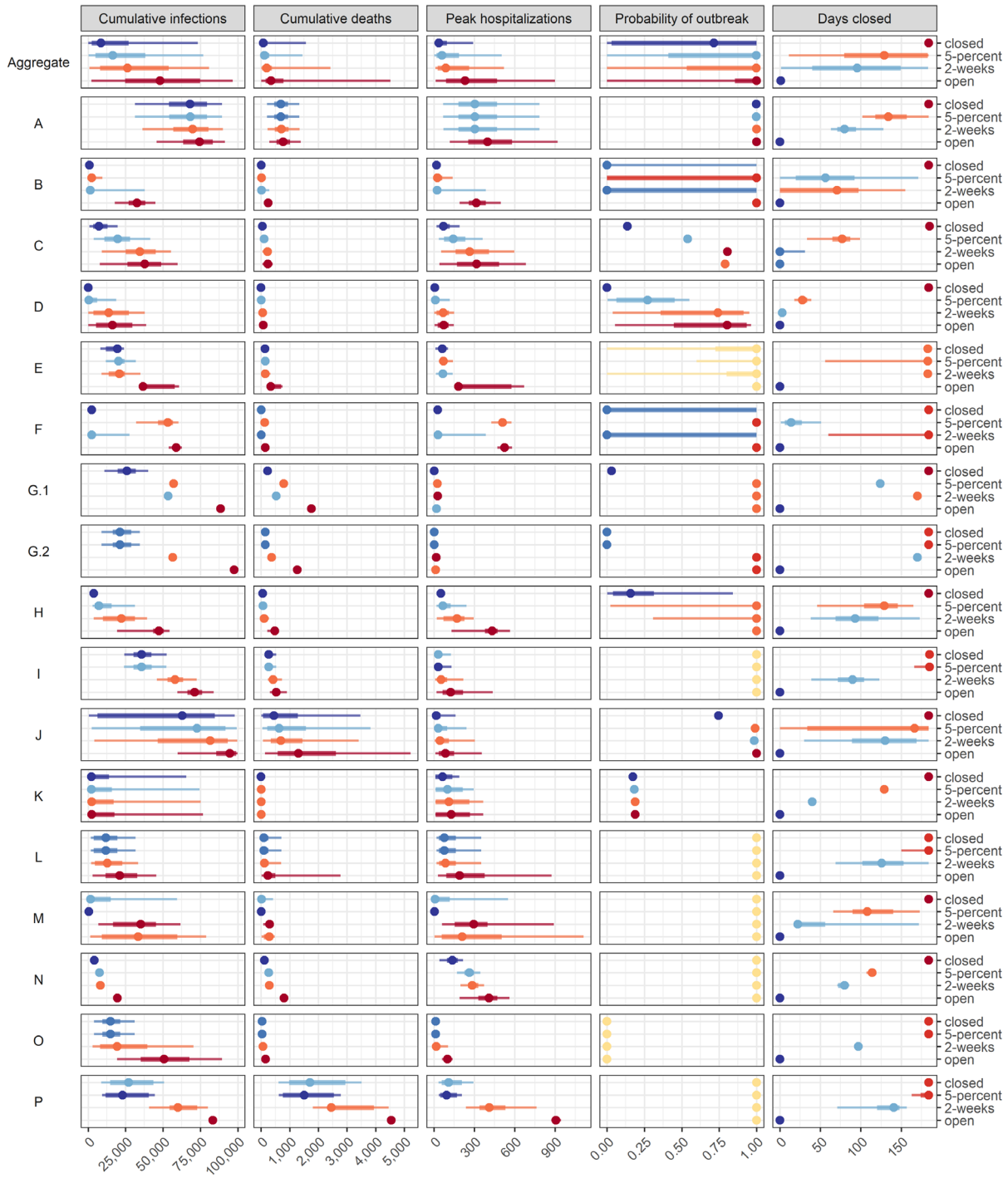
Supplemental Material (SM)

Supplemental Figures, Tables, Model Descriptions, and other materials

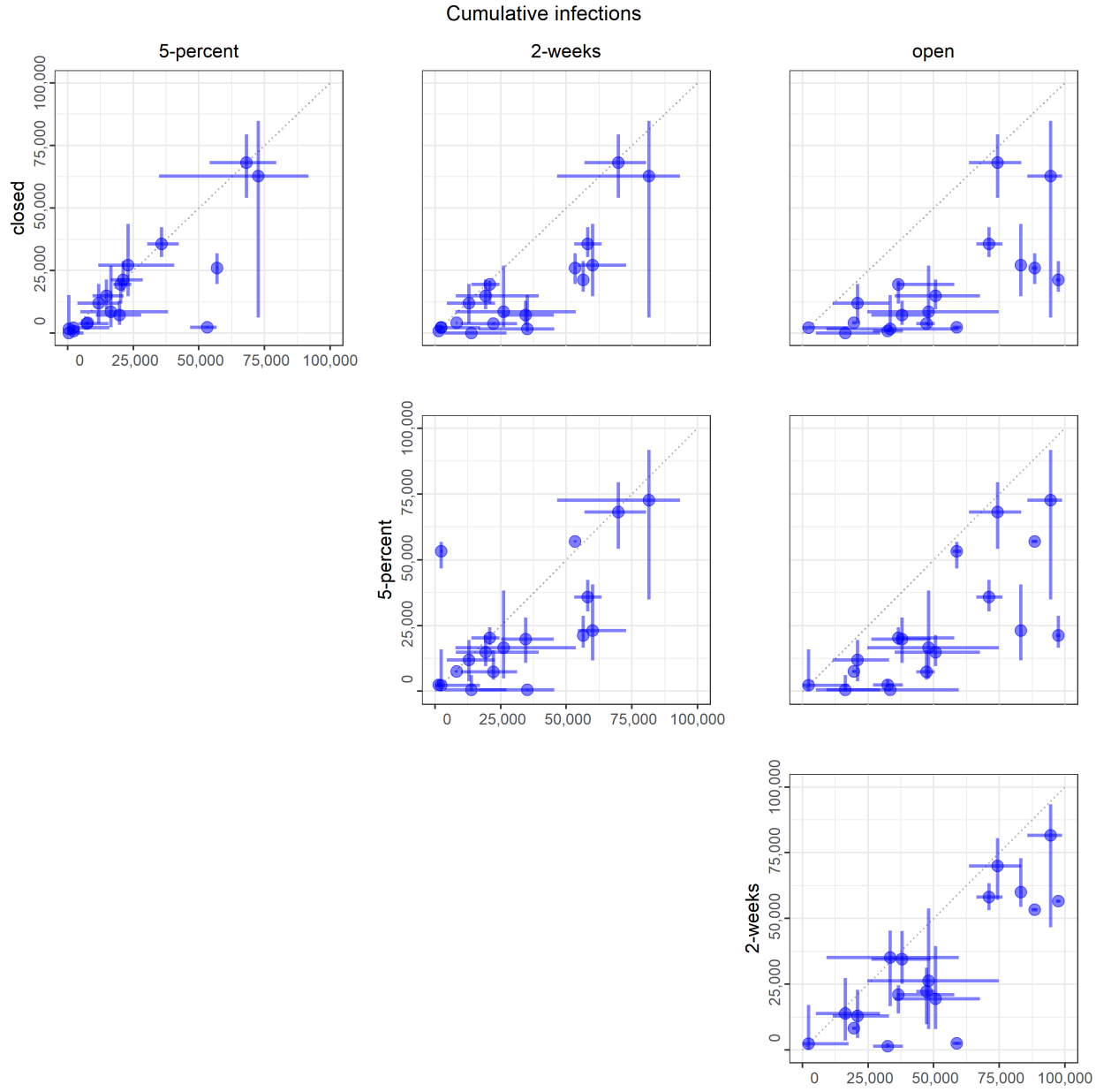
Supplementary Material TOC:

Supplementary Material: Anonymized results from round 2 (Figs. S1 to S8 and Video 1).

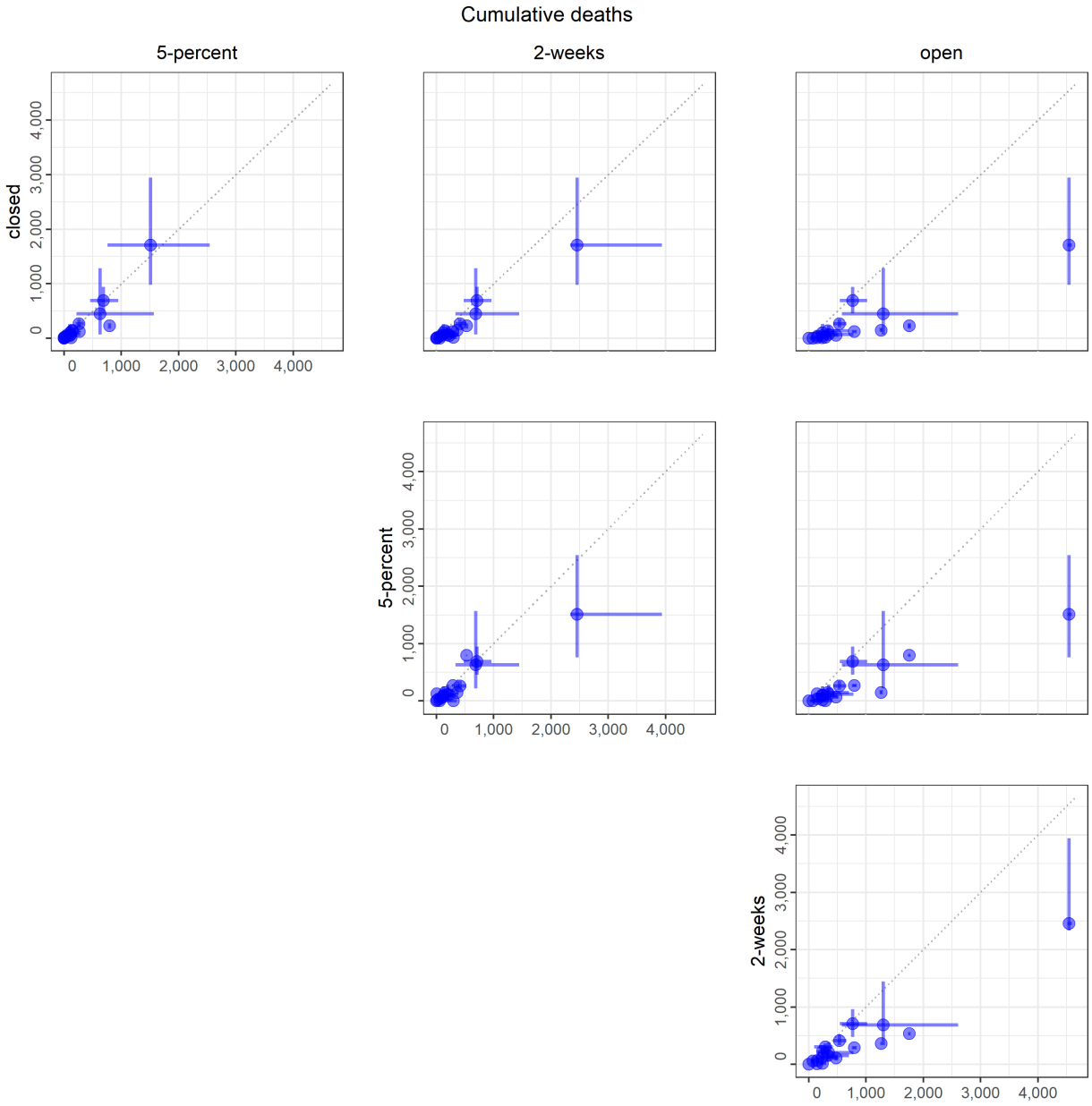


SM Fig. 1: Model results for each target objective and intervention scenario pair organized by model. Median, 50% prediction interval (PI), and 90% PI are indicated as points, thick lines, and thin lines, respectively. Colors denote ranking of each intervention by model for a single objective, where dark blue signifies the lowest value (best performance) and dark red signifies

