Sustained reductions in transmission have led to declining COVID-19 prevalence in King County, WA

Niket Thakkar, Roy Burstein, Daniel Klein, and Mike Famulare
Reviewed by: Jen Schripsema and Hil Lyons
Institute for Disease Modeling, Bellevue Washingon, covid@idmod.org
Results as of 10 a.m. on April 29, 2020

What do we already know?

In our previous report, we estimated that the effective reproductive number in King County as of April 4 remained near one (with a 95% confidence interval from 0.55 to 1.33 and a best estimate of 0.94). This result implied that daily case counts would likely decline slowly and potentially persist at an approximate plateau if societal behavior remained unchanged, putting King County in a situation where relaxation of physical distancing could quickly lead to increased transmission.

What does this report add?

In this report, we update our results for King County with case data from the Washington Disease Reporting System through April 20. We now estimate that the effective reproductive number has been below 1 from March 29 through April 15, with the estimate for April 15 between 0.28 and 1.0 and best estimate 0.64. In addition, we include new results showing model fit to deaths over time and inferences about the prevalence of active infections and cumulative incidence through April 20. By sampling over the widest range of plausible estimates from recent literature for the all-ages infection-fatality ratio to estimate the absolute number of infections in King County since the start of the epidemic, we infer that the prevalence of active COVID-19 infections in King County on April 20 is between 0.1% and 0.68% of the total population, with best estimate 0.29%, and the prevalence time course is consistent with preliminary estimates from the greater Seattle Coronavirus Assessment Network (SCAN). Since January 15, we estimate that cumulative incidence through April 20 was between 0.68% and 4.8%, with best estimate 2.1%.

What are the implications for public health practice?

Our collective efforts to limit physical interaction across society have substantially slowed transmission, but the situation remains sensitive to changes in mitigation and containment policy since the vast majority of the population in King County remains susceptible to COVID-19. Some increased low-risk activity may be compatible with maintaining the effective reproductive number below 1, but continued adherence to physical distancing policies remains necessary to further reduce transmission and bring daily cases to levels where contact tracing and other preventative measures will be most effective.

1   Executive summary

Following the confirmation in late February that COVID-19 was transmitting locally in the Seattle metro area, a suite of physical distancing policies and behavioral recommendations were implemented throughout the month of March. Widespread adherence prevented much of the possible burden of COVID-19, and daily case counts and now deaths have recently been declining in King County. Washington State will soon be permitting some activities previously prohibited, and the consequences of recent changes to policy are not yet known.

Here, we update our estimates of the COVID-19 transmission rate over time in King County, as measured with the effective reproduction number, using updated case and death data from the Washington Disease Reporting System (WDRS) through April 20. **We now estimate that the effective reproductive number (R_e) has likely been consistently below one since March 29, and was between 0.28 and 1.0, with best estimate 0.64, on April 15.** This new estimate is overlapping with the uncertainty interval in our previous report but has increased precision supported by more data.

New to this report are estimates of the prevalence of active infections and cumulative incidence over time. These estimates incorporate the data on deaths over time and estimates from the scientific literature for the fraction of all infections that die from COVID-19 to estimate the total number of infections in King County, and thus the fractional prevalence. There remains a lot of uncertainty in the infection-fatality-ratio (IFR) that measures the fraction of people infected with COVID-19 who will go on to die, with reasonable estimates ranging from roughly 0.2% to 2.4%. Propagating this uncertainty through the model, we estimate that the current prevalence of active infections in King County on April 20 was between 0.10% and 0.68% of the total population, with best estimate 0.29%. Furthermore, the prevalence time course is consistent with preliminary estimates from the greater Seattle Coronavirus Assessment Network (SCAN). Starting from a likely importation to the region on January 15, **we estimate that the cumulative incidence through April 20 was between 0.68% and 4.8%, with best estimate 2.1%. Despite the large uncertainty, we conclude that the vast majority of King County residents are still fully susceptible to COVID-19 infection. Thus, continued physical distancing remains essential to further reduce daily prevalence while testing ramps up and more information-rich data and targeted interventions become available.**

Also included in this report is an expanded technical appendix describing the model in more detail. Accompanying code to generate the figures in this report will be available shortly and linked here in an addendum. (May 22 update: Here's the repository.)

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 continue to use lab testing data provided by Washington State Department of Health (WADoH) through the Washington Disease Reporting System (WDRS), compiled for this report on April 26. Tests are aggregated by specimen collection date, and to hedge against incomplete reporting in the WDRS, we only include data through April 20 in the analysis.

- In this analysis, we also use daily COVID-19 deaths reported to the WDRS over the same time period. These are incorporated into the model by assuming that the infection fatality ratio (IFR) ranges broadly from 0.2% to 2.5% based on published values. Similarly, based on previous studies of early data from China, we also assume 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.

- We assume that a number of COVID-19 importations into King County 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 January 15 to achieve consistency with observed mortality in King County. 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 modeled importations are not reflective of reality. This is a key limitation of our approach that introduces difficult-to-quantify uncertainty.

- In the past, we based R_e estimates on reported COVID-19 positives. Now, to better account for ramp up and subsequent variation in testing, we construct an epidemiological curve by spline smoothing total tests with a 3-day correlation time, which accounts for weekend drops in testing, and multiplying by the observed fraction of tests that were positive (see Figure 5 in the appendix). This better separates changes in R_e from changes in testing volume. We further assume that this approximate epidemiological curve is directly proportional to the number of infectious individuals in the population, with an unknown proportionality constant.

- As before, the case reporting rate (i.e. the daily probability that an infection in King County is reported to the WDRS) is assumed to change over time, with one rate prior to March 10 and a different one thereafter. On top of this, we now explicitly correct for testing decreases over weekends. This has small impacts on results reported to date but more properly accounts for day-to-day variation in testing.

3   Modeling approach

We fit a COVID-specific transmission model to daily testing data and mortality. 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.

