Towards robust, real-time, high-resolution COVID-19 prevalence and incidence estimation

Niket Thakkar, Roy Burstein, and Mike Famulare
Reviewed by: Jen Schripsema
Institute for Disease Modeling, Seattle Washington, covid@idmod.org

Results as of December 13, 2020

What do we already know?

Since April, we have been using compartmental transmission modeling approaches to better understand COVID-19 epidemiology in Washington. These approaches have helped us align observed cases, hospitalizations, and mortality to support public health decision-making, estimate effective reproductive numbers, and quantify disease burden in terms of population prevalence and cumulative incidence. Generally speaking, our estimates have been limited to relatively populous regions of Washington, like King County or the eastern and western halves of the state.

What does this report add?

Heterogeneity among sub-populations within these large regions has been a prominent feature of Washington's COVID-19 epidemiology. Better quantifying and understanding COVID-19 dynamics within and between these sub-populations is therefore a critical next step for model-supported situational awareness.

This report presents progress towards that end. Specifically, we describe a new method that works in concert with our transmission model to jointly estimate weekly COVID-19 prevalence and incidence in 15 geographic sub-regions and in 5 age groups within Washington. This is the highest resolution and most comprehensive burden estimation we have published to date. Moreover, in constructing these estimates, we show that Washington's three COVID-19 waves have had distinct characters, getting more geographically widespread over time and moving from the oldest and youngest age groups into a more representative age distribution. Finally, our burden estimates put us in a position to assess test positivity as a metric for COVID-19 activity, which we show is highly correlated to underlying prevalence across Washington.

What are the implications for public health practice?

Constructing targeted and equitable interventions for a diverse population is a fundamental problem in public health, and accurately quantifying heterogeneity in disease burden is a critical part of this broader public health goal. The method presented in this report is step toward high resolution (in time and in sub-population size), robust COVID-19 burden estimation that can be used regularly to support situational awareness and decision making during this pandemic.

1 Introduction

Interpreting data as we get it has been a persistent challenge throughout the pandemic. Ideally, we want to know who currently has COVID-19 and who those people are passing their infections onto. Said differently, in epidemiological terms, we want estimates of population prevalence and corresponding effective reproductive numbers to construct a more complete picture of risk. Answering these questions in as much detail as possible helps us better deliver care to those who need it without interrupting the lives of those who don't.

Our transmission modeling approaches offer coarse answers to these questions in part because they're restricted to considering relatively large populations. This is partially a data sparsity issue since small populations have lower numbers of cases, hospitalizations, and deaths to facilitate inference, but it is also a fundamental issue of model construction. More concretely, simple transmission models generally consider populations in isolation, and as population sizes of interest get smaller, their connections to

other populations become more and more significant. Detailed transmission models that consider these connections are notoriously difficult to fit to data since we rarely observe who a person got their infection from. As a result, if our goal is real-time situational awareness, we need more robust approaches.

In this report, we present an intuitive approach to this problem that avoids assumptions about group-to-group connections. Essentially, we leverage the transmission models we've used in the past to estimate the total number of infections over time in a large region. Then, breaking that region into grouped sub-populations, we allocate infections to each group in proportion to the number of severe outcomes we've observed in that group, taking care to ensure consistency with the overall transmission model at all times. Given well-defined groups to which severe infections can be assigned, we find that this approach can jointly estimate disease burden and uncertainty in seconds on a laptop.

We demonstrate the utility of this approach by estimating weekly COVID-19 prevalence and incidence in 15 spatial regions within Washington — the most comprehensive burden estimation we have published to date — and in 5 state-wide age groups. With these estimates, we show that each of Washington's three COVID-19 waves has had a distinct character and has become less geographically localized over time. The geographic estimates in particular give us a platform for contextualizing more accessible metrics of COVID-19 activity. To that end, we then show that test positivity and testing volume alone can be used to estimate prevalence heuristically with reasonable accuracy in all 15 geographic regions. This learning provides a bridge to better understand the meaning of readily available statistics when complex modeling is not available.

Taken as a whole, this report is primarily a methodological step forward, presenting a statistical inference approach that works in concert with our transmission models to increase detail in our situational awareness. Going forward, it puts us in a position to ask scientific and epidemiological questions that we plan to pursue in subsequent work.

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 COVID-19 testing, hospitalization, and mortality data collected by the Washington Department of Health (WA DoH) through the Washington Disease Reporting System (WDRS), compiled for this report on December 13. Testing data is aggregated into time series by specimen collection date and mortality data is aggregated by the date of death. To hedge against reporting delays, we use data up to December 4 for transmission model fitting and through December 6 for weekly sub-population estimates.

- Hospital admissions are used a proxy for severe infections, and we aggregate hospital admissions by admission date. Furthermore, 1054 of Washington's COVID-19 deaths (less than 10% of total hospital admissions distributed relatively uniformly over time) were not recorded as hospital admissions in the WDRS. Those are added by specimen collection date to the hospital admission time series throughout this report.

- We fit a compartmental transmission model to the data for the whole state, using the method described in detail in our previous technical report. Key assumptions from that report are applicable here as well. One critical change worth highlighting is that we are now using the age-dependent infection-fatality-ratio estimates published by the CDC instead of the more dated estimates based on data from China which we were using previously.

- We assume that treatment effects have improved outcomes in hospitals and lowered the overall infection-fatality-ratio since March by $\sim 30\%$ as of November. This effect size comes from a separate survival analysis described here.

- Infections in the model are distributed across Washington into groups we chose based on agricultural divisions in eastern Washington and natural resource divisions in western Washington. This is an arbitrary choice, not informed by the epidemiological data in any way. That said, as we continue to learn, the method we use in this report can help us assess the value of different sub-state groupings.

- The sub-state modeling we present requires spatial information, which we have for 99% of cases, hospitalizations, and deaths but only 89% of negative tests. Since our sub-state and age-structured models are fit to hospitalizations, missing negatives do not affect outputs. However, in computing

other measures of COVID-19 activity, like test positivity, our estimates are biased upwards — a test positivity measure reported as 4% may be closer to $4 \times 0.89 = 3.5\%$ if complete data were available. We are currently exploring methods to correct this bias more systematically, but as it is, our estimates of total test volume are in-line with the WA DoH estimates. Dashboards from individual counties may have more complete local information, leading to relatively lower estimates of test positivity.

- Our approach assumes that COVID-19 prevalence varies continuously in time, an assumption that necessarily breaks at high enough spatial and temporal resolution, where it is possible to have zero infections and thus non-continuous dynamics between importations. While we can articulate the problem at extremes (consider for example attempting to estimate hourly prevalence in a single household), we are not yet able to quantify our resolution limitations in general. In sensitivity tests of our approach, this issue was particularly apparent in small groups with sudden spikes in cases in young people (e.g. at a university) and no associated severe outcomes. To avoid this issue, we've been conservative about the number of groups across which burden is allocated, and we've tested the consistency of our estimates with observed data not used for fitting. That said, we will continue to work towards a more quantitative understanding of this issue.

3 Modeling approach

We fit a COVID-specific transmission model to daily testing, hospitalization, and mortality data at the state level. 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 5 days during which infected people are not yet capable of transmission. The choice of a 5-day latent period implicitly assumes that people become infectious on average roughly 1 day before the typical 6-day asymptomatic period ends. After the latent period, we assume that those exposed to COVID-19 are infectious for about 4 days. Note that this represents a change from our previous report, where we assumed that the latent and infectious periods were 4 and 8 days respectively. This change was motivated by tests of model fitting discussed in Appendix A.

- In the model, COVID-19 is introduced to Washington by an unknown number of infectious individuals on January 15 and February 1. On all other days, we assume that local transmission within Washington is the dominant infection route.

