PNAS
www.pnas.org

Supplementary Information for

From predictions to prescriptions: A data-driven response to COVID-19

Dimitris Bertsimas, Leonard Boussioux, Ryan Cory-Wright, Arthur Delarue, Alexandre Jacquillat, Driss Lahlou Kitane, Galit Lukin, Michael Li, Luca Mingardi, Omid Nohadani, Agni Orfanoudaki, Theodore Papalexopoulos, Ivan Paskov, Jean Pauphilet, Omar Skali Lami, Bartolomeo Stellato, Hamza Tazi Bouardi, Kimberly Villalobos Carballo, and Holly Wiberg

Dimitris Bertsimas.
E-mail: dbertsim@mit.edu

This PDF file includes:

Supporting Information Text

1. Supplementary Materials on Clinical Outcomes Database

The clinical outcomes database analyzed in Section 1 aggregates reported data from different hospitals across the world. These hospitals may have different equipment and reporting standards. It was obtained through a human reading process, which is inherently imperfect. For this reason, we now list some important observations and caveats about the database, and refer the reader to our website for a complete list of the 160 papers entered into the database.*

- To minimize human error in data reporting, we have verified some key features with additional scrutiny, including mortality, ICU and hospital length of stay, key symptoms (fever, cough, short breath, fatigue, diarrhea) and common comorbidities (hypertension, diabetes).

- Across papers, subcohort divisions may follow different criteria, including: severity of disease (severe vs. mild), mortality (survivors vs. non-survivors), treatment (intubation vs. non-intubation), comorbidity (diabetic vs. non-diabetic). To retain a large enough number of studies in each category, we classify a population as "mild" if the study classifies it as "not asymptomatic" and "mild", "general", or "non-ICU" and not "severe/critical"; and we classify a population as "severe" if the study classifies it as "severe", "critical", "ICU only" or "non-survivors only".

- Studies in this dataset do not always have the same purpose. For instance, many papers from Italy seem to report data only on non-survivors. In addition, some studies focus on the disease's contagion profile, with little information on mortality, discharge, stay length. Data points from these studies may exhibit a high proportion of missing features.

- We have tried to report all lab values in consistent units. We have included a companion document (Reference Ranges) with corresponding reference ranges to facilitate analysis. There are some instances where the reported lab units seem inconsistent with the expected ranges (e.g. for D-Dimer), but we have generally reported the raw values from the source papers.

- The papers entered in the database do not consistently report confidence intervals alongside population means. For this reason, we have declined to provide confidence intervals for the quantities estimated in this part of the paper.

- We intend to continuously update the database as new papers become available. For this reason, the average values reported in this paper may change as more data becomes available.

Table S1 reports comorbidities, demographics, average lab values and average clinical outcomes among all patients, mild patients and severe patients. This expands Table 2 of the paper by including the number of reported cases in each category.

Finally, Table S2 reports statistics on treatments in different continents. The data are reported at a higher level of granularity in early studies in Asia, which hinders direct comparisons. Still, we observe significant differences in the use of hydroxychloroquine and ventilation between Asia, Europe and North America.

2. Supplementary Materials on Clinical Risk Calculators

A. Clinical Characteristics of Study Population. Our mortality cohort comprises 2,859 patients, 739 (25.8%) of whom deceased during hospitalization. The cohort includes patients from both ASST Cremona and HM Hospitals. Table S3 summarizes the clinical characteristics of the cohort, both in aggregate and broken down by survival status. The reported features are those used in the final model, that is age, gender, 3 vitals values, 13 lab results, and 4 comorbidities.

Our infection cohort comprises 3,135 patients, 1,661 (53.0%) of whom tested positive for COVID-19. This cohort only includes patients from ASST Cremona, as negative tests results were not available from HM Hospitals. Table S4 summarizes the clinical characteristics of the cohort, both in aggregate and broken down by test result. Again, the reported features are those used in the final model, that is age, gender, 4 vitals values, and 14 lab values.

B. Method Details. We construct the feature space by aggregating all clinical features for each of the cohorts. We restricted the features to those that have at most 40% of missing values in both datasets (ASST Cremona and HM Hospitals). Missing values are imputed using k-nearest neighbors imputation method (1). The mortality model consists of 22 features. The infection model has a larger feature space, since we are not limited to common features in both datasets. We restrict this model to the 20 most important features, as determined by the algorithm, to ensure usability and reduce the data entry burden on end-users.

We train models for each of the two outcomes of interest (mortality and infection), both including and excluding lab values. This results in a total of four models, referred to as "mortality with lab", "mortality without lab", "infection with lab", and "infection without lab". We use the XGBoost algorithm to train all models (2). We leverage a Bayesian optimization framework to select the best model parameters, using the mean cross-validation area under the curve (AUC) across 40 random seeds as the loss function. This technique results in a more accurate tuning compared to standard grid search, yielding better performance on the test set. We use Scikit-learn (3) to interface XGBoost and Scikit-optimize (4) to perform the hyperparameter tuning. We tune the following parameters for every model: learning rate, γ, λ, α, minimum child weight, maximum tree depth, number of estimators, and the subsample ratio of columns when constructing each tree. All remaining hyperparameters are set to their default value.

*https://www.covidanalytics.io/dataset

C. Performance Evaluation. Figure S1 reports the average receiver operating curve and precision-recall curve for each model. The results are averaged across models generated from 40 random seeds. The mortality models have higher average AUCs than the infection models, although the infection models are stronger when evaluated on precision and recall. As expected, predictive performance deteriorates when lab values are excluded. Yet, the models without lab values still achieve strong performance. In particular, the AUC of mortality model drops only moderately when lab values are excluded. Both models see a similar loss in precision/recall when lab values are excluded.

Table S5 reports threshold-based metrics, which evaluate the discriminative ability of the calculators at a fixed cutoff. We ensure a sensitivity of at least 90% to reflect the high cost of false negatives (missing a death or an infection). We select the highest corresponding threshold to maximize specificity. The results show that the accuracy of the models spans 65%–80%. The mortality calculator with lab values achieves a specificity of 76%. The infection model with lab values has lower specificity (63%), but better precision (74% vs. 56%).

Finally, Figure S2 displays calibration plots, showing the true event rates as a function of the average predicted probabilities. The x-axis bins the population by average predicted risk, and the y-axis plots the true event rate (percentage of deaths or infection). All four risk calculators are well calibrated across subgroups, as the fits are close to the 45-degree line. The bottom plot shows the distribution of predicted risk values from the models. For the mortality calculators, the mean predicted values fall below 10% for most samples, whereas the infection calculators distribute the risk more evenly across the cohort. This reflects the fact that mortality is less prevalent than infection.

3. Supplementary Materials on DELPHI-pred and DELPHI-presc

A. Formulation of DELPHI-pred.

A.1. General Formulation. The DELPHI model separates people into 11 possible states:

- **Susceptible (S)**: People who have not been infected.
- **Exposed (E)**: People currently infected, but not contagious and within the incubation period.
- **Infected (I)**: People currently infected and contagious.
- **Undetected (U_R) & (U_D)**: People infected and self-quarantined due to the effects of the disease, but not confirmed due to lack of testing. Some of these people recover (U_R) and some die (U_D).
- **Detected, Hospitalized (DH_R) & (DH_D)**: People who are infected, confirmed, and hospitalized. Some of these people recover (DH_R) and some die (DH_D).
- **Detected, Quarantine (DQ_R) & (DQ_D)**: People who are infected, confirmed, and home-quarantined rather than hospitalized. Some of these people recover (DQ_R) and some die (DQ_D).
- **Recovered (R)**: People who have recovered from the disease (and immune).
- **Deceased (D)**: People who have deceased from the disease.

