

One state, many outbreaks: a transmission modeling perspective on current COVID-19 trends in King, Pierce, and Yakima counties

Niket Thakkar and Mike Famulare

Reviewed by: Roy Burstein and Jen Schripsema

Institute for Disease Modeling, Seattle Washington, covid@idmod.org

Results as of 10 a.m. on August 17, 2020

What do we already know?

In April, we published a [detailed technical report](#) describing our COVID-19 transmission model. Since then, we've used that model to estimate COVID-19 prevalence and the effective reproductive number, R_e , in different parts of the state ([here are some examples](#)). Some of our model-based estimates of R_e are now available on [the Washington State COVID-19 risk assessment dashboard](#) and [the King County key indicators dashboard](#).

What does this report add?

From a methods perspective, this report is our [original technical report's](#) sequel. We use the same model structure specified by the same core algorithm, but we've worked towards incorporating more information by better preparing the model's inputs. Specifically, this report introduces methods for incorporating daily hospital admissions and trends in the age distribution of cases, leading to better informed prevalence and R_e estimates.

From an epidemiological perspective, we update our in-depth look at COVID-19 transmission in King County using data up to August 10 from [the Washington Disease Reporting System](#) (WDRS) compiled on August 16. We estimate that on August 5, R_e was between 0.33 and 1.74 with best estimate 1.04. This relatively high level of uncertainty is reflective of the range of transmission rates consistent with both observed daily COVID-19 cases and hospital admissions. Then, by sampling over a time-varying range of infection-fatality-ratios informed by the shifting age distribution of cases, we compare the model to observed COVID-19 mortality and infer the full time course of active COVID-19 infections in King County. We estimate that on August 10 between 0.12% and 0.33%, with best estimate 0.20%, of King County's population was actively infected with COVID-19. Simultaneously, we infer that from January 15 to August 10, between 2.3% and 5.2%, with best estimate 3.5% of King County's population was at some point infected.

Finally, we apply the model independently to nearby Pierce and Yakima counties, and we find that COVID-19 prevalence has grown above previous peak levels in Pierce County while it's fallen in Yakima County throughout July. In conjunction with the uncertainty around $R_e \approx 1$ in King County, which corresponds to a relatively flat trend, this comparison illustrates the heterogeneity in COVID-19 epidemiology across Washington.

What are the implications for public health practice?

These results suggest that in King County the progress made in suppressing COVID-19 through April and May has been retained to some extent through June and July despite increased economic activity. That said, prevalence remains at levels that strain resources for effective outbreak investigation, [increase the challenge of reopening schools](#), and [cause new deaths and hospitalizations every day](#). With more than 95% of the population likely still susceptible to COVID-19, the situation is as precarious as ever. Moreover, King County's situation is not reflective of the state as a whole. While Yakima County continues its downward trend, the epidemic in Pierce County is likely larger than in March.

1 Executive summary

King County’s COVID-19 epidemic is so far a story in three parts. First, in response to evidence of a rapidly growing problem identified in late February, workplace closures, physical distancing recommendations, and finally a stay at home order were implemented throughout March. Then, likely as a result of those efforts, King County’s COVID-19 hospitalizations and later mortality declined through April and May. Finally and most recently, as King County both partially relaxed their COVID-19 restrictions and significantly [expanded their testing program](#) in early June, we have seen a large increase in daily COVID-19 positives over the summer.

In this report, we use an updated transmission model, closely related to the model described in our [previous report](#) but additionally informed by hospital admission data and by the age distribution of COVID-19 cases, to characterize the epidemiological situation in King County. **Overall, we find that King County’s COVID-19 prevalence increased in mid-June, but that recent prevalence levels are still considerably lower than those from late-March.** This suggests that mitigation efforts like masking and physical distancing have had some success in suppressing transmission in King County as [business restrictions have been lifted or relaxed](#). That said, **we further estimate that roughly 95% of King County’s population remains fully susceptible to COVID-19, and as evidenced by the prevalence increases in mid-June, King County is still capable of sustaining an explosive epidemic.** As a result, continued and even increased adherence to physical distancing recommendations and careful relaxation of restrictions will be critical going forward.

Zooming out, we find that **King County’s overall epidemiology is unlikely to be representative of Washington as a whole.** [Heterogeneity has been a prominent feature](#) of Washington’s COVID-19 epidemic, and the drivers of recent trends in cases are no exception. To illustrate this point, we apply our transmission model to neighboring Pierce and Yakima counties, and we uncover three distinct epidemiological situations. **In Pierce County, we estimate that recent local prevalence is as high or higher than in late-March,** suggesting that increased daily COVID-19 positives in June and July were driven by accelerating transmission that is only recently declining again. This heterogeneity teaches us two important lessons: (1) that public health policies must strive to address community-specific nuance and needs, and (2) that even at the county level, **our models likely mask important sub-county differences that have significant implications for public health practice.** As we continue to refine our modeling approaches, we’ll do so with these ideas in mind.

2 Key inputs, assumptions, and limitations

Our modeling approach relies heavily on particular data sources and assumptions, which in turn lead to a number of important limitations. Specifically:

- We use lab testing data provided by Washington State Department of Health (WADoH) through the [Washington Disease Reporting System](#) (WDRS), compiled for this report on August 16. Tests are aggregated by specimen collection date, and to hedge against incomplete reporting in the last few days, we only include data through August 10 in King County and through August 7 in Yakima and Pierce counties.
- We also use daily COVID-19 deaths reported to the WDRS over the same time period, aggregated by date of death. These are incorporated into the model through an all-age average infection-fatality-ratio (IFR) that varies over time based on observed trends in the age distribution of cases, taking [published age-specific IFR values](#) as input. Simultaneously, [based on previous studies of early data from China](#), we assume as before that the time between infection and death is log-normally distributed (log mean 2.8, log standard deviation 0.42), with an average value of roughly 19 days. See Appendix C for more details.
- New to this report, we use daily COVID-19 hospital admissions reported to the WDRS over the same time period and aggregated by admission date. These data are compared to model estimates using an all-age average, time-varying infection-hospitalization ratio (IHR) constructed similarly to the IFR (see Appendix C). Meanwhile, based on [published studies](#), we assume that the time between infection and hospitalization is normally distributed with a mean of 6.1 days and standard deviation of 2.65 days (samples are truncated at zero).
- In the past, we based R_e estimates on an epidemiological curve constructed by multiplying the observed daily fraction of COVID-19 tests that are positive by the 3-day spline smoothed testing

volume. Estimates based on this approach were robust to day-to-day variation in testing volume (e.g. weekend effects) but had difficult-to-correct biases introduced during large-scale changes in testing volume and strategy (e.g. the introduction of free, drive-thru testing sites). In this report, we base R_e on an epidemiological curve constructed the same way but regularized by the trend in observed hospital admissions (see Appendix B for details). While this approach is more robust to changes in testing, it increases reliance on hospital admission data. This implicitly assumes that (1) hospital admission criteria have not changed substantially over time and (2) that hospital admissions are a reflection of severe COVID-19 infections, and are therefore independent of testing strategy to a good approximation. While the resulting model agrees with all the data we have, it is difficult to verify these assumptions more directly, and it is a key limitation of our approach.

- We continue to assume that a number of COVID-19 importations into Washington went undetected by the health system. With the data we have, we cannot estimate how many importations occurred at particular times, so instead we allow the model to have as many imported infections as necessary on prespecified dates to achieve consistency with observed mortality. While we find that this approach yields transmission dynamics consistent with all the data we have and that model outputs are insensitive to reasonable changes in importation timing, we emphasize that the specifics of modeled importations are not reflective of reality. This limitation of our approach introduces difficult-to-quantify uncertainty that likely cannot be resolved without additional data.
- Finally, we note that our models are applied to data aggregated to the county level. This choice is a balance between data sparsity considerations and the needs of public health policy makers; however, analysis at the county level necessarily masks diverse, epidemiologically-relevant circumstance in sub-county communities. [Sub-county heterogeneity certainly exists](#) and has important implications for public health. In its current form, our prevalence model is unable to address these needs.

3 Modeling approach

We fit a COVID-specific transmission model to daily testing, hospitalization, and mortality data. The key modeling assumption is that individuals can be grouped into one of four disease states: susceptible, exposed (latent) but non-infectious, infectious, and recovered. In addition, we assume:

- COVID-19 has a latent period that lasts about 4 days during which infected people are not yet capable of transmission. The choice of a 4 day latent period implicitly assumes that people become infectious on average roughly 1 day before the typical 5-day asymptomatic incubation period ends. After the latent period, those exposed to COVID-19 are infectious for about 8 days. This model describes average infectious dynamics and does not explicitly account for individual level variation in shedding.
- In the model, COVID-19 is introduced by an unknown number of infectious individuals on assumed-known, location-specific dates (January 15 and February 1 in King County, January 20 in Pierce and Yakima counties). On all other days, we assume that community transmission is the dominant infection route.

