Molecular Biology and Evolution, 2026, 43, 1–14
https://doi.org/10.1093/molbev/msag268
Published: 6 November 2026
Methods

Simpler is Not Always Better: Phylodynamic Misspecification and Deep-Learning Corrections

Ruopeng Xie [1], **Olivier Gascuel** [2], **and Anna Zhukova** [3]*

[1]Department of Biology, University of Oxford, Oxford, OX1 3EL, UK
[2]Institut de Systématique, Evolution, Biodiversité (ISYEB), URM 7205 CNRS, Muséum National d'Histoire Naturelle, SU, EPHE & UA, Paris, 75005, France
[3]Modelling of Pathogens: Evolution, Adaptation and Spread, Institut Pasteur, Université Paris Cité, Institut de Biologie de l'ENS (IBENS), École Normale Supérieure, CNRS, INSERM, Université PSL, Paris, 75005, France
***Corresponding author**: E-mail: anna.zhukova@pasteur.fr
Associate editor: Miguel Arenas

Abstract

Phylodynamics bridges the gap between epidemiology and pathogen genetic data. The birth–death (BD) model and its extensions serve to infer the average number of secondary infections R_e and the infection duration *d* from time-scaled pathogen phylogenies. Moreover, more complex models add extra parameters, such as the average length of the incubation period, the proportion of superspreaders in the infected population, or the speed-up in detection after contact-tracing. However, these additional parameters come at an important computational cost: While the simplest, BD, model, has a closed-form solution, its extensions do not and require numerical methods for their likelihood computation. This leads to increased computational times and potential numerical errors. Therefore, the BD model remains the favorite researchers' choice for real dataset analyses, and is often applied even in cases where more complex epidemiological aspects are present. We investigated, using simulations, how model misspecification influences inference of R_e and *d* in the phylodynamic framework. We showed that the use of models not accounting for various epidemiological aspects leads to bias. In particular the simplest, BD, estimator tends to underestimate R_e in the presence of super-spreading, incubation or contact-tracing, which might be dangerous from the public health perspective. In contrast, deep-learning-based estimators for complex models, accounting for multiple epidemiological factors, performed well in our simulations both on the data where those factors were present and where they were absent. This motivates the use of complex epidemiologically realistic estimators, whose design has recently become possible thanks to deep learning.

Keywords phylodynamics, epidemiology, model misspecification, estimation bias, deep learning

Introduction

Many pathogens, especially viruses, evolve rapidly and accumulate mutations in their genomes between transmissions. Phylodynamics (Grenfell et al. 2004; Volz et al. 2013) leverages these mutations to estimate epidemiological parameters. Phylodynamic models bridge the gap between traditional epidemiology and sequence data, and permit estimation of such parameters as the effective reproduction number, R_e, from the topology and time-measured branch lengths of pathogen transmission trees. These trees can be approximated by pathogen phylogenies – i.e. genealogies of the pathogen population (Kozlov et al. 2019; Minh et al. 2020) — time-scaled (To et al. 2016) using the sampling dates of pathogen genomes collected over time. The likelihood of a given transmission tree under phylodynamic model parameters can be represented with differential equations. Model parameter values can then be found by, for instance, maximizing the tree likelihood. The use of the transmission structure empowers the phylodynamic approach with respect to classical epidemiological methods. The latter make inferences from incidence curves and in the presence of imported cases might give appearance of elevated local transmission, while the former can separate local clusters from imported cases using phylogeographic methods (Lemey et al. 2009; Ishikawa et al. 2019; Vaughan et al. 2024).

In the era of rapidly growing pathogen genomic sequence data, the phylodynamic approach is especially promising. However its power to provide valuable insights for preventing epidemic spread comes at a computational cost. A closed form solution of the differential equations exists only for the most

Received: May 7, 2026. **Revised:** October 6, 2026. **Accepted:** October 13, 2026
© The Author(s) 2026. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

simple Birth–Death (BD) model with incomplete sampling (Stadler 2010), and its Skyline version BDSKY (Stadler et al. 2013), where parameter values can change in a piecewise- constant manner. The BD model is similar to the SIR (Susceptible-Infectious-Removed Hethcote 2000) model in classical epidemiology and models pathogen spread in a homogeneous population. However, unlike the SIR model, it only applies to epidemics in the exponential phase, with an unlimited number of susceptible individuals. Under this model, infected individuals I can transmit their pathogen further and eventually get removed. Removal corresponds to the moment when the individual cannot transmit anymore, for example because they healed, became aware of the infection and took action to avoid transmission, or died. For a fraction of infected individuals ρ their pathogen is sampled and sequenced at removal (and becomes available for phylodynamic analyses). Note that sampling therefore occurs through time (Stadler 2010). The BD model permits the estimation of the *effective reproduction number* R_e (i.e. expected number of secondary infections) and the average *infection duration* d (i.e. the time between the moment the individual got infected and the moment they stopped transmitting further). Note that this definition of d does not strictly correspond to the infection duration in the clinical sense (until the pathogen gets undetectable), but rather corresponds to the public health sense (until the pathogen cannot be transmitted further, e.g. because the person decided to self-isolate). The BDSKY model with its piecewise-constant rates further approximates the classical SIR model, as reduced transmission rates at later time-points can model decrease in the susceptible population. More complex versions of the BD model where transmission and removal rates can vary arbitrarily over time can further bridge the gap between the BD and the SIR models. However, not only no closed form solution might exist for their differential equations, but also as highlighted by Louca et al. (2021) under such BD models for any true epidemiological scenario there exists a myriad of alternative "congruent" scenarios that cannot be distinguished using phylogenetic data alone. Identifiability is restored when the number of hidden transmission events along branches is known, the information which can in some cases be inferred from the pathogen sequences (Dieselhorst and Stadler 2026). In the case of (piecewise-)constant BD(SKY) models, giving one of the model parameters as an input also restores identifiability (Stadler 2009; Stadler et al. 2013). This is often the sampling fraction, estimated as the proportion of sampled sequences among the total declared cases. Due to the relative simplicity and speed, the (piecewise-)constant version of the BD model has become widely adopted and remains the first choice for many pathogen analyses. The original BD and BDSKY publications (Stadler 2010; Stadler et al. 2013) have been cited more than a thousand times[1].

The BD model was extended to incorporate complex epidemiological aspects, for instance accounting for an incubation period, super-spreading events or contact-tracing of potentially infected recent partners (Stadler and Bonhoeffer 2013; Stadler et al. 2014; Kühnert et al. 2016; Zhukova and Gascuel 2025). For these models the differential equations need to be solved numerically. Numerical solution is challenging, time consuming and may lead to underflow errors (Scire et al. 2022; Voznica et al. 2022). This results in extensive use of simpler models and ignoring potentially important aspects of real-life epidemics (e.g. incubation). For instance for the phylodynamic analyses of SARS-CoV-2 pandemic in Europe (Hodcroft et al. 2021), New Zealand (Geoghegan et al. 2020), and Brazil (Resende et al. 2021), the implementation of BDSKY model in a Bayesian framework was used instead of the BD Exposed-Infectious (SKY) model, accounting for incubation (which takes 2-5 days (Walsh et al. 2020), i.e. an important time in the total infection duration). In the cases when the epidemiological question is simple (e.g. estimating R_e to assess whether the epidemic is contained ($R_e < 1$) and adjust the lockdown measures), a simpler model choice is indeed tempting (and for very large datasets until recently has been the only one possible).

To our knowledge, the consequences of choosing a simpler homogeneous phylodynamic model in a complex heterogeneous epidemiological setting has only been investigated for the incubation period (Park and Koelle 2026) and showed that not accounting for it leads to underestimation of R_e. Until recently, due to computational limitations, investigating model misspecification essentially meant identifying potential biases in estimates with the simple BD model on complex epidemic data. However, the recent arrival of deep learning (DL) to the phylodynamic domain (Voznica et al. 2022; Thompson et al. 2024; Landis and Thompson 2025; Perez and Gascuel 2026) enabled replacing complex likelihood computation with much simpler transmission tree simulation (for training of a DL estimator). Once trained, DL estimators permit almost instantaneous parameter inference. DL hence enabled parameter estimation under more complex models, and on larger datasets (Voznica et al. 2022). However, its practical applications have not been extensive so far and require ad-hoc implementations and time- and memory-consuming training phase. For instance, only few studies used DL-based estimators for SARS-CoV-2 analyses under more complex and epidemiologically relevant scenarios (e.g. with super-spreading Xie et al. 2024). This raises additional questions that we investigate in this study: Do we need DL estimators for complex models? In which settings do they perform better than simple ones?

In the rest of this article we describe a simulation study that addresses these questions. We generated transmission trees under a wide epidemiological parameter range, covering such epidemics as Ebola, SARS-CoV-2, flu and HIV. These trees varied in size from 200 to 5,000 tips, covering a variety of sequence dataset sizes typically available for such epidemics. The trees featured different combinations of presence or absence of epidemiological aspects potentially relevant to these epidemics (incubation, super-spreading and contact tracing). We estimated the epidemiological parameters on these trees using a simple BD estimator and more complex ones (based on deep-learning). We then compared the accuracy of different estimators for common parameters (such as the average number of secondary infections R_e and the infection time d), and, when applicable, the model-specific ones (such as the fraction of the total infection duration in the incubation, the fraction of the population who are super-spreaders, or the probability to contact-trace). We reported the identified biases and discussed the new opportunities that deep learning opens for phylodynamic analyses of complex epidemics. Finally, to show practical use, we applied our new estimators to Hong Kong's third COVID-19 epidemic wave.

