D.105}$$

and using $\bar{G} = 1 - G$, the first of these equations can be written more succinctly as

$$\tilde{M}(t,l) = \bar{G}(t-l,l)\tilde{S}(t,l,t-l)^{0.5} + \int_0^{t-l} \tilde{S}(t,l,u)^{0.5}g(u,l)du \tag{D.106}$$

The equation for $\tilde{V}(t,l)$ can be calculated in a similar way. The only changes to the derivation are that the first term in Supplementary Equation D.78 is discarded to account for the discarded s in the pgf, and that when adding the mean to move from W to X (in analogue to Supplementary Equation D.87), one no longer needs to add the $\bar{G}(t-l,l)$ term. Thus,

$$\tilde{V}(t,l) = \bar{G}(t-l,l)\tilde{S}(t,l,t-l) + \int_0^{t-l} \tilde{S}(t,l,u)g(u,l)du + \int_0^{t-l} \chi^{\tilde{V}}(t,l,k)\bar{G}(k,l)dk - \tilde{M}(t,l)^2 \tag{D.107}$$

Now, the proof of overdispersion can begin. Firstly, it is helpful to bound $\tilde{M}(t,l)$ above, which can be done as follows. Squaring Supplementary Equation D.106 shows that

$$\tilde{M}(t,l)^2 = \bar{G}(t-l,l)^2\tilde{S}(t,l,t-l) + 2\bar{G}(t-l,l)\tilde{S}(t,l,t-l)^{0.5}\int_0^{t-l} \tilde{S}(t,l,u)^{0.5}g(u,l)du... \tag{D.108}$$

$$... + \Big[\int_0^{t-l} \tilde{S}(t,l,u)^{0.5}g(u,l)du\Big]^2 \tag{D.109}$$

Now, using the Cauchy-Schwarz inequality, we see that

$$\left[\int_0^{t-l} \tilde{S}(t,l,u)^{0.5} g(u,l) du\right]^2 = \left[\int_0^{t-l} (\tilde{S}(t,l,u) g(u,l))^{0.5} (g(u,l))^{0.5} du\right]^2 \tag{D.110}$$

$$\leq \left[\int_0^{t-l} \tilde{S}(t,l,u) g(u,l) du\right] \left[\int_0^{t-l} g(u,l) du\right] \tag{D.111}$$

$$\leq (1 - \bar{G}(t-l,l)) \left[\int_0^{t-l} \tilde{S}(t,l,u) g(u,l) du\right] \tag{D.112}$$

Suppose that $\bar{G}(t-l,l) \neq 1$. Then, using

$$1 = \frac{1}{1-\bar{G}(t-l,l)} - \frac{\bar{G}(t-l,l)}{1-\bar{G}(t-l,l)} \tag{D.113}$$

to split the final term in Supplementary Equation D.108, we find

$$\tilde{M}(t,l)^2 \leq \bar{G}(t-l,l)^2 \tilde{S}(t,l,t-l) + 2\bar{G}(t-l,l)\tilde{S}(t,l,t-l)^{0.5} \int_0^{t-l} \tilde{S}(t,l,u)^{0.5} g(u,l) du ...$$
$$- \frac{\bar{G}(t-l,l)}{1-\bar{G}(t-l,l)} \left[\int_0^{t-l} \tilde{S}(t,l,u)^{0.5} g(u,l) du\right]^2 + \left[\int_0^{t-l} \tilde{S}(t,l,u) g(u,l) du\right] \tag{D.114}$$

To facilitate the remainder of this proof, it is helpful to define

$$Q(t,l) := \int_0^{t-l} \tilde{S}(t,l,u)^{0.5} g(u,l) du \tag{D.115}$$

Note that $Q(t,l) \geq 0$ as $\tilde{S}$ and g are non-negative. Moreover, for fixed t and l, the function $\tilde{S}(t,l,u)^{0.5}$ is non-decreasing in u and hence

$$Q(t,l) \leq \int_0^{t-l} \tilde{S}(t,l,t-l)^{0.5} g(u,l) du = \tilde{S}(t,l,t-l)^{0.5}(1-\bar{G}(t-l,l)) \tag{D.116}$$

Consider the function

$$f(Q) = 2\bar{G}(t-l,l)\tilde{S}(t,l,t-l)^{0.5} Q - \frac{\bar{G}(t-l,l)}{1-\bar{G}(t-l,l)} Q^2 \tag{D.117}$$

for $Q \in [0, \tilde{S}(t,l,t-l)^{0.5}(1-\bar{G}(t-l,l))]$. f is a quadratic, and has a single turning point at

$$f'(Q) = 0 \Rightarrow Q = \tilde{S}(t,l,t-l)^{0.5}(1-\bar{G}(t-l,l)) \tag{D.118}$$

This is an endpoint of the domain of Q and hence the maximum value of $f(Q)$ must occur one of the

endpoints. $f(0) = 0$ and

$$f\left(\tilde{S}(t,l,t-l)^{0.5}(1-\bar{G}(t-l,l))\right) = \bar{G}(t-l,l)(1-\bar{G}(t-l,l))\tilde{S}(t,l,t-l) \tag{D.119}$$

This is non-negative, and hence the maximal value of $f(Q)$.

This can be put into the equation for $\tilde{M}(t,l)^2$ to give

$$\begin{aligned}\tilde{M}(t,l)^2 &\leq \bar{G}(t-l,l)^2\tilde{S}(t,l,t-l) + \bar{G}(t-l,l)(1-\bar{G}(t-l,l))\tilde{S}(t,l,t-l) + \left[\int_0^{t-l} S(t,l,u)g(u,l)du\right] \\ &= \bar{G}(t-l,l)\tilde{S}(t,l,t-l) + \left[\int_0^{t-l} S(t,l,u)g(u,l)du\right] \end{aligned} \tag{D.120}$$

Both the terms on the right-hand side appear in the equation for $\tilde{V}$, and hence, substituting this result in shows that

$$\tilde{V}(t,l) \geq \int_0^{t-l} \chi^{\tilde{V}}(t,l,k)\bar{G}(k,l)dk \tag{D.121}$$

As this holds for all $\bar{G}(t-l,l) < 1$, it must also (under relevant continuity assumptions) hold for $\bar{G}(t-l,l) = 1$, and hence in all cases. Now,

$$\chi^{\tilde{V}}(t,l,k) = \tilde{V}(t,l+k)B(l+k)r(l+k,l) + \tilde{M}(t,l+k)^2\mathbb{E}(Y(l+k,l)^2)r(l+k,l) \tag{D.122}$$

As $Y \geq 1$ by definition, one has

$$B(l+k) = \mathbb{E}(Y(l+k,l)) \leq \mathbb{E}(Y(l+k,l)^2) \tag{D.123}$$

and hence

$$\chi^{\tilde{V}}(t,l,k) \leq \left[\tilde{V}(t,l+k) + \tilde{M}(t,l+k)^2\right]B(l+k)r(l+k,l) = \mathbb{E}(\tilde{Z}(t,l+k)^2)B(l+k)r(l+k,l) \tag{D.124}$$

Finally, as $\tilde{Z}(t,l+k) \geq 0$ and is integer-valued, one has $\tilde{Z}(t,l+k)^2 \geq \tilde{Z}(t,l+k)$ and hence

$$\chi^{\tilde{V}}(t,l,k) \leq \tilde{M}(t,l+k)B(l+k)r(l+k,l) \tag{D.125}$$

Thus,

$$\tilde{V}(t,l) \geq \int_0^{t-l} \tilde{M}(t,l+k)B(l+k)r(l+k,l)\bar{G}(k,l)dk = \tilde{M}(t,l+k) \tag{D.126}$$

which proves weak overdispersion.

To prove strict overdispersion, note that, for Supplementary Equation D.126 to hold to equality, it is necessary that all the inequalities used hold to equality. Thus, in particular, it is necessary that

$$\int_0^{t-l} \tilde{M}(t,l+k)B(l+k)r(l+k,l)\bar{G}(k,l)dk = \int_0^{t-l} \mathbb{E}(\tilde{Z}(t,l+k)^2)B(l+k)r(l+k,l)\bar{G}(k,l)dk \tag{D.127}$$

and hence, as $B(l+k) \geq 1$,

$$r(l+k,l)\bar{G}(k,l) \geq 0 \Rightarrow \mathbb{E}(\tilde{Z}(t,l+k)^2) = \tilde{M}(t,l+k) \tag{D.128}$$

This means that

$$r(l+k,l)\bar{G}(k,l) \geq 0 \Rightarrow \mathbb{E}(\tilde{Z}(t,l+k)(\tilde{Z}(t,l+k)-1)) = 0 \tag{D.129}$$

and hence, as $\tilde{Z}(t,l+k)(\tilde{Z}(t,l+k)-1)$ is a non-negative integer, this means that

$$r(l+k,l)\bar{G}(k,l) \geq 0 \Rightarrow \tilde{Z}(t,l+k)(\tilde{Z}(t,l+k)-1) = 0 \tag{D.130}$$

almost surely. We now show that if $\mathbb{P}(\tilde{Z}(t,l)=1) > 0$, then $\mathbb{P}(\tilde{Z}(t,l)>1) > 0$. This can be done as follows.

Define the set $\mathcal{S}$ to be the possible times at which the initial individual can cause a secondary infection which in turn starts an epidemic that can have non-zero prevalence at time t. Then,

$$\mathcal{S} = \left\{ u \in (l, t-l) : r(l+u,l) > 0, \quad \bar{G}(u,l) > 0 \quad \text{and} \quad \mathbb{P}(Z(t,l+u) > 0) > 0 \right\} \tag{D.131}$$

Note the use of Z rather than $\tilde{Z}$. The first two conditions ensures that the likelihood of the initial individual causing an infection at time u is non-zero (as it must have non-zero rate here, and also a non-zero probability of still being infectious). The third condition ensures that the probability of this secondary infection's infection tree still containing at least one infectious individual at time t is non-zero. It is necessary that

$$\int_{\mathcal{S}} r(l+u,l)\bar{G}(u,l)\mathbb{P}(Z(t,l+u)>0)du > 0 \tag{D.132}$$

as otherwise, $\tilde{Z}(t,l) = 0$ (as this integral sums over all possible epidemics that lead to $\tilde{Z}(t,l) > 0$). Define

$$\mathcal{S}(x) := \mathcal{S} \cap (l, x) \tag{D.133}$$

and the function

$$f(x) = \int_{\mathcal{S}(x)} r(l+u, l)\bar{G}(u, l)\mathbb{P}(Z(t, l+u) > 0)du \tag{D.134}$$

Then, f must be continuous, and so there exists some $y \in (0, t-l)$ such that

$$0 < f(y) < f(t-l) = \int_{\mathcal{S}} r(l+u, l)\bar{G}(u, l)\mathbb{P}(Z(t, l+u) > 0)du \tag{D.135}$$

Thus, there is a non-zero probability of an individual being infected in $(l, l+y)$ causing an epidemic that has non-zero prevalence at time t and, similarly, a non-zero probability of an individual being infected in $(l+y, t)$ causing an epidemic that has non-zero prevalence at time t. Thus, as the infections processes have independent increments and as the initial individual causing an infection in $(l+y, t)$ implies that it must have been infectious for the whole interval $(l, l+y)$, there is a non-zero probability of two such individuals being infected: one in $(l, l+y)$ and one in $(l+y, t)$. Hence,

$$\mathbb{P}(\tilde{Z}(t,l) = 1) > 0 \Rightarrow \mathbb{P}(\tilde{Z}(t,l) > 1) > 0 \tag{D.136}$$

as required. Thus,

$$r(l+k, l)\bar{G}(k, l) \geq 0 \Rightarrow \tilde{Z}(t, l+k) = 0 \tag{D.137}$$

and so

$$\mathbb{E}(\tilde{Z}(t, l+k))r(l+k, l)\bar{G}(k, l) = 0 \quad \forall k \tag{D.138}$$

Thus, we have strict overdispersion, $\tilde{V}(t,l) > \tilde{M}(t,l)$, provided that the epidemic is expanded at time t, as required.

D.5.5 Comparison to a Poisson case

Consider comparing the variance Supplementary Equation D.91 with the variance of an epidemic where infection events are always of size 1 (that is, where the counting process of infections, $N^*(t,l)$ is a Poisson case, meaning $B^*(t) = 1$). Asterisks will be used to denote the quantities relating to this Poisson epidemic.

Suppose that the infectious period is the same in both cases (so $G = G^*$ and $\nu = \nu^*$). To ensure a fair comparison, it is also assumed that the mean number of cases is the same in both cases with

$M(t,l) = M^*(t,l)$. By examining the Supplementary Equation D.72 for the mean, and absorbing κ into ρ in both cases, one can see

$$B(l+u)\rho(l+u) = \rho^*(l+u). \tag{D.139}$$

The variance Supplementary Equation D.91 can now be examined. Firstly, note that

$$\int_0^{t-l} M(t,l+u)B(l+u)r(l+u,l)du = \int_0^{t-l} M^*(t,l+u)r^*(l+u,l)du, \tag{D.140}$$

using the result above and the fact that $M(t,l+u) = M^*(t,l+u)$. Similarly,

$$S(t,l,u) = S^*(t,l,u). \tag{D.141}$$

Thus,

$$\begin{aligned} V(t,l) - V^*(t,l) &= \int_0^{t-l} (\chi^V(t,l,k) - \chi^{V^*}(t,l,k))\bar{G}(k,l)dk && \text{(D.142)} \\ &= \int_0^{t-l} \Big(V(t,l+k) - V^*(t,l+k)\Big)B(l+k)r(l+k,l)\bar{G}(k,l)dk... && \text{(D.143)} \\ &+ \int_0^{t-l} \Big(\mathbb{E}(Y(l+k)^2) - 1\Big)M(t,l+k)^2 r(l+k,l)\bar{G}(k,l)dk && \text{(D.144)} \end{aligned}$$

By defining $\Delta^V(t,l) := V(t,l) - V^*(t,l)$, one can see that this is a renewal equation

$$\Delta^V(t,l) = \int_0^{t-l} \Big(\mathbb{E}(Y(l+k)^2) - 1\Big)M(t,l+k)^2 r(l+k,l)\bar{G}(k,l)dk + \int_0^{t-l} \Delta^V(t,l+k)B(l+k)r(l+k,l)\bar{G}(k,l)dk. \tag{D.145}$$

An important property of this renewal equation is that the part that is independent of Δ^V on the right-hand side grows. That is,

$$\Delta^V(t,l) \geq \int_0^{t-l} \Big(\mathbb{E}(Y(l+k)^2) - 1\Big)M(t,l+k)^2 r(l+k,l)\bar{G}(k,l)dk. \tag{D.146}$$

Thus, even though these two epidemics give the same mean, the difference in their variances is proportional to the square of this mean. This means that models fitted to a Poisson process framework, even without exponential infectious periods, will substantially underestimate the variance of the number of cases (recalling that $\mathbb{E}(Y(l+k)^2) > 1$ in the non-Poisson case).

D.5.6 Large time solutions to the variance equation

To further understand the variance, we consider large time approximate solutions to the variance equation. Note that the level of rigour in this note is lower than the rest of our derivations as the results are derived for illustrative purposes.

It shall be assumed throughout this note that κ has been absorbed into ρ. Moreover, to enable explicit asymptotic solutions to be found, it shall be assumed that ρ, B and $\mathbb{E}(Y^2)$ are constants and that $g = g(t)$. Therefore, all individuals behave identically (in distribution), irrespective of the time at which they were infected. Moreover, it means that $r(l+k,l) = r(k)$, as the rate of infection depends only on the time since the individual has been infected

Under these assumptions, the mean $M(t,l) = M(t-l)$ and the variance $V(t,l) = V(t-l)$ are functions of $t-l$ only. This property will be used when forming the heuristics used in this note.

The final assumption is that $\bar{G}(t)$ has a finite support - that is, $\bar{G}(t) = 0$ for sufficiently large t. This is not strictly necessary, but simplifies the analysis.

Then, for $t >> l$, the mean and variance equations become

$$M(t,l) = \int_0^{t-l} M(t,l+u)B\rho\nu(u)\bar{G}(u)du \tag{D.147}$$

and

$$V(t,l) = \int_0^{t-l} S(t,l,u)g(u,l) + \int_0^{t-l} \chi^V(t,l,k)\bar{G}(k)dk - M(t,l)^2. \tag{D.148}$$

Motivated by the exponential growth of epidemics without susceptible depletion, consider the heuristic

$$M(t,l) = e^{\gamma(t-l)} \tag{D.149}$$

for some growth rate γ (note that in Supplementary Equation D.147, scaling M by a constant does not affect the solution). Then, Supplementary Equation D.147 becomes

$$e^{\gamma(t-l)} = e^{\gamma(t-l)} \int_0^{t-l} e^{-\gamma u} B\rho\nu(u)\bar{G}(u)du. \tag{D.150}$$

Now, assuming that $t-l >> 1$, as the integrand has finite support,

$$e^{\gamma(t-l)} = e^{\gamma(t-l)} \int_0^{\infty} e^{-\gamma u} B\rho\nu(u)\bar{G}(u)du = e^{\gamma(t-l)}H(\gamma), \tag{D.151}$$

where $H(\gamma)$ is a monotonically decreasing function such that $H(-\infty) = \infty$ and $H(\infty) = 0$. It is

necessary that

$$H(\gamma) = 1 \tag{D.152}$$

and, by the above notes on H, there is a unique value for γ (independent of l) such that this holds. We shall henceforth assume that γ is equal to this value.

Note that (by considering the case $\gamma = 0$)

$$\gamma > 0 \Leftrightarrow \int_0^\infty B\rho\nu(u)\bar{G}(u)du > 1 \tag{D.153}$$

and so the epidemic grows if and only if the expected number of cases caused by an individual is greater than 1, as expected.

The variance equation can now be considered. Note that

$$S(t,l,u) = \left[\int_0^u M(t,l+k)Br(k)dk\right]^2 = e^{2\gamma(t-l)}\left[\int_0^u e^{-\gamma k}Br(k)dk\right]^2. \tag{D.154}$$

Hence, the equation for the variance becomes

$$\begin{aligned} V(t,l) = e^{2\gamma(t-l)}\int_0^{t-l}\left[\int_0^u e^{-\gamma k}Br(k)dk\right]^2 g(u)du + \int_0^{t-l} V(t,l+k)Br(k)\bar{G}(k)dk... \\ + e^{2\gamma(t-l)}\int_0^{t-l} e^{-2\gamma k}\mathbb{E}(Y^2)r(k)\bar{G}(k)dk - e^{2\gamma(t-l)}. \end{aligned} \tag{D.155}$$

Note the χ^V term has been split into the two single integrals with integration variable k. This equation motivates a heuristic

$$V(t,l) = Ce^{2\gamma(t-l)}, \tag{D.156}$$

which, again using the fact that the integrands have finite support, results in

$$C = \frac{\int_0^\infty \left[\int_0^u e^{-\gamma k}Br(k)dk\right]^2 g(u)du + \int_0^\infty e^{-2\gamma k}\mathbb{E}(Y^2)r(k)\bar{G}(k)dk - 1}{1 - \int_0^\infty e^{-2\gamma k}Br(k)\bar{G}(k)dk}. \tag{D.157}$$

Note that if $\gamma \leq 0$, this variance approximation is not well-defined (as C is either infinite if $\gamma = 0$ or negative if $\gamma < 0$) and so it is necessary to find another solution. In the $\gamma < 0$ case, $e^{\gamma(t-l)} >> e^{2\gamma(t-l)}$ and a leading-order solution can be found simply from

$$V = e^{\gamma(t-l)}. \tag{D.158}$$

Thus, according to these approximations, the variance grows with the square of the mean in the $\gamma > 0$

(i.e. growing epidemic) case, while it decreases proportionally to the mean in the $\gamma < 0$ (i.e. shrinking epidemic) case. The $\gamma = 0$ case is the bifurcation point between these two solutions and would require further analysis.

In the growing epidemic case, the equation for C is also informative in characterising the effect of the different model parameters on the variance. In particular, it shows that there is a linear relationship between $\mathbb{E}(Y(t)^2)$ and the variance, re-emphasising the point made in the previous subsection that ignoring this parameter can have significant effects on the variance estimate. Moreover, it shows that variance grows rapidly throughout a growing epidemic, remaining proportional to the square of the mean.

D.5.7 Mean and variance for cumulative incidence

The equations for the mean and prevalence of the cumulative incidence of the epidemic can be derived almost identically, as the two generating functions are very similar. The mean equation gains an term from the additional s being differentiated, which is

$$\int_0^{t-l} \mathcal{J}_{(l+u)} \left(\int_0^u \mathcal{Y}_{(l+k,l)(1)} \frac{r(l+k,l)}{\lambda(l+u,l)} dk \right) g(u,l) du = G(t-l,l) \tag{D.159}$$

and hence, the mean equation becomes (using *s to denote cumulative incidence quantities)

$$M^*(t,l) = 1 + \int_0^{t-l} M^*(t,l+u)\rho(l+u)\nu(u)\bar{G}(u,l)du \tag{D.160}$$

Now, the only difference in the equation for W in the case of cumulative incidence is that the term Supplementary Equation D.78 appears in both parts (again due to the extra s term). This can be treated in the same way as χ in the original derivation and so

$$\begin{aligned} W^*(t,l) = 2\int_0^{t-l} \kappa M^*(t,l+u)B(l+u)r(l+u,l)\bar{G}(u,l)du + \bar{G}(t-l,l)\tilde{S}(t,l,t-l)... \\ ... + \int_0^{t-l} \tilde{S}(t,l,u)g(u,l)du + \int_0^{t-l} \chi^*(t,l,k)\bar{G}(k,l)dk \end{aligned} \tag{D.161}$$

Again, following the previous derivation, one can then arrive at

$$\begin{aligned} V^*(t,l) = 1 + 2\int_0^{t-l} \kappa M^*(t,l+u)B(l+u)r(l+u,l)\bar{G}(u,l)du + \bar{G}(t-l,l)\tilde{S}(t,l,t-l)... \\ ... + \int_0^{t-l} \tilde{S}(t,l,u)g(u,l)du + \int_0^{t-l} \chi^{V^*}(t,l,k)\bar{G}(k,l)dk - M^*(t,l)^2 \end{aligned} \tag{D.162}$$

D.6 Likelihood functions

D.6.1 Continuous case

If only the cumulative incidence, $Z(t, l)$, is known at some time t, the full epidemic history - in particular, the times at which each individual was infected, and the times at which they stopped being infectious - are unknown. Thus, it is helpful to derive a likelihood function for each possible sequence of these times.

Perhaps the most intuitive approach would be to treat the times at which each individual was infected as continuous random variables. However, the resultant pdf is complicated by the fact that multiple infections are likely to happen simultaneously if $\mathbb{E}(Y) > 1$, and will have a significant number of Kronecker delta functions to accommodate this, making it complicated both mathematically and practically.

To remedy this, we instead consider three sets of random variables - a vector $\boldsymbol{T}$ of unknown size $n + 1$, which contains the times of all the infection events up to time t; a vector $\boldsymbol{Y}$ also of size $n + 1$, which contains the size of each of these infection events (that is, y_m is the number of individuals that are infected at time τ_m); and a vector $\boldsymbol{D}$ containing the times at which each individual stops being infected. To make the subsequent notation clearer, we shall use a rectangular array $\boldsymbol{X}$ in place of $\boldsymbol{D}$, where X_{ij} will be the time at which the jth individual infected at time T_i stops being infected.

We will suppose that for each $s > u$ and positive integer k

$$\mathbb{P}(N(s + dt, u) - N(s, u) = k) = p_k(s, u)dt + o(dt) \tag{D.163}$$

and that

$$\mathbb{P}(N(s + dt, u) - N(s, u) = 0) = 1 - \sum_{k\geq 1} p_k(s, u)dt + o(dt) = 1 - r(s, u)dt + o(dt) \tag{D.164}$$

as the counting process of jumps in $N(s, u)$ is an inhomogeneous Poisson Process of rate $r(s, u)$ (absorbing the κ into r). We can hence create a likelihood function. Define $\mathbf{1}$ to be a vector of 1's, and choose any vectors $\boldsymbol{\tau}$ and $\boldsymbol{d}$ such that each $\tau_i, d_j \in (0, t)$. Define dt to be small enough so that $\tau_i - \tau_j > dt$ for all $i > j$ and so that $|\tau_i - d_j| > dt$ for all i, j (note that, the set where $\tau_i = d_j$ has zero

measure and can be ignored). Moreover, choose a positive-integer-valued vector $\boldsymbol{y}$. Then,

$$\mathbb{P}(\boldsymbol{T} \in [\boldsymbol{\tau}, \boldsymbol{\tau} + dt\mathbf{1}], \boldsymbol{D} \in [\boldsymbol{d}, \boldsymbol{d} + dt\mathbf{1}], \boldsymbol{Y} = \boldsymbol{y}) = P\Bigg[\Bigg(\bigcap_{k=1}^{n}\{y_k \text{ infections in } [\tau_k, \tau_k + dt]\}\Bigg)... \\ ... \cap \Bigg(\bigcap_{k=0}^{n}\{\text{no infections in } [\tau_k + dt, \tau_{k+1}]\Bigg) \cap \Bigg(\bigcap_{i=0}^{n}\bigcap_{j=1}^{y_i}\{L \in [x_{ij} - \tau_i, x_{ij} - \tau_i + dt]\}\Bigg)\Bigg] \tag{D.165}$$

where $\tau_{n+1} := t$ to reduce notation, x_{ij} is the value of X_{ij} in the case $\boldsymbol{D} = \boldsymbol{d}$ and L is a random variable equal in distribution to the infectious period of an individual. Each of the infection events in the above equation occur on disjoint subintervals of $[0, t]$ and so, as all of the processes $N(t, l)$ have independent increments, and each individual behaves independently of each other and their infectious periods, they can be considered separately. We have

$$\mathbb{P}(y_k \text{ infections in } [\tau_k, \tau_k + dt]) = \sum_{i=0}^{k-1}\sum_{j=0}^{y_i} \mathbb{1}_{\{x_{ij} < \tau_k\}} p_{y_k}(\tau_k, \tau_i)dt + o(dt) \tag{D.166}$$

Here, the $o(dt)$ term contains three components that can be linearised out of the model - the probability that multiple different individuals contribute to the y_k cases (this is $O(dt^2)$); the probabilities of individuals infecting no one in this interval (these are independently $1 - O(dt)$ and hence the $O(dt)$ contribution can be ignored when these probabilities are multiplied together); and the $o(dt)$ terms from the equations defining p_k.

As the counting process of jumps in $N(s, u)$ is an inhomogeneous Poisson Process, and it is only "active" for individual ij up to time x_{ij},

$$\mathbb{P}(\text{no infections in } [\tau_k + dt, \tau_{k+1}]) = \prod_{i=0}^{k}\prod_{j=1}^{y_i} \exp\Bigg(-\int_{\min(x_{ij}, \tau_k)}^{\min(x_{ij}, \tau_{k+1})} r(u, \tau_i)du\Bigg) + O(dt) \tag{D.167}$$

where here, the $O(dt)$ term contains the integral between τ_k and $\tau_k + dt$ of each of the integrands. Taking the products inside the exponential as sums, the various "no infection" terms can be combined together to give

$$P\Bigg(\bigcap_{k=0}^{n}\{\text{no infections in } [\tau_k + dt, \tau_{k+1}]\}\Bigg) = \exp\Bigg(-\sum_{i=0}^{n}\sum_{j=0}^{y_i}\int_{\tau_i}^{\min(t, x_{ij})} r(u, \tau_i)du\Bigg) \tag{D.168}$$

Finally, the infectious period terms can be simply calculated from the pdf, g, of L as

$$\mathbb{P}(L \in [x_{ij} - \tau_i, x_{ij} - \tau_i + dt]) = g(x_{ij} - \tau_i, \tau_i)dt + o(dt) \tag{D.169}$$

Hence, combining all the relevant terms,

$$\mathbb{P}(T \in [\boldsymbol{\tau}, \boldsymbol{\tau} + dt\mathbf{1}], \boldsymbol{D} \in [\boldsymbol{d}, \boldsymbol{d} + dt\mathbf{1}], \boldsymbol{Y} = \boldsymbol{y}) = o(dt^{n+Z(t,l)}) +$$
$$\prod_{k=1}^{n} \left[\left(\prod_{j=1}^{y_k} g(x_{kj} - \tau_k, \tau_k) \right) \left(\sum_{i=0}^{k-1} \sum_{j=0}^{y_i} \mathbb{1}_{\{X_{ij} < \tau_k\}} p_{y_k}(\tau_k, \tau_i) \right) \right] \exp\left(- \sum_{i=0}^{n} \sum_{j=0}^{y_i} \int_{\tau_i}^{\min(t, X_{ij})} r(u, \tau_i) du \right) (dt)^{n+Z(t,l)} \tag{D.170}$$

and thus, taking $dt \to 0$ gives a likelihood function of

$$L(\boldsymbol{\tau}, \boldsymbol{y}, \boldsymbol{d}) = \prod_{k=1}^{n} \left[\left(\prod_{j=1}^{y_k} g(x_{kj} - \tau_k, \tau_k) \right) \left(\sum_{i=0}^{k-1} \sum_{j=0}^{y_i} \mathbb{1}_{\{x_{ij} < \tau_k\}} p_{y_k}(\tau_k, \tau_i) \right) \right] \exp\left(- \sum_{i=0}^{n} \sum_{j=0}^{y_i} \int_{\tau_i}^{\min(t, x_{ij})} r(u, \tau_i) du \right) \tag{D.171}$$

It is simple to substitute in the two examples that have been previously considered. In both cases, $r(a, b) = \rho(a)\nu(a - b)$. In the Poisson case, one has $p_1(a, b) = \rho(a)\nu(a - b)$ and $p_k(a, b) = 0$ for $k > 1$. In the Negative Binomial case, the values of p_k are given by

$$p_k(a, b)dt = \lim_{t \to 0} \left(\frac{\mathbb{P}(T \in [\boldsymbol{\tau}, \boldsymbol{\tau} + dt\mathbf{1}], \boldsymbol{D} \in [\boldsymbol{d}, \boldsymbol{d} + dt\mathbf{1}], \boldsymbol{Y} = \boldsymbol{y})}{dt^{n+Z(t,l)}} \right) \tag{D.172}$$
$$= \mathbb{P}(J_M(a + dt, b) - J_M(a, b) = 1)\mathbb{P}(Y = k) \tag{D.173}$$
$$= \rho(a)\nu(b - a) \left(\frac{(1 - p)^k}{-k \ln(p)} \right) \tag{D.174}$$

D.6.2 Special case (Poisson)

In the Poisson case, $A_{k,i}$ is Poisson distributed with mean $\rho(k)\nu(k - i)$. Hence,

$$\mathcal{A}_k(\boldsymbol{b}, \boldsymbol{y}, \boldsymbol{d}) \sim \text{Poi}\left(\rho(k) \sum_{i=0}^{k-1} \nu(k - i) \sum_{j=1}^{y_i} \mathbb{1}_{\{x_{ij} \leq k\}} \right) := \text{Poi}(\mu_k) \tag{D.175}$$

and so, the more computationally useful log-likelihood is

$$\ell(\boldsymbol{\tau}, \boldsymbol{y}, \boldsymbol{D}) = \sum_{k=1}^{n} (\mu_k \log(y_k) - \mu_k - \log(y_k!)) + \sum_{i=1}^{n} \sum_{j=1}^{y_i} \log(g(x_{ij} - \tau_i, \tau_i)) \tag{D.176}$$

D.6.3 Special case (Negative Binomial)

In the Negative Binomial case,

$$A_{k,i} =_D \sum_{j=1}^{N} Y_j \tag{D.177}$$

where the Y_j are iid logarithmic random variables with a pmf given by Supplementary Equation D.44 that is independent of the properties of the individual, and N is Poisson distributed with mean $\rho(k)\nu(k-i)$. Thus,

$$\mathcal{A} \sim \text{NB}(\phi\mu_k, p) \tag{D.178}$$

where, as before, $p = \frac{\phi}{1+\phi}$ and μ_k is defined in the previous note. Hence, as

$$\log\left[P\Big(NB(a,p)=k\Big)\right] = \sum_{j=0}^{k-1}\log(a+j) + k\log(1-p) + a\log(p) - \log(k!) \tag{D.179}$$

we have

$$\ell(\boldsymbol{\tau}, \boldsymbol{y}, \boldsymbol{D}) = \sum_{k=0}^{n}\left[\log(\phi\mu_k + j) + y_k\log\left(\frac{1}{1+\phi}\right) + \phi\mu_k\log\left(\frac{\phi}{1+\phi}\right) - \log(y_k!)\right] + \sum_{i=1}^{n}\sum_{j=1}^{y_i}\log(g(x_{ij}-\tau_i, \tau_i)) \tag{D.180}$$

D.6.4 Approximating the likelihood

It is difficult to simulate from the likelihoods when the infectious periods of the individuals are unknown because often, $Z(t,l) >> t$ (whereas the other unknowns, $\boldsymbol{\tau}$ and $\boldsymbol{y}$ have only $n \sim t$ parameters). To remedy this, we use an approximation - given an estimate of the function g, we simulate

$$D_i = L_i + \tau_i \quad \text{where } L_i \sim g \tag{D.181}$$

For some $\boldsymbol{D}$, the observed epidemic may be impossible (e.g. if, $D_0 < b_1$, where b_1 is the time that the first infection event occurs). Thus, it is necessary to impose a feasibility condition. Many such conditions are possible, but we use a simple condition by defining

$$L_i^* := \min(L_i, \tau_{i+1} - \tau_i) \tag{D.182}$$

and then define

$$D_i^* := \tau_i + L_i^* \tag{D.183}$$

Given these values, we can then create an approximation, ℓ^* to be

$$\ell^*(\boldsymbol{\tau}, \boldsymbol{y}) \sim \ell(\boldsymbol{\tau}, \boldsymbol{y}, \boldsymbol{D}^*) \tag{D.184}$$

This clearly creates a non-deterministic likelihood as it is dependent on a set of random variables. However, from our simulations, it appears that ℓ^* has a small variance, and so this extra randomness does not significantly affect our calculations.

D.7 Assessing future variance during an epidemic

Many of the equations presented thus far have been concerned with properties of an epidemic started from a single case at a fixed deterministic time. However, it is crucial to be able to calculate the risk from any time during the epidemic, and such a derivation is presented in this note. This derivation is more algebraically involved than the other work in this paper, and so to reduce its length, it will be assumed that $N(t, l)$ is an inhomogeneous Poisson Process, and that $L = \infty$ for each individual. This means that $\boldsymbol{y}$ and $\boldsymbol{D}$ can be ignored when considering the likelihood.

D.7.1 Derivation

Suppose that the prevalence (or, equivalently in this case, cumulative incidence), $Z(t, l) = n + 1$, is known at some point in an epidemic, but that the times at which these infections happened, B_i, are unknown. Note that the notation B_i rather than T_i is used in this note, because these times are now an exact analogue of birth times in a birth-death process. The condition of $n + 1$ rather than n has been chosen as this means that there have been n new infections and will make the following derivation notationally simpler.

Note that the infection time of the initial individual, B_0 is known to be equal to l, but it will be treated identically to the other times to reduce notation. Its marginal pdf is $f_{B_0}(b) = \delta(b - l)$. Following the previous note, the pdf $f_{\boldsymbol{B}}(\boldsymbol{b})$ of the infection times is

$$f_{\boldsymbol{B}}(\boldsymbol{b}) = \frac{1}{\mathbb{P}(Z(t,l) = n)} \prod_{i=1}^{n} \left(\rho(b_i) \sum_{j=0}^{i-1} \nu(b_i - b_j) \right) \exp\left[- \sum_{i=0}^{n} \int_0^{t-b_i} \rho(s + l)\nu(s) ds \right] \tag{D.185}$$

Now, one can write

$$Z(t + s, l) = \sum_{i=0}^{n} Z_i^*(t + s, B_i) \tag{D.186}$$

where $Z_i^*(t + s, B_i)$ counts the infection tree started at the individual infection at b_i, considering only those infections that occurred after time t (that is, if this individual infects someone at time $a < t$, the infections of this second individual will *not* be counted, even if they occur after time t).

This can be rewritten as

$$Z(t+s,l) = \int_{b=0}^{t} \sum_{i=0}^{n} Z_i^*(t+s,b)\mathbb{1}_{\{B_i=b\}} \tag{D.187}$$

where here, $\mathbb{1}$ is the indicator function. Hence,

$$\begin{aligned} \text{var}(Z(t+s,l)) &= \text{var}\bigg(\int_{b=0}^{t} \sum_{i=0}^{n} Z_i^*(t+s,b)\mathbb{1}_{\{B_i=b\}} \bigg) && \text{(D.188)} \\ &= \int_{b=0}^{t} \sum_{i=0}^{n} \text{var}(Z_i^*(t+s,b)\mathbb{1}_{\{B_i=b\}})... \\ &\quad ... + \int_{b=0}^{t}\int_{c=0}^{t} \sum_{i=0}^{n}\sum_{j=0}^{n} \text{cov}\bigg(Z_i^*(t+s,b)\mathbb{1}_{\{B_i=b\}}, Z_j^*(t+s,b)\mathbb{1}_{\{B_j=c\}} \bigg)(\mathbb{1}_{\{(b,i)\neq(c,j)\}}) && \text{(D.189)} \end{aligned}$$

The first term in this equation can be expanded as

$$\begin{aligned} \text{var}(Z_i^*(t+s,b)\mathbb{1}_{\{B_i=b\}}) &= \mathbb{E}(Z_i^*(t+s,b)^2\mathbb{1}_{\{B_i=b\}}^2) - \mathbb{E}(Z_i^*(t+s,b)\mathbb{1}_{\{B_i=b\}})^2 && \text{(D.190)} \\ &= \mathbb{E}(Z_i^*(t+s,b)^2)\mathbb{E}(\mathbb{1}_{\{B_i=b\}}) - \mathbb{E}(Z_i^*(t+s,b))^2\mathbb{E}(\mathbb{1}_{\{B_i=b\}})^2 && \text{(D.191)} \end{aligned}$$

Note that $\mathbb{E}(\mathbb{1}_{\{B_i=b\}})^2 = O(db^2)$ and hence this term has zero measure (as it is only integrated over one dimension). This leaves

$$\text{var}(Z_i^*(t+s,b)\mathbb{1}_{\{B_i=b\}}) = \mathbb{E}(Z_i^*(t+s,b)^2) f_{B_i}(b)db \tag{D.192}$$

where $f_{B_i}(b)$ is the marginal pdf of B_i.