We use a multi-step approach to fit the model to WDRS data. Briefly, we construct an approximate epidemiological curve which simultaneously captures short-term (daily) variability in testing data and long-term (monthly) trends in hospital admissions (see the appendix), and we calculate daily estimates of R_e by applying the method used in our [previous report](#) to this curve. This yields estimates of R_e from March 1 to August 5 in King County. We fill the remaining time from January 15 onward by backward and forward filling. Backward filling is done with the value on March 1, to approximate unabated COVID-19 transmission in the absence of physical distancing and other mitigation efforts, and forward-filling is done with the estimate on August 5.

Given R_e over time, we calculate the expected number of COVID-19 deaths in the model by assuming that the average time to death is 19 days and that the IFR varies in time based on the age-distribution of the population, the weekly age-distribution of positive tests, and [published values](#) of the IFR as a function age (see Appendix C). Then, we minimize the difference between observed mortality and the model average as a function of the number of importations on prespecified dates. Comparisons to daily hospitalizations are treated similarly, with an average time from exposure to hospitalization of roughly 1 week and a time varying IHR based on [published, age-dependent values](#) scaled by an overall, fitted

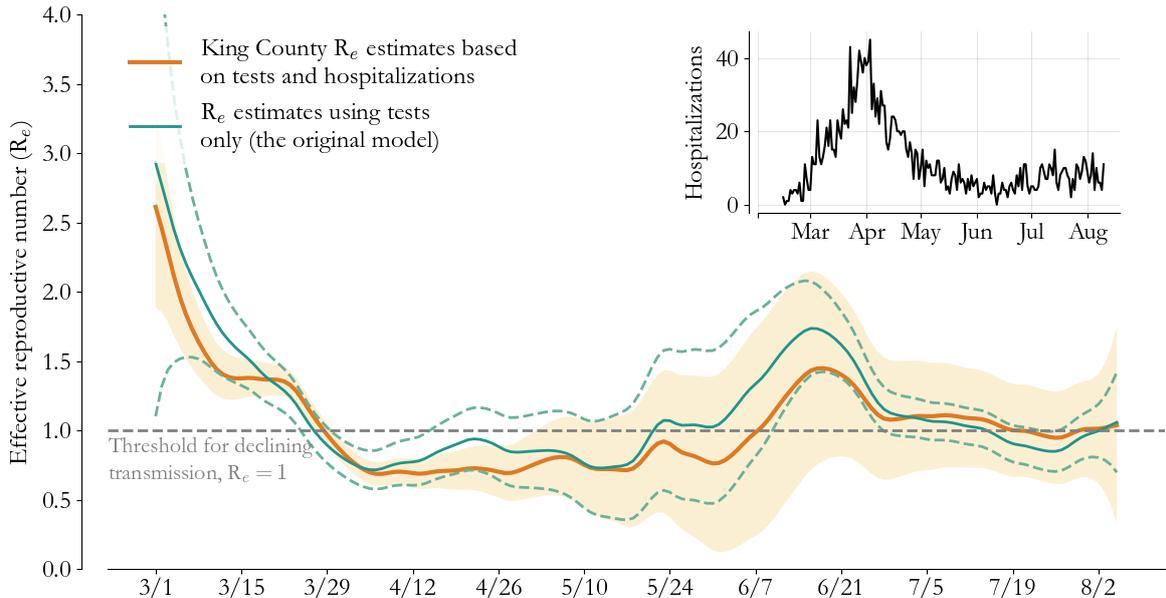


Figure 1: Updated effective reproductive number estimates for King County. The orange line (95% confidence interval shaded) is daily R_e estimates based on testing and hospitalization data. For comparison, estimates from our original method based on testing data only are shown in green (95% CI as dotted lines). While the two methods are generally consistent within uncertainty, our new method has increased confidence in declines from March through April and increased uncertainty more recently, in part because daily hospital admission numbers (inset) remain low relative to late-March.

scale-factor to account for local variation in the probability of hospitalization. Finally, daily estimates of the reporting rate, p_t , are constructed by comparing daily positive COVID-19 tests in the WDRS to the number of infections in the model. These daily estimates are then used to approximate the reporting rate in assumed reporting periods, controlling for weekends, in a standard least squares regression.

This procedure fully specifies the model’s parameters. To estimate the prevalence and cumulative incidence over time consistent with observed testing, hospitalization, and mortality data, we sample the fitted model 10,000 times and summarize the trajectories daily.

4 Modeling COVID-19 transmission dynamics in King County

Updated estimates of the effective reproductive number in King County using our refined model are plotted in Figure 1 in orange (95% confidence interval shaded). Most recently, on August 5, we find that R_e was between 0.33 and 1.74 with best estimate 1.04, indicating that both a growing ($R_e > 1$) and shrinking ($R_e < 1$) COVID-19 epidemic are consistent with recent data through August 10.

To highlight the effect of including hospitalization data, estimates based only on WDRS tests (i.e., our original model) are overlaid in green (95% CI between the dashed lines). Overall, estimates from both methods are consistent within uncertainty, but there are some important structural differences. Notably, our new estimates have lower uncertainty from March to early May, better resolving the effect of big policy changes like the [March 23 Stay Home, Stay Healthy order](#), for example. On the other hand, the hospitalization data introduces additional uncertainty from late May onward. Looking directly at the data (inset), we see a clearly defined peak near April 1 that is also prominent in the testing data (see Figure 2), and model confidence in declining transmission from late March to May is high as a result. Later on, from June onward, relatively low numbers of daily hospitalizations make overall trends difficult to define, leading to increased uncertainty in our model-based estimates of transmission rates.

Just as in our previous technical report, model-based R_e estimates are a starting point for fully specifying a COVID-19 transmission model that can be compared to observed case, hospitalization, and mortality data from King County. Following that same procedure but leveraging time-varying IFR and IHR estimates described in Appendix C, we can construct a transmission model consistent with essentially all of the county-level WDRS data available to us (daily positive tests, hospitalizations, and deaths).