In addition to main functional states, we introduce helper states to calculate a few useful quantities: Total Hospitalized (TH), Total Detected Deceased (DD) and Total Detected Cases (DT). The full mathematical formulation of the model is as follows:

$$
\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}t} &= -\widetilde{\alpha}\gamma(t)S(t)I(t) \\
\frac{\mathrm{d}E}{\mathrm{d}t} &= \widetilde{\alpha}\gamma(t)S(t)I(t) \\
\frac{\mathrm{d}I}{\mathrm{d}t} &= r_i E(t) - r_d I(t) \\
\frac{\mathrm{d}U_R}{\mathrm{d}t} &= r_d(1-\widetilde{p}_{dth})(1-p_d)I(t) - r_{ri}U_R(t) \\
\frac{\mathrm{d}DH_R}{\mathrm{d}t} &= r_d(1-\widetilde{p}_{dth})p_d p_h I(t) - r_{rh}DH_R(t) \\
\frac{\mathrm{d}DQ_R}{\mathrm{d}t} &= r_d(1-\widetilde{p}_{dth})p_d(1-p_h)I(t) - r_{ri}DQ_R(t) \\
\frac{\mathrm{d}U_D}{\mathrm{d}t} &= r_d\widetilde{p}_{dth}(1-p_d)I(t) - \widetilde{r}_{dth}U_D(t) \\
\frac{\mathrm{d}DH_D}{\mathrm{d}t} &= r_d\widetilde{p}_{dth}p_d p_h I(t) - \widetilde{r}_{dth}DH_D(t) \\
\frac{\mathrm{d}DQ_D}{\mathrm{d}t} &= r_d\widetilde{p}_{dth}p_d(1-p_h)I(t) - \widetilde{r}_{dth}DQ_D(t) \\
\frac{\mathrm{d}TH}{\mathrm{d}t} &= r_d p_d p_h I(t) \\
\frac{\mathrm{d}DD}{\mathrm{d}t} &= \widetilde{r}_{dth}(DH_D(t) + DQ_D(t)) \\
\frac{\mathrm{d}DT}{\mathrm{d}t} &= r_d p_d I(t) \\
\frac{\mathrm{d}R}{\mathrm{d}t} &= r_{ri}(U_R(t) + DQ_R(t)) + r_{rh}DH_R(t) \\
\frac{\mathrm{d}D}{\mathrm{d}t} &= \widetilde{r}_{dth}(U_D(t) + DQ_D(t) + DH_D(t))
\end{aligned}
$$

This set of differential equations comprises 11 explicit parameters, defined below. The parameters with a tilde are the parameters that are fitted against historical data for each state; the others are fixed parameters that we estimate using our clinical outcomes database (Section 1).

- $\widetilde{\alpha}$ is the baseline infection rate, constant across all US states.
- $\gamma(t)$ measures the government response and is defined as:

$$\gamma(t) = \frac{2}{\pi} \arctan\left(\frac{-(t-\widetilde{t_0})}{\widetilde{k}}\right) + 1,$$

  where the parameters $\widetilde{t_0}$ and $\widetilde{k}$ capture, respectively, the start date and the strength of the response.
- r_d is the rate of detection. This equals to $\frac{\log 2}{T_d}$, where T_d is the median time to detection (2 days).
- r_i is the rate of infection leaving incubation phase. This equals to $\frac{\log 2}{T_i}$, where T_i is the median time to leave incubation (5 days).
- r_{ri} is the rate of recovery not under hospitalization. This equals to $\frac{\log 2}{T_{ri}}$, where T_{ri} is the median time to recovery when a patient is not under hospitalization (10 days).
- r_{rh} is the rate of recovery under hospitalization. This equals to $\frac{\log 2}{T_{rh}}$, where T_{rh} is the median time to recovery under hospitalization (15 days).
- $\widetilde{r}_{dth}$ is the rate of death.
- $\widetilde{p}_{dth}$ is the mortality rate.
- p_d is the percentage of infection cases detected. This percentage is constant and is set to 20%.
- p_h is the percentage of detected cases hospitalized. This percentage is also constant and set to 15%.

Therefore, we fit on 5 parameters from the list above ($\widetilde{\alpha}, \widetilde{p}_{dth}, \widetilde{r}_{dth}, \widetilde{t_0}, \widetilde{k}$). In addition, we create two additional parameters $\widetilde{k_1}, \widetilde{k_2}$ to account for the initial population in the infected (I) and exposed (E) states. We thus fit seven parameters per state.

The parameters are fitted using non-convex optimization methods, including trust-region methods (5) and the Nelder-Mead method (6). We use historical counts of cases and deaths for the fitting procedure. We use a weighted Mean Squared Error (MSE) metric to account for recency and different types of data. The weighted MSE for a training period of T days is defined as:

$$\begin{aligned}\textbf{Weighted MSE} = &\sum_{t=1}^{T} t \cdot \left(DT(t) - \text{Total Detected Cases on Day } t\right)^2 \\ &+ \lambda^2 \cdot \sum_{t=1}^{T} t \cdot \left(DD(t) - \text{Total Detected Deceased on Day } t\right)^2.\end{aligned}$$

The factor t gives more prominence to more recent data, as recent errors are more likely to propagate into future errors. We set $\lambda = \min\left\{\frac{\text{Total Detected Cases on Day } T}{3 \cdot \text{Total Detected Deceased on Day } T}, 10\right\}$ to balance the fitting between cases and deaths.

A.2. Modeling Government Response. As governments respond to the spread of the epidemic, the rate of infection decreases. We model this by multiplying an initial infection rate with an inverse tangent function, which captures three phases of government response (Figure S3).

- **Phase I**: This phase models the initial response when the government has just started to consider implementing policies. Some people have already changed their behavior in response to early reports, but most people continue business-as-usual activities.
- **Phase II**: This phase is characterized by the sharp decline in infection rate as policies get broadly implemented.
- **Phase III**: This phase reflects the diminishing marginal returns in the decline of the infection rate as the measures reach saturation.

Using parameters $\widetilde{t_0}$ and $\widetilde{k}$, we control the start time and the strength of the measures. We can therefore interpret $\widetilde{t_0}$ as the median day of action, and $\widetilde{k}$ as the median rate of action. This formulation allows us to model, under the same framework, a wide variety of policies—spanning school closures, restriction on mass gatherings, stay-at-home policies, etc.

A.3. DELPHI-pred Validation. Figure S4 shows the projected number of deaths in the United States, with projections made on three different weeks, against historical observations. This complements the corresponding figure in the main text reporting the number of projected vs. actual cases. We see that we were generally able to predict the number of deaths up to 4 weeks ahead with good accuracy. One exception is our prediction made on April 3, which is due to a lack of state-level data on deaths at the time (hence, we had to assume a constant mortality rate per state). But after that, our projections have closely followed historical trends.

B. Formulation of DELPHI-pres.

B.1. Modeling of the Impact of Government Response. Recall that we fit a machine learning model to predict the value of $\gamma(t)$ (fitted by DELPHI-pred), as a function of the policies in place. The objective is to evaluate the impact of each policy on the infection rate in order to simulate its overall effect on the dynamics of the pandemic. Figure S5 shows the resulting regression tree, using state-level data in the United States. The results show that more stringent policies result in lower values of $\gamma(t)$, hence in lower infection rates. For instance, in states with no measure in place, the predicted value of $\gamma(t)$ is 1.304; in states where a stay-at-home policy is in place, the predicted value of $\gamma(t)$ is 0.312; in states where partial social distancing policies are in place, the predicted value of $\gamma(t)$ falls in between.

The main objective of DELPHI-presc is to modify the value of $\gamma(t)$ in DELPHI-pred to account for future changes in social distancing policies, using the values predicted by the tree shown in Figure S5. To this end, we define the following quantities:

- t_c is the time of the policy change.

- k_0 is a *normalized* pair-wise difference between the predicted values of $\gamma(t)$ between policies (with respect to the largest predicted value of $\gamma(t)$ under no measure). For instance, transitioning from stay-at-home to no measure induces an offset $(1.304 - 0.312)/1.304 = +0.761$. All values can be found in Table S6—the offset is positive if the new policy is more lenient, and negative if it is more stringent.
- p_0 is the normalized value of the current policy.

We then correct the government response as follows:

$$\gamma'(t) = \max\left\{\frac{2}{\pi}\arctan\left(-\frac{t-\widetilde{t_0}}{k}\right) + 1 + k_0 \cdot \min\left[\frac{2-\gamma(t_c)}{1-p_0}, \frac{\gamma(t_c)}{p_0}\right], 0\right\}, \quad \forall t \geq t_c.$$

For example, if we are currently in Lockdown and are moving to No measure, then we obtain $k_0 = 0.787$, $p_0 = 0.329/1.544 = 0.213$ and $\gamma(t_c) = \frac{2}{\pi}\arctan\left(-\frac{t_c-\widetilde{t_0}}{k}\right) + 1$.

B.2. Application to Policy Assessment. To assess any policy, we run the DELPHI-pred model (governed by the system of differential equations), using the value of the infection rate derived in Section Eq. (B.1). We report the impact of the different policies on the case count in the main body of the paper. Figure S6 provides a similar visualization of the effect on the death count in the state of New York. We can draw similar observations regarding the impact of the various policies and the impact of the timing of these policies.

4. Supplementary Materials on Ventilator Allocation

We now detail the formulation of the optimization model proposed for ventilator allocation. We begin by specifying the model mathematically, then discuss data sources and parameter calibration.