[1] According to Google scholar (assessed on August 10, 2026).

Materials and methods

BD exposed-infectious model with super-spreading and contact tracing

The BD exposed-infectious model with super-spreading and contact tracing (BDEISS-CT) model (Fig. 1, Table 1) combines the Birth–Death Exposed Infectious (BDEI), Birth-Death with Super-Spreading (BDSS) and Birth-Death with Contact-Tracing (BD-CT) models (Stadler and Bonhoeffer 2013; Kühnert et al. 2016; Zhukova and Gascuel 2025). It models a heterogeneous host population of superspreaders and regular spreaders, infected with a pathogen that features an incubation period. Moreover, at the moment of removal (e.g. due to becoming aware of the infection and self-isolating), the infected individual might notify their contacts (person who infected or people who were infected by the individual in question). Such notification leads to reduced time until removal of the contacts, and hence reduced reproduction number R_e.

Mathematically speaking, this model has six possible states: exposed E (i.e. infected but not yet infectious), infectious regular spreader I, infectious superspreader S, as well as their contact-traced versions: E_C, I_C, and S_C. The regular spreaders (contact-traced or not, I_C or I) transmit the epidemic at a constant *transmission rate* λ, while the superspreaders (S or S_C) transmit it at a higher rate $X_S\lambda$, where $X_S > 1$ represents the *super-spreading transmission increase*. Note, that the model assumes an exponential epidemic phase, with no depletion of susceptible hosts (and hence no decrease in transmission rate over time). The recipient is initially in the exposed state E, and their pathogen needs to go through the incubation before being able to be transmitted further. The incubation is modeled via an *incubation rate* μ. The duration of the *incubation period* can then be calculated as $d_{inc} = 1/\mu$. It can also be expressed as a *fraction of incubation* in the total infection time: $f_E = d_{inc}/d$. The average infection duration d for this model is a sum of the incubation period d_{inc} and the *infectious time* d_{inf}, corresponding to the time interval between the moment the individual got infected and the moment they stop transmitting their pathogen further: $d = d_{inc} + d_{inf}$. Upon incubation a (*superspreader*) *fraction* f_S of individuals becomes superspreaders, while the rest become regular spreaders. If an individual gets contact-traced during the incubation period, their incubation rate does not change (is μ), but upon the incubation such an individual becomes either a contact-traced superspreader (S_C, with a probability f_S) or a contact-traced regular spreader (I_C, with a probability $1 - f_S$). Moreover, infected individuals may also get contact-traced after incubation, in which case their state (I or S) changes to the corresponding contact-traced one (I_C or S_C). The untraced infectious individuals (I and S) are removed (i.e. stop transmitting) at a *removal rate* ψ, while the contact-traced infectious individuals (I_C and S_C) are removed at a higher rate $X_C\psi$, where an $X_C > 1$ represents the *contact-traced removal speed-up*. Upon removal the individual's pathogen may be sampled and sequenced with a *sampling probability* ρ. The exposed individuals (E and E_C) need to incubate first for their pathogen to be detectable, hence their removal rate is zero. Moreover, upon removal (independently of sampling) an individual (in state I, S, I_C or S_C) may notify (all) their contacts with a *contact tracing probability* υ (drawn independently for each contact).

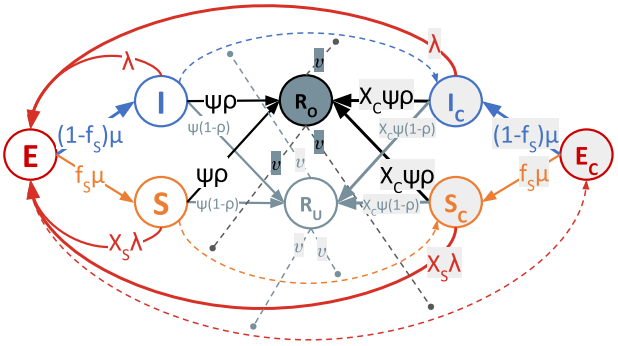


Figure 1 BDEISS-CT model. The model includes three main states, represented as circles – Exposed (E), Infectious regular spreader (I) and infectious Superspreader (S) – as well as their contact-traced versions: E_C, I_C and S_C. It also includes two Removed, non-infectious, states – observed (R_o) and unobserved (R_u). Straight solid lines represent state changes, either due to incubation (e.g. the line from E to I), or due to removal (e.g. the horizontal line from I to R_o due to sampling). Curved dashed lines represent state changes due to contact-tracing (e.g. from I to I_C), they are "catalyzed" (straight dashed lines) by the removal events. Curved solid lines represent transmissions (e.g. the line from I to E). The rates at which different events happen are shown on the corresponding lines (apart from contact-tracing, for which the mechanism is more complex, and only the probability to notify is shown).

Here by contacts of an individual i we mean the person who infected i and people infected by i.

The BDEISS-CT model can be described with the following 8 parameters: $\{R, d, \rho; f_E; f_S, X_S; \upsilon, X_C\}$ (see Table 1 for definitions). Setting $f_E = 0$ removes incubation (EI); setting $f_S = 0$ or $X_S = 1$ removes super-spreading (SS); setting $\upsilon = 0$ or $X_C = 1$ removes contact tracing (-CT). Figure 2 shows how the BD, BDEI, BDSS, BDEISS models and their contact-tracing versions (-CT) are nested within each other and the BDEISS-CT model.

Note that the BD(EI)(SS) models are compartmental and belong to a broader Multi-Type BD model family (Stadler and Bonhoeffer 2013; Kühnert et al. 2016). The epidemiological process under these models is memory-less, where the future events depend only on the current infected individual state, and the transmission tree branches are independent. In the Supplementary Material we explain how R_e and d can be calculated from the rate parameters for a general Multi-Type BD (Stadler and Bonhoeffer 2013; Kühnert et al. 2016) model. In the special case of the BDEISS model, these formulas transform to:

$$R_e = (1 + (X_S - 1)f_S)\lambda d_{inf} \quad (1)$$

$$d = d_{inc} + d_{inf}, \quad where\ d_{inc} = \frac{1}{\mu}, \quad d_{inf} = \frac{1}{\psi}. \quad (2)$$

The CT extensions of these models add the dependency between the notifier branch and the contact branches, and keep the memory of who infected whom. This complicates the calculations, and makes the formula for d_{inf} – and as a consequence the formulas for d and R_e – not easily derivable. We know only that the upper bound for d_{inf} is $1/\psi$ (no CT) and its lower bound is $1/(X_C\psi)$ (notification systematically during the incubation period). Nevertheless, we can extract the values of these parameters

Table 1 Definitions and notations.

Name	Notation	Definition
General concepts:		
Incubation (EI)	–	individual's pathogen is in a latent phase right after the infection, so that the individual cannot transmit until the pathogen incubates
Super-spreading (SS)	–	difference in transmission patterns between individual groups where superspreaders transmit more than regular spreaders
Contact-tracing (CT)	–	possibility to notify contacts (persons infected by or the person who infected the individual in question) leading to their awareness of the infection and stopping further transmission faster than untraced individuals
BDEISS-CT model states:		
exposed state	E	untraced infected individual who cannot yet transmit as their pathogen is in a latent phase
regular infectious state	I	untraced infected individual who can transmit at a standard rate
superspreader infectious state	S	untraced infected individual who can transmit at a higher rate than I
notified exposed state	E_C	contact-traced infected individual who cannot yet transmit as their pathogen is in a latent phase
notified regular infectious state	I_C	contact-traced infected individual who can transmit at a standard rate, and has a higher removal rate (i.e. stops transmitting faster) than I
notified superspreader infectious state	S_C	contact-traced infected individual who can transmit at a higher rate than I_C, and has a higher removal rate (i.e. stops transmitting faster) than S
observed removed state	R_o	previously infectious individual who stopped transmitting and whose pathogen got sampled (and sequenced)
unobserved removed state	R_u	previously infectious individual who stopped transmitting without sampling of their pathogen
BDEISS-CT epidemiological parameters:		
effective reproduction number	R_e	expected number of secondary infections per infected individual
average infection duration	d	average time between the infection and the moment the infected individual stops transmitting further
incubation period	$d_{inc} = 1/\mu$	time between the moment of infection and becoming infectious
infectious time	$d_{inf} = d - d_{inc}$	period when the infected individual can transmit their pathogen
incubation fraction	$f_E = d_{inc}/d$	fraction of infection time spent in incubation
sampling probability	ρ	fraction of removed individuals whose pathogen is sampled (and sequenced) at removal
superspreader fraction	f_S	fraction of superspreaders in the infectious population
super-spreading transmission increase	X_S	increase in transmission rate for superspreaders with respect to regular spreaders
contact tracing probability	υ	for a pair of a removed individual and their contact, probability to notify the contact at removal
contact-traced removal speed-up	X_C	increase in removal rate due to contact-tracing
BDEISS-CT rate parameters:		
regular transmission rate	λ	rate at which regular spreaders transmit their pathogens
superspreader transmission rate	$X_S\lambda$	rate at which superspreaders transmit their pathogens
incubation rate	μ	rate at which exposed individuals become infectious
standard removal rate	ψ	rate at which untraced infectious individuals stop being infectious
notified removal rate	$X_C\psi$	rate at which contact-traced infectious individuals stop being infectious

from the transmission tree simulations (see Supplementary Material for details).