- In the model, COVID-19 transmission begins on January 15 with an unknown number of infectious individuals introduced into the population. On all other days, we assume that community transmission is the dominant infection route.

We use a multi-step approach, described in detail in the appendix, to fit the model to WDRS data. Briefly, we construct an approximate epidemiological curve from the observed fraction of tests that were positive and the total number of tests (smoothed to account for day-to-day variation in testing volume), 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 February 28 to April 15. We fill the time from January 15 onward by backward and forward filling. Backward-filling is done with the average value from February 28 to March 2, to approximate unabated COVID-19 transmission in the absence of physical distancing, and forward-filling is done with the estimate on April 15.

Given R_e over time, we calculate the expected number of COVID-19 deaths in the model by assuming an average IFR of 1% and an average time to death of 19 days, and we minimize the difference between observed mortality and the model average as a function of the number of importations on January 15. 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 before and after March 10, controlling for weekends, in a standard weighted least squares regression.

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

4   Updated estimates of the effective reproductive number

Figure 1 shows our current estimates of the effective reproductive number (black dots, 2 standard deviation error bars) in King County. We estimate that on April 15 R_e was between 0.28 and 1 with best estimate 0.64. Moreover, we estimate that R_e was definitively below 1 from March 29 to April 15.

Comparison between our current estimates and previous results (grey dots, 2 standard deviation error bars) highlights the role of uncertainty in estimating R_e. While point estimates have remained largely consistent up to late March, recent progression of the epidemic in King County gives us more certainty about transmission through early April, leading to estimates confidently below $R_e = 1$. While these estimates are consistent with the uncertainty in our previous results, including our mobility-data-based now-casts (yellow), they offer much more definitive insight into the epidemiological situation in King County. Moreover, as our methods have developed to better account for variation in testing volume (see Figure 5 in the appendix), uncertainty in our most recent estimates is slightly lower throughout time period analyzed.

In past reports, we constructed a relationship between R_e and mobility data to estimate R_e in the days masked by WDRS reporting lags and COVID's latent infectious period. However, as we described in our last report, our preferred mobility covariate is no longer consistent with the epidemiological data, and we're still working to resolve this issue. As a result, in this report, we assume R_e is constant from April 15 forward unless otherwise specified. Meanwhile, from January 15 to February 27, before the analysis period, we assume $R_e \approx 3.02$ based on the average estimated value from February 28 to March 2, before widespread physical distancing.
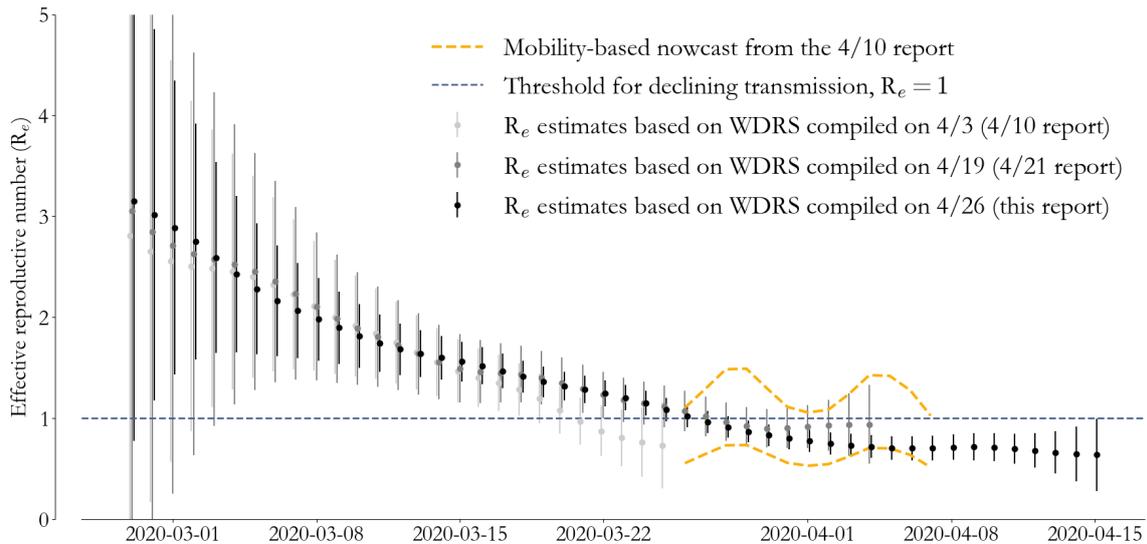
Figure 1: Updated effective reproductive number estimates. Black dots (2 standard deviation error bars) are daily R_e estimates based on testing data from the WDRS, showing declining COVID-19 transmission in King County from March 29 to April 15. Grey dots are similar estimates from previous reports for comparison. In our April 10 report, a now-cast based on mobility data (95% CI in yellow) captured behavior from March 26 to April 7 but with significantly higher uncertainty. (Note also that the declining point estimate for April 11 to April 15 may reflect remaining incomplete data.)

5 Incorporating mortality data gives us information about the COVID-19 reporting system in King County

New in this report, we use daily COVID-19 mortality reported to the WDRS to better characterize the epidemiological situation in King County today. Specifically, in the context of the model, we assume COVID-19 transmission began with an unknown number of importations on January 15, and we estimate the number of importations by minimizing the difference between the average mortality in the model and observed mortality. This restriction to January 15 is made out of necessity: With just recent testing and mortality data, we cannot infer the actual number of importations and their dates since early importations with subsequent community transmission are indistinguishable from a larger number of later importations.