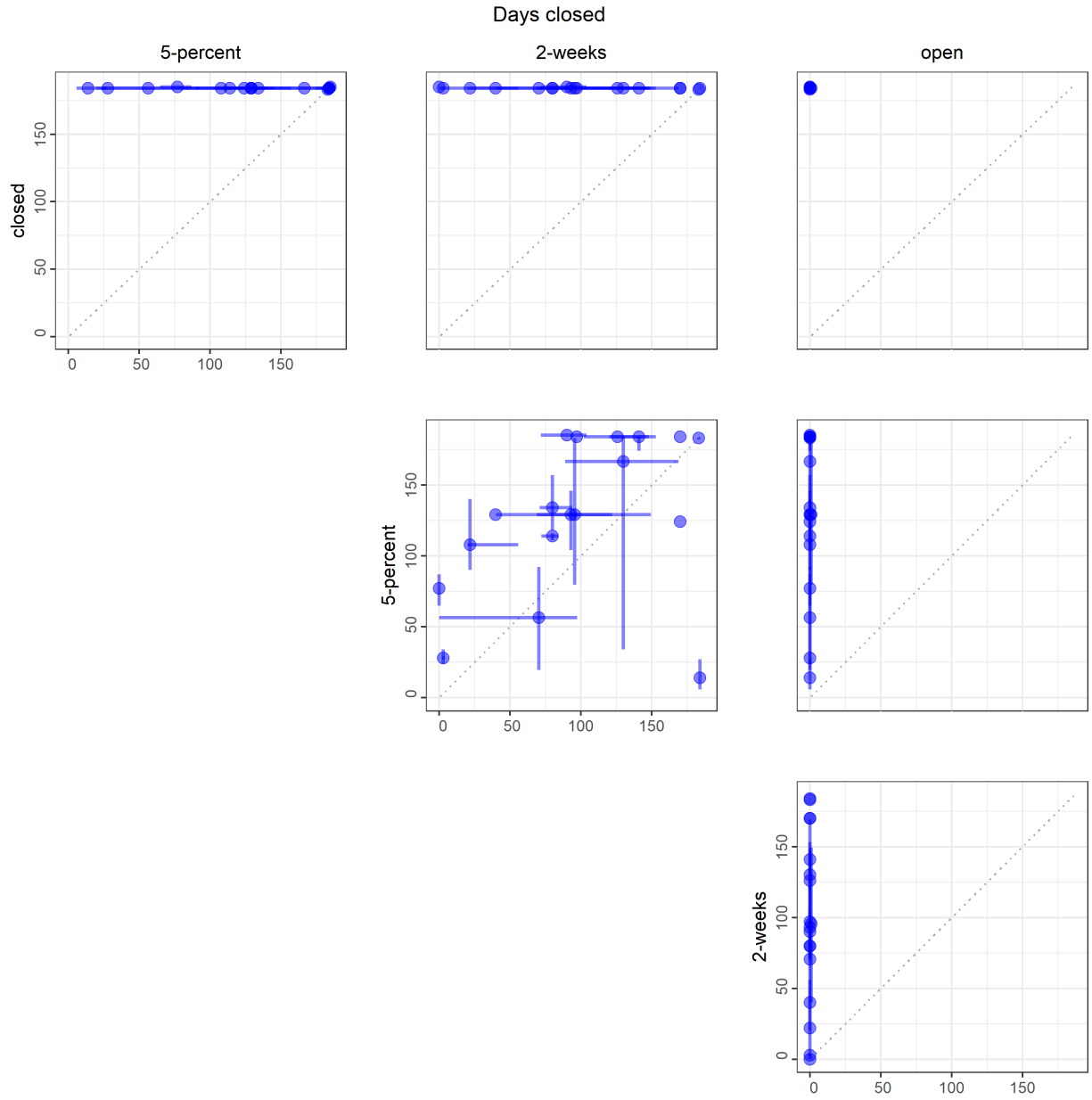
the highest value (worst performance). Ties in ranks are colored as intermediate values. A tie between ranks 1 and 2 and ranks 3 and 4 are shown as an intermediate blue and red, respectively; yellow indicates a tie in ranks across all interventions. Each group is assigned a random, unique identification letter that is specified on the vertical axis.



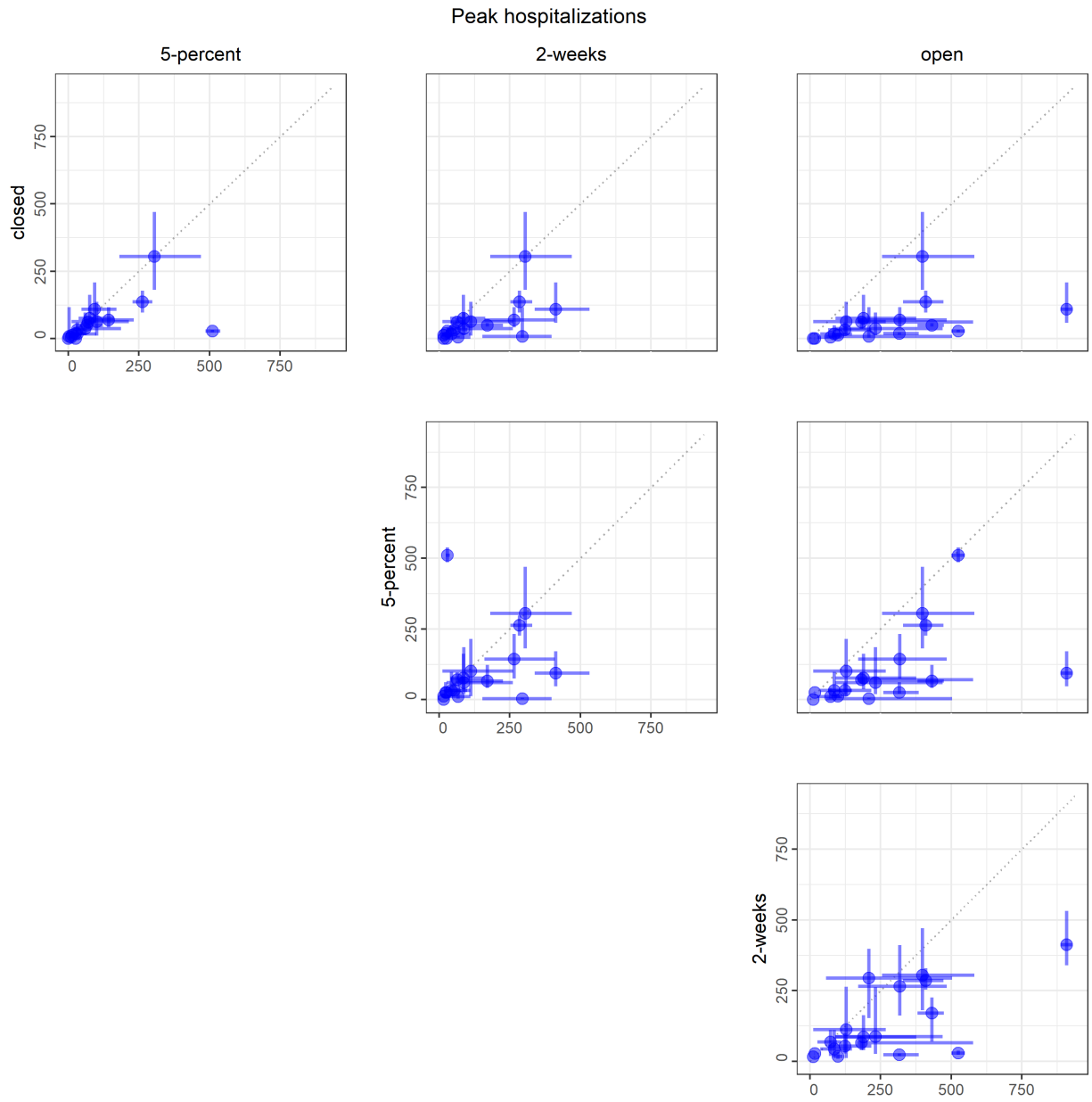
SM Fig. 2: Cumulative infections. Medians (points) and 50% PIs (lines) displayed pairwise by intervention scenario. Each point represents one model.



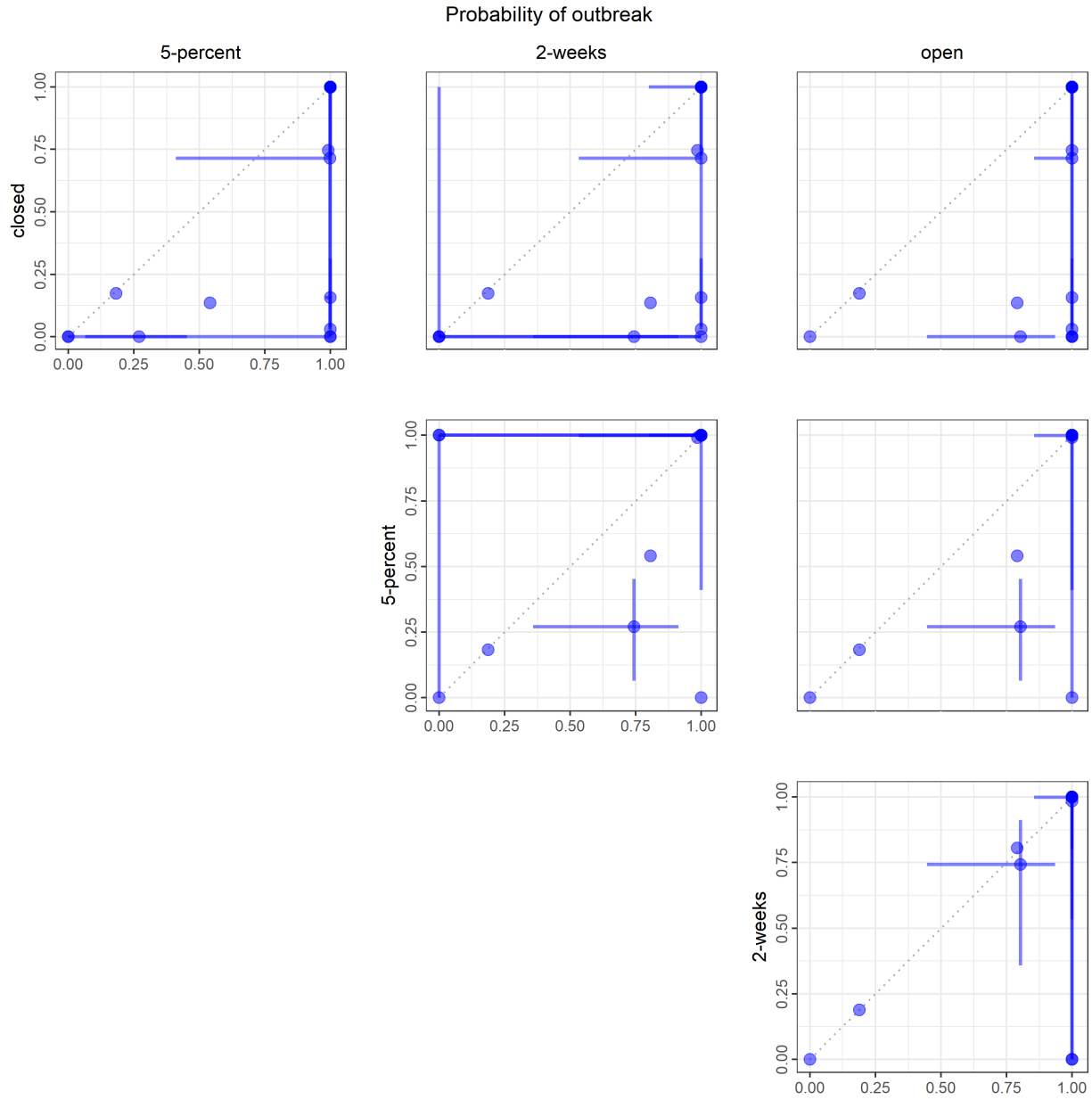
SM Fig. 3: Cumulative deaths. Medians (points) and 50% PIs (lines) displayed pairwise by intervention scenario. Each point represents one model.



