ASKEM 18-Month Milestone: Evaluation Scenarios

USE CASE: EPIDEMIOLOGICAL MODELING, REGNETS, STOCK AND FLOW

Table of Contents

Scenario 1: Modeling SARS-CoV-2 infections in White-tailed Deer

Estimated % of time: Baseline: 20%; Workbench 15%

To date, COVID-19 modeling efforts have focused almost exclusively within human populations. There is now compelling evidence that SARS-CoV-2, the virus that causes COVID-19, spreads from humans to white-tailed deer (Odocoileus virginianus), as well as between white-tailed deer and possibly from white-tailed deer back into human populations. There is evidence that infected white-tailed deer may serve as a reservoir for nearly extinct variants of concern (e.g., Delta variant), some of which are associated with greater clinical severity in human populations. Additional mutations to these nearly extinct variants of concern within the wildlife reservoir could make them more transmissible in addition to causing increased severity of infection.

Decision makers are interested in (1) better characterizing infection dynamics within the white-tailed deer (WTD) population to understand risks such as re-importation of nearly extinct variants of concern back into human populations; (2) understanding the potential efficacy of interventions targeted towards decreasing the deer population infected with SARS-CoV-2, in order to decrease future likelihood of transmission from deer population back to human population (which could potentially drive new Covid waves in the future, were this to happen at any meaningful level). To do this, they have asked you to find a model of Covid in the WTD population, that already can support, or can be modified to support, these types of interventions.

You have identified three compartmental models of SARS-CoV-2 transmission within the WTD population that are relevant to your task. Two are published as pre-prints, and one is a very recent publication. Given the novelty of these models, you want to better understand how they are similar and different in terms of their assumptions, strengths, limitations, and fit-for-purpose.

- Model A: https://doi.org/10.1155/2024/7589509
- Model B: https://doi.org/10.48550/arXiv.2401.10057

- Model C: https://doi.org/10.1101/2023.08.30.555493

For Q1-5, use only the above 3 publications as source material.

1. **Model Extraction**: Begin by extracting the three models, available at the links above. For each model, note the time to extract the model and get it into an executable state that can run a simple test simulation and get sensible results. You may choose the initial conditions and parameter values for the test simulation; they don't need to be realistic, but the results do need to make sense given the values you choose. For workbench modelers, model extraction time may include human-in-the-loop curation, and for baseline modelers, this time may include debugging code. For each model, provide simulation results from your test simulation.

   For baseline modelers, model extraction is defined as the following:
   - Writing out or capturing the equations describing the model, or drawing out the model structure. (You can write them out by hand, but be sure to capture the image for the work product).
   - Writing out definitions of all variables and parameters, with units
   - Finding default values for parameters, initial values for variables, and whatever else is needed to initiate/run the model
   - If the model is not already installed on the VM, find and install code to run it, or produce your own code to run the model. The code should be deposited in your work product SharePoint folder.

   For the workbench modelers, model extraction is defined as the following:
   - Ingesting the model from source paper or code, into the workbench
   - Capturing the set of equations describing the model in the workbench
   - Gathering definitions of all variables and parameters, with units
   - Gathering default values for parameters, initial values for variables, and whatever else is needed to initiate/run the model
   - Ensuring the model is executable in the workbench

2. **Model Comparison**: Do a model comparison based on key differences in assumptions, strengths, limitations, and distinguishing characteristics. Based on this information, rank each model in terms of their relevance and fit-for-purpose in this context.

Model	**Distinguishing characteristics**	**Assumptions**	**Strengths**	**Limitations**	**Rank fit-for-purpose (1 = most suitable; 3 = least suitable), with reasoning**

3. **Structural Model Comparison**: Now perform structural model comparison between each pair of models. By structural comparison, we seek to understand how compartments and transition pathways overlap or diverge between models. Feel free to create diagrams or use equations in your response.

4. **Gap Analysis:** Referring to the high-level decision maker objectives you have been given, what are the key gaps between these existing models and your modeling needs (if any)?

5. Based on Q1-4, select the model you think is the most appropriate **starting point** for supporting decision makers, and explain your reasoning. (Note that there is no single right answer – this scenario is about selecting a model, and being able to justify your choice with evidence.)