Simulated datasets

We used treesimulator (Zhukova and Gascuel 2025) to generate transmission trees under the BDEISS-CT and its seven nested models shown in Fig. 2. For each model, we generated 1,000 trees for the test set and $2^{20} \approx 1{,}000{,}000$ trees for the training+validation set. For each tree the parameter values were drawn uniformly within the bounds listed in Table 2. Note that for model-specific parameters (such as the incubation fraction f_E, which is only applicable to models with incubation), the range for the training set was wider than for the test set (e.g. starting at $f_E = 0$ vs $f_E = 0.01$). This was done to produce model-specific trees in the test data (e.g. with distinguishable incubation period), while allowing for boundary cases in the training set (e.g.

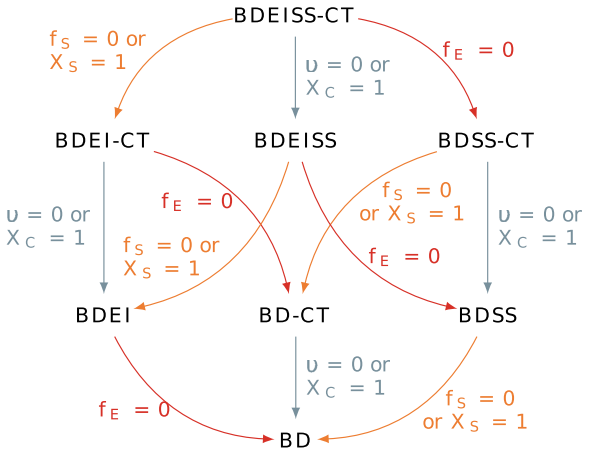


Figure 2 Nesting of BD(EI)(SS)(-CT) models. The arrows link more general models to their nested versions, with the fixed parameter shown along the arrow. The most general, BDEISS-CT, model has 8 parameters: $\Theta_{BDEISS-CT} = \{R, d, \rho, f_E, f_S, X_S, \upsilon, X_C\}$, while the most basic, BD, model has only three parameters: $\Theta_{BD} = \{R, d, \rho\}$.

Table 2 Ranges for parameter values for the trees in the training and test datasets. The BD(EI)(SS)(-CT) models in which each parameter is used are listed in the last column (see Fig. 2 for model nesting).

Parameter	Training set range	Test set range	Models
R_e	[1, 10)	[1, 10)	all
d	[1, 31)	[1, 31)	all
ρ	[0.05, 0.75)	[0.05, 0.75)	all
f_E	[0, 1)	[0.01, 1)	BDEI and its extensions
f_S	[0, 0.5)	[0.05, 0.5)	BDSS and its extensions
X_S	[1, 25)	[5, 25)	BDSS and its extensions
υ	[0, 0.75)	[0.05, 0.75)	BD-CT and its extensions
X_C	[1, 100)	[10, 100)	BD-CT and its extensions

trees without incubation by setting $f_E \approx 0$) and accommodating nesting of the models. The number of tips s^* was drawn uniformly between $s^{(0)} = 2{,}000$ and $S^{(0)} = 5{,}000$ (inclusive). Additionally, to analyze the impact of tree size on estimations, we generated three datasets with smaller tree sizes: $[s^{(1)} = 1{,}000, S^{(1)} = 2{,}000]$, $[s^{(2)} = 500, S^{(2)} = 1{,}000]$, $[s^{(3)} = 200, S^{(3)} = 500]$, where $s^{(i)}$ and $S^{(i)}$ correspond to the minimum and maximum size for the trees in the i-th dataset. The tree $\mathcal{T}_j^{(i)}$ in the dataset $i > 0$ was generated from the corresponding tree $\mathcal{T}_j^{(i-1)}$ by (1) drawing the target number of tips $s_j^* \in [s^{(i)}, S^{(i)}]$, and (2) pruning the tips sampled after the s_j^*-th sampling date from the tree $\mathcal{T}_j^{(i-1)}$.

We chose wide parameter ranges to cover a variety of epidemics. For instance, $R \in [1, 10)$ covers epidemics such as HIV-1 (with R_e between 1 and 2 Rasmussen et al. 2017; Voznica et al. 2022), influenza (with R_e also between 1 and 2 Biggerstaff et al. 2014), SARS-CoV-2 (with a potentially higher R_e, e.g. between 2 and 4 for its Omicron variant Viana et al. 2022), and Ebola (with estimates between 1 and 8 Nash et al. 2024). The infection time $d \in [1, 31)$ covers, on a yearly scale, such epidemics as HIV-1 B (whose average time of progression to AIDS in the absence of treatment is 10 years, and the time before treatment start is shorter). On a daily scale, it corresponds to [1 day, 1 month) and covers influenza (with several days of incubation and infectious time of around 1 week), SARS-CoV-2 (2-5 days of incubation and up to 10 days since the symptom onset Walsh et al. 2020) and Ebola (2-21 days of incubation plus 1–16 days of infectious time Velásquez et al. 2015).

Parameter estimators

We used our maximum-likelihood parameter estimator, bd (Zhukova and Gascuel 2025) for the BD model, to have a reference point for the best achievable performance (as the tree likelihood under the BD model has an analytical formula). This estimator used three different starting points (and kept the result with the highest likelihood), to avoid local maxima. We trained DL-based parameter estimators for all the models shown in Fig. 2, adapting the procedure described in Voznica et al. (2022). The estimator for the BD model was trained on a dataset of $N \approx 900{,}000$ training BD trees and $n \approx 100{,}000$ validation BD trees. For the other models we tried two approaches: the "pure" approach, where, like for BD, the estimator for a model X was trained on $N + n$ (training+validation) trees generated under the same model X; and the "mixed" approach, where the estimator for a model X was trained on a dataset containing trees generated under the model X and all its nested models (the same numbers $N + n$ of total trees equally distributed between different models). For instance, the mixed estimator for the BDEISS model was trained on a dataset of $N/4 + n/4$ BDEISS, $N/4 + n/4$ BDEI, $N/4 + n/4$ BDSS, and $N/4 + n/4$ BD trees. We tried the mixed approach to put more weight on boundary cases in complex models. For instance <1% of the pure BDEISS-CT training trees were BD-like (with f_E, f_S and υ close to zero) vs $\geq 12.5\%$ for the mixed case. As the result, pure estimators specialized less well to their nested models (see for example Table S4). For each model we trained one pure and one mixed DL estimator per simulated tree size range (see "Simulated datasets"). In the following, we report the results for mixed estimators on the largest trees (2,000–5,000 tips), unless otherwise stated.

As in Voznica et al. (2022), the trees were rescaled so that their average branch length became 1, and the target values of the infection duration d were rescaled accordingly. The predicted values of d were converted back to the original scale. We encoded the rescaled trees as vectors of 218 summary statistics, most of which were adapted from Voznica et al. (2022) and some were added to account for peculiarities of the new models, in particular contact-tracing (described in Supplementary Material). These statistics always included the value of the sampling probability ρ (needed for model identifiability). Their values were standardized (by removing the mean and scaling to unit variance, independently for each statistic, based on the training dataset for each estimator). The output values were standardized in the same way. Input features (tree summary statistics) were processed through a three-layer fully-connected network (64, 64, and 32 units) incorporating Layer Normalization, ELU activations, and L2 weight regularization. The output was represented as a Dense layer without activation of the size corresponding to the number of parameters pertinent to the estimator's epidemiological model. For instance for the BDEISS-CT estimator it contained 7×3 neurons: for the

three quantiles of R_e, d, f_E, f_S, X_S, υ, and X_C; while for the BD estimator it contained only 2×3 neurons: for the three quantiles of R_e and d. We used the Adam optimizer with the pinball loss function (Steinwart and Christmann 2011), allowing for estimation of the 0.025^{th}, 0.5^{th} and 0.975^{th} quantiles (i.e. the median and a 95%-confidence interval, CI) for each model parameter. To prevent optimization divergence, global L2 gradient norm clipping (clipnorm = 0.1) was applied to the Adam optimizer, rescaling any gradient vector whose total L2 norm exceeded 0.1 without altering update directions. We set up an adaptive learning rate schedule that monitored the validation loss and decayed the learning rate by 50% after five epochs without improvement, starting at 0.0003 and down to a floor of 10^{-7}. To enhance early training stability, a custom callback linearly scaled the optimizer's learning rate from near zero to its target value (0.0003) across the first five epochs. We set the maximal number of epochs to 1,000 (this value was never reached) and used early stopping (with a patience of 15 epochs, the warm-up epochs ignored) to avoid overfitting. The batch size was set to $2^{12} = 8{,}192$.

To enhance model stability against stochastic initialization, for each setting, five independent deep neural network instances were trained under identical architectural configurations, hyperparameter settings and on the same training datasets, differing solely by their initial random seed. A median ensemble estimator was constructed by computing the point-wise median prediction across the five base models for each sample and output quantile. Our evaluations showed that good cross-seed reproducibility was achieved between these estimators and that the median ensemble consistently filtered out single-seed optimization drift, keeping predictions centered within the optimal seed cluster (see the Supplementary Material and Table S2 for details).

