A compartmental model for epidemic parameter estimation and forecasting, with applications to SARS-CoV-2

Michael Li, Jonathan Dushoff, David J. D. Earn,
Irena Papst, Benjamin M. Bolker
McMaster University

June 12, 2023 @ 13:54

Abstract

Compartmental epidemiological models are widely used to understand, manage, and forecast the SARS-CoV-2 (COVID-19) pandemic. We introduce a new compartmental modeling framework that shares many characteristics with existing models, but includes a number of new and noteworthy features. In particular, it includes a flexible structure based on the *flow matrix* (the *per capita* rates of transitions between compartments) that allows it to be used interchangeably for discrete or continuous time and for deterministic or discrete-state stochastic models; the capacity to set starting conditions based on the expected distribution of states during an exponential phase of the epidemic; automatic computation of $\mathcal{R}_0$ and the mean and dispersion of the generation interval for specified parameters; explicit structures incorporating the intensity of testing and delays between test administration and reporting; time-varying parameters based on breakpoints, spline bases, or external covariates such as cellphone-based mobility indices; and the ability to calibrate model parameters to multiple data streams such as case reports, hospitalization and ICU admission rates. We demonstrate the model by calibrating it to multiple COVID-19 time series (positive tests, negative tests, hospitalizations, and deaths) for each of the Canadian provinces from 2020-02-27 to 2020-08-30. We estimate epidemiological parameters, including the effective reproduction number $\mathcal{R}_t$ over the course of these early six months of the pandemic.

[**DE:** *Target journal? Ideas: PLoS Comp Biol*] [**BB:** *PLoS Comp Biol seems to be shooting very high for this. I'd be fine with PeerJ/PLoS ONE, but would be willing to try harder, e.g. some low-tier-but-not-terrible epi or modeling venue*]

Contents

Overview

[Writing notes here for now so that I can use `fref`. This is not part of the draft.]

May notes

Next step is to decide what results to use and how to frame them. I feel that Figure 4 and Figure 6 are the beginning of a framework that could drive things forward: we tried this and got that. Here are the limitations or concerns – repeat. The big question in my mind is what's a good number/choice of things to show. How far forward in time or complexity are we willing to go? [**BB:** *For my part, not much! I really want to get this out as the **minimal adequate product.***][**DE:** *ditto*]

Reading now, I feel like the Results could be edited to make the supporting cast (most of the other figures) feel more like supporting cast.

If we edit the Results the way I want, I think they will feel thin, so that brings me back to the question: what can we add? [**BB:** *My personal feeling is that we should let them be thin. The forecasting is no longer interesting (I think); what may be worth highlighting are the novel modeling aspects (there are a few).*][**DE:** *agree*]

For even broader framing: here's the stuff we did during the pandemic, some of the problems we encountered, and some of the things we learned. We are describing these lessons because they may be helpful to people in future new outbreaks. We should also find a way to talk about how the development of macpan laid the groundwork for the development of macpan2, and to mention that it is here and we will be writing papers about it. [**BB:** *OK, but not more than a few sentences.*][**DE:** *agree*]

1 Introduction

SARS-CoV-2, the etiological agent of coronavirus disease 2019 (COVID-19), has been circulating in Canada since at least January 2020 (1). Response to the worldwide pandemic (2; 3) has been guided to a substantial extent by mathematical modelling (4).

Here, we present a compartmental framework that was developed over the course of the pandemic. It incorporates the standard epidemiological compartments required to model COVID-19 as well as compartments that track health-care utilization and COVID-induced mortality. Case reporting can be modeled either as a time-delayed convolution of incidence or by enabling a factorial expansion of the model that accounts for the testing status of individuals (5). The model also allows for time-dependent variation in any rate parameter — particularly the transmission rate — and allows this variation to be indexed by external covariates such as cellphone-based mobility metrics, or to follow smooth (spline) curves over time. The same model structure can be run as an ordinary differential equation, a discrete-time deterministic model, or a discrete-time stochastic model. Finally, the model can be used to (1) simulate specific scenarios for planning purposes; (2) calibrate parameters to match multiple input time series such as hospital admissions or occupancy, cases, or deaths; or (3) forecast future epidemic dynamics on the basis of past calibration.

2   Methods

Compartmental structure

The epidemiological structure of the model is based on a susceptible-exposed-infectious-removed (SEIR) model with additional compartments reflecting the biology of COVID-19 and the structure of the health-care system. The COVID-19-specific compartmental structure of the model resembles many other COVID-19 models (6–8) in separating infectious individuals into sub-compartments reflective of the epidemiology of COVID-19; it additionally adds compartments for hospitalized individuals in acute care or intensive care. All symptomatic individuals are presumed to have undergone a period of pre-symptomatic infectiousness (`p`). Infections can be asymptomatic (`a`), mildly or moderately symptomatic (`m`) or severely symptomatic (`s`); all individuals with severe symptoms go to the hospital (acute care or ICU), or die before reaching hospital (e.g., in long-term care facilities). Some fraction [**DE:** *add parameter symbol in brackets*] of individuals who go to the ICU die. Recovered individuals (`R`) are assumed to be immune. The model includes additional compartments that facilitate book-keeping: cumulative hospital admissions (`X`), individuals in acute care after discharge from ICU (`H2`), and cumulative deaths (`D`) (Figure 1). (The version of the model discussed here was developed and used before vaccines were available and introduction of variants of concern and reinfections, we have since expanded the model to include other relevant compartments.)

[**DE:** *It is weird that we do not have a table listing all the parameters of the model. I guess that's why the symbol for the proportion in ICU who die was not given above.*]

This version of the model assumes homogeneous mixing — all classes of infectives contribute additively to the force of infection (the *per capita* infection rate of susceptibles). We assumed hospitalized individuals do not contribute to the force of infection. [**ML:** *re-write this point nicely? it is a strong assumption.*] We did add one feature to the model to account for heterogeneity in susceptibility in the population, which we typically imagine is driven by heterogeneity in exposure (e.g., front-line and essential workers will be infected earlier), but could also be influenced by genetic, immunological, or other factors. A *phe-*

nomenological heterogeneity parameter [**DE:** *not "factor"*] ζ, modifies the force of infection by a factor of $(S(t)/N)^\zeta$. Since $0 < S/N < 1$, a positive value of ζ will make the force of infection decrease as the remaining fraction susceptible decreases, capturing the fact that the most susceptible individuals tend to be infected first, after which the average level of susceptibility in the remaining population decreases.[**DE:** *revised that last sentence, which I kept misreading*] While modelers have considered the dynamics of epidemiological models incorporating incidence functions of this general form (9; 10), previous attempts to model phenomenological heterogeneity have used transmission proportional to an exponential function of prevalence (i.e., $SI/N \times e^{-(I/N)^n}$ where n is a shape parameter) rather than using a power law (11; 12). [**BB:** *JD please check that this makes sense/expand where necessary; check Dwyer? Williams et al 2006: "To allow for heterogeneity in sexual behaviour and to fit the observed asymptotic prevalence of infection, the transmission parameter takes the value λ_0 at the start of the epidemic and declines exponentially at rate α times the prevalence of infection." Granich et al. say "To allow for heterogeneity in sexual behaviour and for the observed steady state prevalence of HIV, we let the transmission decrease with the prevalence, P. If n=1, the decrease is exponential; if $n = \infty$, the decrease is a step function. Both have been used in previous models (5, 29)." Granich et al. also cite (13) (their ref. 29), but I don't see a transmission model in there anywhere ...*]

Internally, the model is defined by a *flow matrix* $\mathbf{M}$, the elements of which specify the *per capita* rates at which individuals move from one compartment to another. For the basic model, the only element of the flow matrix that needs to be recomputed at each time step is the incidence (flow from S to E); all other rates are piecewise constant (they are adjusted instantaneously when policies change, for example). This set-up allows for considerable flexibility:

1. For a numerical differential equation solver, we need the time derivatives of each compartment. The absolute rates are computed by columnwise multiplication by the state vector $\mathbf{s}$ ($\mathbf{F}_{ij} = \mathbf{M}_{ij}\mathbf{s}_j$). The gradient is the difference between the total flows into (column sums of $\mathbf{F}$) and out of (row sums of $\mathbf{F}$) each compartment.

