

# Social distancing and mobility reductions have reduced COVID-19 transmission in King County, WA

Niket Thakkar<sup>1</sup>, Roy Burstein<sup>1</sup>, Hao Hu<sup>2</sup>, Prashanth Selvaraj<sup>1</sup>, and Daniel Klein<sup>1</sup>

<sup>1</sup>Institute for Disease Modeling, <sup>2</sup>Bill & Melinda Gates Foundation

Results as of March 29, 2020 8 p.m.; [covid@idmod.org](mailto:covid@idmod.org)

This work follows a previous study on changes in mobility in the greater Seattle Area, please refer to [that document](#) for more details. Results in this document should be interpreted with respect to the stated assumptions and limitations.

## Executive Summary

To stem the spread of COVID-19, Washington State instituted increasing levels of separation policies between March 11 and 15, 2020, including closing schools and prohibiting large group gatherings. Most recently, Governor Inslee has started the “Stay Home, Stay Healthy” policy - a two week shelter-in-place order for all but essential services. These policy changes were supported by [computational modeling](#), which explored social distancing at 25%, 50%, and 75% levels. In that analysis, only the 75% social distancing scenario was strong enough to bring the effective reproductive number ( $R_e$ ) below 1, causing the number of new infections to decrease over time.

In this report, we quantify the impact of the mid-March policies on reducing COVID-19 transmission in King County, Washington. **Our main result is that the epidemic has slowed, but that more progress is necessary.** This result is clear from epidemiological data alone (tests, diagnoses, and deaths), but this data is lagged due to inherent COVID delays such as the latency period of infection. Still, based on this data for King County, **the trend clearly shows  $R_e$  decreasing from about 2.7 in late February to roughly 1.4 on March 18th.** These estimates come with high uncertainty, and while the trend is encouraging, it remains unlikely that COVID transmission in King County was below the 1.0 threshold on the 18th.

To estimate where we might be more recently (up to March 24th), we have leveraged [Facebook Data for Good Project - Disaster Maps](#), which responds rapidly to policy changes. We have extracted four measures of home-occupancy and daily population flows and fit statistical models to connect these covariates with  $R_e$ . However, the result is indeterminate. Some covariates lead us to believe that  $R_e < 1$ , whereas others suggest  $R_e > 1$ . These and other covariates explain historical data equally well, so **we cannot say with confidence if  $R_e$  is below or above the critical threshold of 1 today.** We are on the cusp, and a few more days of reliable data will give us greater confidence regarding this critical threshold.

This result is greatly encouraging for the people of King County, however compliance with social distancing policies has not been steady (as evidenced by increased mobility over weekends), and thus progress is precarious. Compliance with social distancing policies will remain an important facet of daily life over the coming days and weeks to push the effective reproductive number for COVID-19 in King County as far below 1 as possible.



## Key inputs and assumptions

We use lab testing data provided by Washington State Department of Health (WADoH) through the [Washington Disease Reporting System \(WDRS\)](#). Daily positive COVID-19 tests, negative tests, and COVID-19 mortality were aggregated across testing facilities to county level by WADoH. Tests were assigned to days based on the specimen collection date while mortality was assigned to the date of death. Note that we are using a version of this dataset compiled on March 28. As data is compiled, retrospective changes occasionally occur because of corrections from labs, resolutions of pending tests, and delays in assigning tests to counties. We do not expect these changes to qualitatively change our results for King County, where the overall number of cases is large, and to hedge against this instability, **we use data only up to March 23rd.**

A key assumption of the model is that case data from February 26th to March 23rd can be treated as a sample representative of community transmission in King County, and that daily changes in cases during this period are correlated with changes in the number of active infections instead of changes in the availability or targeting of testing. The case-to-infection rate is estimated from data, and has potential to vary over time, but is assumed constant for this analysis. For the purposes of this work in King County, where the number of positive COVID-19 tests per day have generally increased over this period, and proportion positive has remained steady, we view the assumption of a constant reporting rate as conservative. Scale-up of testing will be interpreted by the model as an increase in the number of COVID-19 infections.