We use a multi-step approach to fit the state-level transmission model, described in detail previously. Once fitted, the model gives us estimates of Washington's infectious population every day and the probability of hospitalization as a function of age. These two estimates are then consumed by our inferential approach. First, using the method described in Appendix C of our previous report, group-specific weekly probabilities of hospitalization are estimated using each group's weekly age distribution of positive COVID-19 tests combined with each group's census age distribution and the state-wide estimate of the IHR by age. Then, we estimate the weekly distribution of COVID-19 burden that best explains observed hospital admissions in each group, ensuring that the distribution sums to 1 at all times. This distribution is used to allocate active infections to each group every week to estimate group prevalence and is used to allocate exposures which are summed to estimate cumulative incidence. For more details, see Appendix B below.

4 Estimating overall COVID-19 prevalence in Washington state

Estimates from the state-level transmission model are shown in Figure 1. In the top three panels, fits (95% interval shaded) to daily positive tests (left), hospital admissions (middle), and deaths (right) show that the model captures state-level trends in essentially all the data (black dots) we have. One of the model's key outputs is the total number of people in Washington actively infected with COVID-19 every day, so-called population prevalence. In the lower panel, the model estimate of underlying population prevalence consistent with the data in Washington shows three distinct waves in the spring, summer, and fall. Alarmingly, we find that recent population prevalence is likely higher than it's ever been and continuing to trend upwards in early December. While the rate of growth slowed in mid-November, concordant with restrictions imposed by the state on November 16, we estimate that transmission rates increased over the Thanksgiving holiday leading to more pronounced rises in prevalence recently.

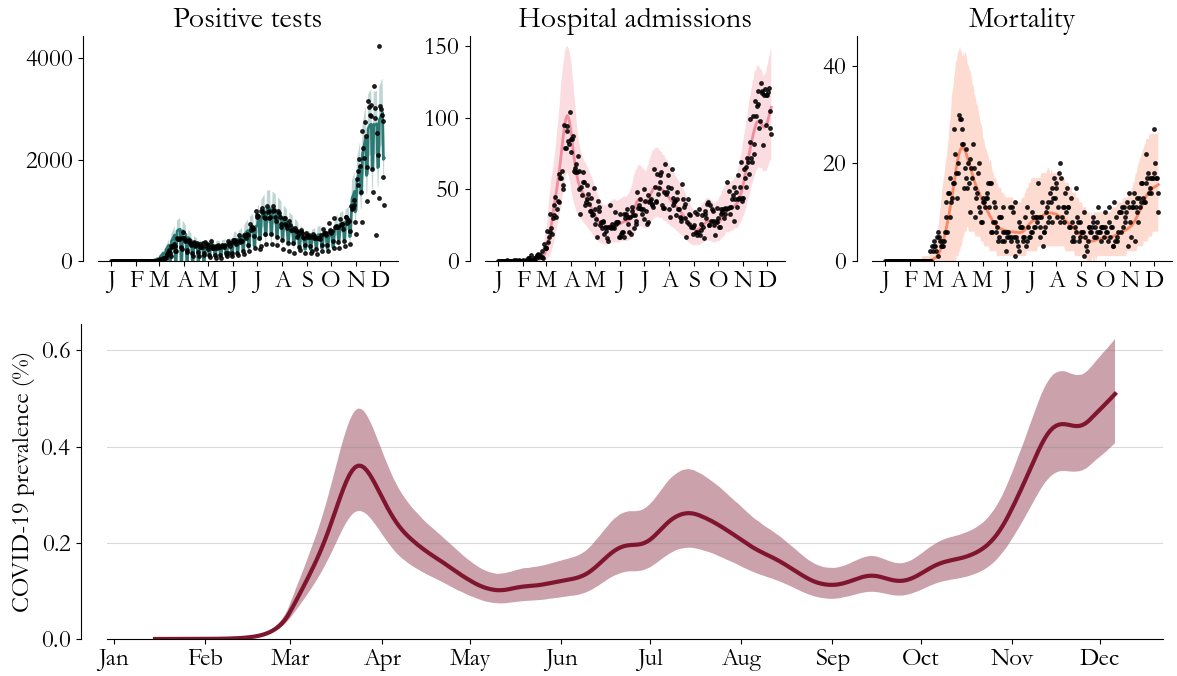

Figure 1: Transmission model for Washington. (Top panels) Using the approach described in our previous report, we construct a COVID-19 transmission model (curves, 95% interval shaded) consistent with observed trends in daily positives tests (black dots, left), hospital admissions (black dots, middle), and deaths (black dots, right). (Bottom) In constructing this model, we estimate the daily number of infections responsible for the observed trends. In Washington, this exposes three distinct waves in the spring, summer, and fall.

In fitting the model, we take as input the CDC estimates of the infection-fatality-ratio (IFR) and published estimates of the infection-hospitalization-ratio (IHR) by age. The latter distribution is scaled by an overall factor to account for local hospital admission criteria; this factor is determined by best fit to the hospitalization time series. As a result, in fitting the transmission model, we also calculate Washington-specific estimates of the proportion of infections that become severe enough to be hospitalized. For the model in Figure 1, we estimate that the all-time, all-age average IHR is 2.73% (95% interval: 2.17% to 3.29%) with an age distribution directly proportional to our published input by assumption. In other words, we estimate that in Washington nearly 1 in 36 people infected with COVID-19 get admitted to the hospital.

5 Distributing infections to capture geographic trends

The model in Figure 1 reconciles data at the state level and offers a relatively coarse picture of the epidemic in Washington. However, given an estimate of the total number of infected people every day and the probability of severe infection as a function of age, we can increase detail in our estimates by allocating burden to best explain observed severe infections in sub-populations.

The approach we have developed is very general, and it could be applied to any type of grouping for which we can assign observations (see Section 8 for an example). For now, we split Washington into 15 regions, shown in Figure 2, with color associated with western Washington (purples), the Puget Sound area (blues), central Washington (oranges), and eastern Washington (yellows). Critically for public health practice, group size need not be evenly distributed, and in this case population in the groups ranges from roughly 100,000 in the Columbia Basin to 1.3 million in north King County.

As described in the appendix, burden allocation is framed as a statistical inference problem, leveraging an assumption that sub-population prevalence varies continuously in time but avoiding assumptions regarding the connections between sub-populations. We visually verify that the model (colors) captures observed trends in weekly hospital admissions (black dots) in Figure 3. Across groups, despite the wide range in group populations, the model captures all 15 observed trends accurately (overall $R^2 = 0.92$). For additional tests of the model's fit, see Appendix C.