2. For a discrete-time model, we compute the vectors of outflows and inflows as above, but simply compute the new states as (original state + (inflow - outflow)Δt).

3. We can also use a discrete-time model with a *hazard correction* [**DE:** *Is there a published paper that uses this terminology? or at least a paper we can cite that uses the idea? I believe that using this to make a discrete-time model is discussed by Hoppensteadt in an old book that I could find.*] to the flows to address the possibility that a state will go negative due to an overly large outflow. Instead of the total *per capita* outflow of a compartment being equal to the sum of *per capita* flow to each other compartment ($f_{\text{tot}} = \sum_i \mathbf{M}_{ij}$), we let the total outflow be

$$f'_{\text{tot}} = 1 - \exp(-f_{\text{tot}}\Delta t); \tag{1}$$

   the individual flows are then adjusted by a factor of $f'_{\text{tot}}/f_{\text{tot}}$. [**DE:** *Notation is not ideal here. f_{tot} depends on j; perhaps f_j would be better? Why f? Maybe M_j? Also f' conveys "derivative". Perhaps $\tilde{f}$ would be better?*] This adjustment accounts for the effects of depletion during the course of a time step.
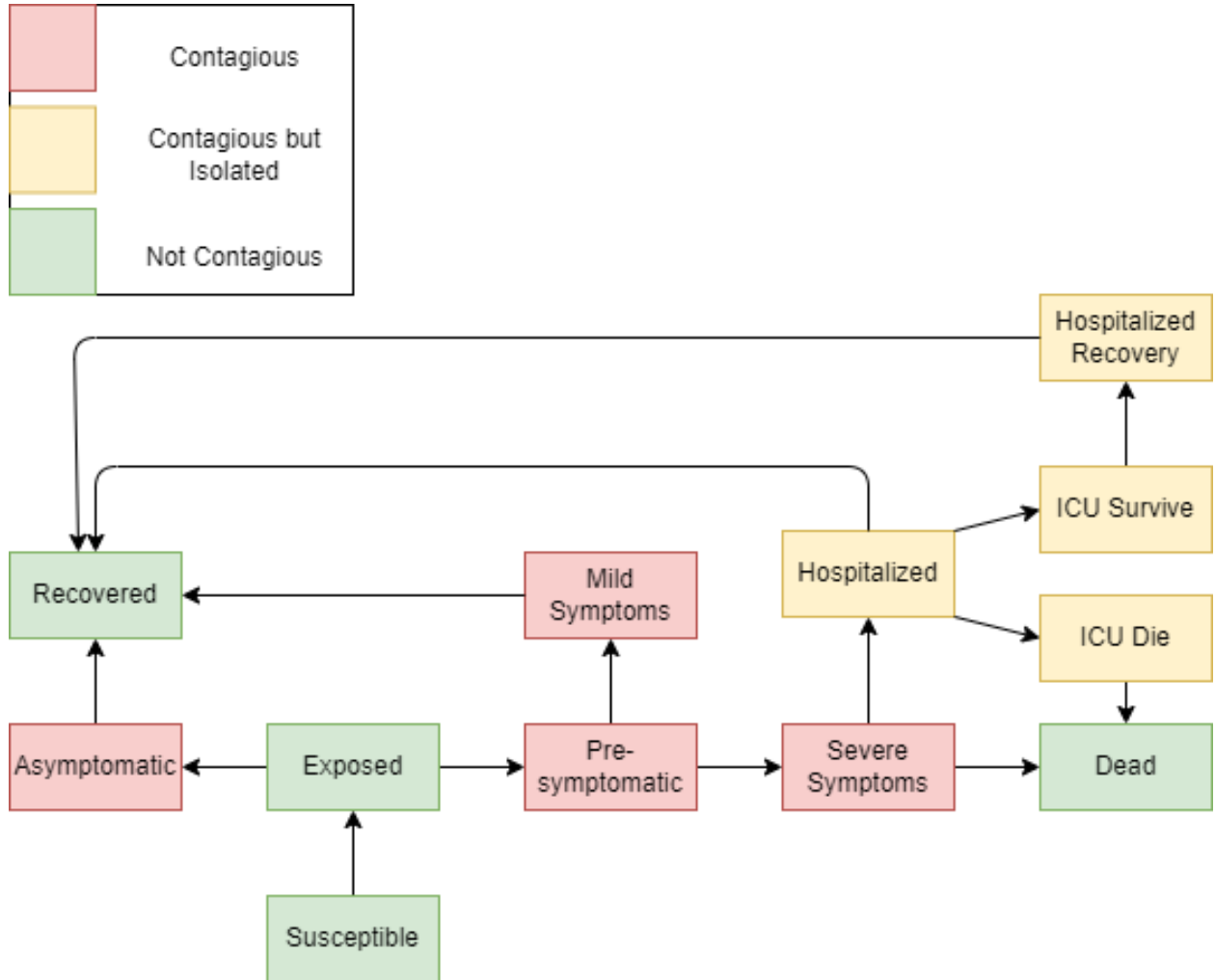
Figure 1: Flow chart for basic compartmental mechanistic transmission model. Compartments: S (susceptible), E (exposed), Ia (asymptomatic infection), Ip (presymptomatic infection), Im (mild/moderately symptomatic infection), Is (severely symptomatic infection), H (hospitalized [acute care]), ICUs (ICU with prognosis of survival), ICUd (ICU with prognosis of death), H2 (acute care after ICU stay). Compartments denoted by rectangles are accumulators, primarily used in the condensation step to compute incidences: R (recovered), D (dead), X (accumulator for cumulative hospital admissions). [**BB:** *prettify this further,* ***or*** *steal something better from Irena?*]

[**ML:** *Stole this diagram from Darren's product paper.*]

4. Finally, if we choose to run a fully stochastic simulation, the flow matrix $\mathbf{M}$ is what we need to sample the flows between compartments as Euler-multinomial deviates (14), which take the hazard correction (1) and use it to compute probabilities for draws from binomial or multinomial random deviates. [**DE:** *Maybe explicitly write the formula for the probabilities in terms of the hazard correction.*]

While the flow matrix description is convenient for most epidemiological dynamics, there are a few epidemiological processes that are more naturally captured by absolute rather than *per capita* rates, for example (1) when intensities of public health interventions such as numbers of tests or vaccines administered are reported by public health agencies; or (2) in models incorporating births or immigration. We have typically handled the former case by taking observed vaccination or testing rates and dividing them by current compartment sizes in order to set the relevant entries in the flow matrix; we have not yet tried to build models including inflows from outside the system (the latter case).[**DE:** *The phrasing "not yet tried" feels odd in a paper. Not sure how to express this. Maybe "the version of the software described here does not include inflows from outside the system (the latter case)."*]

We typically compute the model dynamics deterministically, as a discrete-time model with a hazard correction (1) (option 3 above). Once the trajectories are computed, we reduce the full state vector to a more convenient, collapsed state vector in a step we call *condensation*, for example by summing all of the infectious compartments to a single I state vector, or collapsing the different acute-care (H, $H2$) or ICU (ICU_s, ICU_d) [**DE:** *We should be consistent wrt font used for compartments. We were using* `\texttt` *earlier.*] compartments. As well as allowing us to visualize model results more conveniently, condensation also allows us to compare the simulated state vector to available data streams. In addition to summing compartments, we can also compute incidences as time-lagged differences of accumulator compartments (for example, differencing accumulated deaths D to derive a mortality rate) or perform more complicated operations such as convolution[**DE:** *maybe cite Goldstein et al 2009, PNAS*]. Our main use of convolution is to convert incidence—the force of infection (FOI) multiplied by the number of susceptibles—to a case-reporting (CR) time series:

$$\mathrm{CR}(t) = \sum_i \phi(i)(\mathrm{FOI}(t-i)S(t-i)), \tag{2}$$

where we typically set $\phi(i)$ to be a Gamma distribution with moments chosen to match empirical estimates of case-reporting delays. [**ML:** *Do we have/need a formula for FOI?*] [**BB:** *how did we pick these values? Probably don't have a formal ref but some statement of what we used to guess that this was reasonable would be good ...*]. [**JD:** *Another option here would just be to say where ϕ is typically a Gamma distribution with moments chosen to match literature estimates¿*] [**BB:** *done, but still not clear that there are a lot of such estimates in the literature ...*][**DE:** *I added "empirical", which can cover unpublished estimates.*] When computing case reports from incidence we also assume a case-report proportion `c_prop` to account for the fact that the majority of COVID infections are never reported (15); this value is usually calibrated from data.