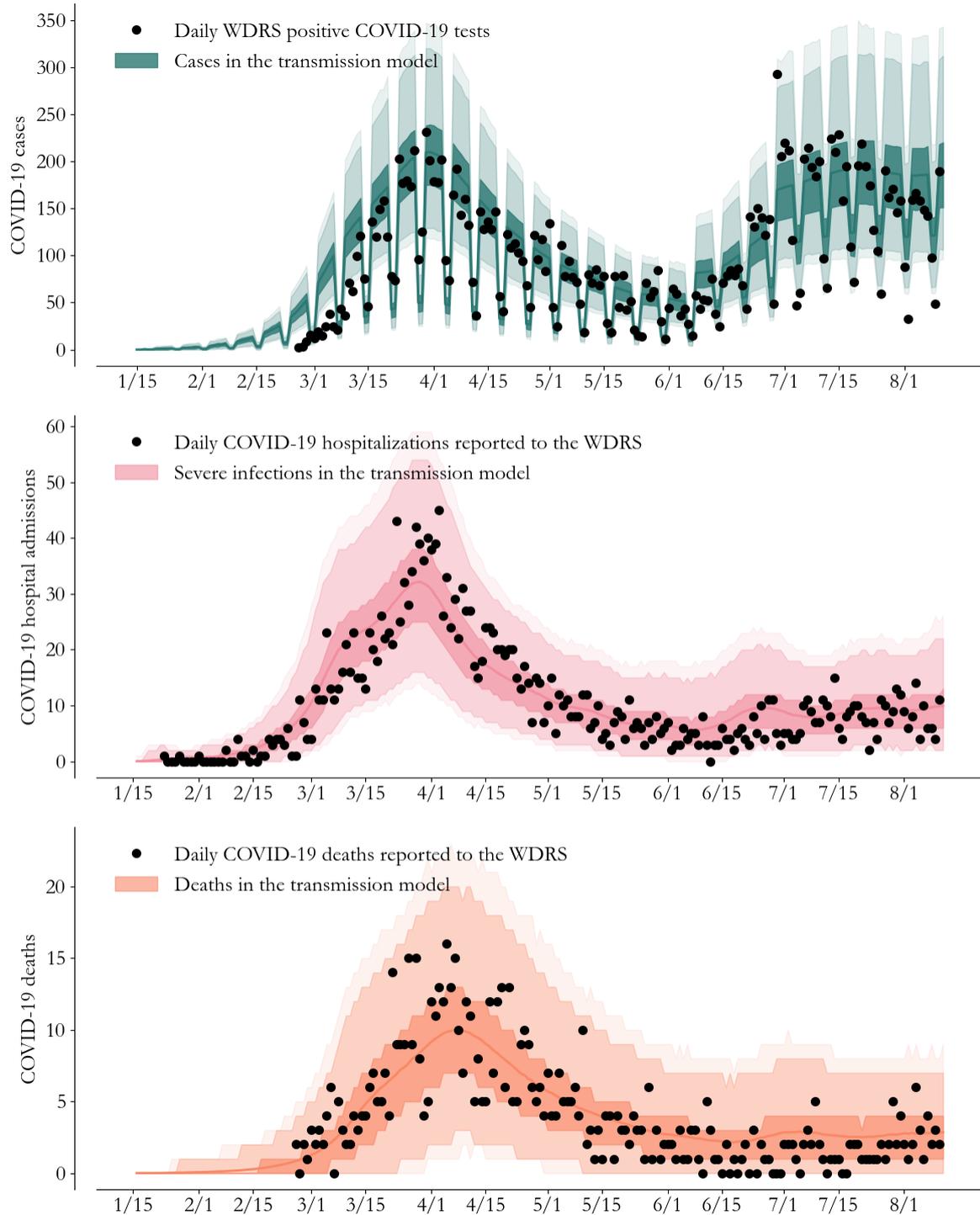


Figure 2: Fitting the transmission model to King County’s testing, hospitalization, and mortality data from the WDRS. **Top panel:** Modeled cases (green, 50% CI dark, 95% CI light, 99% CI lightest) captures the trend in observed, daily COVID-19 cases (black dots) based on R_e estimates and a free number of importations on January 15 and February 1. **Middle panel:** Simultaneously, the model (pink) captures the trend in observed daily hospital admissions (black dots) by assuming hospitalization is reflective of severe disease independent of testing volume. **Bottom panel:** Finally, with the time-varying IFR described in Appendix B, the model (orange) captures the observed trend in daily COVID-19 deaths (black dots).

Model fits to these data are shown in Figure 2. In the top panel, the model (green, 50%, 95%, and 99% CI from dark to light) captures both the March and June waves of cases while accounting for weekend drops in testing volume. In the middle panel, simultaneously, a subset of model infections (pink), based on [published estimates](#) of the probability of severe disease, captures the initial rise, fall, and slight increase in observed hospitalizations. Finally, in the bottom panel, the trend in daily COVID-19 mortality is captured by sampling and lagging model infections (orange). In this case, uncertainty is relatively high since most COVID-19 infections do not result in death. Overall, that the model captures all three trends at once based on published estimates of the age-dependent probability of severe outcomes gives us confidence that the model accurately represents transmission dynamics in King County.

5 COVID-19 prevalence, susceptibility, and case-detection rates in King County

The fitted model gives us daily estimates of the total infected population in King County over the course of the epidemic, shown in Figure 3. In the top panel, infections in the model grow to a March 28 peak where we estimate that 0.49% (0.31% to 0.73% uncertainty interval) of King County’s population was

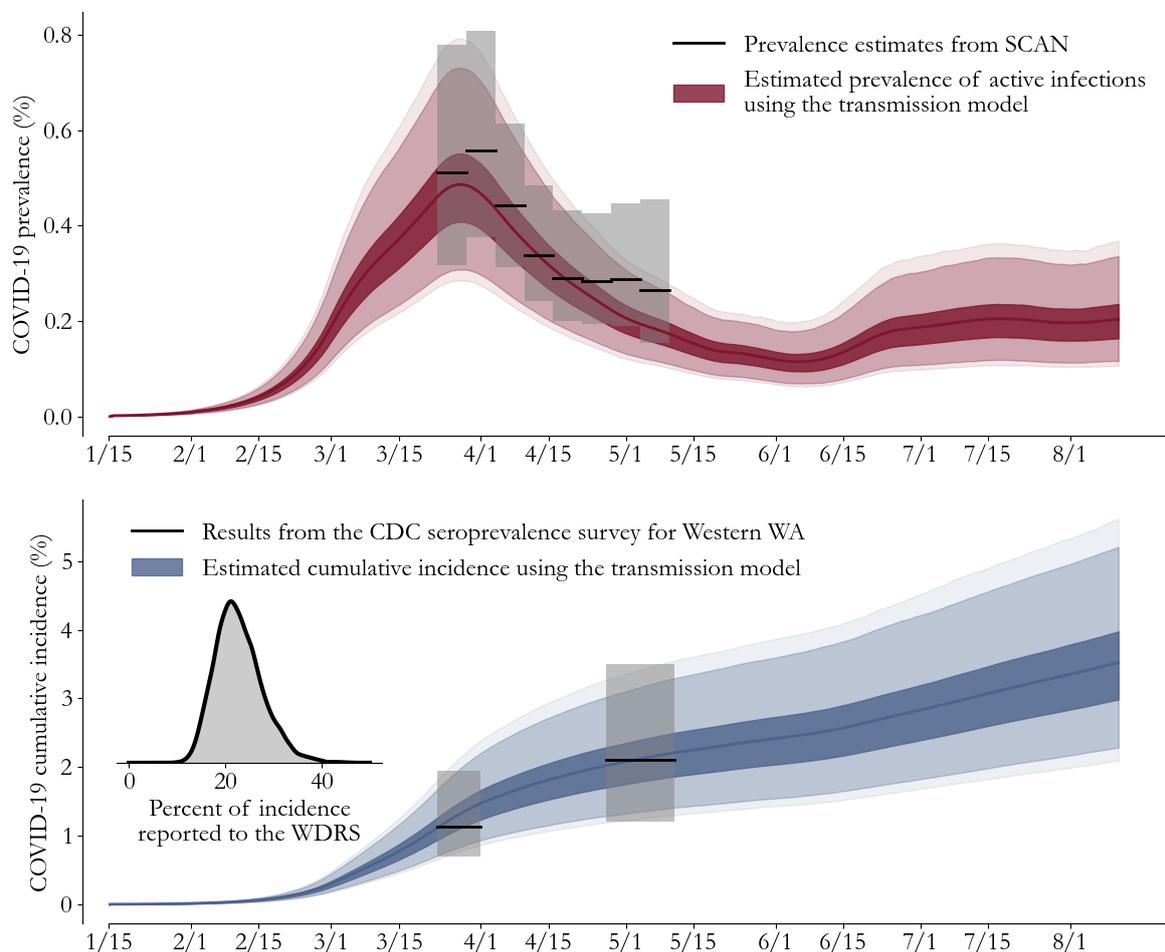


Figure 3: Using the fitted model to estimate prevalence and cumulative incidence. **Top panel:** Model-based prevalence (purple, 50% CI dark, 95% CI light, 99% CI lightest) compared to survey-based estimates from SCAN (mean in black, 95% CI in grey), showing good agreement where survey data was available. **Bottom panel:** Cumulative incidence (blue, intervals as in the top panel) indicates that the vast majority of King County’s population is still susceptible to COVID-19. The CDC’s published serological survey estimates for western Washington (black lines, 95% CI in grey) agree with our estimates. In the inset, comparing total test positives to cumulative incidence implies that over the course of the epidemic so far, roughly 23% of infections were eventually tested positive and reported to the WDRS.

actively infected with COVID-19. Through April and May, likely as a result of physical distancing, masking, and related efforts, our best estimate of King County’s COVID-19 prevalence declined by nearly a factor of 5. Since then, particularly in mid-June, we find that prevalence has grown, and on August 10 we estimate that 0.20% (0.12% to 0.33% uncertainty interval) of the population was actively infected with COVID-19. Notably, despite the mid-June rise, our recent estimate is still likely lower than King County’s historical peak, suggesting that some of the success of mitigation efforts in April and May has been retained in King County.

Aggregating active infections over time allows us to estimate the total fraction of King County’s population no longer completely susceptible to COVID-19, so-called cumulative incidence. This is shown in Figure 3’s lower panel, and on August 10 we estimate that between 2.3% and 5.2% with best estimate 3.5% of King County’s population has at some point been infected. As a result, with more than 95% of the population likely still susceptible, King County remains in a precarious position far from herd immunity. As demonstrated in mid-June, King County can still sustain a growing COVID-19 epidemic, and continued vigilance with regards to masking and physical distancing will be required for the foreseeable future.

Published survey results offer independent checks to our prevalence and cumulative incidence estimates. For prevalence, the [greater Seattle Coronavirus Assessment Network \(SCAN\)](#) conducted community testing from mid-March into May, and their [published estimates](#) (black lines, 95% CI in grey) are in good agreement with our concurrent estimates. Meanwhile, [serological surveys conducted by the CDC](#) for western Washington offer a check on our estimates of cumulative incidence. Both published estimates from late March and early May are in good agreement with our transmission model-based results. These survey estimates are generated using very different principles, and they are not used in any way for model fitting, giving us some added confidence in our approach. As [more survey estimates](#) become available, we will continue to validate our transmission model-based predictions and publish the comparisons.