A. Formulation. We consider S states, indexed by $s = 1, \cdots, S$, and D days, indexed by $d = 1, \cdots, D$.

Data. In formulating the problem, we consider the following data as given:
- $v_{s,d}$ is the demand for ventilators in state s on day d.
- b_s is the base supply of ventilators starting in each state s.
- n_d is the surge supply of ventilators distributed by the federal government on day d.
- $d_{s,s'}$ is the distance between state s and state s'.
- $\tau_{s,s'}$ is the lead time between state s and state s'.

We note two comments regarding these inputs. First, the surge supply n_d corresponds to the number of ventilators that are actually distributed by the federal government on day d: the details of managing the federal stockpile fall beyond the scope of this model. More generally, n_d represents supply available from any exogenous, centralized source. Second, we consider distances between states as a way to encourage transfers between neighboring states. We also calibrate the distances such that $d_{s,s} > 0$ for each state s, to ensure we do not propose meaningless transfers from state s to itself.

Decisions. We define integer decision variables as follows:
- $x_{s,d} \in \mathbb{Z}^+$ is the supply of ventilators in state s on day d.
- $y_{s,s',d} \in \mathbb{Z}^+$ is the number of ventilators sent from state s to state s' on day d.
- $z_{s,d} \in \mathbb{Z}^+$ is the additional supply state s receives from the federal government on day d.
- $w_{s,d} \in \mathbb{Z}^+$ is the shortage of ventilators in state s on day d relative to the demand $v_{s,d}$.
- $\Delta_{s,d} \in \mathbb{Z}^+$ is the shortage of ventilators in state s on day d relative to the demand with a buffer.

Parameters. We define the following parameters, which control different policy trade-offs:
- $f_{\max} \in [0, 1]$ is the maximum fraction of its base supply that each state is willing to share. A value of $f_{\max} = 0$ indicates that states are not willing to share any ventilator with other states; a value of $f_{\max} = 1$ indicates that states are willing to share all their ventilator supply with other states.
- $\alpha \in [0, \infty)$ is the percentage of projected demand that states would like to plan for with a supply buffer. For example, $\alpha = 0.1$ will penalize any solution such that supply falls within 10% of projected demand.
- $\lambda \in [0, \infty)$ is a regularization parameter that captures the trade-off between the financial and logistical cost of interstate transfers with the public health cost of ventilator shortages.
- $t_{\min} \in \mathbb{Z}^+$ is the number of days a ventilator is in use after it is shipped to a new location, allowing to control for excessive transfers.
- $\rho \in [0, 1]$ is a relative cost parameter capturing the relative importance of projected shortages vs. worst-case shortages. Each unit of supply that falls short of the projected demand is assigned a cost of 1. Each unit of supply that exceeds the demand but does not exceed the state's desired supply buffer is assigned a cost of ρ.

Objective. The problem of allocating scarce resources in a pandemic is complex because of the necessity to balance competing interests. We identify two key operational goals: improving public health outcomes, and reducing financial cost. We therefore formulate the ventilator allocation problem with two objectives, minimizing total shortage costs as well as total ventilator transfer costs. Each unit of ventilator shortage is assigned a weight of 1 (for shortage relative to the projected demand) or a weight of $\rho \leq 1$ (for shortage relative to the buffered demand). We formalize the bi-objective problem by means of a penalty on transfers, weighted with a trade-off parameter λ.

$$\min \quad \sum_{s=1}^{S}\sum_{d=1}^{D}(w_{s,d} + \rho\Delta_{s,d}) + \lambda\sum_{s=1}^{S}\sum_{s'=1}^{S}\sum_{d=1}^{D} d_{s,s'}y_{s,s',d}. \qquad [1a]$$

Note that ventilators distributed by the federal government are not penalized, as our model simply treats this source as exogenous.

Constraints.

- Initial supply for each state s:

$$x_{s,0} = b_s, \quad \forall s = 1, \cdots, S. \tag{1b}$$

- The supply in each state s on each day d remains higher than the fraction of its initial supply the state wants to retain:

$$x_{s,d} \geq (1 - f_{\max}) b_s, \quad \forall s = 1, \cdots, S, \ d = 1, \cdots, D. \tag{1c}$$

- The transfers from the federal government cannot exceed the available surge supply on each day d:

$$\sum_{s=1}^{S} z_{s,d} \leq n_d, \quad \forall d = 1, \cdots, D. \tag{1d}$$

- The shortage variable corresponds to the positive part of the difference between ventilator demand and supply, if positive, for each state s and day d:

$$w_{s,d} \geq v_{s,d} - x_{s,d}, \quad \forall s = 1, \cdots, S, \ d = 1, \cdots, D. \tag{1e}$$

- The buffer shortage variable is defined such that the total (actual plus buffer) shortage corresponds to the difference between buffered demand and ventilator supply, if positive, for each state s and day d:

$$w_{s,d} + \Delta_{s,d} \geq (1 + \alpha) v_{s,d} - x_{s,d}, \quad \forall s = 1, \cdots, S, \ d = 1, \cdots, D. \tag{1f}$$

- (Conservation of flow) For each state s and day d, today's supply is equal to yesterday's supply plus what is received today from the government and the other states, minus what is sent to other states, with $\tau_{s',s}$ reflecting shipments' lead times:

$$x_{s,d} = x_{s,d-1} + z_{s,d} + \sum_{s'=1}^{S} y_{s',s,d-\tau_{s',s}} - \sum_{s'=1}^{S} y_{s,s',d}, \quad \forall s = 1, \cdots, S, \ d = 1, \cdots, D. \tag{1g}$$

- (Minimum days in use) For each state s and day d, any incoming ventilator, either from another state or from the federal government, cannot be shipped out for at least t_{min} days. This constraint ensures that ventilators are not transferred too often.

$$\sum_{d'=\max(1,d-t_{\min})}^{d-1} \left(z_{s,d'} + \sum_{s'=1}^{S} y_{s',s,d'} \right) \leq x_{s,d}, \quad \forall s = 1, \cdots, S, \ d = 1, \cdots, D. \tag{1h}$$

- Any state s facing a shortage on day d cannot ship any ventilators to other states on day d. To write this constraint, we define auxiliary binary variables $a_{s,d} \in \{0, 1\}$ indicating if there is a shortage in state s on day d and a parameter $V_{\max}$ providing a trivial upper bound on the number of ventilators a state can ship per day (we use a value of 3,000 which does not restrict the solution, as it exceeds the shortage faced by any state on any given day).

$$w_{s,d} + \Delta_{s,d} \leq v_{s,d}(1 + \alpha) a_{s,d}, \tag{1i}$$

$$\sum_{s'=1}^{S} y_{s,s',d} \leq V_{\max}(1 - a_{s,d}). \tag{1j}$$

B. Data Sources and Parameter Calibration. Our optimization model is complex enough to model high-level dynamics of scarce resource allocation, yet simple enough to only require simple data inputs. We now describe our methodology in collecting the key data necessary to solve this optimization problem.

Demand. The most important input data is the forecasted ventilator demand $v_{s,d}$. Consistent with our end-to-end data-driven approach, and in contrast with other ventilator allocation approaches (7), we develop our own demand forecasts using DELPHI-pred (Section 3 of the main text). Recall that DELPHI-pred does not only estimate the number of cases, but also the number of hospitalizations, equal to $DH_R + DH_D$. We then apply a 25% ratio to estimate the number of ventilators in use—given that, in our clinical outcomes database, 25% of hospitalized patients are on a ventilator. Ultimately, we can use the DELPHI-pred outputs to derive projections of ventilator demand at the state level and at the daily level—consistently with the optimization input $v_{s,d}$.

Supply. It can be difficult to estimate how many ventilators are available in each state as well as at the federal level. For the base supply b_s, we use inventory levels from a 2010 American Medical Association report (8). We adjust this number for population growth, under the assumption that the number of ventilators per capita has remained constant in each state.

Of course, ventilators can also be used to treat non-COVID-19 patients. We assume that 50% of ventilator supply across the board is unavailable due to non-COVID-19 usage, in line with other estimates (7).

In addition, our model takes into account the daily availability n_d of ventilators at the federal level. Estimating this quantity from publicly-available sources is both difficult due to limited data and politically fraught. We use the estimate from the Society for Critical Medicine that the federal stockpile contains at least 12,700 ventilators (9). Some news reports suggest a lower estimate of 10,000 based on some defects in the stockpiled equipment (10), while others suggest an estimate of 16,600 based on older model repairs (11). Based on these reference points, we estimate roughly 13,500 available ventilators and assume that they can be deployed evenly over a month. In other words, we allow 450 ventilators to be deployed each day for 30 days (starting on day 4 to allow for lead times). This gradual release reflects potential operational constraints and strategic considerations of controlling the release of inventory in case of unexpected outbreaks. Yet, given the underlying uncertainty, we perform sensitivity analysis to explore how the model's recommendation varies with the federal stockpile.

