COVID-19: What's New for April 17, 2020

Main updates on IHME COVID-19 predictions since April 13, 2020

Predicting when the current epidemic phase will end: initial estimates on when we could shift to containment strategies in the US

Social distancing policies, which can range from restrictions on large gatherings to strict stay-at-home orders and closure of all non-essential businesses, have been used as a mechanism to substantially reduce the spread – and thus the immediate toll – of COVID-19. We are now entering the phase of the epidemic when government officials are considering when certain types of distancing policies may be eased. With today's release, we provide initial estimates that can serve as an input to such considerations in the US.

These estimates assume that when social distancing policies will be eased, such actions will occur in conjunction with public health containment strategies. Such measures include widespread testing, contact tracing, and isolation of new cases to minimize the risk of resurgence while maintaining at least some social distancing policies to reduce the risk of large-scale transmission (e.g., bans on mass gatherings).

We summarize the data and model updates behind this estimation below; the results can be explored online: https://covid19.healthdata.org

We view these estimates as initial projections. We plan to update them in the coming days and weeks as we, collectively, better understand how COVID-19 is affecting locations throughout the world and how the world in turn is responding to the epidemic. Our aim is to produce the best possible predictions given what we know today – and to continually improve these estimates to support further gains against COVID-19 tomorrow.

More data, improved models, better estimates

Major development 1: Expanding categories of social distancing policies and accounting for mobility

Since our last release, our team has expanded the types of social distancing policies included in our COVID-19 death model. Further, our social distancing covariate now accounts for the effect of different types of measures on population-level movement patterns over time, drawing from mobility data collected from cell phone and online platforms. In combination, these updates now more accurately reflect the effect of social distancing policies enacted – and importantly, how people are changing their behaviors in response to these measures.

With today's release, we now include six categories of social distancing policies. The first four have been reported on and included in our models since our March 30 release, and the additional two policies – mass gathering restrictions and initial business closures – are now types explicitly accounted for in our work (bolded below). Summarized definitions are as follows, with our detailed definitions provided on our FAQ page:

- Stay-at-home order: "Individuals are ordered to stay at home and interactions between individuals from different households prohibited."
- Educational facilities closed: "Education has transitioned to remote-learning options."
- Non-essential services closed: "All stated non-essential services are clearly defined and ordered to cease operations."
- Travel severely limited: "Municipal-level travel restrictions are in place and movement is strictly limited to essential travel. Public transit options are reduced and international borders are closed."
- **Mass gathering restrictions: "Initial implementation of restrictions on people gathering are mandated."** This is a social distancing policy newly included in the IHME COVID-19 estimation effort as of April 17. Recommendations wherein individuals are urged to not congregate en masse without a formal mandate are not included.
- **Initial business closure: "The first phase of non-essential businesses are ordered to cease on-premise operations."** This is a social distancing policy newly included in the IHME COVID-19 estimation effort as of April 17. Examples include the mandated closure of specific groups of business types (e.g., recreational and fitness facilities, entertainment venues, personal care services) without calling for a more sweeping closure of non-essential businesses.

As of today, we include population-level mobility data to better reflect how populations are changing their behavior once distancing mandates are implemented. This is an important update for IHME COVID-19 estimation: we now inform our model predictions by including information on how populations are responding to different distancing measures.

As mentioned in our April 10 update, IHME team members have been processing and testing models using mobility data from Descartes Labs, SafeGraph, and Google (via their COVID-19 Community Mobility Reports) in relation to each type of distancing policy implemented. All three mobility datasets are available for the US, while the Google mobility dataset is the only one that includes European countries.

Each dataset is analyzed separately to estimate the percentage reduction in mobility associated with each of our six social distancing measures. We then use these estimates as weights to construct a single covariate for predicting the epidemic peak in each location. We produce three distinct versions of the social distancing covariate (i.e., one based on data from Descartes Lab, one from SafeGraph, and one from Google). We run the COVID-19 death model for each of the three versions of the social distancing covariate and then ensemble them into a single set of predictions.

Major development 2: Quantifying where COVID-19 daily deaths have peaked and how long their peaks last

Since our last release on April 13, our research team members have adopted a statistical process for identifying which locations have experienced peak daily COVID-19 deaths and how long these peaks last on average. More details on these methods are in a forthcoming paper.