For Q6 and beyond, you may use any additional materials you find in the literature or associated with the 3 models linked at the beginning of the scenario.

6. **Find Parameters:** Find relevant parameter values for the chosen model in Q5. and fill in the following information about sources and quality. You may use any of the papers linked in this scenario, as well as any other literature on SARS-CoV-2 or human/WTD population dynamics. If relevant, you may include multiple rows for the same parameter (e.g. perhaps you find different values from different reputable sources), with a 'summary' row indicating the final value or range of values you decide to use. If there are required parameters for your model that you can't find sources for in the literature, you may find data to calibrate your model with, or make reasonable assumptions on what sensible values could be (with rationale).

Parameter	Parameter Definition	Parameter Units	Parameter Value or Range	Uncertainty Characterization	Sources	Modeler Assessment on Source Quality

7. **Model Forecast**: Now assume you're at the start of a Covid-19 wave in WTD in New York state in the beginning of November, 2021. Set appropriate initial conditions, parameterize your model given the information you found in Q6 and make a 2-month forecast of Covid-19 dynamics in WTD. At the very least, your forecast should include susceptible population, infectious population, and recovered population. If the model you are using has additional components, please include those in the forecast.

8. **Model Modification (Stratification):** If the model that you selected is not able to distinguish between multiple variants, stratify the model to demonstrate that it can be used to simulate at least three (alpha, delta, and omicron) variants.

9. **Interventions**:
    a. Brainstorm three different interventions you could implement in your model that would lower the infectious population in your forecast from Q7. For this problem, you can ignore the feasibility of implementing the interventions in the real world and make hypothetical assumptions (e.g. you can assume treatments or vaccines exist even if they do not in reality). For each intervention, how would you implement it in your chosen model? Which model components are involved?
    b. Implement one of the interventions you considered in the model, redo the forecast from Q7, and comment on the impact of the intervention, compared to the baseline forecast in Q7.

Scenario 1 Summary Table

Question	Inputs	Tasks	Outputs
Q1	Linked papers	• Extract equations • Extract parameter values • Iterate/curate extraction and execute model until a test simulation gives reasonable results	• 3 extracted models with all variables and parameters defined, and with units • Test simulation plots • Time to do model extraction • Time to execute extracted model and plot results
Q2-3	Extracted models	• Model comparison based on assumptions, limitations, strengths, defining characteristics • Structural model comparison	• Completed comparison table • Time to complete table • Time to do structural model comparison
Q4	Results from Q1-3	Identify gaps in the candidate models	Explanation of gaps between candidate models and decision-maker objectives
Q5	Results from Q1-4	Select a model	Selected model, with explanation
Q6	Chosen model from Q5	Find parameters, from publications or other sources	• Completed parameter table • Time to complete table
Q7	• Chosen model from Q5 • Parameters from Q6	• Parameterize model • Set initial conditions • Create 2-month forecast	• Forecast results that include susceptible, infected, and recovered deer populations. • Time to generate forecast
Q8	Parameterized model from Q7	Stratify model by 3 variants	• Stratified model • Time to stratify model
Q9	• Selected model from Q5 • Parameters from Q6	• Brainstorm potential interventions and how they would be implemented • Implement one intervention and compare results with Q7	• Forecast results that include susceptible, infected, and recovered deer populations, with intervention implemented. • Time to generate forecast

Decision-maker Panel Questions

What is your confidence that the modeling team selected a model and associated data appropriate for the decision-making questions under consideration? Select score on a 7-point scale.

1. Very Low
2. Low
3. Somewhat Low
4. Neutral
5. Somewhat High
6. High
7. Very High

Explanation: Decision makers, evaluate whether results include an appropriate model, with relevant parameters, starting values, etc.

The decision-maker confidence score should be supported by the answers to the following questions:

- Did modelers clearly communicate key differences between the chosen model and other candidate models? What kinds of information were provided to help you understand the differences? What kinds of information would you have wanted?
- Are results traceable to sources and did modelers assess source quality?
- Is the model chosen for the scenario appropriate/fit-for-purpose for the given problem? Were assumptions and data associated with the model clearly communicated?

