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

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

April 12, 2023 @ 13:51

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 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 the epidemic.

Contents

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-step deterministic model, or a discrete-time-step stochastic model with dynamical noise. 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 splitting 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 of individuals who go to the ICU die. Recovered individuals (`R`) are assumed to be immune. The model includes additional compartments which 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; we have since expanded the model to include relevant compartments.)

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 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 *phenomenological heterogeneity* 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 as the most susceptible individuals are infected, they lower the average level of susceptibility in the remaining population. 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* $\mathbf{M}$, whose elements 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 constant except when the model specifies a piecewise change.

This set-up allows for considerable flexibility.

- 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.

- 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).

- We can also use a discrete-time model with a *hazard correction* 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*

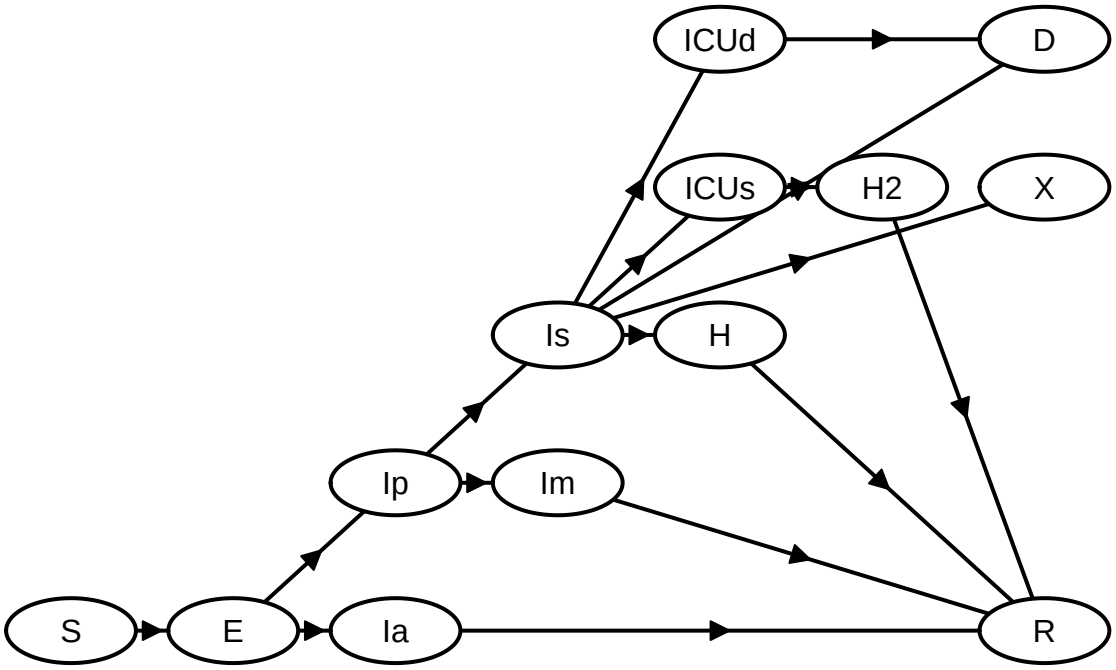

Figure 1: Flow chart for basic compartmental mechanistic transmission model. Compartments: `S` (susceptible), `E` (exposed), `Ia` (asymptomatic infection), `Ip` (presymptomatic infection), `Im` (mild/moderately symptomatic infection), `Is` (severely symptomatic infection), `H` (hospitalized [acute care]), `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?*]

flow to each other compartment ($f_{\text{tot}} = \sum_i \mathbf{M}_{ij}$), we let the total outflow be $f'_{\text{tot}} = 1 - \exp(-f_{\text{tot}}\Delta t)$; the individual flows are then adjusted by a factor of $f'_{\text{tot}}/f_{\text{tot}}$. This adjustment accounts for the effects of depletion during the course of a time step.

- 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 described above and use it to compute probabilities for draws from binomial or multinomial random deviates.

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).

We typically compute the model dynamics deterministically, as a discrete-time model with a hazard correction (the third option 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) 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. Our main use of convolution is to convert incidence, which we compute by multiplying the force of infection by the number of susceptibles, to a case-reporting time series: $\text{CR}(t) = \sum \phi(i)(\text{FOI}(t-i)S(t-i))$, where we typically set $\phi(i)$ to be a Gamma distribution with moments chosen to match estimates of case-reporting delays [**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 . . .*] When computing case reports from incidence we also include 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.

Once condensation is done, the model also allows us to add observation error, which we typically simulate from a negative binomial distribution with a variable-specific dispersion parameter.