Using our point estimate of 3.02 for the reproduction number averaged from February 28 to March 2, prior to any physical distancing, an all-age-average IFR of 1%, and an assumed average time from infection to death of 19 days, our model requires roughly 39 importations on January 15 to be consistent with observed mortality. We do not think that this is a realistic estimate, and the high initial value reflects a few sources of variation that we are not currently modeling. These include the possibility that the reproduction number was slightly higher than our point estimate (but consistent with the uncertainty shown in Figure 1), that an unobserved super-spreading event occurred early on, or that there was more than one importation with genetically-indistinguishable viruses in January. Regardless of these uncertainties in the initial condition, our model is consistent with current data, suggesting that a substantial number of early cases in King County went undetected by the health system. Over time, with data from additional sources, we may be able to more realistically estimate the number and timing of COVID-19 introductions into King County. In this report, we find that this coarse approach yields a model capable of characterizing the epidemiological situation today.

This is shown in Figure 2. In the top panel, mortality from the fitted model (peach, 50%, 95%, and 99% CI shading) is sampled assuming the IFR is gamma distributed (shape 4, scale 0.25), based on a metastudy of published values and time-to-death is log-normally distributed (log mean 2.8, log standard deviation 0.42), based on early data from China. This captures the observed rise and fall in daily deaths reported to the WDRS (black dots). In the middle panel, infections in the model are compared to daily COVID-19 positives