Real pathogen data

We applied our estimators to the transmission trees reconstructed from real pathogen data, described below. We also checked whether these trees resembled the training data for each model by comparing the values of different summary statistics described in the Supplementary Material on the real data to their distributions on the training data. If they were more than five standard deviations away, we considered this to be an incompatibility between the real and training data, and an indication that the estimator's results should be treated with caution.

Hong Kong SARS-CoV-2 wave 3 tree

We applied our estimators to the time-scaled phylogenetic tree of wave 3 of SARS-CoV-2 epidemic in Hong Kong from Xie et al. (2024), representing its exponential phase from May 13 to August 1 2020. This tree contained 460 samples, corresponding to the sampling proportion $\rho = 0.238$. It was resolved with contact-tracing data and previously analyzed with the BDSS model (see more details on these data in Gu et al. 2022; Xie et al. 2024).

Results

The main question that we sought to answer in this study was: If one is only interested in a particular aspect of an epidemic, can one use a simpler model that does not take into account other epidemic aspects? For instance, can one reliably estimate the effective reproduction number R_e with the BD model, even if super-spreading is present in the epidemic? We were also interested in the opposite question: How does a complex estimator perform in a simpler setting? For instance, will a BDSS estimator (which accounts for super-spreading) correctly estimate the absence of super-spreading in a BD epidemic? Do we need to explicitly perform model selection?

To answer these questions we simulated transmission trees under eight epidemic settings representing all the combinations of presence and absence of super-spreading, incubation, and contact tracing (see "Simulated datasets" for more details). We estimated the epidemiological parameters on these trees with the estimators that ranged in complexity from BDEISS-CT (the most complex model, accounting for all of the above aspects) to BD (the most basic model, which only estimates the reproduction number R_e and the infection duration d). Finally we compared the errors and biases for common parameters obtained with different estimators. The results obtained on the 2,000–5,000-tip trees dataset and with estimators trained on a mix of the corresponding and its nested models are shown here: for R_e and d in Fig. 3, for the other parameters in Tables 3–5. The results for the smaller tree size datasets and estimators trained on matched models are shown in Tables S3–S16.

Observed trends

DL-based BD estimator performed as well as the maximum-likelihood (ML) one. DL-based BD estimator has errors, biases and CIs compatible with the ML estimator (Figs. 3 and 4, two left-most blue bars).

Using a simple estimator on complex data leads to biases in estimations, which overall increase with data complexity. For instance the average error for the R_e estimation with the maximum likelihood BD estimator was 3% on the BD data, it increased to 18% on the BDEI data, to 29% on BDEISS data, and finally to 30% on the BDEISS-CT data. The bias also increased, from 0% on the BD data to −12% on the BDEI data, to −29% on the BDEISS data, and finally to −30% on the BDEISS-CT data (Fig. 3, the left-most blue bar). The same trend is observed for the CI coverage: it dropped from 100% on the BD data to only 7% of trees on the BDEISS-CT data (Fig. 4, the left-most blue bar).

BD estimators tend to underestimate R_e when complex epidemic aspects are present (Fig. 3 right panel, two left-most blue bars). This is particularly dangerous from the public health perspective as it suggests that the epidemic is better contained than it really is.

Accounting for additional epidemic aspects even if they are absent increases CI width but not the error, as long as the estimator can infer their absence (e.g. output zero for the incubation or super-spreading fractions, f_E and f_S, or the CT probability υ). For instance, the errors and biases for the estimations of R_e and d on the BD data (first rows in Fig. 3 left panel) are within 3% for all the estimators), and the true R_e value is within their CIs for >95% of trees (first rows in Fig. 4 left panel). Tables 3–5 also show that in the absence of incubation, super-spreading, or contact-tracing the estimators for the correct or more general models correctly infer the values of

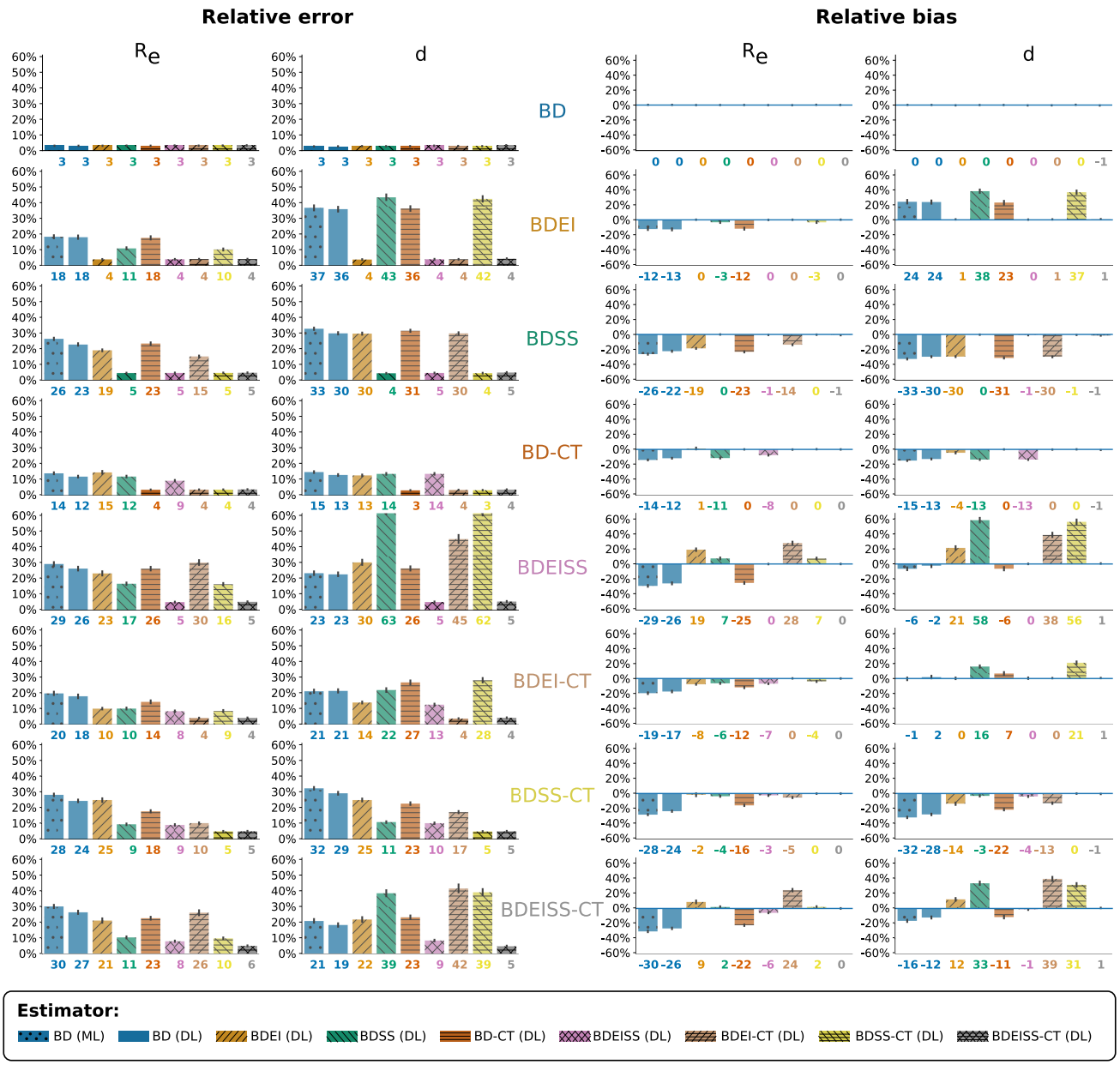


Figure 3 Relative percentage errors (left half) and biases (right half) for effective reproduction number R_e (first and third columns) and average infection time d (second and fourth columns). The errors and biases were estimated on datasets of 1,000 transmission trees with 2,000–5,000-tips generated under different flavors of the BD(EI)(SS)(-CT) model (corresponding to rows, labels shown in the middle). The bars correspond to different estimators (see the legend at the bottom). (ML) and dotted pattern indicate a maximum-likelihood estimator, while (DL) indicates deep-learning estimators. The DL estimators were trained on a mix of trees generated under the corresponding model and its nested models (e.g. BDEI and BD trees for the BDEI estimator). Estimators accounting for incubation have forward-slash pattern, those accounting for super-spreading have backward-slash one, and those accounting for contact-tracing have horizontal dashes. Bar heights correspond to the mean relative error/bias on 1,000 trees (also shown below each bar), while the whiskers correspond to 95% confidence intervals. The estimators that do not correspond or generalize the data model (e.g. BDEI estimator for BDSS data) are shown with semi-transparent colors. (In practice, they correspond to large relative error/bias bars, while the matching and opaque ones correspond to low, almost-zero, bars.)

correspondingly f_E, f_S, or υ close to zero, and their CIs include zero in >95% of cases. However, the CIs of more general estimators may be wider than those of model-specific ones (e.g. 12 for BDEISS-CT vs 10 for BDSS estimator, when estimating f_S on the BDSS data, see Table 4). Finally, the most complex estimator, BDEISS-CT, performs well at estimating all the parameters in all the settings (the last bar in Figs. 3, 4 and the last column of Tables 3–5).