Scenario 2: Improving Forecasts Through Model Updates

Estimated % of time: Baseline 50%; Workbench 40%

It is the end of 2022, and you are supporting a decision maker who is preparing for a winter Covid wave. The winter Covid wave caused by the original Omicron variant just a year earlier (end of 2021 and early 2022), was, at the US country level, the largest of the pandemic so far. Fearing another similar winter wave, the decision maker asks you to do a retrospective analysis of the prior winter. In particular, they want to you try and develop the most accurate model of the original Omicron wave, explore various interventions in the model and assess their effects. For your retrospective analysis, consider the time period of December 1st, 2021, to March 1st, 2022, with the first month (December 1st – 31st, 2021) as the training period, and the remaining time as the test period.

Starting Model: Begin with the following SIRHD model structure (Figure 1) and set of differential equations. For workbench modelers, a version of this may already exist in the workbench; if not, create it. For baseline modelers, see accompanying code in supplementary materials. The general form/structure of the model is below.

Note: Although many compartmental models include an Exposed compartment, we are omitting it for this scenario for simplification reasons. $p_{state1 \to state2}$ represents the probability of moving between the indicated states. $r_{state1 \to state2}$ represent rates for how long processes take (i.e. 1/average time to move between states, e.g. 1/ incubation period, etc.). For this starting model, use the following values as initial guesses for parameter values:

- $\beta = 0.18$ new infections per infected person/day
- $r_{I \to R} = 0.07$/day
- $r_{I \to H} = 0.07$/day
- $r_{H \to R} = 0.07$/day
- $r_{H \to D} = 0.3$/day
- $p_{I \to R} = 0.9$
- $p_{I \to H} = 0.1$
- $p_{H \to R} = 0.87$
- $p_{H \to D} = 0.13$
- To compensate for the fact that we don't have an Exposed compartment in this model, we lower the total population N to 150e6 people, rather than use the actual total population of the United States. This is meant to approximate the situation where some individuals were exercising caution during the winter of 2021-2022, and were not exposed to Covid-19.

For initial conditions please pull values from the gold standard cases and deaths data from the Covid-19 ForecastHub, and HHS hospitalization data from https://healthdata.gov/Hospital/COVID-19-Reported-Patient-Impact-and-Hospital-Capa/g62h-syeh. Let $R(0) = cumulative\ infections - cumulative\ deaths$, as of December 1st, 2021. Let $S(0) = N - I(0) - R(0) - H(0) - D(0)$.

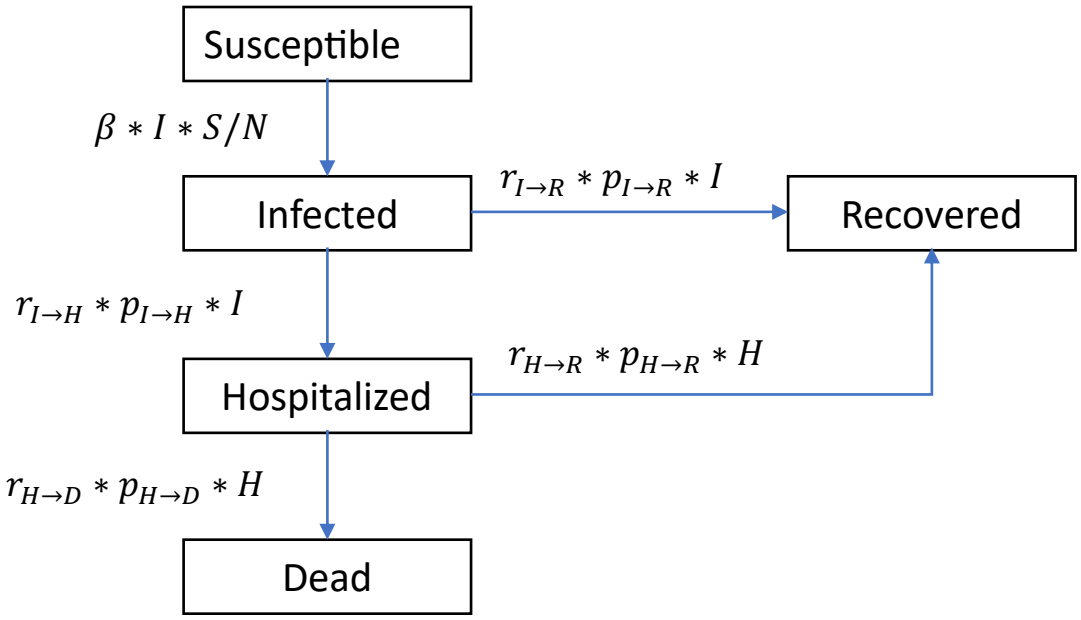

Figure 1. Starting model for Scenario 2: Improving Forecasts Through Model Updates.