## Modeling approach

We fit a COVID-specific transmission model to daily case counts in King County from February 26th to March 23rd. The key modeling assumption is that individuals can be grouped into one of four disease states - susceptible, exposed but non-infectious, infectious, and recovered. In addition, we assume:

- COVID-19 has an asymptomatic period that lasts about 4 days.
- After the asymptomatic period, those exposed to COVID-19 are infectious for about 8 days.
- The probability of testing an infected individual is unknown but roughly constant from February 26th to March 23rd in King County.

We use a multi-step approach to generate point estimates of  $R_e$ , the effective reproductive number. Technical details can be found in the appendix, but concisely, we assume that case data can be scaled up by  $1/p$ , where  $p$  is reporting rate, in order to coarsely approximate the total number of infected individuals<sup>1</sup>. Since COVID-19 infections last roughly 8 days, we expect the number of infecteds to vary with an approximately 8 day timescale, and we smooth the coarse approximations accordingly. A similar procedure is repeated for the number of latently infected individuals (this time smoothing to a roughly 4 day time scale). Comparison of the rates of change of these estimates can then be used to estimate  $R_e$  using the transmission model equations in the appendix. Finally, since the reporting rate is unknown,

---

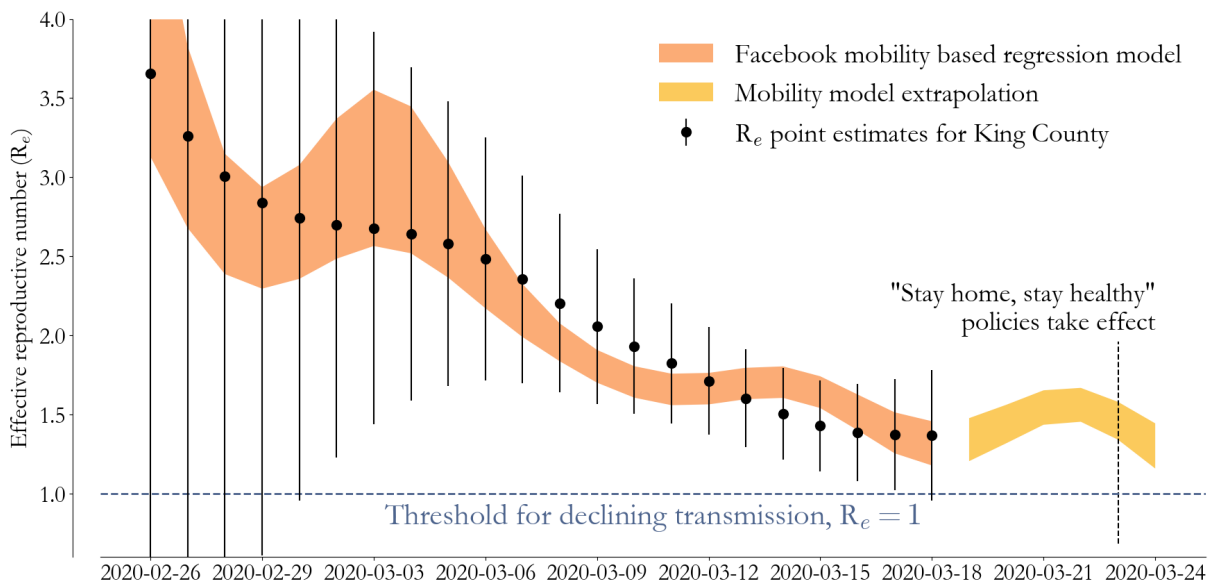
<sup>1</sup> Because the entire population is naive to COVID-19, this is the attack rate.

this procedure is repeated for a range of reporting rates from 0 to 1, and the mean  $R_e$  and uncertainty is collected across all reporting rates.

This procedure gives us high variance, daily estimates of  $R_e$  in King County from February 26th to March 18th. We are unable to estimate  $R_e$  from March 19th to 23rd because of COVID-19's 4 day latent period.

