



Institute for Health Metrics and Evaluation

Data Release Information Sheet

Data Summary

Dataset name: United States COVID-19 Scenarios 2020-2021

Date of release: October 22, 2020

Summary:

This dataset represents estimates of the ongoing COVID-19 pandemic across the 50 U.S. States and DC through 28th February 2021. Projections for total and daily deaths, daily infections, and testing are included with hospital resource use statistics. In total five scenarios are presented: a 'plausible reference scenario,' which assumes social distancing mandates are re-imposed for 6 weeks when a threshold daily death rate of 8 per million is reached; a 'mandates easing' scenario, where mandates are not re-imposed; a 'universal mask-use' scenario, where mask utilization reaches 95% usage in public in every location; a less comprehensive mask scenario of 85% public use of masks ('plausible reference + 85% mask-use' scenario); and a scenario of universal mask wearing in the absence of any additional NPI ('mandate easing + universal mask use'). These projections are produced with a model that incorporates data on observed COVID-19 deaths, hospitalizations, and cases, as well as multiple covariates.

Relevant publications and visualizations:

IHME COVID-19 Forecasting Team. COVID-19 scenarios for the United States. *Nature Medicine*. 22 Oct 2020.

Acknowledgements

Contributing organizations:

- Institute for Health Metrics and Evaluation (IHME)

Funders:

- Bill and Melinda Gates Foundation (BMGF)
- State of Washington
- National Science Foundation (United States)

Suggested Citation:

Institute for Health Metrics and Evaluation (IHME). United States COVID-19 Scenarios 2020-2021. Seattle, United States of America: Institute for Health Metrics and Evaluation (IHME), 2020.

Scenario-Specific Inputs and Results

Inputs and results are stratified into five modeled scenarios

- Plausible Reference Scenario – ZIP FILE
- Mandates Easing Scenario – ZIP FILE
- Universal Mask-Use Scenario – ZIP FILE
- Plausible Reference + 85% Mask-Use Scenario – ZIP FILE
- Mandates Easing + Universal Mask-Use Scenario – ZIP FILE

Each zipped file contains the same files with the following constituent schema

Processed Data Inputs

Processing of input data related to COVID reported deaths, cases, and hospitalizations, occurs prior to modelling. This “elastispliner” process has five steps (outlined in more detail in the supplemental information):

- fit a model of deaths ~ cases
- fit a model of hospitalizations ~ cases
- fit a model to observed deaths and “pseudo-deaths” from steps (a) and (b)
- sample from the residuals of step (c) to create 1000 uncorrelated “noisy” time series
- refits the curve to those data to get 1000 “smooth” time series

This process output two files

- cumulative_elastispliner_noisy.csv - the output of step (d)

Column Name	Contents
location_id	Unique identifier for estimation geography; common identifier across various files in this reference
date	Format yyyy-mm-dd
observed	A binary tag differentiating values derived from input data (i.e. observed == 1) versus those from estimates (i.e. observed == 0)
mean	The mean of the death draws
upper	The 97.5 th percentile of the death draws
lower	The 2.5 th percentile of the death draws
location_name	Name of the estimation geography

- cumulative_elastispliner_smoothed.csv - the output of step (e)

Column Name	Contents
location_id	Unique identifier for estimation geography; common identifier across various files in this reference
date	Format yyyy-mm-dd
observed	A binary tag differentiating values derived from input data (i.e. observed == 1) versus those from estimates (i.e. observed == 0)

Column Name	Contents
mean	The mean of the death draws
upper	The 97.5 th percentile of the death draws
lower	The 2.5 th percentile of the death draws
location_name	Name of the estimation geography

Covariates

- air_pollution_pm_2_5.csv

Column Name	Contents
location_id	Unique identifier for estimation geography; common identifier across various files in this reference
observed	Denotes (with a 1) whether there are observations associated with this estimate. Blank values represent if the location is an aggregate. <i>(note: Washington state is estimated at county-level, so is represented as an aggregate here)</i>
modeled	Denotes (with a 1) whether this estimate is derived from a modelling process
mean	The mean of the draws of the estimate
upper	The 97.5 th percentile of draws of the estimate
lower	The 2.5 th percentile of draws of the estimate
location_name	Name of the estimation geography

Covariate values for ambient particulate matter pollution, measuring PM2.5.