Finally, cumulative incidence in the model gives us an opportunity to estimate what fraction of COVID-19 infections are diagnosed and reported to the WDRS as cases. An overall estimate is shown in Figure 3’s inset, where we find in King County that between 15% and 34% (best estimate 23%) of infections have been reported over the epidemic’s full course to date. This same estimate is generated on a weekly basis in Figure 4, exposing significant time dependence. From April through May, we estimate that roughly 12% of infections were detected by the health-care system. However, after the [testing program significantly expanded on June 5](#), we find that case-detection rates increased considerably, with a best estimate of roughly 28% throughout July.

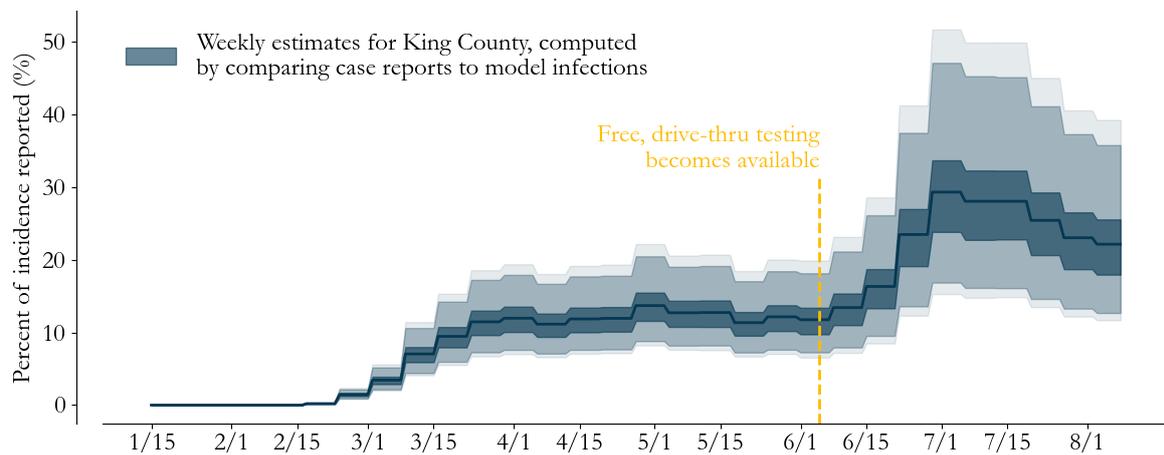


Figure 4: Assessing reporting over time in King County. Weekly totals of reported cases are compared to the weekly average of active infections in the model to construct weekly estimates of the fraction of infections diagnosed and reported to the WDRS. From mid-March to June, roughly 12% of infections were diagnosed, but our estimates rise significantly through June, up to roughly 29% by mid July. This rise is coincident with the June 5 roll-out of free, drive-thru testing sites in King County, suggesting that changes in testing strategy significantly improved the infection-detection system.

6 County-to-county variation is a prominent feature of Washington’s epidemic

Our findings in King County suggest that recent mitigation efforts in combination with the relatively wide availability of testing are suppressing prevalence below March levels as of August 10. However, it is unlikely that this result is representative of the state as a whole.

To illustrate this point, we apply the model to two of King County’s neighbors: Pierce and Yakima. Our prevalence estimates for all three counties are compared in Figure 5, and we find three distinct epidemiological situations. While King County has managed to retain some degree of transmission suppression in June and July, we find evidence of a large second-wave in Pierce County starting in June, with prevalence in mid-July as high or higher than peak levels in late March. Meanwhile, at the same time, prevalence in Yakima County has shown sustained declines from June 1 onward, but it remains at levels comparable to or higher than those in Pierce and King most recently. This suggests that state-wide mitigation efforts have had varying levels of impact from county-to-county.

While we do not fully understand the causes of this diversity in recent trends, its existence speaks overall to the need for care when interpreting COVID-19 case, hospitalization, and mortality data, particularly at high levels of aggregation. Even at the county level, we are masking heterogeneity with implications for public health practice and transmission mitigation strategies.

We can clarify this final point with a simplistic example. Our best estimate for the fraction of infections reported in July in King County is $\sim 30\%$. This estimate averages over the whole county, and it can easily mask significant heterogeneity as a result. For example, consistent with a 30% average, half of the population could have a detection rate of 50% while the other half has a detection rate of only 10%, a 5-fold disparity that our current model would miss. In King County, we know [these types of disparities exist](#), and addressing them is critical for suppressing COVID-19 transmission. At this point, unfortunately, our models are not yet capable of capturing this level of detail.

7 Conclusions

In this report, we’ve refined our transmission modeling approach to incorporate daily hospital admission data and variation in the age-distribution of cases, and we demonstrated that this refined model is a platform for reconciling COVID-19 testing, hospitalization, and mortality data to generate estimates of transmission rates and underlying disease prevalence.

We looked in depth at data from King County, and our main finding was that COVID-19 prevalence increased in mid-June but to levels that are still lower than those from King County’s historical peak in late March. This suggests that relaxed restrictions led to increases in transmission but also that mitigation efforts like masking and physical distancing when out are having some success in King County.

We also applied the same modeling approach to data from Pierce and Yakima counties, and in doing so, we found three distinct epidemiologies. Most prominently, in Pierce County, we found that recent increases in cases were likely driven by increased transmission, and we estimated that COVID-19 prevalence in the county in mid-July was as high or higher than prevalence in late March, the former historical peak. Based on this comparison, we conclude that our findings for King County are not representative of Washington’s situation as a whole.

Heterogeneity between counties and sub-county communities has been a prominent feature of Washington’s epidemic. This has been tragically apparent with outbreaks [disproportionately effecting communities of color](#) like Hispanic communities across the state and [the Marshallese community in Spokane County](#). While our modeling for King County as a whole suggests that the epidemic is under better control than in March, it is important to remember that our 0.2% prevalence estimate corresponds to thousands of currently sick individuals, and that these individuals are [significantly more likely to come from King County’s historically disadvantaged communities](#). From a public health perspective, addressing the needs of these communities should continue to be a top priority. And from a modeling perspective, developing better methods to support these goals is our top priority.