to estimate the daily probability of catching an infectious individual (black dots). A step-wise regression model corrected for relaxed weekend testing aggregates these estimates (yellow) and tells us that a recently infected individual has a roughly 2.2% chance of being tested positive on a given
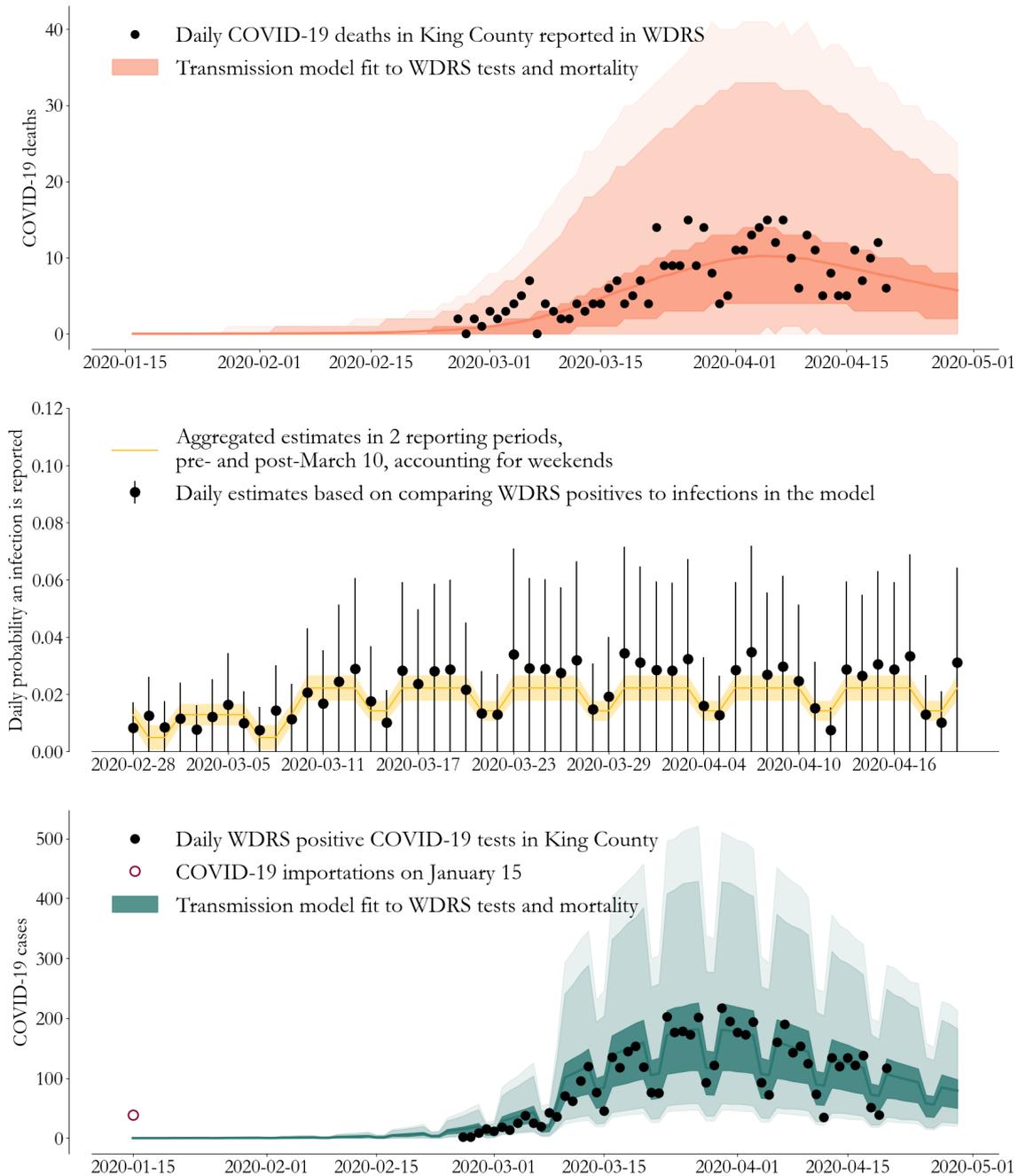
Figure 2: Fitting the transmission model to testing and mortality data from the WDRS. **Top panel:** Modeled mortality (peach, 50% CI dark, 95% CI light, 99% CI lightest) captures the trend in observed COVID-19 mortality (black dots) based on R_e estimates and a free number of importations on January 15. **Middle panel:** Comparing observed COVID-19 positives to modeled infections gives noisy, daily estimates of the probability infections are detected by the health system (black dots, 2 standard deviation error bars). We use a regression model (yellow, mean and 95% CI) to aggregate estimates over time and account for decreased testing over the weekend. **Bottom panel:** With the specified reporting rate, the model (green, intervals as in the top panel) can be compared to observed cases (black dots).
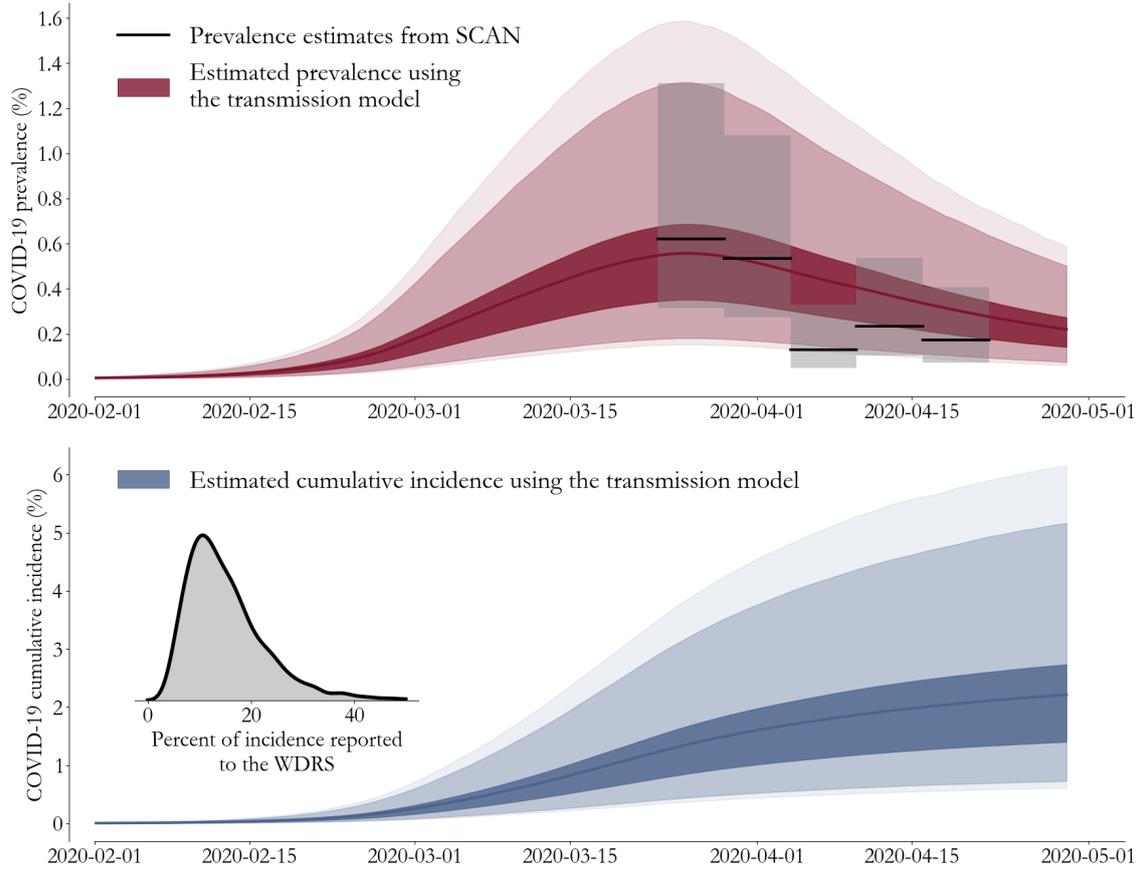
Figure 3: Using the fitted model to estimate prevalence and cumulative incidence. **Top panel:** Model-based prevalence (purple, 50% CI dark, 95% CI light, 99% CI lightest) compared to survey-based estimates from SCAN (mean in black, 95% CI in grey). **Bottom panel:** Cumulative incidence (blue, intervals as in the top panel) indicates that the vast majority of King County's population is still susceptible to COVID-19. In the inset, comparing total test positives by April 20 to cumulative incidence implies roughly 15.4% (5.1% to 36.5% 95% CI) of infections are eventually tested positive and reported to the WDRS.