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 calibrating this value from joint data on cases and hospitalizations),

and imposing a distributed delay between infection and reporting (i.e. via convolution). Here we instead expand the susceptible and all infected compartments factorially to include the possibilities that individuals in those epidemiological classes are 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. (16) thoroughly analyze the epidemiological consequences of this structure in a simpler framework that uses a simple SIR model rather than the COVID-specific compartmental model (Figure 1) as the foundational epidemiological model. 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., $\sum_i w_i X_i/N$), then for a daily *per capita* testing rate ρ we might expect that the corresponding *per capita* testing rate in compartment i would be $T_i = \rho w_i/W$. 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

$$T_i = \frac{\rho \tau w_i}{\tau W + \rho}.$$

(see (16) appendix A.5 for more details). The testing flows out of each untested subcompartment are divided into flows to "positive waiting" and "negative waiting" compartments according to the infection status of the relevant compartment and specificity/sensitivity parameters. If tests are assumed to be perfectly specific and sensitive, then all individuals from non-infectious compartments (`S`, `E`, `R`) enter the "negative waiting" 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)

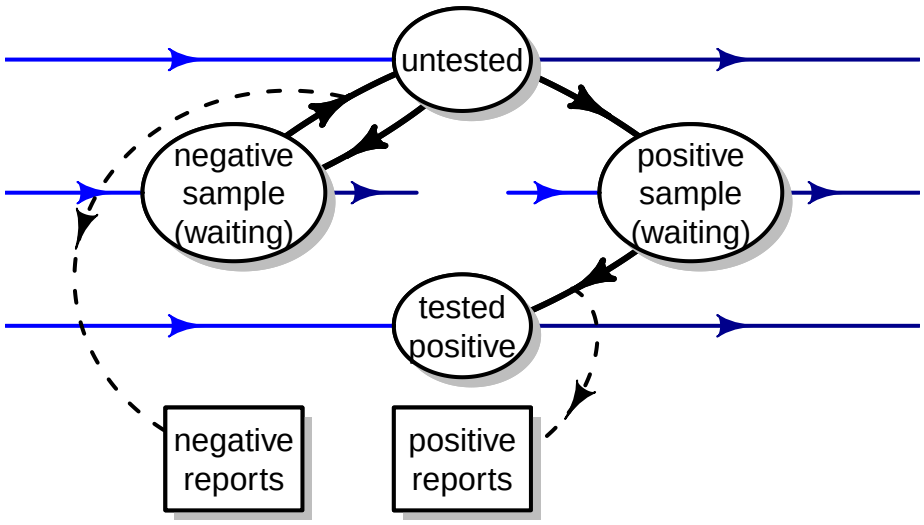

Figure 2: Testing flow. Every epidemiological compartment is subdivided into the four subcompartments shown. Black arrows represent flows between subcompartments due to testing processes (test administration, reporting of tests); blue arrows represent progression between epidemiological compartments. Dashed arrows represent the accumulation of negative and positive test reports, which can be compared against data. [**BB:** *is there a better version of this somewhere?*]
[**ML:** *Irena*]

may have their transmission modified by a specified multiplier (given as model parameters); we assume that hospital transmission is negligible. The time-varying (relative) transmission β_1 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
$$\beta_1(t) \propto \log(M(t))^{p_{\text{mob}}},$$
where $M(t)$ is the relative mobility index at time t, for some power $p_{\text{mob}} > 0$;
- according to an arbitrary spline curve, i.e. a linear combination of components of a B-spline basis.

All of the sub-models for temporal change in the transmission rate can be subsumed under a single log-linear model:

$$\log \beta(t) = \log \beta_0 + \boldsymbol{X}\boldsymbol{c}\,, \tag{1}$$

where $\boldsymbol{X}$ is a model matrix 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. In particular, piecewise breaks correspond to indicator variables for which period an observation falls in; the mobility model 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 $S(j) = (1 + \exp((t - t_{\text{brk}}(j))/s))^{-1}$, 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 $\log(M(t))$. Subsequent columns $2j$ and $2j+1$ are defined as For the first period, Thus, columns $2j-1$ and $2j$ 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 $j+1$. [**BB:** *WZMLi: does this match what you think we did/your description below?*]

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; (2) the value of $\mathcal{R}_0$ 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 (17) for differential equations and (18) 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.

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 of infetion, 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 (19). We can easily compute $\mathcal{R}_0 = \sum_t K$, mean generation interval $\bar{G} = \sum_t K(t)/\mathcal{R}_0 \cdot t$, and generation interval coefficient

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

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

$$\sum_t K(t) e^{-rt} = 1$$

for r.

[**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 (20), 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) 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 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. 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 (21) followed by a final fit with Nelder-Mead.