After condensation, the model also allows us to add observation error, which we typically simulate from a negative binomial distribution with a variable-specific dispersion parameter (16).

Expansion to accomodate testing

At the cost of additional complexity, we can add explicit testing compartments to the model. In the simpler version of the model, we assume that a specified fraction of infections are reported as cases (or we calibrate this fraction from joint data on cases and hospitalizations), and impose a distributed delay between infection and reporting via convolution (2). Here, we instead expand the susceptible and all infected compartments factorially to include the possibilities that individuals in those epidemiological classes have one of four testing statuses:

- untested;

- tested and awaiting negative results;

- tested and awaiting positive results; or

- tested positive.

After receiving negative results (whether true negatives, from individuals in S, or false negatives, from individuals in one of the infectious compartments), individuals cycle back to the untested sub-compartment where they can be tested again. After receiving positive results, they remain in the positively tested sub-compartment; depending on the model parameters, their transmission may be reduced due to self-isolation (controlled by the parameter `iso_t`). We assume here that people waiting for test results do not isolate. Gharouni *et al.* (17) thoroughly analyze the epidemiological consequences of this structure in a simpler framework that expands a basic SIR model rather than the COVID-specific compartmental model (Figure 1) as the foundation. Individuals may progress between epidemiological compartments (e.g., becoming infectious or recovering) while awaiting the results of tests. In general, progressing individuals move to the same testing subcompartment, e.g., from `Im`-negative-waiting to `R`-negative-waiting.

We define a weight vector w across epidemic compartments that gives weight w_a to asymptomatic classes (`S`, `E`, `Ia`, `Ip`, `R`) and 1 to symptomatic classes (i.e., all other classes). If W is the weighted sum of compartment occupancies X_i, i.e.,

$$W = \sum_i w_i X_i / N \,, \tag{3}$$

then for a daily *per capita* testing rate ρ we might expect that the corresponding *per capita* testing rate in compartment i would be $\mathcal{T}_i = \rho w_i / W$. [**DE:** *We should try to keep notation consistent with (17). I changed T_i to $\mathcal{T}_i$. Not sure that's sufficient.*] However, under some extreme conditions (if testing is so extreme that few untested symptomatic people are left) this formulation can allow the *per capita* testing rate to explode. To mitigate this problem we add a maximum daily *per capita* testing rate τ to the model such that

$$\mathcal{T}_i = \frac{\rho \tau w_i}{\tau W + \rho} \,; \tag{4}$$

see Appendix A.5 of (17) for details. The testing flows out of each untested subcompartment are divided into flows to "positive waiting" and "negative waiting" compartments according

7

to the infection status of the relevant compartment and specificity/sensitivity parameters (specificity/sensitivity refers to the probability that a truely negative/positive individual tests negative/positive). [**DE:** *phrasing could probably be improved, but I felt we should define specificity and sensitivity*] If tests are assumed to be perfectly specific and sensitive, then all individuals from non-infectious compartments (S, E, R) enter the "negative waiting" compartment and those from infectious compartments enter the "positive waiting" compartment.

We assume that all individuals admitted to the hospital for COVID-19 are immediately tested. (These tests are not included in the accounting of test distribution above, but in the COVID setting they represent a small fraction of the overall number of tests administered.)

Model parameterization

The model allows for time-varying, piecewise-constant changes in any parameter. Our early analyses focused on changes in time-varying effective reproductive number ($\mathcal{R}_t$) due to behaviour change and non-pharmaceutical interventions, which we model by changing the transmission rate. The transmission rate $\beta(t)$ is taken to be a time-varying function of the form $\beta_0\beta_1(t)$ where β_0 is the baseline value for transmission from symptomatic individuals. Individuals in different symptomatic classes (presymptomatic, asymptomatic, mild, severe) may have their transmission modified by a specified multiplier (given as model parameters); we assume that hospital transmission is negligible. [**DE:** *Can we find a reference to support most transmission occuring in the community / outside hospitals?*] The time-varying (relative) transmission $\beta_1(t)$ can incorporate a variety of different effects, one at a time or in combination:

- abrupt (piecewise) changes on specified dates when control measures are known to have been implemented;

- proportional to a power of observed mobility or some other exogenous proxy for contact behaviour:
$$\beta_1(t) \propto \log(M(t))^{p_{\mathrm{mob}}} \,, \tag{5}$$
where $M(t)$ is the relative mobility index at time t, for some power $p_{\mathrm{mob}} > 0$;

- according to an arbitrary spline curve, i.e., a linear combination of components of a B-spline basis.

All of these sub-models for temporal change in the transmission rate can be subsumed under a single log-linear model:
$$\log\beta(t) = \log\beta_0 + \boldsymbol{X}\boldsymbol{c} \,, \tag{6}$$
where $\boldsymbol{X}$ is a model matrix [**DE:** *not ideal that we have X_i meaning something else entirely above*] that can contain any combination of covariates and $\boldsymbol{c}$ is a vector of covariates determining differences in the log of the relative transmission rate. [**DE:** *Only $\boldsymbol{c}$ can vary in time, right? We should therefore write $\boldsymbol{c}(t)$ in Equation* (6).] In particular, piecewise breaks correspond to indicator variables for which period an observation falls in; the mobility model
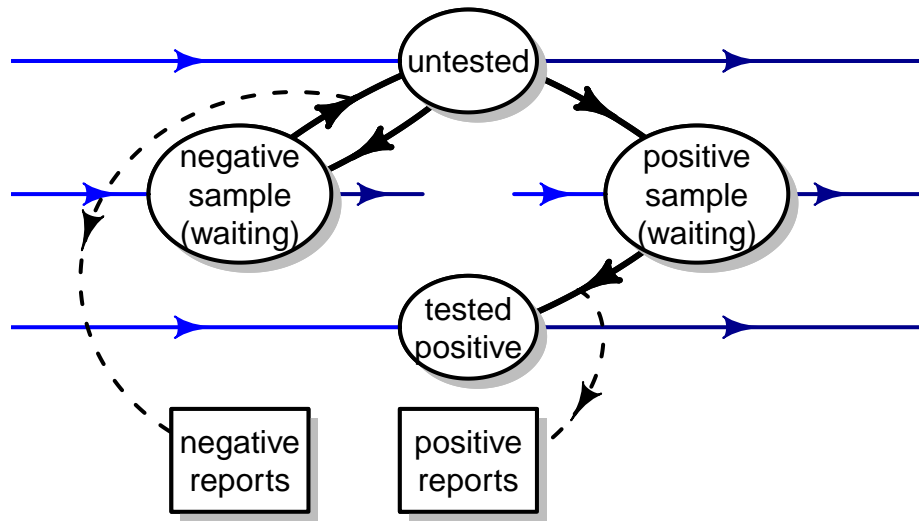
Figure 2: Testing flow. Every epidemiological compartment is subdivided into the four sub-compartments shown. Black arrows represent flows between subcompartments due to testing processes (test administration, reporting of tests); blue arrows represent progression between epidemiological compartments. Dashed arrows represent the accumulation of negative and positive test reports, which can be compared against data.

corresponds to a column containing the log of relative mobility; and the spline model corresponds to a set of columns containing the basis vectors of a B-spline basis with specified knots.

In practice, when we use the model without incorporating covariates we adopt a simpler strategy of providing a list of the breakpoints and the parameters that change at those breakpoints — a special case of the more general log-linear model. In the models illustrated below, we pick breakpoints to denote several time periods $\{P_1, P_2, \ldots\}$ and construct $\boldsymbol{X}$ to allow the effects of mobility, and the baseline contact rate, to change at breakpoints. Specifically, for each breakpoint we define a logistic transition curve

$$S(j) = \frac{1}{1 + e^{[t - t_{\mathrm{brk}}(j)]/s}} \,, \tag{7}$$

where s (set to 3 days) defines the speed of transition. Because the baseline transmission rate is included in the model (as parameter `beta0`), we do not need to include an intercept in $\boldsymbol{X}$; the first column is $\log(M(t))$, the dependence of (log) transmission rate on (log) mobility during the first period. For $j \geq 1$, subsequent columns $2j$ and $2j+1$ of $\boldsymbol{X}$ are defined as $S(j)$ and $S(j) \log M(t)$, respectively. The parameters associated with these columns denote the change in baseline contact rate and the change in the effect of mobility between periods j and $j+1$.