day of their infection (corresponding to 15.4% over the course of their infection, as discussed in the next section). Finally, in the bottom panel, we compare cases in the model (green) to observed cases in the WDRS, demonstrating that the model captures dynamics from late February to today conditional on the importations in January (red circle).

6 Using the model to estimate active and cumulative infections in King County

The fitted model can be used to estimate daily prevalence (the percent of the population actively infected with COVID-19) and cumulative incidence (the percent of the population no longer susceptible to COVID-19). This is shown in Figure 3.

In the top panel, model estimates (purple) show peak prevalence in late March, and we subsequently estimate that between 0.10% and 0.68%, with a best estimate of 0.29%, of the population was actively infected on April 20. These estimates can be compared to published results from SCAN (mean in black, 95% CI in grey), where prevalence is measured through a self-selecting survey, showing reasonable agreement and an independent check of the model estimates from March 23 through April 9.

In the bottom panel, the corresponding cumulative incidence is shown over time. On April 20, we estimate that between 0.68% and 4.8%, with a best estimate of 2.1%, of King County's population was no longer susceptible to COVID-19. This implies that the vast majority, at least 95% of the population,
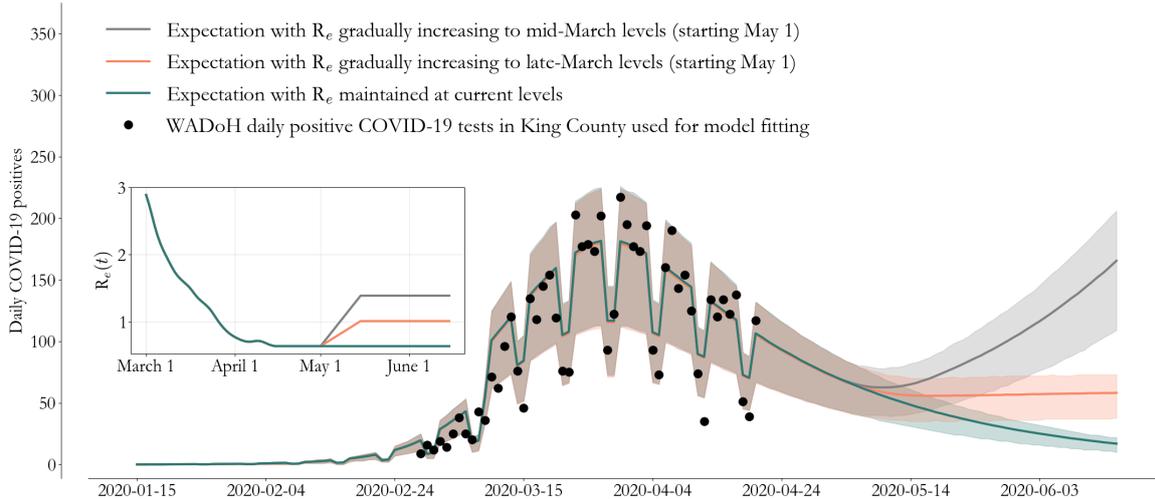
Figure 4: Model-based projections of daily COVID-19 positives under 3 scenarios, shown in the inset. Gradual increase in effective reproductive number to mid-March levels (grey), to late-March levels (red), and maintenance of current levels (green) lead to dramatic changes in projections of COVID-19 cases in King County (50% confidence intervals for each scenario). At current levels of susceptibility, exponential growth of COVID-19 infections is still possible in King County.

remain susceptible, putting King County far from any protection through herd immunity.

Comparing the cumulative incidence to the total number of positive COVID-19 cases in the WDRS as of April 20 allows us to infer what percentage of infections eventually get tested positive. This is shown in Figure 3's inset, where we find that roughly 15% (5% to 37% 95% CI) of infections tested positive and were reported to the WDRS. Note that this number is considerably higher than the probability in Figure 2's second panel since it is relative to the total number of infections instead of actively infectious individuals on a given day. **This estimate implies that a positive test in the WDRS corresponds to between 3 and 19 infections suffered by King County's population over the course of the epidemic to date.**

Even with declining transmission, high susceptibility puts us in a dangerous position. This is illustrated in Figure 4. Under the assumption that testing continues to operate at recent workday levels, we compare estimated daily COVID-19 positives under 3 scenarios (inset): gradual increase of R_e to levels seen in mid-March (grey), gradual increase of R_e to levels seen in late-March (red), and maintenance of current R_e levels associated with continued physical distancing. In the first two scenarios, daily case-counts remain above or comparable to levels we observed before physical distancing policies were implemented in early March. In particular, in the grey scenario associated with significant relaxation of physical distancing polices, King County returns to exponential growth of COVID-19 infections by mid-May. Meanwhile, by maintaining current levels of R_e, it is possible to decrease daily cases to late-February levels by early June.

7 Conclusions

Using WDRS case data through April 20, we infer that the effective reproductive number in King County has remained stable below 1 from March 29 to April 15. This is positive news, indicating clearly that King County is likely experiencing declining COVID-19 transmission.

New to this report, we use data on COVID-19 mortality to subsequently infer population prevalence and cumulative incidence on April 20. **We find that more than 95% of King County's population is still completely susceptible to COVID, stressing the need for continued physical distancing and cautious relaxation of policies to allow for particular, low-risk activities.** If relaxation results in transmission levels observed in mid-March, high susceptibility in King County results in a return to exponential growth in COVID-19 infections.

The model in this report does not describe specific policies or strategies that might allow for relaxed physical distancing with limited increases in transmission. We are working to quantify trade-offs among

strategies with different costs, feasibilities, risks, and benefits, but decision-making under large uncertainty will be a persistent challenge during this pandemic. Information collected by future contact tracing efforts will help clarify the risks associated with different aspects of physical distancing policy, but continued, wide-spread distancing is necessary to bring daily COVID-19 positives to levels where contact tracing efforts are sustainable.

Speaking briefly as members of the King County community: We are proud of the collective response King County has had in the face of this pandemic. The evidence is clear that our efforts have accomplished a great deal, making our community substantially safer and healthier.

A   Fitting a transmission model to case and mortality data

We use the following SEIR model:

$$
\begin{aligned}
S_t &= S_{t-1} - \beta_t S_{t-1} \left(I_{t-1} + z_{t-1} \right) \varepsilon_t \\
E_t &= \beta_t S_{t-1} \left(I_{t-1} + z_{t-1} \right) \varepsilon_t + \left(1 - 1/D_E \right) E_{t-1} \\
I_t &= E_{t-1}/D_E + \left(1 - 1/D_I \right) I_{t-1} \\
C_t &\sim \text{Binomial} \left\{ I_t, p_t \right\} \\
p_t &= a_0 \mathbf{1}(t < t_r) + a_1 \mathbf{1}(t \geq t_r) + a_w \mathbf{1}(\text{weekend})
\end{aligned}
\tag{1}
$$

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_t^2, and C_t are daily observed COVID-19 cases in King County. The case detection rate, p_t, is assumed to have step-wise structure in time enforced by indicator functions $\mathbf{1}$, with independent values before and after $t_r = $ March 10 and a correction for relaxed testing on weekends. 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, 2020. This model is similar to the model from our previous report, but there are some key differences:

- In the past we assumed the transmission rate, β_t, was log-linearly related to mobility-based covariates. In this case, we assume that β_t varies over time with structure inferred directly from positive and negative testing data.

- We assume that the reporting rate, p_t, is a step-function, with unknown, independent values before and after March 10, 2020, corresponding to a significant increase in daily tests in King County. We also add an additional correction for relaxed testing on weekends.

- Finally, we previously assumed z_t, which models COVID-19 importations into King County, was equal to 1 on both January 15 and on February 25. Now, we assume that it is non-zero only on January 15, with unknown value inferred from observed mortality data.