The second term can also be expanded - note that, by the independence of the Z^* terms, for $i \neq j$

$$\text{cov}\bigg(Z_i^*(t+s,b)\mathbb{1}_{\{B_i=b\}}, Z_j^*(t+s,b)\mathbb{1}_{\{B_j=c\}} \bigg) = \mathbb{E}(Z_i^*(t+s,b))\mathbb{E}(Z_j^*(t+s,c))\text{cov}(\mathbb{1}_{\{B_i=b\}}, \mathbb{1}_{\{B_j=c\}}) \tag{D.193}$$

Moreover, if $i = j$, then one has $b \neq c$ and hence

$$\mathbb{1}_{\{B_i=b\}}\mathbb{1}_{\{B_j=c\}} = \mathbb{1}_{\{B_i=b, B_i=c\}} = 0 \tag{D.194}$$

which means

$$\text{cov}\left(Z_i^*(t+s,b)\mathbb{1}_{\{B_i=b\}}, Z_j^*(t+s,b)\mathbb{1}_{\{B_j=c\}}\right) = -\mathbb{E}(Z_i^*(t+s,b))\mathbb{E}(Z_j^*(t+s,c))\mathbb{E}(\mathbb{1}_{\{B_i=b\}})\mathbb{E}(\mathbb{1}_{\{B_j=c\}}) \tag{D.195}$$

$$= \mathbb{E}(Z_i^*(t+s,b))\mathbb{E}(Z_j^*(t+s,c))\text{cov}(\mathbb{1}_{\{B_i=b\}}, \mathbb{1}_{\{B_j=c\}}) \tag{D.196}$$

and hence the Supplementary Equation D.193 holds in all cases. Now, one has, for $i \neq j$

$$\text{cov}(\mathbb{1}_{\{B_i=b\}}, \mathbb{1}_{\{B_j=c\}}) = \mathbb{E}(\mathbb{1}_{\{B_i=b\}}\mathbb{1}_{\{B_j=c\}}) - \mathbb{E}(\mathbb{1}_{\{B_i=b\}})\mathbb{E}(\mathbb{1}_{\{B_j=c\}}) \tag{D.197}$$

$$= \mathbb{E}(\mathbb{1}_{\{B_i=b,B_j=c\}}) - f_{B_i}(b)f_{B_j}(c)dbdc \tag{D.198}$$

$$= (f_{B_i,B_j}(b,c) - f_{B_i}(b)f_{B_j}(c))dbdc \tag{D.199}$$

while if $i = j$ and $b \neq c$, this result also holds, following the convention that

$$f_{B_i,B_i}(b,c) = \delta(b-c)f_{B_i}(b) \tag{D.200}$$

(and hence in this case is zero) where δ is the Kronecker delta.

Thus, in all cases

$$\text{cov}\left(Z_i^*(t+s,b)\mathbb{1}_{\{B_i=b\}}, Z_j^*(t+s,b)\mathbb{1}_{\{B_j=c\}}\right) = \mathbb{E}(Z_i^*(t+s,b))\mathbb{E}(Z_j^*(t+s,c))(f_{B_i,B_j}(b,c) - f_{B_i}(b)f_{B_j}(c))dbdc \tag{D.201}$$

This gives an equation of

$$\begin{aligned}\text{var}(Z(t+s,l)) &= \int_{b=0}^{t}\sum_{i=0}^{n}\mathbb{E}(Z_i^*(t+s,b)^2)f_{B_i}(b)db... \\ ... &+ \int_{b=0}^{t}\int_{c=0}^{t}\sum_{i=0}^{n}\sum_{j=0}^{n}\mathbb{E}(Z_i^*(t+s,b))\mathbb{E}(Z_j^*(t+s,c))(f_{B_i,B_j}(b,c) - f_{B_i}(b)f_{B_j}(c))\mathbb{1}_{\{(b,i)\neq(c,j)\}}dbdc\end{aligned} \tag{D.202}$$

It is more informative to remove the $\mathbb{1}_{\{(b,i)\neq(c,j)\}}$ condition. This can be done by calculating

$$\int_{b=0}^{t}\int_{c=0}^{t}\sum_{i=0}^{n}\sum_{j=0}^{n}\mathbb{E}(Z_i^*(t+s,b))\mathbb{E}(Z_j^*(t+s,c))\Big(f_{B_i,B_j}(b,c)-f_{B_i}(b)f_{B_j}(c)\Big)\mathbb{1}_{\{(b,i)=(c,j)\}}dbdc \tag{D.203}$$

$$=\int_{b=0}^{t}\int_{c=0}^{t}\sum_{i=0}^{n}\mathbb{E}(Z_i^*(t+s,b)\mathbb{E}(Z_i^*(t+s,c))\Big(\delta(b-c)f_{B_i}(b)-f_{B_i}(b)f_{B_i}(c)\Big)\mathbb{1}_{\{b=c\}}dbdc \tag{D.204}$$

$$=\int_{b=0}^{t}\int_{c=0}^{t}\sum_{i=0}^{n}\mathbb{E}(Z_i^*(t+s,b)\mathbb{E}(Z_i^*(t+s,c))\Big(\delta(b-c)f_{B_i}(b)-f_{B_i}(b)f_{B_i}(c)\mathbb{1}_{\{b=c\}}\Big)dbdc \tag{D.205}$$

$$=\int_{b=0}^{t}\sum_{i=0}^{n}\mathbb{E}(Z_i^*(t+s,b))^2 f_{B_i}(b)db \tag{D.206}$$

noting that the second term is bounded and contains $\mathbb{1}_{\{b=c\}}$ which is non-zero only on a null set of the domain of integration (and hence the integral is zero). Thus, absorbing this correction term into the first term in Supplementary Equation D.202,

$$\begin{aligned}\mathrm{var}(Z(t+s,l)) &= \int_{b=0}^{t}\sum_{i=0}^{n}\mathrm{var}(Z_i^*(t+s,b))f_{B_i}(b)db... \\ ... &+ \int_{b=0}^{t}\int_{c=0}^{t}\sum_{i=0}^{n}\sum_{j=0}^{n}\mathbb{E}(Z_i^*(t+s,b))\mathbb{E}(Z_j^*(t+s,c))(f_{B_i,B_j}(b,c)-f_{B_i}(b)f_{B_j}(c))dbdc\end{aligned} \tag{D.207}$$

The advantage of this formulation is that it allows the contributions to the variance from the infection times B_i before time t and from further infections between times t and $t+s$ to be separated. Indeed, note that if the infection times are known (so that $f_{B_i}(b)=\delta(b-b_i)$), one has

$$\int_{b=0}^{t}\int_{c=0}^{t}\sum_{i=0}^{n}\sum_{j=0}^{n}\mathbb{E}(Z_i^*(t+s,b))\mathbb{E}(Z_j^*(t+s,c))(f_{B_i,B_j}(b,c)-f_{B_i}(b)f_{B_j}(c))dbdc$$

$$... = \int_{b=0}^{t}\int_{c=0}^{t}\sum_{i=0}^{n}\sum_{j=0}^{n}\mathbb{E}(Z_i^*(t+s,b))\mathbb{E}(Z_j^*(t+s,c))(\delta(b-b_i)\delta(c-b_j)-\delta(b-b_i)\delta(c-b_j))dbdc \tag{D.208}$$

$$=0 \tag{D.209}$$

noting that the definition of

$$f_{B_i,B_i}(b,c)=f_{B_i}(b)\delta(b-c)=f_{B_i,B_i}(b,c)=\delta(b-b_i)\delta(b-c)=\delta(b-b_i)\delta(c-b_i) \tag{D.210}$$

is consistent in this case. Thus, the second term in Supplementary Equation D.207 is only non-zero when there is uncertainty in the infection times (while, moreover, the first term is only non-zero when there is uncertainty in the infections that occur in the interval $(t,t+s)$, as otherwise $\mathrm{var}(Z_i^*(t+s,b_i))=0$).

To complete the derivation of the variance equation, it is necessary to derive formulae to calculate the quantities $\text{var}(Z_i^*)$. To enable this, define $M^*(t+s, b_i) := \mathbb{E}(Z_i^*(t+s, b_i))$ and $X^*(t+s, b_i) := \mathbb{E}(Z_i^*(t+s, b_i)^2)$ to be the mean and squared mean of the infection tree started from time t by the ith individual.

These quantities can be calculated directly from the mean and variance, M and V, of the "standard case" (where a single initial individual is infected at some time l), considered in previous notes in this appendix. This is possible as, in the context of renewal processes, the quantities Z_i^* are renewal processes where all but the first individuals are identical, and hence are amenable to similar methodology. Indeed, if one supposes that $\{Z(t+s, t+u)\}_{u \leq s}$ are a set of independent realisations of different "standard" epidemics, one has

$$Z_i^*(t+s, t+u) = \int_{u=0}^{s} Z(t+s, t+u) \mathbb{1}_{\{\text{individual } i \text{ infects another individual at time } t+u\}} \tag{D.211}$$

as the newly infected individuals start new, independent and "standard" epidemics. Define

$$\mathcal{I}_u := \mathbb{1}_{\{\text{individual } i \text{ infects another individual at time } t+u\}} \tag{D.212}$$

Hence,

$$\begin{aligned} M^*(t+s, b_i) &= E\left(\int_{u=0}^{s} Z(t+s, t+u) \mathcal{I}_u \right) && \text{(D.213)} \\ &= \int_{u=0}^{s} M(t+s, t+u) \rho(t+u) \nu(t - b_i + u) du && \text{(D.214)} \end{aligned}$$

Moreover,

$$\begin{aligned} X^*(t+s, b_i) &= E\left(\left[\int_{u=0}^{s} Z(t+s, t+u) \mathcal{I}_u \right]^2 \right) && \text{(D.215)} \\ &= E\left(\int_{u=0}^{s} \int_{k=0}^{s} Z(t+s, t+u) \mathcal{I}_u Z(t+s, t+k) \mathcal{I}_k \right) && \text{(D.216)} \end{aligned}$$

Note that, for $k \neq u$, the quantities $Z(t+s, t+u)$ and $Z(t+s, t+k)$ are independent. Moreover, these quantities are all independent from the indicator terms. Thus, it is helpful to split the integral,

giving

$$X^*(t+s,b_i) = \int_{u=0}^{s} E\Big(Z(t+s,t+u)^2\mathcal{I}_u\Big) + \int_{u=0}^{s}\int_{k=0}^{s} E\Big[Z(t+s,t+u)\mathcal{I}_u Z(t+s,t+k)\mathcal{I}_k\Big]\mathbb{1}_{\{u\neq k\}} \tag{D.217}$$

$$= \int_{u=0}^{s} E\Big(Z(t+s,t+u)^2\mathcal{I}_u\Big) + \int_{u=0}^{s}\int_{k=0}^{s} M(t+s,t+u)\mathbb{E}(\mathcal{I}_u)M(t+s,t+k)\mathbb{E}(\mathcal{I}_k)\mathbb{1}_{\{u\neq k\}} \tag{D.218}$$

Now,

$$\int_{u=0}^{s}\int_{k=0}^{s} M(t+s,t+u)\mathbb{E}(\mathcal{I}_u)M(t+s,t+k)\mathbb{E}(\mathcal{I}_k) = \Big[\int_{u=0}^{s} M(t+s,t+u)\mathbb{E}(\mathcal{I}_u)\Big]^2 = M^*(t+s,b_i)^2 \tag{D.219}$$

while

$$\int_{u=0}^{s}\int_{k=0}^{s} M(t+s,t+u)\mathbb{E}(\mathcal{I}_u)M(t+s,t+k)\mathbb{E}(\mathcal{I}_k)\mathbb{1}_{\{u=k\}} = 0 \tag{D.220}$$

as the integrand is bounded and is non-zero only on a null set of the domain of integration. Hence, one has

$$\int_{u=0}^{s}\int_{k=0}^{s} M(t+s,t+u)\mathbb{E}(\mathcal{I}_u)M(t+s,t+k)\mathbb{E}(\mathcal{I}_k)\mathbb{1}_{\{u\neq k} = M^*(t+s,b_i)^2 \tag{D.221}$$

Thus,

$$X^*(t+s,b_i) = \int_{u=0}^{s} E\Big(Z(t+s,t+u)^2\mathcal{I}_u\Big) + M^*(t+s,b_i)^2 \tag{D.222}$$

$$= \int_{u=0}^{s} E\Big(Z(t+s,t+u)^2\Big)\mathbb{E}(\mathcal{I}_u) + M^*(t+s,b_i)^2 \tag{D.223}$$

$$= \int_{u=0}^{s} (V(t+s,t+u) + M(t+s,t+u)^2)\rho(t+u)\nu(t-b_i+u)du + M^*(t+s,b_i)^2 \tag{D.224}$$

Hence, defining $V^*(t+s,b_i) := \text{var}(Z^*(t+s,b_i)) = X^*(t+s,b_i) - M^*(t+s,b_i)^2$, one has

$$V^*(t+s,b_i) = \int_{u=0}^{s} (V(t+s,t+u) + M(t+s,t+u)^2)\rho(t+u)\nu(t-b_i+u)du \tag{D.225}$$

Hence, one has the final form of the variance equation

$$\text{var}(Z(t+s,l)) = \int_{b=0}^{t} \sum_{i=0}^{n} V^*(t+s,b) f_{B_i}(b) db ...$$
$$... + \int_{b=0}^{t} \int_{c=0}^{t} \sum_{i=0}^{n} \sum_{j=0}^{n} M^*(t+s,b) M^*(t+s,c) (f_{B_i,B_j}(b,c) - f_{B_i}(b) f_{B_j}(c)) dbdc \tag{D.226}$$

D.7.2 Bounding the equation

Unlike previous formulae, this is an explicit equation and no recursion is required to get the desired results (although recursion is necessary to calculate the V term in V^*). However, the infection time pdf makes this a difficult equation to evaluate.

However, one can give a simpler upper bound on the variance. Define

$$\nu_{\text{bound}}(u) := \max_{b_i \in [l,t]} (\nu(t - b_i + u)) \tag{D.227}$$

so that

$$M^*(t+s,b_i) \leq \int_0^s M(t+s,t+u)\rho(t+u)\nu_{\text{bound}}(t-b_i+u)du := \mathcal{M}^*(t+s) \tag{D.228}$$

and

$$V^*(t+s,b_i) \leq \int_{u=0}^{s} (V(t+s,t+u) + M(t+s,t+u)^2)\rho(t+u)\nu_{\text{bound}}(u)du := \mathcal{V}^*(t+s) \tag{D.229}$$

so that this is now independent of b_i. Note that the construction of $\nu_{\text{bound}}(u)$ means that it will still decay for large u. Under the assumption that the infection times are roughly deterministic, the second term is zero and so

$$\text{var } (Z(t+s,l)) \leq Z(t,l)\mathcal{V}^*(t+s) \tag{D.230}$$

The covariance term can be added in by noting that

$$\int_{b=0}^{t}\int_{c=0}^{t}\sum_{i=0}^{n}\sum_{j=0}^{n}M^*(t+s,b)M^*(t+s,c)(f_{B_i,B_j}(b,c)-f_{B_i}(b)f_{B_j}(c))dbdc...$$

$$...\leq\int_{b=0}^{t}\int_{c=0}^{t}\sum_{i=0}^{n}\sum_{j=0}^{n}\mathcal{M}^*(t+s)^2f_{B_i,B_j}(b,c)dbdc \tag{D.231}$$

$$\leq\sum_{i=0}^{n}\sum_{j=0}^{n}\int_{b=0}^{t}\int_{c=0}^{t}\mathcal{M}^*(t+s)^2(f_{B_i,B_j}(b,c)+f_{B_i}(b)f_{B_j}(c))dbdc \tag{D.232}$$

$$\leq Z(t,l)^2\mathcal{M}^*(t+s)^2 \tag{D.233}$$

which gives an overall bound of

$$\text{var }(Z(t+s,l))\leq Z(t,l)\mathcal{V}^*(t+s)+Z(t,l)^2\mathcal{M}^*(t+s)^2 \tag{D.234}$$

D.7.3 Special cases

To finish, it is helpful to consider a couple of special cases which may arise when the epidemic is large. If the infection times are mostly independent, then

$$i\neq j\Rightarrow f_{B_i,B_j}(b,c)\sim f_{B_i}(b)f_{B_j}(c) \tag{D.235}$$

while for $i=j$, note that

$$\int_{b=0}^{t}\int_{c=0}^{t}\sum_{i=0}^{n}M^*(t+s,b)M^*(t+s,c)(f_{B_i,B_i}(b,c)-f_{B_i}(b)f_{B_i}(c))dbdc...$$

$$...=\int_{b=0}^{t}\int_{c=0}^{t}\sum_{i=0}^{n}M^*(t+s,b)M^*(t+s,c)(\delta(b-c)f_{B_i}(b)-f_{B_i}(b)f_{B_i}(c))dbdc \tag{D.236}$$

$$=\int_{b=0}^{t}\sum_{i=0}^{n}M^*(t+s,b)^2f_{B_i}(b)db-\sum_{i=0}^{n}\left[\int_b M^*(t+s,b)f_{B_i}(b)db\right]^2 \tag{D.237}$$

and hence

$$\text{var}(Z(t+s,l))\sim\int_{b=0}^{t}\sum_{i=0}^{n}V^*(t+s,b)f_{B_i}(b)db+\int_b\sum_{i=0}^{n}M^*(t+s,b)^2f_{B_i}(b)db-\sum_{i=0}^{n}\left[\int_b M^*(t+s,b)f_{B_i}(b)db\right]^2 \tag{D.238}$$

This is still a complicated equation to compute, although the advantage is that one only needs one-dimensional marginal distributions of the infection times, and hence it is significantly more tractable.

Moreover, the upper bound on the variance can be improved to

$$\text{var}(Z(t+s,l)) \leq Z(t,l)\mathcal{V}(t,l) + Z(t,l)\mathcal{M}(t,l)^2 \tag{D.239}$$

so that it is proportional to $Z(t,l)$, rather than $Z(t,l)^2$.

The simplest case is when the infection times are known - something which may be approximately true if the epidemic is large (and hence has been approximately deterministic in the recent past). In this case, the equation simply reduces to

$$\text{var}(Z(t+s,l)) \sim \sum_{i=0}^{n} V^*(t+s,b_i) \tag{D.240}$$

where b_i are the infection times. In this case, the variance can be simply calculated from the quantities M and V.

D.8 Discrete epidemics

D.8.1 Discrete pgf

Suppose now that the branching process is entirely discrete (and, for convenience, occurs on integer times). For the lifetime, L, of an individual infected at l, define

$$g(u,l) := \mathbb{P}(L = u) \quad \text{and} \quad \overline{G}(u,l) := \mathbb{P}(L \geq u) \tag{D.241}$$

In this discrete setting, it is important to specify exactly inequalities whose strictness is unimportant in the continuous case. In particular, if an individual is infected at time a and has a lifetime of b, it will be considered to be infectious at time $a+b$, and will be counted when calculating prevalence at this time. That is, it can infect others at time $a+b$ (and these individuals will be given infection time $a+b$) but will not be able to infect individuals at time $a+b+1$.

For the counting process of infections, one can in this case work without a separate infection event process and instead simply use the quantities

$$q_u(t,l) := \mathbb{P}\Big(N(t,l) - N(t-1,l) = u\Big) \quad \text{and} \quad \mathcal{Q}_{(t,l)}(s) := E\Big(s^{Q(t,l)}\Big) \tag{D.242}$$

where $Q(t,l)$ has pmf given by $q_u(t,l)$. Hence, each $Q(t,l)$ (which may be zero, unlike Y in the continuous case) gives the number of new infections at time t caused by an individual that was

infected at time l. Now, note that for $u < t - l$, one has

$$E\left(s^{Z(t,l)}\middle|L = u\right) = E\left(s^{\sum_{k=1}^{u}\sum_{i=1}^{Q(l+k,l)} Z_{ik}(l+u,l)}\right) \tag{D.243}$$

where the variables Z_{ik} are iid copies of Z. Note that the variables $Q(l+k,l)$ are independent as $N(t,l)$ has independent increments, meaning that

$$E\left(s^{Z(t,l)}\middle|L = u\right) = \prod_{k=1}^{u} E\left(s^{\sum_{i=1}^{q(l+k,l)} Z_{ik}(l+u,l+k)}\right) \tag{D.244}$$

$$= \prod_{k=1}^{u} \mathcal{Q}_{(l+k,l)}\Big(F(l+u,l+k)\Big) \tag{D.245}$$

Thus, the generating function equation for prevalence can be written as

$$F(t,l) = s\overline{G}(t-l,l)\prod_{k=1}^{t-l}\mathcal{Q}_{(l+k,l)}\Big(F(t,l+k)\Big) + \sum_{u=0}^{t} g_u \prod_{k=1}^{u}\mathcal{Q}_{(l+k,l)}\Big(F(t,l+k)\Big) \tag{D.246}$$

where

$$\overline{G}(t-l,l) = \mathbb{P}(L \geq t-l) \tag{D.247}$$

The form of the generating function for the discrete case is simpler than the continuous one and might be more amenable to computation.

D.8.2 Recovery of the continuous case

Suppose that each step corresponds to a time interval of $dt << 1$. Suppose further that

$$\hat{g}(udt, ldt)dt \sim g_{u,l}, \quad \hat{t} \sim tdt, \quad \text{and} \quad \hat{l} \sim ldt \tag{D.248}$$

where the quantities with a hat are constant. To ensure continuity in probability, it will be assumed that

$$\hat{q}_u(\hat{t},\hat{l})dt \sim q_u(t,l) \quad \forall u \geq 1 \quad \text{and} \quad q_0(t,l) \sim 1 - \sum_{u=1}^{\infty} dt\hat{q}_u(\hat{t},\hat{l}) \tag{D.249}$$

where again, $\hat{q}$ is independent of dt. Now, one has

$$G(t-l,l) = \sum_{u=0}^{t-l} g_{u,l} \sim \sum_{u=0}^{\frac{\hat{t}-\hat{l}}{dt}} \hat{g}_{u,l}(udt)dt \sim \int_0^{\hat{t}-\hat{l}} \hat{g}(u,\hat{l})du := \hat{G}(\hat{t}-\hat{l},l) \tag{D.250}$$

Moreover, one has

$$\mathcal{Q}_{(t,l)}(s) \sim \left(1 - \sum_{u=1}^{\infty} \hat{q}_u(\hat{t},\hat{l})dt\right) + \sum_{u=1}^{\infty} s^u \hat{q}_u(\hat{t},\hat{l})dt = 1 + \sum_{u=1}^{\infty}(s^u - 1)\hat{q}_u(\hat{t},\hat{l})dt \tag{D.251}$$

Using this relation, setting $\hat{k} := kdt$ and Taylor expanding gives

$$\log\left(\prod_{k=1}^{t-l} \mathcal{Q}_{(l+k,l)}(s)\right) \sim \sum_{k=1}^{t-l} \log\left(1 + \sum_{u=1}^{\infty}(s^u - 1)\hat{q}_u(\hat{l}+\hat{k},\hat{l})dt\right) \tag{D.252}$$

$$\sim \sum_{k=1}^{t-l}\sum_{u=1}^{\infty}(s^u - 1)\hat{q}_u(\hat{l}+\hat{k},\hat{l})dt \tag{D.253}$$

$$\sim \int_0^{\hat{t}-\hat{l}} \sum_{u=1}^{\infty}(s^u - 1)\hat{q}_u(\hat{l}+\hat{k},\hat{l})d\hat{k} \tag{D.254}$$

Hence,

$$F(t,l) \sim (1-\hat{G}(\hat{t}-\hat{l}))\exp\left[\int_0^{\hat{t}-\hat{l}} \sum_{u=1}^{\infty}(s^u-1)\hat{q}_u(\hat{l}+\hat{k},\hat{l})d\hat{k}\right] + \int_0^{t-l} \exp\left[\int_0^{\hat{u}} \sum_{w=1}^{\infty}(s^w-1)\hat{q}_w(\hat{l}+\hat{k},\hat{l})d\hat{k}\right]\hat{g}(\hat{u},\hat{l})d\hat{u} \tag{D.255}$$

It is now possible to define the limiting continuous process. Consider a counting process $N(\hat{t},\hat{l})$ in continuous time with independent increments where infection events occur according to a rate function given by

$$r(\hat{t},\hat{l}) = \sum_{u=1}^{\infty} \hat{q}_u(\hat{t},\hat{l}) \tag{D.256}$$

and where, given that an infection event occurs at t from a particle born at l, the infection event is of size $k \geq 0$ with probability

$$\frac{\hat{q}_k(\hat{t},\hat{l})}{\sum_{u=1}^{\infty} \hat{q}_u(\hat{t},\hat{l})}. \tag{D.257}$$

Suppose that $J(\hat{t},\hat{l})$ counts the infection events of this process (and hence is an inhomogeneous Poisson Process of rate $r(\hat{t},\hat{l})$) and that $\mathcal{Y}_{(\hat{t},\hat{l})}$ is the generating function of infection event size given that a infection event occurs at $(\hat{t},\hat{l})$. Note that

$$\int_0^{\hat{t}-\hat{l}} \sum_{u=1}^{\infty} \hat{q}_u(\hat{l}+k,\hat{l})dk = \int_0^{\hat{t}-\hat{l}} r(\hat{l}+k,\hat{l})dk = \mathbb{E}(J(\hat{t},\hat{l})) \tag{D.258}$$

and that

$$\sum_{u=1}^{\infty} s^u \hat{q}_u(\hat{l}+\hat{k},\hat{l}) = \sum_{u=1}^{\infty}\left(\frac{s^u \hat{y}_u(\hat{l}+\hat{k},\hat{l})}{\sum_{m=1}^{\infty}\hat{y}_m(\hat{l}+\hat{k},\hat{l})}\right)\sum_{m=1}^{\infty}\hat{q}_m(\hat{l}+\hat{k},\hat{l}) \tag{D.259}$$

$$= \mathcal{Y}_{(\hat{l}+\hat{k},\hat{l})}(s)\left(\sum_{u=1}^{\infty}\hat{q}_u(\hat{l}+\hat{k},\hat{l})\right) \tag{D.260}$$

$$= \mathcal{Y}_{(\hat{l}+\hat{k},\hat{l})}(s) r(\hat{l}+\hat{k},\hat{l}) \tag{D.261}$$

Hence,

$$\prod_{k=1}^{t-l}\mathcal{Q}_{(l+k,l)}(s) \sim \exp\left[\int_0^{\hat{t}-\hat{l}} r(\hat{l}+k,\hat{l})\mathcal{Y}_{(\hat{l}+k,\hat{l})}(s)dk - \mathbb{E}(J(\hat{t},\hat{l}))\right] \tag{D.262}$$

and so, applying this to Supplementary Equation D.255 shows that the continuous generating function equation is recovered. Note that here, the distribution of Y has been allowed to depend on l (and this is the generating function equation that arises in this case), but the an equation with an l-independent Y will arise if the ratio of each $q_k(t,l)$ and $\sum_{k=1}^{\infty} q_k(t,l)$ are independent of l.

D.8.3 Distinctness from the continuous case

It is important to note that the relaxation of the assumption that N is continuous in probability necessary in considering the discrete case means that the pgf becomes materially different.

Indeed, one can characterise the discrete case through the continuous framework by imposing that

$$r(t,l) = \left(\sum_{u=1}^{\infty} q_u(t,l)\right)\left(\sum_{n=1}^{\infty}\delta(l+n-t)\right) \tag{D.263}$$

as this is gives probability of N increasing (by whatever number) in the discrete case discussed above. Moreover, again allowing Y to depend on l, $Y(t,l)$ has distribution

$$\mathbb{P}(Y(t,l)=k) = \frac{q_k(t,l)}{\sum_{m=1}^{\infty} q_m(t,l)} \tag{D.264}$$

Now, note that

$$\lambda(t,l) = \int_l^t r(s,l)ds = \sum_{n=1}^{\lfloor t-l\rfloor}\sum_{u=1}^{\infty} q_u(l+n,l) \tag{D.265}$$

where $\lfloor m\rfloor$ denotes the largest integer that is smaller than m. Moreover,

$$\int_0^{t-l}\mathcal{Y}_{(l+k,l)}(F(t,l+k))r(l+k,l) = \sum_{n=1}^{\lfloor t-l\rfloor}\sum_{u=1}^{\infty} q_u(l+n,l)\mathcal{Y}_{(l+n,l)}(F(t,l+n)) \tag{D.266}$$

We suppose for a contradiction that the pgf in the continuous case is also valid in this discrete setting.

Hence (taking $\kappa = 1$),

$$F(t,l) = s(1-G(t-l,l))\exp\Bigg[\sum_{n=1}^{\lfloor t-l\rfloor}\sum_{u=1}^{\infty} q_u(l+n,l)\mathcal{Y}_{(l+n,l)}(F(t,l+n)) - \sum_{n=1}^{\lfloor t-l\rfloor}\sum_{u=1}^{\infty} q_u(l+n,l)\Bigg]... \\ ... + \int_0^{t-l}\exp\Bigg[\sum_{n=1}^{\lfloor t-l+u\rfloor}\sum_{m=1}^{\infty} q_m(l+n,l)\mathcal{Y}_{(l+n,l)}(F(t,l+n)) - \sum_{n=1}^{\lfloor t-l+u\rfloor}\sum_{m=1}^{\infty} q_m(l+n,l)\Bigg]g(u,l)du \tag{D.267}$$

Now, note that

$$\mathcal{Y}_{(l+n,l)}(s) = \sum_{m=1}^{\infty}\frac{s^m q_m(l+n,l)}{\sum_{k=1}^{\infty} q_k(l+n,l)} \tag{D.268}$$

$$= \frac{1}{\sum_{k=1}^{\infty} q_k(l+n,l)}\Bigg(\sum_{m=0}^{\infty} s^m q_m(l+n,l) - q_0(l+n,l)\Bigg) \tag{D.269}$$

$$= \frac{1}{\sum_{k=1}^{\infty} q_k(l+n,l)}\Bigg(\mathcal{Q}_{(l+n,l)}(s) - (1-\sum_{k=1}^{\infty} q_k(l+n,l))\Bigg) \tag{D.270}$$

and hence

$$\sum_{n=1}^{\lfloor t-l+u\rfloor}\sum_{m=1}^{\infty} q_m(l+n,l)\mathcal{Y}_{(l+n,l)}(F(t,l+n)) - \sum_{n=1}^{\lfloor t-l+u\rfloor}\sum_{m=1}^{\infty} q_m(l+n,l) \tag{D.271}$$

$$= \sum_{n=1}^{\lfloor t-l+u\rfloor}\Bigg(\mathcal{Q}_{(l+n,l)}(F(t,l+n)) + \sum_{k=1}^{\infty} q_k(l+n,l)\Bigg) \tag{D.272}$$

which means

$$F(t,l) = s(1-G(t-l,l))\exp\Bigg[\sum_{n=1}^{\lfloor t-l\rfloor}\Bigg(\mathcal{Q}_{(l+n,l)}(F(t,l+n)) + \sum_{k=1}^{\infty} q_k(l+n,l)\Bigg)\Bigg]... \tag{D.273}$$

$$+ \int_0^{t-l}\exp\Bigg[\sum_{n=1}^{\lfloor t-l+u\rfloor}\Bigg(\mathcal{Q}_{(l+n,l)}(F(t,l+n)) + \sum_{k=1}^{\infty} q_k(l+n,l)\Bigg)\Bigg]g(u,l)du \tag{D.274}$$

Finally, defining $\mathcal{Q}^*(s) := e^{\mathcal{Q}(s)}$ and turning the integral over g into a discrete sum, we have

$$F(t,l) = s(1-G(t-l,l))\prod_{n=1}^{\lfloor t-l\rfloor}\mathcal{Q}^*(F(t,l+n))e^{\sum_{k=1}^{\infty} q_k(l+n,l)} + \sum_{u=1}^{\lfloor t-l\rfloor}\prod_{n=1}^{\lfloor t-l+u\rfloor}\mathcal{Q}^*(F(t,l+n))e^{\sum_{k=1}^{\infty} q_k(l+n,l)}g(u,l) \tag{D.275}$$

This matches very closely with the pgf in the discrete case, but has some extra terms as expected for the contradiction - firstly, the $\mathcal{Q}^*$ in place of the $\mathcal{Q}$, and also the extra $e^{\sum_{k=1}^{\infty} q_k}$ terms. When taking the small dt limit as in the previous subsection, these anomalies disappear, as

$$e^{\mathcal{Q}(s)} \sim e^{1+\alpha dt} \sim 1 + \alpha dt \sim \mathcal{Q}(s) \tag{D.276}$$

and

$$e^{\sum_{k=1}^{\infty} q_k(l+n,l)} \sim e^{\beta dt} \sim 1 \tag{D.277}$$

for some α and β. Thus, these dissimilarities only appear in the $O(dt^2)$ level (and hence disappear in the small dt limit). However, they will be non-trivial if dt is not small, underlining the importance of the assumption that N is continuous in probability - neglecting such an assumption could lead to materially wrong results in the case of a large step-size.

D.8.4 Discrete likelihood

If the epidemic happens in discrete time, it is significantly easier to calculate the likelihood. Define $A_{k,i}$ to be the number of infections caused at time k by a (still infectious) individual that was infected at time i. Then, the number of infections which occur at time k is given by

$$\mathcal{A}_k(\boldsymbol{y}, \boldsymbol{d}) = \sum_{i=0}^{k-1} \sum_{j=1}^{y_i} A_{k,i}^j \mathcal{I}_{\{x_{ij} \leq k\}} \tag{D.278}$$

where each $A_{k,i}^j$ is an independent copy of $A_{k,i}$ and, similarly to before, x_{ij} is the time at which the jth individual infected at time i stops being infectious. Note that here, as previously in the discrete setting but in contrast to the continuous case, y_i can be zero.

Then, the likelihood is simply given by

$$L(\boldsymbol{y}, \boldsymbol{D}) = \left(\prod_{k=1}^{n} \mathbb{P}(\mathcal{A}_k(\boldsymbol{y}, \boldsymbol{d}) = y_k) \right) \left(\prod_{i=1}^{n} \prod_{j=1}^{y_i} g(x_{ij} - i, i) \right) \tag{D.279}$$

where, as we are in the discrete case, g is now a pmf. This gives a log-likelihood of

$$\ell(\boldsymbol{y}, \boldsymbol{D}) = \sum_{k=1}^{n} \log \left(\mathbb{P}(\mathcal{A}_k(\boldsymbol{y}, \boldsymbol{d}) = y_k) \right) + \sum_{i=1}^{n} \sum_{j=1}^{y_i} \log(g(x_{ij} - i, i)) \tag{D.280}$$

Appendix - Paper V

E.1 Proofs of Theorems 7.1, 7.2 and 7.3

Before beginning the main proof, it is helpful to note some fundamental results about the SIR equations that will be used throughout. Namely, for each $i \in \{1, ..., n\}$ and $t \geq 0$

$$0 \leq S_i(t), I_i(t), R_i(t), S_i^V(t), I_i^V(t), R_i^V(t) \leq N_i \tag{E.1}$$

and

$$\lim_{t\to\infty}(I_i(t)) = \lim_{t\to\infty}(I_i^V(t)) = 0. \tag{E.2}$$

These results are proved in Lemmas E.17 and E.18.

It is first useful to define

$$K_{ij}(t) = \frac{\beta_{ij}^1}{\mu_j^1} R_j(t) + \frac{\beta_{ij}^2}{\mu_j^2} R_j^V(t) \tag{E.3}$$

and

$$L_{ij}(t) := \frac{\beta_{ij}^3}{\mu_i^1} R_j(t) + \frac{\beta_{ij}^4}{\mu_i^2} R_j^V(t). \tag{E.4}$$

Then, the following propositions hold.

E.1.1 An inequality for K_{ij} and L_{ij}

Note that the proof of this proposition requires a significant amount of algebra, and the majority of it has hence been left to lemmas which can be found later in this appendix. However, the key logic of the proof will be presented here.

Also, note that in this paper, a step function is defined to be a function that is piecewise constant on any *bounded* interval of $\Re$. Thus, it may have infinitely many discontinuities, but only finitely many in any bounded interval. This differs from the definition used in some other papers (which impose that a step function is piecewise constant on $\Re$).

Proposition E.1 *Suppose that $U_i(t)$ and $\tilde{U}_i(t)$ are right-continuous step functions. Moreover, suppose that*

$$\beta_{ij}^1 > \beta_{ij}^3 > 0 \quad \forall i, j \in \{1, ..., n\}, \tag{E.5}$$

$$S_i(0)I_i(0) > 0 \quad \forall i \in \{1, ...n\}. \tag{E.6}$$

and that

$$W_i(t) < N_i \quad \forall t \geq 0 \quad \textit{and} \quad \forall i \in \{1, ..., n\} \tag{E.7}$$

Then,

$$K_{ij}(t) \geq \tilde{K}_{ij}(t) \quad \textit{and} \quad L_{ij}(t) \geq \tilde{L}_{ij}(t) \quad \forall t \geq 0. \tag{E.8}$$

Proof: Suppose that the proposition does not hold. Hence, one can define

$$T := \inf \left\{ t : K_{ij}(t) < \tilde{K}_{ij}(t) \quad \text{or} \quad L_{ij}(t) < \tilde{L}_{ij}(t) \quad \text{for some } i, j \in \{1, ..., n\} \right\}. \tag{E.9}$$

Then, there exists some $b \in \{1, .., n\}$ and some real constants κ and η such that the following system of inequalities holds at time T:

$$S_b(T) + S_b^V(T) \leq \tilde{S}_b(T) + \tilde{S}_b^V(T), \tag{E.10}$$

$$I_b(T) + R_b(T) \geq \tilde{I}_b(T) + \tilde{R}_b(T) \tag{E.11}$$

$$R_b(T) \geq \tilde{R}_b(T), \tag{E.12}$$

$$R_b(T) + \kappa R_b^V(T) \leq \tilde{R}_b(T) + \kappa \tilde{R}_b^V(T), \tag{E.13}$$

$$I_b(T) + \eta I_b^V(T) \leq \tilde{I}_b(T) + \eta \tilde{I}_b^V(T), \tag{E.14}$$

$$0 \leq \kappa \leq \eta \leq 1. \tag{E.15}$$

The derivations of inequalities (E.10) - (E.15) are found in Lemmas E.5 - E.8. Moreover,

$$\begin{aligned} &S_b(T) + I_b(T) + R_b(T) + S_b^V(T) + I_b^V(T) + R_b^V(T) = \\ &\tilde{S}_b(T) + \tilde{I}_b(T) + \tilde{R}_b(T) + \tilde{S}_b^V(T) + \tilde{I}_b^V(T) + \tilde{R}_b^V(T), \end{aligned} \tag{E.16}$$

which comes from (7.20). Note that (E.13) in fact holds to equality in this case, but this is not necessary for the proof (and later, the same system will be considered where such an equality is not guaranteed).

By Lemma E.9, the system (E.10) - (E.16) implies that

$$I_b(T) + R_b(T) = \tilde{I}_b(T) + \tilde{R}_b(T), \tag{E.17}$$

$$I_b^V(T) + R_b^V(T) = \tilde{I}_b^V(T) + \tilde{R}_b^V(T), \tag{E.18}$$

$$S_b(T) + S_b^V(T) = \tilde{S}_b(T) + \tilde{S}_b^V(T), \tag{E.19}$$

If $T > 0$, then Lemma E.10 can be used to show that

$$W_k(t) = \tilde{W}_k(t) \quad \forall t \in [0, T] \quad \text{and} \quad \forall k \in \{1, ..., n\} \tag{E.20}$$

while if $T = 0$ then this is immediate. Thus, the two ODE systems are the same up to time T, which means that all variables (in all groups) are equal at time T.

From this point, the proof of Proposition E.1 can be completed by considering the behaviour of the system at time $T + \delta$ for small δ. For sufficiently small δ, $U_i(t)$ and $\tilde{U}_i(t)$ are constant on $[T, T + \delta]$ (as they are step functions) and this condition on δ will be assumed for the remainder of this proof

Define functions Δ_i^f to be

$$\Delta_i^f(t) := f_i(T + t) - \tilde{f}_i(T + t) \quad \text{for} \quad f \in \{S, I, R, S^V, I^V, R^V, W\} \tag{E.21}$$

and note that

$$\Delta_i^f(0) = 0 \quad \forall f \in \{S, I, R, S^V, I^V, R^V, W\}. \tag{E.22}$$

Then, by Lemma E.11, for $t \in [0, \delta]$ and any real numbers x and y

$$\frac{x}{\mu_i^1}\Delta_i^R + \frac{y}{\mu_i^2}\Delta_i^{R^V} = \frac{t^3 S_i(T)(U_i(T) - \tilde{U}_i(T))}{6(N_i - W_i(T))} \left[x \sum_{j=1}^{n} (K'_{ij}(T)) - y \sum_{j=1}^{n} (L'_{ij}(T)) \right] + O(\delta^4). \tag{E.23}$$

Hence, by Lemma E.12,

$$\sum_{j=1}^{n} K_{ij}(t) \geq \sum_{j=1}^{n} \tilde{K}_{ij}(t) \quad \forall t \in [0, T + \delta] \tag{E.24}$$

and

$$\sum_{j=1}^{n} L_{ij}(t) \geq \sum_{j=1}^{n} \tilde{L}_{ij}(t) \quad \forall t \in [0, T + \delta] \tag{E.25}$$

for sufficiently small δ.