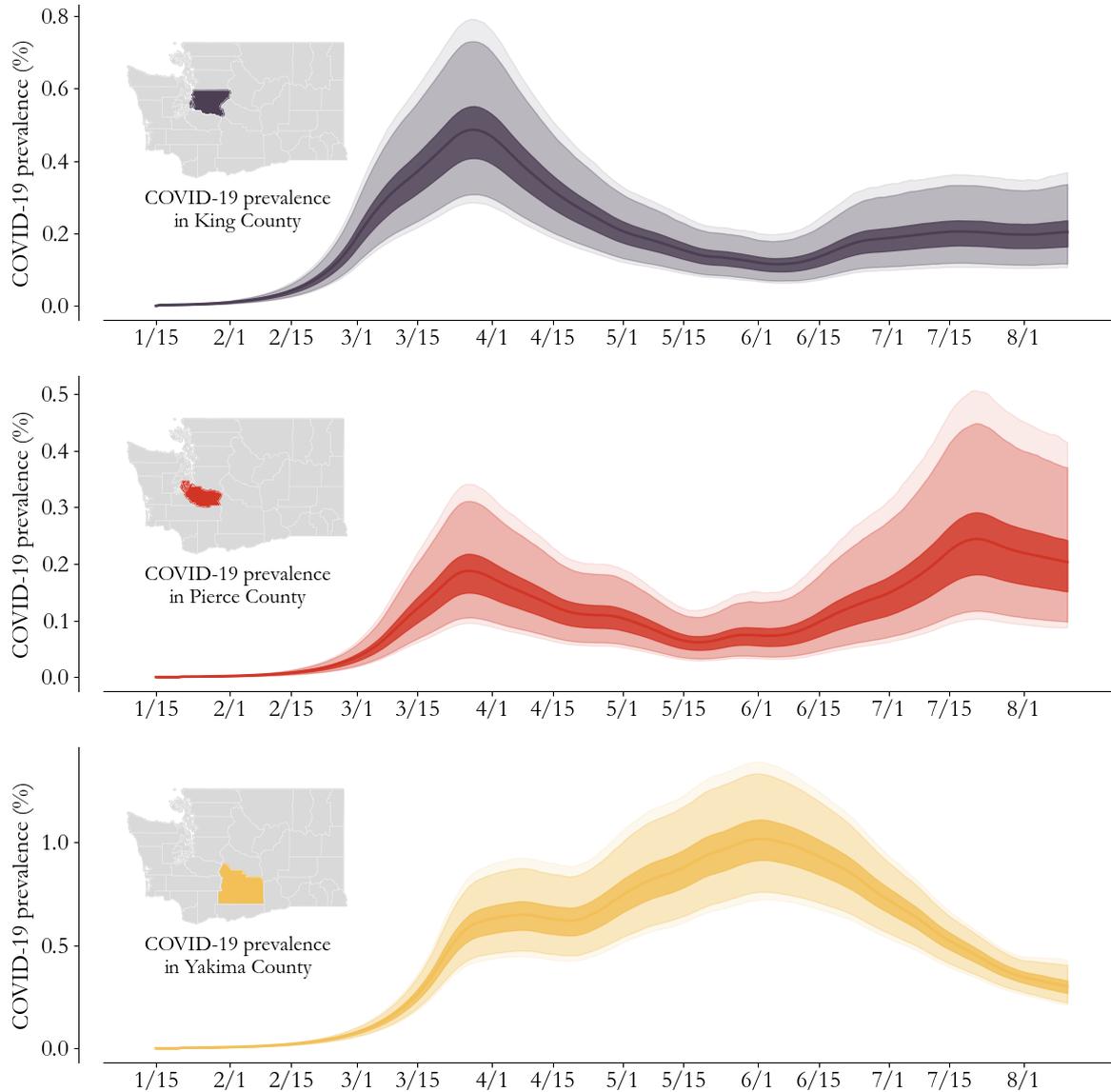


Figure 5: Comparing model-based prevalence estimates for neighboring King (top panel), Pierce (middle panel) and Yakima (bottom panel) counties illustrates three distinct epidemiological situations. While the results of transmission suppression efforts through April and May have in large part been retained in King County, we estimate that Pierce County has experienced a full second wave and has potentially reached a new local prevalence peak. Meanwhile, sustained transmission reductions since June 1 in Yakima County, consistent with efforts to increase mask usage at that time, have continued to suppress prevalence over the last two months. Overall, this comparison highlights that county-to-county heterogeneity is a prominent feature of Washington’s COVID-19 epidemic.

A Updates to our transmission modeling approach

We use the following SEIR model:

$$\begin{aligned}
 S_t &= S_{t-1} - \beta_t S_{t-1} (I_{t-1} + z_{t-1}) \varepsilon_t \\
 E_t &= \beta_t S_{t-1} (I_{t-1} + z_{t-1}) \varepsilon_t + (1 - 1/D_E) E_{t-1} \\
 I_t &= E_{t-1}/D_E + (1 - 1/D_I) I_{t-1} \\
 C_t &\sim \text{Binomial}\{I_t, p_t\} \\
 p_t &= a_w \mathbf{1}(\text{weekend}) + \sum_i a_i \mathbf{1}(t_i < t \leq t_{i+1})
 \end{aligned} \tag{1}$$

where S_t , I_t , and E_t are the number of people who are susceptible, infected, and exposed at time t , $\ln(\varepsilon_t)$ has a zero-mean normal distribution with variance σ_ε^2 , and C_t are daily observed COVID-19 cases. The daily case detection rate, p_t , is assumed to have step-wise structure in time enforced by indicator functions $\mathbf{1}$, with independent values in prespecified reporting periods defined by times t_i and a correction for relaxed testing on weekends. In King County, $t_0 = \text{January 15}$, $t_1 = \text{June 7}$, $t_2 = \text{June 28}$, and $t_3 = \text{August 10}$. We assume $D_E = 4$ days for the latent period, $D_I = 8$ days for the infectious period, and z_t is non-zero only on January 15 and February 1 in King County and January 20 in Pierce and Yakima counties. As mentioned in the main text, this is an arbitrary choice since the timing and number of importations can be traded off without affecting the prevalence time-course where data is available.

Up to adjustment in the assumed timing of reporting period changes and COVID-19 importations, this model is equivalent to the model from [our previous technical report](#). As described in detail there, we fit the model to data hierarchically. First, $\ln(\beta_t)$ and σ_t^2 are estimated using an epidemiological curve based on WDRS data and assumed to be proportional to the underlying infectious population. Then, observed mortality is used to infer the number of importations, z_t , conditional on an infection-fatality-ratio estimate, $\ln(\beta_t)$, σ_t^2 , and the assumption that importation timing is known. Finally, the daily reporting rate, p_t , is estimated by finding a_w and a_i that minimize the discrepancy between the model’s infectious population estimate and observed daily COVID-19 positives in a least-squares sense.

In this report, we introduce three major adjustments to this overall approach:

1. In the past, we based our $\ln(\beta_t)$ and σ_t^2 estimates on an epidemiological curve constructed by smoothing testing volume and multiplying by the observed fraction of tests that are positive. While this led to estimates that were robust to weekend and other short-time-scale fluctuations in testing volume, it was less robust to large-scale, long-term changes in testing volume and strategy (like we saw with free, drive-thru testing in King County). To correct this issue, we pool information from the testing data with daily COVID-19 hospital admissions reported to the WDRS, a measure of severe COVID-19 infections that we assume is independent of testing volume. This information pooling approach is described in detail in Appendix B, but briefly, we find an epidemiological curve that follows hospitalization data on long time-scales (months) but follows testing data on short time-scales (days). This blended epi-curve combines our confidence in global trends in hospitalizations with the lower day-to-day variance in the testing data.
2. Previously, we estimated σ_t^2 where testing data was available (roughly after March 1 in King County), and then back-filled values to define the model during times without testing data. This back-filling led to high transmission variance between January 15 and March 1, since our earliest transmission rate estimates were highly uncertain due to low testing volume. Now, with longer time-series, we back-fill assuming σ_t^2 is equal to the average value where testing data is available. This choice generally leads to lower variance in periods without testing data, tighter model fits to case, mortality, and hospitalization data where we have it, and better constrained uncertainty bounds on key quantities like prevalence and cumulative incidence.
3. Finally, we previously assumed the IFR was widely distributed between 0.2% to 2.5% [based on a meta-study of published values](#). Now, we have adjusted our approach to incorporate the population specific age-distribution and transient time-variation in the age-distribution of cases. This leads to an IFR estimate that accounts for observed demographic trends in the infected population while also estimating the resulting variation in the IFR uncertainty. Moreover, going forward, this better positions the model to incorporate improvements in the standard of care that are not yet well quantified. Our approach is described in detail in Appendix C.

B Estimating transmission rates and variance with testing and hospitalization data

To estimate $\ln(\beta_t)$ and σ_t^2 where testing data is robust (March 1 onward in King County, March 8 onward in Pierce County, and March 20 onward in Yakima County), we apply the RAINIER algorithm (see the appendix [here](#)) to an epidemiological curve that’s proportional to the infectious population up to an unknown proportionality constant. As mentioned above, in the past, we used \tilde{C}_t , the fraction of tests that are positive multiplied by the 3-day spline smoothed daily testing volume. While this approach accounts for day-to-day testing volume variation, we saw recently with the roll-out of free, drive-thru testing sites that it can be biased when testing volume and strategy change dramatically.

To better estimate the epidemiological curve, we incorporate daily COVID-19 hospital admissions, H_t , which give a measure of severe COVID-19 infections and are to a good approximation independent of testing strategy. However, since a large majority of COVID-19 infections don’t result in hospitalization, admission data leads to small numbers and relatively high day-to-day variance. Thus, ideally, we want an epidemiological curve that follows the overall trend in hospitalizations but incorporates the day-to-day trend in testing data, where overall numbers are larger and day-to-day variance is relatively low.

This balance can be framed as an inference problem. Specifically, we want the curve $\tilde{H}_t = f(\mu_t)\tilde{C}_t$ where μ_t is a second order random walk with prior expected variation d_H on the order of months, and f is the logistic function, used to scale the unknowns to be between 0 and 1 since hospitalizations are a subset of cases. We estimate the unknowns by solving

$$\mu_t^* = \min_{\mu_t} \left\{ (H_t - \tilde{H}_t)^2 + \frac{d_H^4}{8} \mu_t^T \Lambda^T \Lambda \mu_t \right\},$$

where all timeseries are column vectors, $\Lambda\mu_t$ is a finite-difference approximation to μ_t ’s second derivative, and the prefactor on the second term comes from equating the expected total-variation in μ_t with that of a sine wave with period d_H . This non-linear optimization problem can be solved numerically, and in practice, we assume $d_H = 28$ days so that μ_t is forced to vary slowly, on a roughly monthly timescale. Note, however, that results throughout this report are robust to reasonable changes in this hyper-parameter.