The model is fit to observed COVID-19 testing and mortality data from the WDRS. This is done in distinct steps (detailed mathematically in the following section):

1. RAINIER, our algorithm for estimating $\ln(\beta_t)$ and therefore R_e from epidemiological data (also described in Appendix 1 of our previous report), is applied to smoothed total tests times the daily fraction of observed tests that were positive (red in Figure 5's top panel). This is in contrast to our previous work, where the algorithm was applied directly to positive cases (dashed black in Figure 5's top panel). We made this change to better account for the ramp up of testing and for weekly oscillations in total tests, shown in Figure 5's second panel. Overall, we think this approximate epi-curve is more reflective of epidemiological changes than positive cases taken at face value; however, we also see that this process sometimes introduces spikes in the epi-curve during weekend periods of relatively lower testing, suggesting that weekend tests may be restricted to a population more likely to be positive. Note finally that we still assume that the epi-curve can be scaled by an unknown constant to coarsely approximate the infectious population over time, and we numerically integrate over that constant.

   Given point estimates and uncertainty for $\ln(\beta_t)$ over the analysis period from February 28 to April 20, we approximate $\sigma_t^2 = \text{Var}[\ln(\varepsilon_t)]$ as $\text{Var}[\ln(\beta_t)]$. This is a cautious assumption since uncertainty in $\ln(\beta_t)$ estimation is added to uncertainty in the transmission process.

   Finally, this approach generates β_t and σ_t estimates from February 28 to April 15. To fill in the remaining time from January 15 to April 20, we backward and forward fill. Backward-filling is done
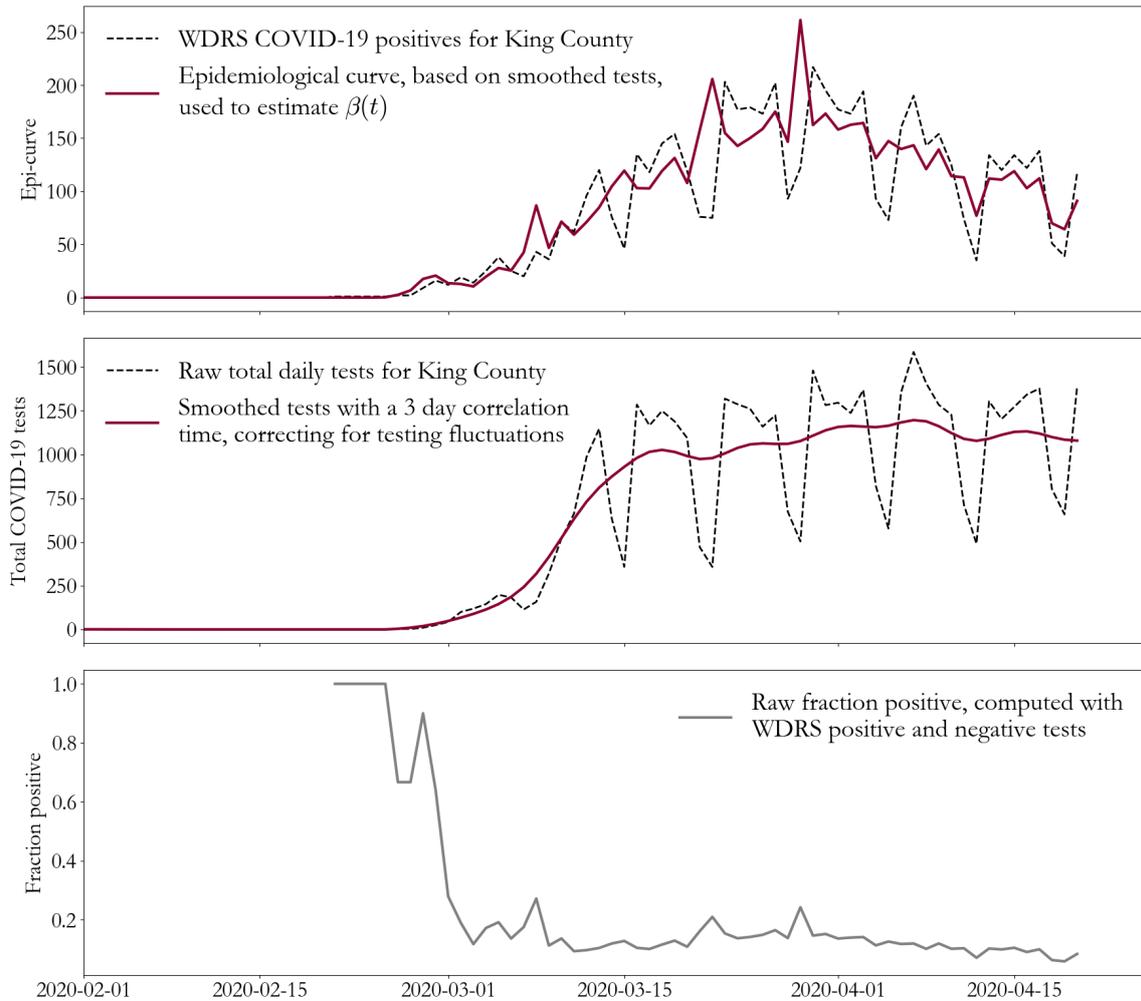
Figure 5: Accounting for changes in testing volume. **Bottom panel:** Fraction of tests that are positive in the WDRS has been relatively stable since late February. **Middle panel:** Daily tests in King County (black dashed line) have clear weekend drops that can be smoothed (red) by a 3-day smoothing spline. **Top panel:** Daily positives (black dashed line) is fraction positive multiplied by total tests. In red is fraction positive multiplied by smoothed total tests, a similar curve corrected for testing fluctuations that we use for epidemiological inference.

with the average estimate up to March 2, approximating unabated transmission before physical distancing. Forward-filling is done using the estimate on April 15.