Why is this important? The duration of COVID-19 epidemic peaks appears to vary widely across locations (e.g., some experience a sharp peak, while others seem to experience a flatter peak where similar peak daily COVID-19 deaths occur for several days before the epidemic curve declines); this is critical to capture. Further, this development is important because the timing and duration of daily death peaks informs our models on the relationship between peak and social distancing covariates.

Detecting COVID-19 daily death peaks. We use the following process to identify which locations have experienced peak daily deaths:

- When a given location reaches its peak, the natural log of the daily death curve should either essentially reach or pass where the curve's tangent line is horizontal. We fit a spline to the natural log of the daily death rate and identify the peak where the slope of the spline is 0.
- These selected locations are then also reviewed by our team to determine the final set of locations with peak COVID-19 daily deaths.

Estimating peak duration. We have sought to capture variations in COVID-19 death peak duration (especially among places with longer peaks) by conducting the following steps:

- Extending the approach for detecting daily death peaks, we fit a concave spline for each location on the natural log of daily deaths. This spline allows for a more flexible functional form that then can capture the flat shape at the peak's top (again, in natural log space). It also prevents noisy data from being overly influential through its concavity assumption.
- After we fit these splines, we compute derivatives of the curve for the natural log of daily death space. We estimate peak duration based on the difference between observations where the relative derivative (i.e., fraction of the maximum observed derivative) crosses the threshold on each side of the epidemic peak.

Major development 3: Updating the COVID-19 death model

Since our last release on April 13, we have implemented two major updates to our COVID-19 death model:

- **First, we now use the recent trend in reported cases to inform predictions of projected deaths.** To achieve this, we compute the ratio of cumulative COVID-19 deaths up to the most recent time period to the cumulative cases reported up to 8 days prior. We use 8 days because this is the median duration between cases and deaths in currently available data. We then use this estimated case-fatality ratio (CFR) for each location to estimate the number of deaths that would likely occur in the coming 8 days based on the number of cases reported in the 8 prior days. These predicted deaths are applied as a leading indicator for our death projections for the coming 8 days. We believe this improves our model predictions as it allows our death model to be informed by recent trends in confirmed cases reported. In other words, if the number of confirmed cases has been increasing in the past few days in a particular location, we want our death model to predict that the number of deaths will also likely increase 8 days later.
- **Second, we now use an extension to our previous model, which can be thought of as a "mixture of Gaussians extension."** Some locations still have relatively few daily death data

points, and we continue to use the model we have been using until now. For locations that have more observations (18 or more data points), we now use an extended death model. To best achieve the balance in ensuring models fit the data, where available (i.e., model flexibility), with the need for model generalizability in projecting potential epidemic trajectories, we use a semi-parametric modeling approach building off our current model (as described below):

- For each location, we fit the long-term death model to inform its shape parameters of alpha and p, with a strong prior on the beta coefficient through the social distancing covariate. The outputs of this model comprise the "atom" used next.
- After fitting the atom, we consider staggered atoms across time: the first atom is centered at the identified epidemic peak, and then the remaining 12 atoms are evenly distributed between 12 days in the past (i.e., before the peak) and 12 days into the future.
- We then fit the death data based on a mixture of these atoms, obtaining non-negative weights that best represent data patterns. The resulting model generalizes the primary model used to date while better capturing signals in the death data – especially when the atom mixtures are asymmetric and exhibit flatter regions.

Overall, this updated approach is well suited to capture the highly variable COVID-19 epidemic trends across locations. This extended model also can be used for generating COVID-19 predictions in the same manner as our original "single atom" case. Note that we estimate uncertainty in the same approach as well, on the basis out-of-sample predictive validity.

Major development 4: Predicting the time when COVID-19 infections fall below 1 prevalent case per million

We are now using estimates from our COVID-19 death models and estimates of infection fatality ratio (IFR) to produce estimates of COVID-19 incidence and prevalence. We then use these estimates to identify the date after which the number of COVID-19 infections is predicted to fall below 1 per 1,000,000 people in each location. This date can be viewed as the earliest time that locations could consider easing social distancing restrictions – conditional on containment measures already in place to avert potential resurgence of the virus. Such necessary containment efforts include extensive testing, robust contact tracing and isolation of new cases, and maintaining restrictions on mass gatherings of people.