The outcome of this inference procedure is shown for King County in Figure 6. In the top panel, \tilde{C}_t is plotted in red with observed C_t as open circles, showing that \tilde{C}_t incorporates variation in the fraction of tests that are positive while following observed cases. In the second panel, H_t is plotted as open circles, illustrating the growing discrepancy in trend between the testing data above and hospitalizations starting in June. In yellow, the inferred \tilde{H}_t is similar to \tilde{C}_t early on but smoothly follows the hospitalization data later. The level of inferred adjustment, $f(\mu_t^*)$ is plotted in the lower panel, showing that the effect of high d_H is variation on the scale of months. As a result, \tilde{H}_t does not follow H_t exactly, and day-to-day variation in \tilde{H}_t mirrors the observed variation in \tilde{C}_t .

Applying this approach to data from Pierce County (Figure 7) and Yakima County (Figure 8) illustrates its flexibility. In Pierce County, similar to King County, hospitalization data leads to adjustment in the overall epi-curve. While \tilde{C}_t has a larger second peak than first, \tilde{H}_t has two peaks of comparable size. However, in King County, the divergence between the case and hospitalization data is more pronounced, and our inference procedure captures this feature of the data with a larger overall change in $f(\mu_t^*)$ over time in King than in Pierce. Meanwhile, in Yakima County, long-term trends in the case and hospitalization data have remained largely consistent throughout the outbreak, and $f(\mu_t^*)$ changes relatively little over time leading to generally consistent trends in \tilde{C}_t and \tilde{H}_t .

In a given setting, this modified epidemiological curve, \tilde{H}_t , is used to determine $\ln(\beta_t)$ and σ_t^2 using the multi-step approach we call RAINIER and have described previously. Briefly, \tilde{H}_t is scaled by a proportionality constant $1/r$ and smoothed to timescale D_I to approximate I_t . A similar procedure is repeated to approximate E_t using timescale D_E , and then those approximations are used in conjunction with Eq. 1 to construct a corresponding approximation to S_t , $\ln(\beta_t)$, and σ_t^2 . This approach produces estimates conditional on r , and it can therefore be used to numerically integrate out r dependence under the assumption that r is uniformly distributed between 0.015 and 0.05 based on a wide range of published values for the infection-hospitalization ratio.

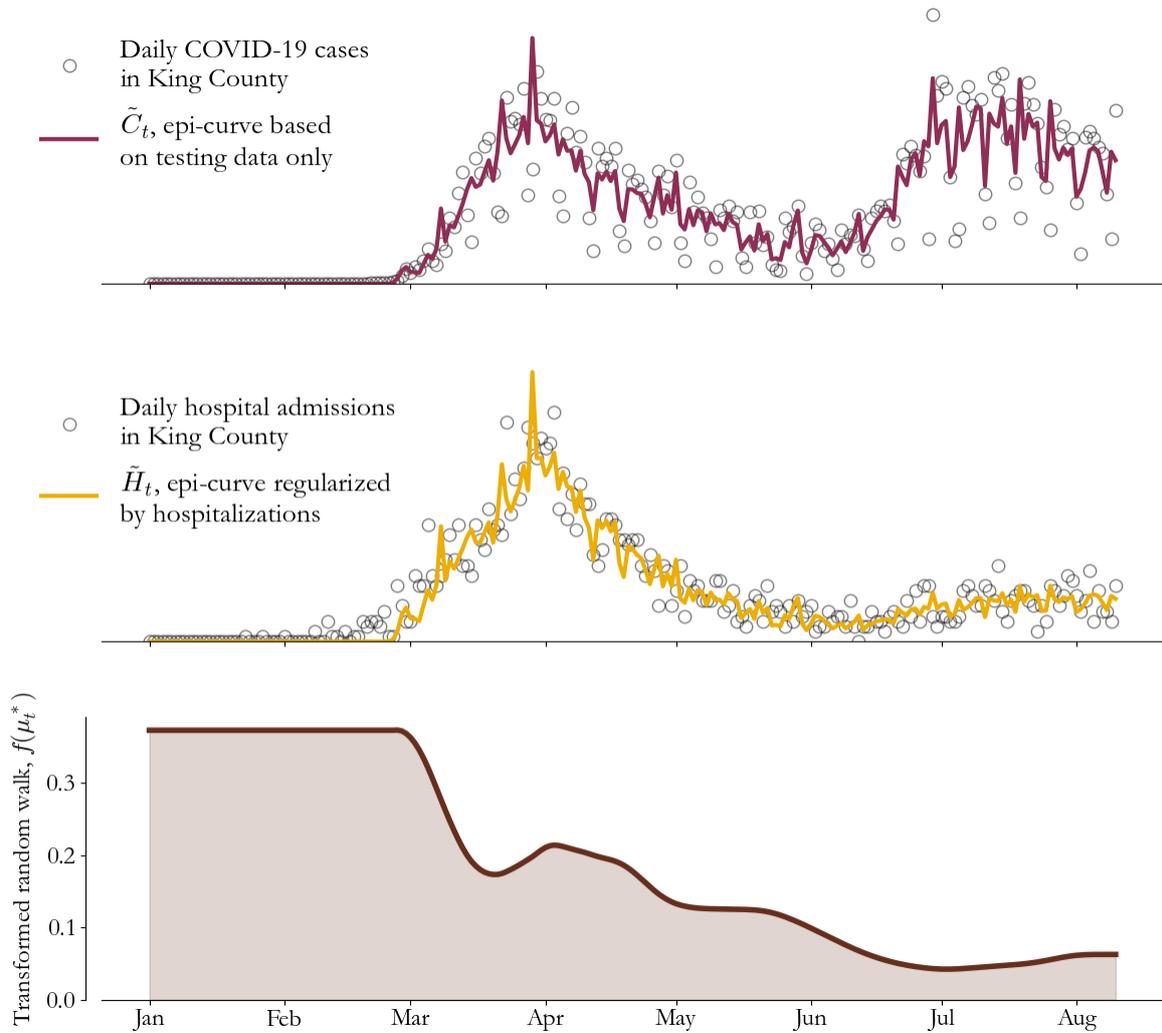


Figure 6: Pooling information from testing and hospitalizations in King County. The testing based epi-curve, \tilde{C}_t , (red, top panel) closely tracks the trend in cases (circles, top panel) and diverges from the trend in daily hospitalizations (circles, middle panel) in June. Multiplying by $f(\mu_t^*)$ (lower panel) smoothly transforms \tilde{C}_t to \tilde{H}_t (yellow, middle panel), which follows the hospitalization data overall but mirrors the variation in \tilde{C}_t . We use \tilde{H}_t to estimate the model's daily transmission rate and variance.