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 non hospitalized mortality) 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 $\beta_0 = \text{logit}^{-1}(-0.2) = 0.45$.

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 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.

In Ontario, we calibrate to deaths, and new confirmations, all of which are available publicly. In the expanded model, we can include time series of both positive and negative tests in our calibration. Reported new confirmations 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 (1)] (two mobility intercept and three slopes), and several basic model parameters (**??**) [**ML:** *E0(initial number of exposed), beta0, nonhosp_mort, zeta(pheomhet), dispersion parameters (report and death)*]. [**ML:** *how does cref work?*] [**BB:** *DJDE/JD/WZMLi: if we can get estimated-parameter table (estparmtab) and literature parameter table (litparmtab) working and properly cross-referenced, then we don't need the information in Mike's comments above. We may need to do some archaeology: were these in the original report to PHAC?*] Other model parameters are taken from the literature (**??**).

We do not include hospitalization and ICU occupancy in our calibration because our model assumes all severe cases go through ICU, whereas many severe cases actually occur 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 properly represented in our model. [**BB:** *I think I got this statement right but am not sure: WZMLi/JD/DJDE, please check for correctness/clarity.*]

Forecasting

Our calibrations yield values for the parameters of our deterministic model. Using the calibrated parameters and the estimated covariance matrix of the sampling distribution of the parameters, we draw 1000 sets of parameters from a multivariate normal distribution (22; 23) and feed them into the deterministic model to generate an ensemble of forecasts. We extend the end date of the calibration window (i.e. the last observed data point for calibration) to create a forecast window. In the forecast window, the user can either input covariates (i.e. relative mobility and/or testing rates, depending on the model) or assume that the last set of inputs remain constant through the forecasting window.

2.1 Data sources

[**DE:** *This section describes all the data available to us, whether we use it or not. This is harmless for the PHAC report, but we'll need to describe only data we when we submit for publication.*] [**ML:** *I am commenting out all the parts we are not using*]

[**ML:** *let's stick with public Ontario for now?*]

Public COVID-19 data for Ontario, Canada

Daily reported data on COVID-19 testing and outcomes (confirmed cases, death, hospital and ICU occupancy) in Ontario are publicly available from the official provincal website. See https://data.ontario.ca/dataset/f4f86e54-872d-43f8-8a86-3892fd3cb5e6/resource/ed270bb8-340b-41f9-a7c6-e8ef587e6d11/download/covidtesting.csv.

However, this dataset does not have testing counts before April 15th. One of us (ML) maintains a public web site containing Canadian COVID-19 data at the provincial level since the beginning of the COVID-19 pandemic. 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.

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

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

Parameter	Estimate
Mobility Intercept Apr 01	-0.440
Mobility Intercept Aug 07	0.095
Mobility Slope	0.655
Mobility Slope Apr 01	0.306
Mobility Slope Aug 07	0.317
Observation Error Report	\gg 1000
Observation Error Death	\gg 1000
E0	10.138
Beta0	1.038
Non-hospitalized Mortality	0.031
Zeta	52.876

Table 1: Parameter estimates for base model calibration. [**BB:** *Need to figure out how this is generated/adjust names to be sensible/augment caption. Parameter names should be left-justified, estimates decimal-point-justified. Where are the (\gg)s coming from? WZMLi?*]

3 Results

[**ML:** *Do we want to include some simulation to show the model works and can do the calibration?*] [**ML:** *figure captions and make better plots*]

3.1 Base model

Figure 4 shows the fit to data of a basic model that calibrates to COVID new cases and death for Ontario from February 24, 2020 (the first covid case reported in Ontario) to August 30, 2020. [**ML:** *do we need a justification of the end date?*] The basic model was chosen to be the base model because reported cases and death time series are most commonly applicable to public data for many juristrictions. This time frame captures the first wave of the COVID outbreak in Ontario. 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. Hospital occupancy and death on the other hand does not fit to the data well after the initial wave. This suggest the proportion hospitalized and death are lower and should be time varying as well.

[**ML:** *keep only forecast plot?*]