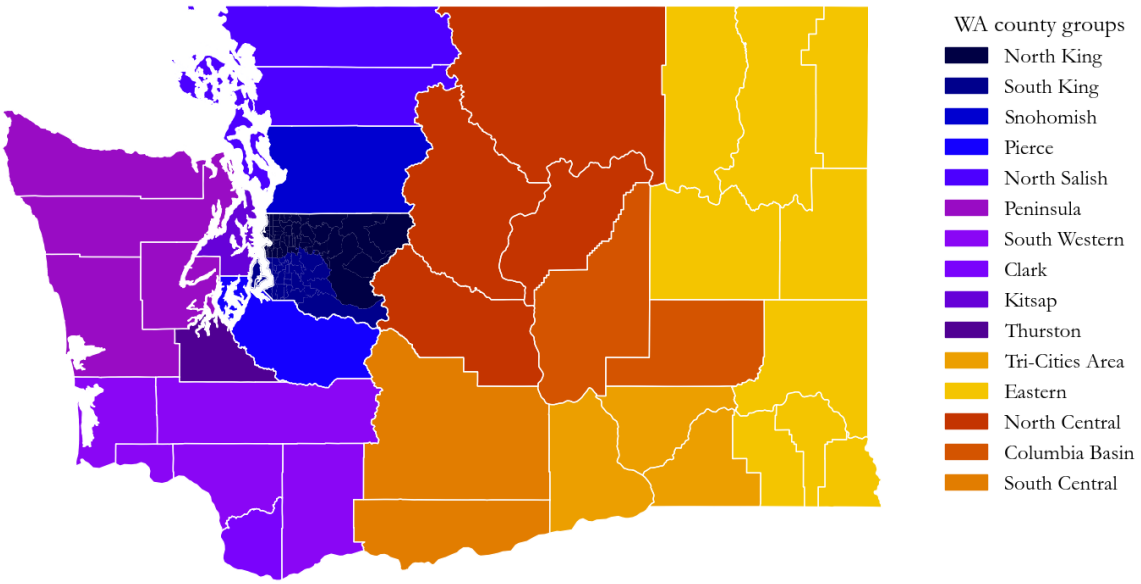

Figure 2: Grouped counties in Washington. We created 15 sub-state groups of counties, loosely divided by color (blue for the Puget Sound area, purple for the west, orange for the center, and yellow for the east). Groups were chosen based on agricultural divisions in eastern and central Washington and based on natural resource divisions for western Washington.

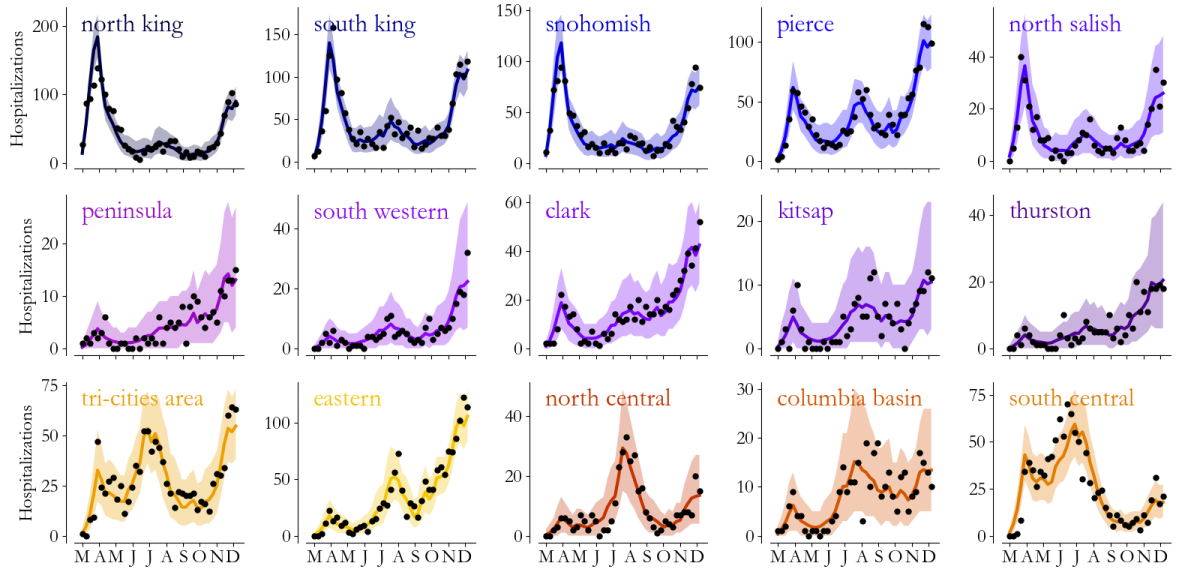

Figure 3: Fits to observed weekly hospital admissions. In all 15 groups, the model estimates (curves, 95% CI shaded) capture the weekly trend in hospital admissions (black dots). Visualized together, the models expose both the variety of trends in Washington and that inference appropriately handles disparity in group populations.

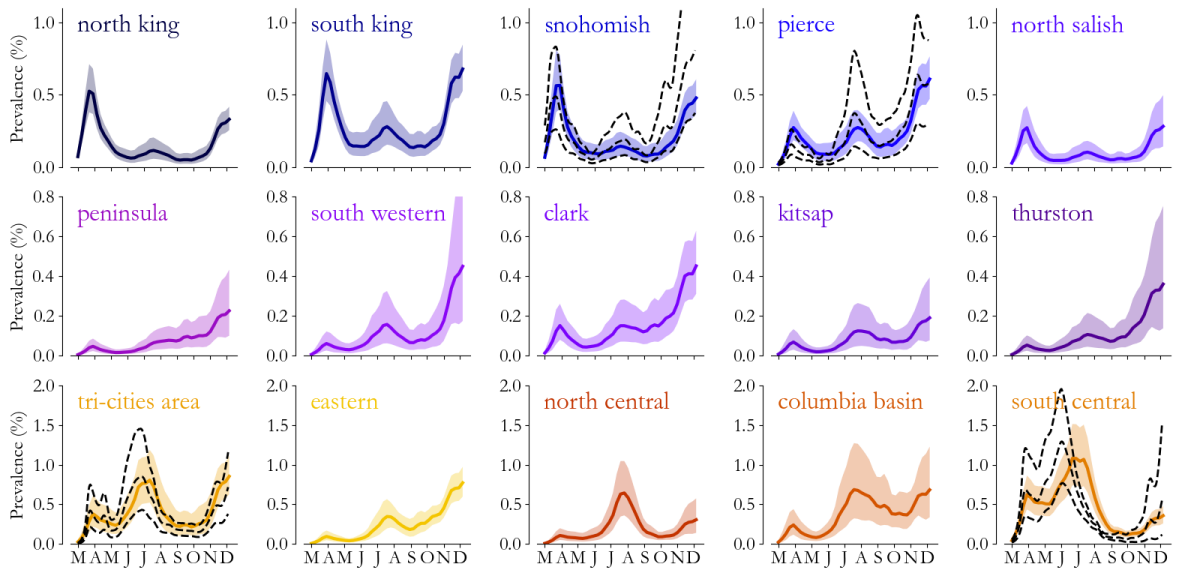

Figure 4: Prevalence estimates across Washington. Overall, different pockets of Washington have had significantly different prevalence time courses. Most recently, all regions show dramatic transmission increases throughout October. To lend confidence to these estimates, we've overlaid prevalence estimates from our transmission model (black dashed lines, mean and 95% interval) in regions where we've published estimates in the past.

6 Washington's COVID waves have so far come in three flavors

Estimating the weekly distribution of COVID-19 infections consistent with observed hospital admissions gives us each group's population prevalence automatically. This is shown in Figure 4.

Overall, Figure 4's most striking feature is the diversity of estimated trends. We find that Washington's first COVID-19 wave was disproportionately concentrated in King and Snohomish counties, as we would have expected based on where COVID-19 was first found in the state. The second wave was largely concentrated in Washington's agricultural belt, with south central, north central, the Columbia basin, and the Tri-Cities area hardest hit during cherry season. Finally, in the fall, rises have been prominent in every region, making this the first wave widely distributed across the state as a whole.

A number of more specific features stand out. In particular:

- Within King County, all 3 waves have disproportionately affected south King County relative to north King County.
- North Salish, the peninsula, and Kitsap County have generally been underrepresented throughout the pandemic.
- Clark County, Thurston County, and the Eastern region were fortunate to avoid a large spring-time wave, but have otherwise had continued transmission from the summer throughout the fall. Summertime mitigation efforts were least successful in these regions.
- In Snohomish, Pierce, the Tri-Cities area, and south central region, areas where we've previously published estimates based on our transmission model, we've overlaid those estimates on Figure 4 (black dashed lines). Generally speaking, our new approach recapitulates our previous estimates but with increased confidence. It is noteworthy, however, that in south central region, there is a pronounced shift in the summer peak, powered by pooled information across the state.
- Finally, early December trends are highly uncertain across the state. While in some regions (north King, eastern, and south central), mid-November deceleration and resulting slowed growth into December is likely, in other regions (Clark, Thurston, south western, and north central), the data remains consistent with rapidly growing prevalence throughout the fall.