2. With β_t and σ_t over the full time period, we find the value of z_t on January 15 that best captures the observed mortality data under the assumption that z_t is zero on all other days. We make this choice for identifiability reasons: With unknown numbers of importations at unknown times, it is not possible to uniquely explain observed mortality. In the future, by incorporating other data sources, we may be able to infer a more realistic set of importations into King County. As it is, our approach gets us a model consistent with transmission dynamics today.

3. Finally, with z_t and β_t specified, we estimate the reporting rate over time, p_t, by comparing positive COVID-19 tests reported in WDRS to infections in the model, estimating the probability that infections are caught by the health system each day. These noisy estimates are input into a standard weighted least squares regression to estimate a_0, a_1, and a_w.

B Additional mathematical details

The algorithm above can be described more formally in terms of probability. Doing so highlights the role of assumptions we make and clarifies the effect of particular approximations.

Our goal is to calculate and maximize the posterior distribution

$$p(\beta_t, p_t, z_t \mid C_t, \tilde{C}_t, M_t),$$

where M_t is the observed mortality over time, $\tilde{C}_t$ is the approximate epi-curve constructed in Figure 5, and the distribution is implicitly conditional on the relationships in equations 1. This is a complicated distribution, difficult to evaluate as a whole and likely with multiple local maxima. To make progress and motivate an optimization algorithm, we break the posterior into conditionally dependent pieces and make strict conditional independence assumptions. More specifically, we approximate

$$
\begin{aligned}
p(\beta_t, p_t, z_t \mid C_t, \tilde{C}_t, M_t) \approx\ & p(\beta_t \mid \tilde{C}_t) \\
& \times p(z_t \mid M_t, \beta_t) \\
& \times p(p_t \mid C_t, \beta_t, z_t),
\end{aligned}
\tag{2}
$$

an approximation that facilitates inference by establishing a hierarchy amongst model parameters and connecting individual parameters to particular subsets of the full dataset. In the remainder of this section, we focus on evaluating and maximizing each of the terms above individually, which corresponds to an approximate posterior maximum since all the terms have value between 0 and 1.

The first term is evaluated by an algorithm we call RAINIER, which was described at a high level in our previous report. Mathematically, we use the relationships in equations 1 to construct β_t in terms of $\mathbf{X}_t = (S_t, E_t, I_t)$, the collection of hidden states. Then, to estimate $\mathbf{X}_t$, we assume that I_t is directly proportional to $\tilde{C}_t$ and E_t is proportional to $D_E \tilde{C}_{t-D_E}/D_I$ both with unknown proportionality constant $1/r$. This latter relationship is motivated by steady-state analysis of equations 1. We then have

$$p(\beta_t \mid \tilde{C}_t) = \int_0^1 dr \int d\mathbf{X}_t\, p(\beta_t \mid \mathbf{X}_t, r) p(\mathbf{X}_t \mid \tilde{C}_t, r) p(r \mid \tilde{C}_t). \tag{3}$$

The last term in the integrand, $p(r \mid \tilde{C}_t)$, can be shown to be uniform under the assumption that $\tilde{C}_t \sim \text{Binomial}\{I_t, r\}$ with the weakly informative prior $r \sim \text{Beta}(2, 1)$ enforcing that r is non-zero. Meanwhile, we construct a Gaussian approximation to $p(\mathbf{X}_t \mid \tilde{C}_t, r)$,

$$
\begin{aligned}
p(\mathbf{X}_t \mid \tilde{C}_t, r) &= p(I_t, E_t \mid \tilde{C}_t, r) p(S_t \mid E_t) \\
&\approx \mathcal{N}(I_t \mid \hat{I}_t, \Sigma_I) \mathcal{N}(E_t \mid \hat{E}_t, \Sigma_E) \mathcal{N}(S_t \mid \hat{S}_t, \Sigma_S)
\end{aligned}
$$

where means $\hat{E}_t$ and $\hat{I}_t$ and covariance matrices Σ_E and Σ_I are constructed using smoothing splines with roughness penalties $D_E^4/8$ and $D_I^4/8$ respectively, enforcing priors on total-variation with expected value equal to that of sine-waves with periods D_E and D_I. Then, rearranging equations 1 gives

$$S_t = S_0 - \sum_{i=1}^{t-1} (E_i - (1 - 1/D_E)E_{i-1})$$

which can be used to calculate $p(S_t \mid E_t)$ through standard manipulations of multivariate Gaussians. To move to the transmission rate, again rearranging equations 1, we have

$$\ln \beta_t + \ln \varepsilon_t = \ln \{E_t - (1 - 1/D_E)E_{t-1}\} - \ln \{I_{t-1} + z_{t-1}\} - \ln S_{t-1},$$

which can be used in a Laplace approximation to the integral over $\mathbf{X}_t$ in equation 3 under the assumption that z_t is known during the analysis period. In this report, as mentioned above, we have a strict prior on z_t setting it to zero on all days but January 15. Finally, with the Laplace approximation to $p(\beta_t \mid \tilde{C}_t, r)$, the integral over r is computed numerically on 50 uniformly spaced quadrature points between 0 and 1.

As described above, this approach is applied only to data from $t_0 =$ February 28 to $t_1 =$ April 20, where the spline approximation to $\mathbf{X}_t$ is guaranteed to remain positive since $\tilde{C}_t$ is far enough from zero. Furthermore, in constructing the approximation to E_t based on delaying data by D_E, we are limited to estimates up to April 15. To fill in the remaining time, we assume $p(\beta_t \mid \tilde{C}_t)$ can be forward and backward filled. Backward-filling is done with the average estimate up to March 2, approximating unabated transmission before physical distancing. Forward-filling is done using the estimate on April 15.