Now, by the definition of T, there exists some t in $[T, T + \delta]$ such that, for some a, b

$$K_{ab}(t) < \tilde{K}_{ab}(t) \quad \text{or} \quad L_{ab}(t) < \tilde{L}_{ab}(t). \tag{E.26}$$

Indeed, from Lemma E.13, there exists some $t \in (T, T + \delta)$ such that

$$R_b(t) + \kappa R_b^V(t) < \tilde{R}_b(t) + \kappa \tilde{R}_b^V(t) \quad \text{and} \quad I_b(t) + \eta I_b^V(t) \leq \tilde{I}_b(t) + \eta \tilde{I}_b^V(t) \tag{E.27}$$

for some

$$0 \leq \kappa \leq \eta \leq 1. \tag{E.28}$$

Now, by Lemmas E.5 - E.7 (which only require the properties (E.24) and (E.25)), the system of inequalities (E.10)-(E.12) holds for group b at time t. These can be combined with (E.27), (E.28) and (E.16) to use Lemma E.9, showing

$$\eta I_b^V(t) + \kappa R_b^V(t) = \eta \tilde{I}_b^V(t) + \kappa \tilde{R}_b^V(t) \tag{E.29}$$

$$I_b(t) + R_b(t) = \tilde{I}_b(t) + \tilde{R}_b(t). \tag{E.30}$$

By adding the inequalities in (E.27) together,

$$R_b(t) + \kappa R_b^V(t) + I_b(t) + \eta I_b^V(t) < \tilde{R}_b(t) + \kappa \tilde{R}_b^V(t) + \tilde{I}_b(t) + \eta \tilde{I}_b^V(t). \tag{E.31}$$

Then, (E.29) and (E.30) show that this must in fact be an equality which is a contradiction. Thus, t cannot exist. This provides a contradiction to the definition of T, and hence finishes the proof of Proposition E.1.

It is now possible to prove Theorem 7.1 under the extra restrictions given in Proposition E.1.

E.1.2 A proof for a restricted parameter and policy set

Proposition E.2 *Under the conditions of Proposition E.1, for any $t \geq 0$ and $i \in \{1, ..., n\}$*

$$I_i(t) + I_i^V(t) + R_i(t) + R_i^V(t) \geq \tilde{I}_i(t) + \tilde{I}_i^V(t) + \tilde{R}_i(t) + \tilde{R}_i^V(t) \tag{E.32}$$

and

$$R_i(t) \geq \tilde{R}_i(t). \tag{E.33}$$

Moreover, for any $\lambda \in [0, 1]$

$$R_i(\infty) + \lambda R_i^V(\infty) \geq \tilde{R}_i(\infty) + \lambda \tilde{R}_i^V(\infty) \tag{E.34}$$

and hence, the objective function is lower for $\tilde{U}$, provided the conditions of Proposition E.1 are met.

Proof: Note that, by Proposition E.1,

$$K_{ij}(t) \geq \tilde{K}_{ij}(t) \quad \text{and} \quad L_{ij}(t) \geq \tilde{L}_{ij}(t) \quad \forall t \geq 0 \tag{E.35}$$

and hence, by Lemma E.5, for each $i \in \{1, ..., n\}$

$$S_i(t) + S_i^V(t) \leq \tilde{S}_i(t) + \tilde{S}_i^V(t). \tag{E.36}$$

Combining this with the conservation of population equation, (E.16), shows that

$$I_i(t) + I_i^V(t) + R_i(t) + R_i^V(t) \geq \tilde{I}_i(t) + \tilde{I}_i^V(t) + \tilde{R}_i(t) + \tilde{R}_i^V(t) \tag{E.37}$$

as required. Now, taking $t \to \infty$ and noting that the infections tend to zero by Lemma E.18 gives

$$R_i(\infty) + R_i^V(\infty) \geq \tilde{R}_i(\infty) + \tilde{R}_i^V(\infty). \tag{E.38}$$

Moreover, by Lemma E.7, for any $t \geq 0$ and $i \in \{1, ..., n\}$

$$R_i(t) \geq \tilde{R}_i(t) \tag{E.39}$$

as required. Also, taking $t \to \infty$ shows that

$$R_i(\infty) \geq \tilde{R}_i(\infty). \tag{E.40}$$

Thus, for any $\lambda \in [0, 1]$

$$\begin{aligned} R_i(\infty) + \lambda R_i^V(\infty) &= (1-\lambda)R_i(\infty) + \lambda(R_i(\infty) + R_i^V(\infty)) && \text{(E.41)} \\ &\geq (1-\lambda)\tilde{R}_i(\infty) + \lambda(\tilde{R}_i(\infty) + \tilde{R}_i^V(\infty)) && \text{(E.42)} \\ &= \tilde{R}_i(\infty) + \lambda\tilde{R}_i^V(\infty) && \text{(E.43)} \end{aligned}$$

as required.

By summing the i inequalities at $t = \infty$ from Proposition E.2 (and using $\lambda = \kappa_i$), Theorem 7.1 holds under the additional conditions given in Proposition E.1. Note that the closure of the set of parameters, initial conditions and vaccination policies which satisfy these conditions is the original set specified in Theorem 7.1. Thus, one can generalise the result with the help of the following proposition.

E.1.3 Continuous dependence

Proposition E.3 *Define the set of functions*

$$\mathcal{F} := \left\{ S_i(t;\epsilon), I_i(t;\epsilon), R_i(t;\epsilon), S_i^V(t;\epsilon), I_i^V(t;\epsilon), R_i^V(t;\epsilon) : i \in \{1,...,n\}, \quad \epsilon, t \geq 0 \right\}, \tag{E.44}$$

where for each fixed ϵ, these functions solve the model equations with parameters

$$\mathcal{P} = \left\{ \beta_{ij}^{\alpha}(\epsilon), \mu_i^{\gamma}(\epsilon) : i,j \in \{1,...,n\}, \quad \alpha \in \{1,2,3,4\}, \quad \gamma \in \{1,2\} \quad \textit{and} \quad \epsilon \geq 0 \right\}, \tag{E.45}$$

initial conditions

$$\mathcal{I} = \left\{ f(0;\epsilon) : i \in \{1,...,n\}, \quad f \in \mathcal{F} \quad \textit{and} \quad \epsilon \geq 0 \right\} \tag{E.46}$$

and vaccination policy $\boldsymbol{U}(t;\epsilon)$. Suppose that

$$|p(\epsilon) - p(0)| \leq \epsilon \quad \forall p \in \mathcal{P}, \tag{E.47}$$

$$|f_i(0;\epsilon) - f_i(0;0)| \leq \epsilon \quad \forall f \in \mathcal{F} \tag{E.48}$$

and that

$$|W_i(t,\epsilon) - W_i(t,0)| \leq \epsilon \quad \forall t \geq 0. \tag{E.49}$$

Moreover, suppose that for each $i \in \{1,...,n\}$ and $\epsilon \geq 0$,

$$U_i(s;\epsilon) \geq 0 \quad \textit{and} \quad \int_0^t U_i(s;\epsilon)ds \leq N_i \quad \forall t \geq 0. \tag{E.50}$$

Then, for each $\delta > 0$ and each $T > 0$ there exists some $\eta > 0$ (that may depend on T and δ) such that

$$\epsilon \in (0,\eta) \Rightarrow |f(t;\epsilon) - f(t;0)| < \delta \quad \forall f \in \mathcal{F} \quad \textit{and} \quad \forall t \in [0,T] \tag{E.51}$$

Proof: The proof is simple but algebraically dense and so is left to Lemma E.22 in the appendices.

This now allows a proof of Theorem 7.1 to be formed.

E.1.4 Theorem 7.1

Theorem 7.1 *Suppose that* $\boldsymbol{U}, \tilde{\boldsymbol{U}} \in C$. *Suppose further that for each* $i \in \{1, ..., n\}$ *and* $t \geq 0$

$$\int_0^t U_i(s)ds \leq \int_0^t \tilde{U}_i(s)ds. \tag{E.52}$$

Then

$$H(\boldsymbol{U}) \geq H(\tilde{\boldsymbol{U}}). \tag{E.53}$$

Proof: Define the parameters $\beta_{ij}^a(\epsilon)$ and $\mu_i^a(\epsilon)$ by

$$\beta_{ij}^a(\epsilon) = \beta_{ij}^a + \frac{\epsilon}{a} \quad \text{and} \quad \mu_i^a(\epsilon) = \mu_i^a. \tag{E.54}$$

This means that, for any $\epsilon > 0$, these parameters satisfy the conditions of Propositions E.1 and E.2. Define, for $\epsilon < 1$, the initial conditions

$$S_i(0;\epsilon) = \begin{cases} S_i(0;0) & \text{if } S_i(0;0), I_i(0;0) > 0 \\ S_i(0;0) + \epsilon N_i & \text{if } S_i(0;0) = 0 \\ S_i(0;0) - \epsilon N_i & \text{if } I_i(0;0) = 0 \end{cases} \tag{E.55}$$

and

$$I_i(0;\epsilon) = N_i - S_i(0;\epsilon). \tag{E.56}$$

Then, the conditions of Propositions E.1 and E.2 are met by these initial conditions for any $\epsilon > 0$.

Now, define the set of points

$$\sigma(\epsilon) := \left\{ n\epsilon : n \in \mathcal{N}_{\geq 0} \right\}. \tag{E.57}$$

Then, define $W_i^*(t;\epsilon)$ to be the first order approximation to the function $\mathcal{W}_i(t;\epsilon) := \max(W_i(t), N_i - \epsilon)$ using the points of $\sigma(\epsilon)$. That is, for each t define

$$K(t;\epsilon) := \inf \left\{ m : m \in \sigma(\epsilon) \quad \text{and} \quad m \geq t \right\} \tag{E.58}$$

and

$$k(t;\epsilon) := \sup \left\{ m : m \in \sigma(\epsilon) \quad \text{and} \quad m \leq t \right\} \tag{E.59}$$

Note that, as $\sigma(\epsilon)$ is nowhere dense, one must have

$$k(t;\epsilon), K(t;\epsilon) \in \sigma(\epsilon) \quad \text{and} \quad k(t;\epsilon) \leq t \leq K(t;\epsilon) \tag{E.60}$$

Then, define

$$W_i^*(t;\epsilon) = (t - k(t;\epsilon))\mathcal{W}_i(k(t;\epsilon);\epsilon) + (K(t;\epsilon) - t)\mathcal{W}_i(K(t;\epsilon);\epsilon). \tag{E.61}$$

Thus, as k and K are constant on any interval not containing a point in $\sigma(\epsilon)$, W_i^* is linear on any interval not containing a point of $\sigma(\epsilon)$ and so its derivative is a step function.

Now, note that, for each t

$$|\mathcal{W}_i(t;\epsilon) - W_i(t)| \leq \epsilon \tag{E.62}$$

and, moreover,

$$t \in S \Rightarrow W_i^*(t;\epsilon) = \mathcal{W}_i(t;\epsilon). \tag{E.63}$$

Also, as U_i is bounded, each W_i (and hence each $\mathcal{W}_i$) are Lipschitz continuous with some Lipschitz constant L. Moreover, each W_i^* is continuous and is differentiable in each interval $(k(t;\epsilon), K(t;\epsilon))$ with a maximal (uniformly bounded) gradient of $U_i(t)$, meaning that W_i^* is also Lipschitz continuous with Lipschitz constant L.

It can now be shown that $|W_i(t) - W_i^*(t;\epsilon)|$ is uniformly bounded in t. For each $t \geq 0$, one can find an element $s \in \sigma(\epsilon)$ such that $|t - s| < \epsilon$. Then,

$$|W_i(t) - W_i^*(t;\epsilon)| \leq |W_i(t) - W_i(s)| + |W_i(s) - W_i^*(s;\epsilon)| + |W_i^*(s;\epsilon) - W_i^*(t;\epsilon)| \tag{E.64}$$

$$\leq L\epsilon + |W_i(s) - \mathcal{W}_i(s;\epsilon)| + L\epsilon \tag{E.65}$$

$$\leq (2L+1)\epsilon \tag{E.66}$$

and so W_i^* converges uniformly to W_i. The same results hold for the analogously defined $\tilde{W}_i^*$. Then, note that, as $\tilde{W}_i(t) \geq W_i(t)$, it must be that $\tilde{\mathcal{W}}_i(t;\epsilon) \geq \mathcal{W}_i(t;\epsilon)$. Thus, it follows that $\tilde{W}_i^*(t;\epsilon) \geq W_i^*(t;\epsilon)$.

This means that Proposition E.2 can be used. Define using stars the variables that come from the $\boldsymbol{U}^*$ and $\tilde{\boldsymbol{U}}^*$ policies. Then, from Proposition 2, for each $t \geq 0$, $\epsilon > 0$ and $i \in \{1, ..., n\}$

$$I_i^*(t;\epsilon) + I_i^{V*}(t;\epsilon) + R_i^*(t;\epsilon) + R_i^{V*}(t;\epsilon) \geq \tilde{I}_i^*(t;\epsilon) + \tilde{I}_i^{V*}(t;\epsilon) + \tilde{R}_i^*(t;\epsilon) + \tilde{R}_i^{V*}(t;\epsilon) \tag{E.67}$$

and

$$R_i^*(t;\epsilon) \geq \tilde{R}_i^*(t;\epsilon). \tag{E.68}$$

Then, taking $\epsilon \to 0$ and using Proposition E.3 (noting that the perturbations to the parameters, initial

conditions and vaccination policies are all bounded by a constant multiple of ϵ) shows that

$$I_i(t) + I_i^V(t) + R_i(t) + R_i^V(t) \geq \tilde{I}_i(t) + \tilde{I}_i^V(t) + \tilde{R}_i(t) + \tilde{R}_i^V(t) \tag{E.69}$$

and

$$R_i(t) \geq R_i^V(t). \tag{E.70}$$

Then, the result follows using the same logic as in the proof of Proposition E.2.

E.1.5 Theorem 7.2

Theorem 7.2 *Suppose that B is differentiable, and that there is an optimal solution $\boldsymbol{U}$ to the optimal vaccination problem. Then, define the function*

$$\chi(t) := \begin{cases} A(t) & \text{if} \quad \int_0^t \chi(s)ds < B(t) \\ \min(A(t), B'(t)) & \text{if} \quad \int_0^t \chi(s)ds \geq B(t) \end{cases} \tag{E.71}$$

and suppose that $\chi(t)$ exists and is bounded. Then, there exists an optimal solution $\tilde{\boldsymbol{U}}$ such that

$$\sum_{i=1}^{n} \tilde{W}_i(t) = \min\left(\int_0^t \chi(s)ds, 1 \right). \tag{E.72}$$

Moreover, if $\chi(t)$ is continuous almost everywhere, there exists an optimal solution $\tilde{\boldsymbol{U}}$ such that

$$\sum_{i=1}^{n} \tilde{U}_i(t) = \begin{cases} \chi(t) & \text{if } \int_0^t \chi(s)ds < 1 \\ 0 & \text{otherwise} \end{cases} \tag{E.73}$$

Proof: Suppose that $\boldsymbol{U}$ is an optimal vaccination policy. To begin, it will be shown that the total vaccination rate χ is indeed a maximal-effort vaccination policy (in the sense that, at each time t, it is impossible to have given out more vaccines than a policy with total overall rate $\chi(t)$).

Claim: $\min\left(1, \int_0^t \chi(s)ds\right) \geq \int_0^t \sum_{i=1}^{n} U_i(s)ds$ for all $t > 0$

Proof: Consider any time $t \geq 0$ such that

$$\int_0^t \chi(s)ds < 1 \tag{E.74}$$

and define the set

$$\mathcal{T} := \left\{ s \leq t : \int_0^s \chi(k)dk \geq B(s) \right\}. \tag{E.75}$$

Suppose that $\mathcal{T} = \emptyset$. Then,

$$\chi(s) = A(s) \quad \forall s \leq t \tag{E.76}$$

and so

$$\int_0^t \chi(s)ds = \int_0^t A(s)ds \geq \int_0^t \sum_{i=1}^n U_i(s)ds. \tag{E.77}$$

Moreover, suppose that $\mathcal{T} \neq \emptyset$ and define

$$\tau := \sup(\mathcal{T}). \tag{E.78}$$

Then,

$$\int_0^\tau \chi(s)ds \geq B(\tau) \geq \int_0^\tau \sum_{i=1}^n U_i(s)ds \tag{E.79}$$

and

$$\int_\tau^t \chi(s)ds = \int_\tau^t A(s)ds \geq \int_\tau^t \sum_{i=1}^n U_i(s)ds \tag{E.80}$$

so that

$$\int_0^t \chi(s)ds \geq \int_0^t \sum_{i=1}^n U_i(s)ds. \tag{E.81}$$

Thus, this holds in all cases for $\int_0^t \chi(s)ds < 1$. Finally, suppose that

$$\int_0^t \chi(s)ds \geq 1. \tag{E.82}$$

Then, one has

$$\min\left(1, \int_0^{t^*} \chi(s)ds\right) = 1 = \sum_{i=1}^n N_i \geq \int_0^{t^*} \sum_{i=1} U_i(s)ds \tag{E.83}$$

and so the claim is proved.

It is now important to show that χ gives a feasible vaccination rate. Note that $\chi(t) \leq A(t)$ by definition.

Claim: $\int_0^t \chi(s)ds \leq B(t)$ for all $t \geq 0$.

Proof: Suppose, for a contradiction, that there exists a t such that

$$\int_0^t \chi(s)ds > B(t). \tag{E.84}$$

Then, define

$$\sigma := \sup\left\{s \leq t : \int_0^t \chi(s)ds \leq B(t)\right\} \tag{E.85}$$

which must exist (as $\int_0^0 \chi(s)ds \leq B(0)$) and satisfy $\sigma < t$, by continuity of $\int_0^t \chi(s)ds$ and $B(t)$. Note that

$$s \in (\sigma, t) \Rightarrow \chi(s) \leq B'(s) \tag{E.86}$$

and so

$$\int_0^t \chi(s)ds \leq \int_0^\sigma \chi(s)ds + \int_\sigma^t B'(s)ds \leq B(\sigma) + (B(t) - B(\sigma)) = B(t), \tag{E.87}$$

which is a contradiction. Thus,

$$\int_0^t \chi(s)ds \leq B(t) \quad \forall t \geq 0 \tag{E.88}$$

as required.

Now, one can create a new optimal vaccination policy with total rate given by χ. Define

$$q(t) = \begin{cases} \inf\left\{s : \int_0^s \sum_{j=1}^n U_j(k)dk = \int_0^t \chi(k)dk\right\} & \text{if this exists} \\ \\ \infty & \text{otherwise} \end{cases} \tag{E.89}$$

so that $q(t)$ represents the earliest time at which $\chi(t)$ vaccines were administered by the $\boldsymbol{U}$ policy. By continuity of the integral, this means that

$$\sum_{i=1}^n W_i(q(t)) = \int_0^{q(t)} \sum_{j=1}^n U_j(k)dk = \int_0^t \chi(s)ds. \tag{E.90}$$

Define further

$$Q := \sup\{t : q(t) < \infty\} \quad \text{and} \quad q_\infty := \lim_{t \to Q}(q(t)) \tag{E.91}$$

so that Q is the earliest time at which all of the vaccines given out by the $\boldsymbol{U}$ policy could have been administered. Note that both Q and q_∞ may be infinite. By taking the limit $t \to Q$, and noting the

left-hand side is bounded by 1,

$$\int_0^{q_\infty} \sum_{j=1}^{n} U_j(k)dk = \int_0^{Q} \chi(k)dk \tag{E.92}$$

Then, the integral of the new vaccination policy, $\tilde{\boldsymbol{W}}$ is given by

$$\tilde{W}_i(t) = \begin{cases} W_i(q(t)) & \text{if } t < Q \\ W_i(q_\infty) + \frac{(N_i - W_i(q_\infty))\int_Q^t \chi(s)ds}{1-\sum_{i=1}^n W_i(q_\infty)} & \text{if } \int_0^t \chi(s)ds < 1 \quad \text{and} \quad t \geq Q \\ N_i & \text{if } \int_0^t \chi(s)ds \geq 1 \quad \text{and} \quad t \geq Q \end{cases} . \tag{E.93}$$

This is well-defined as

$$\sum_{i=1}^{n} W_i(q_\infty) = 1 \Rightarrow \int_0^Q \chi(s)ds = 1 \tag{E.94}$$

and so, in this case, the second part of the definition of χ is never used. It is important to establish for feasibility that each $\tilde{W}_i$ is bounded by N_i.

Claim: $\tilde{W}_i(t) \leq N_i$ for all $t \geq 0$ and all $i \in \{1, ..., n\}$.

Proof: If $t < Q$, then $W_i(q(t)) \leq N_i$ for all $t < Q$ by feasibility of $\boldsymbol{U}$. Otherwise, if $t \geq Q$ and $\int_0^t \chi(s)ds < 1$, then one has

$$W_i(q_\infty) + \frac{(N_i - W_i(q_\infty))\int_Q^t \chi(s)ds}{1-\sum_{i=1}^n W_i(q_\infty)} \leq W_i(q_\infty) + \frac{(N_i - W_i(q_\infty))(1-\int_0^Q \chi(s)ds)}{1-\sum_{i=1}^n W_i(q_\infty)} \tag{E.95}$$

$$= W_i(q_\infty) + \frac{(N_i - W_i(q_\infty))(1-\sum_{i=1}^n W_i(q_\infty))}{1-\sum_{i=1}^n W_i(q_\infty)} \tag{E.96}$$

$$= N_i \tag{E.97}$$

while if $\int_0^t \chi(s)ds \geq 1$ then the result is immediate.

The optimisation problem is framed in terms of $\boldsymbol{U}$ rather than $\boldsymbol{W}$, and so it is important to show that there is some $\tilde{\boldsymbol{U}}$ that integrates to $\tilde{\boldsymbol{W}}$. One can do this by proving the Lipschitz continuity of $\tilde{W}_i$ for each i.

Claim: $\tilde{W}_i(t)$ is Lipschitz continuous for each $i \in \{1, ..., n\}$

Proof: Note that for $s, t < Q$, if M is a bound for χ (which is assumed to exist)

$$|\tilde{W}_i(t) - \tilde{W}_i(s)| = \left| \int_{q(s)}^{q(t)} U_i(k)dk \right| \tag{E.98}$$

$$\leq \left| \int_{q(s)}^{q(t)} \sum_{j=1}^{n} U_j(k)dk \right| \tag{E.99}$$

$$= \left| \int_{s}^{t} \chi(k)dk \right| \tag{E.100}$$

$$\leq |t - s|M \tag{E.101}$$

Moreover, if $s, t > Q$ and $\int_0^t \chi(k)dk, \int_0^s \chi(k)dk < 1$, then

$$|\tilde{W}_i(t) - \tilde{W}_i(s)| \leq \left| \frac{(N_i - W_i(q_\infty)) \int_s^t \chi(s)ds}{1 - \sum_{i=1}^n W_i(q_\infty)} \right| \leq M \left| \frac{(N_i - W_i(q_\infty))}{1 - \sum_{i=1}^n W_i(q_\infty)} \right| |t - s| \tag{E.102}$$

and if $s, t > Q$ and $\int_0^t \chi(k)dk, \int_0^s \chi(k)dk \geq 1$, then $\tilde{W}_i(t) = \tilde{W}_i(s)$. The intermediate cases (where s and t correspond to different cases in the definition of χ) can be proved by combining these bounds.

This means that (for each i) there exists a Lebesgue integrable function $\tilde{U}_i(t)$ such that

$$\frac{d\tilde{W}_i}{dt} = \tilde{U}_i(t) \quad \text{almost everywhere} \tag{E.103}$$

and, for all $t \geq 0$

$$\int_0^t \tilde{U}_i(s)ds = \tilde{W}_i(t) \tag{E.104}$$

A proof of this (for the broader class of absolutely continuous functions) can be found in [504]. One can set $\tilde{U}_i(t)$ to be zero for any t such that $\tilde{W}_i(t)$ is not differentiable. Thus, noting that, where it is differentiable, the derivative of $\tilde{W}_i$ is bounded by its Lipschitz constant, $\tilde{U}_i(t)$ is bounded as required.

Note that, in all cases (as $\sum_{i=1}^n N_i = 1$)

$$\sum_{i=1}^{n} \tilde{W}_i(t) = \min\left(\int_0^t \chi(s)ds, 1 \right) \tag{E.105}$$

and so $\tilde{\boldsymbol{W}}$ does correspond to a maximal vaccination rate. If $\chi(t)$ is continuous almost everywhere, then one can differentiate this relationship at t where each $\tilde{W}_i$ is differentiable and χ is continuous with $\int_0^t \chi(s)ds < 1$ to show that $\sum_{i=1}^n U_i(t) = \chi(t)$. The complement of this set must have zero measure (as it is the finite union of zero measure sets), and so, in this case, one can change the values of each

$U_i(t)$ so that $\sum_{i=1}^n U_i(t) = \chi(t)$ everywhere without changing the value of $\boldsymbol{W}$.

Claim: $\tilde{W}_i(t) \geq W_i(t)$ for all $i \in \{1, ..., n\}$ and $t \geq 0$

Proof: Note that, by maximality of χ, for $t < Q$,

$$\sum_{j=1}^n \tilde{W}_i(t) = \sum_{j=1}^n W_j(q(t)) = \int_0^t \chi(s)ds \geq \sum_{j=1}^n W_j(t) \tag{E.106}$$

If $q(t) \geq t$, then $W_i(q(t)) \geq W_i(t)$ for each i. If $q(t) < t$, then it is necessary that $W_i(q(t)) = W_i(t)$ for each i as W_i is non-decreasing. Thus, $W_i(q(t)) \geq W_i(t)$ for all i and for all $t < Q$.

If $t > Q$ and $\int_0^t \chi(s)ds < 1$, then

$$\tilde{W}_i(t) \geq W_i(q_\infty) \tag{E.107}$$

Now, by definition of Q, it is necessary that

$$\int_0^t \chi(k)dk \geq \int_0^\infty \sum_{j=1}^n U_j(k)dk \quad \forall t > Q \tag{E.108}$$

as otherwise, there must exist some $t > Q$ and some $s < \infty$ such that

$$\int_0^t \chi(k)dk = \int_0^s \sum_{j=1}^n U_j(k)dk \tag{E.109}$$

which means that $q(t) < \infty$. Thus, by continuity, for all $\tau \in (0, t)$, there exists some s such that

$$\int_0^\tau \chi(k)dk = \int_0^s \sum_{j=1}^n U_j(k)dk \tag{E.110}$$

which means $Q \geq t$, which is a contradiction.

Thus, by taking $t \to Q$,

$$\int_0^{q_\infty} \sum_{i=1}^n U_i(k)dk = \int_0^Q \chi(k)dk \geq \int_0^\infty \sum_{j=1}^n U_j(k)dk \tag{E.111}$$

and so

$$\int_{q_\infty}^\infty U_j(k)dk = 0 \Rightarrow W_i(t) = W_i(q_\infty) \quad \forall i \in \{1, ..., n\} \quad \text{and} \quad \forall t \geq q_\infty \tag{E.112}$$

Thus, using (E.107),

$$\tilde{W}_i(t) \geq W_i(t). \tag{E.113}$$

Finally, if $t > Q$ and $\int_0^t \chi(s)ds \geq 1$, then $\tilde{W}_i(t) = N_i \geq W_i(t)$. Thus, for all t and i,

$$\tilde{W}_i(t) \geq W_i(t) \tag{E.114}$$

as required.

Thus, by Theorem 7.1, it is necessary that

$$H(\boldsymbol{U}) \geq H(\tilde{\boldsymbol{U}}) \tag{E.115}$$

and hence, by the optimality of $\boldsymbol{U}$, $\tilde{\boldsymbol{U}}$ is optimal as required.

E.1.6 Theorem 7.3

Theorem 7.3 *Under the assumptions of Theorem 7.2, consider a modified objective function $\mathcal{H}$ given by*

$$\mathcal{H}(\boldsymbol{U}) = H(\boldsymbol{U}) + F(\boldsymbol{W}(\infty)) \tag{E.116}$$

for any function F. Then, with χ defined to be the maximal vaccination effort as in Theorem 7.2, there exists an optimal solution $\tilde{\boldsymbol{U}}$ such that, for some $\tau \geq 0$

$$\sum_{i=1}^{n} \tilde{W}_i(t) = \begin{cases} \int_0^t \chi(s)ds & \text{if } t \leq \tau \\ \\ W_i(\tau) & \text{otherwise} \end{cases}. \tag{E.117}$$

Moreover, if χ is continuous almost everywhere, then there is an optimal solution $\tilde{\boldsymbol{U}}$ such that

$$\sum_{i=1}^{n} U_i(t) = \begin{cases} \chi(t) & \text{if } t \leq \tau \\ 0 & \text{otherwise} \end{cases}. \tag{E.118}$$

Proof: This follows directly from the proof of Theorem 7.2. One can again define $\tilde{\boldsymbol{U}}$ in the interval $(0, Q)$ (where Q is defined in the proof of Theorem 7.2) such that

$$H(\boldsymbol{U}) \geq H(\tilde{\boldsymbol{U}}) \quad \text{and} \quad \int_0^t \sum_{i=1}^{n} \tilde{U}_i(s)ds = \int_0^t \chi(s)ds \quad \forall t < Q \tag{E.119}$$

with the only difference being that now

$$\tilde{U}_i(t) = 0 \quad \forall t \geq Q. \tag{E.120}$$

Thus, as shown in the proof of Theorem 7.2,

$$\boldsymbol{W}(\infty) = \boldsymbol{W}(q_\infty) = \tilde{\boldsymbol{W}}(Q) = \tilde{\boldsymbol{W}}(\infty) \tag{E.121}$$

and so

$$\mathcal{H}(\boldsymbol{U}) \geq \mathcal{H}(\tilde{\boldsymbol{U}}), \tag{E.122}$$

which means $\tilde{\boldsymbol{U}}$ is optimal as required.

E.2 Supplementary Lemmas For Propositions E.1 and E.2 and Theorem 7.2

For the proofs of these lemmas, it is helpful to recall the following definitions of the following variables, which will be extensively used.

$$K_{ij}(t) = \frac{\beta_{ij}^1}{\mu_j^1} R_j + \frac{\beta_{ij}^2}{\mu_j^2} R_j^V, \tag{E.123}$$

$$L_{ij}(t) := \frac{\beta_{ij}^3}{\mu_i^1} R_j + \frac{\beta_{ij}^4}{\mu_i^2} R_j^V \tag{E.124}$$

and

$$\Pi := \left\{ i : \exists t \geq 0 \quad \text{s.t.} \quad I_i(t) > 0 \quad \text{or} \quad I_i^V(t) > 0 \right\}. \tag{E.125}$$

Moreover, note that, under the assumptions of Proposition E.1 and E.2, each $U_i(t)$ is a step function and is therefore piecewise smooth in each bounded interval. Thus, in particular, the derivatives of each of the model variables (and indeed, the derivative of $W_i(t)$) are piecewise continuous in each bounded interval, meaning that each of the model variables is piecewise continuously differentiable in each bounded interval. This means that integration by parts can be performed (in a bounded interval), as will be done extensively throughout the proofs of these lemmas.

E.2.1 Lemma E.4

Lemma E.4 *Suppose that $f(t)$ is a non-increasing, non-negative, continuous and piecewise continuously differentiable function and that the continuous and piecewise continuously differentiable functions*

$g(t)$ and $h(t)$ satisfy $g(0) = h(0)$ and $g(t) \leq h(t)$ for all $t \geq 0$. Then,

$$\int_0^t g'(s)f(s)ds \leq \int_0^t h'(s)f(s)ds. \tag{E.126}$$

Proof: This follows from integrating by parts:

$$\begin{aligned}
\int_0^t g'(s)f(s)ds &= g(t)f(t) - g(0)f(0) - \int_0^t g(s)f'(s)ds && \text{(E.127)}\\
&= g(t)f(t) - h(0)f(0) + \int_0^t g(s)|f'(s)|ds && \text{(E.128)}\\
&\leq h(t)f(t) - h(0)f(0) + \int_0^t h(s)|f'(s)|ds && \text{(E.129)}\\
&\leq h(t)f(t) - h(0)f(0) - \int_0^t h(s)f'(s)ds && \text{(E.130)}\\
&= \int_0^t h'(s)f(s)ds && \text{(E.131)}
\end{aligned}$$

as required.

E.2.2 Lemma E.5

Lemma E.5 *Suppose that*

$$\sum_{j=1}^n K_{ij}(t) \geq \sum_{j=1}^n \tilde{K}_{ij}(t) \quad \textit{and} \quad \sum_{j=1}^n L_{ij}(t) \geq \sum_{j=1}^n \tilde{L}_{ij}(t) \quad \forall i \in \{1, ..., n\} \quad \textit{and} \quad t \in [0, T]. \tag{E.132}$$

Then,

$$S_i(t) + S_i^V(t) \leq \tilde{S}_i(t) + \tilde{S}_i^V(t) \quad \forall t \in [0, T]. \tag{E.133}$$

Proof: To reduce notation in this proof, define

$$\mathcal{K}(t) := \sum_{j=1}^n K_{ij}(t) \quad \text{and} \quad \mathcal{L}(t) := \sum_{j=1}^n L_{ij}(t) \tag{E.134}$$

Note that

$$\begin{aligned}
\frac{d}{dt}(S_i + S_i^V) &= -\sum_{j=1}^n \left(\beta_{ij}^1 I_j + \beta_{ij}^2 I_j^V\right) S_i - \sum_{j=1}^n \left(\beta_{ij}^3 I_j + \beta_{ij}^4 I_j^V\right) S_i^V && \text{(E.135)}\\
&= -\sum_{j=1}^n \left(\beta_{ij}^3 I_j + \beta_{ij}^4 I_j^V\right)(S_i + S_i^V) ... && \text{(E.136)}\\
&\quad - \sum_{j=1}^n \left((\beta_{ij}^1 - \beta_{ij}^3) I_j + (\beta_{ij}^2 - \beta_{ij}^4) I_j^V\right) S_i. && \text{(E.137)}
\end{aligned}$$

Thus,

$$-\sum_{j=1}^{n}\left((\beta_{ij}^{1}-\beta_{ij}^{3})I_j+(\beta_{ij}^{2}-\beta_{ij}^{4})I_j^V\right)S_i=\frac{d}{dt}(S_i+S_i^V)... \tag{E.138}$$

$$+\sum_{j=1}^{n}\left(\beta_{ij}^{3}I_j+\beta_{ij}^{4}I_j^V\right)(S_i+S_i^V) \tag{E.139}$$

$$=\frac{d}{dt}\left((S_i+S_i^V)e^{\mathcal{L}(t)}\right)e^{-\mathcal{L}(t)}. \tag{E.140}$$

This means that

$$S_i(t)+S_i^V(t)=e^{-\mathcal{L}(t)}\left[S_i(0)-\int_0^t e^{\mathcal{L}(s)}\sum_{j=1}^{n}\left((\beta_{ij}^{1}-\beta_{ij}^{3})I_j+(\beta_{ij}^{2}-\beta_{ij}^{4})I_j^V\right)S_i ds\right] \tag{E.141}$$

$$=S_i(0)\left[e^{-\mathcal{L}(t)}-\int_0^t e^{\mathcal{L}(s)-\mathcal{K}(s)-\mathcal{L}(t)}(\mathcal{K}'(s)-\mathcal{L}'(s))\left(\frac{N_i-W_i(s)}{N_i}\right)\right]ds. \tag{E.142}$$

where in the second line, we have used the final-size equation

$$S_i=S_i(0)e^{-\mathcal{K}(s)}\left(\frac{N_i-W_i(s)}{N_i}\right) \tag{E.143}$$

as is derived in the course of Lemma E.6.