## Estimates of effective reproductive number for King County

Using epidemiological data, conditioned on the assumption that changes in case data from February 26th to March 23rd are reflective of epidemiological changes, we find that the effective reproductive number in King County has declined considerably. We estimate that before social distancing began on March 2nd,  $R_e$  was approximately  $2.7 \pm 0.9$ . However, on March 18th, we estimate  $R_e$  is  $1.4 \pm 0.2$ , which is on the cusp of declining transmission. Daily estimates are shown in black with two standard deviation error bars in Figure 1.



**Figure 1.** Daily estimates of the effective reproductive number are computed using WDRS case detection data (black dots, 2 standard deviation error bars). Facebook mobility data can be used to explain  $R_e$  variation by fitting a log-linear regression model with a mobility-based covariate (95% CI in orange). The fitted relationship between  $R_e$  and mobility can be used to extrapolate past inherent delays in the case data due to COVID-19's latent infection period (95% CI in yellow). The recent bump came from heightened mobility during the weekend.

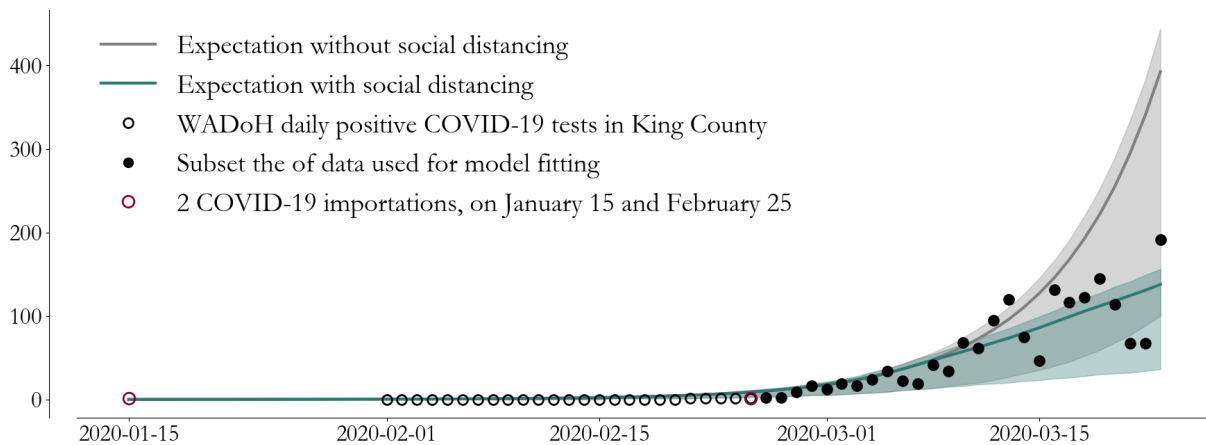
## Incorporating mobility data to explain variance and nowcast

These point estimates, based entirely on the WDRS case data, have relatively large uncertainty that comes from multiple sources: changes in behavior, under-reporting, and randomness in transmission. Each of these sources acts to inflate the uncertainty on the epidemiological situation, particularly in an epidemic's very early days. This motivates the need for connecting case-based  $R_e$  estimates to auxiliary data. In particular, comparing our estimates to mobility data gives us a way to control for some of this uncertainty and generate more confident  $R_e$  estimates. This report leveraged Facebook's Disease

Prevention Maps that are part of its Data for Good program, which are based on aggregated mobility data and use privacy preserving techniques.

This is shown in orange in Figure 1. Facebook mobility data is used to estimate the change in population flux between day and night over time (see the following section for details), and a log-linear model is fit to the  $R_e$  point estimates. **The fitted estimate of  $R_e$  on March 18th decreases to  $1.3 \pm 0.07$ , a value consistent with the case data but with higher certainty.** Finally, connecting the case data to the mobility data allows us to extrapolate into the days masked by COVID-19's latent period, after March 18th in this case. The 95% confidence interval for this extrapolation is shown in yellow, where using this measure of mobility we find that increased mobility over the weekend drives an increase in  $R_e$ .