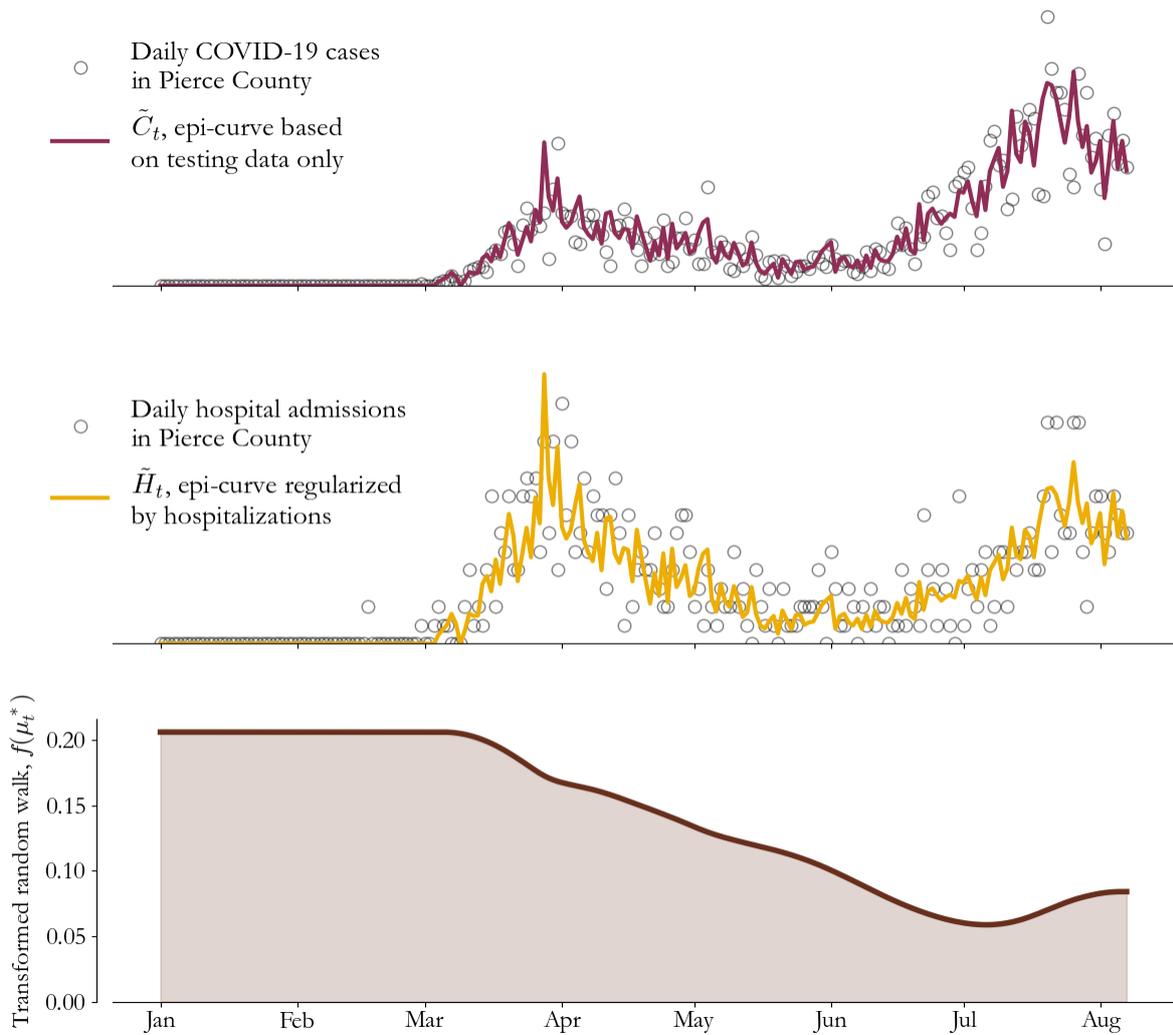


Figure 7: Pooling information from testing and hospitalizations in Pierce County. Similar to King County (see Figure 6), hospitalization data leads to adjustments in the epi-curve’s overall trend and \tilde{H}_t (yellow) has two peaks of comparable size while \tilde{C}_t (red) has a significantly larger second peak. That said, the divergence in trend between the case data (circles, top panel) and hospitalization data (circles, middle panel) is less stark than in King County, and our method captures that with lower overall variation in $f(\mu_t^*)$ (bottom panel).

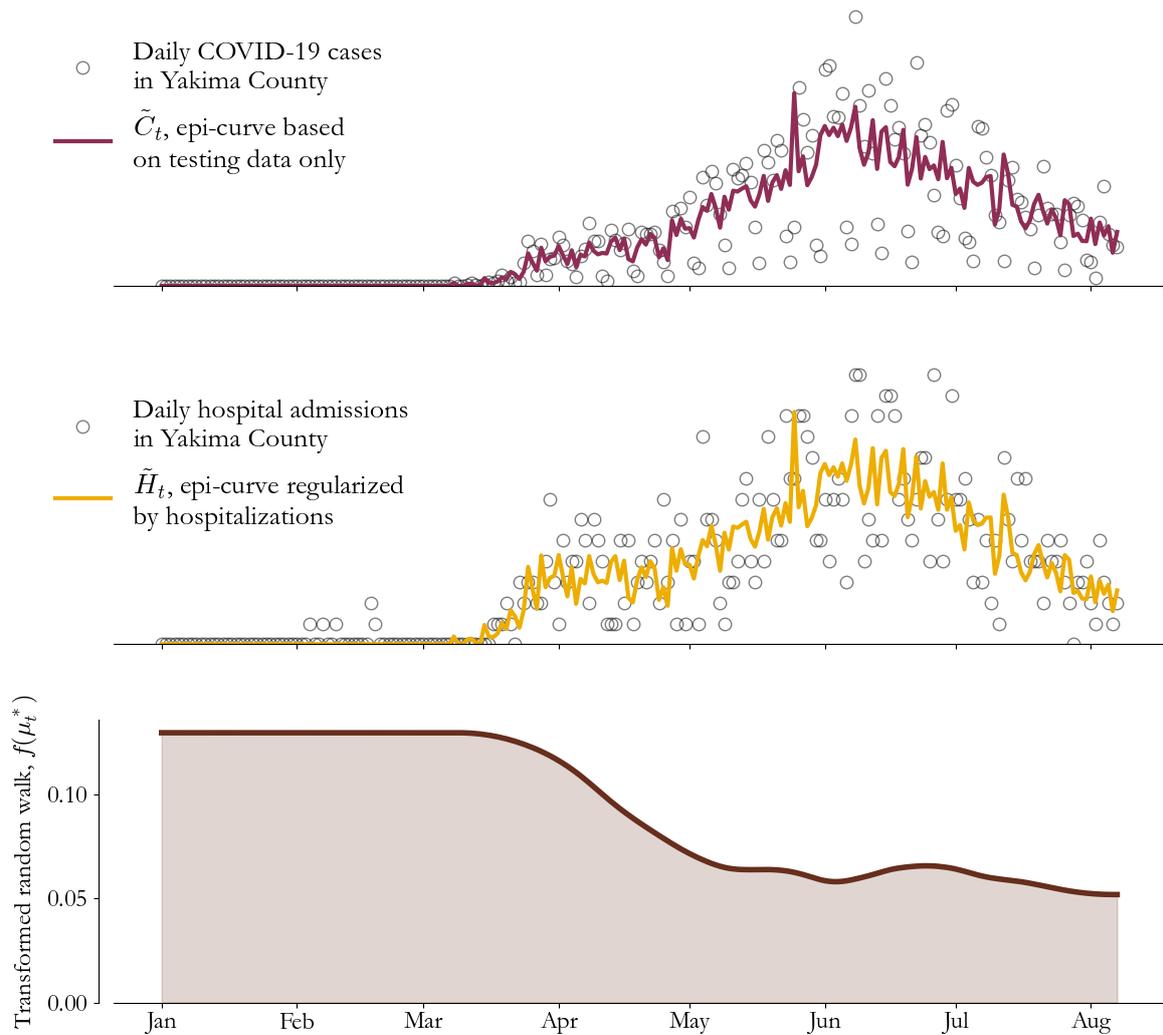


Figure 8: Pooling information from testing and hospitalizations in Yakima County. Of the three counties considered in this report, the divergence in trend between the case data (circles, top panel) and the hospitalization data (circles, middle panel) is lowest in Yakima County, and our approach captures that with relatively low variation in inferred $f(\mu_t^*)$ (bottom panel). As a result, \tilde{H}_t (yellow) and \tilde{C}_t (red) have generally consistent shapes.

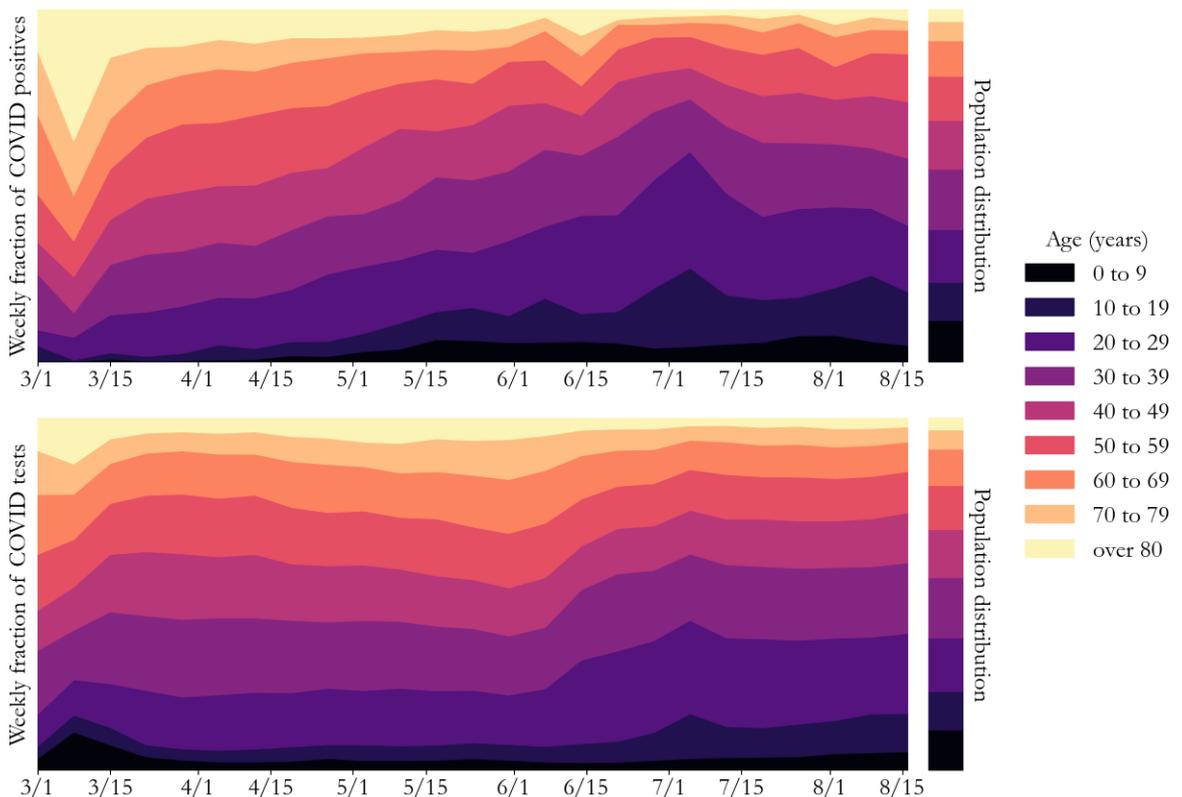


Figure 9: Weekly age distribution of COVID-19 positives (top) and COVID-19 tests (bottom) for King County in 10 year age bins. In early March, older individuals (lighter colors) represented a large fraction of total COVID-19 cases, with significant overrepresentation relative to the age-distribution of the population as a whole (inset). Over time, the age distribution of cases has shifted towards younger individuals (darker colors). That this trend is generally not present in the distribution of tests (albeit with some change coincident with changes in testing strategy on June 5) indicates that the age-distribution of the underlying infected population was shifting.