Accumulating exposures over time allows us to calculate the percent of each group's population no longer fully susceptible to COVID-19, so-called cumulative incidence. This is plotted in Figure 5. Overall, we see that per person, eastern and central Washington (bottom row) have been hardest hit by COVID-19,

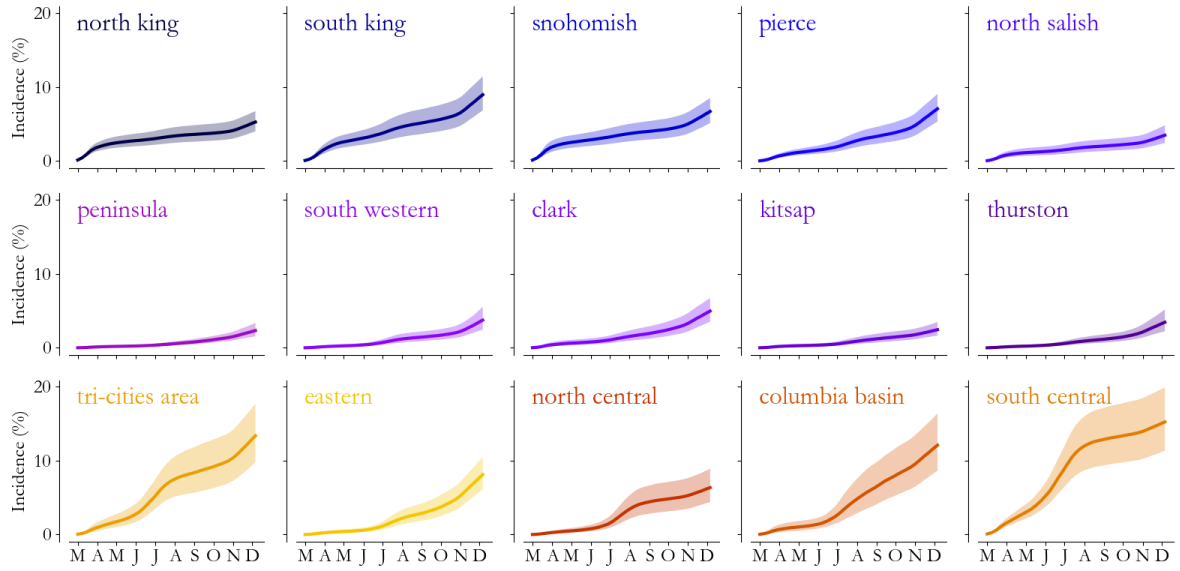

Figure 5: Accumulated, group-level exposures to estimate the percent of the population at some point infected with COVID-19. Eastern Washington regions (bottom row) have been harder hit per person than western Washington regions. Overall, more than 80% of the population in each group is still fully susceptible to COVID-19.

with more than 10% of the population infected at some point in the south central region and the Tri-Cities area. That said, taken as a whole, all groups remain at least 80% susceptible to COVID-19 and most more than 90% susceptible. As a result, burden levels could still grow significantly if left unchecked, and mitigation efforts are as important as ever.

7 Estimates of underlying burden clarify signal in testing data

The analysis described above uses a transmission-model-based methodology which leverages a number of data streams (cases, hospitalizations, deaths, age) to infer underlying COVID-19 burden. Given these estimates in Washington, we're in a position to turn this question around and ask how informative more readily available data is with regards to estimating prevalence. This question is highly relevant as simple metrics based on the aggregated results of population testing are commonly used to inform decision-making where model-informed burden estimates are not readily available. As the premier example, test positivity (confirmed cases/total tests) has been widely used as a proxy for current burden. For example, New York City recently used the '3% rule' as a decision threshold for closing schools. Furthermore, researchers are developing simple, algorithmic test-positivity-based heuristics for approximating underlying burden.

Overall, we find that weekly test positivity explains 77% of the variation in our weekly prevalence estimates, meaning that test positivity alone is reasonably informative of underlying burden in Washington's geographic regions. That said, the relationship between test positivity and underlying burden is modified both by the volume of testing done and the reporting rate (cases/true infections). Reporting rate is an unobserved quantity, but is highly correlated with testing volume, which is typically observable. Along those lines, adding information on test volume improves the proportion of variation explained to 86%. Figure 6 shows that the linear relationship between positivity and prevalence is different depending on the level of testing volume in the population. For this reason, direct interpretation of test positivity as proxy for burden is possible but serves as a coarse and potentially biased approximation. We plan to investigate these findings, and their generalizability, in a forthcoming report.

8 The same statistical principles can be applied to age groups

So far we have focused on estimating COVID-19 burden in spatial groups; however, the method we have developed is purposefully agnostic to the connections between groups, and that allows us to apply the same principles more widely. Primarily as a methodological demonstration, we reapply the method to age groups in this section.