Now, one can see that, as $0\leq W_i(s)\leq N_i$,

$$0\leq\frac{N_i-W_i(s)}{N_i}\leq 1\quad\forall s\geq 0 \tag{E.144}$$

and hence

$$e^{-\mathcal{L}(t)}=1-\int_0^t\mathcal{L}'(s)e^{-\mathcal{L}(s)}ds\leq 1-\int_0^t\mathcal{L}'(s)e^{-\mathcal{L}(s)}\left(\frac{N_i-W_i(s)}{N_i}\right)ds. \tag{E.145}$$

Now, this means that

$$S_i(t)+S_i^V(t)\leq \tag{E.146}$$

$$S_i(0)-S_i(0)\int_0^t\left[\mathcal{L}'(s)e^{-\mathcal{L}(s)}+e^{\mathcal{L}(s)-\mathcal{K}(s)-\mathcal{L}(t)}\sum_{j=1}^{n}\left(\mathcal{K}'(s)-\mathcal{L}'(s)\right)\right]\left(\frac{N_i-W_i(s)}{N_i}\right)ds. \tag{E.147}$$

This allows the use of Lemma E.4. Firstly, note that, as $\mathcal{K}'(s)\geq\mathcal{L}'(s)\geq 0$ and $\tilde{W}_i(s)\geq W_i(s)$, one

has

$$S_i(t) + S_i^V(t) \leq \tag{E.148}$$

$$S_i(0) - S_i(0) \int_0^t \left[\mathcal{L}'(s)e^{-\mathcal{L}(s)} + e^{\mathcal{L}(s)-\mathcal{K}(s)-\mathcal{L}(t)} \sum_{j=1}^{n} \left(\mathcal{K}'(s) - \mathcal{L}'(s)\right)\right] \left(\frac{N_i - \tilde{W}_i(s)}{N_i}\right) ds. \tag{E.149}$$

Moreover,

$$\int_0^t \left[\mathcal{L}'(s)e^{-\mathcal{L}(s)} + e^{\mathcal{L}(s)-\mathcal{K}(s)-\mathcal{L}(t)} \sum_{j=1}^{n} \left(\mathcal{K}'(s) - \mathcal{L}'(s)\right)\right] ds \tag{E.150}$$

$$= 1 - e^{-\mathcal{L}(t)} + e^{-\mathcal{L}(t)} - e^{-\mathcal{K}(t)} \tag{E.151}$$

$$= 1 - e^{-\mathcal{K}(t)} \tag{E.152}$$

$$\geq 1 - e^{-\tilde{\mathcal{K}}(t)} \tag{E.153}$$

$$\geq \int_0^t \left[\tilde{\mathcal{L}}'(s)e^{-\tilde{\mathcal{L}}(s)} + e^{\tilde{\mathcal{L}}(s)-\tilde{\mathcal{K}}(s)-\tilde{\mathcal{L}}(t)} \sum_{j=1}^{n} \left(\tilde{\mathcal{K}}'(s) - \tilde{\mathcal{L}}'(s)\right)\right] ds \tag{E.154}$$

$$\tag{E.155}$$

and $N_i - \tilde{W}_i(s)$ is non-increasing in s. Thus, by Lemma E.4, with

$$g(s) = 1 - e^{-\mathcal{L}(s)} + e^{-\mathcal{L}(t)} - e^{\mathcal{L}(s)-\mathcal{K}(s)-\mathcal{L}(t)}, \tag{E.156}$$

$h(s)$ defined as the tilde version of $g(s)$, and $f(s) := N_i - \tilde{W}_i(s)$, one has

$$\int_0^t \left[\mathcal{L}'(s)e^{-\mathcal{L}(s)} + e^{\mathcal{L}(s)-\mathcal{K}(s)-\mathcal{L}(t)} \sum_{j=1}^{n} \left(\mathcal{K}'(s) - \mathcal{L}'(s)\right)\right] \left(\frac{N_i - \tilde{W}_i(s)}{N_i}\right) ds \tag{E.157}$$

$$\tag{E.158}$$

$$\geq \int_0^t \left[\tilde{\mathcal{L}}'(s)e^{-\tilde{\mathcal{L}}(s)} + e^{\tilde{\mathcal{L}}(s)-\tilde{\mathcal{K}}(s)-\tilde{\mathcal{L}}(t)} \sum_{j=1}^{n} \left(\tilde{\mathcal{K}}'(s) - \tilde{\mathcal{L}}'(s)\right)\right] \left(\frac{N_i - \tilde{W}_i(s)}{N_i}\right) ds.$$

Thus, (as this integral is multiplied by -1 in (E.158)), combining this with (E.158) gives

$$S_i(t) + S_i^V(t) \leq \tilde{S}_i(t) + \tilde{S}_i^V(t) \quad \forall t \in [0, T] \tag{E.159}$$

as required

E.2.3 Lemma E.6

Lemma E.6 *Suppose that*

$$\sum_{j=1}^{n} K_{ij}(t) \geq \sum_{j=1}^{n} \tilde{K}_{ij}(t) \quad \forall i \in \{1, ..., n\} \quad \textit{and} \quad t \in [0, T]. \tag{E.160}$$

Then

$$I_i(t) + R_i(t) \geq \tilde{I}_i(t) + \tilde{R}_i(t) \quad \forall t \in [0, T]. \tag{E.161}$$

To begin, one can write the equation for S_i as

$$\frac{1}{S_i}\frac{dS_i}{dt} = -\sum_{j=1}^{n}(K'_{ij}(t)) - \frac{U_i}{N_i - W_i} \tag{E.162}$$

and hence, integrating

$$\ln(S_i(t)) - \ln(S_i(0)) = -\sum_{j=1}^{n} K_{ij}(t) + \ln(N_i - W_i(t)) - \ln(N_i) \tag{E.163}$$

which implies

$$S_i(t) = \left(\frac{S_i(0)(N_i - W_i(t))}{N_i}\right) e^{-\sum_{j=1}^{n} K_{ij}(t)} \tag{E.164}$$

Using this result shows that

$$\frac{d}{dt}(I_i + R_i) = \sum_{j=1}^{n} \left(\beta_{ij}^1 I_j + \beta_{ij}^2 I_j^V\right) S_i \tag{E.165}$$

$$= \sum_{j=1}^{n} K'_{ij}(t) S_i \tag{E.166}$$

$$= \left[\sum_{j=1}^{n} K'_{ij}(t)\right] \left(\frac{S_i(0)(N_i - W_i(t))}{N_i}\right) e^{-\sum_{j=1}^{n} K_{ij}(t)}, \tag{E.167}$$

Thus,

$$I_i(t) + R_i(t) = I_i(0) + \int_0^t \left[\sum_{j=1}^{n} K'_{ij}(s)\right] \left(\frac{S_i(0)(N_i - W_i(s))}{N_i}\right) e^{-\sum_{j=1}^{n} K_{ij}(s)} ds \tag{E.168}$$

$$\geq \tilde{I}_i(0) + \int_0^t \left[\sum_{j=1}^{n} K'_{ij}(s)\right] \left(\frac{S_i(0)(N_i - \tilde{W}_i(s))}{N_i}\right) e^{-\sum_{j=1}^{n} K_{ij}(s)} ds, \tag{E.169}$$

$$\tag{E.170}$$

using the fact that the initial conditions are the same in both cases and that $W_i \leq \tilde{W}_i$. Now, one can

use the results of Lemma E.4 with

$$g(t) = 1 - \exp\Big(- \sum_{j=1}^{n} K_{ij}(t) \Big), \quad h(t) = 1 - \exp\Big(- \sum_{j=1}^{n} \tilde{K}_{ij}(t) \Big) \tag{E.171}$$

and $f(t) = (N_i - \tilde{W}_i(t))$, noting that

$$\int_0^t \Big[\sum_{j=1}^{n} K'_{ij}(s) \Big] e^{-\sum_{j=1}^{n} K_{ij}(s)} ds = 1 - e^{-\sum_{j=1}^{n} K_{ij}(t)} \tag{E.172}$$

$$\geq 1 - e^{-\sum_{j=1}^{n} \tilde{K}_{ij}(t)} \tag{E.173}$$

$$= \int_0^t \Big[\sum_{j=1}^{n} \tilde{K}'_{ij}(s) \Big] e^{-\sum_{j=1}^{n} \tilde{K}_{ij}(s)} ds \tag{E.174}$$

and that $N_i - \tilde{W}_i(t)$ is non-increasing, we have

$$I_i(t) + R_i(t) \geq \tilde{I}_i(t) + \tilde{R}_i(t) \quad \forall t \in [0, T] \tag{E.175}$$

as required.

E.2.4 Lemma E.7

Lemma E.7 *Suppose that*

$$\sum_{j=1}^{n} K_{ij}(t) \geq \sum_{j=1}^{n} \tilde{K}_{ij}(t) \quad \forall i \in \{1, ..., n\} \quad \textit{and} \quad t \in [0, T]. \tag{E.176}$$

Then,

$$R_i(t) \geq \tilde{R}_i(t) \quad \forall t \in [0, T] \tag{E.177}$$

Proof: The result of Lemma E.6 can be written as

$$\frac{1}{\mu_i^1} \frac{dR_i}{dt} + R_i \geq \frac{1}{\mu_i^1} \frac{d\tilde{R}_i}{dt} + \tilde{R}_i \quad \forall t \in [0, T] \tag{E.178}$$

which implies

$$\frac{d}{dt} \left(R_i e^{\mu_i^1 t} \right) \geq \frac{d}{dt} \left(\tilde{R}_i e^{\mu_i^1 t} \right) \tag{E.179}$$

and hence, after integrating and cancelling exponentials, one finds

$$R_i(t) \geq \tilde{R}_i(t) \quad \forall t \in [0, T] \tag{E.180}$$

as required.

E.2.5 Lemma E.8

Lemma E.8 *Suppose that*

$$T := \inf\left\{t : K_{ij}(t) < \tilde{K}_{ij}(t) \quad or \quad L_{ij}(t) < \tilde{L}_{ij}(t) \quad for\ some\ i,j \in \{1,...,n\}\right\} \tag{E.181}$$

exists. Then, for some $b \in \{1,...,n\}$*, and some real constants* κ *and* η*,*

$$R_b(T) + \kappa R_b^V(T) = \tilde{R}_b(T) + \kappa \tilde{R}_b^V(T), \tag{E.182}$$

$$I_b(T) + \eta I_b^V(T) \leq \tilde{I}_b(T) + \eta \tilde{I}_b^V(T) \tag{E.183}$$

and

$$0 \leq \kappa \leq \eta \leq 1 \tag{E.184}$$

Proof: Suppose that T exists. Then, by continuity, there exists some a and b such that $K_{ab}(T) = \tilde{K}_{ab}(T)$ or $L_{ab}(T) = \tilde{L}_{ab}(T)$. These can be rearranged to give, respectively,

$$R_b(T) + \frac{\mu_b^1 \beta_{ab}^2}{\mu_b^2 \beta_{ab}^1} R_b^V(T) = \tilde{R}_b(T) + \frac{\mu_b^1 \beta_{ab}^2}{\mu_b^2 \beta_{ab}^1} \tilde{R}_b^V(T) \tag{E.185}$$

or

$$R_b(T) + \frac{\mu_b^1 \beta_{ab}^4}{\mu_b^2 \beta_{ab}^3} R_b^V(T) = \tilde{R}_b(T) + \frac{\mu_b^1 \beta_{ab}^4}{\mu_b^2 \beta_{ab}^3} \tilde{R}_b^V(T). \tag{E.186}$$

This can be written as

$$R_b(T) + \kappa R_b^V(T) = \tilde{R}_b(T) + \kappa \tilde{R}_b^V(T), \tag{E.187}$$

where, by the inequality constraints on the β_{ij}^α and μ_i^α

$$\kappa \leq \frac{\mu_b^1}{\mu_b^2}. \tag{E.188}$$

Moreover, note that

$$\frac{d}{dt}\left(R_b + \kappa R_b^V\right) = \mu_b^1 I_b + \frac{\beta_{ab}^2 \mu_b^1}{\beta_{ab}^1} I_b^V \tag{E.189}$$

is a continuous function. Thus, if

$$\frac{d}{dt}\left(R_b + \kappa R_b^V\right)\Big|_{t=T} > \frac{d}{dt}\left(\tilde{R}_b + \kappa \tilde{R}_b^V\right)\Big|_{t=T}, \tag{E.190}$$

then there exists some $\tau > 0$ such that

$$\int_T^{T+\tau} \frac{d}{dt}\Big(R_b(s) + \kappa R_b^V(s)\Big)ds > \int_T^{T+\tau} \frac{d}{dt}\Big(\tilde{R}_b(s) + \kappa \tilde{R}_b^V(s)\Big)ds \quad \forall t \in [0,\tau] \tag{E.191}$$

and hence, in particular

$$R_b(T+t) + \kappa R_b^V(T+t) > \tilde{R}_b(T+t) + \kappa \tilde{R}_b^V(T+t) \quad \forall t \in [0,\tau], \tag{E.192}$$

Thus, it is necessary that there is some b such that

$$\frac{d}{dt}\left(R_b + \kappa R_b^V\right)\Big|_{t=T} \leq \frac{d}{dt}\left(\tilde{R}_b + \kappa \tilde{R}_b^V\right)\Big|_{t=T} \tag{E.193}$$

so

$$I_b(T) + \frac{\kappa \mu_b^2}{\mu_b^1} I^V(T) \leq \tilde{I}_b(T) + \frac{\kappa \mu_b^2}{\mu_b^1} \tilde{I}_b^V(T). \tag{E.194}$$

This can be written as

$$I_b(t) + \eta I_b^V(t) \leq \tilde{I}_b(t) + \eta \tilde{I}_b^V(t), \tag{E.195}$$

where, by (E.188), the fact that $\mu_b^2 \geq \mu_b^1$, and the non-negativity of all parameters,

$$0 \leq \kappa \leq \eta \leq 1 \tag{E.196}$$

as required.

E.2.6 Lemma E.9

For the purposes of this lemma, it is helpful to recall the inequality system (E.10)-(E.16).

$$S_b(T) + S_b^V(T) \leq \tilde{S}_b(T) + \tilde{S}_b^V(T), \tag{E.10}$$

$$I_b(T) + R_b(T) \geq \tilde{I}_b(T) + \tilde{R}_b(T) \tag{E.11}$$

$$R_b(T) \geq \tilde{R}_b(T), \tag{E.12}$$

$$R_b(T) + \kappa R_b^V(T) \leq \tilde{R}_b(T) + \kappa \tilde{R}_b^V(T), \tag{E.13}$$

$$I_b(T) + \eta I_b^V(T) \leq \tilde{I}_b(T) + \eta \tilde{I}_b^V(T), \tag{E.14}$$

$$0 \leq \kappa \leq \eta \leq 1. \tag{E.15}$$

and

$$S_b(T) + I_b(T) + R_b(T) + S_b^V(T) + I_b^V(T) + R_b^V(T) =$$
$$\tilde{S}_b(T) + \tilde{I}_b(T) + \tilde{R}_b(T) + \tilde{S}_b^V(T) + \tilde{I}_b^V(T) + \tilde{R}_b^V(T), \tag{E.16}$$

Lemma E.9 *Suppose that the system (E.10) - (E.16) holds for some $b \in \{1, ..., n\}$ and some $T \geq 0$. Then,*

$$\eta I_b^V(T) + \kappa R_b^V(T) = \eta \tilde{I}_b^V(T) + \kappa \tilde{R}_b^V(T) \tag{E.197}$$
$$I_b(T) + R_b(T) = \tilde{I}_b(T) + \tilde{R}_b(T) \tag{E.198}$$
$$I_b^V(T) + R_b^V(T) = \tilde{I}_b^V(T) + \tilde{R}_b^V(T) \tag{E.199}$$
$$S_b(T) + S_b^V(T) = \tilde{S}_b(T) + \tilde{S}_b^V(T). \tag{E.200}$$

Proof: To begin, note that adding inequalities (E.10), (E.13) and (E.14) gives

$$S_b(T) + S_b^V(T) + R_b(T) + \kappa R_b^V(T) + I_b(T) + \eta I_b^V(T) \leq \tag{E.201}$$
$$\tilde{S}_b(T) + \tilde{S}_b^V(T) + \tilde{R}_b(T) + \kappa \tilde{R}_b^V(T) + \tilde{I}_b(T) + \eta \tilde{I}_b^V(T) \tag{E.202}$$

and then, using (E.16) shows that

$$(\kappa - 1)R_b^V(T) + (\eta - 1)I_b^V(T) \leq (\kappa - 1)\tilde{R}_b^V(T) + (\eta - 1)\tilde{I}_b^V(T). \tag{E.203}$$

Moreover, adding (E.13) and (E.14) shows that

$$I_b(T) + \eta I_b^V(T) + R_b(T) + \kappa R_b^V(T) \leq \tilde{I}_b(T) + \eta \tilde{I}_b^V(T) + \tilde{R}_b(T) + \kappa \tilde{R}_b^V(T) \tag{E.204}$$

and then, using (E.11) shows that

$$\eta I_b^V(T) + \kappa R_b^V(T) \leq \eta \tilde{I}_b^V(T) + \kappa \tilde{R}_b^V(T). \tag{E.205}$$

Now, from the inequality (E.13) combined with the inequality (E.12), it must be the case that

$$R_b^V(T) - \tilde{R}_b^V(T) \leq \frac{1}{\kappa}(\tilde{R}_b(T) - R_b(T)) \leq 0. \tag{E.206}$$

Define

$$x := R_b^V(T) - \tilde{R}_b^V(T) \quad \text{and} \quad y := I_b^V(T) - \tilde{I}_b^V(T) \tag{E.207}$$

so that the system given by (E.15), (E.203), (E.205) and (E.206) reduces to

$$(\kappa - 1)x + (\eta - 1)y \leq 0 \tag{E.208}$$

$$\kappa x + \eta y \leq 0 \tag{E.209}$$

$$x \leq 0 \tag{E.210}$$

$$0 \leq \kappa \leq \eta \leq 1. \tag{E.211}$$

Note first that $x = 0$ implies that $y = 0$ as η and $(\eta - 1)$ have different signs. Thus, in this case, the inequalities (E.208) and (E.209) are in fact equalities.

Suppose instead that $x \neq 0$ (so $x < 0$). The first two of these inequalities can be rearranged (noting the signs of the denominators) to give

$$-\frac{(\kappa - 1)x}{(\eta - 1)} \leq y \leq \frac{-\kappa x}{\eta} \tag{E.212}$$

and so, as $-x > 0$,

$$\frac{(\kappa - 1)}{(\eta - 1)} \leq -\frac{y}{x} \leq \frac{\kappa}{\eta}. \tag{E.213}$$

However, note that

$$\kappa < \eta \Rightarrow \eta\kappa - \eta < \eta\kappa - \kappa \tag{E.214}$$

$$\Rightarrow \eta(\kappa - 1) < \kappa(\eta - 1) \tag{E.215}$$

$$\Rightarrow \frac{\kappa - 1}{\eta - 1} > \frac{\kappa}{\eta} \tag{E.216}$$

and hence, as $\kappa \leq \eta$, for there to be solutions to the inequality (E.213), it is necessary that

$$\kappa = \eta \Rightarrow \frac{-y}{x} = 1 \Rightarrow y = -x. \tag{E.217}$$

This means that the inequalities (E.208) and (E.209) are satisfied to equality in this and hence, from before, all cases. Thus, it is necessary that

$$(\kappa - 1)R_b^V(T) + (\eta - 1)I_b^V(T) = (\kappa - 1)\tilde{R}_b^V(T) + (\eta - 1)\tilde{I}_b^V(T) \tag{E.218}$$

and

$$\eta I_b^V(T) + \kappa R_b^V(T) = \eta \tilde{I}_b^V(T) + \kappa \tilde{R}_b^V(T), \tag{E.219}$$

which is the first required equality. Thus, one can once again add the inequalities (E.13) and (E.14) to give

$$I_b(T) + R_b(T) + \left[\eta I_b^V(T) + \kappa R_b^V(T)\right] \leq \tilde{I}_b(T) + \tilde{R}_b(T) + \left[\eta \tilde{I}_b^V(T) + \kappa \tilde{R}_b^V(T)\right] \tag{E.220}$$

and so

$$I_b(T) + R_b(T) \leq \tilde{I}_b(T) + \tilde{R}_b(T), \tag{E.221}$$

which, combined with (E.11), shows that

$$I_b(T) + R_b(T) = \tilde{I}_b(T) + \tilde{R}_b(T). \tag{E.222}$$

Moreover, one can subtract (E.218) from (E.219) to get

$$I_b^V(T) + R_b^V(T) = \tilde{I}_b^V(T) + \tilde{R}_b^V(T) \tag{E.223}$$

and then, using (E.16) alongside (E.221) and (E.222) shows

$$S_b(T) + S_b^V(T) = \tilde{S}_b(T) + \tilde{S}_b^V(T) \tag{E.224}$$

as required.

E.2.7 Lemma E.10

Note that for this lemma, it will be assumed that each $K_{ij}(t) \geq \tilde{K}_{ij}(t)$, rather than the inequality simply holding for their sums as before.

Lemma E.10 *Under the assumptions of Proposition E.1, suppose that the system of inequalities (E.10) - (E.16) holds for some $b \in \{1, ..., n\}$ and some $T > 0$. Suppose further that*

$$K_{ij}(t) \geq \tilde{K}_{ij}(t) \quad \forall i, j \in \{1, ..., n\}. \tag{E.225}$$

Then,

$$W_i(t) = \tilde{W}_i(t) \quad \forall i \in \{1, ..., n\} \quad \textit{and} \quad \forall t \in [0, T]. \tag{E.226}$$

Proof: By Lemma E.9, the system (E.17) - (E.19) must hold for b. Now, Equation (E.170) in the proof of Lemma E.6 shows that

$$I_b(T) + R_b(T) = \frac{S_b(0)}{N_b} \int_0^T \left[\sum_{k=1}^{n} K'_{bk}(s) \right] (N_b - W_b(s)) e^{-\sum_{j=1}^{n} K_{bk}(s)} ds. \tag{E.227}$$

Now, the equality (E.17) shows

$$I_b(T) + R_b(T) = \tilde{I}_b(T) + \tilde{R}_b(T) \tag{E.228}$$

and hence, after cancelling the non-zero $S_b(0)$ and N_b terms, (E.227) (and its tilde equivalent) shows that

$$\int_0^T \left[\sum_{k=1}^{n} K'_{bk}(s) \right] (N_b - W_b(s)) e^{-\sum_{k=1}^{n} K_{bk}(s)} ds \tag{E.229}$$

$$= \int_0^T \left[\sum_{k=1}^{n} \tilde{K}'_{bk}(s) \right] (N_b - \tilde{W}_b(s)) e^{-\sum_{k=1}^{n} \tilde{K}_{bk}(s)} ds. \tag{E.230}$$

Note that, from Lemma E.20, as $\Pi = \{1, ..., n\}$

$$\tilde{I}_k(s), I_k(s) > 0 \quad \forall k \in \{1, ..., n\} \quad \text{and} \quad s > 0. \tag{E.231}$$

Thus,

$$K'_{bk}(t) \geq \beta^1_{bk} I_j(t) > 0 \quad \forall t > 0. \tag{E.232}$$

In particular,

$$\left[\sum_{j=1}^{n} K'_{bk}(s) \right] e^{-\sum_{j=1}^{n} K_{bk}(s)} > 0 \quad \forall s \in [0, T]. \tag{E.233}$$

Moreover, by continuity of K'_{ik} (as continuous functions attain their bounds on closed intervals), there exists some $m > 0$ such that

$$\left[\sum_{k=1}^{n} K'_{bk}(s) \right] e^{-\sum_{k=1}^{n} K_{bk}(s)} > m \quad \forall s \in [0, T]. \tag{E.234}$$

Hence, as $W_b \leq \tilde{W}_b$

$$\int_0^T \left[\sum_{k=1}^n K'_{bk}(s)\right](N_b - W_b(s))e^{-\sum_{k=1}^n K_{bk}(s)}ds \tag{E.235}$$

$$= \int_0^T \left[\sum_{k=1}^n K'_{bk}(s)\right](N_b - \tilde{W}_b(s) + (\tilde{W}_b(s) - W_b(s))e^{-\sum_{k=1}^n K_{bk}(s)}ds \tag{E.236}$$

$$\geq \int_0^T \left[\sum_{k=1}^n K'_{bk}(s)\right](N_b - \tilde{W}_b(s))e^{-\sum_{k=1}^n K_{bk}(s)}ds + m\int_0^T \tilde{W}_b(s) - W_b(s)ds. \tag{E.237}$$

Finally, as $N - \tilde{W}_b$ is decreasing and for any $t \in [0,T]$,

$$\int_0^t \left[\sum_{k=1}^n K'_{bk}(s)\right]e^{-\sum_{k=1}^n K_{bk}(s)}ds \geq \int_0^t \left[\sum_{k=1}^n \tilde{K}'_{bk}(s)\right]e^{-\sum_{k=1}^n \tilde{K}_{bk}(s)}ds \tag{E.238}$$

one has, by Lemma E.4, setting

$$g(t) = 1 - e^{-\sum_{k=1}^n K_{bk}(t)}, \quad h(t) = e^{-\sum_{k=1}^n \tilde{K}_{bk}(s)} \tag{E.239}$$

and $f(t) = N_b - \tilde{W}_b(s)$,

$$\int_0^T \left[\sum_{k=1}^n K'_{bk}(s)\right](N_b - \tilde{W}_b(s))e^{-\sum_{k=1}^n K_{bk}(s)}ds \tag{E.240}$$

$$\geq \int_0^T \left[\sum_{k=1}^n \tilde{K}'_{bk}(s)\right](N_b - \tilde{W}_b(s))e^{-\sum_{k=1}^n \tilde{K}_{bk}(s)}ds \tag{E.241}$$

$$= \tilde{I}_b(T) + \tilde{R}_b(T) \tag{E.242}$$

and so, combining this with (E.237),

$$I_b(T) + R_b(T) \geq \tilde{I}_b(T) + \tilde{R}_b(T) + m\int_0^T \tilde{W}_b(s) - W_b(s)ds \geq \tilde{I}_b(T) + \tilde{R}_b(T) = I_b(T) + R_b(T). \tag{E.243}$$

Hence,

$$\int_0^T \tilde{W}_b(s) - W_b(s)ds = 0, \tag{E.244}$$

which by continuity means

$$W_b(t) = \tilde{W}_b(t) \quad \forall t \in [0,T] \tag{E.245}$$

Now, moreover, substituting this back into the equality given in (E.230) shows that

$$\int_0^T \left[\sum_{k=1}^n K'_{bk}(s)\right](N_b - W_b(s))e^{-\sum_{k=1}^n K_{bk}(s)}ds \tag{E.246}$$

$$= \int_0^T \left[\sum_{k=1}^n \tilde{K}'_{bk}(s)\right](N_b - W_b(s))e^{-\sum_{k=1}^n \tilde{K}_{bk}(s)}ds. \tag{E.247}$$

Hence, integrating by parts, this shows that

$$0 = (N_b - W_b(T))(e^{-\sum_{k=1}^n K_{bk}(T)} - e^{-\sum_{k=1}^n \tilde{K}_{bk}(T)})... \tag{E.248}$$

$$+ \int_0^T U_b(s)(e^{-\sum_{k=1}^n K_{bk}(s)} - e^{-\sum_{k=1}^n \tilde{K}_{bk}(s)})ds \tag{E.249}$$

Now,

$$\sum_{k=1}^n \tilde{K}_{bk}(s) \geq \sum_{k=1}^n K_{bk}(s) \quad \forall s \in [0,T] \tag{E.250}$$

and so, for equality, it is necessary that

$$(N_b - W_b(T))(e^{-\sum_{k=1}^n K_{bk}(T)} - e^{-\sum_{k=1}^n \tilde{K}_{bk}(T)}) = 0 \tag{E.251}$$

Thus, as it is assumed that $W_b(t) < N_b$ for all $t \geq 0$,

$$e^{-\sum_{k=1}^n K_{bk}(T)} - e^{-\sum_{k=1}^n \tilde{K}_{bk}(T)} = 0 \tag{E.252}$$

and hence, as $K_{bk}(T) \geq \tilde{K}_{bk}(T)$ for all $k \in \{1, ..., n\}$,

$$K_{bk}(T) = \tilde{K}_{bk}(T) \quad \forall k \in \{1, ..., n\} \tag{E.253}$$

Now, suppose that $K'_{bk}(T) > \tilde{K}'_{bk}(T)$ for some k. Then, by continuity and the fact that $T > 0$, it is necessary that there is some $\tau \in (0, T)$ such that

$$\int_{T-\tau}^T K'_{bk}(s)ds > \int_{T-\tau}^T \tilde{K}'_{bk}(s)ds \tag{E.254}$$

which means that

$$K_{bk}(T-\tau) < \tilde{K}_{bk}(T-\tau) \tag{E.255}$$

which is a contradiction to the definition of T. Thus, it is necessary that

$$K'_{bk}(T) \leq \tilde{K}'_{bk}(T) \quad \forall k \in \{1, ..., n\}. \tag{E.256}$$

Dividing (E.253) by β^1_{bk}/μ^1_k and (E.256) by β^1_{bk} shows that the inequality system (E.10) - (E.16) holds for each k (as Lemmas E.5 - E.7 hold for any group) and so, following Lemma E.9 and the previous work of this proof, it is necessary that

$$W_k(t) = \tilde{W}_k(t) \quad \forall t \in [0, T] \tag{E.257}$$

This holds for each k and hence the proof is complete.

E.2.8 Lemma E.11

Lemma E.11 *Define functions* Δ^f_i *to be*

$$\Delta^f_i(t) := f_i(T+t) - \tilde{f}_i(T+t) \quad \textit{for} \quad f \in \{S, I, R, S^V, I^V, R^V, W\} \tag{E.258}$$

and suppose that

$$\Delta^f_i(0) = 0 \quad \forall f \in \{S, I, R, S^V, I^V, R^V, W\}. \tag{E.259}$$

Suppose further that the $U_i(t)$ *are right-continuous step functions. Then, for* $t \in [0, \delta]$ *in the limit* $\delta \to 0$*, and for any* $x, y \in \Re$

$$\frac{x}{\mu^1_i}\Delta^R_i + \frac{y}{\mu^2_i}\Delta^{R^V}_i = \frac{t^3 S_i(T)(U_i(T) - \tilde{U}_i(T))}{6(N_i - W_i(T))}\left[x\sum_{j=1}^{n}(K'_{ij}(T)) - y\sum_{j=1}^{n}(L'_{ij}(T))\right] + O(\delta^4). \tag{E.260}$$

Proof: As the $U_i(t)$ are step functions, for sufficiently small δ, they are constant on the interval $[T, T+\delta]$, so this will be assumed. Note that, for any $i \in \{1, ..., n\}$ and any $t \geq 0$

$$\left|\frac{dS_i}{dt}(t)\right| \leq \left|\sum_{j=1}^{n} S_i(t)\beta^1_{ij} I_j(t)\right| + \left|\frac{S_i(t)}{N_i - W_i(t)} U_i(t)\right| \tag{E.261}$$

$$\leq \left|\sum_{j=1}^{n} N_i \beta^1_{ij} N_j\right| + + |1 \times U_i(t)| \tag{E.262}$$

$$\leq U_i(T) + C, \tag{E.263}$$

where the constant term, C, is independent of t and the vaccination policy. Note the second line follows from the fact that, as $W_i(t) < N_i$,

$$\frac{S_i(t)}{N_i - W_i(t)} = \frac{S_i(0)}{N_i} \exp\left[-\sum_{j=1}^{n} K_{ij}(t)\right] \leq 1. \tag{E.264}$$

Similarly, one can show (by increasing the constant C if necessary) that

$$\left|\frac{dS_i^V}{dt}(t)\right| \leq U_i(T) + C \tag{E.265}$$

$$\left|\frac{dI_i^V}{dt}(t)\right|, \left|\frac{dR_i^V}{dt}(t)\right|, \left|\frac{dI_i}{dt}(t)\right|, \left|\frac{dR_i}{dt}(t)\right| \leq C \tag{E.266}$$

$$\left|\frac{dW_i}{dt}(t)\right| \leq U_i(T). \tag{E.267}$$

Then, for $t \in (0, \delta)$ and $f \in \{S, I, R, S^V, I^V, R^V, W\}$

$$|f_i(T+t) - f_i(T)| = \left| \int_T^{T+t} \frac{df_i}{dt}(s)ds \right| \leq (C + U_i(T))\delta \tag{E.268}$$

so that, in particular

$$f_i(T+t) = f_i(T) + O(\delta) \quad \forall f \in \{S, I, R, S^V, I^V, R^V, W\}. \tag{E.269}$$

Now,

$$\frac{d\Delta_i^S}{dt} = -\sum_{j=1}^{n}(K_{ij}S_i - \tilde{K}_{ij}\tilde{S}_j) + \frac{S_iU_i}{N_i - W_i} - \frac{\tilde{S}_i\tilde{U}_i}{N_i - \tilde{W}_i}. \tag{E.270}$$

Using (E.259) and (E.269), this equation linearises to

$$\frac{d\Delta_i^S}{dt}(t) = \frac{S_i(T)(U_i(t+T) - \tilde{U}_i(t+T))}{N_i - W_i(T)} + O(\delta). \tag{E.271}$$

Noting that

$$U_i(t+T) - \tilde{U}_i(t+T) = U_i(T) - \tilde{U}_i(T) \quad \forall t \in [0, \delta] \tag{E.272}$$

this means that

$$\frac{d\Delta_i^S}{dt} = \frac{S_i(T)(U_i(T) - \tilde{U}_i(T))}{N_i - W_i(T)} + O(\delta) \tag{E.273}$$

and so (for $t < \delta$)

$$\Delta_i^S(t) = t\frac{S_i(T)(U_i(T) - \tilde{U}_i(T))}{N_i - W_i(T)} + O(\delta^2). \tag{E.274}$$

Now, one can linearise the equation for Δ_i^I. Note that

$$\frac{d\Delta_i^I}{dt} = \sum_{j=1}^{n}(K'_{ij}S_i - \tilde{K}'_{ij}\tilde{S}_i) + \mu_i^1(I_i - \tilde{I}_i) \tag{E.275}$$

and so, with

$$I_i(t+T) = I_i(T) + O(\delta) \tag{E.276}$$

and similar expressions for other variables,

$$\frac{d\Delta_i^I}{dt} = O(\delta) \Rightarrow \Delta_i^I(t) = O(\delta^2) \quad \text{for } t < \delta \tag{E.277}$$

Now, one can linearise in a different way. Note that

$$\tilde{I}_i(T+t) = I_i(T+t) + O(\delta^2) \quad \text{and} \quad \tilde{I}_i^V(T+t) = I_i^V(T+t) + O(\delta^2) \tag{E.278}$$

so

$$\tilde{K}'_{ij}(T+t) = K'_{ij}(T+t) + O(\delta^2). \tag{E.279}$$

Thus,

$$\frac{d\Delta_i^I}{dt}(T+t) = \sum_{j=1}^{n}(K'_{ij}(T+t)S_i(T+t) - \tilde{K}'_{ij}(T+t)\tilde{S}_i(T+t)) + \mu_i^1\Delta_i^I(T+t) + O(\delta^2) \tag{E.280}$$

$$= \Delta_i^S(t)\sum_{j=1}^{n}(K'_{ij}(T+t)) + O(\delta^2) \tag{E.281}$$

$$= t\frac{S_i(T)(U_i(T) - \tilde{U}_i(T))}{N_i - W_i(T)}\sum_{j=1}^{n}(K'_{ij}(T) + O(\delta)) + O(\delta^2) \tag{E.282}$$

$$= t\frac{S_i(T)(U_i(T) - \tilde{U}_i(T))}{N_i - W_i(T)}\sum_{j=1}^{n}(K'_{ij}(T)) + O(\delta^2) \tag{E.283}$$

and hence

$$\Delta_i^I = \frac{t^2}{2}\frac{S_i(T)(U_i(T) - \tilde{U}_i(T))}{N_i - W_i(T)}\sum_{j=1}^{n}(K'_{ij}(T)) + O(\delta^3). \tag{E.284}$$

Thus,

$$\frac{d\Delta_i^R}{dt} = \Delta_i^I\mu_i^1 \Rightarrow \Delta_i^R(t) = \frac{\mu_i^1 t^3}{6}\frac{S_i(T)(U_i(T) - \tilde{U}_i(T))}{N_i - W_i(T)}\sum_{j=1}^{n}(K'_{ij}(T)) + O(\delta^4). \tag{E.285}$$

Now, note that

$$\frac{d(\Delta_i^S + \Delta_i^{S^V})}{dt} = O(\delta) \tag{E.286}$$

as this derivative has no explicit dependence on U. Thus, in particular,

$$\Delta_i^S + \Delta_i^{S^V} = O(\delta^2) \tag{E.287}$$

and so

$$\Delta_i^{S^V} = -t\frac{S_i(T)(U_i(T) - \tilde{U}_i(T))}{N_i - W_i(T)} + O(\delta^2). \tag{E.288}$$

Then, as before (as the equation for $\frac{dI_i}{dt}$ is the same as that for $\frac{dI_i^V}{dt}$, but with S_i^V instead of S_i, μ_i^1 instead of μ_i^2 and K_{ij} instead of L_{ij})

$$\frac{d\Delta_i^{I^V}}{dt}(T+t) = -t\frac{S_i(T)(U_i(T)-\tilde{U}_i(T))}{N_i - W_i(T)}\sum_{j=1}^{n}(L'_{ij}(T)) + O(\delta^2), \tag{E.289}$$

which means

$$\Delta_i^{I^V} = -\frac{t^2}{2}\frac{S_i(T)(U_i(T)-\tilde{U}_i(T))}{N_i - W_i(T)}\sum_{j=1}^{n}(L'_{ij}(T)) + O(\delta^3) \tag{E.290}$$

and hence

$$\Delta_i^{R^V}(t) = -\frac{\mu_i^2 t^3}{6}\frac{S_i(T)(U_i(T)-\tilde{U}_i(T))}{N_i - W_i(T)}\sum_{j=1}^{n}(L'_{ij}(T)) + O(\delta^4). \tag{E.291}$$

Thus,

$$\frac{x}{\mu_i^1}\Delta_i^R + \frac{y}{\mu_i^2}\Delta_i^{R^V} = \frac{t^3 S_i(T)(U_i(T)-\tilde{U}_i(T))}{6(N_i - W_i(T))}\left[x\sum_{j=1}^{n}(K'_{ij}(T)) - y\sum_{j=1}^{n}(L'_{ij}(T))\right] + O(\delta^4) \tag{E.292}$$

as required.

E.2.9 Lemma E.12

Lemma E.12 *Suppose that*

$$T := \inf\left\{t : K_{ij}(t) \geq \tilde{K}_{ij} \quad or \quad L_{ij}(t) \geq \tilde{L}_{ij}(t) \quad for\ some\ i,j \in \{1,...,n\}\right\} \tag{E.293}$$

exists. Define functions Δ_i^f *to be*

$$\Delta_i^f(t) := f_i(T+t) - \tilde{f}_i(T+t) \quad for \quad f \in \{S, I, R, S^V, I^V, R^V, W\} \tag{E.294}$$

and suppose that

$$\Delta_i^f(0) = 0 \quad \forall f \in \{S, I, R, S^V, I^V, R^V, W\}. \tag{E.295}$$

Suppose further that the $U_i(t)$ *are right-continuous step functions,* $\Pi = \{1,...,n\}$ *and that*

$$\beta_{ij}^1 > \beta_{ij}^3 > 0 \quad and \quad I_i(0) > 0 \quad \forall i,j \in \{1,...,n\}. \tag{E.296}$$

Then,

$$\sum_{j=1}^{n} K_{ij}(t) \geq \sum_{j=1}^{n} \tilde{K}_{ij}(t) \quad \forall t \in [0, T+\delta] \tag{E.297}$$

and

$$\sum_{j=1}^{n} L_{ij}(t) \geq \sum_{j=1}^{n} \tilde{L}_{ij}(t) \quad \forall t \in [0, T+\delta], \tag{E.298}$$

for sufficiently small δ.

Proof: By Lemma E.11, with $x = \beta_{li}^1$ and $y = \beta_{li}^2$ for some $l \in \{1, ..., n\}$

$$\frac{\beta_{li}^1}{\mu_i^1}\Delta_i^R + \frac{\beta_{li}^2}{\mu_i^2}\Delta_i^{R^V} = \frac{t^3 S_i(T)(U_i(T) - \tilde{U}_i(T))}{6(N_i - W_i(T))}\left[\beta_{li}^1 \sum_{j=1}^{n}(K'_{ij}(T)) - \beta_{li}^2 \sum_{j=1}^{n}(L'_{ij}(T))\right] + O(\delta^4). \tag{E.299}$$

Now, as $\beta_{li}^1 \geq \beta_{li}^2$, $\beta_{li}^1 > 0$ and $K'_{ij}(t)$ and $L'_{ij}(t)$ are non-negative

$$\beta_{li}^1 \sum_{j=1}^{n}(K'_{ij}(T)) - \beta_{li}^2 \sum_{j=1}^{n}(L'_{ij}(T)) \leq 0 \Rightarrow \sum_{j=1}^{n}(K'_{ij}(T)) \leq \sum_{j=1}^{n}(L'_{ij}(T)). \tag{E.300}$$

Noting that

$$K'_{ij}(T) \geq L'_{ij}(T) \quad \forall j \in \{1, ..., n\}, \tag{E.301}$$

(E.300) requires

$$K'_{ij}(T) = L'_{ij}(T) \quad \forall j \in \{1, ..., n\} \tag{E.302}$$

which, from the definitions of K' and L' requires

$$\beta_{ij}^1 I_j(T) + \beta_{ij}^2 I_j^V(T) = \beta_{ij}^3 I_j(T) + \beta_{ij}^4 I_j^V(T). \tag{E.303}$$

Thus, as $I_j(T) > 0$ (as $\Pi \in \{1, ..., n\}$) and $\beta_{ij}^2 I_j^V(T) \geq \beta_{ij}^4 I_j^V(T)$, it is necessary that

$$\beta_{ij}^1 \leq \beta_{ij}^3, \tag{E.304}$$

which is a contradiction. Thus,

$$\beta_{ij}^1 \sum_{j=1}^{n}(K'_{ij}(T)) - \beta_{ij}^2 \sum_{j=1}^{n}(L'_{ij}(T)) > 0 \tag{E.305}$$

which means

$$S_i(T)U_i(T) < S_i(T)\tilde{U}_i(T) \Rightarrow \frac{\beta_{ij}^1}{\mu_i^1}\Delta_i^R + \frac{\beta_{ij}^2}{\mu_i^2}\Delta_i^{R^V} = -C\delta^3 + O(\delta^4) \tag{E.306}$$

for some positive constant C. Now, if

$$S_i(T)U_i(T) > S_i(T)\tilde{U}_i(T) \tag{E.307}$$

then, necessarily, $U_i(T) > \tilde{U}_i(T)$. Thus, as $\Delta_i^W(0) = 0$, one will have

$$W_i(T+t) > \tilde{W}_i(T+t) \tag{E.308}$$

for sufficiently small t, which is a contradiction. Moreover, if

$$S_i(T)U_i(T) = S_i(T)\tilde{U}_i(T) \quad \forall i \in \{1, ...n\} \tag{E.309}$$

then the vaccination policies are the same in the interval $[T, T+\delta]$, as for each i, either $S_i(T) = 0$ (in which case there is no more vaccination in group i so $U_i(T) = \tilde{U}_i(T) = 0$) or $U_i(T) = \tilde{U}_i(T)$. Thus, the disease trajectories are the same, which contradicts the definition of T, as then $K_{ij}(T+t) = \tilde{K}_{ij}(T+t)$ and $L_{ij}(T+t) = \tilde{L}_{ij}(T+t)$ for all $t \in [0, \delta]$.