Derived parameters and parameter setting

It is useful to be able to compute several quantities derived from the parameters of a model, in particular:

1. the dominant eigenvector of the system in the exponential growth phase (as described below, we use this eigenvector to set sensible initial conditions for all state variables, including those that cannot be observed);

2. the basic reproduction number $\mathcal{R}_0$ and the intrinsic growth rate r;

3. the mean and coefficient of variation of the generation interval.

In principle we could compute these values directly from the flow matrix, by constructing the Jacobian matrix and the next-generation matrix and performing the appropriate eigenvector/eigenvalue calculations (as discussed by (18) for differential equations and (19) for discrete-time systems). However, we found it simpler to derive these values by simulation.

To compute the eigenvector, we run a simulation where we set the outflow from the susceptible compartment to zero (while maintaining the *inflow* from S to E), which mimics the dynamics near the disease-free equilibrium where susceptible depletion is negligible. After running the simulation for a long time (100 steps by default), the state vector is close to the eigenvector. We use this value to set starting conditions when starting near the beginning of the epidemic. It is easy to specify a scalar value (say, 1% of the population) to indicate the initial size of the epidemic; by distributing these individuals according to the calculated eigenvector, we reduce numerical instability at the beginning of our simulations.

10

To compute the other summaries ($\mathcal{R}_0$, r, and moments of the generation interval) we rely on the fact that the software saves the force of infection at each time step. We simulate the progression of a *single* individual through the infection process, i.e., setting the population size to 1 and setting the initial state to $E = 1$ with all other compartments empty. Simulating deterministically then generates a time course of the probability that an individual is in any given box at a particular age [**ML:** *stage? or infection state?*] of infection, and thus also the expected force of infection generated by a single infectious individual, as a function of time since infection. This vector $K(t)$ is the same as the transmission kernel in a renewal equation (20). We can easily compute

$$\mathcal{R}_0 = \sum_t K(t) \, , \tag{8}$$

mean generation interval

$$\bar{G} = \sum_t \frac{t \, K(t)}{\mathcal{R}_0} \, , \tag{9}$$

and generation interval coefficient of variation,

$$\mathrm{CV}(G) = \frac{\sqrt{\frac{1}{\mathcal{R}_0} \sum_t K(t)(t - \bar{G})^2}}{\bar{G}} \, . \tag{10}$$

The growth rate can be computed by numerically solving for r in the Euler-Lotka equation,

$$\sum_t K(t) e^{-rt} = 1 \, . \tag{11}$$

[**JD:** *Suggest dropping the detailed equations for Gbar and CV, but instead adding g(t) and saying that we also calculate its mean and CV.*] [**JD:** *Suggested easy ref is ChamDush15: http://dx.doi.org/10.1098/rspb.2015.2026 (instead of current Cham18)*] [**BB:** *JD: please go ahead and implement this change*]

In addition to its use in summarizing a given set of parameters, the computation of $\mathcal{R}_0$, r, and the generation interval is useful as an initial step in calibrating the model. Estimates of these summary statistics are more broadly available (21), and more epidemiologically relevant, than the more specific mechanistic parameters describing the relative infectiousness and duration of each of the different infectious compartments (although this detailed information is still important for determining the effectiveness of interventions like contact tracing). We typically start with mechanistic parameters gathered from the literature and pre-calibrate them to specified target values of r (which is easy to estimate from the observed initial growth rate of the epidemic in a region) and the mean generation interval by adjusting β_0 (baseline transmission) and simultaneously scaling the values of all of the epidemiological transition rates (σ, γ_s, γ_m, γ_a; see § 2) by a single factor until the target values are achieved.

Calibration

Once we have run a deterministic model simulation for a particular set of parameters (including a starting number infected, distributed across non-susceptible classes according to

11

the exponential-phase eigenvector computed as described above) and we have some set of time series data to calibrate against (for example, case reports and hospital admissions), we can calculate a log-likelihood (e.g., (22)). We assume that every observation is independently negative binomially distributed, with a series-specific estimated dispersion parameter (i.e., the variability in cases, hospital admissions, *etc.*, will differ). We use standard nonlinear optimization algorithms built into R, such as Nelder-Mead, to find the maximum likelihood estimates; when we have had difficulty with numerical instability, we have performed an initial fit with differential evolution (23) followed by a final fit with Nelder-Mead. [**ML:** *Do we need a line about process/dynamical error here? Say we have it or we are not including it.*]

A link function can be added for any parameter in the model to constrain it to a sensible domain; the user specifies this by adding an appropriate prefix to the name of the parameter in the list of starting values for parameters to be calibrated. For example, specifying `log_beta0 = -1` would specify that the baseline transmission parameter β_0 should be calibrated on the log scale, ensuring that the value of β_0 is always positive, and using a starting value of -1 on the log scale (i.e., initial $\beta_0 = \exp(-1)$). Specifying `logit_nonhosp_mort = -0.2` would specify that the value of `nonhosp_mort` (the fraction of mortality that occurs outside hospitals) should be fitted on a logit, or log-odds, scale, ensuring that it is bounded between 0 and 1, and using a starting value of $\beta_0 = \text{logit}^{-1}(-0.2) = 0.45$. [**JD:** *Maybe make this number a bit smaller?*]

The model includes a general framework for adding a prior probability distribution for any parameter, using any distribution available in R. For example, `dbeta(nonhosp_mort, 2, 2)` would specify a $\text{Beta}(2, 2)$ prior for `nonhosp_mort`. We do not need to adopt a fully Bayesian framework to make use of priors; instead, we can think of them as convenient regularizing factors to keep the model-fitting process numerically stable. If we do want to be Bayesian, then the fitting procedure described above will return maximum *a posteriori* (MAP) parameter values, not a sample from the full posterior distribution as is standard with frameworks that use Markov chain Monte Carlo. [**DE:** *A standard ref is needed for MCMC. (22) or something else?*]

For the province of Ontario, Canada, in 2020 we calibrated to deaths, and new confirmations, all of which were available publicly. In the expanded model, we can include time series of both positive and total tests (in our calibration. Reported new confirmations (daily positive tests) are the most reliable and voluminous source of epidemic information. Unfortunately, they are also subject to many inherent biases, including substantial variation over time in **testing intensity** (*i.e.,* tests *per capita* per day).

We simultaneously estimate the temporal pattern of the transmission rate [Equation (6)] (two mobility intercepts and three slopes), and several basic model parameters (Table 1). Other model parameters are taken from the literature (§ 2).

We do not include hospitalization and ICU occupancy in our calibration because our model assumes all severe cases go through ICU, whereas during the early pandemic period that we focus on here, many severe cases actually occurred entirely in Long Term Care Facilities (LTCFs); thus, our model would not properly match the observed time series of hospitalization and ICU occupancy, or at least would fail to account for severe cases outside of hospitals. Furthermore, capacity limitations in the health care system lead to ceilings on hospitalization that are not represented in our model. [**BB:** *I think I got this statement*

Parameter	Value	Meaning
C_a	0.667	Relative asymptomatic transmissibility
C_p	1	Relative pre-symptomatic transmissibility
C_m	1	Relative mildly symptomatic transmissibility
C_s	1	Relative severely symptomatic transmissibility
α	0.39	Proportion of infections that are asymptomatic
σ	0.303	1 / mean latent period
γ_a	0.143	1 / mean days in asymptomatic infectious class
γ_m	0.143	1 / mean days in mildly symptomatic infectious class
γ_s	0.175	1 / mean days in severely symptomatic infectious class
γ_p	0.833	1 / mean days in pre-symptomatic infectious class
ρ	0.1	1 / mean days in acute care
δ	0	Proportion of acute care patients who die
ϕ_1	0.76	Proportion of severe infections that do not require ICU
ϕ_2	0.26	Proportion of ICU patients who die
ψ_1	0.05	1 / mean days in ICU if survive
ψ_2	0.125	1 / mean days in ICU if die
ψ_3	0.2	1 / mean days post-ICU until discharge
c_{prop}	0.1	Fraction of infections reported
W_{asymp}	0.1	Relative testing intensity in asymptomatic compartments