Accounting for at least some of the epidemic aspects is usually better than not accounting for anything. For instance, the error and bias for R_e on BDSS-CT data were lower for the BDSS and BD-CT estimators (error: 9% and 18%, bias: −4% and −16%) than for the BD ones (24–28% and −24– −28%), while with the correct estimator (BDSS-CT) their values were the lowest (5% and 0%, row BDSS-CT in Fig. 3). Moreover, the BDSS and BD-CT CIs included the true R_e value for respectively 62% and 38% of trees, while the

BD-estimated CIs only included it for 7–16% of trees. The correct estimator's CIs (BDSS-CT) included the true R_e value for 96% of trees (row BDSS-CT in Fig. 4 left panel).

Unaccounted-for epidemic aspects might get confused with something else. In particular, *incubation for super-spreading, and vice versa* (Table 4, rows BDEI and BDEI-CT, columns BDSS and

Table 3 Estimation metrics (errors, biases, CI coverage and width) for the incubation fraction f_E for 2,000–5,000-tip transmission trees generated under different models (rows) and different estimators (columns).

	BDEI		BDEISS		BDEI-CT		BDEISS-CT	
				Error (bias)				
BD	**0**	**(+0)**	**0**	**(+0)**	**0**	**(+0)**	**1**	**(+1)**
BDSS	14	(+14)	**0**	**(+0)**	24	(+24)	**0**	**(+0)**
BD-CT	0	(+0)	1	(+1)	**0**	**(+0)**	**1**	**(+1)**
BDSS-CT	4	(+4)	0	(+0)	24	(+24)	**0**	**(+0)**
BDEI	**2**	**(−1)**	**2**	**(−1)**	**2**	**(−1)**	**3**	**(−1)**
BDEISS	21	(+20)	**3**	**(−0)**	26	(+26)	**3**	**(−0)**
BDEI-CT	18	(−18)	24	(−24)	**2**	**(−1)**	**3**	**(−2)**
BDEISS-CT	16	(+11)	15	(−15)	25	(+25)	**4**	**(+0)**
				CI coverage (width)				
BD	**99**	**(3)**	**99**	**(4)**	**99**	**(4)**	**99**	**(5)**
BDSS	45	(9)	**100**	**(2)**	14	(11)	**99**	**(3)**
BD-CT	93	(2)	48	(2)	**99**	**(3)**	**98**	**(5)**
BDSS-CT	72	(6)	95	(2)	5	(9)	**100**	**(5)**
BDEI	**96**	**(10)**	**96**	**(12)**	**97**	**(11)**	**96**	**(13)**
BDEISS	14	(9)	**96**	**(14)**	10	(11)	**97**	**(15)**
BDEI-CT	21	(10)	22	(12)	**97**	**(13)**	**96**	**(16)**
BDEISS-CT	23	(9)	34	(13)	8	(8)	**96**	**(19)**

The top half of the table reports mean absolute errors multiplied by 100 – $100|f_{E\,estimated} - f_{E\,true}|$ – and in parenthesis the corresponding biases – $100(f_{E\,estimated} - f_{E\,true})$ – estimated on 1,000 trees generated under each dataset for each estimator.
The bottom half of the table reports CI coverage – percentage of trees for which the real parameter value was within the estimated CI – and in parenthesis the corresponding mean CI width – $100(f_{E\,97.5\%} - f_{E\,2.5\%})$ – estimated on 1,000 trees generated under each dataset for each estimator.
The errors and biases of the estimators corresponding to or generalizing the model that generated the data are shown in bold.
The first and third groups of rows contain data-generating models with $f_E = 0$ (BD, BDSS, BD-CT and BDSS-CT), the second and fourth groups of rows contain those with $f_E \geq 0$.

Table 4 Estimation metrics (errors, biases, CI coverage and width) for the super-spreader fraction f_S for 2,000–5,000-tip transmission trees generated under different models (rows) and different estimators (columns).

	BDSS		BDEISS		BDSS-CT		BDEISS-CT	
				Error (bias)				
BD	**0**	**(+0)**	**0**	**(+0)**	**0**	**(+0)**	**0**	**(+0)**
BDEI	6	(+6)	**0**	**(+0)**	6	(+6)	**0**	**(+0)**
BD-CT	1	(+1)	7	(+7)	**0**	**(+0)**	**0**	**(+0)**
BDEI-CT	11	(+11)	21	(+21)	10	(+10)	**1**	**(+1)**
BDSS	**2**	**(−0)**	**2**	**(−0)**	**2**	**(−0)**	**2**	**(+0)**
BDEISS	18	(−18)	**2**	**(−0)**	18	(−18)	**3**	**(−0)**
BDSS-CT	4	(−3)	4	(−2)	**2**	**(−0)**	**2**	**(−0)**
BDEISS-CT	17	(−17)	7	(−6)	17	(−17)	**3**	**(−1)**
				CI coverage (width)				
BD	**99**	**(14)**	**99**	**(18)**	**99**	**(17)**	**99**	**(18)**
BDEI	46	(13)	**100**	**(17)**	37	(14)	**99**	**(18)**
BD-CT	92	(21)	52	(29)	**99**	**(17)**	**99**	**(22)**
BDEI-CT	35	(13)	38	(24)	33	(12)	**98**	**(22)**

(*continued*)

Table 4 Continued

	BDSS		BDEISS		BDSS-CT		BDEISS-CT	
BDSS	**97**	**(10)**	**96**	**(11)**	**97**	**(11)**	**97**	**(12)**
BDEISS	14	(7)	**97**	**(12)**	14	(6)	**98**	**(14)**
BDSS-CT	74	(9)	78	(11)	**97**	**(10)**	**98**	**(12)**
BDEISS-CT	12	(7)	56	(10)	16	(7)	**97**	**(17)**

The top half of the table reports mean absolute errors multiplied by 100 – $100|f_{S_{estimated}} - f_{S_{true}}|$ – and in parenthesis the corresponding biases – $100(f_{S_{estimated}} - f_{S_{true}})$ – estimated on 1,000 trees generated under each dataset for each estimator.
The bottom half of the table reports CI coverage – percentage of trees for which the real parameter value was within the estimated CI – and in parenthesis the corresponding mean CI width – $100(f_{S_{97.5\%}} - f_{S_{2.5\%}})$ – estimated on 1,000 trees generated under each dataset for each estimator.
The errors and biases of the estimators corresponding to or generalizing the model that generated the data are shown in bold.
The first and third groups of rows contain data-generating models with $f_S = 0$ (BD, BDEI, BD-CT and BDEI-CT), the second and fourth groups of rows contain those with $f_S \geq 0$.

Table 5 Estimation metrics (errors, biases, CI coverage and width) for the contact-tracing probability υ for 2,000–5,000-tip transmission trees generated under different models (rows) and different estimators (columns).

	BD-CT		BDEI-CT		BDSS-CT		BDEISS-CT	
				Error (bias)				
BD	**0**	**(+0)**	**0**	**(+0)**	**0**	**(+0)**	**0**	**(+0)**
BDEI	0	(+0)	**0**	**(+0)**	0	(+0)	**0**	**(+0)**
BDSS	5	(+5)	46	(+46)	**0**	**(+0)**	**0**	**(+0)**
BDEISS	2	(+2)	47	(+47)	0	(+0)	**1**	**(+1)**
BD-CT	**3**	**(−0)**	**3**	**(+0)**	**3**	**(−0)**	**3**	**(+0)**
BDEI-CT	22	(−21)	**3**	**(−0)**	24	(−23)	**4**	**(−1)**
BDSS-CT	12	(+10)	28	(+28)	**3**	**(−0)**	**4**	**(−0)**
BDEISS-CT	21	(−14)	28	(+28)	23	(−23)	**6**	**(+0)**
				CI coverage (width)				
BD	**99**	**(12)**	**100**	**(14)**	**100**	**(10)**	**100**	**(16)**
BDEI	98	(4)	**99**	**(15)**	97	(4)	**99**	**(17)**
BDSS	76	(30)	9	(48)	**99**	**(12)**	**99**	**(17)**
BDEISS	86	(16)	4	(48)	99	(8)	**99**	**(25)**
BD-CT	**96**	**(15)**	**96**	**(16)**	**96**	**(16)**	**98**	**(18)**
BDEI-CT	27	(12)	**96**	**(19)**	27	(14)	**97**	**(23)**
BDSS-CT	46	(22)	4	(17)	**96**	**(19)**	**96**	**(21)**
BDEISS-CT	33	(18)	8	(19)	36	(20)	**95**	**(32)**

The top half of the table reports mean absolute errors multiplied by 100 – $100|\upsilon_{estimated} - \upsilon_{true}|$ – and in parenthesis the corresponding biases – $100(\upsilon_{estimated} - \upsilon_{true})$ – estimated on 1,000 trees generated under each dataset for each estimator.
The bottom half of the table reports CI coverage – percentage of trees for which the real parameter value was within the estimated CI – and in parenthesis the corresponding mean CI width – $100(\upsilon_{97.5\%} - \upsilon_{2.5\%})$ – estimated on 1,000 trees generated under each dataset for each estimator.
The errors and biases of the estimators corresponding to or generalizing the model that generated the data are shown in bold.
The first and third groups of rows contain data-generating models with $\upsilon = 0$ (BD, BDEI, BDSS and BDEISS), the second and fourth groups of rows contain those with $\upsilon \geq 0$.