Now, note that

$$\sum_{i=1}^{n} K_{li}(t) - \sum_{i=1}^{n} \tilde{K}_{li}(t) = \sum_{i=1}^{n} \left(\frac{\beta_{li}^1}{\mu_i^1} \Delta_i^R + \frac{\beta li^2}{\mu_i^2} \Delta_i^{R^V} \right) = -\sum_{i=1}^{n} E_i \delta^3 + O(\delta^4), \tag{E.310}$$

where $E_i > 0$ if $U_i(T) < \tilde{U}_i(T)$ and $E_i = 0$ otherwise. Thus, in particular

$$\sum_{i=1}^{n} E_i > 0 \tag{E.311}$$

and hence

$$\sum_{i=1}^{n} K_{li}(t) - \sum_{i=1}^{n} \tilde{K}_{li}(t) = -\sum_{i=1}^{n} E_i \delta^3 + O(\delta^4) < 0 \tag{E.312}$$

for sufficiently small δ. Thus,

$$\sum_{j=1}^{n} K_{ij}(t) \geq \sum_{j=1}^{n} \tilde{K}_{ij}(t) \quad \forall t \in [0, T+\delta] \tag{E.313}$$

and, by identical arguments (using $x = \beta_{li}^3$ and $y = \beta_{li}^4$ in Lemma E.11)

$$\sum_{j=1}^{n} L_{ij}(t) \geq \sum_{j=1}^{n} \tilde{L}_{ij}(t) \quad \forall t \in [0, T+\delta] \tag{E.314}$$

as required.

E.2.10 Lemma E.13

Lemma E.13 *Suppose that*

$$T := \inf\left\{t : K_{ij}(t) < \tilde{K}_{ij}(t) \quad or \quad L_{ij}(t) < \tilde{L}_{ij}(t) \quad for\ some\ i,j \in \{1,...,n\}\right\}. \tag{E.315}$$

Then, for any $\delta > 0$, there exists some $t \in (T, T+\delta)$ and some real parameters $0 \leq \kappa \leq \eta \leq 1$ such that

$$R_b(t) + \kappa R_b^V(t) < \tilde{R}_b(t) + \kappa \tilde{R}_b^V(t) \quad and \quad I_b(t) + \eta I_b^V(t) \leq \tilde{I}_b(t) + \eta \tilde{I}_b^V(t). \tag{E.316}$$

Proof: Firstly, note that by the definition of T, for each $\delta > 0$, there must exist $i, j \in \{1, ..., n\}$ and $t \in (0, \delta)$ such that

$$K_{ij}(T+t) < \tilde{K}_{ij}(T+t) \quad \text{or} \quad L_{ij}(T+t) < \tilde{L}_{ij}(T+t). \tag{E.317}$$

That is, there is some $b \in \{1, .., n\}$ such that

$$R_b(T+t) + \kappa R_b^V(T+t) < \tilde{R}_b(T+t) + \kappa \tilde{R}_b^V(T+t) \tag{E.318}$$

where

$$\kappa \leq \frac{\mu_b^1}{\mu_b^2}. \tag{E.319}$$

Note that

$$\mu_b^1 I_b(t) + \kappa \mu_b^2 I_b^V(t) = \frac{d}{dt}\left(R_b(t) + \kappa R_b^V(t)\right). \tag{E.320}$$

Now, define

$$\Delta_i^f(t) := f_i(T+t) - \tilde{f}_i(T+t) \quad \forall f \in \{I, I^V, R, R^V\} \tag{E.321}$$

and

$$\tau := \sup\{s \in [0,t] : \Delta_b^R(s) + \kappa \Delta_b^{R^V}(s) \geq 0\} \tag{E.322}$$

which exists as $\Delta_b^R(0) + \kappa \Delta_b^{R^V}(0) = 0$. Note that $\tau < t$ by (E.318). Note also that by continuity, it is necessary that

$$\Delta_b^R(\tau) + \kappa \Delta_b^{R^V}(\tau) = 0. \tag{E.323}$$

Now, by the mean value theorem (as $\Delta_b^R + \kappa \Delta_b^{R^V}$ is continuously differentiable), there exists an s in

the non-empty interval (τ, t) such that

$$\mu_1^b \Delta_b^I(s) + \kappa \mu_b^2 \Delta_b^{I^V}(s) = \frac{1}{t-\tau}\left[(\Delta_b^R(t) + \kappa \Delta_b^{R^V}(t)) - (\Delta_b^R(\tau) + \kappa \Delta_b^{R^V}(\tau))\right] \tag{E.324}$$

$$= \frac{1}{t-\tau}\left[\Delta_b^R(t) + \kappa \Delta_b^{R^V}(t)\right] \tag{E.325}$$

$$< 0 \tag{E.326}$$

while also

$$\Delta_b^R(s) + \kappa \Delta_b^{R^V}(s) < 0, \tag{E.327}$$

by definition of τ. Thus, defining $\eta := \kappa \frac{\mu_b^2}{\mu_b^1} \leq 1$,

$$\Delta_b^R(s) + \kappa \Delta_b^{R^V}(s) < 0 \quad \Delta_b^I(s) + \eta \Delta_b^{I^V}(s) \geq 0 \quad \text{and} \quad 0 \leq \kappa \leq \eta \leq 1 \tag{E.328}$$

as required.

E.2.11 Lemma E.14

Lemma E.14 *Consider two non-negative functions $A(t)$ and $B(t)$ such that $B(t)$ is non-decreasing and differentiable with a Lebesgue integrable derivative $B'(t)$ satisfying*

$$\int_0^t B'(s)ds = B(t) - B(0) \quad \forall t \geq 0. \tag{E.329}$$

Suppose further that for each $T \geq 0$, one can partition the interval $[0, T]$ into a finite number of subintervals $S_1^A, ..., S_m^A$ and $S_1^B, ..., S_k^B$ such that

$$s \in \bigcup_{i=1}^{m} S_i^A \Leftrightarrow A(s) > B'(s) \tag{E.330}$$

Then, there exists a unique function $\chi(t)$ for $t \geq 0$ such that

$$\chi(t) := \begin{cases} A(t) & \text{if} \quad \int_0^t \chi(s)ds < B(t) \\ \min(A(t), B'(t)) & \text{if} \quad \int_0^t \chi(s)ds \geq B(t) \end{cases} \tag{E.331}$$

Proof: χ can be constructed for each of the subintervals S_i^A and S_i^B. Note first that,

$$t \in S_i^B \Rightarrow B'(t) \geq A(t) \Rightarrow \chi(t) = A(t) \tag{E.332}$$

Now, suppose that $t \in S_i^A$ for some i. Then, as S_i^A is an interval, one can suppose $S_i^A = [c_i, d_i]$. Define

$$\tau := \inf\left(\left\{s \in S_i^A : B(s) \leq \int_0^{c_i} \chi(u)du + \int_{c_i}^{s} A(u)du\right\} \cup \{d_i\}\right). \tag{E.333}$$

If $\tau = d_i$, then one has (uniquely) $\chi(t) = A(t)$ in S_i^A. Otherwise, one has (again uniquely)

$$\chi(t) = A(t) \quad \forall t \in [c_i, \tau] \quad \text{and} \quad \chi(t) = B'(t) \quad \forall t \in [\tau, d_i] \tag{E.334}$$

Uniqueness can be demonstrated as follows. If $\chi(t) = B'(t)$ for some $t \in [c_i, \tau]$, then it is necessary (as $A(t) > B'(t)$ so $\chi(t) \neq A(t)$ in this case)

$$\int_0^t \chi(s) \geq B(t) \tag{E.335}$$

As $A(t) \geq B'(t)$ in S_i^A, so $\chi(t)$ is bounded by $A(t)$, the previous inequality can be extended to give

$$B(t) \leq \int_0^t \chi(s) \leq \int_0^{c_i} \chi(u)du + \int_{c_i}^{s} A(u)du \tag{E.336}$$

which contradicts the definition of τ. A similar argument stands to prove uniqueness in $[\tau, d_i]$.

Thus, χ is uniquely defined in each of the finite number of intervals and hence in $[0, T]$ for each T and hence, it is uniquely defined for all t as required.

E.3 Results on the SIR Equations

This section presents a variety of results on the SIR equations which are used in the proofs of the theorems in this paper. Many of them are well-known and widely used in the literature, but this appendix aims to provide a source of formal definitions and proofs of these results.

Before the results can be proved, it is necessary to establish two lemmas on differential equations.

E.3.1 Lemma E.15

Lemma E.15 *Suppose that $H(t)$ is a continuous non-negative $n \times n$ matrix for $t \geq 0$ and that $\boldsymbol{a} \in \Re^n$. Then, suppose that a function $\boldsymbol{u} : \Re \to \Re^n$ satisfies*

$$\boldsymbol{u}(t) \leq \boldsymbol{a} + \int_0^t H(s)\boldsymbol{u}(s)ds \quad \forall t \geq 0. \tag{E.337}$$

Then,

$$\boldsymbol{u}(t) \leq \left(1 + \int_0^t V(t,s)H(s)ds\right)\boldsymbol{a}, \tag{E.338}$$

where the matrix $V(t,s)$ *satisfies*

$$V(t,s) = I_n + \int_s^t H(k)V(k,s)dk \tag{E.339}$$

and I_n *is the* $n \times n$ *identity matrix.*

Proof: This theorem is a special case of the theorem proved in [505] where (in the notation of [505]), x, y and z have been replaced by t, s and k respectively, $G(t)$ has been set to be the identity matrix and x^0 has been set to zero.

E.3.2 Lemma E.16

Lemma E.16 *Consider a continuous, time-dependent, matrix* $A(t)$ *which satisfies*

$$A(t)_{ij} \geq 0 \quad \forall t \geq 0 \quad \textit{and} \quad \forall i \neq j \tag{E.340}$$

and a constant matrix B *that satisfies*

$$B_{ij} \geq 0 \quad \forall t \geq 0 \quad \textit{and} \quad \forall i \neq j. \tag{E.341}$$

Then, suppose that each element of $A(t)$ *is non-increasing with* t *and that*

$$A(t)_{ij} \geq B_{ij} \quad \forall t \geq 0 \quad \textit{and} \quad \forall i \neq j. \tag{E.342}$$

Moreover, define a non-negative initial condition $\boldsymbol{v}$ *and suppose that* $\boldsymbol{y}$ *and* $\boldsymbol{z}$ *solve the systems*

$$\frac{d\boldsymbol{y}}{dt} = A(t)\boldsymbol{y} \quad \textit{and} \quad \frac{d\boldsymbol{z}}{dt} = B\boldsymbol{z} \tag{E.343}$$

with

$$\boldsymbol{y}(0) = \boldsymbol{z}(0) = \boldsymbol{v} \geq \boldsymbol{0}. \tag{E.344}$$

Then,

$$\boldsymbol{y}(t) \geq \boldsymbol{z}(t) \geq \boldsymbol{0} \quad \forall t \geq 0. \tag{E.345}$$

Proof: To begin, define

$$\mu := \min_i \left(B_{ii} \right) \tag{E.346}$$

so that, defining

$$A^*(t) := A(t) + \mu I \quad \text{and} \quad B^* := B + \mu I, \tag{E.347}$$

where I is the identity matrix, A^* and B^* are non-negative matrices. Moreover, note that

$$\frac{d\boldsymbol{y}}{dt} + \mu \boldsymbol{y} = A^*(t)\boldsymbol{y} \tag{E.348}$$

and so

$$e^{-\mu t} \frac{d}{dt} \left(e^{\mu t} \boldsymbol{y} \right) = A^*(t)\boldsymbol{y}. \tag{E.349}$$

Thus, define

$$\boldsymbol{y}^*(t) := e^{\mu t} \boldsymbol{y}(t) \tag{E.350}$$

so

$$\frac{d\boldsymbol{y}^*}{dt} = A^*(t)\boldsymbol{y}^*. \tag{E.351}$$

Similarly, defining

$$\boldsymbol{z}^*(t) := e^{\mu t} \boldsymbol{z}(t) \tag{E.352}$$

gives

$$\frac{d\boldsymbol{z}^*}{dt} = B\boldsymbol{z}^* \tag{E.353}$$

while, moreover,

$$\boldsymbol{y}^* \geq \boldsymbol{z}^* \Leftrightarrow \boldsymbol{y} \geq \boldsymbol{z} \quad \text{and} \quad \boldsymbol{z}^* \geq \boldsymbol{0} \Leftrightarrow \boldsymbol{z} \geq \boldsymbol{0}. \tag{E.354}$$

Thus, it is simply necessary to prove that the results of this lemma hold when $A(t)$ and B are non-negative matrices.

Now, it is helpful to note that, as the off-diagonal entries of $A(t)$ and B are non-negative, the two differential systems are totally positive [506]. Thus, in particular, as $\boldsymbol{v}$ is non-negative,

$$\boldsymbol{y}(t), \boldsymbol{z}(t) \geq \boldsymbol{0} \quad \forall t \geq 0, \tag{E.355}$$

which proves one of the required inequalities. Now, one can also note that

$$\frac{d}{dt}\Big(\boldsymbol{y}-\boldsymbol{z}\Big)=A(t)\boldsymbol{y}-B\boldsymbol{z}. \tag{E.356}$$

As $A(t)$ is assumed to be non-negative, and $\boldsymbol{y}$ is non-negative,

$$\frac{d}{dt}\Big(\boldsymbol{y}-\boldsymbol{z}\Big)\geq B(\boldsymbol{y}-\boldsymbol{z}). \tag{E.357}$$

Defining $\boldsymbol{\zeta}:=\boldsymbol{z}-\boldsymbol{y}$ and integrating gives

$$\boldsymbol{\zeta}(t)\leq\int_0^t B(s)\boldsymbol{\zeta}(s)ds, \tag{E.358}$$

noting that $\boldsymbol{\zeta}=\mathbf{0}$. Hence, by Lemma E.15, one has

$$\boldsymbol{\zeta}(t)\leq\mathbf{0}\Rightarrow\boldsymbol{y}\geq\boldsymbol{z} \tag{E.359}$$

as required.

E.3.3 Lemma E.17

Lemma E.17 *Define the set of functions*

$$\mathcal{F}_i(t):=\Big\{S_i(t),I_i(t),R_i(t),S_i^V(t),I_i^V(t),R_i^V(t)\Big\}. \tag{E.360}$$

Then, for all $t\geq 0$ *and* $i\in\{1,...,n\}$,

$$0\leq f\leq N_i\quad\forall f\in\mathcal{F}_i(t). \tag{E.361}$$

Proof: Noting that

$$\sum_{f\in\mathcal{F}_i(t)} f=N_i, \tag{E.362}$$

it is simply necessary to show that (for each t and i)

$$f(t)\geq 0\quad\forall f(t)\in\mathcal{F}_i(t). \tag{E.363}$$

Now, note that

$$\frac{dS_i}{dt}=-\sum_{j=1}^{n}(\beta_{ij}^1 I_j+\beta_{ij}^2 I_j^V)S_i-\frac{U_i(t)S_i}{N_i-W_i(t)}, \tag{E.364}$$

which means

$$\frac{d}{dt}\left(S_i \exp\left[\sum_{j=1}^{n}(\frac{\beta_{ij}^1}{\mu_j^1}R_j + \frac{\beta_{ij}^2}{\mu_j^2}R_j^V) - \ln(N_i - W_i)\right]\right) = 0 \tag{E.365}$$

and hence (using the initial conditions)

$$S_i(t) = \frac{S_i(0)(N_i - W_i(t))}{N_i}\exp\left(\sum_{j=1}^{n}\left[\frac{\beta_{ij}^1}{\mu_j^1}R_j + \frac{\beta_{ij}^2}{\mu_j^2}R_j^V\right]\right). \tag{E.366}$$

As $W_i(t) \leq N_i$ by construction, this means that

$$S_i(t) \geq 0 \quad \text{as required.} \tag{E.367}$$

Now, note that

$$\frac{dS_i^V}{dt} = -\sum_{j=1}^{n}(\beta_{ij}^3 I_j + \beta_{ij}^4 I_j^V)S_i^V + \frac{U_i(t)S_i}{N_i - W_i(t)} \geq -\sum_{j=1}^{n}(\beta_{ij}^3 I_j + \beta_{ij}^4 I_j^V)S_i^V \tag{E.368}$$

so that

$$\frac{d}{dt}\left(S_i^V \exp\left[\sum_{j=1}^{n}\left(\frac{\beta_{ij}^3}{\mu_j^1}R_j + \frac{\beta_{ij}^4}{\mu_j^2}R_j^V\right)\right]\right) \geq 0, \tag{E.369}$$

which means (as $S_i^V(0) = 0$)

$$S_i^V(t)\exp\left[\sum_{j=1}^{n}\left(\frac{\beta_{ij}^3}{\mu_j^1}R_j(t) + \frac{\beta_{ij}^4}{\mu_j^2}R_j^V(t)\right)\right] \geq 0 \tag{E.370}$$

and hence

$$S_i^V(t) \geq 0 \quad \text{as required.} \tag{E.371}$$

Now, define the vector

$$\boldsymbol{y} := \begin{pmatrix} \boldsymbol{I} \\ \boldsymbol{I}^V \end{pmatrix} \tag{E.372}$$

Then, one can rewrite the equations for I_i and I_i^V in the form

$$\frac{d\boldsymbol{y}}{dt} = M(\boldsymbol{S}(t), \boldsymbol{S}^V(t))\boldsymbol{y} \tag{E.373}$$

for some matrix M, where, from the previous results

$$M_{ij} \geq 0 \quad \forall i \neq j. \tag{E.374}$$

Thus, from Lemma E.16,

$$\boldsymbol{y}(t) \geq \mathbf{0} \quad \forall t \geq 0. \tag{E.375}$$

Then,

$$\frac{dR_i}{dt} = \mu_i^1 I_i \geq 0 \quad \text{so} \quad R_i(t) \geq 0 \tag{E.376}$$

and similarly,

$$R_i^V(t) \geq 0 \tag{E.377}$$

and so the proof is complete.

E.3.4 Lemma E.18

Lemma E.18 *For each i,*

$$\lim_{t\to\infty}(I_i(t)) = \lim_{t\to\infty}(I_i^V(t)) = 0. \tag{E.378}$$

Proof: Firstly, suppose

$$\lim_{t\to\infty}(\inf\{I_i(s) : s \geq t\}) = Q, \tag{E.379}$$

noting this infimum exists as I_i is bounded below by 0, and the limit exists as the sequence of infima given $s \leq t$ is non-decreasing and bounded above by N_i. If $Q \neq 0$, there exists some $m > 0$ and some t such that for all $s \geq t$

$$I_i(s) \geq m \Rightarrow \frac{dR_i}{dt}(s) \geq m\mu_i^1 \Rightarrow R_i\left(t + \frac{2N_i}{m\mu_i^1}\right) > N_i \tag{E.380}$$

which contradicts Lemma E.17. Thus, $Q = 0$ and so there exists some sequence t_n such that

$$\lim_{n\to\infty}(t_n) = \infty \quad \text{and} \quad \lim_{n\to\infty}(I(t_n)) = 0. \tag{E.381}$$

Now note that $S_i(t)$ is non-increasing and bounded and that $R_i(t)$ and $(S_i^V(t) + I_i^V(t) + R_i^V(t))$ are non-decreasing and bounded. Thus, their limits as $t \to \infty$ must exist and be finite, so in particular

$$\lim_{t\to\infty}(I_i(t)) = \lim_{t\to\infty}(N_i - S_i(t) - R_i(t) - S_i^V(t) - I_i^V(t) - R_i^V(t)) \tag{E.382}$$

must exist. Thus, by (E.381), the only possible limit is 0 so

$$\lim_{t\to\infty}(I_i(t)) = Q = 0 \tag{E.383}$$

as required. By noting that $S_i(t) + S_i^V(t)$ is non-increasing and that $I_i(t) + R_i(t)$ and $R_i^V(t)$ are

non-decreasing, an identical argument shows that

$$\lim_{t\to\infty}(I_i^V(t)) = 0. \tag{E.384}$$

E.3.5 Lemma E.19

Lemma E.19 *Suppose that $I_i(t) > 0$ for some $t \geq 0$ and some $i \in \{1, ..., n\}$. Then,*

$$I_i(s) > 0 \quad \forall s > t. \tag{E.385}$$

An analogous result holds for $I_i^V(t)$.

Proof: Note that

$$\frac{dI_i}{dt} \geq -\mu_i^1 I_i \tag{E.386}$$

and so

$$\frac{d}{dt}\left(e^{\mu_i^1 t} I_i(t)\right) \geq 0 \tag{E.387}$$

which means, for any $s > t$

$$e^{\mu_i^1 s} I_i(s) \geq e^{\mu_i^1 t} I_i(t) \tag{E.388}$$

and hence

$$I_i(s) > 0 \tag{E.389}$$

as required. The same argument then works for $I_i^V(t)$ as well (with a μ_i^2 instead of a μ_i^1).

E.3.6 Lemma E.20

Lemma E.20 *Define*

$$\Pi := \left\{i : \exists t \geq 0 \quad s.t. \quad I_i(t) > 0 \quad or \quad I_i^V(t) > 0\right\}. \tag{E.390}$$

Moreover, define

$$\Pi^0 := \{i : I_i(0) > 0\} \tag{E.391}$$

and the n by n matrix M by

$$M_{ij} = S_i(0)\beta_{ij}^1. \tag{E.392}$$

Then, define the connected component C of Π^0 in M as follows. The index $i \in \{1, ..., n\}$ belongs to C if any only if there is some sequence $a_1, ..., a_k$ such that

$$a_j \in \{1, ..., n\} \quad \forall j \in \{1, ..., k\}, \tag{E.393}$$

$$M_{a_1,a_2} M_{a_2,a_3} ..., M_{a_{k-1}a_k} > 0 \tag{E.394}$$

and

$$a_1 = i \quad \textit{and} \quad a_k \in \Pi^0. \tag{E.395}$$

Then,

(a) $i \in C \Rightarrow I_i(t) > 0 \quad \forall t > 0.$

(b) $\Pi = C \cup \Pi^0.$

Thus, in particular,

$$i \in C \cup \Pi^0 = \Pi \Leftrightarrow I(t) > 0 \quad \forall t > 0. \tag{E.396}$$

Proof: (a): The proof will proceed by induction. For $k \geq 1$, define P^k is the set of elements of C that are connected to an element of Π^0 by a sequence of length at most k. Then, note that

$$P^k \subseteq P^{k+1} \quad \forall k \geq 1 \tag{E.397}$$

and

$$P^{n^2} = C \tag{E.398}$$

as there are n^2 elements in M. (Thus, if $i \in C$ then there must be a sequence of length at most n^2 connecting i with an element in Π^0 as any loops can be ignored.)

The inductive hypothesis is that

$$i \in P^k \Rightarrow I_i(t) > 0 \quad \forall t > 0. \tag{E.399}$$

The explanation of the base case will be left until the end of the proof. Suppose that this claim holds for some $k \geq 0$. If $P^{k+1} = P^k$, then

$$i \in P^{k+1} \Rightarrow i \in P^k \Rightarrow I_i(t) > 0 \quad \forall t > 0 \tag{E.400}$$

and so the inductive step is complete. Otherwise, consider any $i \in P^{k+1}/P^k$. Then, there exists some

j such that

$$M_{ij} > 0 \quad \text{and} \quad j \in P^k. \tag{E.401}$$

Thus, by continuity, for sufficiently small τ,

$$t < \tau \Rightarrow S_i(t)\beta_{ij}^1 > 0 \tag{E.402}$$

and indeed, by Boundedness Theorem, there exists some $\chi > 0$ such that

$$S_i(t)\beta_{ij}^1 > \chi \quad \forall t \in [0, \tau]. \tag{E.403}$$

Now, choose any $\epsilon \in [0, \tau]$. By Boundedness Theorem, $I_j(t)$ is bounded and achieves its maximum, θ_ϵ in the interval $[0, \epsilon]$. Moreover, $\theta_\epsilon > 0$ as $I_j(t) > 0$ in $(0, \epsilon)$ by assumption. Thus, by continuity, there exists some non-empty region $(\delta_\epsilon, \Delta_\epsilon)$ such that

$$t \in (\delta_\epsilon, \Delta_\epsilon) \Rightarrow I_j(t) > \frac{\theta_\epsilon}{2}. \tag{E.404}$$

Thus, in particular

$$\int_0^\epsilon S_i(t)\beta_{ij}^1 I_j(t)dt \geq \chi \int_{\delta_\epsilon}^{\Delta_\epsilon} I_j(t)dt \geq \frac{\chi\theta_\epsilon}{2}(\Delta_\epsilon - \delta_\epsilon) > 0. \tag{E.405}$$

Now, note that

$$\frac{dI_i}{dt} \geq S_i(t)\beta_{ij}^1 I_j(t) - \mu_i^1 I_i(t). \tag{E.406}$$

Suppose for a contradiction that $I_i(t) = 0$ for all $t \in [0, \epsilon]$. Then,

$$\frac{dI_i}{dt} \geq S_i(t)\beta_{ij}^1 I_j(t) \Rightarrow I_i(\epsilon) \geq I_i(0) + \frac{\chi M_\epsilon}{2}(\Delta_\epsilon - \delta_\epsilon) \tag{E.407}$$

and hence,

$$I_i(\epsilon) > 0, \tag{E.408}$$

which is a contradiction. Thus, there exists a $t \in [0, \epsilon]$ such that $I_i(t) > 0$ and hence, by Lemma E.19,

$$I_i(t) > 0 \quad \forall t \in [\epsilon, \infty). \tag{E.409}$$

Thus, as ϵ was any constant in the region $(0, \tau)$, and $\tau > 0$, this means that

$$I_i(t) > 0 \quad \forall t > 0 \tag{E.410}$$

as required.

Finally, note that the base case $k = 1$ can be proved in exactly the same way, except now $j \in \Pi^0$ (but this still means that $I_j(t) > 0$ for all $t > 0$ by Lemma E.19), and so **(a)** has been proved.

(b): The previous work has shown that

$$C \subseteq \Pi. \tag{E.411}$$

Hence, as clearly $\Pi^0 \subseteq \Pi$, this means that

$$C \cup \Pi^0 \subseteq \Pi \tag{E.412}$$

and so it suffices to prove that

$$\Pi \subseteq C \cup \Pi^0. \tag{E.413}$$

That is, it suffices to prove

$$i \notin C \cup \Pi^0 \Rightarrow I_i(t) = I_i^V(t) = 0 \quad \forall t \geq 0. \tag{E.414}$$

To check that this solution satisfies the equations, one notes that, in this case, if $i \notin C \cup \Pi^0$, then

$$\frac{dI_i}{dt} = \sum_{j=1}^{n} S_i(t)\beta_{ij}^1 I_j(t) + \sum_{j=1}^{n} S_i(t)\beta_{ij}^2 I_j^V(t) - \mu I_i(t) \tag{E.415}$$

$$= \sum_{j \in C \cup \Pi^0} S_i(t)\beta_{ij}^1 I_j(t) + \sum_{j \in C \cup \Pi^0} S_i(t)\beta_{ij}^2 I_j^V(t) \tag{E.416}$$

and, similarly,

$$\frac{dI_i^V}{dt} = \sum_{j \in C \cup \Pi^0} S_i^V(t)\beta_{ij}^3 I_j(t) + \sum_{j \in C \cup \Pi^0} S_i(t)\beta_{ij}^4 I_j^V(t), \tag{E.417}$$

as $I_j(t) = I_j^V(t) = 0$ for all $j \notin C \cup \Pi^0$.

Now, suppose that $i \notin C \cup \Pi^0$ and $j \in C \cup \Pi^0$. Then, by definition of C, this means that

$$M_{ij} = S_i(0)\beta_{ij}^1 = 0 \quad \forall j \in C \cup \Pi^0 \tag{E.418}$$

and hence, as S_i is non-increasing and non-negative

$$S_i(t)\beta_{ij}^1 = 0 \quad \forall j \in C \cup \Pi^0. \tag{E.419}$$

Now, as $\beta_{ij}^1 \geq \beta_{ij}^2 \geq 0$, this means that

$$S_i(t)\beta_{ij}^2 = 0 \quad \forall j \in C \cup \Pi^0 \tag{E.420}$$

so that

$$\sum_{j \in C \cup \Pi^0} S_i(t)\beta_{ij}^1 I_j(t) + \sum_{j \in C \cup \Pi^0} S_i(t)\beta_{ij}^2 I_j^V(t) = 0, \tag{E.421}$$

which means

$$\frac{dI_i}{dt} = 0 \quad \text{as required.} \tag{E.422}$$

Moreover, as $S_i^V(0) = 0$, it is necessary that

$$(S_i(0) + S_i^V(0))\beta_{ij}^1 = 0 \quad \forall j \in C \cup \Pi^0 \tag{E.423}$$

so, as $(S_i + S_i^V)\beta_{ij}^1$ is non-increasing and non-negative

$$(S_i(t) + S_i^V(t))\beta_{ij}^1 = 0 \quad \forall j \in C \cup \Pi^0 \tag{E.424}$$

and hence, as $S_i(t)$ is non-negative

$$(S_i^V(t))\beta_{ij}^1 = 0 \quad \forall j \in C \cup \Pi^0. \tag{E.425}$$

Thus, as $\beta_{ij}^1 \geq \beta_{ij}^3 \geq \beta_{ij}^4 \geq 0$, one has

$$\sum_{j \in C \cup \Pi^0} S_i^V(t)\beta_{ij}^3 I_j(t) + \sum_{j \in C \cup \Pi^0} S_i(t)\beta_{ij}^4 I_j^V(t) = 0 \tag{E.426}$$

and hence

$$\frac{dI_i^V}{dt} = 0 \quad \text{as required.} \tag{E.427}$$

Then, one can separately solve the system for all $j \in C \cup \Pi^0$ as the equations will now be independent of any indices $i \notin C \cup \Pi^0$ (as they only depend on these indices via the I_i and I_i^V terms, which are identically zero). Thus, by the uniqueness of solution, one must have

$$i \in C \cup \Pi^0 \Rightarrow I_i(t) = I_i^V(t) = 0 \quad \forall t \geq 0 \tag{E.428}$$

and hence part **(b)** is proved. Thus, the lemma has been proved.

E.3.7 Lemma E.21

Lemma E.21 *Consider a set* $C = [a_1, b_1] \times [a_2, b_2] \times ... \times [a_n, b_n]$ *that is a Cartesian product of real intervals. Suppose that* $f : \Re^n \to \Re$ *is differentiable with bounded derivatives in* C*. Then,* f *is Lipschitz continuous on* C *- that is, there exists some* $L > 0$ *such that*

$$|f(\boldsymbol{x}) - f(\boldsymbol{y})| \leq L \sum_{i=1}^{n} |x_i - y_i| \quad \forall \boldsymbol{x}, \boldsymbol{y} \in C. \tag{E.429}$$

Proof: Note that, by assumption, for each i,

$$\frac{\partial f}{\partial x_i} \quad \text{is bounded in } C, \tag{E.430}$$

so define the global bound for all i to be M. Choose some $\boldsymbol{x}, \boldsymbol{y} \in C$. Define the points $\boldsymbol{p}^k \in C$ for $k = 0, 1, ..., n$ by

$$p_i^k = \begin{cases} y_i & \text{if } i \leq k \\ x_i & \text{otherwise} \end{cases} \tag{E.431}$$

and define the curve γ_i to be the straight line joining the point $\boldsymbol{p}^{i-1}$ to the point $\boldsymbol{p}^i$. As C is a product of intervals, the γ_i lie entirely in C.

Define Γ to be the union of the curves γ_i, so that Γ joins $\boldsymbol{p}^0 = \boldsymbol{x}$ to $\boldsymbol{p}^n = \boldsymbol{y}$. Then

$$|f(\boldsymbol{x}) - f(\boldsymbol{y})| = \left| \int_\Gamma \nabla f \cdot d\boldsymbol{x} \right| \tag{E.432}$$

$$= \left| \sum_{i=1}^{n} \int_{\gamma_i} \nabla f \cdot d\boldsymbol{x} \right| \tag{E.433}$$

$$= \left| \sum_{i=1}^{n} \int_{s=x_i}^{s=y_i} \frac{\partial f}{\partial x_i}(\boldsymbol{p}^{i-1} + (s - x_i)\boldsymbol{e}_i) ds \right| \tag{E.434}$$

$$\leq \sum_{i=1}^{n} \left| \int_{s=x_i}^{s=y_i} \frac{\partial f}{\partial x_i}(\boldsymbol{p}^{i-1} + (s - x_i)\boldsymbol{e}_i) ds \right| \tag{E.435}$$

$$\leq \sum_{i=1}^{n} \sup_{\boldsymbol{s} \in C} \left| \frac{\partial f}{\partial x_i}(\boldsymbol{s}) \right| |y_i - x_i| \tag{E.436}$$

$$\leq M \sum_{i=1}^{n} |y_i - x_i| \tag{E.437}$$

where $\boldsymbol{e}_i$ is the ith canonical basis vector. Hence, the required Lipschitz continuity holds with $M = L$.

E.3.8 Lemma E.22

Lemma E.22 *Define the set of functions*

$$\mathcal{F} := \left\{ S_i(t;\epsilon), I_i(t;\epsilon), R_i(t;\epsilon), S_i^V(t;\epsilon), I_i^V(t;\epsilon), R_i^V(t;\epsilon) : i \in \{1, ..., n\}, \quad \epsilon, t \geq 0 \right\}, \tag{E.438}$$

where for each fixed ϵ, these functions solve the model equations with parameters

$$\mathcal{P} = \left\{ \beta_{ij}^{\alpha}(\epsilon), \mu_i^{\gamma}(\epsilon) : i, j \in \{1, ..., n\}, \quad \alpha \in \{1, 2, 3, 4\}, \quad \gamma \in \{1, 2\} \quad \textit{and} \quad \epsilon \geq 0 \right\}, \tag{E.439}$$

initial conditions

$$\mathcal{I} = \left\{ f(0;\epsilon) : i \in \{1, ..., n\}, \quad f \in \mathcal{F} \quad \textit{and} \quad \epsilon \geq 0 \right\} \tag{E.440}$$

and vaccination policy $\boldsymbol{U}(t;\epsilon)$. Suppose that

$$|p(\epsilon) - p(0)| \leq \epsilon \quad \forall p \in \mathcal{P}, \tag{E.441}$$

$$|f_i(0;\epsilon) - f_i(0;0)| \leq \epsilon \quad \forall f \in \mathcal{F} \tag{E.442}$$

and that

$$|W_i(t,\epsilon) - W_i(t,0)| < \epsilon \quad \forall t \geq 0. \tag{E.443}$$

Moreover, suppose that for each $i \in \{1, ..., n\}$ *and* $\epsilon \geq 0$,

$$U_i(s;\epsilon) \geq 0 \quad \textit{and} \quad \int_0^t U_i(s;\epsilon)ds \leq N_i \quad \forall t \geq 0. \tag{E.444}$$

Then, for each $\delta > 0$ *and each* $T > 0$ *there exists some* $\eta > 0$ *(that may depend on* T *and* δ*) such that*

$$\epsilon \in (0, \eta) \Rightarrow |f(t;\epsilon) - f(t;0)| < \delta \quad \forall f \in \mathcal{F} \quad \textit{and} \quad \forall t \in [0, T] \tag{E.445}$$

Proof: To begin, it is helpful to note that, by Lemma E.17,

$$f(t;\epsilon) \in [0, \max(N_i)] \quad \forall f \in \mathcal{F} \quad \text{and} \quad t \geq 0 \tag{E.446}$$

and that, by assumption on the feasibility of U_i

$$W(t;\epsilon) \in [0, \max(N_i)] \quad \forall t \geq 0. \tag{E.447}$$

Moreover, as the parameter values converge, it can be assumed that

$$p(\epsilon) \in [\alpha, \beta] \quad \forall \epsilon \geq 0 \quad \text{and} \quad p \in \mathcal{P} \tag{E.448}$$

for some $\alpha, \beta \geq 0$. Moreover, it can be assumed that, as each $\mu_i^a > 0$, there is some $\gamma > 0$ such that $\mu_i^a(\epsilon) > \gamma$ for all $\epsilon \geq 0$.