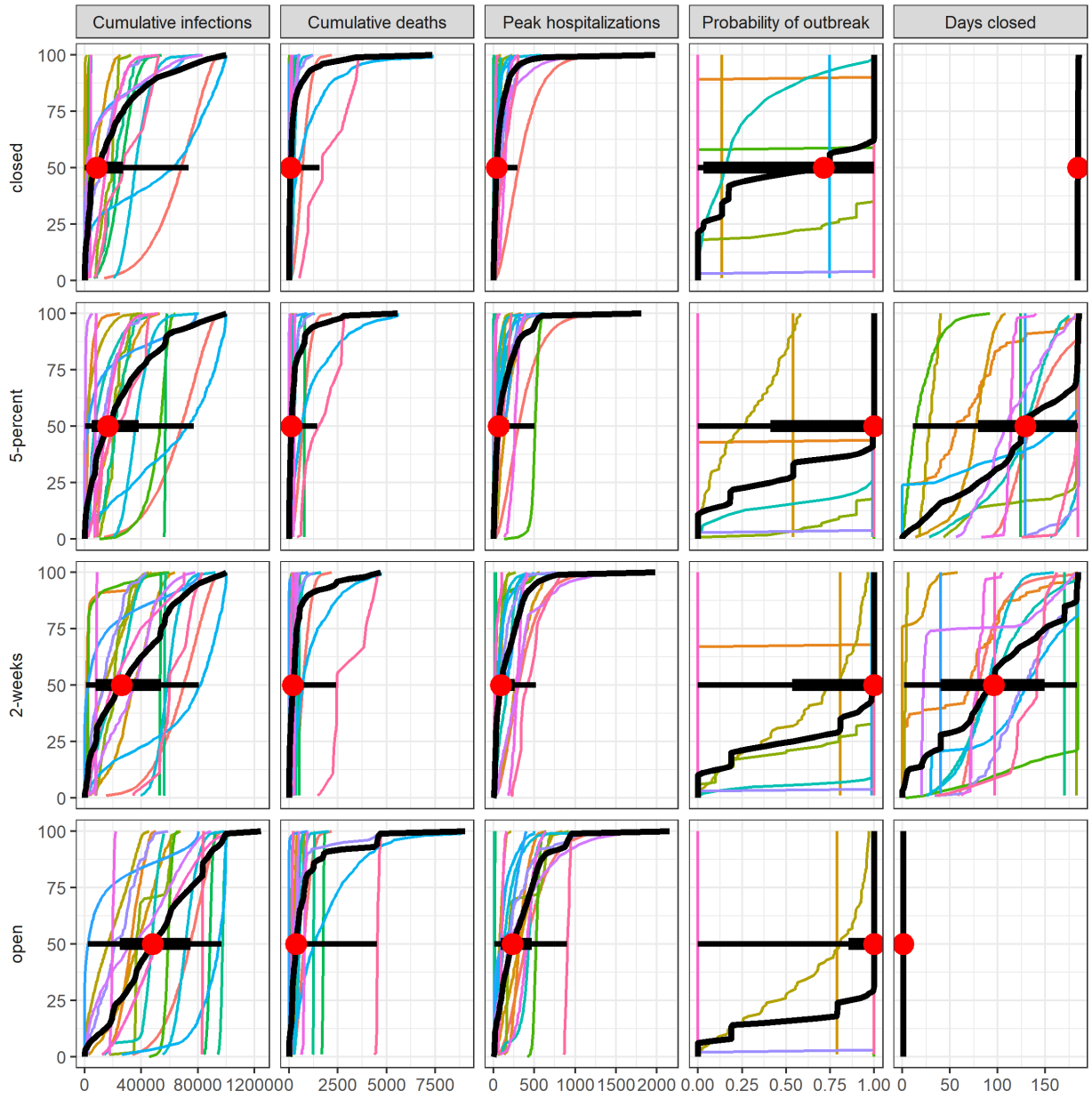
SM Fig. 4: Days closed for non-essential workplaces. Medians (points) and 50% PIs (lines) displayed pairwise by intervention scenario. Each point represents one model.



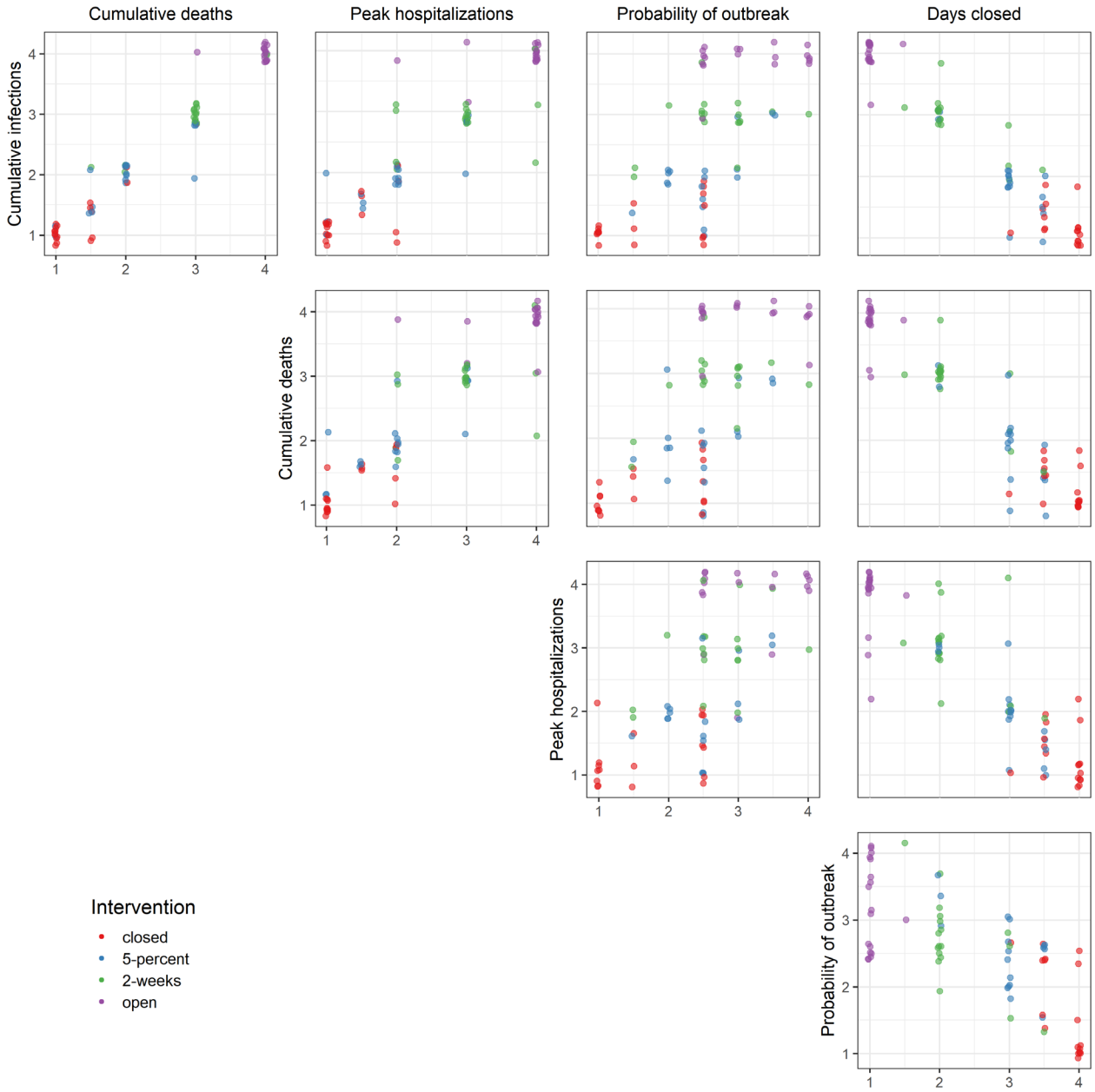
SM Fig. 5: Peak number of hospitalized cases. Medians (points) and 50% PIs (lines) displayed pairwise by intervention scenario. Each point represents one model.



SM Fig. 6: Probability of outbreak. Median (points) and 50% PI (lines) displayed pairwise by intervention scenario. Each point represents one model.



SM Fig. 7: Cumulative distribution functions (CDFs) across models and for the aggregate. Each colored line shows the quantile distribution for a single model. The aggregate CDF is shown in black with median, 50% PI, and 90% PI indicated as red points, thick lines, and thin lines respectively.



SM Fig. 8: Scatter plots of intervention ranks for a given pair of objectives. Rank ties are shown as intermediate numerical values (e.g. a tie between 1 and 2 is shown as 1.5). For visual clarity, shaded points are jittered around the discrete rank values.

Supplementary Material: Resolution of linguistic uncertainty in structured discussion between rounds 1 and 2.

The group discussion between modeling rounds is an efficient method to reduce linguistic uncertainty resulting from differing interpretations of the problem setting. As well as allowing a common definition of "peak" and other terms, as described in the main text, other sources of unanticipated uncertainty were resolved. For example, one modeling group asked for clarification on the definition of 'death.' There was a thorough discussion of the options that different groups had considered or used (reported only; reported plus probable; reported, probable and co-morbidities; or, also indirect deaths, such as those from unrelated causes in patients choosing not to go to the ER during a pandemic). We agreed as a group to use all deaths due to COVID-19 disease-induced mortality, regardless of reporting. This way of counting deaths is based on infection status, not testing status, and can include comorbidities but not indirect deaths, as we are only focusing on people who have been infected with SARS-CoV-2 and died from their infection.

The first round also provided some important checks and balances on the consistency of objective and intervention interpretations across groups, i.e., were the same definitions of workplace closures used? In the first round, some groups used the May 15 to Nov 15 timeframe, others based start dates on declarations of a State of Emergency or stay-at-home orders, and one group implemented a weighting for essential and non-essential business closures and associated compliance issues explicitly (SM Fig. 9). Including a metric that should be consistent across models allowed us to check for and remove linguistic uncertainty in round 2 submissions that would have limited our ability to compare the rankings of interventions between models and objectives.

In addition to resolving linguistic uncertainty, the first round provided information on the utility of the interventions themselves. We initially requested results for reopening after declining to 1% of peak – round 1 results suggested this condition would rarely, if ever, be met, and thus we altered the intervention to trigger at 5% of peak, instead. Typically, such changes in interventions would be made in consultation with decision makers (as part of Fig. 1, loop A).