369 *right but am not sure: WZ/JD/DE, please check for correctness/clarity.*] [**DE:** *I removed*
370 *"properly" since we make no attempt to represent capacity limits to my knowledge. What*
371 *you describe accurately reflects our rationale in 2020, but we are currently happy to calibrate*
372 *to hosp and ICU. Can we rephrase in a way that is consistent with us later deciding that*
373 *fitting to hosp and ICU is OK? Is my additional phrase "during the early pandemic period*
374 *that we focus on here" a solution?*] [**ML:** *RE:Ben, this is reads fine to me.*]

Forecasting

376 Our calibrations yield values for the parameters of our deterministic model (Table 1). Using
377 the calibrated parameters and the estimated covariance matrix of the sampling distribution
378 of the parameters, we draw many (typically 1000) sets of parameters from a multivariate
379 normal distribution (22; 24) and feed them into the deterministic model to generate an
380 ensemble of forecasts. We extend the end date of the calibration window (i.e., the last
381 observed data point for calibration) to create a forecast window. In the forecast window, the
382 user can either investigate scenarios by inputting covariates (i.e., relative mobility and/or
383 testing rates, depending on the model) or assume that the last set of inputs remains constant
384 through the forecasting window to obtain a *status quo* forecast.

2.1 Data sources

Public COVID-19 data for Ontario, Canada

Daily reported data on COVID-19 testing and outcomes (confirmed cases, deaths, hospital and ICU occupancies) are publicly available for Ontario on the official provincial website: https://data.ontario.ca/dataset/f4f86e54-872d-43f8-8a86-3892fd3cb5e6/resource/ed270bb8-340b-41f9-a7c6-e8ef587e6d11/download/covidtesting.csv. However, this dataset does not include testing counts before 15 April 2020. Since the beginning of the COVID-19 pandemic (and until the present at the time of writing), one of us (ML) has maintained a public web site containing Canadian COVID-19 data at the provincial level. Data are frequently downloaded from a variety of sources and cleaned. See https://wzmli.github.io/COVID19-Canada/.

Mobility data

We use mobility data from Apple[1] and Google[2]. From these data we derive a **relative mobility index** using the "driving" index from Apple and the "retail and recreation" and "workplaces" indices from Google; we compute a 7-day moving average of these indices, rescale all of them to have a baseline (pre-pandemic) value of 1.0, and average the three indices (equally weighted), to obtain an overall index for each day.

[**DE:** *The "important dates" section is in* `cuts.tex`. *Do we really not use those dates at all?*] [**ML:** *We didn't use it because mobility is handling most of the changes in transmission.*] [**ML:** *We replaced it with the important dates when we gave up on mobility later on. I do prefer mobility for this paper for a more generalized approach to incorprating different exogenous data (I guess same arugement applies to important dates as well).*]

---

[1] https://raw.githubusercontent.com/ActiveConclusion/COVID19_mobility/master/apple_reports/applemobilitytrends.csv

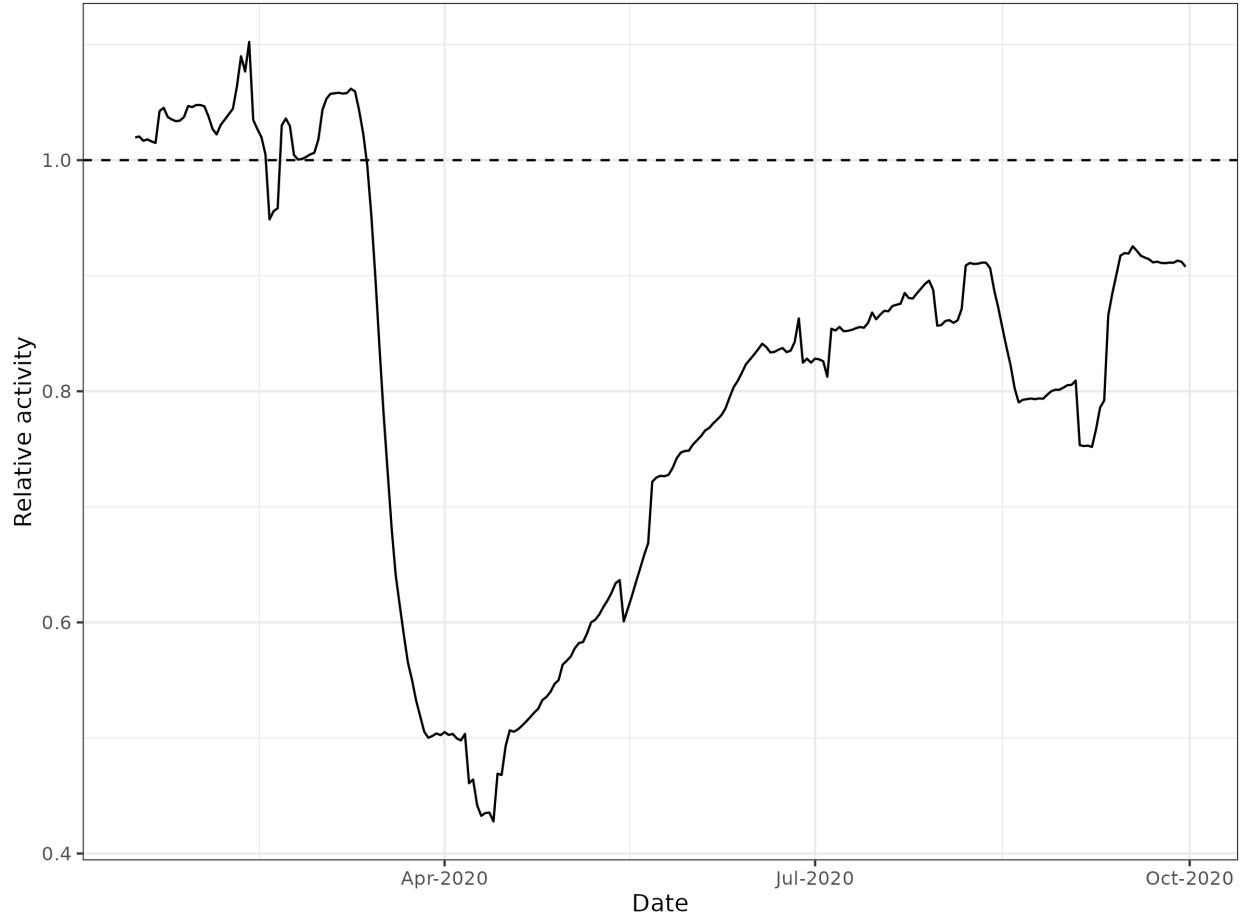[2] https://www.gstatic.com/covid19/mobility/Global_Mobility_Report.csv

Figure 3: Ontario mobility [**BB:** *WZ: a previous version of this graph looked very different, with giant day-of-week effects. The current version I'm seeing is much smoother ... do we know what happened/which was right?*] [**ML:** *Need to add the mobility breakdates in the diagram; the model uses a different slope and intercept at different breakdates.*]

3 Results

We used the Ontario data and fitted it to two different models to demonstrate the capability of this modeling framework. The first model is our simplest model, following the schematic illustrated in Figure 1. This model calibrates to new confirmation and death time series and includes mobility with two additional breaks and phenomenological heterogeneity. We will refer to this as the "base model". The second model extends the base model by incorporating the testing structure illustrated in Figure 2. We will refer to the second model as the "testify model". Both models were used to forecast 30 days ahead. We did not explicitly include break dates (i.e., time-varying transmission rates) to match different public health measures and restrictions in place within the fitting window, the transmission rate is a function of time-varying intercepts and slopes on mobility index which is a sensible proxy for changes in behaviour responding to public health restrictions.[**DE:** *That was true in 2020 but is no longer true. We should rephase. We could perhaps say "was a good proxy... during the focal*

15

time period" or something like that, but we then need to say this stopped being true eventually and some reference(s) to support this would be very helpful.] [**ML:** *This is incorrect, we did use it in mobility, but we have to double check if it is used to match policy change vs goodness of fit?!? for the mobility pattern.*]

[**ML:** *Do we want to include some simulation to show the model works and can do the calibration?*] [**DE:** *I think that would strengthen the paper, if people have the strength for it. While I'm always pushing for testing against simulations, I am most concerned about getting this paper finished and submitted soon. If a fit to a simulation can be achieved quickly, then great.*] [**ML:** *Pretty drainned, happy to delete this suggestion unless the results are "really" thin*]

3.1 Base model

Figure 4 shows the fit to new confirmations and death time series from Febrary 24 2020 to August 30 2020. This time frame captures the first wave of the COVID outbreak and the beginning of the second wave in Ontario. The model capture to fit the new confirmations data well within the fitting window but not for the death time series. The model has adaqualt flexibility to capture the effective transmission parameter with mobility data and changes in mobility effects for each mobility break [**ML:** *need to reword*], and mortality parameters were assumed to be constant at the optimized level resulting in a overestimation of deaths.