Below we summarize the different components that constitute this major development for our COVID-19 prediction platform:

- **Threshold of 1 infection per 1 million:** We have chosen this threshold – 1 prevalent COVID-19 infection per 1,000,000 population – to represent a conservative estimate of the number of infections each location could reasonably try to identify via active case detection and contact tracing in order to prevent COVID-19 resurgence. We arrived at this threshold based on these assumptions: (1) for testing, contact tracing, and isolation to be effective, each infected individual must infect less than one other individual; and (2) currently many locations have limited capacity to test and contact trace, so a lower prevalence of infection would increase the odds of success.

- **Estimating IFR:** Prior to today's release, we were drawing from age-specific data on the deaths among cases detected in South Korea, which had a very extensive testing program to inform infection fatality ratio (IFR) estimates. Over the last week, we have been able to review data from all locations where extensive COVID-19 testing has occurred and where information on detected infections and age-specific deaths have also been reported. Based on this review, we then conducted a random-effects meta-regression on all location data and used these estimates to inform age-specific IFR. Notably, the lowest IFRs were recorded among passengers on the Diamond Princess cruise ship. Drawing from the random-effects meta-regression, we used these pooled estimates and random effect from the location with the lowest IFR (i.e., the Diamond Princess).

- **Estimating COVID-19 incident and prevalent infections:** We use these IFR estimates to derive COVID-19 infections based on death estimates. We draw from our analysis of peak daily deaths and timing of these peaks, alongside our time series of reduced mobility patterns, to estimate 16-20 days from COVID-19 infection to death. We then estimate daily incidence of COVID-19 infections per day.

  To convert daily estimates of incident cases to prevalent infections, we estimated two categories of duration: "Group A," or people who sought testing/treatment or self-isolated; and "Group B," people who did not seek treatment or self-isolate. We estimate the proportion of each population in Groups A and B based on testing rates in each location. Testing rates were approximated by taking the ratio of confirmed cases to estimated infections 8 days ago. For "Group A" duration estimates, we used the model parameter of 8 days from hospitalization to death among COVID-19 patients who died. This implied infection prevalence of 9 to 13 days outside of hospitalization, or a total of 16 to 20 days minus the 8 days in-hospital. For "Group B" – individuals who did not seek treatment or self-isolate – minimal data currently exist on infectiousness duration in the absence of care-seeking or behavioral change. Subsequently, we use the 16- to 20-day duration for this population group in terms of prevalent infections.

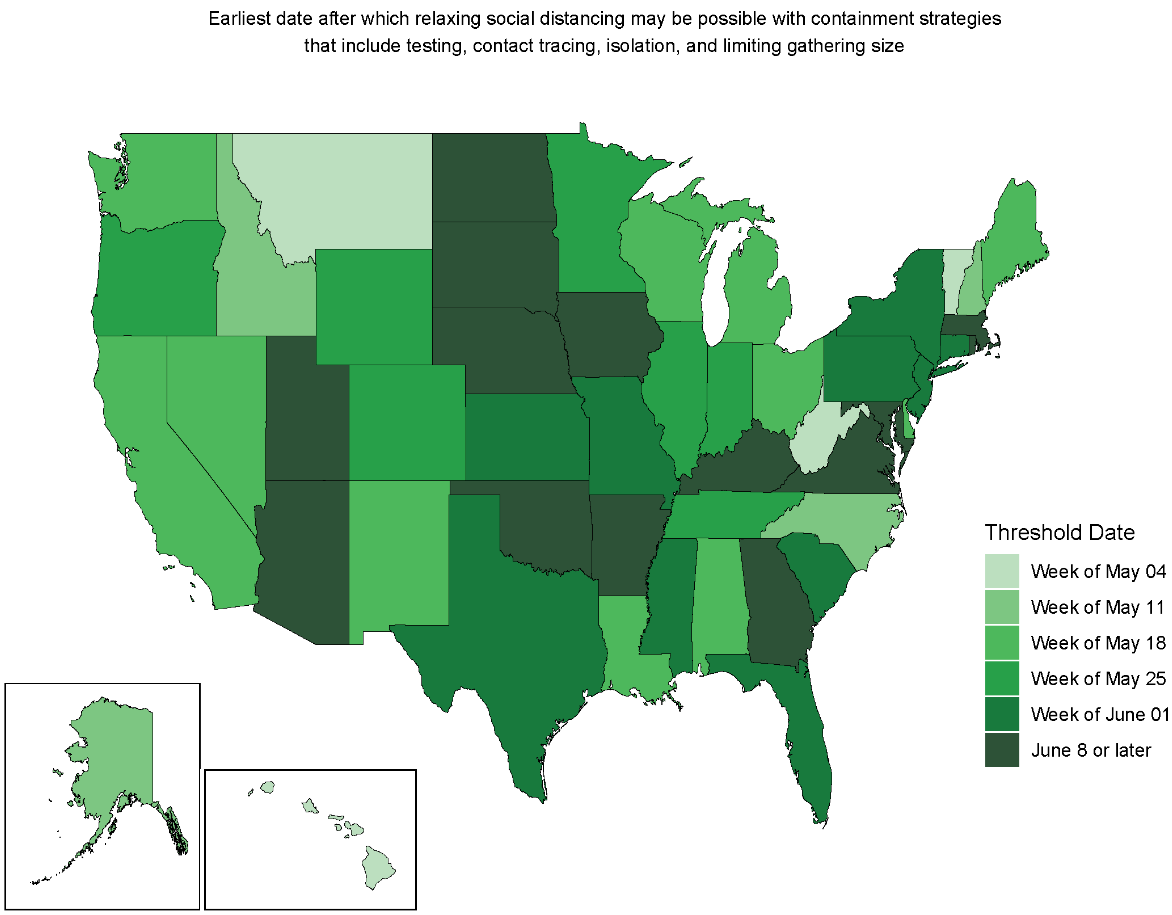
  All together, we model the distribution of prevalent infections per day. To identify the date after which easing social distancing may be possible – below the 1 prevalent infection per million threshold – we took the first date after the upper bound of the 95% uncertainty interval for all-age prevalence predictions was lower than 1 infection per 1,000,000.