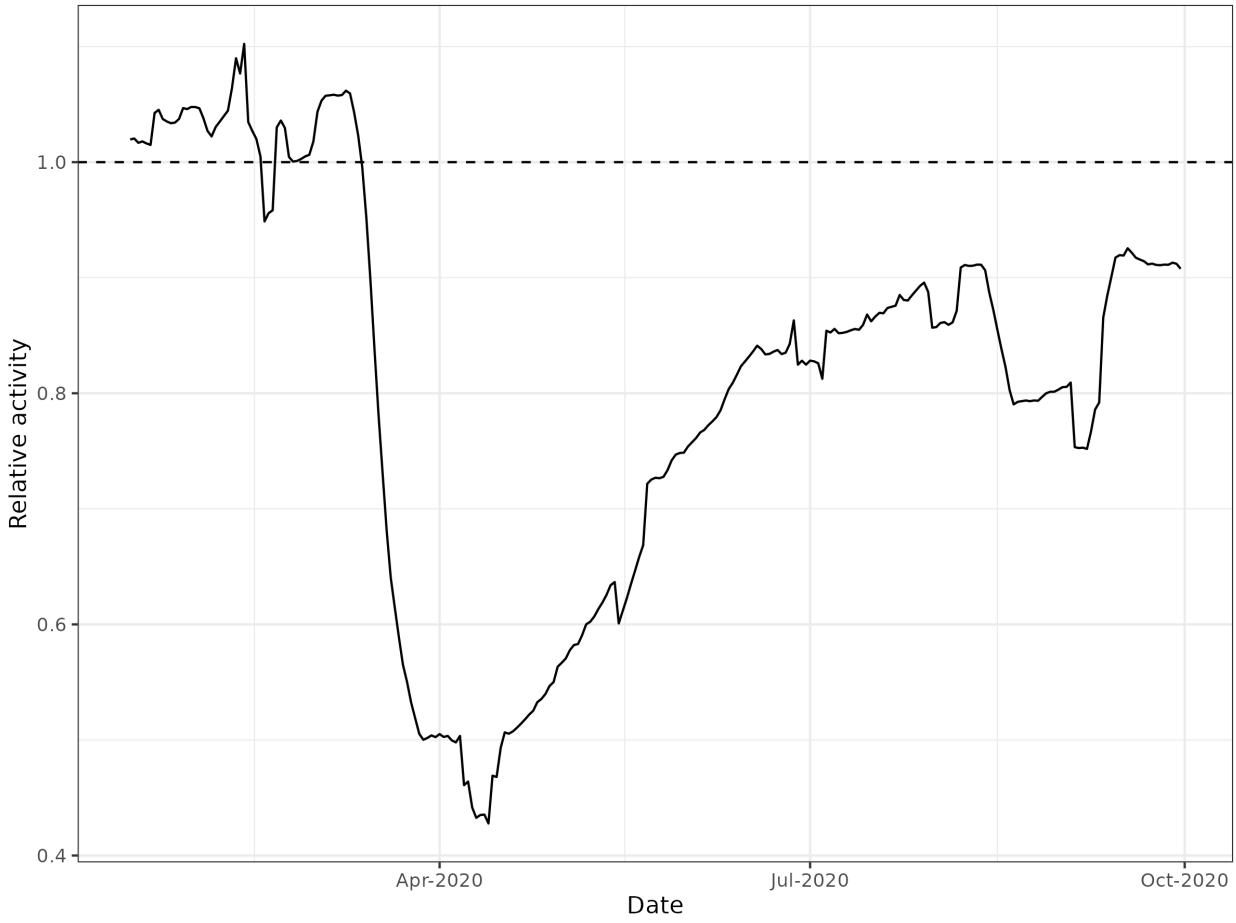

Figure 3: Ontario mobility [**BB:** *WZMLi: a previous version of this graph looked very different, with giant day-of-week effects. The current version I'm seeing is much smoother … do we know what happened/which was right?*]