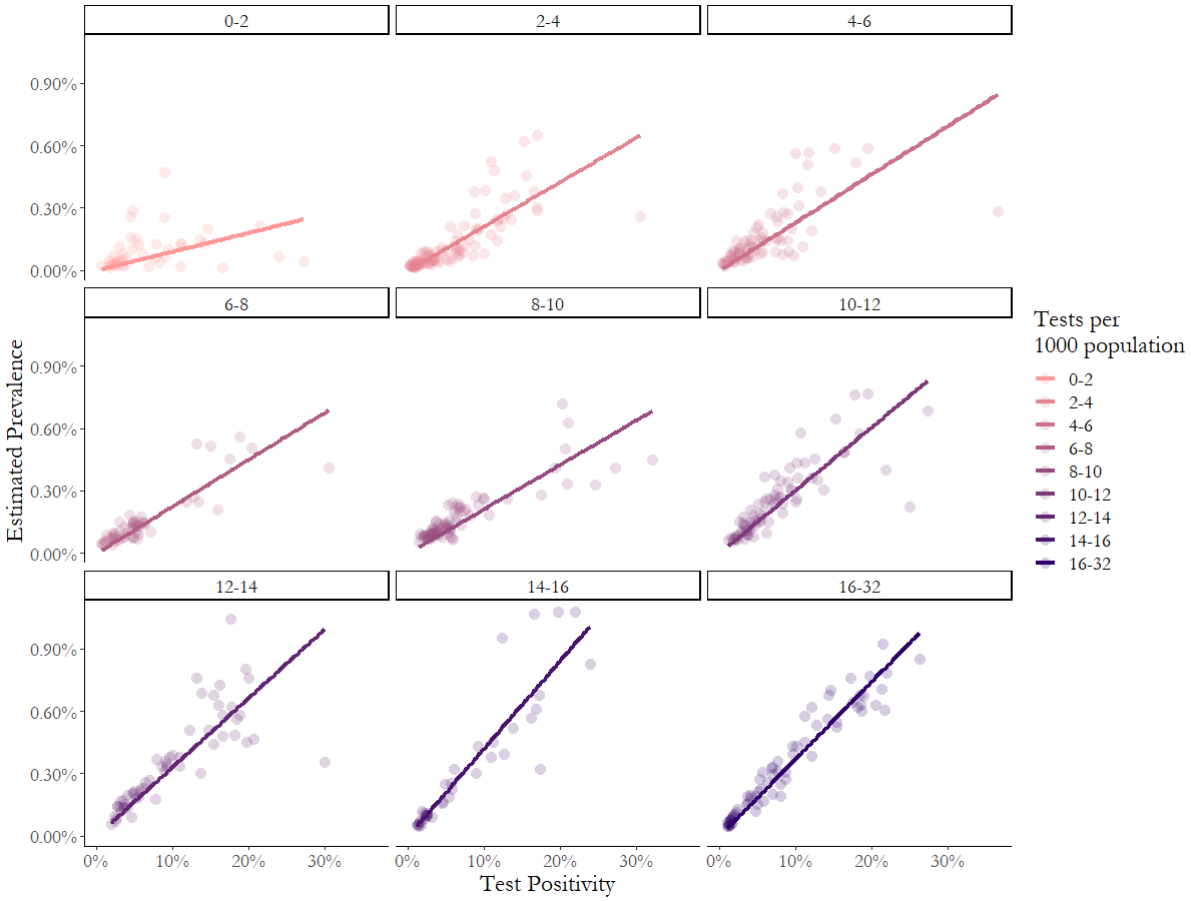

Figure 6: Estimated weekly prevalence (with uncertainty suppressed for clarity), as a function of test positivity, stratified by testing volume (weekly tests per 1000 population). Each point is a group-week. Test positivity is a reliable linear predictor of prevalence, but the relationship varies with testing volume.

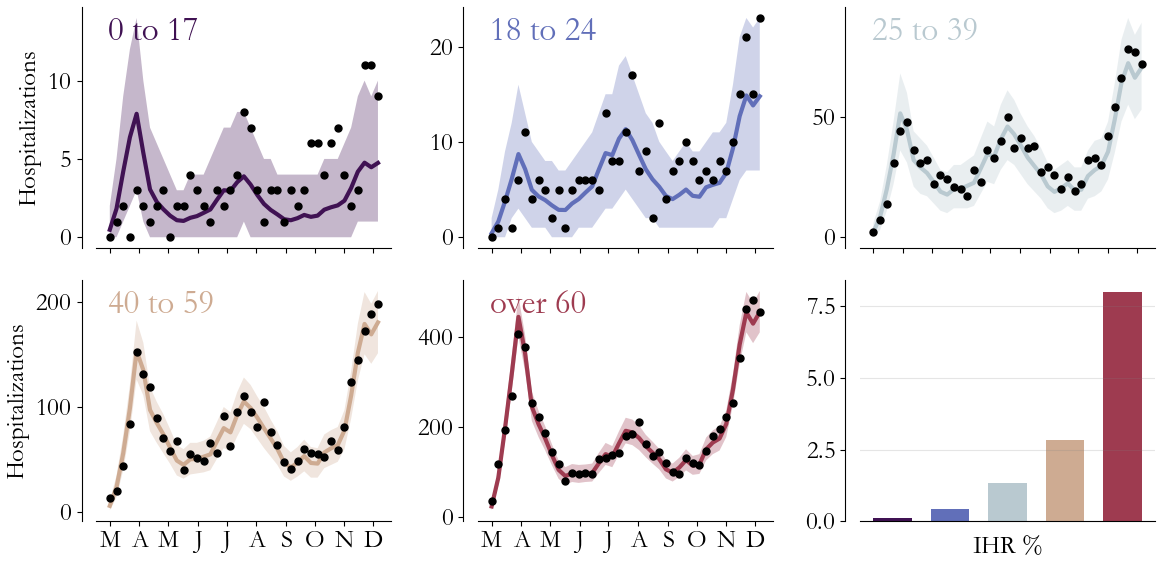

Figure 7: Applying our approach to age groups in Washington. Allocating COVID-19 burden from the state-level model to these 5 age groups, accounting for the differences in the probability of hospital admission given infection (bars), leads to age-structured inferences (colors, 95% CI shaded) that capture observed, weekly hospitalizations (black dots) since March with an all-age $R^2 = 0.99$.

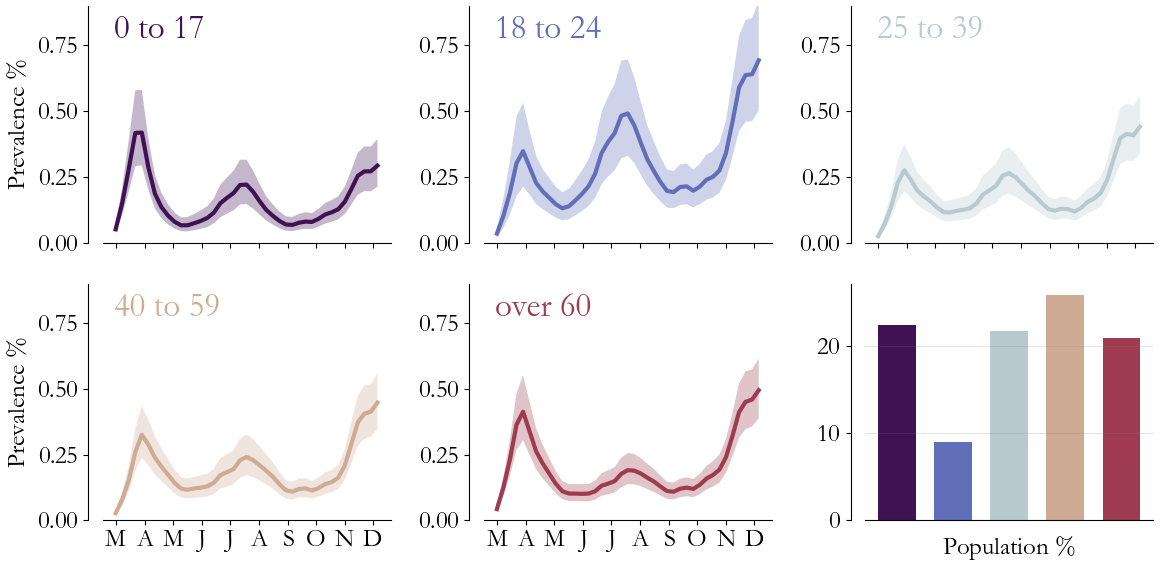

Figure 8: Age-structured prevalence estimates in Washington. Compared to the geographic estimates, there is significantly less heterogeneity across age groups, with each group's estimated prevalence time course showing 3 distinct waves. This suggests that once COVID-19 is established in a particular location, transmission is widely age-distributed.

As before, starting with the state-level prevalence and IHR age distribution estimated in section 4, we allocate burden to five age bins in proportion to observed hospital admissions in Figure 7 accounting for the age dependence of COVID-19 severity. The model captures weekly trends in admissions ($R^2 = 0.99$) despite the non-uniform bin sizes and the dramatic differences in bin-specific IHR.

Weekly, age-distributed COVID-19 prevalence is shown in Figure 8. Overall, unlike the spatial groupings, we see significantly less diversity in trends, and each age group has a pronounced three-wave structure. As a result, comparison of the two sub-state groupings suggests that COVID-19 has taken time to spread geographically while being widely age distributed once established.

That said, the age-distributed prevalence estimates highlight a number of epidemiological features:

- Since the summer wave, 18- to 24-year-olds have had the highest prevalence of the age groups considered. While summertime mitigation efforts were successful in this group, the autumn rise has been the fastest of the age groups.
- While the first wave disproportionately affected children under 18, this age group has been under-represented since. The high prevalence in the first wave is particularly striking since it isn't clearly reflected by observed hospitalizations at that time (see Figure 7), suggesting that 2 to 3 severe infections in children per week may have been misdiagnosed early on. The model arrives at this estimate based on observed hospitalizations in other age-groups and the enforced consistency with the state-level model.
- Prevalence in 25- to 39-year-olds and in 40- to 59-year-olds has been largely consistent throughout the pandemic.
- Adults over 60 were dramatically under-represented in the summer wave, but prevalence has unfortunately rebounded in this group in the fall.