The 30 day forecast is expected to increase; when over-laying actual observed time series for that time-frame, the observed time series are lower than predicted but within 95% prediction interval for the new confirmations and missed all the death observations.
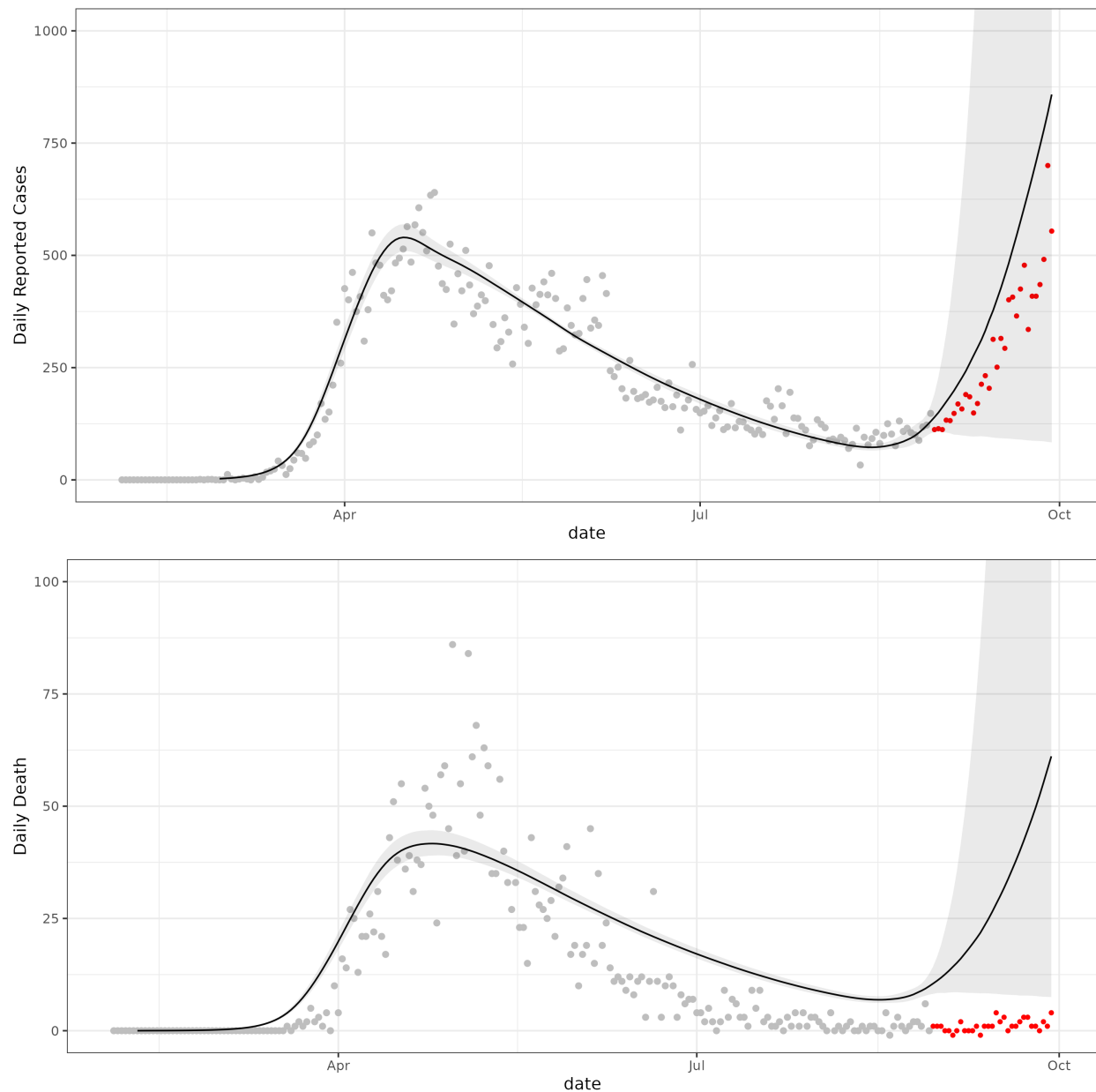
[**ML:** *add a comment mobility is freezed?*]

Figure 4: Ontario base forecast. [**BB:** *need to explain what's going on here; at the very least the caption should say (even if it's kind of obvious) that gray points represent calibration points, red are forecast points. Why is the death forecast so far off? Should we just leave it out (probably)? (WZ ...)*]
[**JD:** *More comfortable leaving it out after we figure it out. It's bizarre. I could be convinced that it's water under the bridge, but it's worth a look.*]

Parameter	Estimate	Meaning
$E(0)$	9.203	Initial number exposed
β_0	1.070	Baseline (non-intervention) transmission across categories
eta	0.073	Probability of mortality without hospitalization
ζ	40.050	Phenomenological heterogeneity parameter
θ_{report}	$\gg 1000$	Negative binomial dispersion parameter (case reports)
θ_{death}	$\gg 1000$	Negative binomial dispersion parameter (deaths)
β_1	-0.621	Relative change in transmission after Apr 01
β_2	0.218	Relative change in transmission after Apr 07
$\beta_{\text{mob},0}$	0.845	Mobility power
$\beta_{\text{mob},1}$	0.153	Change in mobility power after Apr 01
$\beta_{\text{mob},2}$	1.407	Change in mobility power after Apr 07

Table 1:   Parameter estimates for base model calibration. [**BB:** *JD/DE: should some parameters (change in transmission etc.) be back-transformed? Are parameter names comprehensible? Should we add confidence intervals/std errs? We should explain that dispersion model is converging to Poisson; do we need to justify/explain the very high value of ζ (phenom het)?*][**DE:** *We definitely need to comment on the value of ζ; it seems absurdly high.*] [**DE:** *CIs would be nice.*]

3.2 Testify model

The base model is relatively simple but does not account for testing practices. Testing strategies changed frequently over the course of the pandemic due to availability of testing facilities and shifting regulations on eligibility for testing.

Figure 5 shows the results from calibrating the model to positive tests, with the testing flow incorporated and with observed testing rates fed into the model as a known series. The variation in the predicted line is driven by high-frequency changes in testing rate, including day-of-week effects. It also shows the daily testing in Ontario in the calibration time frame. The vertical line is the cut-off between the calibration window (before September 1st 2020) and 30 days beyond it for the forecast window. The testing intensity over the forecast window is assumed to be constant at the last observed testing intensity on August 30th 2020 (since the actual increase in testing that occurred during the forecast window was not known at the time of the forecast). The forecast under predicts the number of new positive cases in the forecast time-frame. The actual testing intensitive did increase during the forecast period.