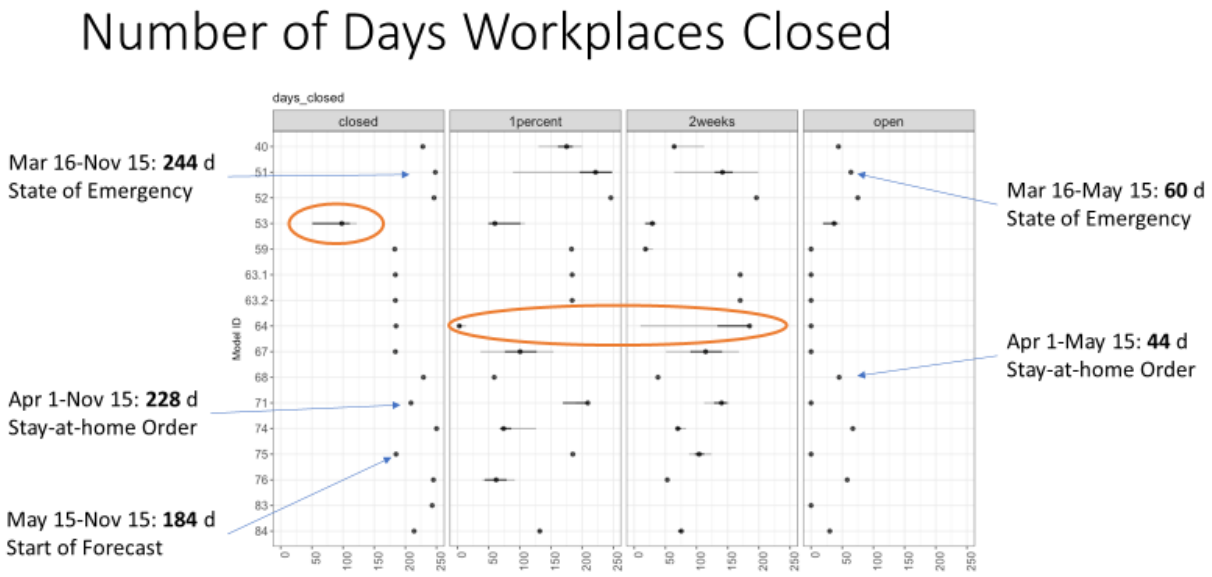
Deliberately, consensus on scientific uncertainty was not required. In fact, model results were presented anonymously to reduce the pressure to conform to other groups' expectations and hence to avoid 'groupthink,' and other cognitive biases, engendering a more comprehensive expression of legitimate scientific uncertainty. We thus encouraged modeling groups to adjust their models to reflect unknown aspects of the transmission and intervention implementation process to more fully express genuine scientific and logistical uncertainty.

Due to the opposing effects of decreasing linguistic uncertainty and maintaining or increasing expression of scientific uncertainty, it was difficult to draw conclusions about the source of model-level changes in expressions of uncertainty between rounds 1 and 2. To begin to assess model-level changes, we compared the lengths of inter-quartile ranges (IQRs) (SM Figs 10-12) within groups by round as well as the ratio of IQR length between each model and the corresponding aggregate distribution. The clearest examples of model incorporation of additional scientific uncertainty in round 2 were the models that provided point estimates in round 1 (length of IQR = 0) that subsequently expanded these estimates to distributions in round 2. Requiring distributions rather than point estimates necessarily increased the degree of expressed

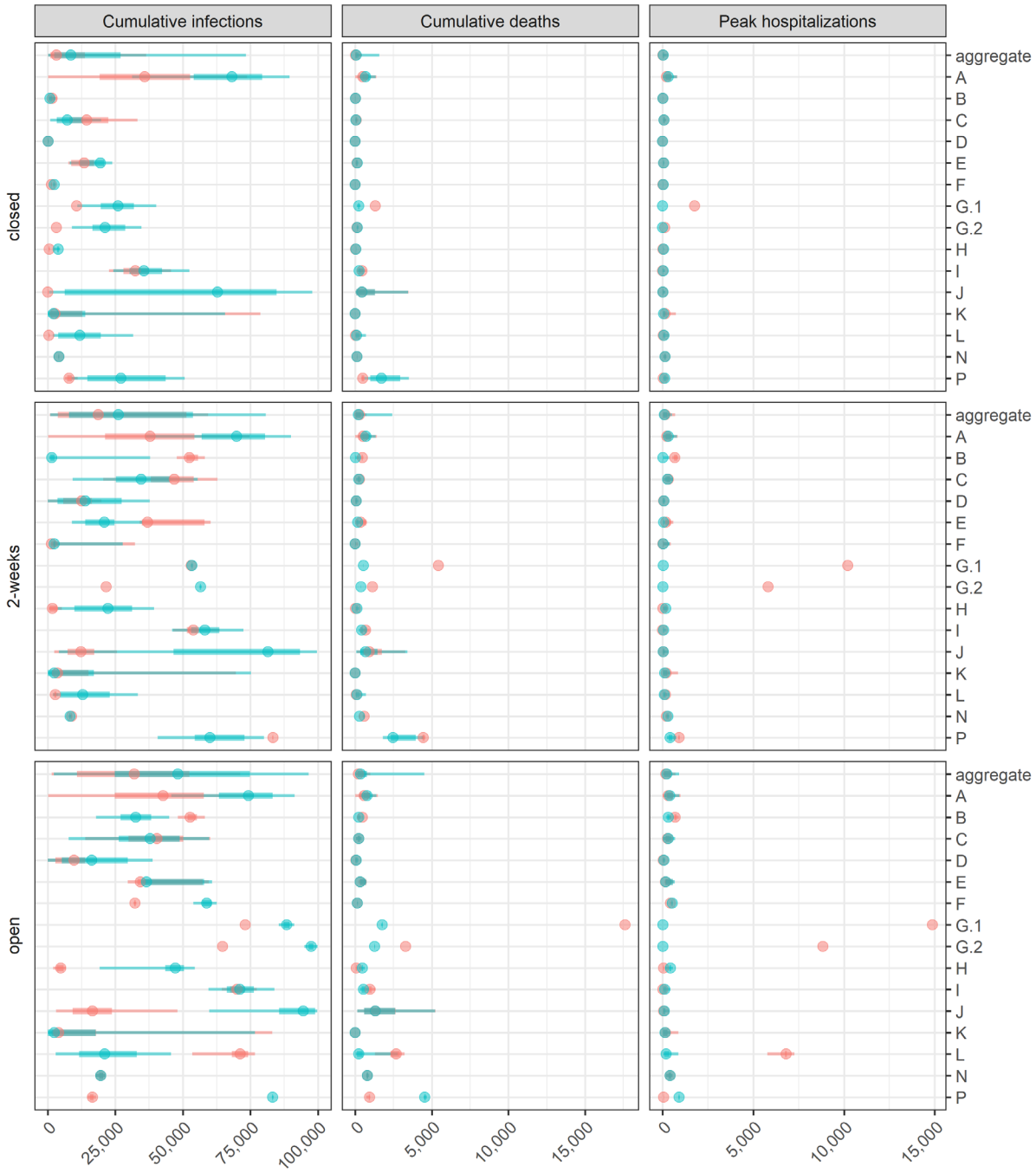
uncertainty. However, even in these models, we observed decreases in uncertainty (presumably in linguistic uncertainty) as the point estimates account for the majority of outliers in round 1 (SM Fig. 10).

For each objective-intervention pair considered in both rounds, the length of the aggregate IQR was greater than the median length of the corresponding model IQRs (SM Fig. 11). The degree of uncertainty (as measured by IQR lengths) for the majority of models increased towards that of the corresponding objective-intervention aggregate distribution from round 1 to round 2 (see the clustering of points near the orange dashed line in SM Fig. 11 round 2).

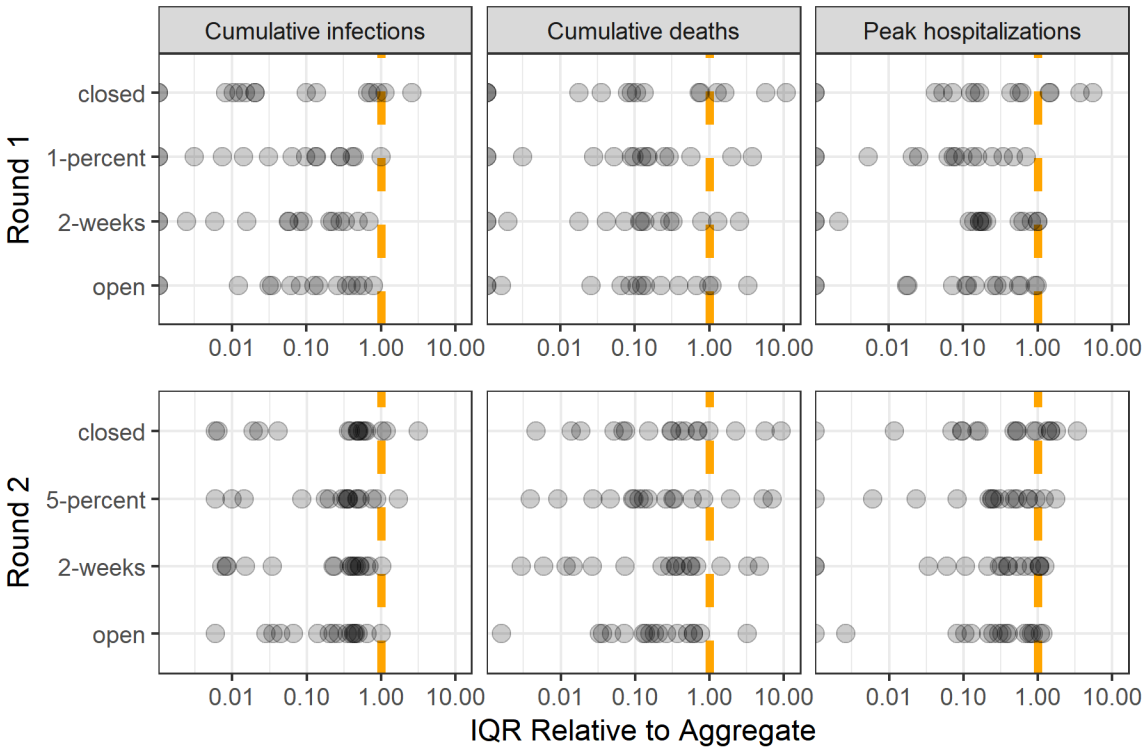
Implementation of the open and closed interventions did not rely on a definition of "peak". In SM Fig. 12, we observed that the ratio of IQRs (IQR(model)/IQR(aggregate)) between rounds tended to be closer to one than the 2-week intervention, which required a definition of peak (SM Fig. 12). We also note that decreases in the IQR length for the aggregate distribution were observed for all objectives in the 2-week scenario (i.e. aggregate ratio of $IQR_{R2}/IQR_{R1} < 1$). Changes observed in the open scenario (cumulative infections, cumulative deaths, and peak hospitalization ratios observed are 1.20, 1.02, and 0.949 respectively) were moderate compared to those in the closed scenario (cumulative infections, cumulative deaths, and peak hospitalization ratios observed are 1.93, 1.92, and 1.54 respectively). Note that an analogous comparison for the alternative peak intervention was not possible, given the change from 1-percent of the peak to 5-percent of the peak between rounds.