Key findings from today's release (April 17, 2020)

A focus on when US states could consider easing social distancing if containment measures are in place

- Based on our initial estimates, the earliest that some states may be able to ease presently implemented distancing policies – conditional on strong containment measures – appears to be in early to mid-May (lighter greens in the map below). This means, given the current epidemic trajectories and assuming these states have instituted robust containment strategies (e.g., widely available testing, contact tracing and case-based isolation, restricting mass gatherings), states including Alaska, Hawaii, Idaho, Montana, New Hampshire, North Carolina, Vermont, and

West Virginia could potentially fall below 1 prevalent case of COVID-19 per 1,000,000 (1 million) around the first or second week of May.



- These initial projections suggest states with among the highest total COVID-19 deaths to date – including New York, New Jersey, and Connecticut – may be able to safely consider easing their currently enacted distancing policies, conditional on implementing strong containment strategies, as early as late May or early June. Other states, such as Louisiana, Michigan, and Washington, may fall below the 1 prevalent infection per 1,000,000 threshold around mid-May.

- Based on current trajectories, 15 states may need to wait until June 8 or even beyond that time before they fall below the 1 prevalent infection per 1,000,000 threshold. These projections could change as new data become available and/or different policies are implemented.

A focus on COVID-19 death predictions

United States

- **Predicted peak for daily COVID-19 deaths.** Nationally, the predicted peak for daily COVID-19 deaths appeared to be on April 15, reaching 2,481. This is a slightly higher and later peak than our last release on April 13: based on data through April 12, it seemed as if the national peak would occur around April 13 at 2,150 deaths (estimate range of 464 to 7,084). Updates in available data and corresponding model improvements are the main drivers of this small shift.

- At the state level, much more variation remains in terms of projected COVID-19 daily death peaks – as well as identifying which states may have passed their epidemic peaks in recent days. In the table below are states that either appeared to reach their peaks in daily COVID-19 deaths or are projected to potentially reach their peaks before the end of April.

  Note that here and in our visualization tool we report on predicted peak dates and corresponding predictions for COVID-19 daily deaths. Reported daily death data may be higher for earlier dates, but because of known data lags and heaping on certain days (as discussed in prior updates), we focus on predicted peak dates to avoid issues with highly noisy data.