- lri_mortality.csv

Column Name	Contents
location_id	Unique identifier for estimation geography; common identifier across various files in this reference
observed	Denotes (with a 1) if this is the level of estimation, or if the estimated location is an aggregate. <i>(note: Washington state is estimated at county-level, so is represented as an aggregate here)</i>
modeled	Denotes (with a 1) whether this estimate is derived from a modelling process
mean	The mean of the draws of the estimate
upper	The 97.5 th percentile of draws of the estimate
lower	The 2.5 th percentile of draws of the estimate
location_name	Name of the estimation geography

Covariate values for age-standardized adult (15+) lower respiratory infection mortality. Singular estimates for 2019.

- mobility.csv

Column Name	Contents
location_id	Unique identifier for estimation geography; common identifier across various files in this reference
date	Format yyyy-mm-dd
observed	Denotes (with a 1) whether there are observations associated with this estimate. Missing values indicate the location shown is an aggregate location (ex: United States, WA state)
modeled	Denotes (with a 1) whether this estimate is derived from a modelling process
mean	The mean of the draws of the estimate
upper	The 97.5 th percentile of draws of the estimate
lower	The 2.5 th percentile of draws of the estimate
location_name	Name of the estimation geography

Covariate values for relative mobility. Reported and estimated daily.

- pneumonia.csv

Column Name	Contents
location_id	Unique identifier for estimation geography; common identifier across various files in this reference
date	Format yyyy-mm-dd
observed	Denotes (with a 1) whether there are observations associated with this estimate. Missing values denote the location shown is an aggregate location.
modeled	Denotes (with a 1) whether this estimate is derived from a modelling process
mean	The mean of the draws of the estimate
upper	The 97.5 th percentile of draws of the estimate
lower	The 2.5 th percentile of draws of the estimate
location_name	Name of the estimation geography

Covariate values for pneumonia seasonality. Reported and estimated daily.

- proportion_over_2_5k.csv

Column Name	Contents
location_id	Unique identifier for estimation geography; common identifier across various files in this reference
observed	Denotes (with a 1) whether there are observations associated with this estimate. Missing values denote the location shown is an aggregate location.
modeled	Denotes (with a 1) whether this estimate is derived from a modelling process
mean	The mean of the draws of the estimate
upper	The 97.5 th percentile of draws of the estimate
lower	The 2.5 th percentile of draws of the estimate

Column Name	Contents
location_name	Name of the estimation geography

Covariate values for proportion of the population living in settings with population density greater than 2,500 people per square kilometer. Singular estimates.

- proportion_under_100m.csv

Column Name	Contents
location_id	Unique identifier for estimation geography; common identifier across various files in this reference
observed	Denotes (with a 1) whether there are observations associated with this estimate. Missing values denote the location shown is an aggregate location.
modeled	Denotes (with a 1) whether this estimate is derived from a modelling process
mean	The mean of the draws of the estimate
upper	The 97.5 th percentile of the death draws
lower	The 2.5 th percentile of the death draws
location_name	Name of the estimation geography

Covariate values for proportion of the population living in settings at altitudes less than 100m above sea level. Singular estimates.

- smoking_prevalence.csv

Column Name	Contents
location_id	Unique identifier for estimation geography; common identifier across various files in this reference
observed	Denotes (with a 1) whether there are observations associated with this estimate. Missing values denote the location shown is an aggregate location.
modeled	Denotes (with a 1) whether this estimate is derived from a modelling process
mean	The mean of the draws of the estimate
upper	The 97.5 th percentile of the death draws
lower	The 2.5 th percentile of the death draws
location_name	Name of the estimation geography

Covariate values for the adult (15+ years) age-standardized smoking prevalence. Singular estimates for 2019.