Given the mobility based estimates of  $R_e$ , we use a transmission model to compare case detections with and without  $R_e$  declines. COVID-19 importations into King County are based on [current information from NextStrain](#). At this early stage of the outbreak, uncertainty remains high, but in the absence of social distancing and at the same testing capacity, our expectation for positive COVID-19 tests on March 23rd is roughly 3 times higher than it is with social distancing.

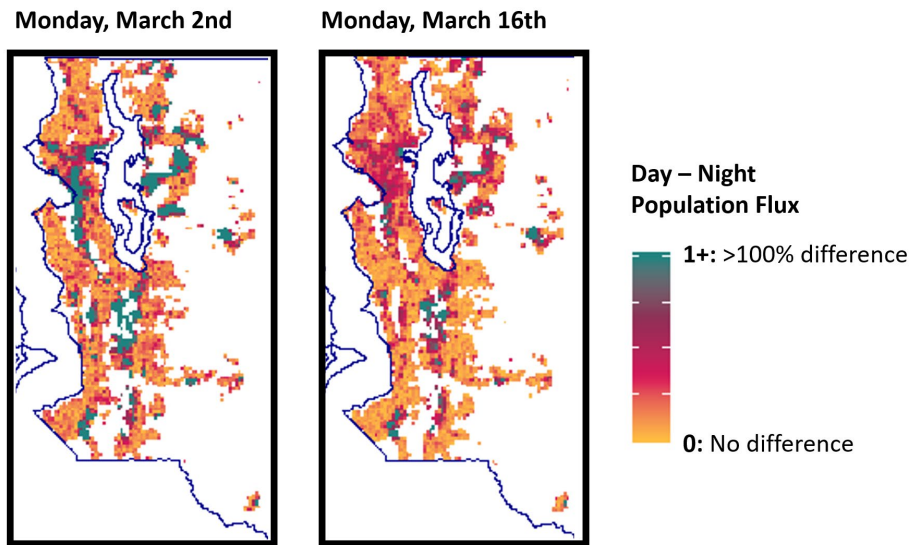


**Figure 2.** Given  $R_e$  over time, we can parameterize a COVID-19 transmission model to estimate social distancing's effect on observed cases. While expectations for transmission with a mobility-based social distancing model (green line) agree better with observations (black dots) than expectations for transmission without distancing (grey line), uncertainty (50% confidence intervals are shaded for each model) remains high, emphasizing the need for sustained adherence to aggressive social-distancing policies.