BDSS-CT; Table 3, rows BDSS and BDSS-CT, columns BDEI and BDEI-CT). *Super-spreading and contact-tracing* might be confused one for another as well (Table 5, rows BDSS and BDEISS, columns BD-CT and BDEI-CT; Table 4, rows BD-CT and BDEI-CT, columns BDSS and BDEISS). However, *incubation does not get confused with CT, and vice versa* (Table 5, row BDEI, columns BD-CT and BDSS-CT; Table 3, row BD-CT, columns BDEI and BDEISS).

The above confusion trends can also be explained using patterns in the transmission trees. Incubation and super-spreading leave similar internal branch patterns. Incubation leads to length differences between sister branches, as right after the transmission (internal node) the transmitter (one of the child branches) is ready to transmit again or be removed, while the recipient (the other child branch) needs to undergo incubation first. Under super-spreading, when the transmitter and the recipient states are different, the same pattern arises for internal sister branches: the one corresponding to a regular spreader is on average longer. This is due to regular spreaders transmitting less over the same infectious time, hence taking longer between transmissions. (Note, that internal branches represent times between transmissions.) CT, on the contrary, does not change the times between transmissions, hence not leading to such a pattern.



Figure 4 95%-CI coverage (left half, percentage of trees for which the true parameter value is within the estimated interval) and relative width (right half, $(p_{97.5\%} - p_{2.5\%})/p_{true}$ for $p \in \{R_e, d\}$) for effective reproduction number R_e (first and third columns) and average infection time d (second and fourth columns). The CIs were estimated on datasets of 1,000 transmission trees with 2,000-5,000-tips generated under different flavors of the BD(EI)(SS)(-CT) model (corresponding to rows, labels shown in the middle). The bars correspond to different estimators (see the legend at the bottom). (ML) and dotted pattern indicate a maximum-likelihood estimator, while (DL) indicates deep-learning estimators. The DL estimators were trained on a mix of trees generated under the corresponding model and its nested models (e.g. BDEI and BD trees for the BDEI estimator). Their CIs were estimated with pinball loss (Steinwart and Christmann 2011). For the BD (ML) estimator the CIs are approximate: they were obtained from the CIs of the rate parameters (λ and ψ) as $R_{e,97.5\%} = \lambda_{97.5\%}/\psi_{2.5\%}$; $R_{e,2.5\%} = \lambda_{2.5\%}/\psi_{97.5\%}$; $d_{97.5\%} = 1/\psi_{2.5\%}$; $d_{2.5\%} = 1/\psi_{97.5\%}$. The rate parameter CIs were estimated using Wilks' method (Wilks 1938). Estimators accounting for incubation have forward-slash pattern, those accounting for super-spreading have backward-slash one, and those accounting for contact-tracing have horizontal dashes. Bar heights correspond to the CI coverage/mean relative width on 1,000 trees (also shown below each bar), while the whiskers for the width correspond to 95% confidence intervals. The estimators that do not correspond or generalize the data model (e.g. BDEI estimator for BDSS data) are shown with semi-transparent colors.

Super-spreading and CT produce similar patterns in tip time differences. Under CT, when an infected individual (the notifier) is removed, their notified contacts are removed soon afterwards. This speeds up the contacts' removal and therefore shortens the gap between the notifier's removal time and theirs – a gap that is smaller, on average, than the one between a transmitter and a recipient when no notification takes place. Super-spreading also shortens the removal-time gap between a superspreader transmitter and their first recipients, compared to regular transmitters. But here the mechanism is different: because superspreaders and regular transmitters have the same infectious period, the shorter gap must reflect the timing of transmission itself. Under the BDSS model, superspreaders transmit sooner after infection, which reduces the time between their removal and their recipients'. Incubation, by contrast, has the opposite effect: it lengthens the tip time difference by adding the incubation period (as explained above).

Having larger trees reduces uncertainty and increases accuracy with correct models. Indeed, CIs are narrower and errors are lower on larger datasets, when the estimator model matches (or generalizes) the model that generated the data (Tables S3–S16). For instance, the error for d with the BDEI(SS)(-CT) estimators on BDEI trees dropped from 9–14% to 4–5%, and the CI width from 49–68% to 19–26%, as the tree sizes increased from 200–500 tips to 2,000–5,000 tips (Tables S4 and S11, row BDEI, columns BDEI, BDEISS, BDEI-CT and BDEISS-CT). However, for

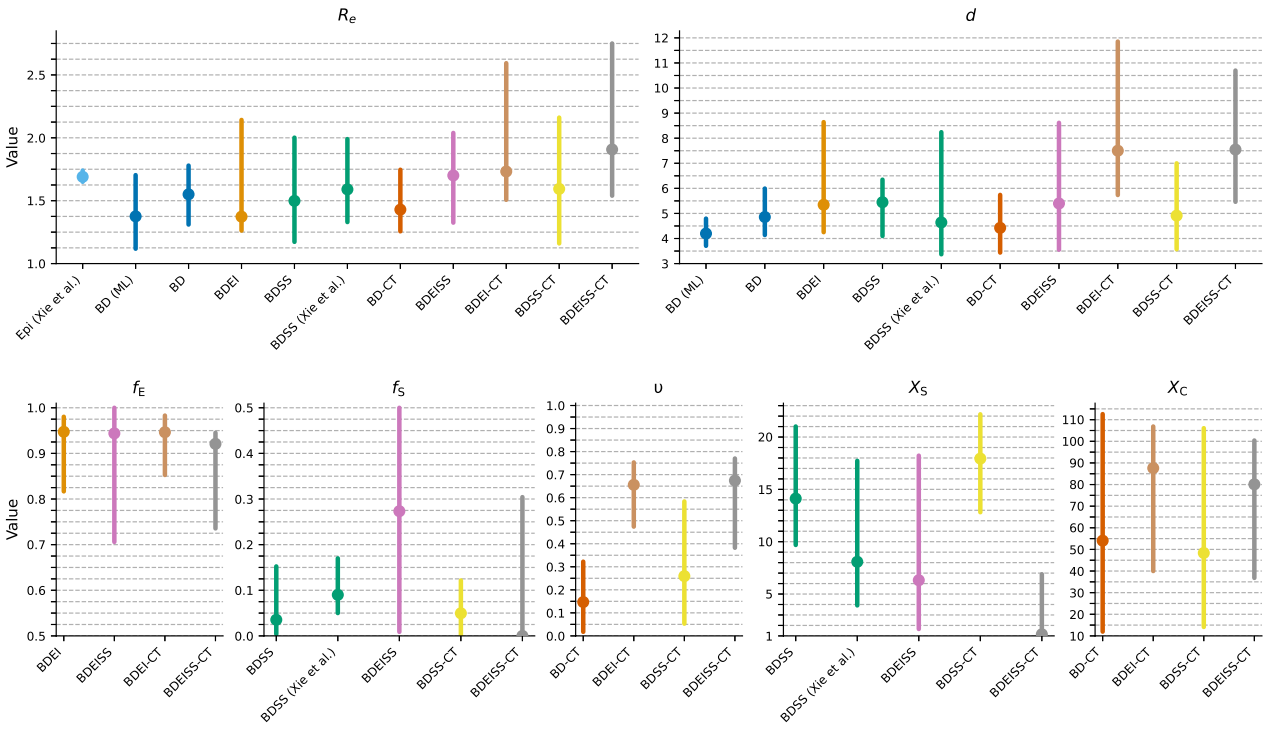


Figure 5 Hong Kong SARS-CoV-2 wave 3 epidemiological parameters and their CIs estimated with different models. Each parameter is represented by a separate plot (parameter name shown in title). The dots correspond to the median estimates, while the whiskers cover the 95% CIs inferred with different estimators. The estimator names are shown on the *x*-axis. "BD (ML)" is a maximum-likelihood estimator for the BD model (Zhukova and Gascuel 2025). Estimators labeled only with the model name correspond to the DL estimators described in this study. They were trained on nested models for 200-500-tip trees. "BDSS (Xie *et al.*)" is a DL-based inference with the BDSS model, while "Epi (Xie *et al.*)" is an epidemiological inference from incidence data, both described in Xie et al. (2024).

mismatched estimators, the errors overall did not decrease, while the CI widths did, e.g. the errors for d with the BD-CT estimator on BDEI data stayed around 36% for all tree sizes, though the CI width decreased from 52–55% to 20–25% (Tables S4 and S11, row BDEI, column BD-CT). This is dangerous, as mismatched estimators become more and more certain in their biased predictions.

***X_C* and *X_S* are the most difficult parameters to estimate.** The average errors were more than 10% for all the estimators (Tables S8–S9), and many CIs close to or wider than 100% (Tables S15–S16). This is not surprising as there are few traces of these parameters in the tree (i.e. few notified external branches for X_C and few superspreader internal branches for X_S). This behavior was also observed in previous studies (Voznica et al. 2022; Xie et al. 2024; Perez and Gascuel 2026).