$$\frac{dS}{dt} = -\beta * I * \frac{S}{N}$$
$$\frac{dI}{dt} = \beta * I * \frac{S}{N} - r_{I\to R} * p_{I\to R} * I - r_{I\to H} * p_{I\to H} * I$$
$$\frac{dR}{dt} = r_{I\to R} * p_{I\to R} * I + r_{H\to R} * p_{H\to R} * H$$
$$\frac{dH}{dt} = r_{I\to H} * p_{I\to H} * I - r_{H\to R} * p_{H\to R} * H - r_{H\to D} * p_{H\to D} * H$$
$$\frac{dD}{dt} = r_{H\to D} * p_{H\to D} * H$$

1. **Model Calibration**: Using the given parameter values as initial guesses, calibrate the starting model, with data from the first month of the retrospective analysis: December 1st, 2021, through December 31st, 2021. You may decide which parameter values you are confident about and don't need to calibrate, and the min/max ranges for the ones you would like to calibrate. Include plots of your calibrated model outputs, compared to actual data, for this time period.

2. **Single Model Forecast**:
    a. Using your calibrated model, forecast cases, hospitalizations, and deaths, for the test period (January 1st, 2022 – March 1st, 2022).
    b. Plot your forecast against actual observational data from this time period, and calculate Absolute Error.
    c. How does your forecast's Absolute Error metric over the first 4 weeks of this time period, compare against forecasts during this time period, from other

compartmental models in the Covid-10 ForecastHub? Compare specifically against the UCLA-SuEIR and BPagano models. You can find forecast data and error scores for these two models, in the supplementary materials. All model forecasts in the ForecastHub are located here: https://github.com/reichlab/covid19-forecast-hub/tree/master/data-processed

3. **Ensemble Forecast**: You hypothesize that the β parameter should actually be time-varying and reflects interaction and transmission processes that change considerably throughout the course of the analysis period. A crude way to account for a time-varying parameter is to create an ensemble from different configurations of the same model that sufficiently explores the range of a time-varying parameter.
    a. Create 3 different configurations of the model from Q1, each with a different value of β that is constant. Combine these configurations into an ensemble. You can parameterize each configuration, or the combined ensemble model, using any approach you'd like, including weighting coefficients that change over time.
    b. Forecast cases, hospitalizations, and deaths, for the test period (January 1st, 2022 – March 1st, 2022).
    c. For each outcome (cases, hospitalizations, deaths), plot your forecast against actual observational data from this time period, and calculate Absolute Error.
    d. How does your forecast's Absolute Error metric over the first 4 weeks of this time period compare against one of the ForecastHub ensembles (e.g. 'COVIDhub-4_week_ensemble'). You can find forecast data and error scores for this ensemble, in the supplementary materials. All forecast data from the ForecastHub ensembles are here: https://github.com/reichlab/covid19-forecast-hub/tree/master/data-processed
    e. How does your forecast performance compare against the results of Q2?

4. **Model Update**: Now update your model to include vaccination. Ensure that this is done in a way that can support interventions around vaccinations (e.g. incorporate a vaccination policy or requirement that increases rate of vaccination). For this question, only consider one vaccine type and assume one dose of this vaccine is all that's required to achieve 'fully vaccinated' status. You will consider multiple doses in a later question.