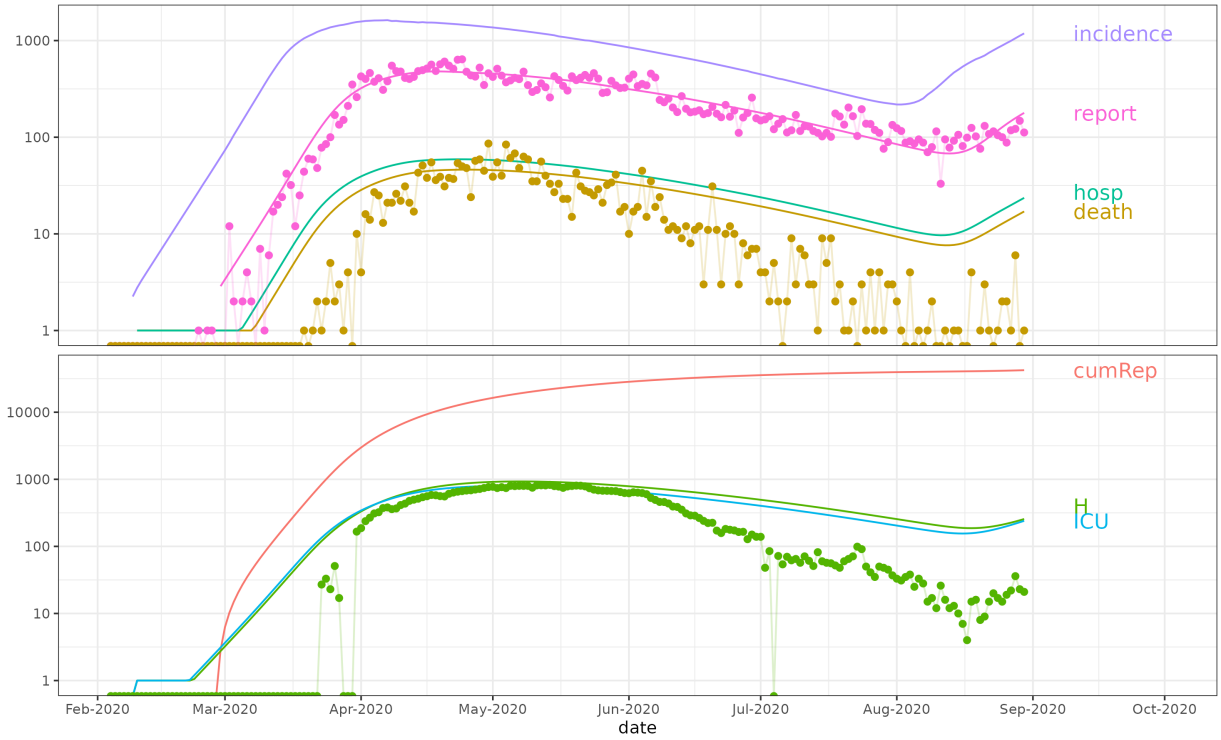

Figure 4: Ontario calibration [**BB:** *Take out incidence, cumRep? Mention that death and hospitalization rates declined (why?) and that we are only calibrating to reports/not allowing time-variation in hosp and mortality rates?*]