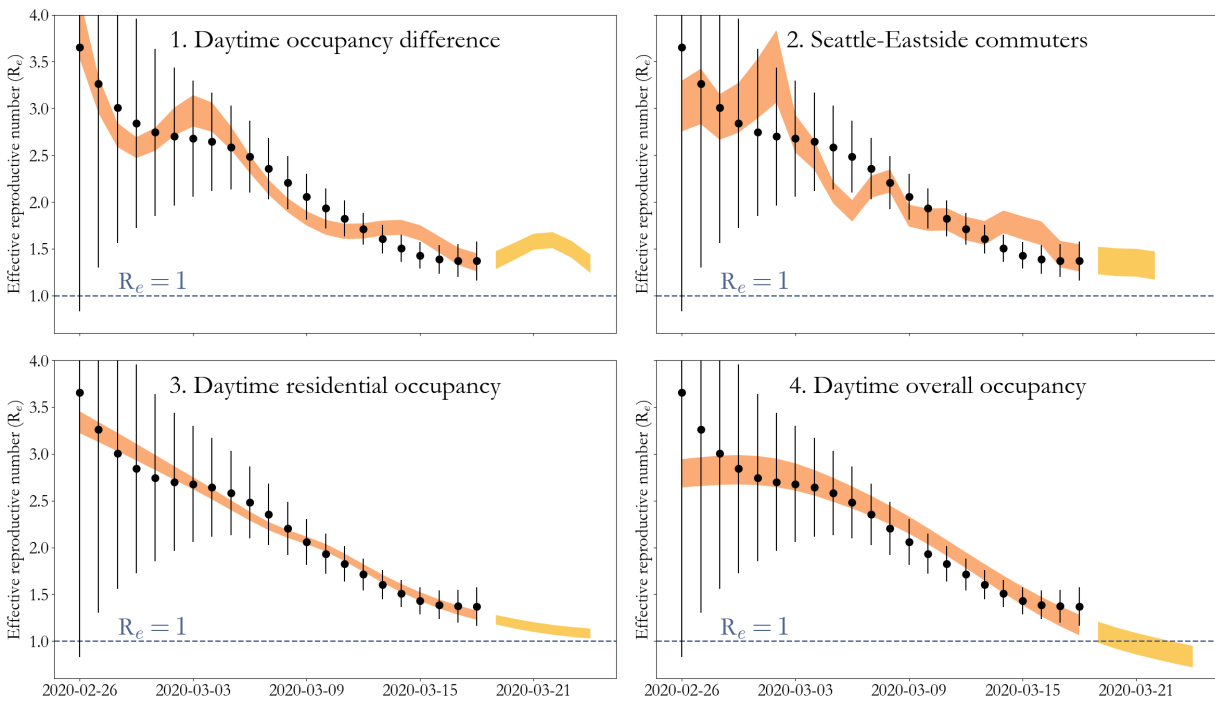
## Quantifying mobility in King County

[Facebook's Disease Prevention Maps](#), which are part of its Data for Good program, have provided aggregated mobility data to allow us to construct a measure of mobility that captures change in **population flux between day and night**. As fewer people leave home to go to work, for example, workplaces are empty both during the day and at night, and residential areas have higher occupancy during the day and night. So as social distancing intensifies, the flux in night-day population diminishes. Additionally, we created three more covariate measures that capture mobility in different ways

(discussed in the appendix in more detail). We used Facebook data through March 24th. The figure below illustrates the day-night population flux on two days, two weeks apart.



**Figure 3.** Illustrating the change in day-nite population flux on two days two weeks apart - showing large reductions in day-night population flux across the region. Anything greater than 1 represents a lot of flux (>100% of the night-time occupancy population shifts during the day). Close to zero represents no difference in day and night. We use the population-weighted average over the whole county as our main mobility covariate. Data are masked in tiles with <10 observed occupants.



**Figure 4.** Variation in epidemiological estimates of  $R_e$  for King County (black dots, 1 standard deviation error bars on all 4 panels) can be explained well with a number of mobility measures (95% CI in orange for all measures). This highlights uncertainty in our expectations today (95% CI in yellow for all 4 models). Some metrics suggest  $R_e$  increases, some suggest decreases, and others

are flat. As more case data is collected, we will better understand how best to quantify mobility in relation to COVID-19 transmission.

In Figure 4 above, we show the change in  $R_e$  over time superimposed with the regression scaled mobility covariates. All four measures of mobility yield a reasonable fit to  $R_e$ . We can then use the mobility covariates to project (i.e. “nowcast”)  $R_e$  through March 24. Depending on how we measure mobility, we see that  $R_e$  could have taken several potential paths. Due to these conflicting projections, we cannot say with confidence if  $R_e$  is above or below one today.

## Limitations

Like all modeling work, this analysis is not without significant limitations. To list a few:

- We have assumed the reporting rate is constant during the period of data evaluated.
- The fitting procedure does not incorporate mortality data.
- We have not adjusted for testing specificity (ill patients in hospital vs. general public).
- Only one mobility covariate is used at a time, and the nowcasted results are sensitive to this choice.
- Age or other heterogeneity in acquisition or transmission is not modeled.
- Mobility is only a proxy for transmission, which could occur within households as well.
- There is uncertainty in how COVID-19 was introduced into King County, and we have not explored that uncertainty here.

## Conclusions

These results show that the King County COVID-19 epidemic has slowed over the past month. We estimate that on March 1st, before behavior change, the effective reproductive number was  $2.7 \pm 0.9$ . **The reproductive number has fallen to  $1.4 \pm 0.2$  on March 18th.** These results are based on epidemiological data alone.