5. **Find Parameters**: Your updated model from Q3 should have additional variables and new parameters. What is the updated parameter table that you will be using? As with scenario 1, you may include multiple rows for the same parameter (e.g. perhaps you find different values from different reputable sources), with a 'summary' row indicating the final value or range of values you decide to use. If there are required parameters for your model that you can't find sources for in the literature, you may find data to calibrate your model with, or make reasonable assumptions on what sensible values could be (with rationale). You may use any sources, including the following references on vaccine efficacy for Moderna, Pfizer, and J&J vaccines.
    - Estimates of decline of vaccine effectiveness over time https://www.science.org/doi/10.1126/science.abm0620

- CDC Vaccine Efficacy Data https://covid.cdc.gov/covid-data-tracker/#vaccine-effectiveness
- Vaccination data sources https://data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-United-States-Jurisdi/unsk-b7fc

Parameter	Parameter Definition	Parameter Units	Parameter Value or Range	Uncertainty Characterization	Sources	Modeler Assessment on Source Quality

6. **Model Checks**: Implement common sense checks on the model structure and parameter space to ensure the updated model and parameterization makes physical sense. Explain the checks that were implemented. For example, under the assumption that the total population is constant for the time period considered:
    a. Demonstrate that population is conserved across all compartments
    b. Ensure that the total unvaccinated population over all states in the model, can never increase over time, and the total vaccinated population over all states in the model, can never decrease over time.
    c. What other common-sense checks did you implement? Are there other checks you would have liked to implement but it was too difficult to do so?

7. **(Optional) Single Model Forecast**: Using your updated model, forecast cases, hospitalizations, and deaths, for the test period (January 1st, 2022 – March 1st, 2022).
    a. Plot your forecast against actual observational data from this time period, and calculate Absolute Error.
    b. How does your forecast's Absolute Error metric over the first 4 weeks of this time period, compare against forecasts during this time period, from other compartmental models in the Covid-10 ForecastHub? Compare specifically against the UCLA-SuEIR and BPagano models.
    c. How does your forecast performance compare with the one in Q2? If the forecast performance has improved or gotten worse, why do you think this is?

8. **Model Update**: During this time period, access to at-home testing was vastly expanded, through distribution of free antigen tests and requirements for insurance to cover at-home tests for free. Update your model from Q4 to incorporate testing by modifying the β parameter between a susceptible and infected person by the following factor: $(1 - test\_access * (1 - test\_dec\_transmission))$, where *test_dec_transmission* is defined as the net decrease testing has on transmission, and *test_access* is the percentage of the general population who is likely to take a test after suspected exposure, due to increased accessibility or lowered costs, etc. For this question, assume testing has the effect of decreasing transmission between susceptible and infected populations by 25%, due to infected people choosing not to gather or interact with others, based on test outcomes confirming their infection status. Assume *test_access* increases linearly from 25% of the total population at the start of the retrospective analysis period, to 50% by the end of the period.

9. **Model Stratification**: The decision maker you're supporting is exploring targeted vaccination campaigns to boost vaccination rates for specific subpopulations. To support these questions, you decide to further extend the model from Q8, by considering several demographic subgroups, as well as vaccination dosage. Stratify the model by the following dimensions:
    - Vaccination dosage (1 or 2 doses administered)
    - Age group
    - Sex
    - Race or Ethnicity

    To inform initial conditions and rates of vaccination, efficacy of vaccines, etc., consider the subset of vaccination datasets from the starter kit listed in 'Scenario2_VaccinationDatasets.xlsx' (in the supplementary materials). Where initial conditions are not available for a specific subgroup, make a reasonable estimate based on percentages from Census sources (e.g. https://www.census.gov/quickfacts/fact/table/US/PST045223). Where parameters for specific subgroups are unavailable, generalize based on the ones that are available. Choose the number of age and race/ethnicity groups based on the data that is available.

10. **Model Checks**: Implement common sense checks on model structure and parameter space to ensure the updated model and parameterization from Q9 is structurally sound and makes physical sense. Explain the checks that were implemented. For example, under the assumption that the total population is constant for the time period considered:
    a. Demonstrate that population is conserved across all disease compartments, and within each demographic group (age, sex, race/ethnicity).
    b. Ensure that the total unvaccinated population and unvaccinated population within each age group, can never increase over time, and the total vaccinated population and vaccinated population within each age group, can never decrease over time.
    c. What other common-sense checks did you implement? Are there others you would have liked to implement but were too difficult?