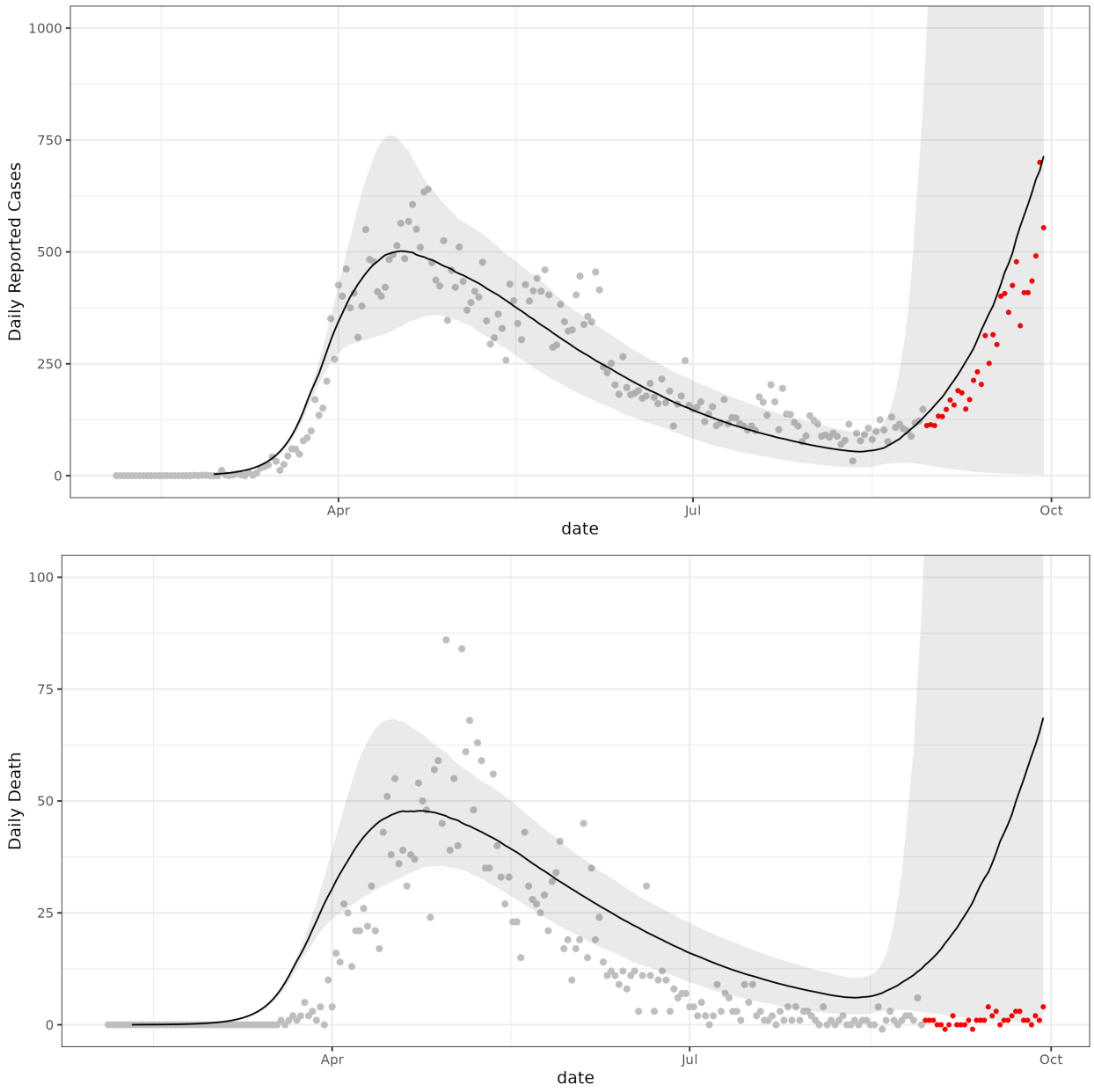

Figure 5: Ontario base forecast. [**BB:** *need to explain what's going on here; at the very least the caption should say (even if it's kind of obvious) that gray points represent calibration points, red are forecast points. Why is the death forecast so far off? Should we just leave it out (probably)? (WZMLi ...)*]

3.2 Incorporating testing flows

The base model is relative simple but does not account for testing practices. Testing strategies changed continuously over the course of the pandemic due to availability of testing facilities and shifting regulations on eligibility for testing. Figure 6 shows the daily testing in Ontario in the calibration time frame. The verticle line is the cut-off between the calibration window (before September 1st 2020) and 30 days ahead for the forecast window. The testing intensity over the forecast window is assumed to be constant as the last observed testing intensity on August 30th 2020.

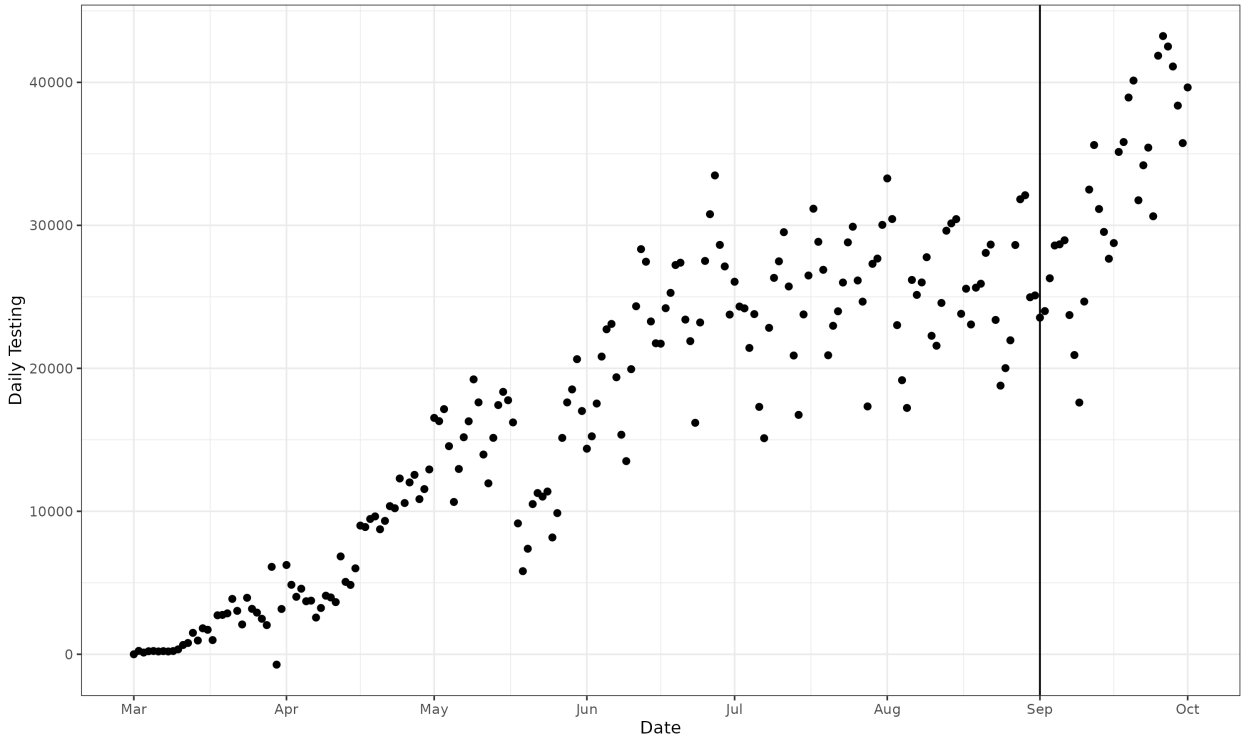

Figure 6: Ontario Testing Intensity

Figure 7 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.