State	Predicted peak date of daily COVID-19 deaths	Predicted daily COVID-19 deaths at peak: average projection (estimate range)
Massachusetts	April 15	264
Connecticut	April 15	197
Idaho	April 15	6
Maine	April 15	4
New York	April 16	837
New Jersey	April 16	362
Illinois	April 16	124
California	April 16	96
North Carolina	April 16	20
Alabama	April 16	15
Delaware	April 16	9
Arkansas	April 16	4
Vermont	April 16	4
West Virginia	April 16	2
Hawaii	April 17	5 (0 to 17)
Maryland	April 20	46 (4 to 219)
Rhode Island	April 25	18 (2 to 82)
Utah	April 25	7 (1 to 26)

- **Predictions for cumulative deaths.** Across the US, projected cumulative COVID-19 deaths could reach 60,308 (estimate range of 34,063 to 140,381) during the epidemic's first wave. Today's release is somewhat lower than the average US predictions for cumulative COVID-19 deaths published on April 13 (68,841, with an estimate range of 30,188 to 175,965), though the uncertainty intervals still overlap considerably.

  This change is at least partially driven by higher cumulative estimates for a subset of states, particularly in New York and New Jersey. At the same time, other states like Massachusetts, Connecticut, Georgia, and Florida now have lower average projections for their cumulative daily COVID-19 deaths through the epidemic's first wave.

Our improvements to the death model, as described above, are the primary reason why such changes have occurred since the April 10 release. By incorporating the trend in cases alongside COVID-19 deaths in our model, many locations are now predicted to have longer peaks and are taking longer to move down the epidemic curve to zero deaths. Subsequently, these places now have higher projections for cumulative COVID-19 deaths through the first wave.

State	Predictions for cumulative COVID-19 deaths through the first wave from our April 17 release (today)	Predictions from our April 13 release	Change of average values since the April 13 release*
New York	21,812 (13,623 to 42,798)	14,542 (11,008 to 23,000)	↑ 7,271 deaths
New Jersey	6,952 (4,160 to 14,367)	4,407 (2,684 to 10,033)	↑ 2,546 deaths
Michigan	3,304 (2,131 to 6,780)	2,373 (1,738 to 4,097)	↑ 930 deaths
Massachusetts	3,236 (1,289 to 9,426)	8,219 (1,680 to 25,347)	↓ 4,984 deaths
Connecticut	2,732 (1,163 to 8,601)	5,426 (1,344 to 15,397)	↓ 2,694 deaths
Illinois	2,259 (1,212 to 5,054)	1,248 (780 to 2,483)	↑ 1,010 deaths
Pennsylvania	1,707 (914 to 4,555)	2,005 (794 to 6,171)	↓ 297 deaths
Louisiana	1,685 (1,269 to 2,767)	1,141 (905 to 1,729)	↑ 544 deaths
California	1,658 (1,068 to 3,548)	1,483 (852 to 3,143)	↑ 175 deaths
Georgia	1,369 (670 to 3,828)	3,718 (1,081 to 10,693)	↓ 2,349 deaths
Florida	1,363 (775 to 3,430)	4,748 (1,250 to 13,759)	↓ 3,385 deaths
Texas	957 (472 to 2,520)	2,704 (631 to 8,552)	↓ 1,747 deaths
Virginia	763 (277 to 2,465)	1,188 (298 to 3,971)	↓ 425 deaths
Kentucky	407 (160 to 1,213)	1,067 (212 to 3,274)	↓ 660 deaths
Missouri	362 (188 to 1,027)	1,713 (420 to 5,124)	↓ 1,351 deaths

*Change estimates do not include uncertainty; they are only based on the average value. If prediction values' uncertainty intervals (the numbers reported in parentheses) overlap a lot across different releases, changes in these estimates are not considered substantively different.

Europe

- **Predicted peak for daily COVID-19 deaths.** Based on the latest available data, it appears that a number of EEA countries may have already experienced their epidemic peaks. These locations include Italy, Cyprus, and Spain in late March to early April, and several in early to mid-April (e.g., Greece, Switzerland, Denmark, France, the Netherlands, Norway, Belgium, Malta).

- In the table below are EEA countries that may have recently experienced their peaks in daily COVID-19 deaths or are projected to potentially reach their peaks before the end of April. As we note above, here and in our visualization tool, we report on predicted peak dates and corresponding predictions for COVID-19 daily deaths. Reported daily death data may be higher for earlier dates, but because of known data lags and heaping on certain days (as discussed in prior updates), we focus on predicted peak dates to avoid issues with highly noisy data.