9 Conclusions

The prevalence models in this report enrich our understanding of Washington's COVID-19 epidemic by providing quantitative insight into something we can't see — prevalence and incidence that describe how infections are distributed in our community over time — based on information we can see like cases, hospitalizations, deaths, and the scientific literature. The technical advance in this report provides a flexible workflow to sharpen transmission modeling's resolution, and the workflow facilitates asking new

scientific questions. This report's key insight is that it's useful to pool evidence across large regions to get precise estimates of how many people are infected and then distribute those estimates across sub-regions in proportion to the most comparable and reliable local data to infer heterogeneity (in this case, hospitalization). On spatial inference, estimates in low population density regions are now statistically stable and precision is improved in the higher density regions we've modeled before. And because the method is one of accounting, not population structure, it is easy to ask new questions, as demonstrated by the analysis of prevalence by age which indicates for example that children must've been frequently infected in the first wave despite low rates of testing in that age group.

Prevalence estimates are also useful for putting other questions in context. For example, we showed that the fraction of tests positive for COVID-19 is generally proportional to the prevalence, but that the exact numerical relationship varies with testing volume. From this, we can conclude that in Washington, it is reasonable to look at trends in test positivity as evidence of changes in transmission, and that increases in test positivity do not necessarily imply that case ascertainment rates are falling. That said, one should be careful about comparing test positivity across populations with differences in testing rates and policies — the trend in positivity is meaningful but not necessarily the level of positivity.

Our estimates of cumulative incidence show considerable variation across the state, but they are nowhere higher than at most 20%. Thus, nearly a year after COVID-19 first arrived in our state, the majority of the population in every geographic region studied here is still susceptible to infection. This estimate reflects successes in our ability to control COVID-19 through non-pharmaceutical interventions, reveals the still-huge need for vaccination to eventually end this pandemic, and quantifies the catastrophic risk COVID-19 continues to pose if we do not maintain control until then.

The heterogeneity among sub-groups is also informative about variations in control. All spatial regions studied show periods of increasing and declining prevalence, but the details vary in important ways. The south central region of the state, including Yakima County, experienced the largest per-capita outbreak through the harvest season but had squashed it to nearer zero than most other places since March, before starting to lose control again. In contrast, the Olympic peninsula region has never really maintained control of COVID transmission, but has also not yet had an explosive region-wide outbreak. This likely reflects the regions advantages of low population density and reduced mixing.

We are continuing to refine and better understand our COVID-19 models, and this report's primary goal is to describe new and updated methods we have developed to make more detailed epidemiological inferences. Going forward, higher resolution prevalence estimation will be a routine part of our situational awareness support for Washington's Department of Health. Moreover, these detailed estimates will put us in a position to better understand societal connections relevant to COVID-19 transmission and associated ways for public health to better address Washington's needs.

A Estimating state-level COVID-19 prevalence

We use the following SEIR model:

$$\begin{aligned}
S_t &= S_{t-1} - \beta_t S_{t-1} \left(I_{t-1} + z_{t-1}\right) \varepsilon_t \\
E_t &= \beta_t S_{t-1} \left(I_{t-1} + z_{t-1}\right) \varepsilon_t + \left(1 - 1/D_E\right) E_{t-1} \\
I_t &= E_{t-1}/D_E + \left(1 - 1/D_I\right) I_{t-1} \\
C_t &\sim \text{Binomial}\left\{I_t, p_t\right\} \\
H_{t+d_H} &\sim \text{Binomial}\left\{\beta_t S_{t-1} I_{t-1} \varepsilon_t, \alpha\text{IHR}_t\right\} \\
F_{t+d_F} &\sim \text{Binomial}\left\{\beta_t S_{t-1} I_{t-1} \varepsilon_t, \text{IFR}_t\right\}
\end{aligned} \tag{1}$$

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, β_t is daily COVID-19 transmission rate, D_E and D_I are the latent and infectious duration respectively, and z_t is non-zero only on January 15 and February 1. This model has three observation processes. Specifically, C_t are daily observed COVID-19 cases with daily case detection rate, p_t, assumed to have step-wise structure in time with independent values in prespecified reporting periods and a correction for relaxed testing on weekends. Meanwhile, hospitalizations, H_t, and fatalities, F_t, are computed with time-varying IHR and IFR based on published age distributions of the probability of severe outcomes, accounting for the time from exposure to outcome, d_H and d_F respectively.

As described in detail in our previous technical report, we fit the model to data hierarchically. First, $\ln(\beta_t)$ and σ_t^2 are estimated using an epidemiological curve based on WDRS hospitalizations and cases.

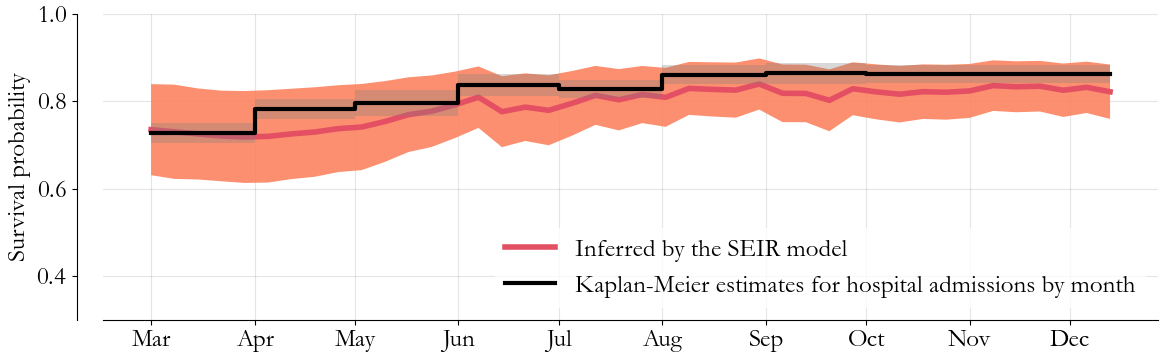

Figure 9: COVID-19 survival given hospital admission. Subsetting the WDRS hospitalization data by admission month and computing Kaplan-Meier survival curves gives us estimates of the raw, monthly survival probability (black, 95% CI in grey) given hospital admission. In our transmission model, we can compute a comparable estimate using our inferred IHR and IFR over time (i.e. the survival probability is 1−IFR/IHR). With treatment effects incorporated and accounting for variation in the age distribution of the infected population using the method of our previous report, we capture the overall trend in survival while simultaneously fitting the daily timeseries data at the state level (see Figure 1).

This curve is 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 specified by minimizing the L2 discrepancy between the model's infectious population estimate and observed daily COVID-19 positives in each reporting period with an adjustment for weekends. Similarly, the overall scale factor for the IHR by age distribution, α, which models location-specific hospital admission practices, is determined by best fit between the model's average hospitalization estimate and the observed time series.

In this report, we introduce four adjustments to this overall approach:

1. Unlike our previous reports, where we used age-distributed IFR estimates based on data from Asia, we've switched to using the CDC's published estimates. These estimates are used in the same way as before, accounting for shifts in the age distribution of cases via the method described in Appendix C of our previous report.

2. We additionally use a Cox proportional hazard model to quantify the effects that advances in treatment have had on COVID-19 mortality rates over time. This approach is described in detail in our associated report. Briefly, we apply the survival model to hospital admissions to estimate monthly hazard ratios accounting for sex, comorbidities, and age. Overall, we find that treatment advances have lowered the average IFR by $\sim 30\%$ relative to March as of November. Monthly hazard ratios are applied directly to the age-adjusted IFR over time to incorporate this effect, and we find that the transmission model captures the observed survival probability accurately (see Figure 9) while simultaneously fitting the time series data.