[**ML:** *We (I) probably should make another forecast using the testing intensity of the forecast window. Thoughts?*]

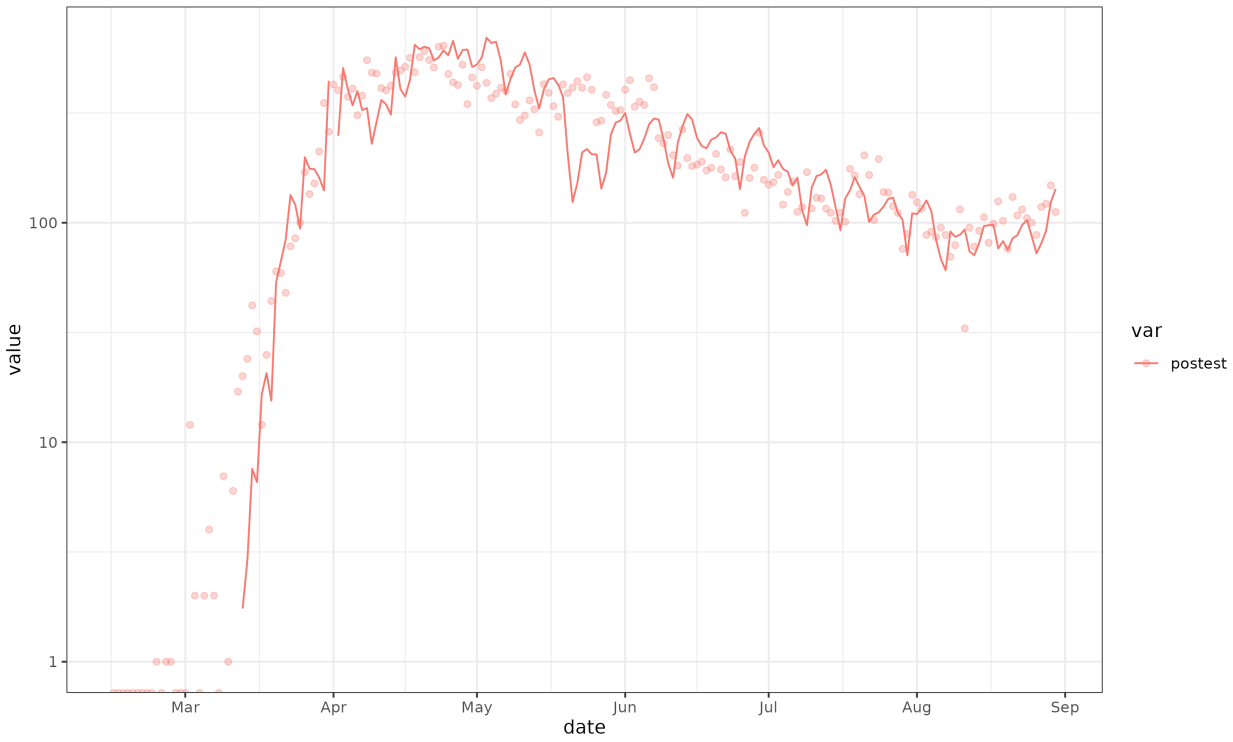

Figure 7: Calibration of positive tests [**BB:** *is there a nice way to combine Figures 6 and 7? Maybe a two-y-axis plot?*]

Parameter	Estimate
Mobility Intercept Apr 01	-0.652
Mobility Intercept Aug 07	-0.949
Mobility Slope	0.032
Mobility Slope Apr 01	0.611
Mobility Slope Aug 07	-2.493
Observation Error	\gg 1000
E0	37.553
Beta0	0.677
Zeta	1.41

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 ...)*]

