



covideestim model updates

9-29-2020

We have introduced an updated version of the model that generates estimates for covideestim. The goals of this update were to make the model faster and more stable, and reflect changes in COVID-19 science and epidemiology since our initial release. We introduced these changes at the same time as we began reporting county-level outcomes, and a number of these changes were needed so that we could successfully estimate outcomes for all these new jurisdictions.

The list below describes the major changes included in this update. We expect that there will be periodic model updates in the future, and we will document these changes when introduced.

model Change in starting point of model

The initial version of our model specified a flexible function for the number of new infections each day, operationalized as a geometric random walk for the daily change in the number of new infections. In the revised version of the model, this has been replaced by a penalized cubic b-spline for the log of R_t , the effective reproductive number. This replacement reduces the number of parameters required in the model, which is important for improving runtimes as the time-series data become longer. The spline knots are evenly spaced every 4 days, retaining substantial flexibility.

model New approach for estimating R_t

The initial version of our model calculated R_t from the estimated time-series of symptomatic cases, using functions provided in the [EpiEstim](#) package (Thompson et al, Epidemics 2019). In the revised version of the model, we simulate R_t directly, as described above, and so no longer need to use EpiEstim to back-calculate R_t from other results.

model More flexible approach to case ascertainment for symptomatic, non-severe cases

The initial version of the model assumed a simple functional relationship for the fraction of symptomatic, non-severe cases that were detected, which was based on time-series data on test-positivity. While this approach worked for most jurisdictions, we found instances where it did not adequately capture the relationship between cases and deaths, particularly as the epidemic progressed. In the new version of the model, we have replaced this with a cubic b-spline for the logit of the probability of detection for symptomatic non-severe cases, with knots evenly spaced every 21 days.

This probability is bounded between zero and the probability of diagnosis for severe cases, under the assumption that the probability of diagnosis is always higher for severe vs. non-severe cases.

model Allowance for diagnosis of asymptomatic cases

The initial version of the model assumed that diagnosis was only possible for symptomatic cases. In the revised version of the model we have relaxed this assumption to allow for diagnosis of asymptomatic cases. This probability is assumed to be a fraction of the probability of diagnosis of symptomatic, non-severe cases. This new parameter is operationalized with a Beta(2,18) prior, with mean value of 0.1. While the probability of diagnosis for asymptomatic cases is likely low, these individuals will contribute to case counts (such as through testing of contacts of diagnosed cases, or occupational testing), and this revision will make the model more robust in situations where a high fraction of cases are detected.

model Allowance for imported infections

The initial version of the model assumed that all infections were due to transmission within the modeled jurisdiction. In the revised version of the model we relax this assumption to allow for imported cases, which are given a half-Normal prior distribution equivalent to 0.5 imported infections per day. This addition has no effect in established epidemics, but produced more credible R_t estimates for some early epidemics.

model Assumption about epidemiology prior to first data point

The initial version of the model made no assumption about the trajectory of reported cases and deaths in the period preceding the first reported COVID-19 case. While this worked for most epidemics, it produced implausible results in a small number of counties, where observed data appeared to show a declining epidemic at the start of the time series. In the revised model we have added a penalty function to limit the expected number reported cases and deaths arising in the model burn-in period. This enforces the assumption that there were no COVID-19 diagnoses prior to those included in the reported data.

model Revised prior distributions for natural history parameters

In the initial version of the model, the prior distributions for natural history parameters were based on our review of the literature, favoring systematic reviews, local data, and stronger study designs where possible. In the revised version of the model we have revised prior distributions for some parameters to follow the synthesized evidence reported in the CDC's [COVID-19 Pandemic Planning Scenarios](#).

data New data source

For state-level estimation, we have been using data from the COVID-Tracking Project. County-level data are not currently available from this source, and so for county-level estimates we use data from the Johns Hopkins University COVID-19 [Data Repository](#).

model More efficient code

In the revised model, we made multiple small edits to reduce computation time. These included (i) replacing loops with vector operations, (ii) truncating all delay distributions at 60 days, and (iii) setting the shape and scale parameters for reporting delay distributions (time from diagnosis or death to when this event is reported) at fixed values.

website New workflow engine

We now use [Nextflow](#) to run all ~3200 counties and 51 states on AWS Batch.

website New website features

County-level estimates can now be inspected on a map, and in greater detail by clicking an individual county to add it to the “tray.” Please keep in mind that maps do not visualize uncertainty. Forthcoming website changes will include additional county-level detail.