Distances and lead times. We compute the interstate distance $d_{s,s'}$ as the Euclidean distance between the centers of states s and s', and we let the lead time parameter $\tau_{s,s'}$ equal 3 days for every pair of states. Our choice of a conservative uniform lead time for shipments is motivated by simplicity concerns. This could be improved, in future work, to better reflect efficiencies in the US shipping infrastructure.

Trade-off parameters.

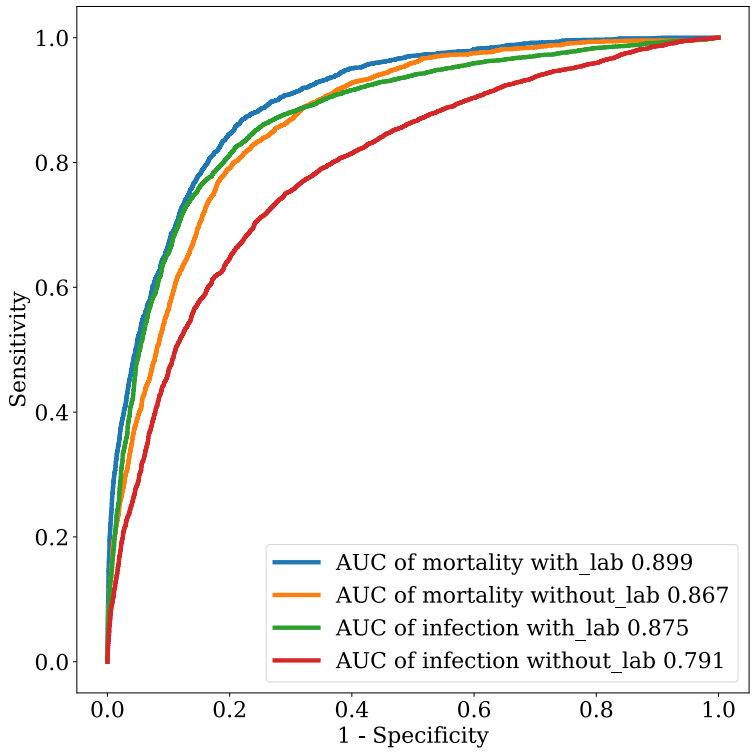
- Understanding the impact of states' willingness to share ventilators with other states is a key takeaway from our model. In Figure 7B, we vary the fraction $f_{\max}$ of each state's pooling supply between $\{0\%, 5\%, 10\%, 15\%, 20\%\}$. Results indicate there is little additional efficiency to be gained from states sharing more than 20% of their supply.
- Understanding the relative importance of federal surge supply compared to interstate transfers is another interesting takeaway from our model. In Figure 7A, we show the effects of removing federal surge supply, or preventing interstate transfers, on ventilator shortage reduction. We show more detailed sensitivity analysis results in Section C.
- We vary the parameter λ to derive the the Pareto-optimal frontier of the trade-off between inter-state transfers vs. ventilator shortages (for instance in Fig. 7B). As λ tends to zero, transfers incur no cost other than rendering the ventilators unavailable during shipment; as λ tends to infinity, transfers become heavily discouraged.
- The parameter α models uncertainty in the demand forecast as well as robustness to inefficiency in ventilator allocation within each state. We vary the percentage α of buffered demand within $\{0\%, 5\%, 10\%, 20\%\}$ in Section C.

Finally, we choose the following values for the remaining two parameters, which have a small impact on the final solution.

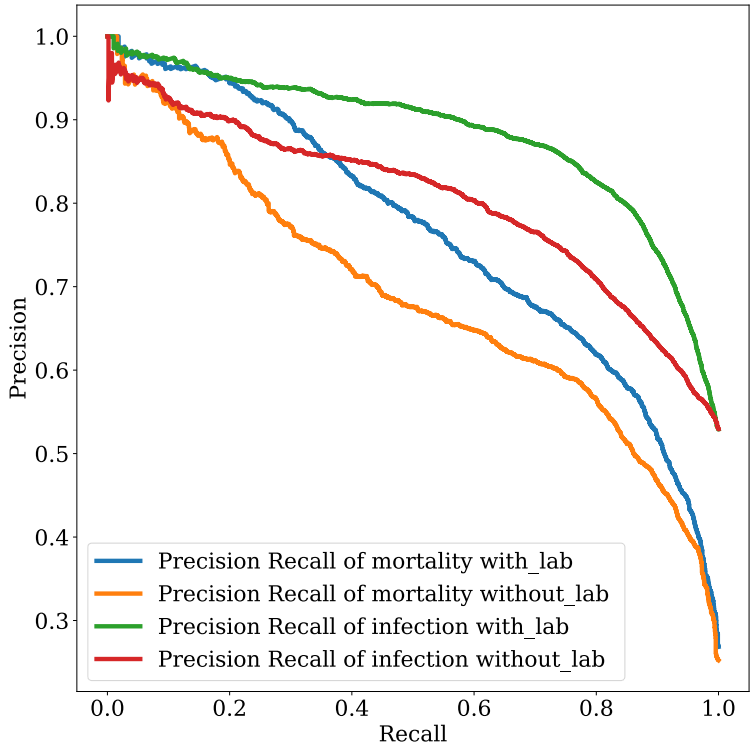
- We set the value of $t_{\min}$ to 10 days, based on the clinical outcomes database (Section 1).
- We set ρ, the cost of shortage with respect to buffered demand relative to the cost of shortage with respect to real demand, to 0.25.

C. Sensitivity analysis. In the main text, we discuss the impact of the federal surge and inter-state coordination on alleviating ventilator shortages, showing that ventilator shortages can be eliminated through limited transfers among states and from the federal government. The results also suggest trade-offs between the number of inter-state transfers, the amount of ventilator shortages, and the fraction of ventilators that each state is willing to share (captured by the $f_{\max}$ parameter). We report additional results to underscore the trade-offs between inter-state transfers, the ventilator supply from the federal government, and the robustness of the solution—that is, the exposure of the states to ventilator shortages if demand exceeds our projections.

To this end, Figure S7 shows the Pareto-optimal frontier between the model's two objectives, inter-state transfer distance and ventilator shortage, as a function of two model parameters: α (capturing the demand buffer that states would like to plan for) and a *surge supply multiplier* (capturing by how much the federal government's ventilator supply varies from our estimates). Note that the buffer α does not impact the number of inter-state transfers and the amount of ventilator shortages too significantly—suggesting that the solution can be made robust at limited overall costs. In contrast, as the surge supply is decreased, the number of required ventilator transfers and the amount of ventilator shortages increase—highlighting the need for stronger cooperation between states as federal supply drops.



(a) Average ROC Curves



(b) Average Precision Recall Curves

Fig. S1. Bootstrapped results on ROC and Precision Recall curves for all calculators on the testing set.