- testing.csv

Column Name	Contents
location_id	Unique identifier for estimation geography; common identifier across various files in this reference
date	Format yyyy-mm-dd
observed	Denotes (with a 1) whether there are observations associated with this estimate, 0 represents the value is an

Column Name	Contents
	estimate/projection. Missing values denote the location shown is an aggregate location.
modeled	Denotes (with a 1) whether this estimate is derived from a modelling process
mean	The mean of the draws of the estimate
upper	The 97.5 th percentile of the death draws
lower	The 2.5 th percentile of the death draws
location_name	Name of the estimation geography

Covariates values for COVID testing rates per capita. Reported and estimated daily.

Model Parameters

- beta_scaling_parameters.csv

Column Name	Contents
location_id	Unique identifier for estimation geography; common identifier across various files in this reference
scaling_parameter	Categorical field describing the parameter. Consists of “deaths”, “window_size”, “fit_final”, “pred_start”, “scale_init”, “history_days_start”, “history_days_end”, “log_beta_residual_mean”, “draw”, “log_beta_residual_mean_offset”, and “scale_final”
mean	The mean value of the scaling parameter
upper	The 97.5 th percentile of draws of the estimate
lower	The 2.5 th percentile of draws of the estimate
location_name	Name of the estimation geography

The calculated adjustments to forecasted betas based on the log_beta_residuals and the algorithm described in the Supplementary Information.

- coefficients.csv

Column Name	Contents
location_id	Unique identifier for estimation geography; common identifier across various files in this reference
covariate	Name of the relevant covariate. Names are consistent with corresponding covariate file name. Additionally, report “intercept”
mean	The mean value of the coefficient
upper	The 97.5 th percentile of draws of the estimate
lower	The 2.5 th percentile of draws of the estimate
location_name	Name of the estimation geography

The fitted values of the coefficients for each covariate by draw across all locations from the beta regression

- log_beta_residuals.csv

Column Name	Contents
location_id	Unique identifier for estimation geography; common identifier across various files in this reference
date	Format yyyy-mm-dd
mean	Mean value of the residual
upper	The 97.5 th percentile of draws of the estimate
lower	The 2.5 th percentile of draws of the estimate
location_name	Name of the estimation geography

The calculated difference between the observed betas from the first stage of the SEIR model and the predicted betas from the beta regression.

Estimated Outcomes

- betas.csv

Column Name	Contents
location_id	Unique identifier for estimation geography; common identifier across various files in this reference
date	Format yyyy-mm-dd
mean	The mean value of the estimate
upper	The 97.5 th percentile of draws of the estimate
lower	The 2.5 th percentile of draws of the estimate
location_name	Name of the estimation geography

The final adjusted forecasted betas based upon the beta regression and the adjustments from the scaling algorithm.

- cumulative_deaths.csv

Column Name	Contents
location_id	Unique identifier for estimation geography; common identifier across various files in this reference
date	Format yyyy-mm-dd
observed	A binary tag differentiating values derived from input data (i.e. observed == 1) versus those from estimates (i.e. observed == 0)
mean	The mean value of the estimate. When observed == 1, this value is derived from cumulative_elastipliner_smoothed.csv. When observed == 0, this value is derived from the SEIR model.
upper	The 97.5 th percentile of draws of the estimate
lower	The 2.5 th percentile of draws of the estimate
location_name	Name of the estimation geography

The cumulative total of estimated deaths due to COVID-19. Estimated daily.

- cumulative_infections.csv

Column Name	Contents
location_id	Unique identifier for estimation geography; common identifier across various files in this reference
date	Format yyyy-mm-dd
mean	The mean value of the estimate
upper	The 97.5 th percentile of draws of the estimate
lower	The 2.5 th percentile of draws of the estimate
location_name	Name of the estimation geography

The cumulative total of estimated infections due to COVID-19. Estimated daily.