Country	Predicted peak date of daily COVID-19 deaths	Predicted daily COVID-19 deaths at peak: average projection (estimate range)
Sweden	April 15	170
Finland	April 15	8
Germany	April 16	315
Ireland	April 16	42
Portugal	April 17	40 (4 to 156)
Romania	April 17	34 (4 to 178)
Slovakia	April 20	43 (5 to 169)
United Kingdom	April 21	1,285 (185 to 4,173)
Latvia	April 28	3 (0 to 10)

- **Predictions for cumulative deaths.** Cumulative COVID-19 death projections changed for a number of EEA countries since our April 13 release, but the direction of these changes varied across locations. The table below compares today's release – April 17 – with estimates from April 13 for countries whose average cumulative projections are 1,000 COVID-19 deaths or higher for either release date.

- While the uncertainty intervals still overlap across the two releases (as shown in parentheses), for at least some countries these changes were fairly substantial. The five countries with the highest mean predictions for cumulative COVID-19 deaths all saw cumulative death predictions increase, while others – especially the Netherlands and Sweden – saw decreases.

  As mentioned above, these changes are primarily driven by the major improvements to our death model. In addition, for some EEA countries, data updates (i.e., incorporating updated time series of COVID-19 deaths by location, some of which experienced large lags in reporting or did not fully account for deaths occurring in nursing homes and the inclusion of the number of reported cases as a leading predictor in the death model) as well as the inclusion of more social distancing policy types and accounting for changes in mobility (e.g., this was initially introduced for the Netherlands in our April 13 release) likely contributed to their changed estimates published today.

Country	Predictions for cumulative COVID-19 deaths through the first wave from our April 17 release (today)	Predictions from our April 13 release	Change of average values since the April 13 release*
United Kingdom	37,521 (17,625 to 89,385)	23,791 (14,076 to 50,820)	↑ 13,730 deaths
Italy	26,007 (23,589 to 31,056)	21,130 (20,488 to 22,311)	↑ 4,878 deaths
Spain	23,680 (20,269 to 31,608)	18,713 (17,563 to 21,386)	↑ 4,967 deaths
France	22,555 (19,455 to 29,314)	17,448 (15,497 to 22,550)	↑ 5,107 deaths
Belgium	8,039 (5,416 to 15,180)	6,041 (4,054 to 11,665)	↑ 1,998 deaths
Netherlands	6,814 (4,035 to 14,051)	15,834 (6,207 to 35,969)	↓ 9,019 deaths
Sweden	5,890 (1,965 to 16,883)	18,322 (6,780 to 44,694)	↓ 12,432 deaths
Germany	4,957 (3,697 to 9,379)	7,332 (3,364 to 19,542)	↓ 2,375 deaths
Switzerland	1,584 (1,360 to 2,121)	5,794 (1,502 to 21,267)	↓ 4,211 deaths
Denmark	683 (354 to 1,637)	1,669 (657 to 4,051)	↓ 986 deaths
Poland	646 (337 to 2,020)	2,540 (514 to 9,390)	↓ 1,894 deaths

*Change estimates do not include uncertainty; they are only based on the average value. If prediction values' uncertainty intervals (the numbers reported in parentheses) overlap a lot across different releases, changes in these estimates are not considered substantively different.

Data updates since our last release on April 13, 2020

Data and locations

- For all currently included locations, we have added reported data points on COVID-19 deaths and available information on social distancing policies through April 16 at 5:00 pm PDT.
- Currently included locations are the United States (national level) and 50 states plus the District of Columbia, as well as EEA countries and Switzerland. Three EEA countries – Germany, Italy, and Spain – also have subnational estimates at the first administrative level.

What's in the development pipeline for IHME COVID-19 predictions

Before we introduce new model components or improvements to our current analytical platform for predictions, IHME's COVID-19 development team members test these additions or changes.

Based on currently available data and model testing progress, we aim to include the following in the coming days:

- Initial COVID-19 projections for a subset of Latin American countries. Data collation and processing for a wider set of locations and countries worldwide are also in progress.

Our team continues to work on initial infectious disease compartmental models that simulate if and how groups of people move from being susceptible, exposed, infected, and recovered (often referred to as SEIR). We will provide updates on timing of these models.

A note of thanks

None of these estimation efforts is possible without the tireless data collection and collation efforts of individuals throughout the world. Your work in hospitals, health care organizations, local health departments, and state and national public health agencies, among others, is invaluable.

We thank you for your dedication to fighting the coronavirus pandemic and we appreciate your willingness to share data and collaborate with the IHME COVID-19 team.

For all COVID-19 resources at IHME, visit http://www.healthdata.org/covid.
Questions? Requests? Feedback? Please contact covid19@healthdata.org.