However, there is no condition on the maximal difference (at a point) between $U_i(t;\epsilon)$ and $U_i(t;0)$. To avoid this problem, it is helpful to consider the variable $S_i^O := S_i + S_i^V$ instead of S_i^V. Then, the equations for S_i and S_i^O can be written as

$$S_i(t;\epsilon) = \frac{S_i(0)(N_i - W_i(t;\epsilon))}{N_i} \exp\left[-\sum_{j=1}^{n}\left(\frac{\beta_{ij}^1(\epsilon)R_j(t;\epsilon)}{\mu_j^1(\epsilon)} + \frac{\beta_{ij}^2(\epsilon)R_j^V(t;\epsilon)}{\mu_j^2(\epsilon)}\right)\right] \tag{E.449}$$

$$\frac{dS_i^O(t;\epsilon)}{dt} = -\sum_{j=1}^{n}\left[\left(\beta_{ij}^1(\epsilon)I_j(t;\epsilon) + \beta_{ij}^2(\epsilon)I_j^V(t;\epsilon)\right)S_i(t;\epsilon)\right] \tag{E.450}$$

$$-\sum_{j=1}^{n}\left[\left(\beta_{ij}^3(\epsilon)I_j(t;\epsilon) + \beta_{ij}^4(\epsilon)I_j^V(t;\epsilon)\right)(S_i^O(t;\epsilon) - S_i(t;\epsilon))\right]. \tag{E.451}$$

Then, one can define

$$\boldsymbol{v} := (\boldsymbol{S}^O, \boldsymbol{I}, \boldsymbol{I}^V, \boldsymbol{R}, \boldsymbol{R}^V)^T \tag{E.452}$$

and $\boldsymbol{p}(\epsilon)$ to be a vector of the elements of $\mathcal{P}$ at some $\epsilon \geq 0$. Then, (substituting for $\boldsymbol{S}$), the model

equations can be written in the form

$$\frac{d\boldsymbol{v}(t;\epsilon)}{dt} = \boldsymbol{\Phi}(\boldsymbol{v}(t;\epsilon), \boldsymbol{W}(t;\epsilon), \boldsymbol{p}(\epsilon)) \tag{E.453}$$

where Φ is a smooth function. Thus, from Lemma E.21, there exists some constant L such that, for $\boldsymbol{v}$, $\boldsymbol{W}$ and $\boldsymbol{p}$ within the closed bounded feasible set of values and any $j \in \{1, ..., 5n\}$,

$$|\boldsymbol{\Phi}(\boldsymbol{v}, \boldsymbol{W}, \boldsymbol{p})_j - \boldsymbol{\Phi}(\boldsymbol{v}^*, \boldsymbol{W}^*, \boldsymbol{p}^*)_j| \leq L \left(\sum_{i=1}^{5n} |v_i - v_i^*| + \sum_{i=1}^{n} |W_i - W_i^*| + \sum_{i=1}^{4n^2+2n} |p_i - p_i^*| \right). \tag{E.454}$$

Thus, in particular, this means that

$$\begin{aligned}
\frac{d}{dt}\Big(|v_j(t;\epsilon) - v_j(t;0)| \Big) &\leq \left| \frac{d}{dt}\Big(v_j(t;\epsilon) - v_j(t;0) \Big) \right| && \text{(E.455)} \\
&\leq \Big| \boldsymbol{\Phi}(\boldsymbol{v}(t;\epsilon), \boldsymbol{W}(t;\epsilon), \boldsymbol{p}(\epsilon))_i - \boldsymbol{\Phi}(\boldsymbol{v}(t;0), \boldsymbol{W}(t;0), \boldsymbol{p}(\epsilon))_i \Big| && \text{(E.456)} \\
&\leq L\Big(\sum_{i=1}^{5n} |v_i(t;\epsilon) - v_i(t;0)| + \sum_{i=1}^{n} |W_i(t;\epsilon) - W_i(t;0)| ... && \text{(E.457)} \\
&+ \sum_{i=1}^{4n^2+2n} |p_i - p_i^*| \Big). && \text{(E.458)}
\end{aligned}$$

Now, adding these $5n$ inequalities together, one seems that

$$\begin{aligned}
&\frac{d}{dt}\Big(\sum_{i=1}^{5n} |v_i(t;\epsilon) - v_i(t;0)| \Big) \\
&\leq 5nL \left(\sum_{i=1}^{5n} |v_i(t;\epsilon) - v_i(t;0)| + \sum_{i=1}^{n} |W_i(t;\epsilon) - W_i(t;0)| + \sum_{i=1}^{4n^2+2n} |p_i - p_i^*| \right) && \text{(E.459)}
\end{aligned}$$

and hence

$$\begin{aligned}
&\sum_{i=1}^{5n} \left[\frac{d}{dt} \left(e^{-5nLt} |v_i(t;\epsilon) - v_i(t;0)| \right) \right] && \text{(E.460)} \\
&\leq 5nLe^{-5nLt} \left(\sum_{i=1}^{n} |W_i(t;\epsilon) - W_i(t;0)| + \sum_{i=1}^{4n^2+2n} |p_i - p_i^*| \right) && \text{(E.461)} \\
&\leq (15n^2 + 20n^3) L\epsilon e^{-5nLt}. && \text{(E.462)}
\end{aligned}$$

Thus, integrating (and using the fact that the initial conditions differ by at most ϵ)

$$e^{-5nLt}\sum_{i=1}^{5n}|v_i(t;\epsilon)-v_i(t;0)| \leq \sum_{i=1}^{5n}|v_i(0;\epsilon)-v_i(0;0)| + (3n+4n^2)\epsilon(1-e^{-5nLt}) \tag{E.463}$$

$$\leq 5n\epsilon + (3n+4n^2)\epsilon(1-e^{-5nLt}) \tag{E.464}$$

which means

$$\sum_{i=1}^{5n}|v_i(t;\epsilon)-v_i(t;0)| \leq 5n\epsilon e^{5nLt} + (3n+4n^2)\epsilon(e^{5nLt}-1) \tag{E.465}$$

and hence, for each $i \in \{1,...,5n\}$

$$|v_i(t;\epsilon)-v_i(t;0)| \leq 5n\epsilon e^{5nLt} + (3n+4n^2)\epsilon(e^{5nLt}-1). \tag{E.466}$$

The right-hand side is non-decreasing in t (as $L>0$) so, taking

$$\epsilon < \frac{\delta}{5ne^{5nLt}+(3n+4n^2)(e^{5nLt}-1)} \tag{E.467}$$

ensures that the required inequalities hold for $\boldsymbol{I}$, $\boldsymbol{I}^V$, $\boldsymbol{R}$ and $\boldsymbol{R}^V$ for all $s \leq t$. Now, note also that $S_i(t;\epsilon)$ is a smooth function of $W_i(t;\epsilon)$, $\boldsymbol{v}(\epsilon)$, $S_i(0;\epsilon)$ and $\boldsymbol{p}$ so that there exists an L' such that

$$|S_i(t;\epsilon)-S_i(0;\epsilon)| < L'\Bigg(\sum_{i=1}^{5n}|v_i-v_i^*| + \sum_{i=1}^{n}|W_i-W_i^*|... \tag{E.468}$$

$$+\sum_{i=1}^{4n^2+2n}|p_i-p_i^*| + |S_i(0;\epsilon)-S_i(0;0)|\Bigg) \tag{E.469}$$

$$< L'\epsilon\Big[5ne^{5nLt}+(3n+4n^2)(e^{5nLt}-1)+(3n+4n^2)+1\Big] \tag{E.470}$$

$$:= \chi(t)\epsilon \tag{E.471}$$

and so, as $\chi(t)$ is non-decreasing in t, taking

$$\epsilon < \frac{\delta}{\chi(t)} \tag{E.472}$$

gives the required inequalities for $\boldsymbol{S}$ for all times $s \leq t$. Finally, note that

$$|S_i^V(t;\epsilon)-S_i^V(t;0)| = |S_i^O(t;\epsilon)-S_i^O(t;0)-S_i(t;\epsilon)+S_i(t;0)| \tag{E.473}$$

$$\leq |S_i^O(t;\epsilon)-S_i^O(t;0)| + |S_i(t;\epsilon)-S_i(t;0)| \tag{E.474}$$

$$\leq +5n\epsilon e^{5nLt} + (3n+4n^2)\epsilon(e^{5nLt}-1) + \epsilon\chi(t) \tag{E.475}$$

and so, as the right-hand side is increasing in t, taking

$$\epsilon < \frac{\delta}{5ne^{5nLt} + (3n + 4n^2)(e^{5nLt} - 1) + \chi(t)} \tag{E.476}$$

gives the required inequalities for $\boldsymbol{S}^V$ for all times $s \leq t$ and hence completes the proof.

Appendix - Paper VI

F.1 Proof of Theorem 8.1

Note that, throughout this section, the tilde will be removed from the rescaled p_i terms to reduce notation (and hence all p_i terms used here will be assumed to be rescaled).

Theorem 8.1 *Suppose that for all* $\epsilon > 0$

$$N_1(\epsilon) = \epsilon, \quad S_1(0;\epsilon) = \epsilon\sigma, \quad I_1(0;\epsilon) = (1-\sigma)\epsilon \quad \textit{and} \quad p_1(\epsilon) = \frac{1}{\epsilon} \tag{F.1}$$

for some $\sigma \in (0,1)$. *Suppose that all other parameter values and initial conditions are independent of* ϵ.

Consider any vaccination policy given by $\boldsymbol{U}(t;\epsilon)$ *and suppose that there exists fixed* $\alpha, \tau, w > 0$ *such that*

$$W_1(\tau;\epsilon) < \alpha\epsilon \quad \textit{and} \quad \sum_{i=1}^{n} W_i(\tau;\epsilon) > w \quad \forall \epsilon > 0. \tag{F.2}$$

Define a new policy $\tilde{\boldsymbol{U}}(t;\epsilon)$

$$\tilde{U}_1(t;\epsilon) = \begin{cases} \sum_{i=1}^{n} U_i(t) & \textit{if } \sum_{i=1}^{n} W_i(t;\epsilon) \leq \epsilon \\ 0 & \textit{otherwise} \end{cases} \tag{F.3}$$

and, for $i \neq 1$

$$\tilde{U}_i(t;\epsilon) = \begin{cases} 0 & \textit{if } \sum_{i=1}^{n} W_i(t;\epsilon) \leq \epsilon \\ U_i(t;\epsilon) & \textit{otherwise} \end{cases}. \tag{F.4}$$

Suppose that for each $i \in \{1,...,n\}$ *and* $t \geq 0$,

$$|W_i(t;0) - W_i(t;\epsilon)| \leq \epsilon. \tag{F.5}$$

Define

$$\Pi(\epsilon) := \{i : \exists t \geq 0 \quad s.t. \quad I_i(t;\epsilon) > 0\} \tag{F.6}$$

and suppose that $\Pi(\epsilon) = \{1,...,n\}$ *for any* $\epsilon > 0$ *and that* $\Pi(0) = \{2,...,n\}$. *Finally, suppose that there exists an* $i \in \{2,...,n\}$ *such that*

$$\beta_{1i}^{1} > \beta_{1i}^{3} \geq 0. \tag{F.7}$$

Then, the policy $\tilde{\boldsymbol{U}}$ is feasible and for sufficiently small ϵ,

$$H(\boldsymbol{U}(t;\epsilon)) > H(\tilde{\boldsymbol{U}}(t;\epsilon)). \tag{F.8}$$

Proof: It is first important to prove that the $\tilde{\boldsymbol{U}}$ is feasible. Firstly,

$$\sum_{i=1}^{n} \tilde{U}_i(t;\epsilon) \leq \sum_{i=1}^{n} U_i(t;\epsilon) \tag{F.9}$$

which, as $\boldsymbol{U}$ is feasible, means that the supply and rate constraints are satisfied. Moreover, as each $U_i(t;\epsilon) \geq 0$,

$$\tilde{U}_i(t;\epsilon) \geq 0 \quad \forall i \in \{1,...,n\}. \tag{F.10}$$

Also, for $i \neq 1$,

$$\tilde{U}_i(t;\epsilon) \leq U_i(t;\epsilon) \Rightarrow \tilde{W}_i(t;\epsilon) \leq W_i(t;\epsilon) \leq N_i. \tag{F.11}$$

Finally, define

$$t^* := \sup\{t : \sum_{i=1}^{n} W_i(t;\epsilon) \leq \epsilon\} \in \Re \cup \{\infty\} \tag{F.12}$$

and then

$$U_1(t;\epsilon) \leq \int_0^{t^*} \sum_{i=1}^{n} U_i(s;\epsilon) ds \leq \epsilon = N_1 \tag{F.13}$$

as required.

Define $S_i(t;\epsilon)$ to be the number of susceptibles given the parameters $N_1(\epsilon)$, $S_1(0;\epsilon)$ and $I_1(0;\epsilon)$ and the vaccination policy $\boldsymbol{U}(t;\epsilon)$, and define $\tilde{S}_i(t;\epsilon)$ to be the number of susceptibles given the parameters $N_1(\epsilon)$, $S_1(0;\epsilon)$ and $I_1(0;\epsilon)$ and the vaccination policy $\tilde{\boldsymbol{U}}(t;\epsilon)$. Use similar definitions for the other variables in the model.

F.1.1 Proposition F.1

Proposition F.1 *For each $t \geq 0$ and $i \in \{1,...,n\}$,*

$$|\tilde{W}_i(t;\epsilon) - \tilde{W}_i(0;\epsilon)| \leq 2\epsilon. \tag{F.14}$$

Proof: Firstly, note that

$$\tilde{W}_1(t;\epsilon) \leq \epsilon \tag{F.15}$$

so

$$|\tilde{W}_1(t;\epsilon) - \tilde{W}_1(0;\epsilon)| \leq \epsilon. \tag{F.16}$$

Now, suppose that $i \neq 1$. Then, for each $\epsilon, t \geq 0$, with t^* defined as in (F.12),

$$|W_i(t;\epsilon) - \tilde{W}_i(t;\epsilon)| = \left| \int_0^t U_i(s)ds - \int_{t^*}^{\max(t,t^*)} U_i(s)ds \right|. \tag{F.17}$$

If $t < t^*$, then

$$|W_i(t;\epsilon) - \tilde{W}_i(t;\epsilon)| \leq \left| \int_0^t U_i(s)ds \right| \leq \left| \int_0^{t^*} U_i(s)ds \right| \leq \epsilon \tag{F.18}$$

while if $t \geq t^*$, then,

$$|W_i(t;\epsilon) - \tilde{W}_i(t;\epsilon)| = \left| \int_0^{t^*} U_i(s)ds \right| \leq \epsilon. \tag{F.19}$$

Thus, noting

$$W_i(t;0) = \tilde{W}_i(t;0) \tag{F.20}$$

and using (F.5),

$$|\tilde{W}_i(t;\epsilon) - \tilde{W}_i(0;\epsilon)| \leq |\tilde{W}_i(t;\epsilon) - W_i(t;\epsilon)| + |W_i(t;\epsilon) - W_i(t;0)| \leq \epsilon + \epsilon = 2\epsilon \tag{F.21}$$

as required.

F.1.2 Proposition F.2

Next, it is helpful to consider the continuous dependence of the final size of the epidemic on the initial conditions and the vaccination policy. A weaker result is proved in [50] (and an almost identical version is referenced in this appendix as Lemma F.6). However, that result only holds for finite times, and extending it to hold for the final sizes requires a significant amount of extra work.

Proposition F.2 *Suppose that the U_i have uniformly bounded support for each $\epsilon > 0$. Moreover, for each of the model variables, f_i, suppose that*

$$|f_i(0;\epsilon) - f_i(0;0)| < K\epsilon \tag{F.22}$$

for some constant K and that

$$|W_i(t;\epsilon) - W_i(t;0)| < K'\epsilon \tag{F.23}$$

for some constant K'. Finally, suppose all parameters are independent of ϵ with the exception that

$N_1(\epsilon) = \epsilon$. *Then, for each* $\delta > 0$, *there exists some* $\Delta > 0$ *such that*

$$\epsilon \in [0, \Delta] \Rightarrow |f_i(\infty; \epsilon) - f_i(\infty; 0)| < \delta \quad \forall f \in \{I_i(t;\epsilon), I_i^V(t;\epsilon), R_i(t;\epsilon), R_i^V(t;\epsilon)\}. \tag{F.24}$$

Note that this holds both in the case of Theorem 8.1 (where $N_1 \to 0$, $\Pi(\epsilon) = \{1, ..., n\}$ *for* $\epsilon > 0$ *and* $\Pi(0) = \{2, ..., n\}$*) or, in the case where each* N_i *is independent of* ϵ *(by adding a disconnected group of size* ϵ*).*

Proof: Choose any $\delta > 0$. Now, it is possible to write the system for $\boldsymbol{I}$ and $\boldsymbol{I}^V$ in the form

$$\frac{d\boldsymbol{J}(t;\epsilon)}{dt} = \boldsymbol{M}(t;\epsilon)\boldsymbol{J}(t;\epsilon), \tag{F.25}$$

where $\boldsymbol{M}$ depends on the values of $\boldsymbol{S}(t;\epsilon)$, $\boldsymbol{S}^V(t;\epsilon)$, β_{ij}^{α} and μ_i^{α} and

$$\boldsymbol{J} = \begin{pmatrix} \boldsymbol{I} \\ \boldsymbol{I}^V \end{pmatrix}. \tag{F.26}$$

Hence, in particular, by using Proposition F.1 and Lemma F.6 for any fixed $t \geq 0$,

$$\lim_{\epsilon \to 0}(M(t;\epsilon)) = M(t;0). \tag{F.27}$$

Moreover, if the support of each $U_i(t;\epsilon)$ is bounded by t_U (which exists by assumption), then for $t > t_U$, each $S_i(t;\epsilon)$ and $S_i^V(t;\epsilon)$ is non-increasing in t and so $\boldsymbol{M}(t;\epsilon)$ is non-increasing. As it is bounded below, it therefore must converge to some matrix $\boldsymbol{M}(\infty;\epsilon)$, and, for $t > t_U$,

$$\frac{d\boldsymbol{J}(t;\epsilon)}{dt} \geq \boldsymbol{M}(\infty;\epsilon)\boldsymbol{J}(t;\epsilon). \tag{F.28}$$

Hence, by Lemma E.16,

$$\boldsymbol{J}(t_U + t';\epsilon) \geq e^{t'\boldsymbol{M}(\infty;\epsilon)}\boldsymbol{J}(t_U;\epsilon). \tag{F.29}$$

Moreover, by Lemma E.18,

$$\lim_{t' \to \infty}\left(\boldsymbol{J}(t_U + t';\epsilon)\right) = 0 \tag{F.30}$$

and hence (by non-negativity)

$$\lim_{t' \to \infty}\left(e^{t'\boldsymbol{M}(\infty;\epsilon)}\boldsymbol{J}(t_U;\epsilon)\right) = 0. \tag{F.31}$$

Now, define

$$\max_{i,\alpha}(\mu_i^{\alpha}) := \mu_{\max} \tag{F.32}$$

and then define

$$\boldsymbol{\mathcal{M}}(\infty;0) := \boldsymbol{M}(\infty;0) + \mu_{\max}\boldsymbol{\mathcal{I}}_{2n}, \tag{F.33}$$

where $\boldsymbol{\mathcal{I}}_{2n}$ is the $2n$ by $2n$ identity matrix. Thus, in particular, $\boldsymbol{\mathcal{M}}(\infty;0)$ is non-negative and so

$$e^{\boldsymbol{M}(\infty;0)} = e^{-\mu_{\max}} e^{\boldsymbol{\mathcal{M}}(\infty;0)} \tag{F.34}$$

is non-negative as the exponential of a non-negative matrix is non-negative (as it is a weighted sum of powers of that matrix with positive weights). Thus, by Perron-Frobenius theory, summarised in [507], there exists a real non-negative eigenvalue $\lambda(\infty;0)$ (called the Perron eigenvalue) of $e^{M(\infty;0)}$ such that any other eigenvalues $\rho(\infty;0)$ satisfy

$$|\rho(\infty;0)| \leq |\lambda(\infty;0)| \tag{F.35}$$

so, in particular

$$\Re(\rho(\infty;0)) \leq \Re(\lambda(\infty;0)). \tag{F.36}$$

Claim: $0 < |\lambda(\infty;0)| < 1$

Proof: Note that $\lambda(\infty;0) > 0$, as

$$\text{trace}\left(e^{\boldsymbol{M}(\infty;0)}\right) \geq \text{trace}\left(e^{-\mu_{\max}}\boldsymbol{\mathcal{I}}_{2n}\right) > 0 \tag{F.37}$$

and thus, $\lambda(\infty;0) \neq 0$.

From [507], there is a non-negative eigenvector, $\boldsymbol{v}$, with eigenvalue $\lambda(\infty;0)$. Now, $\boldsymbol{v}$ must be an eigenvector of $\boldsymbol{M}(\infty;0)$ (as eigenvectors of a matrix and its exponential are the same). Thus, there is some $\lambda^*(\infty;0)$ such that

$$\boldsymbol{M}(\infty;0)\boldsymbol{v} = \lambda^*(\infty;0)\boldsymbol{v}. \tag{F.38}$$

In particular, writing $\boldsymbol{v} = (v_1, ..., v_{2n})^T$

$$\lambda^*(\infty;0)v_1 = (\boldsymbol{M}(\infty;0)\boldsymbol{v})_1 = -\mu_1^1 v_1 \tag{F.39}$$

and thus, either $\lambda^*(\infty;0) = -\mu_1^1 < 0$ or $v_1 = 0$. Suppose first that $\lambda^*(\infty;0) = -\mu_1^1$. Then, this means

that (as the eigenvalues of $e^{\boldsymbol{M}(\infty;0)}$ are the exponentials of the eigenvalues of $\boldsymbol{M}(\infty;0)$),

$$|\lambda(\infty;0)| = |e^{-\mu_1^1}| < 1. \tag{F.40}$$

Similarly, $v_{n+1} \neq 0$ implies that

$$|\lambda(\infty;0)| = |e^{-\mu_1^2}| < 1. \tag{F.41}$$

Thus, suppose for the remainder of the proof of this claim that $v_1 = v_{n+1} = 0$. Now, for $i \leq n$, the entries on the ith row of $\boldsymbol{M}(\infty;0)$ are given by

$$M(\infty;0)_{ij} = \begin{cases} S_i(\infty;0)\beta_{ij}^1 - \delta_{ij}\mu_i^1 & \text{if} j \leq n \\ \\ S_{i-n}(\infty;0)\beta_{i(j-n)}^3 & \text{if} j > n \end{cases} \tag{F.42}$$

and for $i > n$, they are given by

$$M(\infty;0)_{ij} = \begin{cases} S_i^V(\infty;0)\beta_{ij}^2 & \text{if} j \leq n \\ \\ S_{i-n}^V(\infty;0)\beta_{i(j-n)}^4 - \delta_{ij}\mu_i^2 & \text{if} j > n \end{cases}, \tag{F.43}$$

where δ_{ij} is the Kronecker delta.

Now, as $\Pi(0) = \{2, ..., n\}$, by Lemma E.20, it is necessary that

$$J_i(t;0) > 0 \quad \forall t > 0 \quad \text{and} \quad i \in \{2, ..., n\}. \tag{F.44}$$

Moreover, if $I_i^V(t;0) = 0$ for some $t > 0$, then, by Lemma F.7, as $\Pi(0) = \{2, ..., n\}$, it is necessary that

$$S_i^V(t;0)\beta_{ji}^3 = S_i^V(t;0)\beta_{ji}^4 = 0 \quad \forall j \in \{2, ..., n\} \tag{F.45}$$

and so, if $t \geq t_U$, then this implies

$$S_i^V(\infty;0)\beta_{ji}^3 = S_i^V(\infty;0)\beta_{ji}^4 = 0 \quad \forall j \in \{2, ..., n\}. \tag{F.46}$$

Thus, in this case, for $j \notin \{1, n+1\}$

$$M(\infty;0)_{ij} = -\mu_{(i-n)}^2\delta_{ij}. \tag{F.47}$$

Therefore, suppose $\boldsymbol{J}_i(t_U;0) = 0$ for some $i \notin \{1, n+1\}$ (and so, necessarily, $i \in \{n+2, ..., 2n\}$). Then,

$$(\boldsymbol{M}(\infty;0)\boldsymbol{v})_i = \sum_{j=1}^{2n} M(\infty;0)_{ij} v_j \tag{F.48}$$

$$= M\infty;0)_{i1} v_1 + M(\infty;0)_{i(n+1)} v_{(n+1)} + M(\infty;0)_{ii} v_i \tag{F.49}$$

$$= -\mu_i^2 v_i \tag{F.50}$$

and so

$$|\lambda(\infty;0)| = |e^{-\mu_i^2}| < 1. \tag{F.51}$$

Consequently, this holds if any $\boldsymbol{J}_i(t_U;0) = 0$. Conversely, suppose that $\boldsymbol{J}_i(t_U;0) \neq 0$ for all $i \notin \{1, (n+1)\}$. Then, there exists some $\alpha > 0$ and some non-negative vector $\boldsymbol{w}$ such that

$$\boldsymbol{J}(t_U;0) = \alpha \boldsymbol{v} + \boldsymbol{w}. \tag{F.52}$$

Therefore, for any positive integer n,

$$e^{n\boldsymbol{M}(\infty;0)}\boldsymbol{J}(t_U;0) = e^{n\boldsymbol{M}(\infty;0)}(\alpha\boldsymbol{v} + \boldsymbol{w}) = \lambda(\infty;0)^n \alpha \boldsymbol{v} + e^{n\boldsymbol{M}(\infty;0)}\boldsymbol{w} \geq \lambda(\infty;0)^n \alpha \boldsymbol{v}. \tag{F.53}$$

Now, $\boldsymbol{v}$ is an eigenvector so it has a non-zero component, which means that, using (F.31),

$$\left(\lim_{n\to\infty}(e^{nM(\infty;0)}\boldsymbol{J}(t_U;0)) = \boldsymbol{0}\right) \Rightarrow \left(\lim_{n\to\infty}(\lambda(\infty;0)^n \alpha \boldsymbol{v}) = \boldsymbol{0}\right) \Rightarrow \left(|\lambda(\infty;0)| < 1\right) \tag{F.54}$$

and so $|\lambda(\infty;0)| < 1$ holds in all cases, which finishes the proof of this claim.

Claim: *There exists some T dependent on δ and some constant X independent of δ such that $\int_T^\infty J_i(s;\epsilon)ds \leq X\delta$*

Proof: Now, the exponentials of the eigenvalues of $\boldsymbol{M}(\infty;0)$ are the eigenvalues of $e^{\boldsymbol{M}(\infty;0)}$ which means that, if $\eta(\infty;0)$ is an eigenvalue of $\boldsymbol{M}(\infty;0)$ then there exists some $\kappa > 0$ such that

$$|e^{\eta(\infty;0)}| \leq |\lambda(\infty;0)| < e^{-4\kappa} < 1 \Rightarrow |e^{\Re(\eta(\infty;0))}| < e^{-4\kappa} \Rightarrow \Re(\eta) < -4\kappa \tag{F.55}$$

and so all eigenvalues of $\boldsymbol{M}(\infty;0)$ have strictly negative real part. Thus, by continuous dependence of eigenvalues on the matrix, as $\boldsymbol{M}(t;0)$ converges to $\boldsymbol{M}(\infty;0)$ as $t \to \infty$, there exists some $T > t_U$

such that

$$\Re(\eta(t;0)) < -2\kappa \quad \forall t > T \tag{F.56}$$

where $\eta(t;0)$ is an eigenvalue of $M(t;0)$. Now, fix $\delta > 0$. From Lemma E.18, by choosing T to be sufficiently large, one can assume that

$$J_i(T;0) < \delta \quad \forall i \in \{1, ..., 2n\}. \tag{F.57}$$

Moreover, there exists some Δ (which is dependent on T) such that

$$\Re(\eta(T;\epsilon)) < -\kappa \quad \forall \epsilon \in [0, \Delta]. \tag{F.58}$$

Now, similarly, by choosing Δ to be sufficiently small, one can assume that by Lemma F.6

$$|J_i(t;\epsilon) - J_i(t;0)| < \delta \quad \forall t < T \Rightarrow |J_i(T;\epsilon)| < 2\delta \quad \forall i \in \{1, ..., 2n\} \quad \text{and} \quad \forall \epsilon \in [0, \Delta] \tag{F.59}$$

and, for all $t < T$,

$$|R_i(T;\epsilon) - R_i(T;0)|, |R_i^V(T;\epsilon) - R_i^V(T;0)| < \delta \quad \forall i \in \{1, ..., 2n\}, \quad \text{and} \quad \forall \epsilon \in [0, \Delta]. \tag{F.60}$$

Now, for any $t > 0$,

$$\boldsymbol{M}(t + T;\epsilon) \leq \boldsymbol{M}(T;\epsilon). \tag{F.61}$$

Thus, as the solution to the system

$$\frac{d\boldsymbol{z}}{dt} = \boldsymbol{M}(T;\epsilon)\boldsymbol{z}, \quad \boldsymbol{z}(0) = \boldsymbol{J}(T;\epsilon) \tag{F.62}$$

is

$$\boldsymbol{z}(t) = e^{\boldsymbol{M}(T;\epsilon)t}\boldsymbol{J}(T;0); \tag{F.63}$$

one has, by Lemma E.16,

$$\boldsymbol{J}(t + T;\epsilon) \leq e^{\boldsymbol{M}(T;\epsilon)t}\boldsymbol{J}(T;0). \tag{F.64}$$

Now, noting that $\boldsymbol{M}(T;\epsilon)$ is invertible as all its eigenvalues have strictly negative real part, for any

$t > 0$

$$\int_T^{t+T} \boldsymbol{J}(s;\epsilon)ds \leq \int_0^t e^{\boldsymbol{M}(T;\epsilon)s}\boldsymbol{J}(T;\epsilon)ds \tag{F.65}$$

$$= \boldsymbol{M}(T;\epsilon)^{-1}(e^{M(T;\epsilon)t}\boldsymbol{J}(T;\epsilon) - \boldsymbol{J}(T;\epsilon)) \tag{F.66}$$

and so, taking t to ∞ and noting that all eigenvalues of $e^{M(T;\epsilon)}$ have real part less than 1 shows that

$$\int_T^{\infty} \boldsymbol{J}(s;\epsilon)ds \leq -\boldsymbol{M}(T;\epsilon)^{-1}\boldsymbol{J}(T;\epsilon). \tag{F.67}$$

Now, each element of $\boldsymbol{M}(t;\epsilon)$ is uniformly bounded (for any bounded range of ϵ and all $t \geq 0$) as the parameters and variables are uniformly bounded. Thus, by expressing the inverse in terms of determinants of sub-matrices of $\boldsymbol{M}(t;\epsilon)$ (each of which must be uniformly bounded as $\boldsymbol{M}(t;\epsilon)$ is uniformly bounded) by Cramer's rule ([508]), one can see that there exists a constant M^* such that for each i and j,

$$\det(\boldsymbol{M}(t;\epsilon)) \neq 0 \Rightarrow |\boldsymbol{M}(t;\epsilon)^{-1}_{ij}| \leq \left|\frac{M^*}{\det(M(t;\epsilon))}\right|. \tag{F.68}$$

Note that

$$|\det(\boldsymbol{M}(T;\epsilon))| = \left|\prod_{\lambda \text{ eigenvalue of } \boldsymbol{M}(T;\epsilon)} (\lambda)\right| \geq \kappa^n \tag{F.69}$$

because all eigenvalues of $M(T;\epsilon)$ have real part at most $-\kappa$ and hence modulus at least κ. Thus, there exists some constant X (independent of δ) such that for each i and j,

$$\left|\boldsymbol{M}(T;\epsilon)^{-1}_{ij}\right| \leq \frac{X}{4n}. \tag{F.70}$$

Thus, by the conditions on $\boldsymbol{J}(T;\epsilon)$,

$$\int_T^{\infty} J_i(s;\epsilon)ds \leq X\delta \tag{F.71}$$

which completes the proof of this claim.

As all the parameters and variables are uniformly bounded for all ϵ, there exists a constant Y (independent of δ) such that

$$\left|\frac{dJ_i}{dt}\right| \leq Y \quad \forall i \in \{1, ..., 2n\}. \tag{F.72}$$

Now, suppose there exists some $J_i(t;\epsilon) > \delta^{\frac{1}{3}}$ for $t > T$ and $\epsilon \in [0, \eta_1]$. Then, by non-negativity of J_i

$$\int_T^{\infty} J_i(s;\epsilon)ds \geq \int_t^{t+\delta^{\frac{1}{2}}} J_i(s;\epsilon)ds \geq \int_0^{\delta^{\frac{1}{2}}} \delta^{\frac{1}{3}} - Ysds = \delta^{\frac{5}{6}} - \frac{Y}{2}\delta. \tag{F.73}$$

Thus, taking δ sufficiently small such that

$$\delta^{\frac{5}{6}} - \frac{Y}{2}\delta > X\delta \tag{F.74}$$

gives a contradiction. This means that, for each $i \in \{1, ..., 2n\}$

$$J_i(t;\epsilon) \leq \delta^{\frac{1}{3}} \quad \forall t \geq T \quad \text{and} \quad \forall \epsilon \in [0, \Delta] \tag{F.75}$$

and hence, combining this with (F.59) (and assuming $\delta < 1$ so $\delta < \delta^{\frac{1}{3}}$) shows that

$$|J_i(t;\epsilon) - J_i(t;0)| \leq \delta^{\frac{1}{3}} \quad \forall t \quad \text{and} \quad \forall i \in \{1, ..., 2n\}. \tag{F.76}$$

Moreover, by (F.71), for any $t > 0$

$$|R_i(T+t;\epsilon) - R_i(T+t;0)| \leq |R_i(T;\epsilon) - R_i(T;0)| + |R_i(T;0) - R_i(T+t;0)| \tag{F.77}$$

$$\leq \delta + |R_i(T+t;\epsilon) - R_i(T;\epsilon)| + |R_i(T+t;0) - R_i(T;0)| \tag{F.78}$$

$$\leq \delta + 2X\mu_i^1(\epsilon)\delta + 2X\mu_i^1(0)\delta \tag{F.79}$$

$$\leq X^*\delta \tag{F.80}$$

for some constant X^*, alongside an identical result for R_i^V. Combining this with (F.60) (and redefining $\delta \to \delta^3$), the result of the proposition is proved.

F.1.3 Theorem 8.1

Note that Proposition F.2 also holds for the vaccination policies $\tilde{\boldsymbol{U}}(t;\epsilon)$, using Proposition F.1. Thus, one can define a function $\delta(\epsilon)$ such that for all sufficiently small ϵ

$$|f_i(t;\epsilon) - f_i(t;0)|, |\tilde{f}_i(t;\epsilon) - f_i(t;0)| \leq \delta(\epsilon) \quad \forall f \in \{I, I^V, R, R^V\} \tag{F.81}$$

and

$$\delta(\epsilon) = o(1) \quad \text{as } \epsilon \to 0. \tag{F.82}$$

Then, using, for example

$$|R_i(\infty;\epsilon) - \tilde{R}_i(\infty;\epsilon)| \leq |R_i(\infty;\epsilon) - \tilde{R}_i(\infty;0)| + |R_i(\infty;\epsilon) - \tilde{R}_i(\infty;0)| \tag{F.83}$$

(as $R(\infty;0) = \tilde{R}(\infty;0)$) shows that

$$|R_i(\infty;\epsilon) - \tilde{R}_i(\infty;\epsilon)|, |R_i^V(\infty;\epsilon) - \tilde{R}_i^V(\infty;\epsilon)| < 2\delta(\epsilon) \quad \forall \epsilon \in [0,\eta] \tag{F.84}$$

which means

$$\left| \sum_{j=2}^{n} p_j \left(R_j(\infty;\epsilon) + \kappa_j R_j^V(\infty;\epsilon) \right) - \sum_{j=2}^{n} p_j \left(\tilde{R}_j(\infty;\epsilon) + \kappa_j \tilde{R}_j^V(\infty;\epsilon) \right) \right| = O(\delta). \tag{F.85}$$

Thus, the aim of the remainder of the proof is to show that the leading order changes to $R_1(\infty;\epsilon)$ are of exactly $O(\epsilon)$, and so $p_1 R_1(\infty;\epsilon)$ changes by an $O(1)$ amount, meaning these changes to the objective function will eventually dominate the other changes given in (F.85). This can be done by taking advantage of the fact that the quantities $f_1(t;\epsilon)$ are small, and so a linearised version of the equations for group 1 can be used.