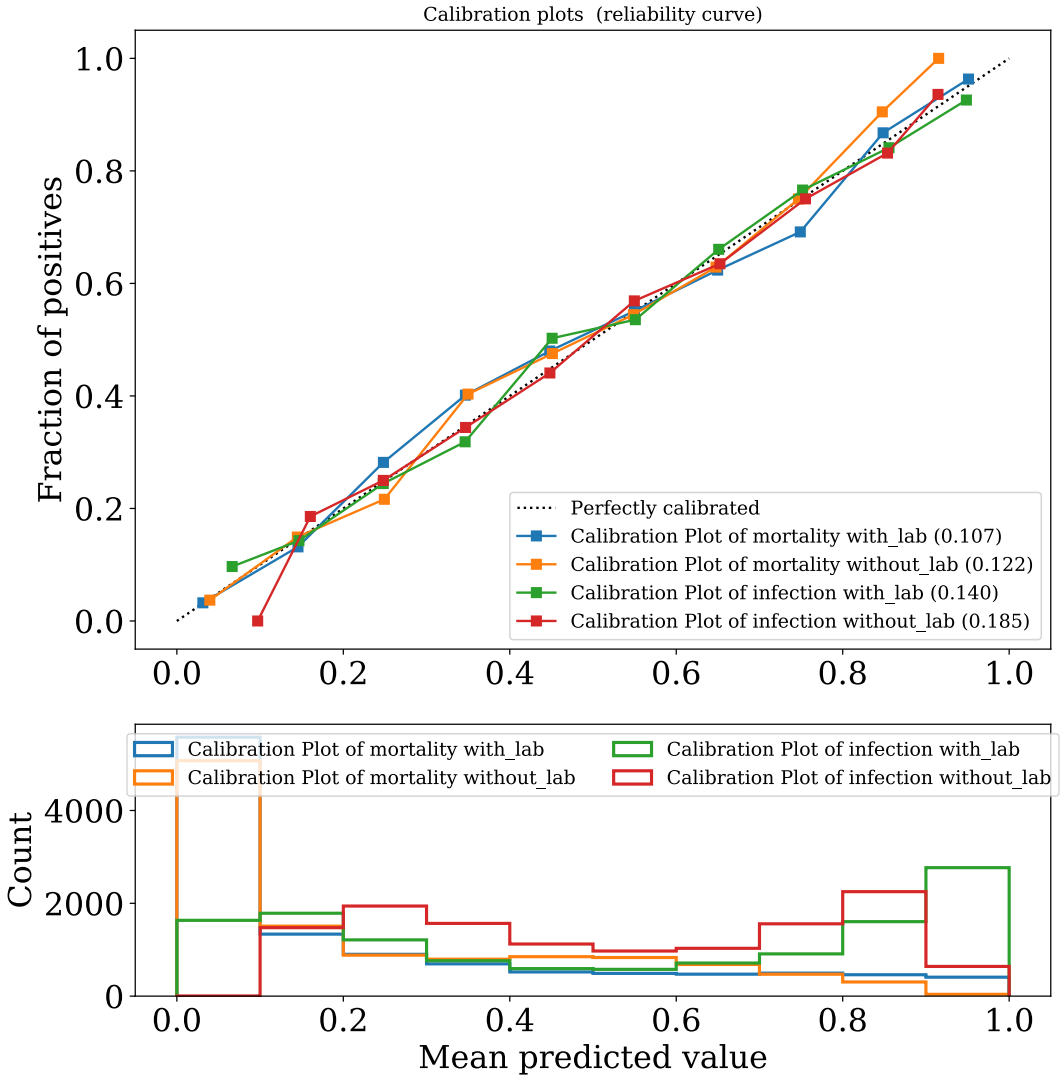


Fig. S2. Bootstrapped Results on the Calibration Curves for both risk calculators on the testing set. The intervals are: [0,10%], (10,20%], (20,30%], …, (90,100%]. The event rates are plotted against the bin mid-points. An ideal event rate is marked by the dotted 45 degree line.

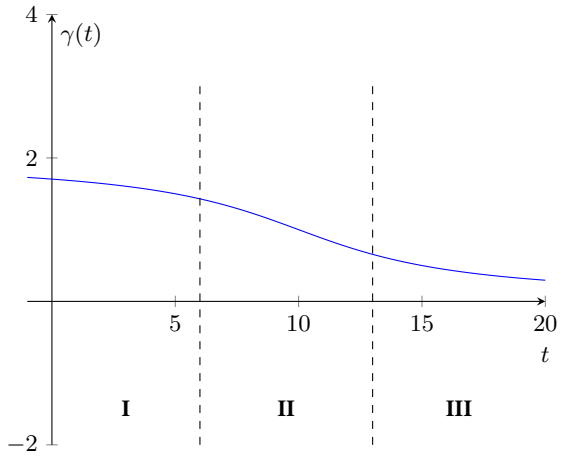


Fig. S3. Illustration of $\gamma(t) = \frac{2}{\pi} \arctan(-\frac{t-10}{5}) + 1$ (i.e., $\widetilde{t_0} = 10$ and $k = 5$).

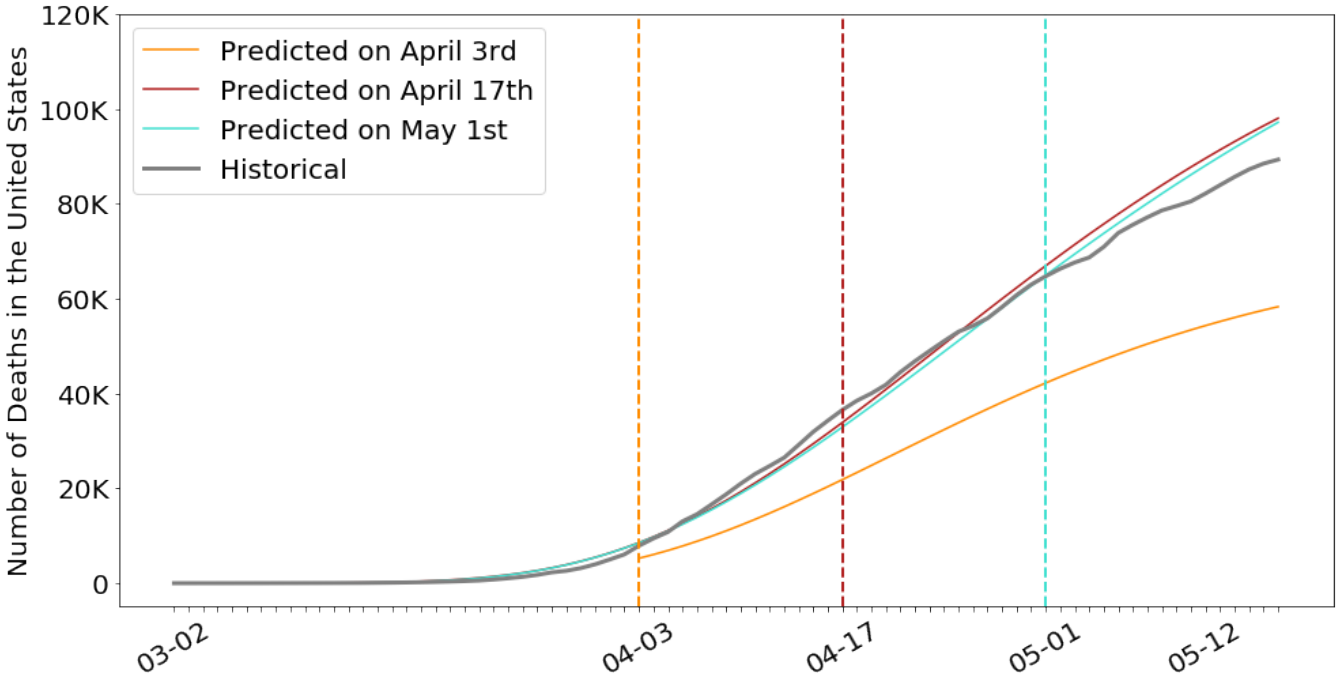


Fig. S4. Cumulative number of deaths in the United States according to our projections made at different points in time, against actual observations