DL predictors trained on a mix of nested models generalize better. As the BDEISS-CT training data contained <1% of BD-like trees (with f_E, f_S and υ close to zero), it is not surprising that it performed worse on the BD dataset than the BDEISS-CT estimator trained on nested models, with $\geq 12.5\%$ of them containing BD-like data (Tables S3–S16). Strong performance of nested-model-trained estimators suggests that an explicit model selection step is not required. Rather than adopting a two-step approach – in which a model finder first selects the most appropriate model for the input tree, followed by a model-specific estimator to infer the parameters (e.g. as in Voznica et al. 2022) – one can instead train a single estimator on data generated from all models simultaneously (provided each model is represented by enough examples). This estimator will then implicitly perform model selection and infer the appropriate parameters directly from the input data (as the mixed BDEISS-CT estimator here).

Application to SARS-CoV-2 wave 3 epidemic in Hong Kong

For the SARS-CoV-2 tree, BD and BD-CT models were incompatible according to the summary statistics check (see "Real pathogen data" for details): median length of internal branches, variance of tip branch lengths and the 25^{th} percentile of the sum of branch lengths of four chains were more than five standard deviations away from the corresponding mean values of the training data. We report the predictions of these estimators for completeness, but the check suggests they should not be trusted. Moreover, as the other estimators pass the check, the mismatch between the BD(-CT) training data and the SARS-CoV-2 tree is most likely due to the presence of complex epidemiological aspects in the latter.

Using the BDEISS-CT estimator for the SARS-CoV-2 wave 3 epidemic in Hong Kong (Fig. 5, Table S16), we obtained the reproduction number $R_e = 1.91$ $(1.5 - 2.8)$. This point estimate is higher than the ones estimated with all the other models, as well as from a previous study on this data, using the BDSS model and an epidemiological approach from incidence data (Xie et al.

2024). This is compatible with the trends obtained on simulated data: our simulations showed that simpler models tend to underestimate R_e. We estimated the infection duration $d = 7$ $(5 - 11)$ days, most of which (92 (74 − 95) %) corresponds to incubation. The latter part could be explained by the high CT probability 0.7 (0.4 − 0.8) with a very large (80 (37 − 100) times) increase in removal rate. This means that once notified, the infected person immediately takes care to prevent further transmission (and if notified during the incubation period does not transmit at all). We estimated that there is potentially no super-spreading ($f_S = 0$ $(0 - 0.3)$), and even if there is, the difference in transmission rate is not large ($X_S = 1$ $(1 - 7)$ times). Assuming no super-spreading, the BDEI-CT estimator gives very similar values for all the parameters to BDEISS-CT: $R_e = 1.73$ $(1.5 - 2.6)$; $d = 7$ $(5 - 12)$ [days] out of which 95 (85 − 98)% is incubation; $\upsilon = 0.7$ $(0.5 - 0.8)$ with 88 (40 − 107) times faster removal after notification.

As we have seen on simulated data, unaccounted contact-tracing might be confused with super-spreading, which explains why our BDEISS estimator and a previous study (Xie et al. 2024) using a BDSS estimator detected super-spreading in these data. While the R_e estimate by the BDEISS and the BDEI-CT estimators are close, the underlying mechanisms are very different. In the case of super-spreading (the BDEISS model), the estimated R_e and d parameters translate to the infectious period $d_{inf} = 0.3$ [days], with an average transmission rate of 5.6 [times per day]. This transmission rate corresponds to the basic transmission rate $\lambda = 2.3$ [times per day] for regular spreaders I, and an increased rate of 14.5 [times per day] for superspreaders (accounting for 27% of the infected population). The infectious time seems to be short, though might be due to the person's decision to self-isolate or wear a mask soon after the onset of symptoms. In the case of contact tracing (BDEI-CT), the estimated parameters translate to the average infectious period $d_{inf} = 0.4$ [days] and a transmission rate $\lambda = 4.3$ [times per day]. The average infectious period of 0.4 days here means that untraced spreaders I typically stay infectious longer than that, while contact-traced ones I_C – shorter (potentially taking care to wear a mask even before the end of incubation). Therefore, with comparable average infectious times and transmission rates, the BDEISS model suggests lower base transmission rate (for I) and its increase for superspreaders over homogeneous infectious time, while the BDEI-CT model suggests homogeneous transmission rate over heterogeneous infectious time, with shorter time for contact-traced individuals.

Discussion

We investigated, using simulated data, the performance of epidemiological parameter estimators of varying complexity in the setting of different epidemiological model misspecifications. On one hand, our results suggest that using simpler estimators (such as BD) on data that features complex epidemic features (e.g. super-spreading) leads to bias in parameter estimation. This bias is dangerous from the public health perspective: For instance using the BD estimator on the data featuring super-spreading, contact tracing or incubation, leads to underestimation of the reproduction number R_e by 10–50%, and hence potential conclusion that the epidemic is better contained that it really is. This results agrees with a previous study of effects of non-accounting for incubation by Park and Koelle (2026). On the other hand, our results suggest that using a more complex estimator on simpler data (e.g. accounting for contact-tracing when it is not present) leads to accurate estimates (e.g. contact tracing probability of zero), however with wider confidence intervals. A similar conclusion was recently achieved for several models in macroevolution domain (Barido-Sottani and Morlon 2026), as well as for nucleotide substitution models for phylogenetic inference (Guimarães Fabreti and Höhna 2023) in the Bayesian context, provided well-behaved priors are used. However, as our simulations show, to achieve this behavior with complex models in the DL-based context, it is important to have boundary-case data (i.e. corresponding to the nested models) well represented in the training dataset. Moreover, our results show that there is no need for explicit model selection: an estimator trained on the data generated from multiple models performs model selection implicitly and infers the parameters appropriate to the input data.

Our study hence raises awareness of the importance of accounting for as many aspects of the epidemic as possible. Until recently, this strategy was hindered by the mathematical and computational complexity of the models in question. This led to excessive use of the simple BD(SKY) models, and potential underestimates of the key epidemiological parameter values. However, the recent arrival of deep learning to the phylodynamic domain (Voznica et al. 2022; Thompson et al. 2024; Landis and Thompson 2025; Perez and Gascuel 2026) enabled replacing complex likelihood computation with much simpler transmission tree simulation (for training of a deep learning predictor). Deep learning thus opened the way for the design of, and parameter estimation under, more complex models. In particular, the CT-aware models in our study did not have an explicit formula for R_e and d, however, these parameter values could be extracted from simulations, and the DL-based predictors managed to accurately estimate them. Estimation of these parameters for CT-aware models would not be possible in the likelihood-based setting.

For our study we trained proof-of-concept deep learning estimators for complex models that did not have a likelihood-based alternative. We used a simple neural network architecture and did not try to achieve the best possible performance for these estimators. Nevertheless even these versions of complex-model estimators outperformed the estimators for simpler models, and especially the maximum likelihood estimator for the BD model. The performance gap might further increase with a better deep learning architecture (Perez and Gascuel 2026) or larger training dataset size.

We applied our estimators to an empirical dataset (SARS-CoV-2 wave 3 epidemic in Hong-Kong) and obtained biologically reasonable estimates, which were also coherent with the trends observed on simulated data (e.g. lower R_e for simpler models, contact-tracing potentially mistaken for super-spreading).

Our investigation also raises new questions for future studies: Where should one stop when accounting for complex epidemic aspects? And how to ensure all the needed aspects are accounted for? In particular, in the models we considered all the parameters were constant. This is rarely the case in real epidemics, and can become an additional source of bias – or even non-identifiability (Louca et al. 2021) – in the estimations. Assessing the errors in

estimations with BD(EI)(SS)(-CT) models on BD(EI)(SS)(-CT)-SKY data – featuring piecewise-constant changes in rates over time – would be the natural next step. We could detect the mismatch between the training data for the BD and BD-CT estimators and the real SARS-CoV-2 tree by looking at summary statistics. This check is definitely useful, but its performance depends on the choice of summary statistics. For instance, we could not detect a mismatch for the BDSS model, while the presence of incubation is well-established for SARS-CoV-2. Therefore, additional tests will need to be designed in the future.

Conclusion

In conclusion, our results suggest that accounting for as many potential aspects (super-spreading, incubation, contact tracing) of an epidemic as possible is key even for estimation of epidemiological parameters that do not directly depend on these aspects (e.g. the reproduction number R_e). Deep learning helps remove the mathematical difficulty in designing such estimators. If, however, a simpler estimator is used on a potentially more complex epidemic, one needs to keep in mind potential biases in the results (e.g. underestimation of R_e). These biases might be present in previous phylodynamic studies that had to use the simpler BD(SKY) model estimator due to its closed-form ODE solution and scalability to large sequence datasets. Simulated data allows to assess potential biases in estimation with such simpler models.

Acknowledgments

We would like to thank the High-Performance Computational Cluster team of Institut Pasteur.

Author contributions

A.Z. and R.X. conceived the experiments, A.Z. conducted the experiments, analysed the results and wrote the manuscript. R.X. and O.G. critically reviewed the manuscript, and discussed the project progress.

Supplementary material

Supplementary material is available at *Molecular Biology and Evolution* online.

Funding

This work is supported in part by funds from the Marie Skłodowska-Curie Actions (Project No. 101203810, R.X.) and from the French National Research Agency (Project No. ANR-25-CE45-1990 EmPaTHIc, A.Z.).