Mobility data improves the model by explaining variance and enabling nowcasting. Using a mobility statistic based on Facebook data, we can re-interpret the epi-only result. **The mobility-enhanced estimate of  $R_e$  on March 18th decreases to  $1.3 \pm 0.07$ , a value consistent with the case data but more certain.** However, this finding changes slightly as we try different mobility covariates.

The mobility covariates also allow us to extend the analysis beyond March 18th. However, the result is sensitive to which covariate we use for the extrapolation. Some covariates show a strong increase in mobility over the weekend (March 21 and 22), resulting in an increase in predicted COVID incidence. News media [noted](#) heightened outdoor activities and group gatherings corresponding to nice weather. **These nowcasts act as a warning that our progress is precarious, and they indicate that Governor Inslee’s Stay Home, Stay Healthy order was timely and necessary.**

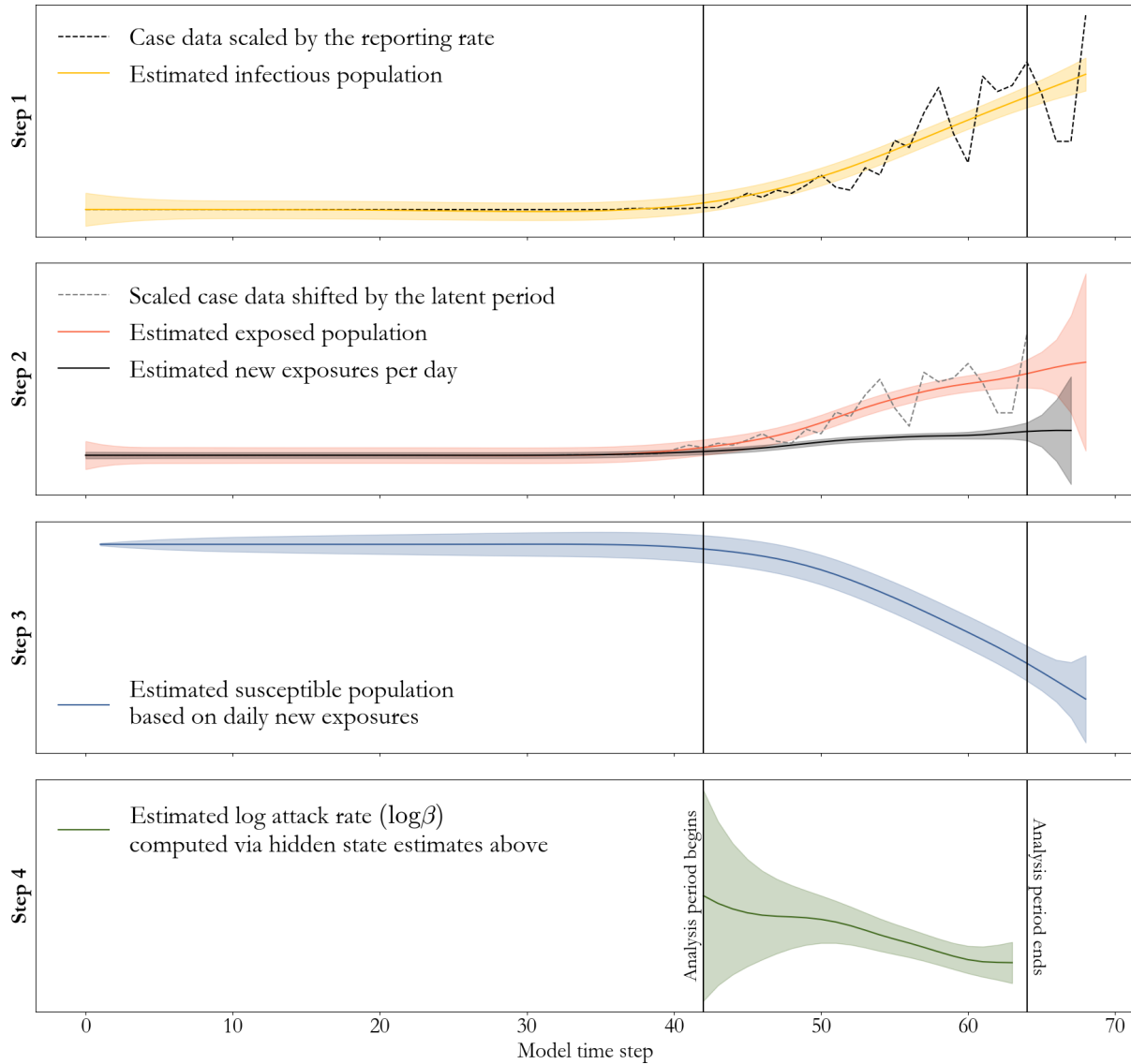
## Appendix 1: Estimating the reproductive number from case data