C Estimating time-variation in the probability of observing severe outcomes

As described by others at the [state-level](#) and [nationally](#), shifts in the age distribution of COVID-19 positives have been a prominent feature of the epidemic. This is shown for King County in Figure 9. In the top panel, weekly fractions of total COVID-19 positives are broken into 10-year age groups, showing a stark transition from older age bins (light colors) to younger age bins (darker colors) from April to July before the trend reverses and levels off. Comparison to the population’s age distribution (inset) shows that 20 to 29 year-olds are particularly over-represented in recent case data. Meanwhile, in the bottom panel, total COVID-19 tests are similarly visualized, showing that the trend in cases cannot be fully explained by a trend in targeted testing, suggesting that King County’s underlying infectious population was getting younger overall from April to July. From our model’s perspective, our prevalence estimates depend heavily on the IFR, which in turn has strong, documented dependence on age.

We take published values for COVID’s age-dependant IFR, $p(d|I, a)$, i.e. the probability of dying (d) given that an individual is infected (I) and is in age-bin a , as a starting point. Our goal is to calculate an all age-average IFR,

$$\theta \equiv p(d|I) = \sum_a p(d|I, a)p(a|I),$$

where the sum is over all age-bins and $p(a|I)$ is the age-distribution of the infected population. Generally speaking, $p(a|I)$ is unknown and difficult to measure since many infections go undetected by the health system. In our dynamic model, $p(a|I) \approx p(a|P)$, the age-distribution of the population from census, since any susceptible individual in the population is equally likely to get infected as any other by assumption.

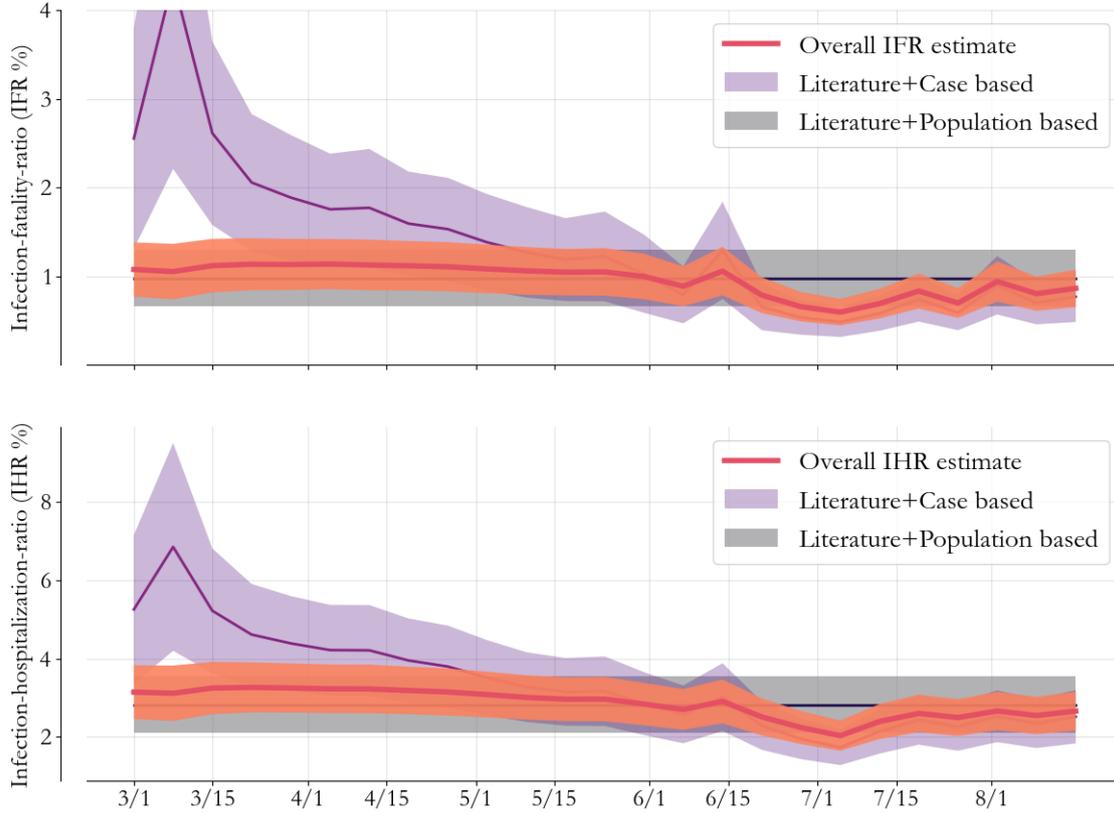


Figure 10: Estimating transient changes in the IFR (top panel) and IHR (bottom panel). While the probability of severe disease as a function of age is generally related to COVID-19’s biology and therefore constant in time to a good approximation, shifts in the infected population’s age distribution can induce shifts in the overall probability of observing severe outcomes from week to week. We take the age-distribution of severity from scientific literature, and we reconcile two estimates of the age distribution of the infected population (census-based in grey, case-distribution-based in purple) for an overall, time-varying estimate of the IFR and IHR (orange). Early on, severity was more likely than expectations based on the population average since cases were concentrated in older individuals. Then, as the epidemic moved into younger age groups, the probability of observing severe outcomes became transiently lower from mid-June to mid-July.

Still, over short-time intervals, $p(a|I)$ can drift away from $p(a|P)$ significantly. The age distribution of cases, $p(a|C)$, offers a noisy estimate of $p(a|I)$ capable of capturing these drifts.

These two approximations to $p(a|I)$ can be reconciled. To that end, we estimate θ for weekly cohorts of the infectious population by assuming $p(d|I, a)$ is uniformly distributed between the published uncertainty intervals and $p(a|C) \approx p(a|I)$ is a Dirichlet distribution with concentration parameter $C_a + 1$ where C_a is a vector of the number of COVID-19 positives in each age bin that week. Meanwhile, we assume that uncertainty in the census is negligible and that $p(a|P)$ is a delta-function centered at P_a , the fraction of the population in age bin a . We sample from each of these distributions 10,000 times to construct two Gaussian approximations to θ , one with high variability based on cases and one static over time based on the census. Convolving these distributions is equivalent to treating the census distribution as a prior and the case-based distribution as a likelihood in a Bayesian inference scheme, leading to an overall estimate of the all-age-average IFR. This process is repeated independently week-to-week.

The top panel of Figure 10 demonstrates the result of this process for King County. In purple (mean with 95% CI shaded), $p(a|I) \approx p(a|C)$ gives high variance estimates that change dramatically over time, reflective of the drift in cases from older age groups to younger age groups. Meanwhile, in grey, $p(a|I) \approx p(a|P)$ is static over time. The overall estimate in orange lies higher than the prior in March, but by a small amount since testing volume was low and variance in the case-based estimate is high. Later, when variance in the case-based estimate decreases, the overall estimate follows the downwards trend and rebound from June through July closely. We use this overall estimate as input to our prevalence model, and in practice we linearly interpolate the weekly estimates to the model’s daily time step and we

back-fill time before March 1 with the census-based estimate.

A similar approach is used to construct estimates of the all-age-average IHR, shown in Figure 10's bottom panel. One key difference, however, is that $p(h|a, I)$, the age-dependent IHR, is taken from published values but scaled by an age-independent scale factor reflective of setting-specific differences in hospital admission criteria. This scale factor is inferred by minimizing the difference between the observed hospitalization time course and modeled hospitalizations, and we find published values need to be scaled by ~ 0.42 in King County, ~ 0.62 in Pierce County, and ~ 0.52 in Yakima County. Note, however, that the IHR does not effect our estimate of prevalence directly (unlike the IFR), and is only used in the comparison to hospitalization data shown in Figure 2's second panel.