- daily_deaths.csv

Column Name	Contents
location_id	Unique identifier for estimation geography; common identifier across various files in this reference
date	Format yyyy-mm-dd
observed	A binary tag differentiating values derived from input data (i.e. observed == 1) versus those from estimates (i.e. observed == 0)
mean	The mean value of the estimate. When observed == 1, this value is derived from cumulative_elastipliner_smoothed.csv. When observed == 0, this value is derived from the SEIR model.
upper	The 97.5 th percentile of draws of the estimate
lower	The 2.5 th percentile of draws of the estimate
location_name	Name of the estimation geography

The estimated daily number of deaths due to COVID-19. Estimated daily.

- daily_infections.csv

Column Name	Contents
location_id	Unique identifier for estimation geography; common identifier across various files in this reference
date	Format yyyy-mm-dd
mean	The mean value of the estimate
upper	The 97.5 th percentile of draws of the estimate
lower	The 2.5 th percentile of draws of the estimate
location_name	Name of the estimation geography

The estimated daily number of infections due to COVID-19. Estimated daily.

- r_effective.csv

Column Name	Contents
location_id	Unique identifier for estimation geography; common identifier across various files in this reference
date	Format yyyy-mm-dd
mean	The mean value of the estimate
upper	The 97.5 th percentile of draws of the estimate
lower	The 2.5 th percentile of draws of the estimate
location_name	Name of the estimation geography

The estimated effective reproductive number. Estimated daily.

Further Common Model Inputs

These data are available in the file: IHME_USA_COVID_2020_2021_FULL_DATA_Y2020M10D09.CSV

Column Name	Contents
location_id	Unique identifier for estimation geography; common identifier across various files in this reference
Province/State	Name of the specific estimation geography
Country/Region	Name of the Country or Region the specific estimation geography can be found in
Date	Format yyyy-mm-dd
Confirmed	Number of COVID-19 cases reported for the specific estimation geography, post revision for noted corrections. Blank cells denote no reported values on that day
Deaths	Number of COVID-19 attributed deaths reported for the specific estimation geography, post revision for noted corrections. Blank cells denote no reported values on that day
Hospitalizations	Number of COVID-19 attributed hospitalizations reported for the specific estimation geography, post revision for noted corrections. Blank cells denote no reported values on that day

Ad-hoc corrections to the raw data time series are available in this file:

IHME_USA_COVID_2020_2021_RAW_DATA_CORRECTIONS_Y2020M10D09.CSV

Column Name	Contents
location_id	Unique identifier for estimation geography; common identifier across various files in this reference
location	Name of the estimation geography
date	Format dd.mm.yyyy
value	Number of corresponding units flagged for revision
unit	COVID-19 metric affected. Limited to "cases", "deaths", "hospitalizations"
note	Open field providing context for why the revision took place

Mandates orders and dates of enactment and easing are available in this file:
 IHME_USA_COVID_2020_2021_MANDATES_CLOSURES_REOPENING_Y2020M10D09.CSV

Column Name	Contents
location	Name of the estimation geography
location_id	Unique identifier for estimation geography; common identifier across various files in this reference
date_checked	Last day of data seeking refresh. Format dd.mm.yyyy
Any Gathering Restrictions	Date on which any gathering restriction was enacted (regardless of threshold gathering size, or focus on particular gathering type). Format dd.mm.yyyy. When a mandate has not been imposed this is denoted by "not implemented". When this mandate is immediately replaced by a "Stay at Home" order this is denoted by "full implementation"
Source Any Gathering Restrictions	Reference and link to source used to determine date of enactment of order
Stay at Home	Date on which a Stay at Home order was enacted. Format dd.mm.yyyy. When a mandate has not been imposed this is denoted by "not implemented".
Source Stay at Home	Reference and link to source used to determine date of enactment of order
Schools Close	Date on which schools (including K-12 grades, and higher education entities) were mandated to close. Format dd.mm.yyyy. When a mandate has not been imposed this is denoted by "not implemented".
Source Schools Close	Reference and link to source used to determine date of enactment of order
Any Business Closures	Date on which the first restrictions on any type of business were mandated. Format dd.mm.yyyy. When a mandate has not been imposed this is denoted by "not implemented". When the first mandates imposed on businesses were coincident with the imposition of an all non-essential business closure this is denoted by "full implementation"
Source Business Closures	Reference and link to source used to determine date of enactment of order
Non-Essential Business Closures	Date on which all non-essential businesses were mandated to close. Format dd.mm.yyyy. When a mandate has not been imposed this is denoted by "not implemented".
Source Non-Essential Business Closures	Reference and link to source used to determine date of enactment of order
Travel Restrictions	Date on which severe local restriction on travel were implemented. This includes, but is not limited to: reduction in public transit, restrictions to movement within and between municipalities, requirements for prior authorization for travel. Format dd.mm.yyyy. When a mandate has not been imposed this is denoted by "not implemented".