3. In our previous report, we described an inference approach blending hospitalization data at long time scales with COVID-19 positives at short time scales to create the epidemiological curve upon which model fitting is based. We continue to use this method; however, the input hospitalization data is weighted by the age-adjusted IHR over time relative to the expected IHR based on the census age distribution. This better accounts for variation in the relative probability of observing a hospitalization on a given day.

4. Finally, we have changed D_E from 4 to 5 days and D_I from 8 to 4 days. This change was based on the grid-search of model fits to observed hospitalizations shown in Figure 10. Overall, we find compelling evidence that models with comparable D_E and D_I outperform models with pronounced duration differences. Weighing the results of this brute-force approach with published viral load studies and the CDC estimate that the mean time from exposure to symptom onset is 6 days, we selected the best performing model with $D_E = 5$ days (to account for 1 day of pre-symptomatic transmission, as before). We tested these new durations in other settings, namely in King County, Snohomish County, Pierce County, and the Tri-Cities and south central regions (see Figure 2).

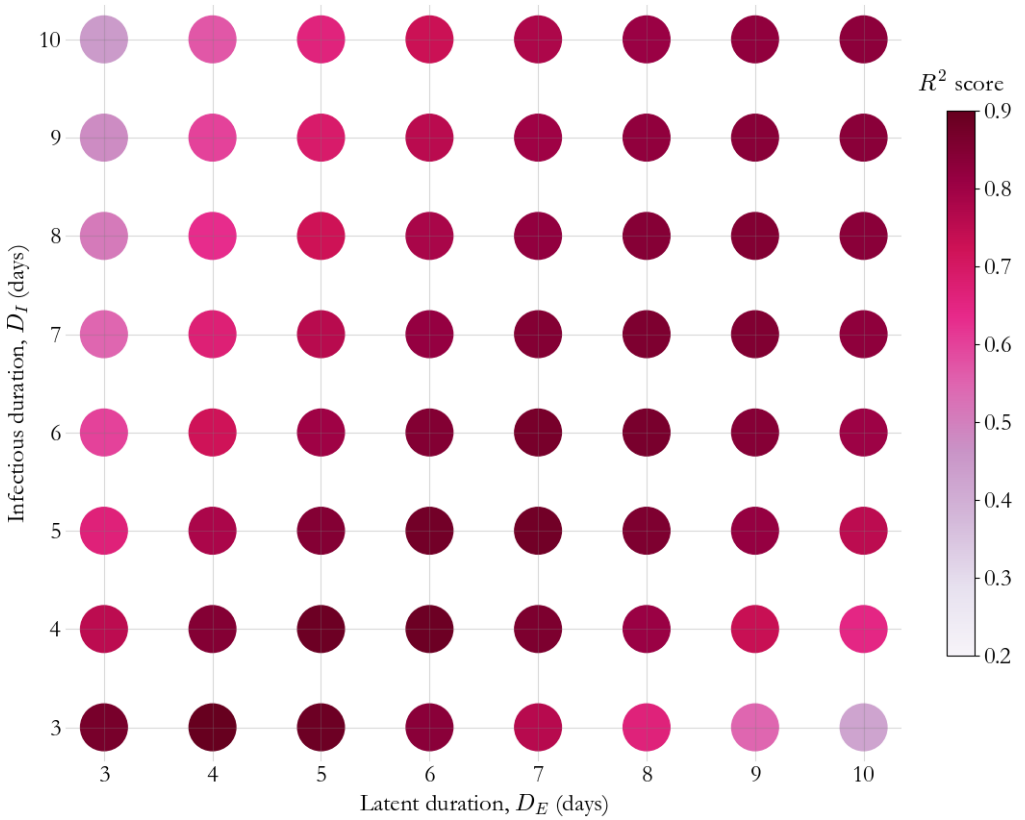

Figure 10: Testing the transmission model's response to variation in D_E and D_I. State-level transmission models were fit conditional on different D_E and D_I pairs (dots) to compare goodness of fit to the observed hospitalization times series as a function of the latent and infectious duration (colors). Overall, we find that models with $D_E \sim D_I$ outperform models with large duration differences.

Overall, we found that the $D_E = 5$ and $D_I = 4$ model was generally the highest performing choice with D_E consistent with the literature (i.e. fixed at 5 days).

B An intuitive approach for allocating the transmission model's infections to sub-populations

The fitted transmission model in the previous section gives us an overall estimate of the infected population in a region that we assume is large enough to sustain uninterrupted community transmission and therefore be considered in isolation. While the fitted transmission model captures observed trends in cases, hospitalizations, and mortality at the region level, it averages over significant heterogeneity among sub-populations.

This report presents a simple way to estimate COVID-19 burden in sub-populations given a fitted model for the full population. Succinctly, we allocate infections to groups in proportion to observed severe outcomes over time, taking care to ensure that the total number of infections in the population is consistent with the overall transmission model. Here, we describe this process in mathematical detail.

We are given data h_t^g, the number of hospitalizations observed in group g at time t, and we assume that both the total number of infections at time t, I_t, and the probability of severe outcomes for group g at time t, p_t^g are known. We model the distribution of infections across groups at time t as $s(\vec{\theta}_t)$, where $\vec{\theta}_t$ is a vector of random walks for each group at time t and $s(\cdot)$ is the soft-max function ensuring that the distribution sums to 1 at all times. Our goal is to estimate random walks θ_t^g for groups $g = 1, ..., G$ and times $t = 1, ..., T$.

To get started, consider $\vec{\theta}_t$ at a specific time. The posterior distribution

$$\begin{aligned} p(\vec{\theta}_t|\vec{h}_t, I_t, \vec{p}_t) &\propto p(\vec{h}_t|I_t, \vec{\theta}_t, \vec{p}_t)p(\vec{\theta}_t|I_t, \vec{p}_t) \\ &= \int dI_t^1 ... dI_t^G p(\vec{h}_t|\vec{I}_t, \vec{p}_t)p(\vec{I}_t|\vec{\theta}_t, I_t)p(\vec{\theta}_t|I_t, \vec{p}_t), \end{aligned} \tag{2}$$

where in the second line we've introduced group specific infections, I_t^g, and made some sensible conditional independence assumptions (that h_t^g is independent of θ_t^g given I_t^g for example). Note that all vector valued quantities represent the relevant collections over groups and have length G. As we mentioned above, we model

$$p(\vec{I}_t|\vec{\theta}_t, I_t) = \prod_g \delta(I_t^g - s(\vec{\theta}_t)^g I_t),$$

and we further assume that

$$p(h_t^g|I_t^g, p_t^g) = \text{Binom}\left\{h_t^g|I_t^g, p_t^g\right\}.$$

Thus, given I_t^g and p_t^g, h_t^g is conditionally independent of observed hospitalizations at all other times and in all other groups. Collecting terms over time and taking the integrals gives

$$p(\mathbf{\Theta}|\mathbf{H}, I_t, \mathbf{P}) = p(\mathbf{\Theta}) \prod_{g,t} \text{Binom}\left\{h_t^g|s(\vec{\theta}_t)^g I_t, p_t^g\right\} \tag{3}$$

where bold, capital letters represent $T \times G$ matrices of the associated quantities. The prior $p(\mathbf{\Theta})$ is the product of the priors over time in Eq. 2. As mentioned, we model $\mathbf{\Theta}$ as a collection of G random walks in time, implying that

$$p(\mathbf{\Theta}) \propto \exp\left\{-\frac{1}{2}\text{Tr}[\mathbf{\Theta}^{\mathrm{T}}\Lambda\mathbf{\Theta}]\right\},$$

where $\text{Tr}[\mathbf{\Theta}^{\mathrm{T}}\Lambda\mathbf{\Theta}]$ is the finite-difference approximation to the total variation in each group's random walk, and Λ is scaled such that the expected value of each random walk's total variation is that of a sine wave with period τ. Taken together with Eq. 3, we have a fully specified posterior distribution for $\mathbf{\Theta}$ which can be used for Bayesian inference.