[**DE:** *I suggest using open circles during the forecast window, as we do for observed cases and hospitalizations.*]
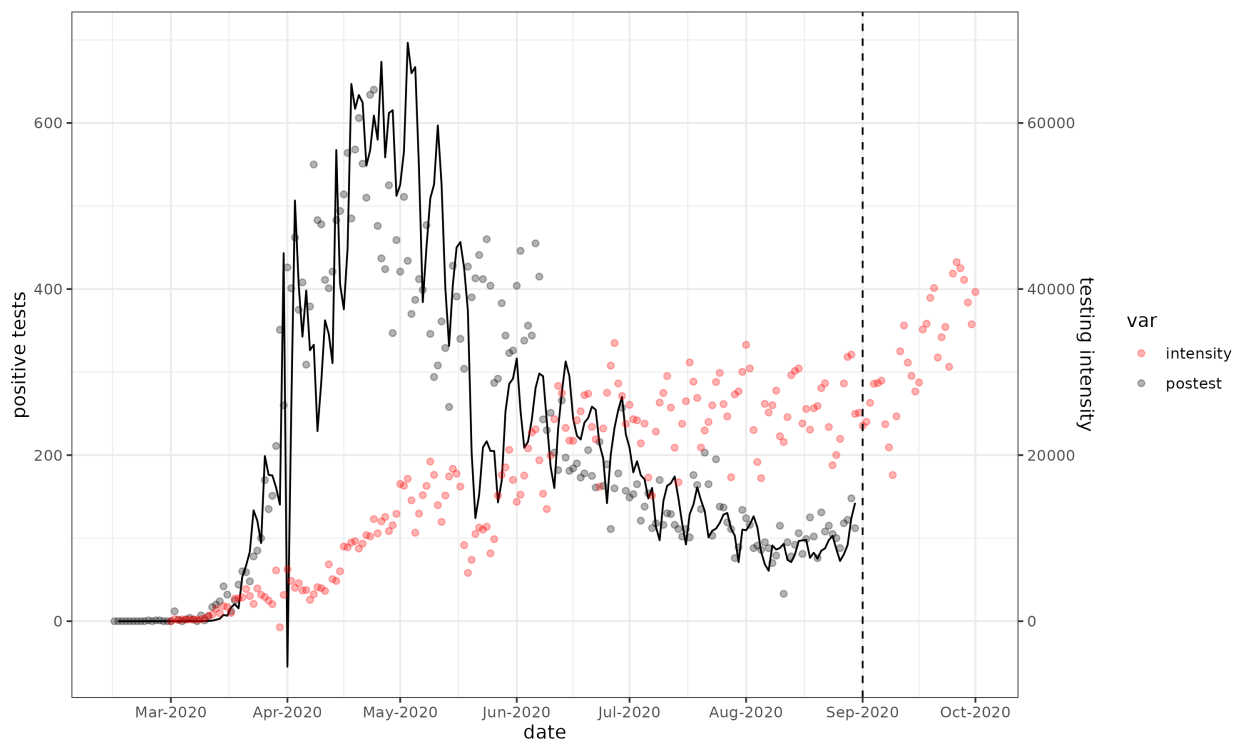

Figure 5: Calibration of positive tests and testing intensity

[**ML:** *We (I) probably should make another forecast using the testing intensity of the forecast window. Thoughts?*] [**DE:** *Yes it would be good to see how using the correct testing intensity affects the forecast. This needs to be expressed carefully to avoid confusion. We need to get across that this emphasizes how much better the forecast would be with advance knowledge of how testing intensity would change over the forecast period.*]
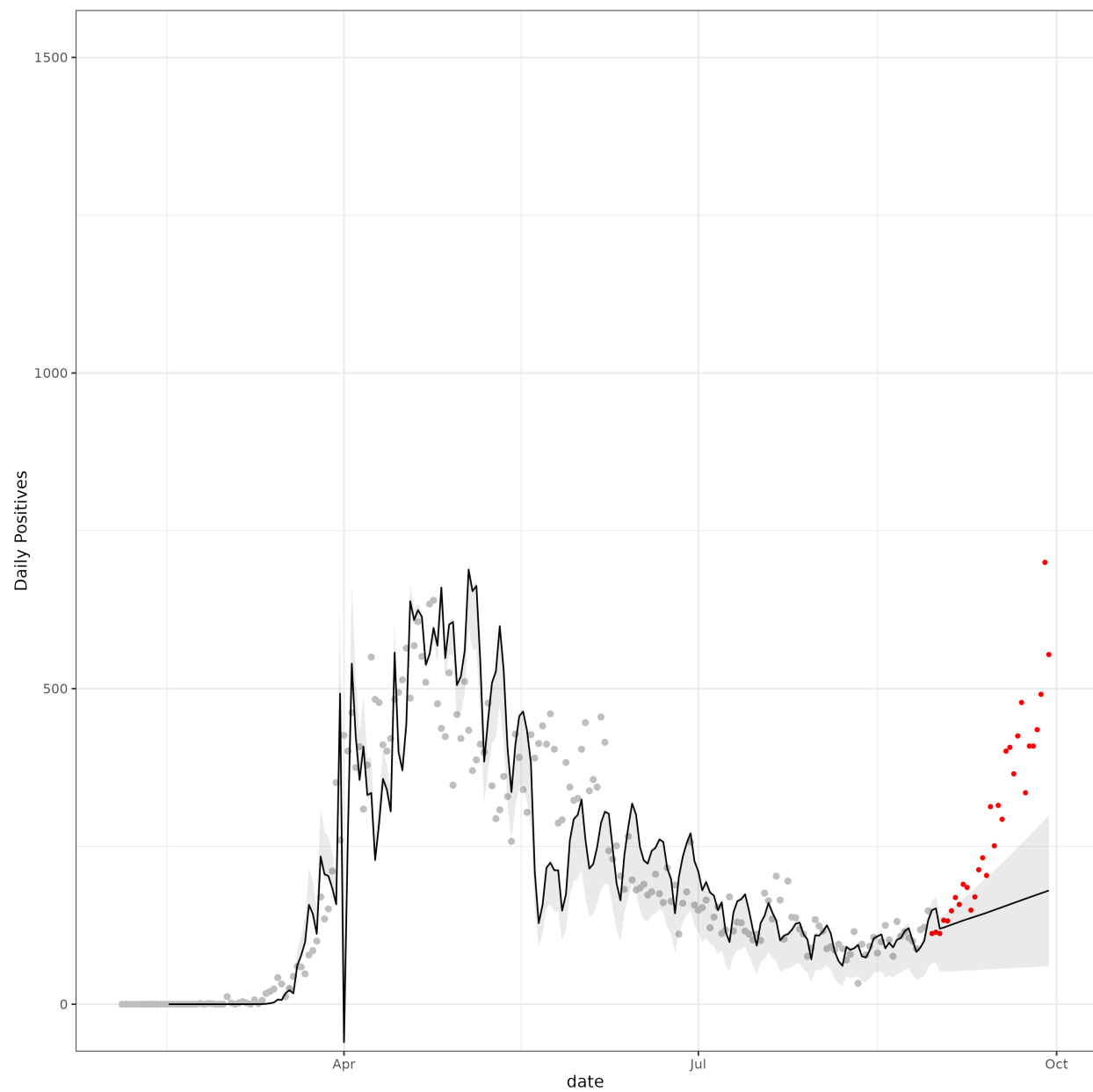
19

Figure 6: Ontario calibration with testing flows

Parameter	Estimate	Meaning
$E(0)$	3.634	Initial number exposed
β_0	0.990	Baseline (non-intervention) transmission across categories
ζ	3.910	Phenomenological heterogeneity parameter
θ_{postest}	$\gg 1000$	Negative binomial dispersion parameter (positive tests)
β_1	-0.583	Relative change in transmission after Apr 01
β_2	-0.710	Relative change in transmission after Apr 07
$\beta_{\text{mob},0}$	1.405	Mobility power
$\beta_{\text{mob},1}$	0.767	Change in mobility power after Apr 01
$\beta_{\text{mob},2}$	-1.669	Change in mobility power after Apr 07

Table 2: Parameter estimates for testify model calibration. [**BB:** *Contrast with non-testify results. Maybe instead of this table we can include a figure that shows the two values along with confidence intervals? (Have to be a little careful about the scaling - i.e., which parameters it makes sense to show on a common scale ...)*] [**DE:** *Can we explain the factor of 10 difference in ζ?*]

4 Discussion

We have fitted two models that vary in complexity to SARS-CoV-2 daily reporting and death time series data for Ontario, Canada, from 1 March to 30 August 2020. Our simple base model is typicalof compartmental epidemic models used to study infectious disease epidemics. The testify model [**DE:** *The "testify" terminology has not been defined. Do we want to define it and use it because we like the term? or should we just refer to the "model with testing flows" or something like that so referees don't think we're goofy?*] is a practical expansion of the base model that allows us to take advantage of data on time-varying testing practices and policies.