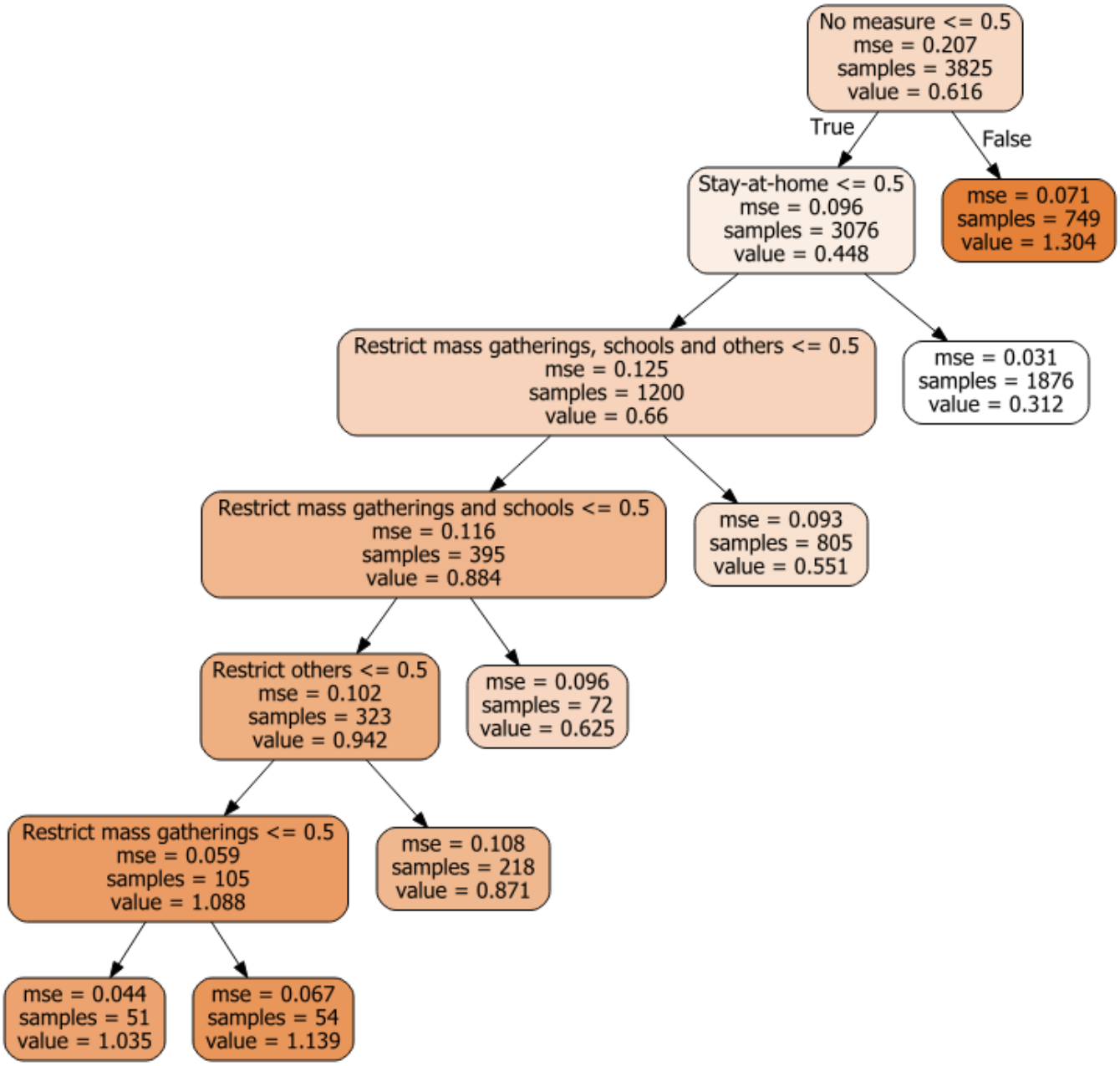
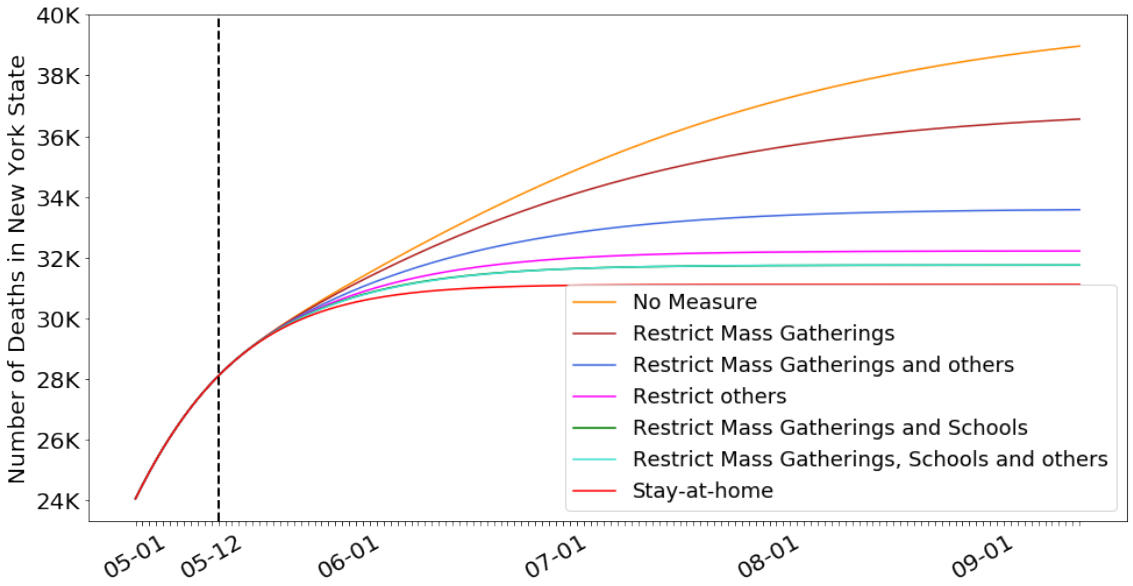
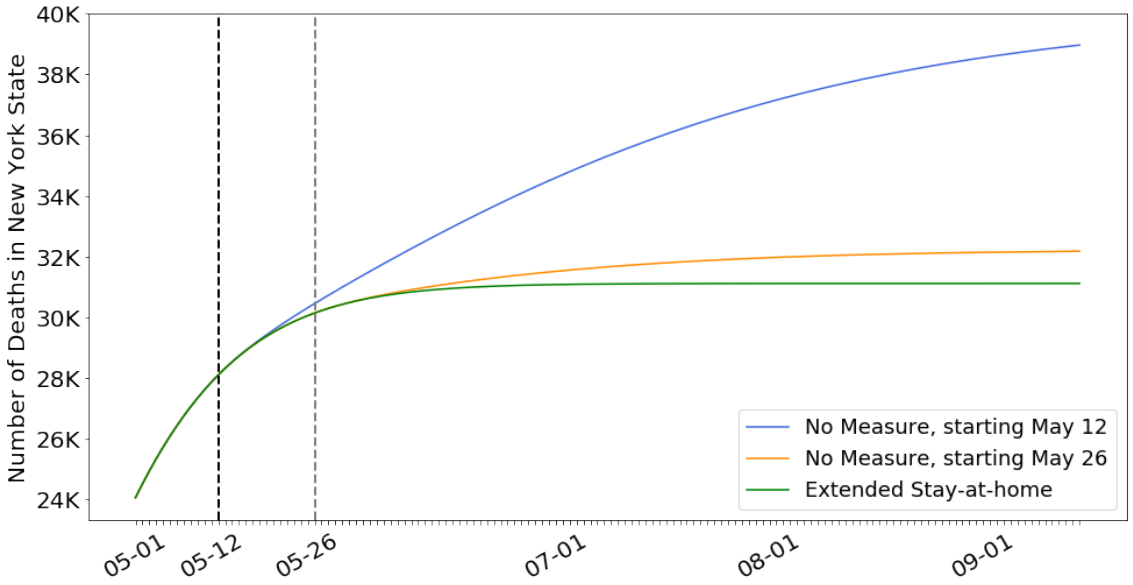


Fig. S5. Regression Tree (CART) predicting an average value of $\gamma(t)$ for each policy.