Conflicts of interest

No competing interest is declared.

Data availability

The Snakemake (Köster and Rahmann 2012) pipelines and source code for training and test dataset generation, tree encoding into summary statistics, DL-based estimator training, and epidemiological parameter inference with DL- and maximum-likelihood-based estimators are available on GitHub at github.com/modpath/bdeissct. The repository also contains the newick files with transmission trees used for tests, Hong Kong SARS-CoV-2 wave 3 tree, as well as the pretrained DL models. The DL-based estimators are also available as a command-line program and a Python 3 package via PyPi (bdext), and via Docker/Singularity (evolbioinfo/bdext).

References

Barido-Sottani J, Morlon H. Integrating fossils samples with heterogeneous diversification rates: a combined multi-type fossilized birth-death model. *Syst Biol*. 2026:75:1175–1195. https://doi.org/10.1093/sysbio/syag011.

Biggerstaff M, Cauchemez S, Reed C, Gambhir M, Finelli L. Estimates of the reproduction number for seasonal, pandemic, and zoonotic influenza: a systematic review of the literature. *BMC Infect Dis*. 2014:14:480. https://doi.org/10.1186/1471-2334-14-480.

Dieselhorst T, Stadler T. Information on hidden birth events restores identifiability in phylodynamic inference [preprint]. arXiv:2604.17926. 2026. https://doi.org/10.48550/arXiv.2604.17926.

Geoghegan JL *et al.* Genomic epidemiology reveals transmission patterns and dynamics of SARS-CoV-2 in Aotearoa New Zealand. *Nat Commun*. 2020:11:6351. https://doi.org/10.1038/s41467-020-20235-8.

Grenfell BT *et al.* Unifying the epidemiological and evolutionary dynamics of pathogens. *Science*. 2004:303:327–332. https://doi.org/10.1126/science.1090727.

Gu H *et al.* Genomic epidemiology of SARS-CoV-2 under an elimination strategy in Hong Kong. *Nat Commun*. 2022:13:736. https://doi.org/10.1038/s41467-022-28420-7.

Guimarães Fabreti L, Höhna S. Nucleotide substitution model selection is not necessary for Bayesian inference of phylogeny with well-behaved priors. *Syst Biol*. 2023:72:1418–1432. https://doi.org/10.1093/sysbio/syad041.

Hethcote HW. The mathematics of infectious diseases. *SIAM Rev*. 2000:42:599–653. https://doi.org/10.1137/S0036144500371907.

Hodcroft EB *et al.* Spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *Nature*. 2021:595:707–712. https://doi.org/10.1038/s41586-021-03677-y.

Ishikawa SA, Zhukova A, Iwasaki W, Gascuel O. A fast likelihood method to reconstruct and visualize ancestral scenarios. *Mol Biol Evol*. 2019:36:2069–2085. https://doi.org/10.1093/molbev/msz131.

Köster J, Rahmann S. Snakemake-a scalable bioinformatics workflow engine. *Bioinformatics*. 2012:28:2520–2522. https://doi.org/10.1093/bioinformatics/bts480.

Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood

phylogenetic inference. *Bioinformatics*. 2019:35:4453–4455. https://doi.org/10.1093/bioinformatics/btz305.

Kühnert D, Stadler T, Vaughan TG, Drummond AJ. Phylodynamics with migration: a computational framework to quantify population structure from genomic data. *Mol Biol Evol*. 2016:33:2102–2116. https://doi.org/10.1093/molbev/msw064.

Landis MJ, Thompson A. phyddle: software for exploring phylogenetic models with deep learning. *Syst Biol*. 2025:74:1007–1019. https://doi.org/10.1093/sysbio/syaf036.

Lemey P, Rambaut A, Drummond AJ, Suchard MA. Bayesian phylogeography finds its roots. *PLoS Comput Biol*. 2009:5:e1000520. https://doi.org/10.1371/journal.pcbi.1000520.

Louca S, McLaughlin A, MacPherson A, Joy JB, Pennell MW. Fundamental identifiability limits in molecular epidemiology. *Mol Biol Evol*. 2021:38:4010–4024. https://doi.org/10.1093/molbev/msab149.

Minh BQ *et al.* IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 2020:37:1530–1534. https://doi.org/10.1093/molbev/msaa015.

Nash RK *et al.* Ebola virus disease mathematical models and epidemiological parameters: a systematic review. *Lancet Infect Dis*. 2024:24:e762–e773. https://doi.org/10.1016/S1473-3099(24)00374-8.

Park Y, Koelle K. Common misspecification of the generation interval leads to reproduction number underestimation in phylodynamic inference. *Epidemics*. 2026:55:100910. https://doi.org/10.1016/j.epidem.2026.100910.

Perez MF, Gascuel O. PhyloCNN: improving tree representation and neural network architecture for deep learning from trees in phylodynamics and diversification studies. *Syst Biol*. 2026:75:776–795. https://doi.org/10.1093/sysbio/syaf082.

Rasmussen DA, Kouyos R, Günthard HF, Stadler T. Phylodynamics on local sexual contact networks. *PLoS Comput Biol*. 2017:13:1–23. https://doi.org/10.1371/journal.pcbi.1005448.

Resende PC *et al.* Evolutionary dynamics and dissemination pattern of the SARS-CoV-2 lineage B.1.1.33 during the early pandemic phase in Brazil. *Front Microbiol*. 2021:11:615280. https://doi.org/10.3389/fmicb.2020.615280.

Scire J, Barido-Sottani J, Kühnert D, Vaughan TG, Stadler T. Robust phylodynamic analysis of genetic sequencing data from structured populations. *Viruses*. 2022:14:1648. https://doi.org/10.3390/v14081648.

Stadler T. On incomplete sampling under birth-death models and connections to the sampling-based coalescent. *J Theor Biol*. 2009:261:58–66. https://doi.org/10.1016/j.jtbi.2009.07.018.

Stadler T. Sampling-through-time in birth-death trees. *J Theor Biol*. 2010:267:396–404. https://doi.org/10.1016/j.jtbi.2010.09.010.

Stadler T, Bonhoeffer S. Uncovering epidemiological dynamics in heterogeneous host populations using phylogenetic methods. *Philos Trans R Soc B Biol Sci*. 2013:368:20120198–-20120198. https://doi.org/10.1098/rstb.2012.0198.

Stadler T, Kühnert D, Bonhoeffer S, Drummond AJ. Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proc Natl Acad Sci U S A*. 2013:110:228–233. https://doi.org/10.1073/pnas.1207965110.

Stadler T, Kühnert D, Rasmussen DA, du Plessis L. Insights into the early epidemic spread of ebola in sierra leone provided by viral sequence data. *PLoS Curr*. 2014:6:ecurrents.outbreaks.02bc6d927ecee7bbd33532ec8ba6a25f. https://doi.org/10.1371/currents.outbreaks.02bc6d927ecee7bbd33532ec8ba6a25f.

Steinwart I, Christmann A. Estimating conditional quantiles with the help of the pinball loss. *Bernoulli*. 2011:17:211–225. https://doi.org/10.3150/10-BEJ267.

Thompson A, Liebeskind BJ, Scully EJ, Landis MJ. Deep learning and likelihood approaches for viral phylogeography converge on the same answers whether the inference model is right or wrong. *Syst Biol*. 2024:73:183–206. https://doi.org/10.1093/sysbio/syad074.

To T-H, Jung M, Lycett S, Gascuel O. Fast dating using least-squares criteria and algorithms. *Syst Biol*. 2016:65:82–97. https://doi.org/10.1093/sysbio/syv068.

Vaughan TG, Scire J, Nadeau SA, Stadler T. Estimates of early outbreak-specific SARS-CoV-2 epidemiological parameters from genomic data. *Proc Natl Acad Sci U S A*. 2024:121:e2308125121. https://doi.org/10.1073/pnas.2308125121.

Velásquez GE *et al.* Time from infection to disease and infectiousness for ebola virus disease, a systematic review. *Clin Infect Dis*. 2015:61:1135–1140. https://doi.org/10.1093/cid/civ531.

Viana R *et al.* Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature*. 2022:603:679–686. https://doi.org/10.1038/s41586-022-04411-y.

Volz EM, Koelle K, Bedford T. Viral phylodynamics. *PLoS Comput Biol*. 2013:9:e1002947. https://doi.org/10.1371/journal.pcbi.1002947.

Voznica J *et al.* Deep learning from phylogenies to uncover the transmission dynamics of epidemics. *Nat Commun*. 2022:13:3896. https://doi.org/10.1038/s41467-022-31511-0.

Walsh KA *et al.* The duration of infectiousness of individuals infected with SARS-CoV-2. *J Infect*. 2020:81:847–856. https://doi.org/10.1016/j.jinf.2020.10.009.

Wilks SS. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann Math Stat*. 1938:9:60–62. https://doi.org/10.1214/aoms/1177732360.

Xie R *et al.* Integrating contact tracing data to enhance outbreak phylodynamic inference: a deep learning approach. *Mol Biol Evol*. 2024:41:msae232. https://doi.org/10.1093/molbev/msae232.

Zhukova A, Gascuel O. Accounting for contact tracing in epidemiological birth-death models. *PLoS Comput Biol*. 2025:21:e1012461. https://doi.org/10.1371/journal.pcbi.1012461.