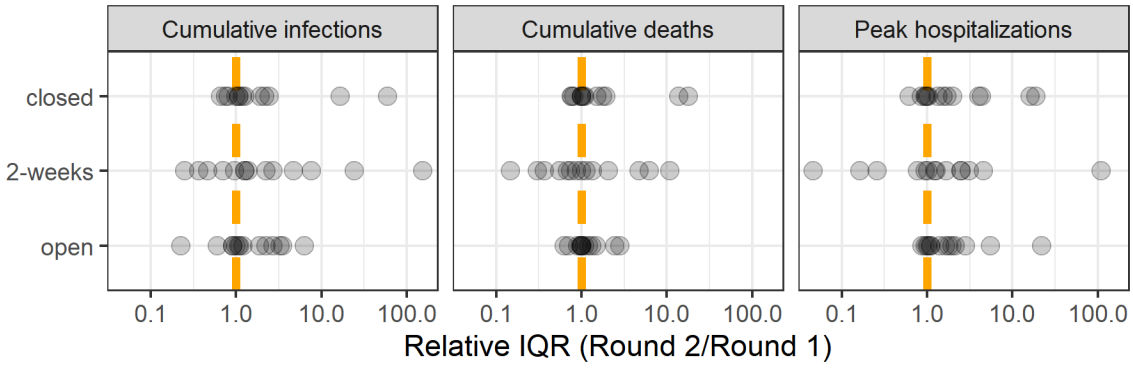
SM Fig. 9: Resolution of linguistic uncertainty about the number of days non-essential workplaces are closed in the discussion following round 1 of modeling (note that model IDs changed between rounds); figure is of a slide from the group discussion after round 1. See main text Fig. 3 column 5 for days of non-essential workplace closure results from round 2. Ovals highlight points of discussion about different ways of capturing uncertainty for days workplaces are closed and unusual results about intermediate interventions.



SM Fig. 10: Team and aggregate values for each intervention and objective pair. Round 1 and round 2 results displayed in red and blue respectively. Since the 1-percent intervention from round 1 was updated to a 5-percent intervention in round 2, results for these interventions have been omitted from this comparison. Also note that two models were excluded from this analysis, as they submitted incomplete results in round 1. After the discussion between rounds 1 and 2, these groups were able to provide complete and comparable results. Additionally, in at least one case, some of the differences can be attributed to model error fixes between rounds.



SM Fig. 11: Comparison between model-specific IQR lengths and the length of the IQR for the aggregate distribution (i.e. length(IQR_team)/length(IQR_aggregate)) shown on a logarithmic scale. Results are grouped by round, intervention, and objective. Round is indicated on the left axis. Columns indicate the objective and rows indicate the intervention. The dashed orange line highlights the point at which there is no difference between the model-specific IQR lengths between rounds 1 and 2 (points to the left indicate a model IQR less than that of the corresponding aggregate and *vice versa* for points to the right).



SM Fig. 12: Round comparison of IQR length by team, calculated as the ratio of the length of IQRs between rounds 1 and 2 (i.e. length(IQR_{R2}) / length(IQR_{R1})) shown on a logarithmic scale. Note that in the first round, two models (G.1 and G.2) submitted point estimates for each intervention and metric. Since point estimates are such that length(IQR) = 0, the relative IQR (compared to round 1) is infinity and thus not shown here. Similarly, there is not a point representing cumulative deaths in the closed scenario for group K since the corresponding length(IQR) = 0. Because the 1-percent intervention from round 1 was changed to a 5-percent intervention in round 2, the corresponding results have been omitted from this comparison. The dashed orange line highlights the point at which there is no difference between the model-specific IQR lengths between rounds 1 and 2 (points to the left indicate a lower R2 IQR than that of the corresponding group's R1 submission, and vice versa for points to the right).

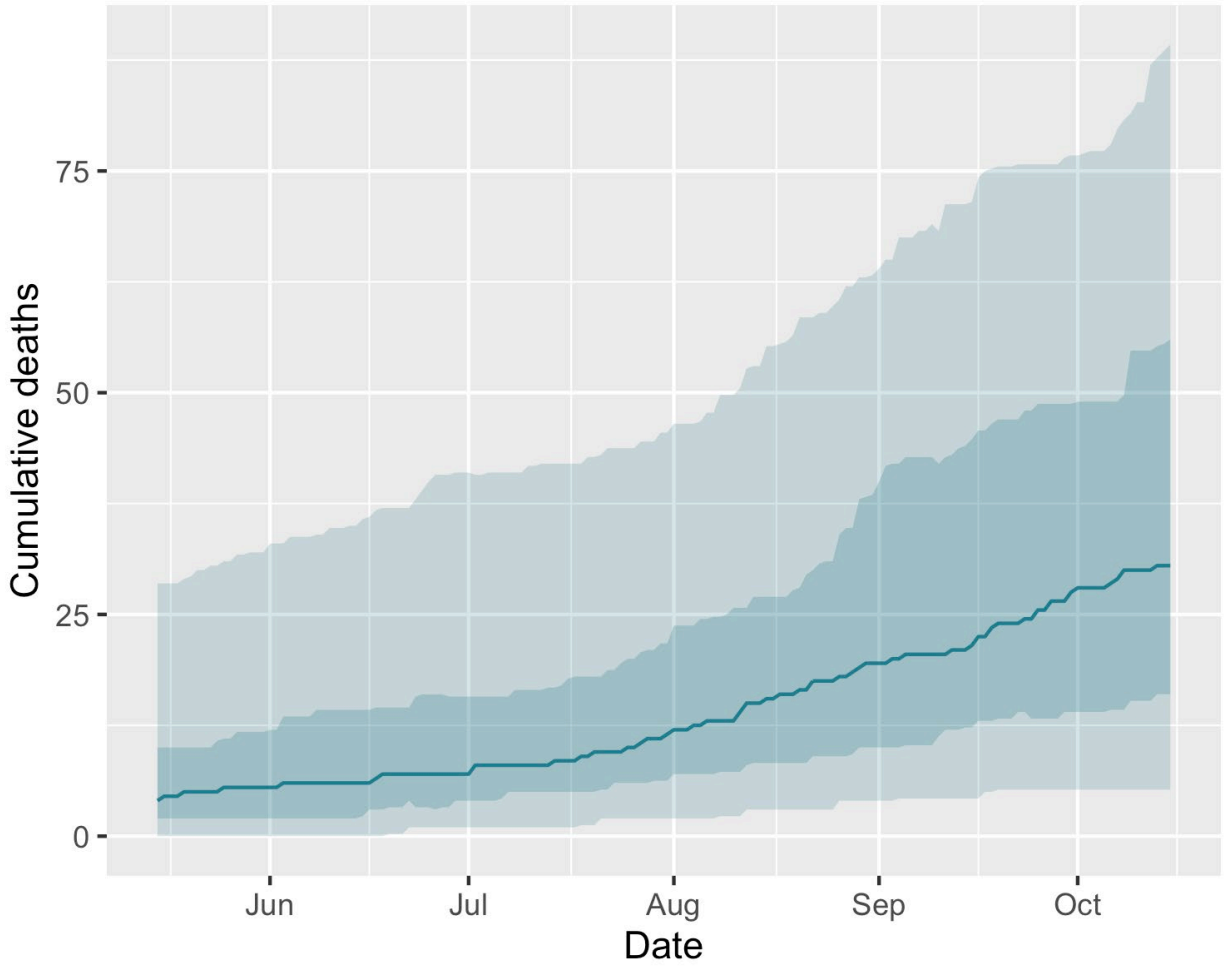
Supplementary Material: Comparison of county data with aggregate model results

The modeling exercise was motivated by a U.S. county representative of mid-sized counties with populations of approximately 100,000 people, with limited mobility and stay at home orders in place until at least May 15, 2020. Here, we compare aggregate model results with reported data from counties meeting the target county description.

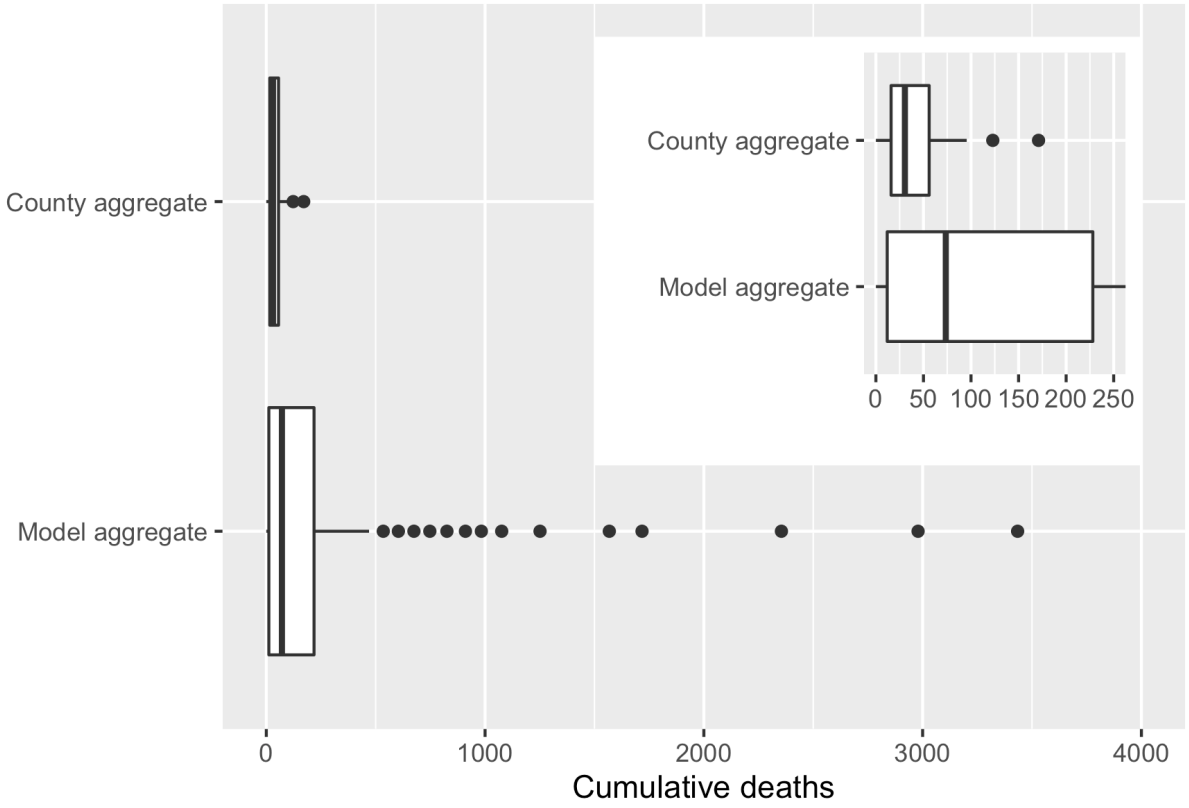
We first selected the 99 U.S. counties with population sizes between 90,000 and 110,000 using data from the Johns Hopkins University COVID-19 dashboard (Dong, Du, and Gardner 2020). From this subset, we then selected counties with stay at home orders in place until at least May 15, 2020 (data from Killeen et al. (2020), Keystone Strategy (2020), and NACO county emergency declarations (2020)), and changes in mobility in line with stay at home orders, i.e., less than 50% increase from baseline retail mobility, less than 25% increase in baseline work mobility, and less than 5% decline from baseline residential mobility (data from Google COVID-19 Community Mobility Reports). This resulted in a subset of 84 counties. Finally, from this subset, we determined the set of counties implementing a fully 'closed' intervention (with stay at home orders in place from May 15, 2020 to present day (October 15, 2020 as of submission) and mobility patterns suggesting those orders were followed). We found 66 counties that met these criteria. No counties were found to be fully open during this period, and it was not possible to determine if any counties implemented the '2-week' or '5-percent' interventions. We compared aggregate cumulative deaths (reported deaths only) with modeled cumulative deaths (all COVID-19 deaths) under the closed intervention for the 66 counties following the 'closed' intervention. Cumulative reported deaths for the 66 counties under the closed intervention were summarized in 100 quantiles, the same format requested from model groups (See SM Figs 13 - 14, below).

Note that model results were for the entire period May 15, 2020 to November 15, 2020 and data were only available through the present day (October 15, 2020 as of submission). Further, we are comparing *reported* deaths (from data) with *all* COVID-19 deaths, not just reported deaths (from model results). We did not assume a reporting rate, but expect a higher number of model predicted cumulative deaths. Crucially, our results represent the realization of one pandemic across 66 counties in comparison to multiple model realizations across a wide range of uncertainty. Thus, the model uncertainty will necessarily be higher than the observed uncertainty. The model mean will likely also be higher, as the right-skewed uncertainty will increase the mean.