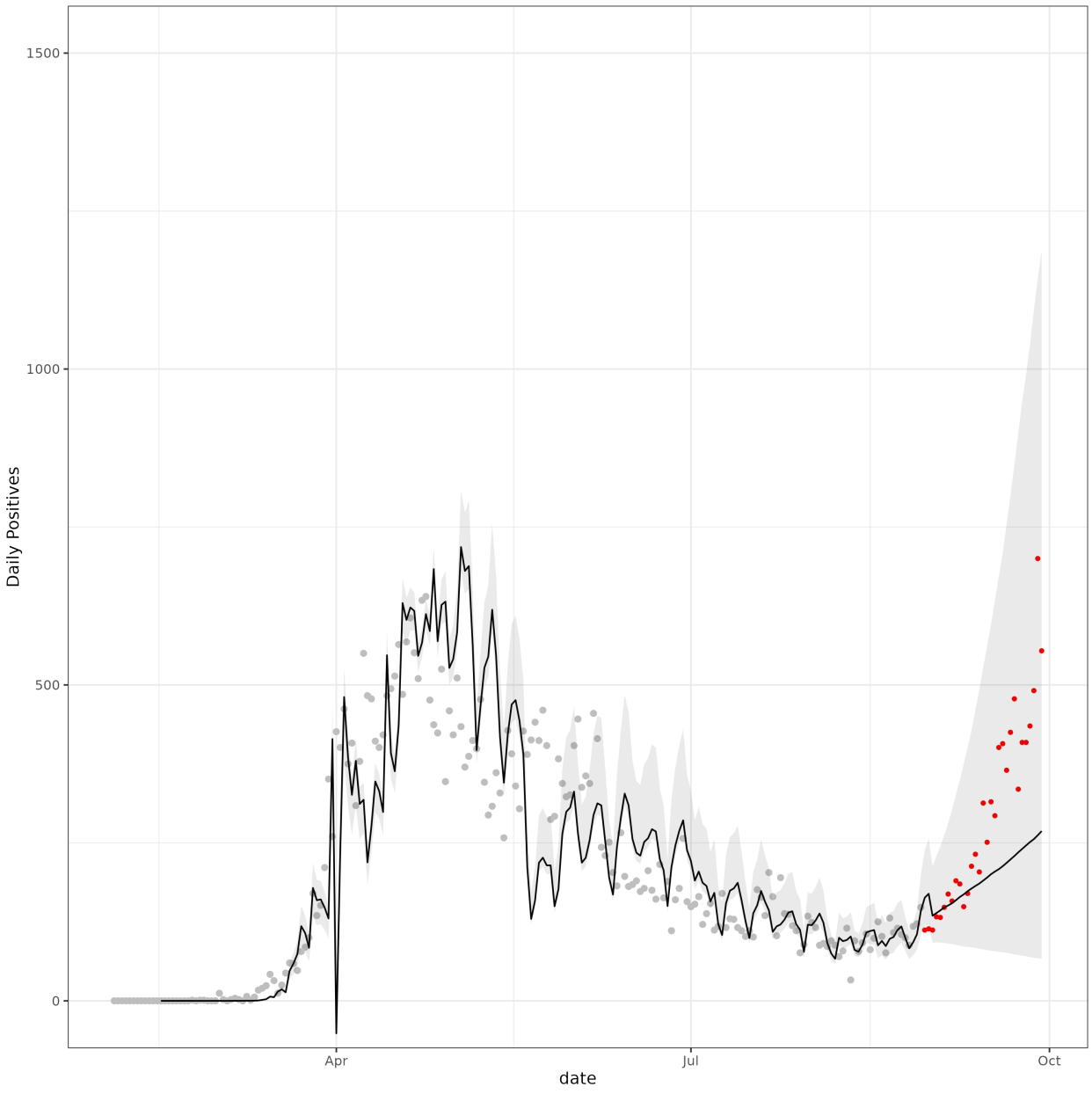

Figure 8: Ontario calibration with testing flows

4 Discussion

We have fitted two models varying in complexity to SARS-CoV-2 daily reporting and death time series data for Ontario Canada from March 1st to Aug 30th 2020. The simple base model is the typical simplistic type of compartmental epidemic model to infectious disease epidemics. The testify model is a practical addition to the base model that allows the model to cooperate with different testing practices and policy changes.

4.1 Limitations

Our model assumes homogeneous mixing of the population. No age-related or spatial contact structure is considered. 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.

4.2 Conclusions

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.

5 To do

- find and clean up `estparmtab`/`litparmtab` (DJDE, WZMLi?)
- clean up Tables 1 and 2 (WZMLi?)
- should we include fits to simulations or not?
- Clean up figures, improve captions
- Discussion and conclusions
- output/include figs as PDF rather than PNG (JD, WZMLi?)

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;382:1199-207.

[3] Fauci AS, Lane HC, Redfield RR. Covid-19 – Navigating the Uncharted. New England Journal of Medicine. 2020;382:1268-9.

[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;584:257-61.

[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):E497-505.

[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):294-8. 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):359-80. 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):1450-8. 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):319-48. 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;6:123-32. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7718109/.

[16] 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.

[17] 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.

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

[19] 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):3258-78. Available from: https://epubs.siam.org/doi/10.1137/18M1186411.

[20] Park SW, Bolker BM, Champredon D, Earn DJ, 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.

[21] 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.

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

[23] 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.