[**ML:** *Talk about the assumptions of the forecast period for each model. Freezing vs extending/predicting mobility and testing intensity.*]

4.1 Limitations

Our model assumes homogeneous mixing of the population. No age-related or spatial contact structure is considered. [**DE:** *The following sentence needs to be rewritten. I don't understand what was intended to be expressed.*] We did not include hospitalization in the calibrations due to reporting from capacity issues which became a serious issue with the admissions and occpuancy underreporting the severity of the pandemic. In addition, long-term care facilities (LTCFs), where many elderly people in Canada have died without going to hospital, have not been modelled explicitly. While we do not anticipate any qualitative differences in results, explicitly creating compartments and calibrating to data for LTCFs would likely allow us to match ICU occupancy and forecast pressure on ICUs more accurately but will not resolve the issues on hospital intensity. [**DE:** *the last phrase needs to be rephrased. What is "hospital intensity"?*] [**ML:** *How busy the hospitals get. I.e. addmission only shows how many the hosp admit but not how many people how actually showed up that needs treatment.*]

4.2 Conclusions

[**DE:** *This conclusion paragraph leaves a lot to be desired. I think we should discuss what we think the key take-home messages should be, and then write a new conclusion paragraph. We should also state that we have continue to develop the software during the pandemic, and will describe further innovations in subsequent papers.*]

SARS-CoV-2 continues to be a global burden after 3 years. Here we demonstrated a simple modeling framework that can flexibly fit epidemic data and easily allow additional complexities to account for features during the epidemic. We learned two things about fitting epidemic data. First, it is important to have ways to incorporate time-varying parameterizations fitting changes in the observed data or behavioural changes. Time-varying transmission rates played an important role in our calibration. Second, allowing for testing intensity to disaggregate the transmission rates through time (i.e. does increase in positive cases/confirmation due to an increase in transmission or testing capacity). The flexibility of incorporating various time-vary parameterizations allows users to make more realistic features and scenario exploration.

22

[**DE:** \input{nextsteps.tex} *has been commented out. There may be useful comments in there that would help with Discussion/Conclusions.*]

5   To do

- base model is calibrating to new cases and death

- testify is calibrating to positive test only

- consistency of model (death forecast looks bad because the mort parameters are not time-varying), we can drop death and only calibrate to cases and this is now compariable with testify.

- Bt and Rt plot? Even if they are a bit redundant, they are a bit different imo. If we include it, we need to make the plot and need a paragraph in the results.

- find and clean up `estparmtab`/`litparmtab` (done)

- clean up Tables 1 and 2 (WZ? WZ: Combine it into a single table)

- should we include fits to simulations or not? (Now sure what this means)

- Clean up figures, improve captions

- Discussion and conclusions

- output/include figs as PDF rather than PNG (JD/WZ?) [**ML:** *output ok, brute force the actual non-output graphics from PNG to PDF?*]

- [**DE:** *In the original PHAC report, we showed fits to all the large provinces. I'm OK with discussing Ontario only, but just checking that change was intentional. I've move the commented-out appendix to* `cuts.tex`.]

References

[1] Ontario Confirms First Case of Wuhan Novel Coronavirus; 2020. Available from: https://news.ontario.ca/mohltc/en/2020/01/ontario-confirms-first-case-of-wuhan-novel-coronavirus.html.

[2] Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. New England Journal of Medicine. 2020;382:1199–1207.

[3] Fauci AS, Lane HC, Redfield RR. Covid-19 – Navigating the Uncharted. New England Journal of Medicine. 2020;382:1268–1269.

[4] Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature. 2020;584:257–261.

[5] Friston KJ, Parr T, Zeidman P, Razi A, Flandin G, Daunizeau J, et al. Dynamic causal modelling of COVID-19. Wellcome Open Research. 2020 Aug;5:89. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7431977/.

[6] Childs ML, Kain MP, Harris MJ, Kirk D, Couper L, Nova N, et al. The impact of long-term non-pharmaceutical interventions on COVID-19 epidemic dynamics and control: the value and limitations of early models. Proceedings of the Royal Society B. 2021;288(1957):20210811.

[7] Tuite AR, Fisman DN, Greer AL. Mathematical modelling of COVID-19 transmission and mitigation strategies in the population of Ontario, Canada. Cmaj. 2020;192(19):E497–E505.

[8] Kain MP, Childs ML, Becker AD, Mordecai EA. Chopping the tail: How preventing superspreading can help to maintain COVID-19 control. Epidemics. 2021 Mar;34:100430. Available from: https://www.sciencedirect.com/science/article/pii/S1755436520300487.

[9] Wilson EB, Worcester J. Damping of Epidemic Waves. Proceedings of the National Academy of Sciences. 1945 Sep;31(9):294–298. Publisher: Proceedings of the National Academy of Sciences. Available from: https://www.pnas.org/doi/abs/10.1073/pnas.31.9.294.

[10] Liu Wm, Hethcote HW, Levin SA. Dynamical behavior of epidemiological models with nonlinear incidence rates. Journal of Mathematical Biology. 1987 Sep;25(4):359–380. Available from: https://doi.org/10.1007/BF00277162.

[11] Williams BG, Lloyd-Smith JO, Gouws E, Hankins C, Getz WM, Hargrove J, et al. The Potential Impact of Male Circumcision on HIV in Sub-Saharan Africa. PLOS Medicine. 2006 Jul;3(7):e262. Publisher: Public Library of Science. Available from: https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.0030262.

[12] Granich RM, Gilks CF, Dye C, De Cock KM, Williams BG. Universal voluntary HIV testing with immediate antiretroviral therapy as a strategy for elimination of HIV transmission: a mathematical model. The Lancet. 2009 Jan;373(9657):48–57. Available from: http://www.sciencedirect.com/science/article/pii/S0140673608616979.

[13] Williams BG, Korenromp EL, Gouws E, Schmid GP, Auvert B, Dye C. HIV Infection, Antiretroviral Therapy, and CD4+ Cell Count Distributions in African Populations. The Journal of Infectious Diseases. 2006 Nov;194(10):1450–1458. Available from: https://doi.org/10.1086/508206.

[14] Bretó C, He D, Ionides EL, King AA. Time Series Analysis via Mechanistic Models. The Annals of Applied Statistics. 2009;3(1):319–348. Publisher: Institute of Mathematical Statistics. Available from: https://www.jstor.org/stable/30244243.

[15] Dougherty BP, Smith BA, Carson CA, Ogden NH. Exploring the percentage of COVID-19 cases reported in the community in Canada and associated case fatality ratios. Infectious Disease Modelling. 2020 Dec;6:123–132. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7718109/.

[16] Lindén A, Mäntyniemi S. Using the negative binomial distribution to model overdispersion in ecological count data. Ecology. 2011;92(7):1414–1421.

[17] Gharouni A, Abdelmalek FM, Earn DJD, Dushoff J, Bolker BM. Testing and Isolation Efficacy: Insights from a Simple Epidemic Model. Bulletin of Mathematical Biology. 2022 Jun;84(6):66. Available from: https://link.springer.com/10.1007/s11538-022-01018-2.

[18] van den Driessche P, Watmough J. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Mathematical Biosciences. 2002 Nov;180(1):29–48. Available from: http://www.sciencedirect.com/science/article/pii/S0025556402001086.

[19] Caswell H. Matrix Population Models: Construction, Analysis and Interpretation. Sunderland, MA: Sinauer; 2000.

[20] Champredon D, Dushoff J, Earn DJD. Equivalence of the Erlang SEIR epidemic model and the renewal equation. SIAM Journal on Applied Mathematics. 2018;78(6):3258–3278. Available from: https://epubs.siam.org/doi/10.1137/18M1186411.

[21] Park SW, Bolker BM, Champredon D, Earn DJD, Li M, Weitz JS, et al. Reconciling early-outbreak estimates of the basic reproductive number and its uncertainty: framework and applications to the novel coronavirus (SARS-CoV-2) outbreak. Journal of the Royal Society Interface. 2020;17(168):20200144.

[22] Bolker BM. Ecological models and data in R. Princeton University Press; 2008.

[23] Mullen K, Ardia D, Gil D, Windover D, Cline J. DEoptim: An R Package for Global Optimization by Differential Evolution. Journal of Statistical Software. 2011;40(6):1–26.

[24] Krinsky I, Robb AL. Three methods for calculating the statistical properties of elasticities: A comparison. Empirical Economics. 1991 Jun;16(2):199–209. Available from: https://doi.org/10.1007/BF01193491.