County comparison results provided in the SM will be updated once data are available for the entire period May 15, 2020 to November 15, 2020, for counties that remain closed.



SM Fig. 13: Summary of cumulative reported deaths for counties similar to the model context and following the closed intervention. Median reported cumulative deaths (solid line), 50% IQR (darker shaded area), and 90% IQR (lighter shaded area) for the subset of 66 counties following the closed intervention from May 15 to October 15.



SM Fig. 14: Comparison of aggregate reported county data to model results for the closed intervention. Boxplot of cumulative reported deaths from 66 U.S. counties from May 15 to October 15 (median: 31; 50% IQR: 16, 56; 90% IQR: 3, 59) and model results for cumulative deaths from May 15 to November 15 (median: 73; 50% IQR: 12, 228; 90% IQR: 2, 1568). Inset shows overlap of box area for the two plots.

Supplementary Material: Checklist Data

SM Table 1: Contributed model descriptions. Name, description (including links to model code where available), diagram, calibration method, other non-pharmaceutical interventions (NPIs) included in the model, additional data sources used, previous use cases for the model (both for COVID-19 in other settings and other disease systems) and references for each of the 17 models. Categories which were not relevant were excluded.

CoMo Collaborative COVID-19 Model	
Description	Age-structured, SEIR compartmental model with infected compartments stratified by symptoms, severity and treatment seeking and access. Code available: https://github.com/ocelhay/como
Additional NPIs included	Social distancing, Isolation (post infection), Stay-at-home (voluntary), Handwashing, Travel ban
Additional data sources used	National data on hospital, ICU, and ventilator availability; Data on U.S. household size from the American Community Survey; Data from China and New York City for healthcare parameterization; Age-structure mixing matrices for Work, School, and Home from Prem, Cook, and Jit (2017); Vital surveillances from The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team (2020); NYS Governor Cuomo Daily presentation (April 23, 2020); List of countries by hospital beds (Wikipedia)
Previous use cases	RSV in Thailand
References	Covid-19 International Modelling Consortium 2020
Co-authors	Ricardo Agus, Lisa White, Nathaniel Hupert (PI)
Acknowledgements	Wirichida Pan-Ngun, PhD, Olivier Celhay, Vruj Patel, Lior Shtayer
Covasim	
Description	Stochastic agent-based model, including age-structured mixing, susceptibility to infection and health outcomes; transmission networks in different social layers; variable intrahost viral dynamics; presymptomatic, symptomatic, and asymptomatic transmission; hospitalizations (regular and intensive care); and multiple non-pharmaceutical and testing interventions. Code available: https://github.com/institutefordiseasemodeling/covasim
Diagram	See Kerr et al. 2020, Fig. 1
Calibration	Parameters were calibrated by optimizing the L1 relative error norm of positive diagnoses, number of deaths, and number of tests using global optimization package, *Optuna*
Additional NPIs included	Social distancing, Isolation (post infection), and school, workplace, and community closures based on stay-at-home and state-of-emergency orders.
Previous use cases	COVID-19 in Africa, Europe, Oceania, and North America
References	Kerr et al. 2020
Co-authors	Rafael C. Nez, Katherine Rosenfeld, Gregory R. Hart, Daniel Klein, Cliff C. Kerr (PI)
Acknowledgements	Dina Mistry, Prashanth Selvaraj, Jamie A. Cohen, Michael Famulare, Robyn M. Stuart, Romesh Abeysuriya
EvoNet SARS2	
Description	Stochastic, place-based model in which agents travel to different locations within the community. Agents can infect others within homes, schools, workplaces and other regular gathering spots, as well as during random walks within the community.

Calibration	Created ~1000 parameter sets each with 52 uniformly distributed parameters. For each parameter set, we considered a range of transmission probabilities. Interventions were simulated for parameter sets for which case and death counts came within range of the county data up to May 15th.
Additional NPIs included	Social distancing, Quarantine (post exposure), Isolation (post infection), Stay-at-home (voluntary), Age-specific interventions (e.g., isolation of elderly)
Previous use cases	HIV
Co-authors	John Mittler (PI)
Acknowledgements	Joshua Herbeck, James Murphy, Neil Abernethy, Sarah Stansfield, Molly Reid, Steven Goodreau
Funding	NIH grants R01AI108490 and R01 GM125440
JHU-CDDEP Bayesian Three-stage ODE Model	
Description	Bayesian, mechanistic ODE-based compartmental model composed of three transmission stages with varied force of infection pre-lockdown, lockdown, post-lockdown. Each stage corresponds with lockdown phases and social distancing measures that might be imposed by public health policymakers.
Calibration	Bayesian inference was conducted using MCMC-based method was used to fit the model to confirmed cases and deaths. Parameter ranges were estimated form the posterior distribution. Prior distribution was assumed to be uniformly distributed
References	Lin et al. 2020
Co-authors	Gary Lin, Yupeng Yang, Eili Klein
Acknowledgements	Anindya Bhaduri, Max Pinz, and the U.S. Centers for Disease Control and Prevention (CDC) Modeling in Infectious Diseases Network
LANL1-EpiCast	
Description	Agent based model with communities, households, and workplaces
Calibration	Transmission rates were varied in burn in period (March to May) to try to model actual county statistics. Burn in transmission rates calculated by parameter testing the model were much higher than previous experience fitting covid-19 (0.43), and were thus scaled down to an assumptive 0.2. This is potentially due to very low testing rates in initial stages and the existence of many more cases than were validated, thus explaining apparent excessively rapid growth in case numbers.
Additional NPIs included	Social distancing, Quarantine (post exposure), Isolation (post infection), Stay-at-home (voluntary), Stay-at-home (mandatory, e.g., government-ordered)
Additional data sources used	CDC disease statistics; past model calibration experience for COVID
Previous use cases	Flu and smallpox in the U.S.
References	Germann et al. 2006; Halloran et al. 2008; Germann et al. 2019
Co-authors	Chrysm Watson Ross, Tim Germann, Geoffrey Fairchild, Sara Del Valle (PI)
Funding	The LANL team was partially funded by the Laboratory Directed Research Development Program at Los Alamos National Laboratory (20200698ER and 20200697ER) and was supported by the DOE Office of Science through the National Virtual Biotechnology Laboratory, a consortium of DOE national laboratories focused on response to COVID-19, with funding provided by the Coronavirus CARES Act. Los Alamos National Laboratory is operated by Triad National Security, LLC under Contract No. 89233218CNA000001 with the U.S. Department of Energy. The content is solely the responsibility of the authors and does not

	necessarily represent the official views of the sponsors. The funders had no role in study design, data collection, analysis, decision to publish, or preparation of manuscript.
LANL2-Age Structured ODE	
Description	Age-structured compartmental ODE model. Stochasticity is incorporated by selecting parameters randomly from uniform distributions for each run, where the parameter ranges are determined from literature.
Diagram	See Spencer 2020
Previous use cases	COVID-19 in New Mexico
References	Spencer 2020; Spencer et al. 2020
Co-authors	Rosalyn Cherie Rael, Julie Spencer, Isabel Crooker, Carrie Manore (PI)
Funding	The LANL team was partially funded by the Laboratory Directed Research Development Program at Los Alamos National Laboratory (20200698ER and 20200697ER) and was supported by the DOE Office of Science through the National Virtual Biotechnology Laboratory, a consortium of DOE national laboratories focused on response to COVID-19, with funding provided by the Coronavirus CARES Act. Los Alamos National Laboratory is operated by Triad National Security, LLC under Contract No. 89233218CNA000001 with the U.S. Department of Energy. The content is solely the responsibility of the authors and does not necessarily represent the official views of the sponsors. The funders had no role in study design, data collection, analysis, decision to publish, or preparation of manuscript. The publication is approved for release LA-UR-20-27777.
MESALab-FOSP	
Description	Integer order generalized SEIR compartmental models with power law infection rates and age structure
References	Guo et al. 2020
Co-authors	Lihong Guo, Yanting Zhao, YangQuan Chen (PI)
MESALab-FOSP2	
Description	Fractional order generalized SEIR compartmental models with power law infection rates
References	Guo et al. 2020
Co-authors	Lihong Guo, Yanting Zhao, YangQuan Chen (PI)
NEU-MOBS	
Description	Stochastic, age-structured, compartmental model, including symptomatic and asymptomatic transmissions, as well as hospitalizations.
Calibration	Calibration of R0 and initial date performed using reported deaths
Additional NPIs included	Social distancing, Stay-at-home (mandatory, e.g., government-ordered)
Additional data sources used	Age structure contact patterns from highly detailed macro (census) and micro (survey) data on key socio-demographic features
References	Mistry et al. 2020
Co-authors	Kunpeng Mu, Ana Pastore y Piontti, Alessandro Vespignani (PI)
Acknowledgements	Matteo Chinazzi, Jessica T. Davis, Xinyue Xiong
Funding	AV, APyP, KM acknowledge the support of the McGovern Foundation, Google Cloud and Google Cloud Research Credits program.
NIH-FDA SICR	