There are a few additional practical considerations worth highlighting for completeness.

- First, we use the transmission model's expected value for I_t to then compute $\mathbf{\Theta}^*$, the parameters that maximize Eq. 3, and an associated covariance matrix. This gives us a Gaussian approximation to the posterior distribution which we can sample to more completely propagate the transmission model's uncertainty in I_t to individual estimates at the group level.

- Second, the soft-max function is degenerate in the sense that $s(\vec{\theta}_t) = s(\vec{\theta}_t + c)$ where c is an arbitrary constant added to $\vec{\theta}_t$ element-wise. To stabilize the optimization, we set $\theta_t^G = 0$ for all t so that other group's random walks can be interpreted relative to group G and the posterior distribution has a unique maximum. In testing the approach, we found that results were insensitive to the group chosen as the baseline.

- Third, throughout the report, we set $\tau = 6$ weeks. This was chosen by testing values between 2 and 10 weeks and selecting the value with the highest posterior probability and sharpest posterior mode. In testing however, we found that the results were generally similar across reasonable choices for τ.

- Finally, optimization is carried out using scipy's implementation of BFGS to minimize the negative log posterior. This process is assisted by analytically computing the gradient of $\ln p(\mathbf{\Theta}|\mathbf{H}, I_t, \mathbf{P})$ with respect to $\mathbf{\Theta}$, which we found dramatically increases speed and stability relative to optimization using finite-difference approximations to the gradient. To be transparent, with $L \equiv \ln p(\mathbf{\Theta}|\mathbf{H}, I_t, \mathbf{P})$, we calculate

$$\begin{aligned} \frac{\partial L}{\partial \theta_t^g} =& (-2\Lambda\mathbf{\Theta})_t^g + \left[\psi(s(\vec{\theta}_t)^g I_t + 1) - \psi(s(\vec{\theta}_t)^g I_t - h_t^g + 1) + \ln(1 - p_t^g)\right] s(\vec{\theta}_t)^g I_t \\ &- s(\vec{\theta}_t)^g I_t \sum_{g'} \left[\psi(s(\vec{\theta}_t)^{g'} I_t + 1) - \psi(s(\vec{\theta}_t)^{g'} I_t - h_t^{g'} + 1) + \ln(1 - p_t^{g'})\right] s(\vec{\theta}_t)^{g'}, \end{aligned}$$

  for all t and g where ψ is the digamma function.

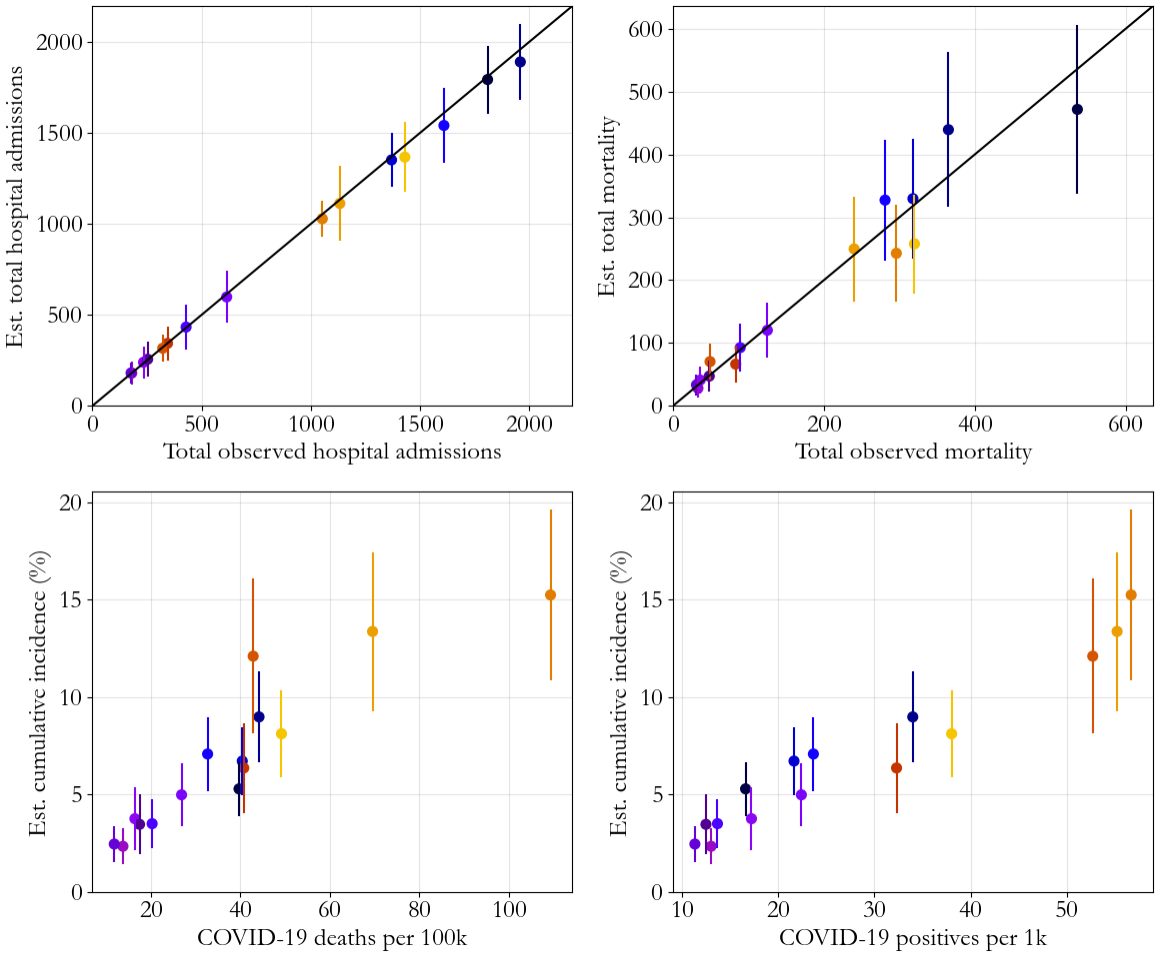

Figure 11: Four high-level tests of model performance. (Top row) The model captures cumulative, observed hospitalizations and mortality across geographic regions (colors as per Fig. 2, 2 standard deviation error bars), despite being fit to hospitalizations alone. (Bottom row) Model-based cumulative incidence estimates are highly correlated to mortality per 100k (Pearson correlation 0.91) and positives per 1k (Pearson correlation 0.97), two additional correlates of COVID-19 burden.

One final note: A critical feature of this approach is that it is agnostic to how groups are defined. For example, while we concentrate on spatial collections above, at no point in the above analysis do we leverage spatial correlation (which could be included in Λ if we needed). This facilitates application of the same method to age groups in section 8, and we could in principle apply this approach to groups defined in any way as long as we have the associated data on hospital admissions. While we do not explore that idea at length here, we plan to do so in the future.

C Additional tests of our geographic estimates

In the main text, we show that the model estimates capture observed trends in weekly hospital admissions (Figure 3). Here we further test the approach's cumulative estimates.

Figure 11 shows four such tests. In the top two panels, model-based estimates of total hospitalizations and total deaths (colors as per Figure 2, 2 standard deviation error bars) are compared to observations, and in both cases, the estimates capture the distribution across the state. In the bottom two panels, model-based cumulative incidence estimates are compared to data-based correlates of COVID-19 burden, namely deaths per 100k population and positive tests per 1k population. Our cumulative incidence estimates are highly correlated to these measures that were not used for model fitting, increasing our confidence in our estimates overall.