Column Name	Contents
Source Travel Restrictions	Reference and link to source used to determine date of enactment of order
Lifting Travel	Date on which prior Travel Restrictions were ended. Format dd.mm.yyyy. If this mandate had not been previously imposed this is denoted by "not implemented". If this mandate is still being enacted this is denoted as "still implemented"
Source Lifting Travel	Reference and link to source used to determine date of enactment of order
Ease Education Closures	Date on which prior Schools Close restrictions were eased. This can include, but is not limited to, reopening of only a portion of a sector (e.g. specific grades allowed back to school), or orders that allow for some, but not all, students to attend in-person schooling. Format dd.mm.yyyy. If this mandate had not been previously imposed this is denoted by "not implemented". If this mandate is still being enacted this is denoted as "still implemented"
Source Ease Education Closures	Reference and link to source used to determine date of easing of order
Open Education	Date on which prior Schools Close restrictions were ended. Format dd.mm.yyyy. If this mandate had not been previously imposed this is denoted by "not implemented". If this mandate is still being enacted this is denoted as "still implemented"
Source Open Education	Reference and link to source used to determine date of ending of order
Ease NE Business Closures	Date on which prior Non-Essential Business Closures were eased. This includes, but is not limited to, mandates that allow some, but not all, previously closed businesses to reopen fully. Format dd.mm.yyyy. If this mandate had not been previously imposed this is denoted by "not implemented". If this mandate is still being enacted this is denoted as "still implemented"
Source Ease NE Business Closures	Reference and link to source used to determine date of easing of order
No Business Closure	Date on which prior business closures were entirely ended. Format dd.mm.yyyy. If this mandate had not been previously imposed this is denoted by "not implemented". If this mandate is still being enacted this is denoted as "still implemented"
Source No Business Closure	Reference and link to source used to determine date of ending of order

Column Name	Contents
Ease Stay At Home	Date on which prior Stay At Home orders were eased. This includes, but is not limited to, mandates that allow some, but not all, previously required individuals to stay at home, to no longer be required to do so. Format dd.mm.yyyy. If this mandate had not been previously imposed this is denoted by “not implemented”. If this mandate is still being enacted this is denoted as “still implemented”
Source Ease Stay At Home	Reference and link to source used to determine date of easing of order
No Gathering Restrictions	Date on which prior gatherings restrictions were entirely ended. Format dd.mm.yyyy. If this mandate had not been previously imposed this is denoted by “not implemented”. If this mandate is still being enacted this is denoted as “still implemented”

Data Input Sources

This file contains relevant metadata about the input sources as suggested in the [Guidelines for Accurate and Transparent Health Estimates Reporting \(GATHER\)](#), a statement that promotes best practices in reporting health estimates.

IHME_USA_COVID_2020_2021_DATA_INPUT_SOURCES_Y2020M10D22.PDF

Additional Information

Terms and Conditions

<http://www.healthdata.org/about/terms-and-conditions>

Contact information

To request further information about this dataset, please contact IHME:

Address:

Institute for Health Metrics and Evaluation
Population Health Building/Hans Rosling Center
3980 15th Ave. NE, Seattle, WA 98195 USA
UW Campus Box #351615

Telephone: +1-206-897-2800

Fax: +1-206-897-2899

Email: data@healthdata.org

Website: www.healthdata.org

These files may be updated periodically, so we appreciate hearing feedback or additional information about how these data are being used.