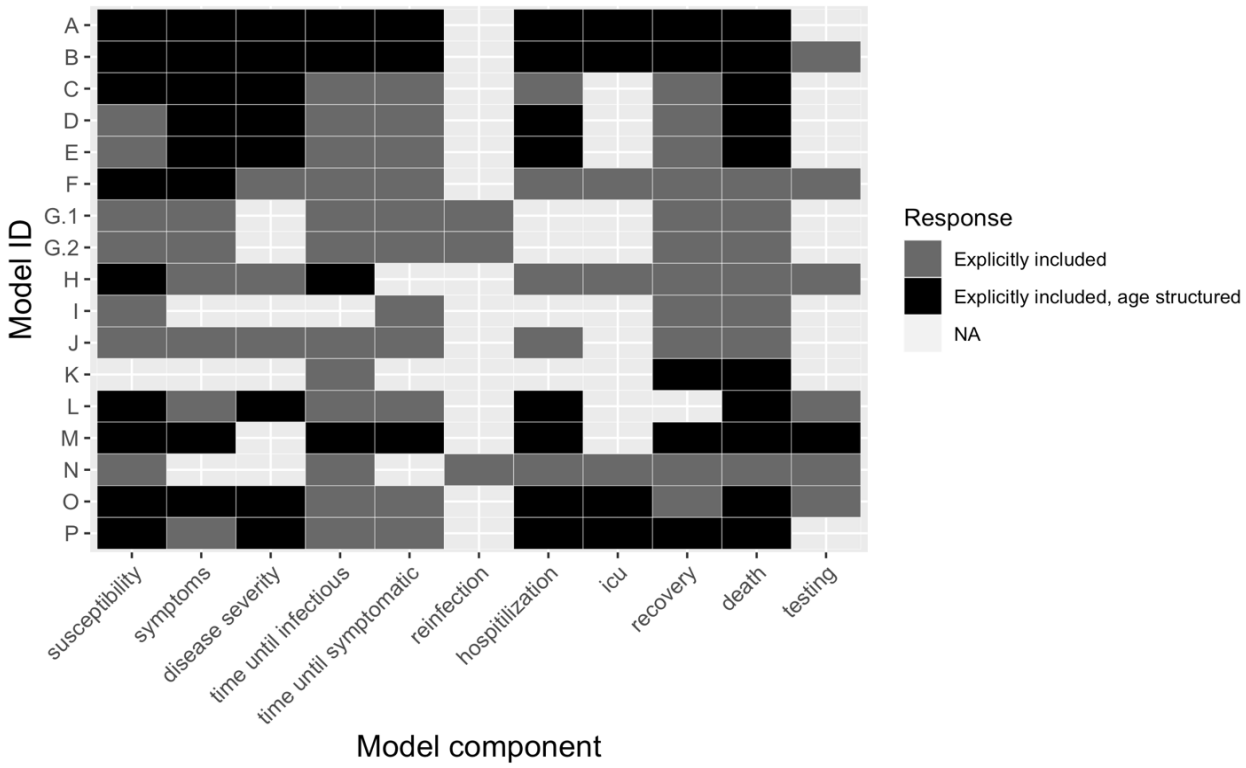
Description	Compartmental model with compartments for Susceptible, Infected, Case (C), case Recovered (R), and case Dead (D). The mean dynamics of the compartments are governed by an ODE system. The likelihood for the rate of appearance of C, R, and D are given by a negative binomial distribution where the dispersion parameter is a fitted parameter. Code available: https://github.com/ccc1685/covid-19
Diagram	See Fig. 1 in Chow et al. (2020)
Calibration	Priors were obtained from posteriors of fits to New York and Maryland
Previous use cases	COVID-19 globally, data permitting
References	Chow et al. 2020
Co-authors	Joshua C Chang, Richard C Gerkin, Shashaank Vattikuti, Artur Belov, Osman Yogurtcu, Carson C Chow (PI)
Acknowledgements	Hong Yang
Funding	CCC and SV were supported by the Intramural Program of the NIH, NIDDK. RCG was supported by NIDCD, NINDS, and NSF. This work utilized the computational resources of the NIH HPC Biowulf cluster (http://hpc.nih.gov).
Notre Dame-FRED	
Description	Agent-based model, FRED (Framework for Reconstructing Epidemic Dynamics), with updated epidemiological parameters based on studies to date. FRED explicitly models transmission dynamics of a pathogen in a realistic population, and allows for the impacts of NPIs to be modeled explicitly (e.g., school closure results in agents representing children staying home). Code available: https://github.com/confunguido/covid19_ND_forecasting
Calibration	Disease specific parameters were calibrated to the number of daily deaths in PA. Adams County was then simulated to estimate the rate of importations from the state incidence and a scaling factor to google mobility trends. Parameters were uniformly sampled for each step of the calibration using a sobol design sequence (pomp package in R). Then, the likelihood based on the daily number of deaths was calculated.
Additional NPIs included	Isolation (post infection)
Additional data sources used	NY times data to match the daily deaths of the state of PA as a pre-fitting step (https://github.com/nytimes/covid-19-data); Google mobility trends
Previous use cases	Several diseases; originally developed by University of Pittsburgh to model the 2009 influenza pandemic
References	Grefenstette et al. 2013
Co-authors	Guido España, Sean Cavany, Rachel Oidtman, T. Alex Perkins (PI)
Acknowledgements	Alan Costello, Annaliese Wieler, Anita Lerch, Carly Barbera, Marya Poterek, Quan Tran
Funding	This work was supported by an NSF RAPID grant to TAP (DEB 2027718), an Arthur J. Schmitt Fellowship and Eck Institute for Global Health Fellowship to RJO. We thank the University of Notre Dame Center for Research Computing for computing resources.
UCLA-SuEIR	
Description	New epidemic compartmental model (SuEIR) based on the standard SEIR model that also takes into account untested/unreported cases. The model is trained by machine learning algorithms based on reported historical data. Project website: https://covid19.uclaml.org/
Diagram	See Fig. 1 in Zou et al. (2020)

Additional NPIs included	Quarantine (post exposure), Isolation (post infection), Stay-at-home (voluntary), Stay-at-home (mandatory, e.g., government-ordered), Age-specific interventions (e.g., isolation of elderly)
References	Zou et al. 2020
Co-authors	Difan Zou, Weitong Zhang, Lingxiao Wang, Pan Xu, Jinghui Chen, Quanquan Gu (PI)
UF COVID-ABM	
Description	Spatially explicit, agent-based model simulating a community of individuals based on census, workplace, and school data. The movement of each person during a simulated day takes place among a set of pre-assigned local places. Pathogen exposure events occur probabilistically when a susceptible person co-localizes with an infectious person and exposures can be resisted, or result in asymptomatic, mild, severe and/or critical infection
Additional NPIs included	Social distancing, Stay-at-home (voluntary), School closures
Additional data sources used	The American Community Survey 5-year dataset; geographical coordinates and the business type from the National Corporation Directory; North American Industry Classification System to identify essential vs non-essential businesses; University of Florida GeoPlan Center shapefile and data from the National Center for Education Statistics to locate schools
Previous use cases	Dengue in Yucatan, Mexico
References	Hladish et al. 2020; Hladish et al. 2018; Hladish et al. 2016; Flasche et al. 2016
Co-authors	Kok Ben Toh, Arlin Stoltzfus, Carl Pearson, Dianela Perdomo, Alexander Pillai, Sanjana Bhargava, Thomas Hladish (PI)
UNCC LSTM	
Description	Data-driven, stochastic SI model utilizing a deep learning recurrent neural network with multivariate LSTM architecture. The model was calibrated using COVID-19 epidemic data in another region with ending of the epidemic to guide the model to learn how the epidemic could eventually phase out.
Calibration	Transfer learning was used to let the LSTM learn how the epidemic would eventually end from another region, explore the RNN structure and hyperparameters, and apply them to tune the model for the modeled region
Additional data sources used	COVID-19 data from another region where the epidemic has (presumably) ended.
Co-authors	Daniel Janies, Rajib Paul, Shi Chen (PI)
Acknowledgements	Tinghao Feng
UT-SEPAYHR	
Description	Stochastic, age- and risk-structured compartmental model that includes susceptible, exposed, presymptomatic, asymptomatic, symptomatic, hospitalized, and recovered states (SEPAYHR). The model is simulated using a hybrid approach with a deterministic initial phase (up to 20 total symptomatic cases) followed by a stochastic phase.
Diagram	See Fig. A1 in Duque et al. (2020)
Calibration	Basic reproductive number (R_t) was estimated using provided and transmission probability was estimated using a next-generation matrix approach based on the model structure and R_t. Epidemic start date was based on the time to first death implied by the estimated Rt and transmission probability. Transmission reduction due to social distancing was estimated with a nonlinear least squares fitting procedure in the SciPy/Python package. Detection rate was estimated using the provided data and published estimates of age-structured infection fatality ratios.

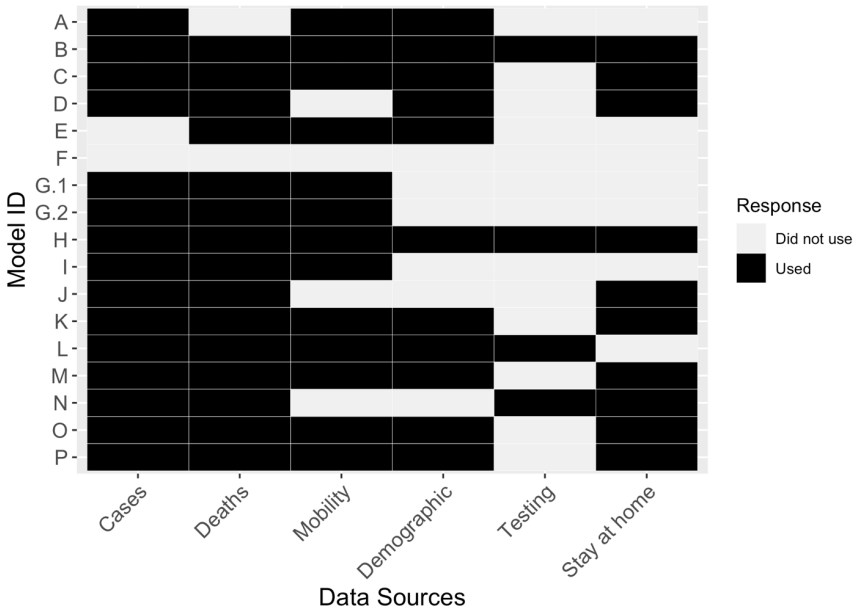
Additional NPIs included	Social distancing, Stay-at-home (voluntary), Stay-at-home (mandatory, e.g., government-ordered)
References	Duque et al. 2020
Co-authors	Kelly Pierce, Remy Pasco, Lauren Ancel Meyers (PI)
Acknowledgements	Spencer Fox, Zhanwei Du, Ethan Ho, Greg Zynda, Jawon Song
Funding	CDC contract 75D-301-19-C-05930, and NIH grant 3R01AI151176-01S1
UW-THINKLAB-SEIQRD	
Description	Compartmental model, consisting of 6 compartments: Susceptible (S), Exposed (E), Infectious (I), Quarantined (Q), Recovered (R) and Dead (D). Transitions between compartments are formulated using deterministic functions in discrete time steps and parameters governing transitions are assumed to change stochastically on a daily basis (except for predetermined parameters).
Calibration	Particle filtering is used to update the distribution of parameter estimates on a daily basis while case and death data is available (i.e. by May 15).
Additional data sources used	National average hospital beds per capita from World Health Organization (https://www.who.int/data/gho/data/indicators/indicator-details/GHO/hospital-beds-(per-10-000-population)
Co-authors	Xiangyang Guan, Cynthia Chen (PI)
VT Childs Lab	
Description	Deterministic, compartmental ODE system. Parameter sets are chosen using Latin Hypercube Sampling and refined based on comparison to data.
Calibration	Parameters were chosen from given ranges via Latin Hypercube Sampling (LHS)
Additional NPIs included	Social distancing, Isolation (post hospitalization)
Co-authors	Lauren M Childs (PI)
Acknowledgements	Kate Langwig, Leah Johnson, Eyvindur Ari Palsson, Julie Blackwood
Funding	LMC acknowledges support from National Science Foundation grant No. 2029262.

SM Table 2: Importation rate. Most models did not include an importation rate after any initial seeding. Models that did maintained a relatively small importation rate, per the elicitation setting.