Before beginning this process, it is helpful to note the following. From (F.56) in the proof of Proposition F.2, there exists some $T^* > t_U$ independent of δ and ϵ such that

$$\lambda(T^*;0) < e^{-2\kappa} < 1 \tag{F.86}$$

where $\lambda(T^*;0)$ is the (necessarily real and non-negative) Perron eigenvalue of $e^{M(T^*;0)}$ (and is the exponential of the $\eta(\infty;0)$ referenced in (F.56)). Moreover, by the continuity of eigenvalues on the entries of the matrix, there exists some $\Delta > 0$ such that the analogously defined $\lambda(T^*;\epsilon)$ also satisfies

$$\lambda(T^*;\epsilon) < e^{-\kappa} < 1 \quad \forall \epsilon \in [0,\Delta]. \tag{F.87}$$

Now, note that, for $t \geq T^* > t_U$, the matrix $M(t;\epsilon)$ and hence the matrix $e^{M(t;\epsilon)}$ is non-increasing. Thus, as $e^{M(t;\epsilon)}$ is non-negative (as proved in Proposition F.2), it is necessary from Perron-Frobenius theory ([507]) that its Perron eigenvalue, $\lambda(t;\epsilon)$ satisfies

$$\lambda(t;\epsilon) \leq \lambda(T^*;\epsilon) < e^{-\kappa} < 1. \tag{F.88}$$

Then, following the method used to derive (F.67), one has, for any $t \geq T^*$

$$\int_t^{\infty} I_1(t;\epsilon)dt \leq (\boldsymbol{M}(t;\epsilon)^{-1}\boldsymbol{J}(t;\epsilon))_1 \quad \forall \epsilon \in [0,\Delta]. \tag{F.89}$$

This is exactly the same equation as (F.67), except that here, T^* is independent of δ (as no conditions on $\boldsymbol{J}(T;0)$ are assumed). Now, note that

$$M(t;0)_{1j} = -\mu_1^1 \delta_{1j} \quad \text{and} \quad M(t;0)_{(n+1)j} = -\mu_1^2 \delta_{(n+1),j} \tag{F.90}$$

where here δ_{ij} is the Kronecker delta. This means that, for any vector $\boldsymbol{y}$, the equation

$$\boldsymbol{M}(t;0)\boldsymbol{x} = \boldsymbol{y} \tag{F.91}$$

must satisfy

$$x_1 = \frac{-y_1}{\mu_1^1} \quad x_{n+1} = -\frac{y_{n+1}}{\mu_1^2} \quad \text{and} \quad \boldsymbol{x} = \boldsymbol{M}^{-1}\boldsymbol{y}. \tag{F.92}$$

Thus, in particular

$$\boldsymbol{M}_{1j}^{-1}(t;0) = \frac{-1}{\mu_1^1}\delta_{1j} \quad \text{and} \quad \boldsymbol{M}_{(n+1)j}^{-1}(t;0) = \frac{-1}{\mu_1^2}\delta_{(n+1)j}, \tag{F.93}$$

where here δ_{ij} denotes the Kronecker delta. Now, note that, as the inverse of a matrix is a rational function of its entries,

$$\boldsymbol{M}^{-1}(t;0) = \boldsymbol{M}^{-1}(t;\epsilon) + O(\epsilon) \tag{F.94}$$

and hence

$$\boldsymbol{M}_{1j}^{-1}(t;0) = \frac{-1}{\mu_1^1}\delta_{1j} + O(\epsilon). \tag{F.95}$$

Moreover, defining

$$\mu_{\min} := \min\{\mu_i^1, \mu_i^2\}, \tag{F.96}$$

there must exist a $T(\epsilon) \in \left(T^*, T^* + \frac{2n}{\delta^{\frac{1}{3}}\mu_{\min}}\right)$ such that for each i,

$$I_i(T(\epsilon);\epsilon) < \delta^{\frac{1}{3}} N_i(\epsilon). \tag{F.97}$$

Otherwise,

$$\sum_{i=1}^{n} \frac{d}{dt}\left(\frac{R_i(t;\epsilon)}{\mu_i^1 N_i(\epsilon)} + \frac{R_i^V(t;\epsilon)}{\mu_i^2 N_i(\epsilon)}\right) \geq \sum_{i=1}^{n}\left(\frac{\mu_i^1 I_i(t;\epsilon)}{\mu_i^1 N_i(\epsilon)} + 0\right) \geq \delta^{\frac{1}{3}} \quad \forall t \in \left(T^*, T^* + \frac{2n}{\delta^{\frac{1}{3}}\mu_{\min}}\right) \tag{F.98}$$

and integrating this between T^* and $T^* + \frac{2n}{\delta^{\frac{1}{3}}\mu_{\min}}$ gives

$$\sum_{i=1}^{n}\left[\frac{R_i\left(T^* + \frac{2n}{\delta^{\frac{1}{3}}\mu_{\min}};\epsilon\right)}{\mu_i^1 N_i(\epsilon)} + \frac{R_i^V\left(T^* + \frac{2n}{\delta^{\frac{1}{3}}\mu_{\min}};\epsilon\right)}{\mu_i^2 N_i(\epsilon)}\right] \geq \frac{2n\delta^{\frac{1}{3}}}{\delta^{\frac{1}{3}}\mu_{\min}} > \frac{n}{\mu_{\min}}. \tag{F.99}$$

Thus, as $\frac{\mu_{\min}}{\mu_i^\alpha} \leq 1$ for each i and α,

$$\sum_{i=1}^{n}\left[\frac{R_i\left(T^* + \frac{2n}{\delta^{\frac{1}{3}}(\mu_{\min}+1)};0\right) + R_i^V\left(T^* + \frac{2n}{\delta^{\frac{1}{3}}(\mu_{\min}+1)};0\right)}{N_i(\epsilon)}\right] > n \tag{F.100}$$

which means, for some i

$$\frac{R_i\left(T^* + \frac{2n}{\delta^{\frac{1}{3}}(\mu_{\min}+1)};0\right) + R_i^V\left(T^* + \frac{2n}{\delta^{\frac{1}{3}}(\mu_{\min}+1)};0\right)}{N_i(\epsilon)} > 1, \tag{F.101}$$

which is a contradiction as the total population size in group i cannot exceed $N_i(\epsilon)$ by definition of $N_i(\epsilon)$. Thus, for each $\epsilon \in [0, \Delta]$,

$$\int_{T(\epsilon)}^{\infty} I_1(t;\epsilon)dt \leq (\boldsymbol{M}(T;\epsilon)^{-1}\boldsymbol{J}(T(\epsilon);\epsilon))_1 \tag{F.102}$$

$$= \begin{pmatrix} O(1) & O(\epsilon) & ... & O(\epsilon) \end{pmatrix} \begin{pmatrix} O(\epsilon\delta^{\frac{1}{3}}) \\ O(\delta^{\frac{1}{3}}) \\ . \\ . \\ . \\ O(\delta^{\frac{1}{3}}) \end{pmatrix} \tag{F.103}$$

$$= O(\epsilon\delta^{\frac{1}{3}}) \tag{F.104}$$

while similarly

$$\int_{T(\epsilon)}^{\infty} I_1^V(t;\epsilon)dt = O(\epsilon\delta^{\frac{1}{3}}). \tag{F.105}$$

Moreover,

$$\int_0^{T(\epsilon)} \delta\epsilon dt = O(\epsilon\delta^{\frac{2}{3}}). \tag{F.106}$$

These results allow for the linearisation to be carried out. To reduce notation, define

$$T := T(\epsilon). \tag{F.107}$$

Now, to begin the linearisation, define

$$X(t) = \sum_{j=1}^{n} \left[\beta_{1j}^{1} I_j(t;0) + \beta_{1j}^{2} I_j^V(t;0) \right], \tag{F.108}$$

which is the leading order infective force on group 1. By Proposition F.2,

$$X(t) = \sum_{j=1}^{n} \left[\beta_{1j}^{1} I_j(t;\epsilon) + \beta_{1j}^{2} I_j^V(t;\epsilon) \right] + O(\delta). \tag{F.109}$$

Then, as $S_1(t;\epsilon) \leq \epsilon$,

$$\frac{dI_1}{dt}(t;\epsilon) + \mu_1^1 I_1(t) = S_1(t;\epsilon)X(t) + O(\delta\epsilon). \tag{F.110}$$

Now, note that

$$\begin{aligned}
R_1(\infty;\epsilon) &= \mu_1^1 \int_0^\infty I_1(t;\epsilon)dt && \text{(F.111)}\\
&= \mu_1^1 \int_0^T I_1(t;\epsilon)dt + \mu_1^1 \int_T^\infty I_1(t;\epsilon)dt && \text{(F.112)}\\
&= \int_0^T \left(S_1(t;\epsilon)X(t) - \frac{dI_1}{dt}(t;\epsilon) + O(\epsilon\delta) \right) dt + O(\delta^{\frac{1}{3}}\epsilon) && \text{(F.113)}\\
&= I_1(0;\epsilon) - I_1(T) + \int_0^T S_1(t;\epsilon)X(t)dt + O(\delta^{\frac{1}{3}}\epsilon) && \text{(F.114)}\\
&= I_1(0;\epsilon) + \int_0^T S_1(t;\epsilon)X(t)dt + O(\delta^{\frac{1}{3}}\epsilon). && \text{(F.115)}
\end{aligned}$$

Now, the equations for I^V are of the same form, but with S^V in place of S and a different leading order infection function $Y(t)$ given by

$$Y(t) = \sum_{j=1}^{n} \left[\beta_{1j}^{3} I_j(t;0) + \beta_{ij}^{4} I_j^V(t;0) \right]. \tag{F.116}$$

Thus, an analogous derivation (noting that $I^V(0;\epsilon) = 0$) shows that

$$R_1^V(\infty;0) = \int_0^T Y(t) S_1^V(t;\epsilon)dt + O(\epsilon\delta^{\frac{1}{3}}) \tag{F.117}$$

while analogous results hold for $\tilde{R}_1$ and $\tilde{R}_1^V$ (with $\tilde{S}_1$ and $\tilde{S}_1^V$ in place of S_1 and S_1^V). Now, note that

$$S_1(t;\epsilon) = S_1(t;\epsilon)\left(\frac{N_1(\epsilon) - W_1(t;\epsilon)}{N_1(\epsilon)}\right)\exp\left[-\sum_{j=1}^{n}\left(\frac{\beta_{1j}^1 R_j(t;\epsilon)}{\mu_j^1} + \frac{\beta_{1j}^2 R_j^V(t;\epsilon)}{\mu_j^2}\right)\right] \tag{F.118}$$

$$= \sigma(N_1(\epsilon) - W_1(t;\epsilon))\exp\left[-\sum_{j=1}^{n}\left(\frac{\beta_{1j}^1 R_j(t;\epsilon)}{\mu_j^1} + \frac{\beta_{1j}^2 R_j^V(t;\epsilon)}{\mu_j^2}\right)\right]. \tag{F.119}$$

Define

$$P(t) := \exp\left[-\sum_{j=1}^{n}\left(\frac{\beta_{1j}^1 R_j(t;0)}{\mu_j^1} + \frac{\beta_{1j}^2 R_j^V(t;0)}{\mu_j^2}\right)\right] \tag{F.120}$$

and then, note that by Proposition F.2

$$P(t) = \exp\left[-\sum_{j=1}^{n}\left(\frac{\beta_{1j}^1 R_j(t;\epsilon)}{\mu_j^1} + \frac{\beta_{1j}^2 R_j^V(t;\epsilon)}{\mu_j^2}\right)\right] + O(\delta) \tag{F.121}$$

which means (as $(N_1(\epsilon) - W_1(t;\epsilon)) \leq \epsilon$ and $\sigma < 1$)

$$S_1(t;\epsilon) = \sigma(N_1 - W_1(t;\epsilon))P(t) + O(\delta\epsilon) \tag{F.122}$$

with an identical result for $\tilde{S}$. It is helpful to note for later that, as $W_1(t;\epsilon) \leq \tilde{W}(t;\epsilon)$, this means that

$$S_1(t;\epsilon) \geq \tilde{S}_1(t;\epsilon) + O(\delta\epsilon). \tag{F.123}$$

Now, this means

$$\int_0^T X(t)S_1(t;\epsilon)dt = \int_0^T X(t)\sigma(N_1 - W_1(t;\epsilon))P(t)dt + O(\epsilon\delta^{\frac{2}{3}}) \tag{F.124}$$

and so

$$R_1(\infty;\epsilon) = I_1(0;\epsilon) + \int_0^T X(t)\sigma(N_1 - W_1(t;\epsilon))P(t)dt + O(\epsilon\delta^{\frac{1}{3}}). \tag{F.125}$$

Now, note that

$$\int_0^T X(t)\sigma(N_1 - W_1(t;\epsilon))P(t)dt = \left(\int_0^\tau + \int_\tau^T\right)\left(X(t)\sigma(N_1 - W_1(t;\epsilon))P(t)dt\right) \tag{F.126}$$

and that, as $W_1(t;\epsilon) \leq \tilde{W}_1(t;\epsilon)$,

$$\int_\tau^T X(t)\sigma(N_1 - W_1(t;\epsilon))P(t)dt \geq \int_\tau^T X(t)\sigma(N_1 - \tilde{W}_1(t;\epsilon))P(t)dt. \tag{F.127}$$

Now, define $z(\epsilon)$ to be

$$z(\epsilon) = \inf\left\{t : \sum_{i=1}^{n} W_i(t) = \epsilon\right\}. \tag{F.128}$$

Note that, for $\epsilon < w$, z exists and is bounded above by τ as

$$\sum_{i=1}^{n} W_i(\tau) = w. \tag{F.129}$$

Now, define a fixed value

$$z_0 := z\left(\frac{w}{2}\right) \tag{F.130}$$

so that, by continuity of W, $z_0 < \tau$ (and is independent of ϵ). Suppose that $\epsilon < \frac{w}{2}$ (which will be assumed for the rest of the proof). Note that

$$\int_0^{z_0} X(t)\sigma(N_1 - W_1(t;\epsilon))P(t)dt \geq \int_0^{z_0} X(t)\sigma(N_1 - \tilde{W}_1(t;\epsilon))P(t)dt \tag{F.131}$$

and that

$$\int_{z_0}^{\tau} X(t)\sigma(N_1 - \tilde{W}_1(t;\epsilon))P(t)dt = 0 \tag{F.132}$$

as $\tilde{W}_1(t;\epsilon) = N_1$ for all $t > z(\epsilon)$. Moreover, by (F.2)

$$\int_{z_0}^{\tau} X(t)\sigma(N_1 - W_1(t;\epsilon))P(t)dt \geq (1-\alpha)\epsilon\sigma \int_{z_0}^{\tau} X(t)P(t)dt. \tag{F.133}$$

Now, note that $P(t)$ is strictly positive for $t > 0$ as it is an exponential, while, as $\beta_{1j} > 0$ for some $j \neq 1$,

$$X(t) \geq \beta_{ij} I_j(t;0) > 0 \quad \text{as } j \in \Pi(0). \tag{F.134}$$

Thus,

$$(1-\alpha)\int_{z_0}^{\tau} X(t)P(t)dt > 0 \tag{F.135}$$

and this is independent of ϵ. This means that

$$R_1(\infty;\epsilon) - \tilde{R}(\infty;\epsilon) \geq \epsilon(1-\alpha)\int_z^{\tau} X(t)P(t)dt + O(\epsilon\delta^{\frac{1}{3}}) = \epsilon(1-\alpha)\int_z^{\tau} X(t)P(t)dt + o(\epsilon) \tag{F.136}$$

and so the leading order change in $R_1(\infty;\epsilon)$ is indeed of order exactly ϵ.

Now, it is important to check the leading order change in $R_1^V(\infty;\epsilon)$. Note that, as $S_1(t;\epsilon)$ and

$S_1^V(\epsilon)$ are at most ϵ,

$$\frac{d}{dt}\left(S_1(t;\epsilon)+S_1^V(t;\epsilon)\right) = -X(t)S_1(t;\epsilon) - Y(t)S_1^V(t;\epsilon) + O(\epsilon\delta). \tag{F.137}$$

Using (F.122), this can be written as

$$\frac{d}{dt}\left(S_1(t;\epsilon)+S_1^V(t;\epsilon)\right) + Y(t)(S_1(t;\epsilon)+S_1^V(t;\epsilon)) = (Y(t)-X(t))S_1(t;\epsilon) + O(\epsilon\delta). \tag{F.138}$$

This equation can be integrated by defining

$$\mathcal{Y}(t) := \int_0^t Y(s)ds \tag{F.139}$$

so that

$$\frac{d}{dt}\left(e^{\mathcal{Y}(t)}(S_1(t;\epsilon)+S_1^V(t;\epsilon)\right) = e^{\mathcal{Y}(t)}(Y(t)-X(t))S_1(t;\epsilon) + O(\epsilon\delta). \tag{F.140}$$

Thus, for any $t \le T$

$$\begin{aligned} S_1(t;\epsilon)+S_1^V(t;\epsilon) =& e^{-\mathcal{Y}(t)}(S_1(0;\epsilon)+S_1^V(0;\epsilon)) \\ &+ \int_0^t e^{\mathcal{Y}(s)-\mathcal{Y}(t)}(Y(s)-X(s))S_1(s;\epsilon)ds + O(\epsilon\delta^{\frac{2}{3}}) \end{aligned} \tag{F.141}$$

which means that

$$\begin{aligned} &S_1(t;\epsilon)+S_1^V(t;\epsilon)-\tilde{S}_1(t;\epsilon)-\tilde{S}_1^V(t;\epsilon) = \\ &\int_0^t e^{\mathcal{Y}(s)-\mathcal{Y}(t)}(Y(s)-X(s))\Big(S_1(s;\epsilon)-\tilde{S}_1(s;\epsilon)\Big)ds + O(\epsilon\delta^{\frac{2}{3}}) \end{aligned} \tag{F.142}$$

Thus,

$$\int_0^t Y(s)\Big[S_1(s;\epsilon)+S_1^V(s;\epsilon)-\tilde{S}_1(s;\epsilon)-\tilde{S}_1^V(s;\epsilon)\Big]ds = \tag{F.143}$$

$$\int_0^t\int_0^s Y(s)e^{\mathcal{Y}(k)-\mathcal{Y}(s)}(Y(k)-X(k))\Big(S_1(k;\epsilon)-\tilde{S}_1(k;\epsilon)\Big)dkds + O(\epsilon\delta^{\frac{1}{3}}) \tag{F.144}$$

$$= \int_0^t\int_k^t \Big[Y(s)e^{-\mathcal{Y}(s)}\Big]e^{\mathcal{Y}(k)}(Y(k)-X(k))\Big(S_1(k;\epsilon)-\tilde{S}_1(k;\epsilon)\Big)dsdk + O(\epsilon\delta^{\frac{1}{3}}) \tag{F.145}$$

$$= \int_0^t (e^{-\mathcal{Y}(k)}-e^{-\mathcal{Y}(t)})e^{\mathcal{Y}(k)}(Y(k)-X(k))\Big(S_1(k;\epsilon)-\tilde{S}_1(k;\epsilon)\Big)dk + O(\epsilon\delta^{\frac{1}{3}}) \tag{F.146}$$

$$= \int_0^t (1-e^{\mathcal{Y}(k)-\mathcal{Y}(t)})(Y(k)-X(k))\Big(S_1(k;\epsilon)-\tilde{S}_1(k;\epsilon)\Big)dk + O(\epsilon\delta^{\frac{1}{3}}). \tag{F.147}$$

Now, note that, as $\mathcal{Y}$ is non-decreasing, and non-negative

$$0 \leq 1 - e^{\mathcal{Y}(k)-\mathcal{Y}(t)} \leq 1 - e^{-\mathcal{Y}(t)}. \tag{F.148}$$

Moreover, one has

$$\mathcal{Y}(t) = \int_0^t \sum_{j=1}^n \left[\beta_{1j}^3 I_j(s;0) + \beta_{1j}^4 I_j^V(s;0)\right] ds \tag{F.149}$$

$$= \sum_{j=1}^n \left[\frac{\beta_{1j}^3 R_j(t;0)}{\mu_j^1} + \frac{\beta_{1j}^4 R_j(t;0)}{\mu_j^2}\right] \tag{F.150}$$

$$\leq \sum_{j=1}^n \left[\frac{\beta_{1j}^3 N_j(1)}{\mu_j^1} + \frac{\beta_{1j}^4 N_j(1)}{\mu_j^2}\right] \tag{F.151}$$

and so $\mathcal{Y}(t)$ is bounded above by some constant (for $\epsilon \leq 1$). This in turn means that there exists some $\mathcal{Y}^*$ such that

$$1 - e^{-\mathcal{Y}(t)} \leq \mathcal{Y}^* < 1. \tag{F.152}$$

Thus, as $Y(t) - X(t) \leq 0$ and $S_1(k;\epsilon) \geq \tilde{S}_1(k;\epsilon) + O(\delta\epsilon)$, for any $k \leq t$

$$\begin{aligned}&\int_0^t Y(s)\left[S_1(s;\epsilon) + S_1^V(s;\epsilon) - \tilde{S}_1(s;\epsilon) - \tilde{S}_1^V(s;\epsilon)\right] \\ &\geq \mathcal{Y}^* \int_0^t (Y(k) - X(k))\left(S_1(k;\epsilon) - \tilde{S}_1(k;\epsilon)\right) dk + O(\epsilon\delta^{\frac{1}{3}}).\end{aligned} \tag{F.153}$$

Now, adding the equations (F.115) and (F.117) together gives

$$R_1(\infty;\epsilon) + R_1^V(\infty;\epsilon) = I_1(0;\epsilon) + \int_0^T X(t)S_1(t;\epsilon) + Y(t)S_1^V(t;\epsilon)dt + o(\epsilon). \tag{F.154}$$

Note that

$$X(t)S_1(t;\epsilon) + Y(t)S_1^V(t;\epsilon) = (X(t) - Y(t))S_1(t;\epsilon) + Y(t)(S_1(t;\epsilon) + S_1^V(t;\epsilon)) \tag{F.155}$$

and hence

$$\begin{aligned}R_1(\infty;\epsilon) + R_1^V(\infty;\epsilon) =& I_1(0;\epsilon) + \\ &\int_0^T (X(t) - Y(t))S_1(t;\epsilon) + Y(t)(S_1(t;\epsilon) + S_1^V(t;\epsilon))dt + o(\epsilon).\end{aligned} \tag{F.156}$$

This means that

$$R_1(\infty;\epsilon) + R_1^V(\infty;\epsilon) - \tilde{R}_1(\infty;\epsilon) - \tilde{R}_1^V(\infty;\epsilon)$$
$$\geq (1-\mathcal{Y}^*)\int_0^T (X(t)-Y(t))\Big(S_1(t;\epsilon) - \tilde{S}_1(t;\epsilon)\Big)dt + O(\epsilon\delta^{\frac{1}{3}}). \tag{F.157}$$

Now, as there is some $i \neq 1$ such that

$$\beta_{1i}^1 > \beta_{1i}^3 \geq 0 \tag{F.158}$$

and (as $i \neq 1$), $i \in \Pi(0)$ which means that

$$\beta_{1i}^1 I_i(t) > \beta_{1i}^3 I_i(t) \quad \forall t > 0. \tag{F.159}$$

This means that $X(t) > Y(t)$ for all $t > 0$ and hence

$$\int_0^T (X(t) - Y(t))dt > 0. \tag{F.160}$$

Thus, following the arguments from before, one can see that

$$\int_0^t (X(s)-Y(s))\Big(S_1(s;\epsilon) - \tilde{S}_1(s;\epsilon)\Big)ds > \epsilon(1-\mathcal{Y}^*)\int_{z_0}^{\tau}(X(t)-Y(t))P(t)dt + o(\epsilon) \tag{F.161}$$

where the leading order term is positive as required (as $P(t)$ is positive). Hence, from (F.157)

$$R_1(\infty;\epsilon) + R_1^V(\infty;\epsilon) - (\tilde{R}_1(\infty;\epsilon) + \tilde{R}_1^V(\infty;\epsilon))$$
$$\geq (1-\mathcal{Y}^*)\epsilon(1-\alpha)\int_{z_0}^{\tau}(X(t)-Y(t))P(t)dt + o(\epsilon). \tag{F.162}$$

Thus, for any $\kappa_1 \in [0,1]$, combining (F.136) and (F.162)

$$R_1(\infty) + \kappa_1 R_1^V(\infty) = \kappa_1(R_1(\infty) + R_1^V(\infty)) + (1-\kappa_1)R_1(\infty) \tag{F.163}$$
$$\geq \epsilon\int_{z_0}^{\tau}(1-\alpha)P(t)\Big[(1-\mathcal{Y}^*)\kappa_1(X(t)-Y(t)) + (1-\kappa_1)X(t)\Big]dt$$
$$+ \kappa_1\tilde{R}_1^V(\infty) + \tilde{R}_1(\infty) + o(\epsilon). \tag{F.164}$$

Thus, recalling (F.85) and that $p_1 = \frac{1}{\epsilon}$

$$H(\boldsymbol{U}) \geq H(\tilde{\boldsymbol{U}}) + \int_{z_0}^{\tau}(1-\alpha)[\kappa_1(X(t)-Y(t)) + (1-\kappa_1)X(t)]dt + o(1) \tag{F.165}$$

for some constant K. Moreover, for sufficiently small ϵ,

$$\int_{z_0}^{\tau} \alpha[\kappa_1(X(t) - Y(t)) + (1 - \kappa_1)X(t)]dt + o(1) > 0 \tag{F.166}$$

and hence

$$H(\boldsymbol{U}(t;\epsilon)) > H(\tilde{\boldsymbol{U}}(t;\epsilon)), \tag{F.167}$$

as required.

F.2 Proof of Theorem 8.2

Note that, throughout this section, the tilde will be removed from the rescaled p_i terms to reduce notation (and hence all p_i terms used here will be assumed to be rescaled).

Recall from the main text that, using the results in [50], if one defines

$$\chi(t) := \begin{cases} A(t) & \text{if} \quad \int_0^t A(s)ds < B(t) \\ \min(A(t), B'(t)) & \text{if} \quad \int_0^t A(s)ds \geq B(t) \end{cases}, \tag{F.168}$$

then (assuming that there is an optimal solution, and under mild smoothness conditions on $\boldsymbol{U}$, A and B) there must be an optimal solution satisfying

$$\sum_{i=1}^{n} W_i(t) = \max\left(\int_0^t \chi(s)ds, 1\right). \tag{F.169}$$

Theorem 2 *With the definitions of Theorem 8.1, suppose additionally that*

$$\sum_{j=2}^{n} (\beta_{1j}^1 - \beta_{1j}^3) I_j(0;\epsilon) > 0. \tag{F.170}$$

That is, the initial difference between the infective force on vaccinated and unvaccinated members of the population is positive. Suppose further that

$$\sigma = 1. \tag{F.171}$$

Suppose an optimal vaccination policy for each ϵ is given by $\overline{\boldsymbol{U}}(t;\epsilon)$ and suppose that $\overline{\boldsymbol{U}}(t;\epsilon)$ has uniformly bounded finite support. Then, there exists an η depending only on α, τ, w and the model parameters such that, for any $\boldsymbol{U}$ satisfying the condition (F.2) as defined in Theorem 8.1

$$\epsilon \in (0, \eta) \Rightarrow H(\boldsymbol{U}(t;\epsilon)) > H(\overline{\boldsymbol{U}}(t;\epsilon)). \tag{F.172}$$

Moreover, there is a sequence of optimal vaccination policies $\overline{\boldsymbol{U}}(t;\epsilon)$ *satisfying*

$$\lim_{\epsilon\to 0}\left(\frac{\overline{W}_1(t;\epsilon)}{\epsilon}\right)=1 \quad \forall t \quad s.t. \quad \int_0^t \chi(s)ds>0. \tag{F.173}$$

To make things clearer in the course of this proof, note that H will be written as

$$H(\boldsymbol{U};\epsilon) \tag{F.174}$$

where the ϵ refers to the size of the population N_1 under consideration.

F.2.1 Proposition F.3

It remains to show that, for sufficiently small ϵ and fixed α, τ and w, there is no $\boldsymbol{U}$ satisfying the conditions (F.2) that is the optimal vaccination policy.

To do this, it is necessary to prove the function $H(\boldsymbol{U}(t;\epsilon);\epsilon)$ is non-increasing in ϵ. This uses the work of [50] as the main result of that paper gives a method of finding inequalities between the objective values of different vaccination policies. However, it is a meaningful extension, as here the population sizes are not identical when objective values are compared.

Proposition F.3 *Suppose that* $I_1(0;\epsilon)=0$ *for all* ϵ. *Consider, for* $\epsilon\leq 1$ *any bounded vaccination policy* $\boldsymbol{U}(t;\epsilon)$ *given by*

$$U_1(t;\epsilon)=\epsilon U_1(t;1) \quad and \quad U_i(t;\epsilon)=U_i(t;1) \quad \forall i\neq 1. \tag{F.175}$$

Then, if $H(\boldsymbol{U}(t;\epsilon);\epsilon)$ *is the value of the objective function for a given value of* ϵ,

$$\epsilon>\epsilon' \Rightarrow H(\boldsymbol{U}(t;\epsilon);\epsilon)\geq H(\boldsymbol{U}(t;\epsilon');\epsilon'). \tag{F.176}$$

Proof: Note that this proof is similar in structure to that of Lemma F.8, and so some long (though simple) calculations are summarised.

Consider a new epidemic, denoted by hats, with $n+1$ groups, where groups 2, 3, ... , n are unchanged. The quantities in this epidemic depend both on ϵ and ϵ', though we do not explicitly write this dependence for brevity.

We suppose that the only parameters that change in group 1 are that

$$\hat{N}_1 = \hat{S}_1(0;\epsilon) = \frac{\epsilon'}{\epsilon} S_1(0;\epsilon) = \epsilon' \tag{F.177}$$

$$\hat{U}_1(t) = \frac{\epsilon'}{\epsilon} U_1(t;\epsilon) \tag{F.178}$$

$$\hat{p}_1 = \frac{\epsilon}{\epsilon'} p_1(\epsilon) = \frac{1}{\epsilon'} \tag{F.179}$$

We suppose that group $n+1$ has exactly the same properties as group 1, with the exception that

$$\hat{N}_{n+1} = \hat{S}_{n+1}(0;\epsilon) = \frac{\epsilon - \epsilon'}{\epsilon} S_1(0;\epsilon) = \epsilon - \epsilon' \tag{F.180}$$

$$\hat{U}_{n+1}(t) = \frac{\epsilon - \epsilon'}{\epsilon} U_1(t;\epsilon) \tag{F.181}$$

$$\hat{p}_{n+1} = 0 \tag{F.182}$$

By adding the equations for groups 1 and $n+1$ and assuming uniqueness of solution, we can see that

$$\hat{f}_1(t) + \hat{f}_{n+1}(t) = f_1(t;\epsilon) \quad \forall f \in \{S, I, R, S^V, I^V, R^V\} \tag{F.183}$$

Moreover, we can see that the equations and initial conditions are satisfied by

$$\hat{f}_1(t) = \frac{\epsilon'}{\epsilon} f_1(t;\epsilon) \quad \text{and} \quad \hat{f}_{n+1}(t) = \frac{\epsilon - \epsilon'}{\epsilon} f_1(t;\epsilon) \quad \forall f \in \{S, I, R, S^V, I^V, R^V\} \tag{F.184}$$

Finally, we note that $\hat{H}(\hat{\boldsymbol{U}}) = H(\boldsymbol{U}(t;\epsilon);\epsilon)$ as, for example

$$\hat{p}_1 \hat{R}_1(\infty) = \left(\frac{\epsilon}{\epsilon'} p_1(\epsilon)\right)\left(\frac{\epsilon'}{\epsilon} R_1(\infty;\epsilon)\right) = p_1(\epsilon) R_1(\infty;\epsilon) \tag{F.185}$$

and $\hat{p}_{n+1} = 0$ so group $n+1$ does not contribute (directly) to the objective.

Now, consider, for some $\Delta > 0$, a new vaccination policy $\boldsymbol{u}$ where (again omitting the dependence on ϵ and ϵ')

$$\boldsymbol{u}_i(t;\Delta) = \boldsymbol{U}_i(t;\epsilon) \quad \forall i < n+1 \quad \text{and} \quad \boldsymbol{u}_{n+1}(t) = \begin{cases} \frac{1}{\Delta} & \text{if } \Delta t < \epsilon - \epsilon' \\ 0 & \text{otherwise} \end{cases} \tag{F.186}$$

That is, we vaccinate all members of group $n+1$ increasingly quickly as $\Delta \to 0$. We use lower-case letters to denote the various quantities under this policy. For sufficiently small Δ, we know that

$$w_{n+1}(t;\Delta) \geq \hat{W}_{n+1}(t) \tag{F.187}$$

as, assuming $A(t)$ is bounded by some $\mathcal{A}$

$$\hat{W}_{n+1}(t) \leq \min(\epsilon - \epsilon', \mathcal{A}t) \tag{F.188}$$

and so for $\Delta < \frac{1}{\mathcal{A}}$, (F.187) holds. Hence, by [50], we know that, for such sufficiently small Δ

$$h(\boldsymbol{u}(t;\Delta)) \leq \hat{H}(\hat{\boldsymbol{U}}) \tag{F.189}$$

Then, following the method of Lemma F.8, we see that the objective function $h(\boldsymbol{u}(t;\Delta))$ converges to the objective function of the same epidemic as in the $\hat{\boldsymbol{U}}$ case, but with the initial conditions in group $n+1$ changed to

$$\hat{s}_{n+1}(0) = 0 \quad \text{and} \quad \hat{s}^V_{n+1}(0) = \epsilon - \epsilon' \tag{F.190}$$

and with no vaccination in group $n+1$. By adding an empty group $n+1$ to our original epidemic at ϵ' (that is, where the vaccination policy was $\boldsymbol{U}(t;\epsilon')$) we see that the only difference between this original epidemic and the epidemic given by $\boldsymbol{u}$ is that the initial population of S^V_{n+1} is non-zero in the latter case (as, in particular, by construction in the $\boldsymbol{u}$ case, there are now ϵ' people initially in group 1, and the vaccination in group 1 has the appropriate scaling). Thus, using Lemma F.8 shows that

$$H(\boldsymbol{U}(t;\epsilon');\epsilon') \leq h(\boldsymbol{u}) \leq \hat{H}(\hat{\boldsymbol{U}}) \leq H(\boldsymbol{U}(t;\epsilon);\epsilon) \tag{F.191}$$

which completes the proof.

F.2.2 Theorem 8.2

This allows the overall proof of Theorem 8.2. The proof will rely on Theorem 8.1, which allows the creation of an $O(1)$ decrease in the objective function by reducing ϵ. By comparing a sequence of policies satisfying (F.2) with a sequence that does not satisfy (F.2) and using Proposition F.3, one can then create a sequence of optimal policies such that the associated sequence of objective values decreases by at least a fixed quantity at each step (and thus will eventually become negative, giving a contradiction).

Suppose (for a contradiction) that Theorem 8.2 does not hold for some fixed α, τ and w. Thus, for any $\eta > 0$, there is an $\epsilon \in (0, \eta)$ such that, for some $\boldsymbol{U}$ satisfying (F.2),

$$H(\boldsymbol{U}(t;\epsilon);\epsilon) \leq H(\overline{\boldsymbol{U}}(t;\epsilon);\epsilon). \tag{F.192}$$

By optimality of $\overline{\boldsymbol{U}}(t;\epsilon)$, (F.192) must in fact be an equality, and so it can be assumed that $\boldsymbol{U}(t;\epsilon) = \overline{\boldsymbol{U}}(t;\epsilon)$, which will be done in the remainder of this proof (that is, if for some ϵ there is an optimal solution satisfying (F.2), then it will be assumed that $\overline{\boldsymbol{U}}$ satisfies (F.2)). Thus, there is some ϵ_0 such that

$$H(\overline{\boldsymbol{U}}(t;\epsilon_0);\epsilon_0) \leq H(\tilde{\overline{\boldsymbol{U}}}(t;\epsilon_0);\epsilon_0) \tag{F.193}$$

where $\tilde{\overline{\boldsymbol{U}}}$ is defined by (F.4). Now, for $\epsilon < \epsilon_0$, define $\boldsymbol{U}^0(t;\epsilon)$ by

$$U_1^0(t;\epsilon) = \frac{\epsilon}{\epsilon_0}\overline{U}_1(t;\epsilon_0) \quad \text{and} \quad \overline{U}_i(t;\epsilon) = U_i^0(t;\epsilon_0) \quad \forall i \neq 1 \tag{F.194}$$

and note that this means that

$$\boldsymbol{U}^0(t;\epsilon_0) = \overline{\boldsymbol{U}}(t;\epsilon_0) \quad \forall t \geq 0. \tag{F.195}$$

By (F.163) in the proof of Theorem 8.1, (noting that the required conditions hold, as $\frac{W_1^0(t;\epsilon)}{\epsilon}$ is independent of ϵ), there exists some $\delta_1 > 0$ such that, for all $\epsilon < \delta_1$,

$$H(\boldsymbol{U}^0(t;\epsilon);\epsilon) > H(\tilde{\boldsymbol{U}}^0(t;\epsilon);\epsilon) \tag{F.196}$$

$$+ \frac{1}{2}\int_{z_0}^{\tau}(1-\alpha)P(t)\Big[(1-\mathcal{Y}^*)\kappa_1(X(t)-Y(t)) + (1-\kappa_1)X(t)\Big]dt. \tag{F.197}$$

where

$$X(t) = \sum_{j=1}^{n}\beta_{1j}^1 I_j(t;\epsilon) + \beta_{1j}^3 I_j^V(t;\epsilon), \tag{F.198}$$

$$Y(t) = \sum_{j=1}^{n}\beta_{1j}^2 I_j(t;\epsilon) + \beta_{1j}^4 I_j^V(t;\epsilon) \tag{F.199}$$

and

$$P(t) = \exp\left[-\sum_{j=1}^{n}\left(\frac{\beta_{1j}^1 R_j(t;0)}{\mu_j^1} + \frac{\beta_{1j}^2 R_j^V(t;0)}{\mu_j^2}\right)\right]. \tag{F.200}$$

Note that z_0, τ and $\mathcal{Y}^*$ are independent of $\boldsymbol{U}^0$, but $X(t)$, $Y(t)$ and $P(t)$ are not. However, note that

$$\frac{dI_i(t;\epsilon)}{dt} \geq -\mu_i^1 I_i(t;\epsilon) \tag{F.201}$$

and so

$$X(t) - Y(t) \geq \sum_{j=2}^{n}(\beta_{1j}^1 - \beta_{1j}^3)e^{-\mu_j^1 t} I_j(0;\epsilon) > 0, \tag{F.202}$$

by the assumption (F.170), giving a bound that is independent of $\boldsymbol{U}^0$. Moreover,

$$X(t) \geq X(t) - Y(t) > 0. \tag{F.203}$$

Finally, for $\epsilon \leq 1$,

$$P(t) \geq \exp\left[-\sum_{j=1}^{n} \left(\frac{\beta_{1j}^1 N_j(\epsilon)}{\mu_j^1} + \frac{\beta_{1j}^2 N_j(\epsilon)}{\mu_j^2} \right) \right] \geq \exp\left[-\sum_{j=1}^{n} \left(\frac{\beta_{1j}^1 N_j(1)}{\mu_j^1} + \frac{\beta_{1j}^2 N_j(1)}{\mu_j^2} \right) \right] > 0 \tag{F.204}$$

and this bound is again independent of $\boldsymbol{U}^0$. Thus,

$$H(\boldsymbol{U}^0(t;\epsilon);\epsilon) > H(\tilde{\boldsymbol{U}}^0(t;\epsilon);\epsilon) + K \quad \forall \epsilon < \delta_1 \tag{F.205}$$

for some constant $K > 0$ where this is now independent of $\boldsymbol{U}^0$. Now, by assumption, there must exist some $\epsilon_1 \in (0, \delta_1)$ such that $\overline{\boldsymbol{U}}(t;\epsilon_1)$ meets the conditions (F.2) so

$$H(\overline{\boldsymbol{U}}(t;\epsilon_1);\epsilon_1) \leq H(\tilde{\overline{\boldsymbol{U}}}(t;\epsilon_1);\epsilon_1) \tag{F.206}$$

while by optimality

$$H(\overline{\boldsymbol{U}}(t;\epsilon_1);\epsilon_1) \leq H(\tilde{\boldsymbol{U}}^0(t;\epsilon_1);\epsilon_1) < H(\boldsymbol{U}^0(t;\epsilon_1);\epsilon_1) - K. \tag{F.207}$$

Now, moreover, note that by Proposition F.3,

$$H(\boldsymbol{U}^0(t;\epsilon_1);\epsilon_1) \leq H(\boldsymbol{U}^0(t;\epsilon_0);\epsilon_0) = H(\overline{\boldsymbol{U}}(t;\epsilon_0);\epsilon_0) \tag{F.208}$$

and so

$$H(\overline{\boldsymbol{U}}(t;\epsilon_1);\epsilon_1) \leq H(\overline{\boldsymbol{U}}(t;\epsilon_0);\epsilon_0) - K. \tag{F.209}$$

Now, this can be continued iteratively so that, for any $n \geq 0$,

$$H(\overline{\boldsymbol{U}}(t;\epsilon_n);\epsilon_n) \leq H(\overline{\boldsymbol{U}}(t;\epsilon_0);\epsilon_0) - Kn \tag{F.210}$$

However, this means that eventually, one finds

$$H(\overline{\boldsymbol{U}}(t;\epsilon_n);\epsilon_n) < 0 \tag{F.211}$$

which is a contradiction. Thus, for each fixed α, w and τ, there must exist some η such that for any $\epsilon \in (0, \eta)$, the optimal solution does not satisfy (F.2).

Now, suppose that $\int_0^t \chi(s)ds > 0$ and suppose $\overline{\boldsymbol{U}}(t;\epsilon)$ is an optimal solution for each value of ϵ such that, for each t

$$\sum_{i=1}^{n} \overline{W}_i(t;\epsilon) = \min\left(\int_0^t \chi(s)ds, 1\right) \tag{F.212}$$

(note that this can be assumed by Theorem 7.2 in [50]). Now, suppose that, for some t

$$\lim_{\epsilon\to 0}\left(\frac{\overline{W}_1(t;\epsilon)}{\epsilon}\right) \neq 1 \quad \text{and} \quad \min\left(\int_0^t \chi(s)ds, 1\right) > 0. \tag{F.213}$$

This means that there exists some $\delta > 0$ such that there is a subsequence ϵ_m satisfying

$$\frac{\overline{W}_1(t;\epsilon_m)}{\epsilon_m} < 1-\delta < 1 \quad \text{and} \quad \lim_{m\to\infty}(\epsilon_m) = 0 \tag{F.214}$$

noting that

$$\frac{\overline{W}_1(t;\epsilon_m)}{\epsilon_m} \leq 1 \quad \forall \epsilon_m > 0. \tag{F.215}$$

However, this means that for each m, $\overline{\boldsymbol{U}}(t;\epsilon_m)$ satisfies the condition (F.2) with $\tau = t$, $\alpha = 1-\delta$ and $w = \min\left(\int_0^t \chi(s)ds, 1\right)$. This is a contradiction to the previous part of the proof (as $\lim_{m\to\infty}(\epsilon_m) = 0$) and hence

$$\lim_{\epsilon\to 0}\left(\frac{W_1^*(t;\epsilon)}{\epsilon}\right) = 1 \quad \forall t \quad \text{s.t.} \quad \min\left(\int_0^t \chi(s)ds, 1\right) > 0, \tag{F.216}$$

as required.