Conditional on β_t, the second term in equation 2 can be calculated using Bayes' theorem,

$$p(z_t \mid M_t, \beta_t) \propto p(M_t \mid z_t, \beta_t) \prod_{i \neq 1/15} \delta(z_i),$$

where the prior enforces the assumption that z_t is nonzero only on January 15. This strict choice is made out of necessity: With only mortality data from late February to today, timing and number of importations are not jointly identifiable since fewer importations can always be assumed to have occurred earlier leading to similar mortality closer to today.

Given β_t and z_t, the average mortality in the model can be calculated using the mean trajectory associated with equations 1. Specifically, we calculate the average number of new exposures, scale by the average 1% infection-fatality-ratio, and lag by the average time to death of 19 days. Estimation of the one non-zero element of z_t is carried out by assuming $p(M_t \mid z_t, \beta_t)$ is a least-squares likelihood and numerically minimizing the difference between the average model morality and M_t. While this approach gets us a working point estimate for z_t, uncertainty is poorly captured by the least-squares approximation, and for now we use the point estimate only. In the future, we will explore the possibility of more sophisticated inference of z_t using other sources of data, but for the time being, z_t should be thought of as a modeled set of importations that yields the appropriate dynamics later on, not as an inference of the actual COVID-19 introductions into King County.

Finally, to calculate $p(p_t \mid C_t, \beta_t, z_t)$, we sample the model for I_t trajectories. Mathematically,

$$
\begin{aligned}
p(p_t \mid C_t, \beta_t, z_t) &= \prod_{i=t_0}^{t_1} \int dI_i \, p(p_i \mid C_i, I_i) p(I_i \mid \beta_t, z_t) \\
&\approx \prod_{i=t_0}^{t_1} \frac{1}{N} \sum_{j=1}^{N} \mathrm{Beta}(C_i + 1, I_{j,i} - C_i + 1) \\
&\approx \prod_{i=t_0}^{t_1} \mathcal{N} \left(p_i \,\middle|\, \mathrm{E}\left[\frac{C_i + 1}{I_i + 1}\right], \mathrm{Var}\left[\frac{C_i + 1}{I_i + 1}\right] \right)
\end{aligned}
$$

where in the second line we calculate $p(p_i \mid C_i, I_i)$ by exploiting conjugacy of Beta and Binomial distributions and the integrals over I_i are approximated using $N = 10,000$ sample trajectories. The sum over Beta distributions is taken by coarsely approximating them by their mean in the final line, where the expectation and variance is over samples. This completely specifies the posterior distribution in equation 2 up to the approximations and assumptions we have made.

To fully specify the transmission model, sample means and variances of p_t for each time step can then be used in a weighted least squares regression to estimate parameters a_0, a_1, and a_w. Then, with the complete model, the distribution of $\mathbf{X}_t$ was sampled 10,000 times to estimate prevalence over time. Distributions of cases and mortality can also be computed from these samples by additionally sampling the binomial reporting model or the distributions of infection-fatality ratios and time-to-death specified in the main text.
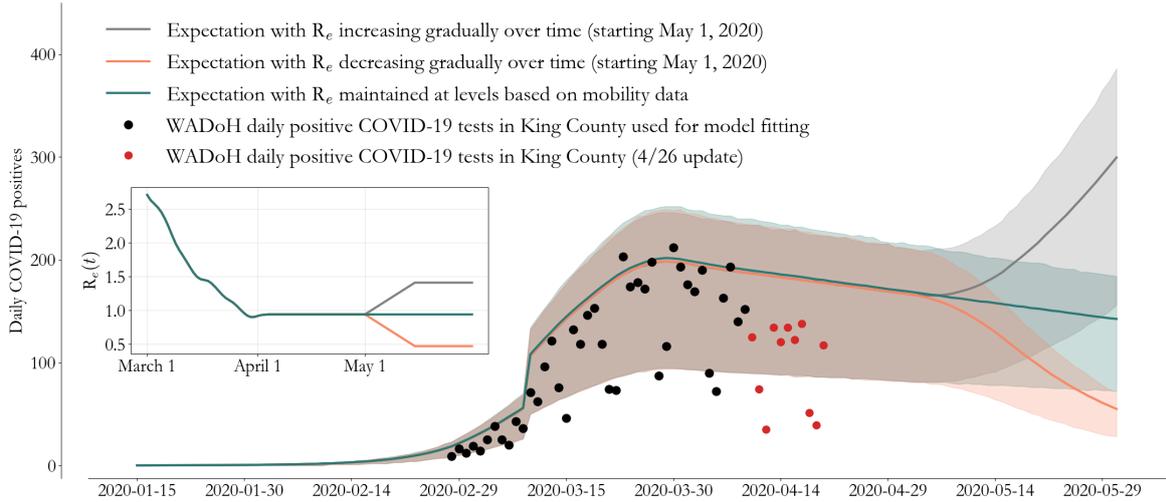
Figure 6: Previous model-based projections of cases under 3 scenarios, shown in the inset. Gradual increase in effective reproductive number (grey), maintenance (green), and gradual decrease (red) lead to dramatic changes in projections of COVID-19 cases in King County (50% confidence intervals for each scenario). This is a replication of Figure 3 from our previous report, with new data superimposed. Red dots show new daily positives from the most recent WDRS update, and indicate that current data are on the lower end of projections under all three scenarios.

C Comparing projections from the previous report to today

In our previous report, we used a transmission model to forecast 3 scenarios similar to those described above, with increasing (grey), constant (green), and decreasing (red) transmission associated with corresponding changes in physical distancing. Here, we use the updated data to evaluate those projections.

Updated data, the red dots in Figure 6, are on the low end of our projections under all three scenarios. This illustrates why, in Figure 1, our updated model has more confidence that R_e was definitively below 1 from March 29 to April 15. More generally, this comparison highlights the difficulty in definitive forecasting: All three scenarios did equally well in predicting case data, and future transmission still depends entirely on societal behavior.