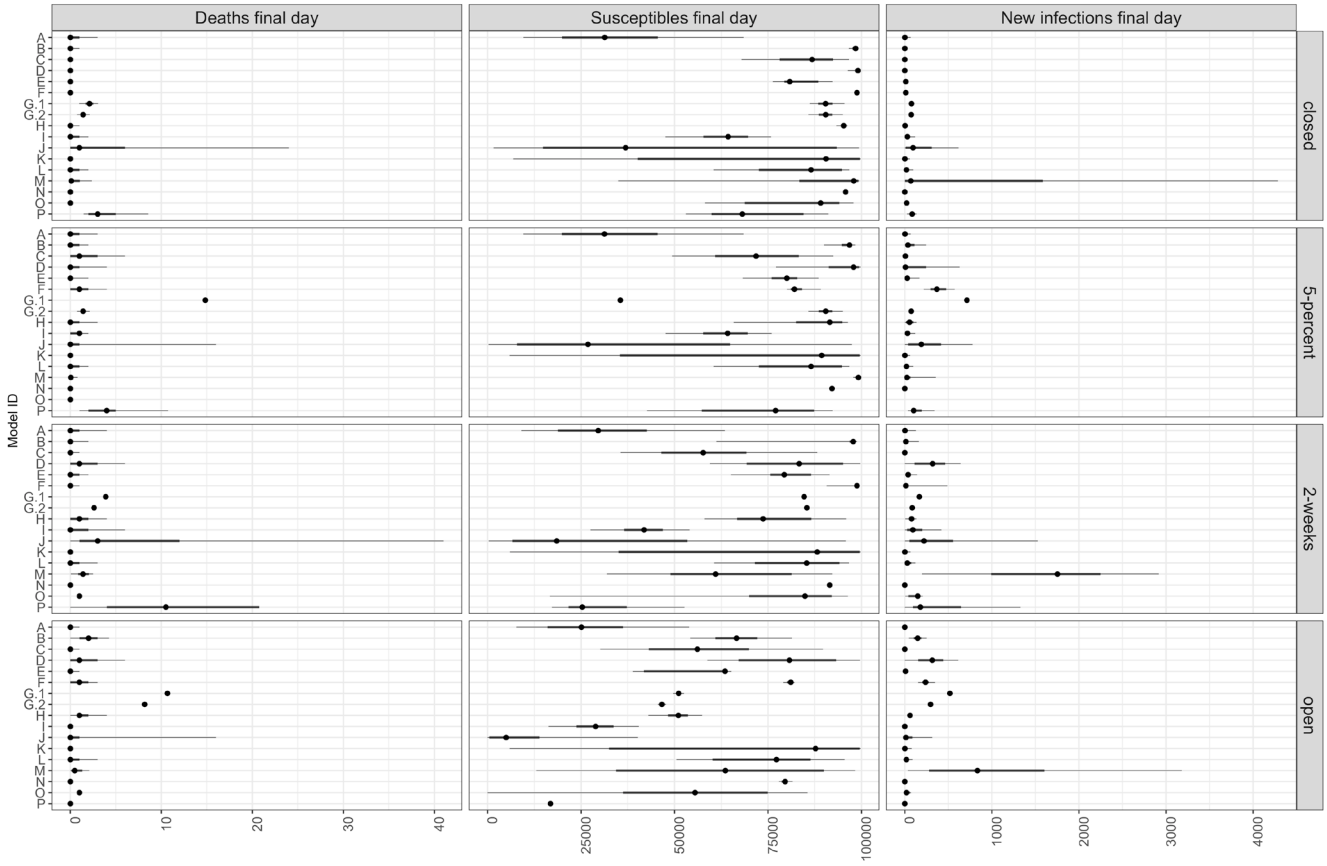
Model ID	Importation rate (or None)
A	None
B	None
C	None
D	None
E	0.14 cases / day
F	1 exposure / day (with a probability of resistance)
G.1	None
G.2	None
H	None
I	None (after initial seeding)
J	None (after initial seeding)
K	None
L	0 – 0.5 cases / day
M	None
N	None
O	Varied
P	None



SM Fig. 15: Description of model components and structure by model. Participants were asked to indicate which model components were included in their model (from a given set) and whether any component was structured by age and/or gender and/or sex as part of the submission checklist. No model included any components structure by gender and/or sex. Twelve of the 17 included at least one component structured by age.



SM Fig. 16: Data sources used for each model. Participants were asked to indicate which of the provided datasets were used for any part of the model (e.g., for calibration, training, fitting etc.) as part of the submission checklist. All but one model used at least two of the provided data. Model F used only external data sources (provided data was used solely to better understand the intent of the exercise).



SM Fig. 17: Projected number of deaths, people who are susceptible, and new infections under each scenario for the final day of the forecast. Participants reported the 5th, 25th, 50th, 75th, and 95th quantiles for the number of deaths, susceptibles, and new infections on the final day (November 15, 2020) under each scenario. All models started with similar initial susceptibles.

Supporting references

Chow, C.C., Chang, J.C., Gerkin, R.C., and Vattikuti, S. (2020). Global Prediction of Unreported SARS-CoV2 Infection from Observed COVID-19 Cases. MedRxiv, May 5, 2020. https://doi.org/10.1101/2020.04.29.20083485.

Covid-19 International Modelling Consortium. (2020). COVID-19 App v15.1.3. https://comomodel.net/

Dong, E., Du, H. and Gardner, L., 2020. An interactive web-based dashboard to track COVID-19 in real time. The Lancet infectious diseases, 20(5), pp.533-534.

Du, Z, Wang, X, Pasco, R, Pignone, M, Fox, S, and Meyers, L (2020). Covid-19 healthcare demand projections: 22 texas cities. https://sites.cns.utexas.edu/sites/default/files/cid/files/covid-19_analysis_for_austin_march2020.pdf, 03-2020.

Duque, D., Tec, M., Morton, D.P., Scott, J., Yang, H., Pasco, R., Pierce, K., Fox, S.J., Pignone, M., Hudson, P., and Meyers, L.A. (2020). Staged Strategy to Avoid Hospital Surge and Preventable Mortality, while Reducing the Economic Burden of Social Distancing Measures. Presented to the city of Austin and Travis County on May 18, 2020. https://sites.cns.utexas.edu/sites/default/files/cid/files/staged_strategy_avoid_surge_texas.pdf?m=1590438337

Flasche, S., Jit, M., Rodrguez-Barraquer, I., Coudeville, L., Recker, M., Koelle, K., Milne, G., Hladish, T. J., Perkins, T. A., Cummings, D. A. T., Dorigatti, I., Laydon, D. J., Espaa, G., Kelso, J., Longini, I., Lourenco, J., Pearson, C. A. B., Reiner, R. C., Mier-y-Tern-Romero, L., Vannice, K., & Ferguson, N. (2016). The long-term safety, public health impact, and cost-effectiveness of routine vaccination with a recombinant, live-attenuated dengue vaccine (Dengvaxia): A model comparison study. PLoS Medicine, 13(11). https://doi.org/10.1371/journal.pmed.1002181

Germann, T.C., Gao, H., Gambhir, M., Plummer, A., Biggerstaff, M., Reed, C., and Uzicanin, A. (2019). School dismissal as a pandemic influenza response: When, where and for how long?, Epidemics 28, 100348. https://doi.org/10.1016/j.epidem.2019.100348

Germann, T.C., Kadau, K., Longini, I.M. and Macken, C.A. (2006). Mitigation Strategies for Pandemic Influenza in the United States, Proc. Natl. Acad. Sci. (USA) 103, 5935-5940. https://doi.org/10.1073/pnas.0601266103

Google LLC "Google COVID-19 Community Mobility Reports". https://www.google.com/covid19/mobility/ Accessed: August 10, 2020.

Grefenstette, J.J., Brown, S.T., Rosenfeld, R., DePasse, J., Stone, N.T., Cooley, P.C., Wheaton, W.D., Fyshe, A., Galloway, D.D., Sriram, A. and Guclu, H., 2013. FRED (A Framework for Reconstructing Epidemic Dynamics): an open-source software system for modeling infectious diseases and control strategies using census-based populations. BMC public health, 13(1), pp.1-14.

Guo L, Zhao Y, Chen Y Q. (2020). Management strategies and prediction of COVID-19 by a fractional order generalized SEIR model[J]. medRxiv. https://doi.org/10.1101/2020.06.18.20134916

Halloran, M.E., Ferguson, N.M., Eubank, S., Longini Jr., I.M., Cummings, D.A.T., Lewis, B., Xu, S., Fraser, C., Vullikanti, A., Germann, T.C., Wagener, D., Beckman, R., Kadau, K.,

Barrett, C., Macken, C.A., Burke, D.S., and Cooley, P. (2008). Modeling Targeted Layered Containment of an Influenza Pandemic in the USA, Proc. Natl. Acad. Sci. (USA) 105, 4639-4644. https://doi.org/10.1073/pnas.0706849105

Hladish, T. J., Pearson, C. A., Chao, D. L., Rojas, D. P., Recchia, G. L., Gmez-Dants, H., Halloran, M. E., Pulliam, J. R. C., & Longini, I. M. (2016). Projected impact of dengue vaccination in Yucatn, Mexico. PLoS neglected tropical diseases, 10(5). https://doi.org/10.1371/journal.pntd.0004661

Hladish, T. J., Pearson, C. A., Rojas, D. P., Gomez-Dantes, H., Halloran, M. E., Vazquez-Prokopec, G. M., & Longini, I. M. (2018). Forecasting the effectiveness of indoor residual spraying for reducing dengue burden. PLoS neglected tropical diseases, 12(6), e0006570. https://doi.org/10.1371/journal.pntd.0006570

Hladish, T. J., Pearson, C. A., Toh, K. B., Rojas, D. P., Manrique-Saide, P., Vazquez-Prokopec, G. M., Halloran, M. E., & Longini, I. M. (2020). Designing effective control of dengue with combined interventions. Proceedings of the National Academy of Sciences, 117(6), 3319-3325. https://doi.org/10.1073/pnas.1903496117

Kerr, C.C., Stuart, R.M., Mistry, D., Abeysuriya, R.G., Hart, G., Rosenfeld, K., Selvaraj, P., Nunez, R.C., Hagedorn, B., George, L. and Izzo, A. (2020) Covasim: an agent-based model of COVID-19 dynamics and interventions. medRxiv. https://doi.org/10.1101/2020.05.10.20097469

Keystone Strategy "COVID-19 Intervention Data" https://github.com/Keystone-Strategy/covid19-intervention-data/ Accessed: August 10, 2020

Killeen, B.D., Wu, J.Y., Shah, K., Zapaishchykova, A., Nikutta, P., Tamhane, A., Chakraborty, S., Wei, J., Gao, T., Thies, M. and Unberath, M., 2020. A County-level Dataset for Informing the United States' Response to COVID-19. arXiv preprint arXiv:2004.00756.

Lin, G., Strauss, A.T., Pinz, M., Martinez, D.A., Tseng, K.K., Schueller, E., Gatalo, O., Yang, Y., Levin, S.A. and Klein, E.Y. (2020) Explaining the Bomb-Like Dynamics of COVID-19 with Modeling and the Implications for Policy. medRxiv. https://doi.org/10.1101/2020.04.05.20054338v1

"List of countries by hospital beds" Wikipedia. https://en.wikipedia.org/wiki/List_of_countries_by_hospital_beds

Mordecai, E, Childs, M, Kain, M, Harris, M, Ritchie, J, Couper, L, Delwel, I, and Nova, N (2020). Potential Long-Term Intervention Strategies for COVID-19. https://covid-measures.github.io/

Mistry, D, Litvinova, M, Chinazzi, M, Fumanelli, L, Gomes, MF, Haque, SA, Liu, QH, Mu, K, Xiong, X, Halloran, ME and Longini Jr, IM, 2020. Inferring high-resolution human mixing patterns for disease modeling. arXiv preprint arXiv:2003.01214. Accessed: April 4, 2020.

NACO "Counties and COVID-19 Safer at Home Orders" https://www.naco.org/resources/featured/counties-and-covid-19-safer-home-orders Accessed: August 10, 2020

The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team (2020). The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases

(COVID-19)—China, 2020. China CDC Weekly, 2(8), pp.113-122. http://weekly.chinacdc.cn/en/article/id/e53946e2-c6c4-41e9-9a9b-fea8db1a8f51

Prem, K., Cook, A.R. and Jit, M., 2017. Projecting social contact matrices in 152 countries using contact surveys and demographic data. PLoS computational biology, 13(9), p.e1005697. https://doi.org/10.1371/journal.pcbi.1005697

Spencer, J.A. (2020) Chronic and Acute Respiratory Pathogens: Evolutionary and Epidemiological Characteristics of Tuberculosis, Influenza-like Illness, and COVID-19. (Doctoral Thesis, Biology, University of New Mexico, Albuquerque, New Mexico).

Spencer, J., Shutt, D. P., Moser, S. K., Clegg, H., Wearing, H. J., Mukundan, H., & Manore, C. A. (2020). Epidemiological parameter review and comparative dynamics of influenza, respiratory syncytial virus, rhinovirus, human coronavirus, and adenovirus. *medRxiv*.

Zou, D., Wang, L., Xu, P., Chen, J., Zhang, W., and Gu, Q. (2020). Epidemic Model Guided Machine Learning for COVID-19 Forecasts in the United States. https://doi.org/10.1101/2020.05.24.20111989