F.3 Proof of Theorem 8.3

Recall the definitions from the main text.

$$\beta'_{ij} = \begin{cases} \beta^1_{ij} & \text{if } i,j \leq n \\ \beta^2_{i(n-j)} & \text{if } i \leq n < j \leq 2n \\ \beta^3_{(n-i)j} & \text{if } j \leq n < i \leq 2n \\ \beta^4_{(n-i)(n-j)} & \text{if } n < i,j \leq 2n \end{cases}, \tag{F.217}$$

$$\mu'_i = \begin{cases} \mu^1_i & \text{if } i \leq n \\ \mu^2_{(i-n)} & \text{if } n < i \leq 2n \end{cases}, \tag{F.218}$$

$$p'_i = \begin{cases} p_i & \text{if } i \leq n \\ \kappa_{(i-n)} p_{(i-n)} & \text{if } n < i \leq 2n \end{cases}, \tag{F.219}$$

$$Q_{ij} = \frac{1}{1 - e^{-\sum_{j=1}^{2n} \frac{\beta'_{ij}}{\mu'_j} R_j(\infty;0)}} \left[\delta_{ij} + \frac{S_i(0;0) e^{-\sum_{j=1}^{2n} \frac{\beta'_{ij}}{\mu'_j} R_j(\infty;0)} \beta'_{ij}}{\mu'_j} \right], \tag{F.220}$$

and

$$\boldsymbol{x} = \boldsymbol{Q}^{-T} \boldsymbol{p}' \quad \text{and} \quad y_i = \frac{S_i(0;0)}{N_i} (x_{i+n} - x_i) \quad \forall i \in \{1, ..., n\}. \tag{F.221}$$

Theorem 3 *Suppose that, for all* $\epsilon > 0$

$$B(t;\epsilon) = \epsilon \quad \forall t \geq 0 \tag{F.222}$$

and that all other parameter values and initial conditions are independent of ϵ*. Suppose that* $A(t)$ *is a continuous function with*

$$A(0) > 0 \tag{F.223}$$

and that the matrix M *is invertible. Assuming that* ϵ *is sufficiently small so that it exists, define*

$$\tau(\epsilon) := \inf \left\{ t : \int_0^t A(s) ds = \epsilon \right\}. \tag{F.224}$$

Suppose that $\boldsymbol{U}$ *satisfies the condition*

$$\sum_{i=1}^{n} U_i(s) = \min \left(\int_0^t \chi(s) ds, 1 \right) \tag{F.225}$$

where χ *is defined in (F.169). Then, for sufficiently small* ϵ*, the objective function is given by*

$$H(\boldsymbol{U}(t;\epsilon)) = H(\boldsymbol{0}) + \boldsymbol{y}^T \boldsymbol{W}(\tau(\epsilon);\epsilon) + o(\epsilon). \tag{F.226}$$

Moreover, if there is a unique element of $\boldsymbol{y}$ *equal to the minimum of* $\boldsymbol{y}$ *then the optimal vaccination policy (to leading order in* ϵ*) is uniquely given by*

$$U_i(t;\epsilon) = \begin{cases} A(t) & \text{if } i = \min\{y_i : i \in \{1, ..., n\}\} \quad \text{and} \quad \int_0^t A(s) ds < \epsilon \\ 0 & \text{otherwise} \end{cases}. \tag{F.227}$$

F.3.1 Proposition F.4

Note that the n-group model can be considered as a $2n$-group model once vaccination has finished - an idea that is formalised in the below proposition. This moderately extends previous work by incorporating the initial vaccination policy into the final size equation, but is not a major advancement on well-known results found in books such as [492].

Proposition F.4 *Define for* $i \in \{1, ..., n\}$,

$$(S_{n+i}, I_{n+i}, R_{n+i}) := (S_i^V, I_i^V, R_i^V). \tag{F.228}$$

Define further

$$\sigma_i(\epsilon) = \begin{cases} -\frac{S_i(0;0)W_i(\tau(\epsilon))}{N_i} & \text{if } i \leq n \\ \frac{S_{i-n}(0;0)W_{i-n}(\tau(\epsilon))}{N_{i-n}} & \text{if } n < i \leq 2n \end{cases} \tag{F.229}$$

and

$$\rho_i(\epsilon) := R_i(\infty;\epsilon) - R_i(\infty;0) \quad \forall i \in \{1, ..., 2n\}. \tag{F.230}$$

Then, $\rho_i(\epsilon)$ *is* $o(1)$ *as* $\epsilon \to 0$ *and*

$$\sigma_i = \frac{\rho_i + S_i(0;0)e^{-\sum_{j=1}^{2n}\frac{\beta'_{ij}}{\mu'_j}R_j(\infty;0)}\left(\sum_{j=1}^{2n}\frac{\beta'_{ij}}{\mu'_j}\rho_j\right) + o(\sigma_i) + \sum_{j=1}^{2n} o(\rho_j) + O(\epsilon^2)}{1 - e^{-\sum_{j=1}^{2n}\frac{\beta'_{ij}}{\mu'_j}R_j(\infty;0)}}. \tag{F.231}$$

Proof: As A is continuous, there is some region $(0, \delta)$ such that

$$\frac{A(0)}{2} < A(t) < 2A(0) \tag{F.232}$$

and hence

$$\int_0^\delta A(t)dt > \frac{\delta A(0)}{2}. \tag{F.233}$$

This lower bound is independent of ϵ and hence, for sufficiently small ϵ,

$$\int_0^\delta A(t)dt > \epsilon. \tag{F.234}$$

Now, by assumption,

$$\sum_{i=1}^{n} U_i(t;\epsilon) = \begin{cases} A(t) & \text{if } \int_0^t A(s)ds < \epsilon \\ 0 & \text{otherwise} \end{cases}. \tag{F.235}$$

By continuity and the definition of $\tau(\epsilon)$,

$$\int_0^{\tau(\epsilon)} A(t)dt = \epsilon \tag{F.236}$$

and note that it is necessary that $\tau(\epsilon) = O(\epsilon)$ as

$$\tau(\epsilon) \leq \frac{2\epsilon}{A(0)} \tag{F.237}$$

for sufficiently small ϵ.

Now, all of the variables are bounded independently of ϵ in the interval $[0, \tau(\epsilon)]$ (including $\boldsymbol{U}$, which is bounded by $2A(0)$). Moreover, assuming $N_i > 0$ for each $i \in \{1, ..., n\}$,

$$N_i - W_i > N_i - \epsilon > \frac{\min_i(N_i)}{2} \tag{F.238}$$

for sufficiently small ϵ. Thus, in particular, all of the derivatives of the model variables are bounded and so

$$S_i(\tau(\epsilon); \epsilon) = S_i(0; 0) + O(\epsilon) \tag{F.239}$$

with analogous results for the other model variables, noting that the initial conditions are identical in each case. Thus, in particular,

$$\frac{dS_i}{dt}(t; \epsilon) = \frac{dS_i}{dt}(0; \epsilon) - \frac{S_i(0; 0)(U_i(t; \epsilon) - U_i(0; \epsilon))}{N_i - W_i(0; \epsilon)} + O(\epsilon) \quad \forall t \in (0, \epsilon), \tag{F.240}$$

noting that the $U_i(t; \epsilon)$ are the only quantities that can change by an $O(1)$ amount in $O(\epsilon)$ time. Now, one can set $U_i(0; \epsilon) = 0$ to reduce notation (noting that the model depends only on the integral of U_i). Moreover, as $W_i(0; \epsilon) = 0$, the initial conditions are independent of ϵ and $\tau(\epsilon) = O(\epsilon)$, integrating gives

$$S_i(\tau(\epsilon); \epsilon) = S_i(0; 0) + \tau(\epsilon)\frac{dS_i}{dt}(0; 0) - \frac{S_i(0; \epsilon)W_i(\tau(\epsilon); \epsilon)}{N_i} + O(\epsilon^2). \tag{F.241}$$

Similarly,

$$S_i^V(\tau(\epsilon); \epsilon)) = S_i^V(0; 0) + \tau(\epsilon)\frac{dS_i^V}{dt}(0; 0) + \frac{S_i(0; 0)W_i(\tau(\epsilon); \epsilon)}{N_i} + O(\epsilon^2) \tag{F.242}$$

while for the other model variables, f_i, there is no $O(1)$ change to the derivative so

$$f_i(\tau(\epsilon); \epsilon) = f(0; 0) + \tau(\epsilon)\frac{df_i}{dt}(0; 0) + O(\epsilon^2). \tag{F.243}$$

Now, for times $t \geq \tau(\epsilon)$, one has $U_i(t;\epsilon) = 0$ and so a standard multi-group SIR model (with initial conditions given by the model variables evaluated at time $\tau(\epsilon)$) is recovered. Thus, in particular, the final number infected can be formulated in terms of a final size equation, following the work of [492] among others. Define, for $i \in \{1, ..., n\}$,

$$(S_{n+i}, I_{n+i}, R_{n+i}) = (S_i^V, I_i^V, R_i^V). \tag{F.244}$$

This new $2n$ group model has the same behaviour as the original model if the parameters are

$$\beta'_{ij} = \begin{cases} \beta^1_{ij} & \text{if } i,j \leq n \\ \beta^2_{i(n-j)} & \text{if } i \leq n < j \\ \beta^3_{(n-i)j} & \text{if } j \leq n < i \\ \beta^4_{(n-i)(n-j)} & \text{if } n < i,j \end{cases}, \quad \mu'_i = \begin{cases} \mu^1_i & \text{if } i \leq n \\ \mu^2_{(i-n)} & \text{if } i > n \end{cases} \tag{F.245}$$

and

$$p'_i = \begin{cases} p_i & \text{if } i \leq n \\ \kappa_{(i-n)} p_{(i-n)} & \text{if } i > n \end{cases}. \tag{F.246}$$

Thus, integrating the S_i equation between $\tau(\epsilon)$ and $t + \tau(\epsilon)$ gives

$$\frac{d}{dt}\left(\log(S_i)\right) = -\sum_{j=1}^{2n} \frac{\beta'_{ij}}{\mu'_j} \frac{dR_j}{dt} \tag{F.247}$$

$$\Rightarrow \ln(S_i(t+\tau(\epsilon);\epsilon)) = \ln(S_i(\tau(\epsilon);\epsilon)) - \sum_{j=1}^{2n} \frac{\beta'_{ij}}{\mu'_j}\left[R_j(t+\tau(\epsilon);\epsilon) - R_j(\tau(\epsilon);\epsilon)\right] \tag{F.248}$$

$$\Rightarrow S_i(t+\tau(\epsilon);\epsilon) = S_i(\tau(\epsilon);\epsilon) e^{-\sum_{j=1}^{2n} \frac{\beta'_{ij}}{\mu'_j}\left[R_j(t+\tau(\epsilon);\epsilon) - \tau(\epsilon)\frac{dR_j}{dt}(0;0)\right]} + O(\epsilon^2) \tag{F.249}$$

as $R_j(0;0) = 0$ for each j. Now, note that for any $t \geq 0$,

$$S_i(\tau(\epsilon);\epsilon) + I_i(\tau(\epsilon);\epsilon) + R_i(\tau(\epsilon);\epsilon) = S_i(t+\tau(\epsilon);\epsilon) + I_i(t+\tau(\epsilon);\epsilon) + R_i(t+\tau(\epsilon);\epsilon) \tag{F.250}$$

and hence, taking $t \to \infty$ and using Lemma E.18 shows that

$$S_i(\tau(\epsilon);\epsilon) + I_i(\tau(\epsilon);\epsilon) + R_i(\tau(\epsilon);\epsilon) = S_i(\infty;\epsilon) + R_i(\infty;\epsilon). \tag{F.251}$$

Hence, by (F.243),

$$S_i(\tau(\epsilon);\epsilon) + I_i(0;0) + \tau(\epsilon)\left[\frac{dI_i}{dt}(0;0) + \frac{dR_i}{dt}(0;0)\right] = S_i(\infty;\epsilon) + R_i(\infty;\epsilon) + O(\epsilon^2). \tag{F.252}$$

Now, substituting this into the limit of (F.249) as $t \to \infty$ shows that

$$\begin{aligned} R_i(\infty;\epsilon) =& S_i(\tau(\epsilon);\epsilon) + I_i(0;0) + \tau(\epsilon)\left[\frac{dI_i}{dt}(0;0) + \frac{dR_i}{dt}(0;\epsilon)\right] \\ & - S_i(\tau(\epsilon);\epsilon)e^{-\sum_{j=1}^{2n}\frac{\beta'_{ij}}{\mu'_j}\left[R_j(\infty;\epsilon)-\tau(\epsilon)\frac{dR_j}{dt}(0;0)\right]} + O(\epsilon^2). \end{aligned} \tag{F.253}$$

By treating this model as a model that has initial conditions given by the variable values at time $\tau(\epsilon)$, one sees that these initial conditions differ from the initial conditions of the $\epsilon = 0$ model by $O(\epsilon)$ (where no vaccination occurs in either case). This means that Proposition F.2 can be used (as the vaccination policies $\boldsymbol{U}$ must have uniformly bounded finite support for sufficiently small ϵ) and so there exists some function $\delta(\epsilon)$ such that, for all sufficiently small ϵ,

$$|R_j(\infty;\epsilon) - R_j(\infty;0)| < \delta(\epsilon) \quad \forall j \quad \text{and} \quad \delta(\epsilon) = o(1). \tag{F.254}$$

Thus, in particular, one can define functions $\rho_j(\epsilon)$ such that

$$R_j(\infty;\epsilon) = R_j(\infty;0) + \rho_j(\epsilon) \quad \forall j \in \{1,..,2n\} \tag{F.255}$$

and

$$\rho_j(\epsilon) = o(1) \quad \text{as } \epsilon \to 0. \tag{F.256}$$

Furthermore, defining σ_i such that

$$\sigma_i(\epsilon) = \begin{cases} -\frac{S_i(0;0)W_i(\tau(\epsilon))}{N_i} & \text{if } i \leq n \\ \frac{S_{i-n}(0;0)W_{i-n}(\tau(\epsilon))}{N_{i-n}} & \text{if } n < i \leq 2n \end{cases} \tag{F.257}$$

gives

$$S_i(\tau(\epsilon);\epsilon) = S_i(0;0) + \tau(\epsilon)\frac{dS_i}{dt}(0;0) + \sigma_i(\epsilon) + O(\epsilon^2) \quad \forall i \in \{1,..,2n\}. \tag{F.258}$$

Now, when $\sigma_i(\epsilon) = 0$ for all i, it must be the case that $\rho_i(\epsilon) = 0$ for all i as the final size is unchanged (as no vaccination has taken place). Thus, in this case, (F.253) can be linearised to give

$$\begin{aligned} O(\epsilon^2) = \tau(\epsilon)\Bigg[&\frac{dS_i}{dt}(0;0) + \frac{dI_i}{dt}(0;0) \\ &+ \frac{dR_i}{dt}(0;0)e^{-\sum_{j=1}^{2n}\frac{\beta'_{ij}}{\mu'_j}R_j(\infty;0)}\left(-\frac{dS_i}{dt}(0;0) + S_i(0;0)\sum_{j=1}^{2n}\frac{\beta'_{ij}}{\mu'_j}\frac{dR_j}{dt}(0;0)\right)\Bigg]. \end{aligned} \tag{F.259}$$

Note that this equality does indeed hold, as in the no vaccination case

$$\frac{dS_i}{dt}(0;0) + \frac{dI_i}{dt}(0;0) + \frac{dR_i}{dt}(0;0) = 0 \tag{F.260}$$

is the conservation of population law, while

$$-\frac{dS_i}{dt}(0;0) + S_i(0;0)\sum_{j=1}^{2n}\frac{\beta'_{ij}}{\mu'_j}\frac{dR_j}{dt}(0;0) = -\frac{dS_i}{dt}(0;0) + S_i(0;0)\sum_{j=1}^{2n}\beta'_{ij}I_j(0;0) = 0. \tag{F.261}$$

This means that, for non-zero σ_i, all terms not dependent on σ_i or ρ_i cancel and so the linearisation becomes

$$\begin{aligned}\rho_i = \sigma_i - \sigma_i e^{-\sum_{j=1}^{2n}\frac{\beta'_{ij}}{\mu'_j}R_j(\infty;0)} - S_i(0;0)e^{-\sum_{j=1}^{2n}\frac{\beta'_{ij}}{\mu'_j}R_j(\infty;0)}\sum_{j=1}^{2n}\frac{\beta'_{ij}}{\mu'_j}\rho_j \\ + o(\sigma_i) + \sum_{j=1}^{2n} o(\rho_j) + O(\epsilon^2)\end{aligned} \tag{F.262}$$

and so

$$\sigma_i = \frac{\rho_i + S_i(0;0)e^{-\sum_{j=1}^{2n}\frac{\beta'_{ij}}{\mu'_j}R_j(\infty;0)}\left(\sum_{j=1}^{2n}\frac{\beta'_{ij}}{\mu'_j}\rho_j\right) + o(\sigma_i) + \sum_{j=1}^{2n} o(\rho_j) + O(\epsilon^2)}{1 - e^{-\sum_{j=1}^{2n}\frac{\beta'_{ij}}{\mu'_j}R_j(\infty;0)}} \tag{F.263}$$

as required.

F.3.2 Proposition F.5

The result of Proposition F.4 can be written as a system of equations for vectors $\boldsymbol{\sigma}$ and $\boldsymbol{\rho}$

$$\boldsymbol{\sigma} = \boldsymbol{Q}\boldsymbol{\rho} + o(\boldsymbol{\sigma}) + \sum_{j=1}^{2n} o(\rho_j) + O(\epsilon^2) \tag{F.264}$$

for some matrix $\boldsymbol{Q}$ with non-zero determinant by assumption. However, it is important to establish the dominant balance in these equations, which is done through the following proposition, another result that the authors believe is novel to the literature.

Proposition F.5

$$\rho_i(\epsilon) = O(\epsilon) \quad \forall i \in \{1, ..., 2n\}. \tag{F.265}$$

Proof: Suppose that this does not hold. Thus, there must be some sequence ϵ_m such that, for some i

$$\lim_{m\to\infty}\left(\frac{\rho_i(\epsilon_m)}{\epsilon_m}\right)=\infty \quad \text{and} \quad \lim_{m\to\infty}(\epsilon_m)=0. \tag{F.266}$$

Define $J^*(\epsilon)$ such that

$$J^*(\epsilon)=\operatorname{argmax}\left\{|\rho_j(\epsilon)|:j\in\{1,...,2n\}\right\}. \tag{F.267}$$

Now, by the finiteness of $\{1,...,2n\}$, there exists some subsequence ϵ_{m_k} and some fixed $J\in\{1,...,2n\}$ such that

$$J^*(\epsilon_{m_k})=J \quad \forall k. \tag{F.268}$$

For notational convenience, assume that the original sequence ϵ_m has this property. Note that

$$\lim_{m\to\infty}\left(\frac{\sigma_j(\epsilon_m)}{\rho_J(\epsilon_m)}\right)=\lim_{m\to\infty}\left(\frac{\sigma_j(\epsilon_m)}{\epsilon_m}\times\frac{\epsilon_m}{\rho_J(\epsilon_m)}\right)=0, \tag{F.269}$$

as $\sigma_j(\epsilon)=O(\epsilon)$ and $\epsilon=o(\rho_i(\epsilon))\leq o(\rho_J(\epsilon))$. Moreover,

$$\lim_{m\to\infty}\left(\frac{O(\epsilon_m^2)}{\rho_J(\epsilon_m)}\right)=\lim_{m\to\infty}\left(\epsilon_m\times\frac{O(\epsilon_m)}{\rho_J(\epsilon_m)}\right)=0, \tag{F.270}$$

$$\lim_{m\to\infty}\left(\frac{o(\sigma_j(\epsilon_m)))}{\rho_J(\epsilon_m)}\right)=\lim_{m\to\infty}\left(o(1)\times\frac{\sigma_j(\epsilon_m)}{\rho_J(\epsilon_m)}\right)=0 \tag{F.271}$$

and

$$\left|\lim_{m\to\infty}\left(\frac{o(\rho_j(\epsilon_m))}{\rho_J(\epsilon_m)}\right)\right|\leq\lim_{m\to\infty}\left(\left|\frac{o(\rho_j(\epsilon_m))}{\rho_j(\epsilon_m)}\right|\right)=0. \tag{F.272}$$

Note that there is some abuse of notation in these calculations, but, for example, an $O(\epsilon^2)$ term in the limit represents any function which is $O(\epsilon^2)$. Thus, dividing (F.264) by $\rho_J(\epsilon_m)$ and taking m to ∞ shows that

$$\lim_{m\to\infty}\left(\frac{\boldsymbol{Q}\boldsymbol{\rho}}{\rho_J(\epsilon_m)}\right)=\mathbf{0}. \tag{F.273}$$

Define

$$\hat{\boldsymbol{\rho}}(\epsilon):=\frac{\boldsymbol{\rho}(\epsilon)}{\sum_{j=1}^{2n}|\rho_j(\epsilon)|} \tag{F.274}$$

and note that

$$\left|\left(\frac{\sum_{j=1}^{2n}|\rho_j(\epsilon_m)|}{\rho_J(\epsilon_m)}\right)\right|\in[1,2n] \tag{F.275}$$

and thus remains finite and non-zero. Thus,

$$\mathbf{0} = \lim_{m\to\infty}\left(\frac{\boldsymbol{Q}\boldsymbol{\rho}}{\rho_J(\epsilon_m)}\right) \tag{F.276}$$

$$= \lim_{m\to\infty}\left(\frac{\sum_{j=1}^{2n}|\rho_j(\epsilon_m)|}{\rho_J(\epsilon_m)} \times \boldsymbol{Q}\hat{\boldsymbol{\rho}}(\epsilon_m)\right), \tag{F.277}$$

which means

$$\mathbf{0} = \lim_{m\to\infty}\left(\boldsymbol{Q}\hat{\boldsymbol{\rho}}(\epsilon_m)\right). \tag{F.278}$$

However, note that

$$\sum_{j=1}^{2n}|\hat{\rho}_i(\epsilon)| = 1 \tag{F.279}$$

and hence the sequence $\hat{\boldsymbol{\rho}}$ is bounded. Thus, by the Bolzano-Weierstrass Theorem, there must be some subsequence m_k such that $\lim_{k\to\infty}(\hat{\boldsymbol{\rho}}(\epsilon_{m_k}))$ exists and is equal to some $\boldsymbol{\rho}^*$ where

$$\sum_{j=1}^{2n}|\rho_j^*| = 1. \tag{F.280}$$

However, then, by continuity and the fact that $\boldsymbol{Q}$ is invertible,

$$\boldsymbol{Q}\boldsymbol{\rho}^* = \mathbf{0} \Rightarrow \boldsymbol{\rho}^* = \mathbf{0} \tag{F.281}$$

which is a contradiction to (F.280) as required. Thus, it must be the case that $\rho(\epsilon) = O(\epsilon)$

F.3.3 Theorem 8.3

Combining Proposition F.5 with the fact that $\sigma_i = O(\epsilon)$ means that (F.264) can be written as

$$\boldsymbol{\sigma} = \boldsymbol{Q}\boldsymbol{\rho} + o(\epsilon). \tag{F.282}$$

Thus, one can multiply the equation by $\boldsymbol{Q}^{-1}$ to get

$$\boldsymbol{\rho} = \boldsymbol{Q}^{-1}\boldsymbol{\sigma} + o(\epsilon). \tag{F.283}$$

Hence, given vectors $\boldsymbol{p}$ and $\boldsymbol{q}$ where

$$\boldsymbol{p}_i := p_i \quad \text{and} \quad \boldsymbol{q}_i = p_i\kappa_i \quad \forall i \in \{1, ..., n\}, \tag{F.284}$$

the change to the objective function is given by

$$(\boldsymbol{p},\boldsymbol{q})^T\boldsymbol{\rho} = (\boldsymbol{p},\boldsymbol{q})^T\left[\boldsymbol{Q}^{-1}\boldsymbol{\sigma} + o(\epsilon)\right] \tag{F.285}$$

$$:= \boldsymbol{x}^T\boldsymbol{\sigma} + o(\epsilon). \tag{F.286}$$

Now, note that, for $i \in \{1,...,n\}$,

$$\sigma_i = -\frac{S_i(0;0)W_i(\tau(\epsilon);\epsilon)}{N_i} \tag{F.287}$$

while, for $i \in \{n+1,...,2n\}$

$$\sigma_i = -\sigma_{i-n}. \tag{F.288}$$

Hence, one can write (F.286) as

$$(\boldsymbol{p}^T,\boldsymbol{q}^T)\boldsymbol{\rho} = \boldsymbol{y}^T\boldsymbol{W}(\tau(\epsilon);\epsilon) + o(\epsilon), \tag{F.289}$$

where

$$\boldsymbol{y} = \frac{S_i(0;0)}{N_i}\left[-(x_1,....x_n)^T + (x_{n+1},...,x_{2n})^T\right], \tag{F.290}$$

as required by Theorem 8.3. The only restriction is that all the W_i are non-negative and that

$$\sum_{i=1}^{n} W_i(\tau(\epsilon);\epsilon) = \epsilon \tag{F.291}$$

and so the optimisation problem becomes

$$\min\{\boldsymbol{y}^T\boldsymbol{w} : \boldsymbol{w} \geq \boldsymbol{0} \quad \text{and} \quad \sum_{i=1}^{n} w_i = \epsilon\}. \tag{F.292}$$

Now, by Theorem 7.2 it must be the case that the objective function is non-increasing in $\boldsymbol{w}$. Thus, in particular, one must have

$$\boldsymbol{y} \leq \boldsymbol{0} \tag{F.293}$$

as otherwise, if $y_i > 0$ then setting $\boldsymbol{w} = \epsilon\boldsymbol{e}_i$ (where $\boldsymbol{e}_i$ is the ith canonical basis vector) means that

$$H(\boldsymbol{U}(t;\epsilon)) = H(\boldsymbol{U}(t;0)) + y_i\epsilon + o(\epsilon) \tag{F.294}$$

and so, for sufficiently small ϵ,

$$H(\boldsymbol{U}(t;\epsilon)) > H(\boldsymbol{U}(t;0)) \tag{F.295}$$

which is a contradiction. Hence, $\boldsymbol{y} \leq \mathbf{0}$ which means that the optimisation problem is an example of a continuous knapsack problem and one can readily see that a solution given is by

$$w_i^* = \begin{cases} \epsilon & \text{if } i = \min\{y_i\} \\ 0 & \text{otherwise} \end{cases}. \tag{F.296}$$

As this minimum is unique by assumption, this is the unique leading order optimal solution to the optimisation problem.

A technical note is that this only proves the form of the optimal solution to leading order. Indeed, if

$$w_i = w_i^* + o(\epsilon), \tag{F.297}$$

then the optimal objective value is unchanged to leading order. Hence, this restriction is given in the statement of the theorem (although in practice is unimportant).

F.4 Supplementary Lemmas

F.4.1 Lemma F.6

Lemma F.6 *Define the set of functions*

$$\mathcal{F} := \left\{ S_i(t;\epsilon), I_i(t;\epsilon), R_i(t;\epsilon), S_i^V(t;\epsilon), I_i^V(t;\epsilon), R_i^V(t;\epsilon) : i \in \{1,...,n\}, \quad \epsilon, t \geq 0 \right\}, \tag{F.298}$$

where for each fixed ϵ, these functions solve the model equations with parameters

$$\mathcal{P} = \left\{ \beta_{ij}^{\alpha}(\epsilon), \mu_i^{\gamma}(\epsilon) : i,j \in \{1,...,n\}, \quad \alpha \in \{1,2,3,4\}, \quad \gamma \in \{1,2\} \quad \textit{and} \quad \epsilon \geq 0 \right\}, \tag{F.299}$$

initial conditions

$$\mathcal{I} = \left\{ f(0;\epsilon) : i \in \{1,...,n\}, \quad f \in \mathcal{F} \quad \textit{and} \quad \epsilon \geq 0 \right\} \tag{F.300}$$

and vaccination policy $\boldsymbol{U}(t;\epsilon)$. Suppose further that the population sizes are independent of ϵ, except in group 1 where $N_1(\epsilon)$ satisfies

$$|N_1(\epsilon) - N_1(0)| \leq \epsilon \quad \textit{and} \quad \frac{S_1(0;\epsilon)}{N_1} = \sigma \tag{F.301}$$

for some constant σ.

Suppose that

$$|p(\epsilon) - p(0)| \leq \epsilon \quad \forall p \in \mathcal{P}, \tag{F.302}$$

$$|f_i(0;\epsilon) - f_i(0;0)| \leq \epsilon \quad \forall f \in \mathcal{F} \tag{F.303}$$

and that

$$|W_i(t,\epsilon) - W_i(t,0)| < \epsilon \quad \forall t \geq 0. \tag{F.304}$$

Moreover, suppose that for each $i \in \{1, ..., n\}$ *and* $\epsilon \geq 0$,

$$U_i(s;\epsilon) \geq 0 \quad \textit{and} \quad \int_0^t U_i(s;\epsilon)ds \leq N_i \quad \forall t \geq 0. \tag{F.305}$$

Then, for each $\delta > 0$ *and each* $T > 0$ *there exists some* $\eta > 0$ *(that may depend on* T *and* δ*) such that*

$$\epsilon \in (0,\eta) \Rightarrow |f(t;\epsilon) - f(t;0)| < \delta \quad \forall f \in \mathcal{F} \quad \textit{and} \quad \forall t \in [0,T]. \tag{F.306}$$

Proof: An almost identical result was proved in Lemma E.22 from Paper V with the only exception being that N_1 can vary in this example. However, note that by replacing $\frac{S_1(0;\epsilon)}{N_1(\epsilon)}$ with σ, this lemma can be proved identically.

F.4.2 Lemma F.7

The following lemma is a new result, proved using similar techniques to results in [50] such as Lemma E.20.

Lemma F.7 *Suppose that* $i \in \Pi$, *with* Π *defined as in Lemma E.20. Then, for* $t > 0$,

$$I_i^V(t) = 0 \Rightarrow S_i^V(t)\beta_{ji}^3 = S_i^V(t)\beta_{ji}^4 = 0 \quad \forall j \in \Pi. \tag{F.307}$$

Proof: Suppose that there exists some t and some $i, j \in \Pi$ such that

$$S_i^V(t)\beta_{ji}^3 > 0 \quad \text{and} \quad I_i^V(t) = 0. \tag{F.308}$$

Then, by continuity, there exists some $a < t$ such that

$$S_i^V(s)\beta_{ji}^3 > 0 \quad \forall s \in (a,t). \tag{F.309}$$

Moreover, by Lemma E.19, it is necessary that

$$I_i^V(s) = 0 \quad \forall s \in (a,t), \tag{F.310}$$

while, by Lemma E.20

$$I_j(t) > 0 \quad \forall s \in (a,t) \tag{F.311}$$

and hence (using the fact that $I_i^V(s) = 0 \quad \forall s \in (a,t)$)

$$\frac{dI_i^V}{dt} \geq S_i^V(s)\beta_{ji}^3 I_j(t) > 0 \quad \forall s \in (a,t) \tag{F.312}$$

and so

$$I_i^V(t) > I_i^V(a) = 0, \tag{F.313}$$

which is a contradiction as required. The final equality then follows as $\beta_{ji}^3 \geq \beta_{ji}^4 \geq 0$.

F.4.3 Lemma F.8

The following result extends the main theorem from [50] in a similar way to Proposition F.3 to provide an additional inequality on the objective values from the optimal vaccination problem.

Lemma F.8 *Suppose that the disease trajectories $\boldsymbol{S}$ and $\tilde{\boldsymbol{S}}$ are given by the same model equations, parameters, vaccination policy $\boldsymbol{U}$ and initial conditions except for the fact that*

$$S_1^V(0) > \tilde{S}_1^V(0). \tag{F.314}$$

Then, if the objective functions are denoted by H and $\tilde{H}$ for the two policies,

$$H(\boldsymbol{U}) \geq \tilde{H}(\boldsymbol{U}). \tag{F.315}$$

Proof: Define a new disease model, denoted by hats where a new group $(n+1)$ is added in such that its unvaccinated compartments behave like the vaccinated compartments of group 1 and its vaccinated compartments are perfectly immune from the disease. That is,

$$\hat{\beta}_{(n+1)j}^1 = \beta_{1j}^3, \quad \hat{\beta}_{(n+1)j}^2 = \beta_{1j}^4, \quad \text{and} \quad \hat{\beta}_{(n+1)j}^3 = \hat{\beta}_{(n+1)j}^4 = 0 \quad \forall j \in \{1,...n\}, \tag{F.316}$$

$$\hat{\beta}_{j(n+1)}^1 = \beta_{j1}^3 \quad \hat{\beta}_{j(n+1)}^2 = \beta_{j1}^4 \quad \text{and} \quad \hat{\beta}_{j(n+1)}^3 = \hat{\beta}_{j(n+1)}^4 = 0 \quad \forall j \in \{1,...,n\}, \tag{F.317}$$

$$\beta_{(n+1)(n+1)}^{\alpha} = 0 \quad \forall \alpha \in \{1,2,3,4\} \tag{F.318}$$

and

$$\hat{\mu}^1_{n+1} = \mu^2_1 \quad \text{and} \quad \hat{\mu}^2_{n+1} = 1. \tag{F.319}$$

Suppose further that all other parameter values are identical, and that the only differences in the initial conditions is that

$$\hat{S}^V_1(0) = \tilde{S}^V_1(0) \quad \text{and} \quad S_{n+1}(0) = S^V_1(0) - \tilde{S}^V_1(0) > 0. \tag{F.320}$$

Then, note that

$$\begin{aligned}\frac{d(\hat{S}^V_1 + \hat{S}_{n+1})}{dt} &= -\sum_{j=1}^{n+1}\left[\hat{S}^V_1(\hat{\beta}^3_{1j}\hat{I}_j + \hat{\beta}^4_{1j}\hat{I}^V_j) + \hat{S}_{n+1}(\hat{\beta}^1_{(n+1)j}\hat{I}_j + \hat{\beta}^2_{(n+1)j}\hat{I}^V_j)\right]\cdots \\ &\cdots - \frac{\hat{S}_{n+1}\hat{U}_{n+1}}{\hat{N}_{n+1} - \hat{W}_{n+1}} \\ &= -(\hat{S}^V_1 + \hat{S}_{n+1})\sum_{j=1}^{n+1}\left[\hat{\beta}^3_{1j}\hat{I}_j + \hat{\beta}^4_{1j}\hat{I}^V_j\right] - \frac{\hat{S}_{n+1}\hat{U}_{n+1}}{\hat{N}_{n+1} - \hat{W}_{n+1}}.\end{aligned} \tag{F.321}$$

Moreover, for $i \neq 1$

$$\frac{d}{dt}(\hat{S}_i) = -\hat{S}_i\sum_{j=1}^{n+1}\left[\beta^1_{ij}\hat{I}_i + \beta^2_{ij}\hat{I}^V_i\right] - \frac{\hat{S}_i\hat{U}_i}{\hat{N}_i - \hat{W}_i} \tag{F.322}$$

$$= -\hat{S}_i\left(\sum_{j=2}^{n}\left[\beta^1_{ij}\hat{I}_i + \beta^2_{ij}\hat{I}^V_i\right] + \beta^1_{ij}\hat{I}_1 + \beta^2_{ij}(\hat{I}^V_1 + \hat{I}_{n+1})\right) - \frac{\hat{S}_i\hat{U}_i}{\hat{N}_i - \hat{W}_i}. \tag{F.323}$$

Thus, with similar calculations for $\hat{I}$, $\hat{I}^V$, $\hat{R}$ and $\hat{R}^V$, by the initial conditions and by the uniqueness of solution, in the case that $\hat{U}_{n+1} = 0$,

$$\hat{S}^V_1 + \hat{S}_{n+1} = S^V_1 \quad \hat{I}^V_1 + \hat{I}_{n+1} = I^V_1 \quad \text{and} \quad \hat{R}^V_1 + \hat{R}_{n+1} = R^V_1. \tag{F.324}$$

Thus, setting

$$p_{n+1} = p_1\kappa_1, \tag{F.325}$$

this means that

$$\hat{H}(\hat{\boldsymbol{U}}) = H(\boldsymbol{U}) \tag{F.326}$$

for any $\hat{\boldsymbol{U}}$ such that $\hat{U}_{n+1} = 0$ and $\hat{U}_i = U_i$ for any $i \neq n$.

Now, define a vaccination policy $\hat{\boldsymbol{U}}^*(t;\Delta)$ such that

$$\hat{U}_i^*(t;\Delta) = \hat{U}_i(t) \quad \forall t \geq 0 \quad \text{and} \quad i \neq n+1 \tag{F.327}$$

and

$$\hat{U}_{n+1}^*(t;\Delta) = \begin{cases} \frac{1}{\Delta}\left(S_1^V(0) - \tilde{S}_1^V(0)\right) & \text{if } t \leq \Delta \\ 0 & \text{otherwise} \end{cases}. \tag{F.328}$$

Then, this means that

$$\hat{S}_{n+1}(\Delta;\Delta) = 0 \quad \text{and} \quad \hat{S}_{n+1}^V(\Delta;\Delta) = S_1^V(0) - \tilde{S}_1^V(0) + O(\Delta) \tag{F.329}$$

while all other variable values at time Δ differ by at most $O(\Delta)$ from their initial values. Thus, define by an overbar the model given by the initial conditions which are the same as those in the hat model, but with

$$\overline{S}_{n+1}(0) = 0 \quad \text{and} \quad \overline{S}_{n+1}^V = S_1^V(0) - \tilde{S}_1^V(0). \tag{F.330}$$

Suppose also that the vaccination policy in this case is equal to $\boldsymbol{U}$, which is the pointwise limit of the vaccination policy $\hat{\boldsymbol{U}}^*(t;\Delta)$ (for $t > 0$). Then, using Proposition F.2, by considering the values of the variables $\hat{f}$ at time Δ to be the initial conditions, one finds that for any finite time t,

$$\lim_{\Delta \to 0} (\hat{H}(\boldsymbol{U}^*(t;\Delta))) = \overline{H}(\boldsymbol{U}). \tag{F.331}$$

Note this holds as it is assumed that $\boldsymbol{U}$ is bounded and so

$$|W_i(t+\Delta;\Delta) - W_i(\Delta;\Delta) - W_i(t)| = O(\Delta). \tag{F.332}$$

Moreover, note that the only difference between the bar model and the tilde model is in group $(n+1)$. However, by the fact that $\beta_{ij}^3 = \beta_{ij}^4 = 0$ if $(n+1) \in \{i,j\}$, the value of $\overline{S}_{n+1}^V$ is constant and the other variable values are independent of it. Thus, by the uniqueness of solution, this means that

$$\overline{H}(\boldsymbol{U}) = \tilde{H}(\boldsymbol{U}). \tag{F.333}$$

Finally, note that by Theorem 7.1, it must be necessary that for any $\Delta > 0$

$$\hat{H}(\boldsymbol{U}(t;\Delta)) \leq \hat{H}(\boldsymbol{U}(t;\infty)) = H(\boldsymbol{U}), \tag{F.334}$$

where $\Delta = \infty$ corresponds to no vaccination taking place in group $(n+1)$ (and hence the original objective function H is recovered). Thus,

$$\tilde{H}(\boldsymbol{U}) \leq H(\boldsymbol{U}), \tag{F.335}$$

as required.

Appendix - Other Work

This appendix summarises the other papers and preprints which I have published during my DPhil.

The paper "Sherlock - A flexible, low-resource tool for processing camera-trapping images" [62], published in *Methods in Ecology and Evolution*, describes a new tool for classifying images from camera traps. It has particular use as a pre-processing tool, as it is able to accurately remove a substantial proportion of empty images, reducing the need for either manual classification or more computationally expensive classification models.

The paper "Analysis of a Double Poisson Model for Predicting Football Results in Euro 2020" [63], published in *PLoS ONE*, provides a mixture of analytic and practical results on the use of the common double Poisson model in football match prediction. It then applies this methodology to the Euro 2020 football tournament and analyses its performance.

The paper "Optimal loading of hydrogel-based drug-delivery systems", [509], in collaboration with Matthew Hennessy from the University of Bristol, published in *Applied Mathematical Modelling*, provides a novel framework for approximating desired drug-release profiles from hydrogel-based delivery systems. This was an extension of my MMath dissertation submitted in 2021. This increases the flexibility of current hydrogels and should reduce the number of new hydrogels that need to be developed - an expensive and time-consuming process.

The preprint "Continuous football player tracking from discrete broadcast data" [61], develops a novel algorithm for approximating continuous player trajectories from discrete, incomplete, unlabelled observations. This has the potential to expand the reach of continuous tracking data to a wide range of professional and semi-professional football clubs.

Finally, I also made a key contribution to the paper "Unifying incidence and prevalence under a time-varying general branching process" [60], published in the *Journal of Mathematical Biology*, proving that the renewal equation for mean prevalence derived under the novel framework in this paper yields the same solutions as the mean prevalence in the standard branching process framework.

The article "Barely a passing resemblance: why women's football stands out from the crowd", published in *Significance* [510] compares match event data from the 2022 Men's and 2023 Women's World Cup. By examining the risk and threat of passes, it shows that the women's game has substantially more focus on attacking, leading to much more engaging football.