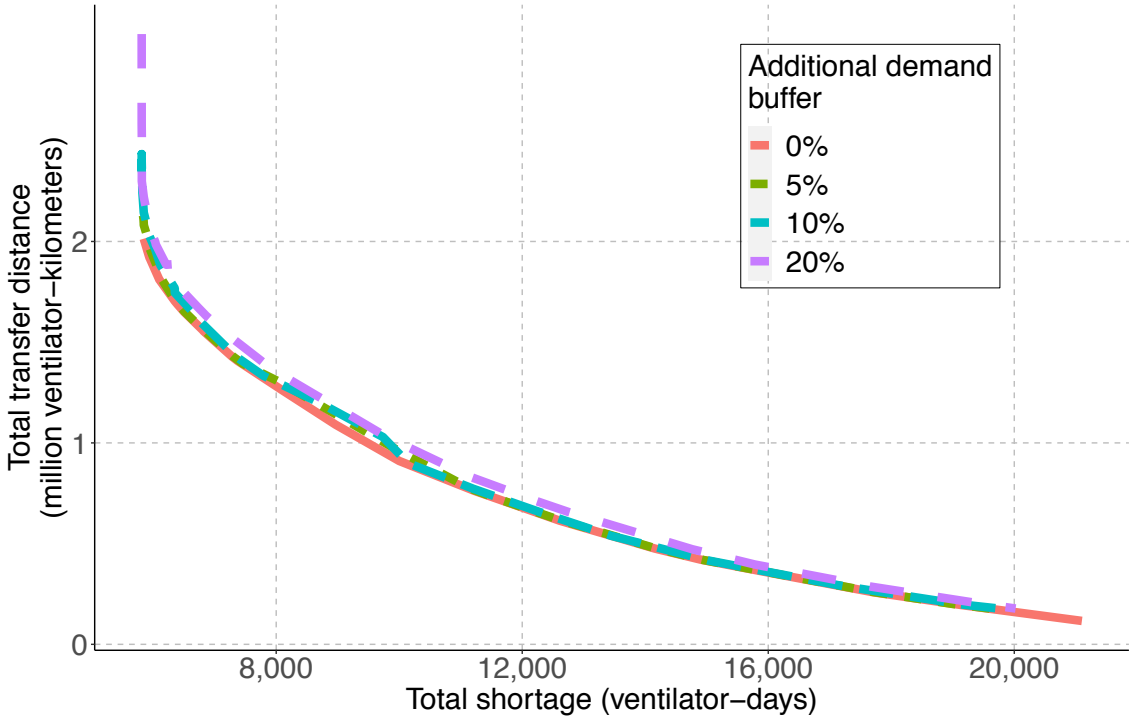


(a) Impact of different policies on the future number of deaths, in NY

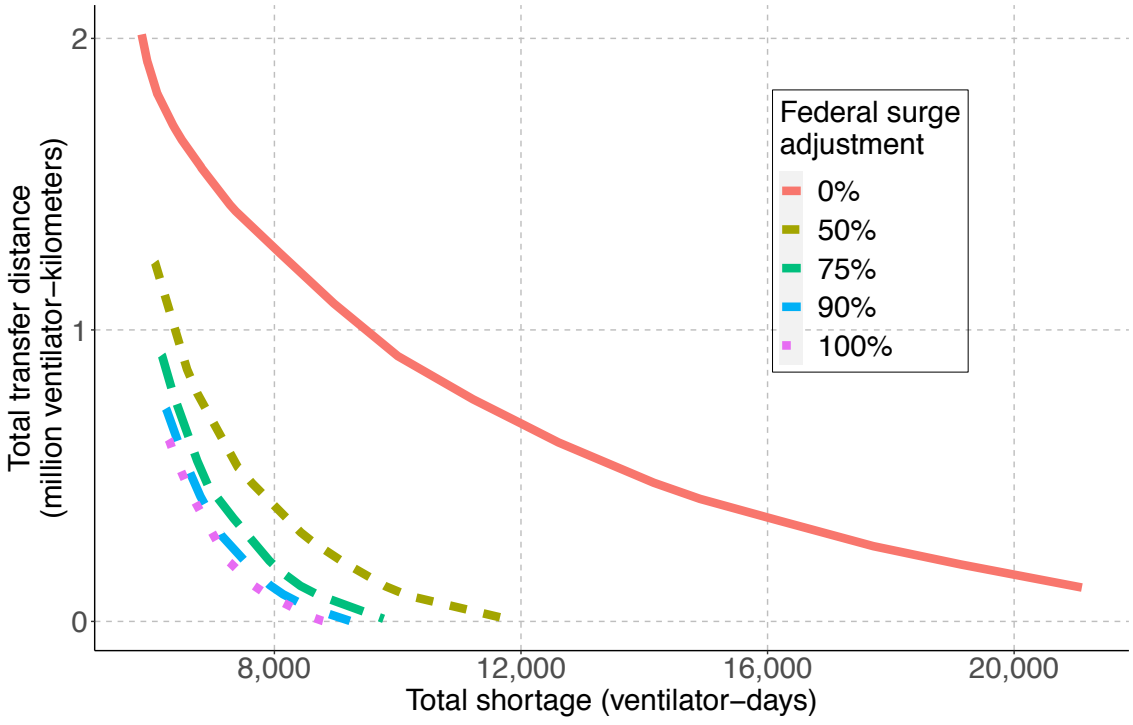


(b) Impact of the timing of policies on the future number of deaths, in NY.

Fig. S6. Impact of different policies on the future number of deaths in the State of New York, for different policies and policy start dates.



(a) Impact of additional buffer



(b) Impact of federal surge

Fig. S7. Influence of additional buffer and federal surge availability on ventilator shortage and ventilator transfers.

Table S1. Comorbidities, demographics, average lab values, average length of stay and projected mortality among COVID-19 patients, in aggregate and broken down into mild/severe patients.

Feature	All Patients		Mild Patients		Severe Patients	
	No. Report	Avg.	No. Report.	Avg.	No. Report.	Avg.
Demographics						
Male (%)	131,200	53.0%	9,570	48.8%	10,120	68.7%
Age (years)	119,000	51.3	8,022	46.1	9,685	68.2
White/European (%)	55,490	22.2%	10,120	9.7%	9,887	63.9%
African American (%)	55,490	5.4%	10,120	3.5%	9,887	2.5%
Asian (%)	55,320	51.3%	10,300	80.2%	9,933	31.2%
Hispanic/Latino	50,630	19.9%	8,017	0%	9,107	0%
Multiple ethnicities/other	55,190	3.6%	10,120	6.9%	9,887	2.7%
Comorbidities						
Smoking history	27,900	16.1%	6,080	12.2%	1,973	16.6%
Hypertension	38,390	35.9%	8,252	15.2%	8,449	54.4%
Diabetes	39,790	20.8%	8,396	6.8%	8,818	26.1%
Cardio Disease	40,030	12.4%	8,028	3.0%	9,540	20.3%
COPD	34,150	6.0%	6,297	2.8%	8,727	10.0%
Cancer	29,170	7.2%	6,259	3.2%	8,355	12.9%
Liver Disease	18,300	2.8%	1,875	2.3%	6,832	3.5%
Cebrovascular	6,830	9.8%	3,245	2.7%	1,360	24.8%
Kidney Disease	35,500	5.7%	6,152	1.2%	8,139	10.8%
Lab values						
WBC Count (10^9/L)	19,970	6.41	5,403	5.07	2,305	6.80
Neutrophil Count (10^9/L)	12,500	4.72	2,236	5.12	1,410	5.78
Platelet Count (10^9/L)	12,125	195.7	5,165	184.0	2,105	170.4
ALT (U/L)	14,467	29.0	2,840	24.6	2,428	31.1
AST (U/L)	14,214	37.3	2,766	27.1	2,366	45.7
BUN (mmol/L)	4,822	5.22	1,700	4.18	1,138	6.86
Creatinine (μmol/L)	8,504	63.08	2,529	66.0	2,454	56.4
CRP Count (mg/L)	17,090	76.5	2,573	18.9	2,339	94.1
Interleukin-6 (pg/mL)	2,582	24.57	1,127	4.17	552	38.63
Procalcitonin (ng/mL)	14,750	2.26	1,468	1.85	1,969	4.81
D-Dimer (mg/L)	13,330	38.81	2,478	8.04	2,401	165.9
Length of Stay (days)	16,010	10.7	4,131	14.0	5,642	7.97
Proj. Mortality (%)	111,700	11.7%	7,428	0.4%	9,146	74.0%

Table S2. Count and prevalence of treatments among COVID-19 patients, broken down per continent (Asia, Europe, North America). A "-" indicates that fewer than 100 patients in a subpopulation reported on this symptom.

Treatment	Asia		Europe		North America	
	No. Report	Prev. (%)	No. Report.	Prev. (%)	No. Report.	Prev. (%)
Kaletra	5,665	35.2%	–	–	–	–
Oseltamivir	5,901	25.1%	–	–	–	–
Remdesivir	337	47.4%	–	–	868	10.3%
Arbidol	5,902	34.8%	–	–	–	–
Interferon	3,647	51.8%	–	–	–	–
Hydroxychloroquine	6,008	0.7%	–	–	1,235	61.7%
Invasive Ventilation	7,945	8.0%	75,120	4.8%	5,840	19.3%
Proj. Mortality	12,820	16.7%	79,750	9.9%	19,060	15.8%

Table S3. Characteristics of study population for mortality prediction model.

	All ($N = 2,831$)	Survivors ($N = 2,120$)	Non-Survivors ($N = 711$)	P-Value
Age	68.0 (57.0-79.0)	63.0 (54.0-74.0)	81.0 (73.2-86.0)	1.28E-185
Female *	1095.0 (38.7%)	868.0 (40.9%)	227.0 (31.9%)	1.18E-05
Heart Rate	89.0 (79.0-101.0)	90.0 (80.0-102.0)	87.0 (78.0-100.0)	1.29E-03
Oxygen Saturation	94.0 (90.0-96.0)	94.4 (92.0-96.0)	88.5 (80.0-93.6)	3.16E-37
Temperature (F)	98.4 (97.5-99.7)	98.4 (97.5-99.6)	98.8 (97.7-100.0)	2.42E-04
Alanine Aminotransferase	27.0 (17.0-44.0)	27.8 (17.5-45.0)	25.5 (16.0-41.0)	3.77E-02
Aspartate Aminotransferase	36.0 (25.0-55.0)	34.0 (24.4-51.0)	45.0 (30.0-69.0)	1.55E-11
Blood Glucose	118.0 (105.0-141.0)	115.0 (103.4-133.0)	134.0 (113.0-171.0)	1.12E-22
Blood Urea Nitrogen	17.0 (12.6-25.2)	15.0 (11.5-20.0)	29.5 (20.3-47.2)	1.02E-65
C-Reactive Protein	74.2 (29.1-149.5)	58.6 (22.7-119.3)	141.1 (72.0-223.1)	4.76E-50
Creatinine	1.0 (0.8-1.2)	0.9 (0.7-1.1)	1.3 (1.0-1.8)	2.84E-36
Hemoglobin	13.9 (12.7-15.0)	14.0 (12.9-15.0)	13.5 (12.0-14.7)	9.11E-10
Mean Corpsular Volume	87.8 (84.9-91.0)	87.5 (84.7-90.4)	89.3 (85.8-92.7)	2.80E-08
Platelets	201.0 (156.0-263.0)	206.0 (160.0-266.5)	185.0 (141.0-246.8)	6.62E-08
Potassium	4.1 (3.7-4.4)	4.0 (3.7-4.4)	4.1 (3.7-4.6)	1.43E-04
Prothrombin Time (INR)	1.1 (1.0-1.2)	1.1 (1.0-1.2)	1.1 (1.0-1.3)	3.20E-05
Sodium	137.1 (135.0-140.0)	137.0 (135.0-139.4)	138.0 (135.0-141.0)	5.65E-08
White Blood Cell Count	6.8 (5.2-9.2)	6.5 (5.0-8.7)	8.0 (5.7-11.4)	3.00E-15
Cardiac dysrhythmias *	200.0 (7.1%)	127.0 (6.0%)	73.0 (10.3%)	6.50E-04
Chronic kidney disease *	65.0 (2.3%)	33.0 (1.6%)	32.0 (4.5%)	3.67E-04
Heart disease *	125.0 (4.4%)	80.0 (3.8%)	45.0 (6.3%)	1.10E-02
Diabetes *	345.0 (12.2%)	234.0 (11.0%)	111.0 (15.6%)	2.73E-03
Mortality *	711 (25.1%)	0 (0%)	711 (100%)	–

* Count (proportion) is reported for binary variables.