We use the following SEIR model:

$$\begin{aligned}S_t &= S_{t-1} - \beta(m)S_{t-1}(I_{t-1} + z_{t-1}) \varepsilon_t \\E_t &= \beta(m)S_{t-1}(I_{t-1} + z_{t-1}) \varepsilon_t + (1 - 1/D_e)E_{t-1} \\I_t &= E_{t-1}/D_e + (1 - 1/D_i)I_{t-1} \\C_t &\sim \text{Binom}(I_t, p) \\ \log(\beta(m)) &= \theta_0 + \theta_1 m\end{aligned}$$

Where  $S_t$ ,  $I_t$ , and  $E_t$  are the number of people who are susceptible, infected and exposed at time  $t$ .  $\varepsilon_t$  has a zero-mean log-normal distribution, and  $p$  is the case detection rate. 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 15th, 2020 and February 25th, 2020, corresponding to the two Washington clades on NextStrain.  $\beta(m)$  is the infection rate,  $m$  represents the movement covariate, and  $\theta_0$  and  $\theta_1$  represent the coefficients used to calculate the infection rate by regressing the movement covariate against case-based estimates.

Unknown parameters such as the infection rate  $\beta(m)$  are estimated from daily case data. As shown in Figure S1, this is a multi-step process. The critical steps are 1 and 2. In step 1, case data is scaled by a proposed reporting rate and smoothed to timescale  $D_i$  to construct an approximation of  $I_t$ . Then, in step 2, the same case data is shifted by  $D_e$ , scaled (by  $D_e/pD_i$ ), and smoothed to timescale  $D_e$  to construct an approximation of  $E_t$ . Finally, in steps 3 and 4, corresponding approximations for  $S_t$  and  $\log(\beta(t))$  are constructed using the SEIR equations. This algorithm estimates  $\log(\beta(t))$  conditional on the reporting rate, and it can therefore be used to numerically integrate out reporting rate dependence. These reporting-rate independent point estimates of  $\log(\beta(t))$  can finally be used to estimate  $\theta_0$ ,  $\theta_1$ , and the variance in transmission in a standard log-normal regression problem given covariate  $m$ .

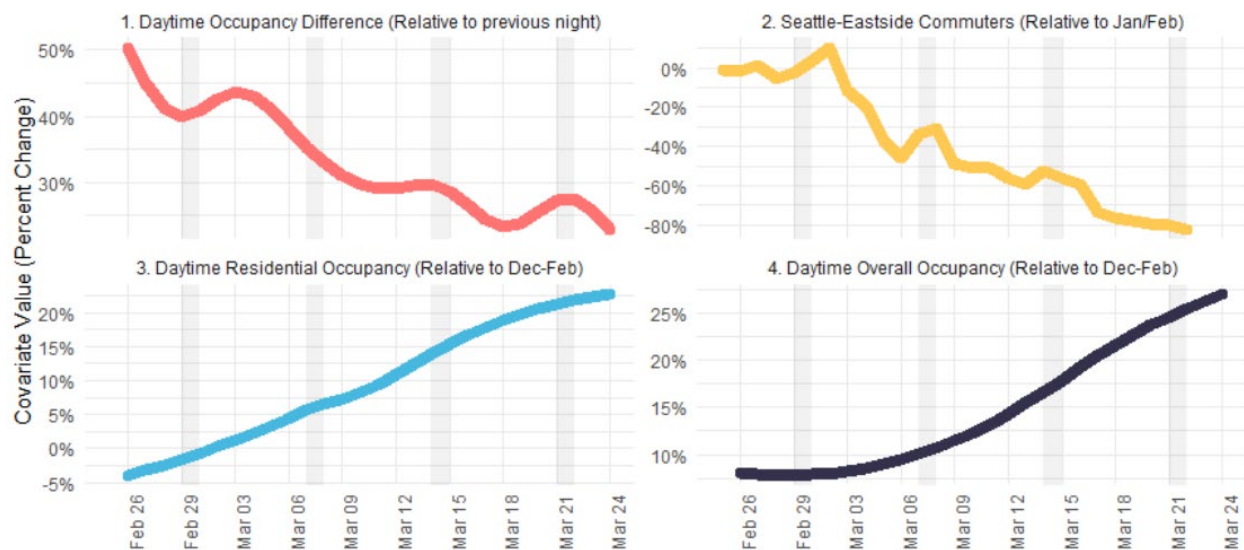


**Figure S1.** Unknown parameters such as infection rate are inferred using RAINIER in a multi-step process. **Step 1** - case data is scaled by an inferred reporting rate and smoothed to timescale  $D_i$  to obtain  $I_t$ . **Step 2** - case data is shifted by  $D_e$  scaled and smoothed to timescale  $D_e$  to obtain an estimate for  $E_t$  and calculate new exposures. **Step 3** - the susceptible population on a given day is then calculated by subtracting the new exposures from the previous day's susceptible population. **Step 4** - calculation of attack rate via hidden states.



## Appendix 2: Comparing measures of mobility

We used data from the [Facebook Data for Good Project - Disaster Maps](#) to track changes in population and mobility between regions over time. These data are collected from mobile users with location services enabled and are aggregated to coarse geographic levels as anonymous counts of users; individual users cannot be identified. Using these data, we make comparisons over a time period from February 26 through March 24. We emphasize we are not measuring reductions in social contact, but rather changes in mobility and the places where people are spending their days and evenings. Though as populations spend more time in residential areas and away from shared public and work spaces, it is likely they are coming in contact with fewer people. These datasets are described in detail [here](#), and a broader discussion of the utility of these data for COVID-19 research [here](#).



**Figure S2.** Facebook mobility data can be used to construct measures of King County mobility in multiple ways. Here we showcase 4 options, and at this point, we are unable to determine which (if any) is most related to COVID-19 transmission.

In addition to **population flux between day and night**, we created three additional mobility measures from the Facebook data and compared results. All variables fit well to transmission reduction, but showed variability in projecting beyond the data (nowcasting). Briefly:

- **Residential Occupancy (#3 in figure S2):** Percent difference in daytime occupancy of single family zoned areas of Seattle each day compared to the baseline period. This measures the extent to which people spend their days at home, compared to the 3 month period prior to February 26th. This covariate was smoothed using a spline smoother. For smoothing we excluded weekend days, because the measure is relative to a baseline period and during normal weekends people tend to stay at home more regardless of social distancing policy measures.
- **Commuters (#2 in figure S2):** Percent difference in volume of commuters between Seattle and the Eastside compared to the baseline period. This measures the percent reduction in commuters each day compared to a baseline period in the month and half preceding February 26th.

- **Daytime occupancy change (#4 in figure S2):** Percent difference (absolute) in daytime occupancy across the whole county compared to the baseline period. This is an overall measure of how different the places people spend their days are, compared to the pre-epidemic period.

This covariate was also smoothed using a generalized additive model excluding weekend days.

Finally, the **population flux between day and night (#1 in figure S2)** covariate is measured as the absolute value of the difference between day and night population occupancy, divided by nighttime population occupancy for each tile (where observed population was >10 people) covering the county. This absolute relative difference was then averaged across all tiles for each day, weighted by population occupancy. The overall average was smoothed using a spline smoother. For this one, we included weekends, as this measure was not relative to a baseline period.

For more details on our use and analysis of these datasets, see our [report](#) on characterizing mobility changes in King County.