Table S4. Characteristics of study population for infection test prediction model.

	All ($N = 3,135$)	No Infection ($N = 1,474$)	Infection ($N = 1,661$)	P-Value
Age	63.0 (49.0-78.0)	58.0 (42.0-78.0)	66.0 (55.0-78.5)	9.00E-28
Female *	1444.0 (46.1%)	777.0 (52.7%)	667.0 (40.2%)	1.70E-12
Heart Rate	88.5 (78.0-100.5)	89.0 (78.0-101.2)	88.0 (78.2-100.0)	6.13E-01
Oxygen Saturation	95.4 (91.8-97.0)	96.5 (94.8-97.5)	94.2 (89.6-96.4)	1.68E-31
Respiratory Frequency	18.0 (16.0-19.0)	18.0 (16.0-18.0)	18.0 (16.0-20.0)	9.64E-21
Temperature	98.3 (97.5-99.5)	97.7 (97.2-98.7)	99.0 (97.9-100.0)	4.51E-80
Alanine Aminotransferase	22.0 (15.0-37.0)	19.0 (13.0-30.0)	27.0 (18.0-43.0)	6.59E-09
Aspartate Aminotransferase	29.0 (21.0-47.0)	23.0 (19.0-31.0)	37.0 (26.0-57.0)	1.20E-20
Blood Urea Nitrogen	17.0 (13.0-25.0)	16.0 (12.0-22.0)	18.0 (13.0-27.0)	3.78E-05
Calcium	9.3 (8.9-9.7)	9.6 (9.2-9.9)	9.0 (8.7-9.4)	1.90E-96
C-Reactive Protein	31.0 (3.4-107.6)	4.7 (1.1-35.4)	69.8 (23.2-152.3)	1.28E-83
Creatinine	0.9 (0.8-1.2)	0.9 (0.7-1.1)	1.0 (0.8-1.2)	2.19E-05
Hemoglobin	13.5 (12.3-14.7)	13.4 (12.1-14.6)	13.6 (12.5-14.8)	5.70E-05
Mean Corpsular Volume	87.2 (84.0-90.3)	87.7 (84.2-90.7)	86.8 (83.9-90.0)	1.43E-01
Platelets	223.0 (174.0-285.0)	241.0 (198.0-297.0)	202.0 (156.0-266.0)	1.41E-18
Red Cell Distrbution Width	13.2 (12.5-14.3)	13.2 (12.5-14.5)	13.1 (12.4-14.0)	1.59E-06
Sodium	139.0 (137.0-141.0)	140.0 (138.0-142.0)	139.0 (136.0-141.0)	1.12E-14
Prothrombin Time (INR)	1.0 (1.0-1.1)	1.1 (1.0-1.1)	1.0 (1.0-1.1)	8.96E-01
Total Bilirubin	0.6 (0.5-0.8)	0.6 (0.4-0.9)	0.6 (0.5-0.8)	8.83E-03
White Blood Cell Count	7.6 (5.8-10.1)	8.7 (7.0-11.1)	6.6 (5.1-8.7)	7.59E-38
COVID-19 Positive Test *	1661 (53.0%)	0 (0%)	1661 (100%)	–

* Count (proportion) is reported for binary variables.

Table S5. Performance evaluation summary. Average results across 40 random seeds are reported along with 95% confidence intervals. A minimum threshold of 90% sensitivity is enforced.

Model Type	Lab Values	Threshold	Accuracy	Sensitivity	Specificity	Precision	Negative predictive value	False positive rate
Mortality	Present	17.45 (15.61,19.29)	79.29 (77.47,81.11)	90.38 (90.38,90.38)	75.66 (73.25,78.07)	55.86 (53.46,58.25)	95.97 (95.84,96.1)	24.34 (21.93,26.75)
Mortality	Absent	12.66 (11.11,14.2)	70.85 (68.45,73.24)	90.38 (90.38,90.38)	64.53 (61.37,67.7)	46.15 (43.96,48.33)	95.3 (95.04,95.55)	35.47 (32.3,38.63)
Infection	Present	28.32 (26.63,30.02)	77.58 (76.48,78.68)	90.36 (90.36,90.36)	63.24 (60.9,65.58)	73.59 (72.33,74.85)	85.26 (84.79,85.73)	36.76 (34.42,39.1)
Infection	Absent	27.51 (26.55,28.47)	66.31 (65.37,67.25)	90.36 (90.36,90.36)	39.32 (37.33,41.32)	62.64 (61.85,63.44)	78.12 (77.28,78.96)	60.68 (58.68,62.67)

Table S6. Values of normalized offset computations to correct the estimation of $\gamma(t)$. Policies include: (i) *No measure* ("none"); (ii) *Restrict mass gatherings* ("R-MG"); (iii) *Restrict others* ("R-O"); (iv) *Authorize schools, restrict mass gatherings and others* ("R-MG-O"); (v) *Restrict mass gatherings and schools* ("R-MG-S"); (vi) *Restrict mass gatherings, schools and others* ("R-MG-S-O"); and (vii) *Stay-at-home* ("SAH").

from/to	none	R-MG	R-O	R-MG-O	R-MG-S	R-MG-S-O	SAH
none	0	-0.127	-0.332	-0.206	-0.521	-0.577	-0.761
R-MG	+0.127	0	-0.206	-0.080	-0.294	-0.451	-0.634
R-O	+0.332	+0.206	0	+0.126	-0.189	-0.245	-0.429
R-MG-O	+0.206	+0.080	-0.126	0	-0.314	-0.371	-0.554
R-MG-S	+0.521	+0.294	+0.189	+0.314	0	-0.057	-0.240
R-MG-S-O	+0.577	+0.451	+0.245	+0.371	+0.057	0	-0.183
SAH	+0.761	+0.634	+0.429	+0.554	+0.240	+0.183	0

Bibliography

1. O Troyanskaya, et al., Missing value estimation methods for dna microarrays. *Bioinformatics* **17**, 520–525 (2001).
2. T Chen, C Guestrin, XGBoost: A scalable tree boosting system in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16. (ACM, New York, NY, USA), pp. 785–794 (2016).
3. F Pedregosa, et al., Scikit-learn: Machine learning in Python. *J. machine Learn. research* **12**, 2825–2830 (2011).
4. T Head, et al., scikit-optimize/scikit-optimize: v0.5.2 (2018).
5. RH Byrd, JC Gilbert, J Nocedal, A trust region method based on interior point techniques for nonlinear programming. *Math. programming* **89**, 149–185 (2000).
6. JC Lagarias, JA Reeds, MH Wright, PE Wright, Convergence properties of the nelder–mead simplex method in low dimensions. *SIAM J. on optimization* **9**, 112–147 (1998).
7. S Mehrotra, H Rahimian, M Barah, F Luo, K Schantz, A model of supply-chain decisions for resource sharing with an application to ventilator allocation to combat COVID-19. *Nav. Res. Logist. (NRL)* (2020).
8. L Rubinson, et al., Mechanical ventilators in us acute care hospitals. *Disaster medicine public health preparedness* **4**, 199–206 (2010).
9. NA Halpern, KS Tan, VT SCCM, Us icu resource availability for covid-19. *Soc. Critical Care Medicine, March* **25** (2020).
10. New York Times, A Ventilator Stockpile, With One Hitch: Thousands Do Not Work (https://www.nytimes.com/2020/04/01/us/politics/coronavirus-ventilators.html) (2020).
11. New York Times, The U.S. Tried to Build a New Fleet of Ventilators. The Mission Failed. (https://www.nytimes.com/2020/03/29/business/coronavirus-us-ventilator-shortage.html) (2020).