11. **Single (Stratified) Model Forecast**: Using your updated model from Q9, forecast cases, hospitalizations, and deaths, for the test period (January 1st, 2022 – March 1st, 2022).
    a. Plot your forecast against actual observational data from this time period, and calculate Absolute Error. Use observational data aggregated to the general population as well as granular data for individual demographic groups. Plot outcomes for individual demographic groups, as well as the total population.
    b. How does your forecast's Absolute Error metric over the first 4 weeks of this time period, compare against forecasts during this time period, from other compartmental models in the Covid-10 ForecastHub? Compare specifically against the UCLA-SuEIR and BPagano models.

c. How does your forecast performance compare with the one in Q2? If the forecast performance has improved or worsened, why do you think this is?

12. **Interventions**: Now that you have a model that can support targeted interventions, the decision maker you support asks you to explore what would have happened during the retrospective analysis period, had these interventions been implemented at that time.
    a. With respect to your forecast from Q11, which demographic group had the worst outcomes during the retrospective period, and therefore should be targeted with interventions such as vaccine campaigns, or increased community outreach to make testing more widely available and encouraged?
    b. Implement an intervention that targets testing-related parameters (e.g. programs to increase access to tests, distribution of free tests, etc.) at the start of the forecast period, and redo the forecast from Q11. For a 1% increase in a test-related parameter (that has a net positive impact), what's the impact of the intervention on the forecast trajectory, for the affected demographic group identified in Q12a, as well as for the overall population?
    c. Implement another intervention that targets vaccination rate(s), at the start of the forecast period, and redo the forecast from Q11. For a 1% increase in vaccination rate, what's the impact of the intervention on the forecast trajectory, for the affected demographic group identified in Q12a, as well as for the overall population?

Scenario 2 Summary Table

Question	Inputs	Tasks	Outputs
Q1	• Model code OR model in workbench • Parameter value initial guesses • Training dataset for given date range	Calibrate model with the data	• Calibrated model • Plot of calibrated model results and training data • Time to calibrate model
Q2, Q7, Q11	• Model from Q1 (or Q6 or Q9) • Test dataset • Forecast data from ForecastHub models	• Create forecast of cases, hospitalizations, and deaths • Plot forecast against test dataset • Calculate Absolute Error • Get Absolute Error for ForecastHub models	• Plot of forecast against test data • Absolute Error metric for your forecast and comparison with Absolute Error from ForecastHub models • Time to generate forecast
Q3	Calibrated model from Q1	• Create ensemble forecast from 3 different configurations of Q1 model • Plot forecast against test dataset • Calculate Absolute Error metrics	• Plot of ensemble forecast against test data • Absolute Error metric for your ensemble model and comparison with Absolute Error for ForecastHub ensemble • Time to generate ensemble forecast

		• Calculate Absolute Error for ForecastHub ensemble model	
Q4	Calibrated model from Q1	Update model to include vaccination	• Updated model • Time to make model updates
Q5	Updated model from Q4, that includes vaccination	Find new or updated parameters and fill out table	• Completed parameter table • Time to complete table
Q6	• Updated model from Q4 • Parameters from Q5	• Set parameters in model • Implement model checks	• Results from model checks • Time to execute model checks
Q8	Parameterized model from Q6	Update model to include testing	• Updated model • Time to make model updates
Q9	Updated model from Q8	• Stratify model • Find initial conditions and new parameters from suggested datasets	• Stratified model • Time to stratify model • Time to parameterize updated model
Q10	Stratified model from Q9	Implement model checks	• Results from model checks • Time to execute model checks
Q12	• Model from Q10 • Forecast from Q11	• Identify target demographic group that had the worst outcomes • Implement intervention targeting testing-related parameters and redo forecast • Implement intervention targeting vaccination rates ad redo forecast	• Simulations with each intervention separately, and comparison with Q11 forecast • Time to implement interventions • Time to simulate updated forecasts with interventions

Decision-maker Panel Questions

1. What is your confidence that the modeling team developed an appropriate model and associated parameter space to sufficiently explore the scenario/problem? Select score on a 7-point scale.
    1. Very Low
    2. Low
    3. Somewhat Low
    4. Neutral
    5. Somewhat High
    6. High
    7. Very High

Explanation: The scenario involves updating or modifying a model, and decision makers will evaluate whether this was done in a sensible way and whether the final model can support all the questions asked in the scenario.

The decision-maker confidence score should be supported by the answers to the following questions:

- Did modelers clearly explain the changes being made and key differences between the original and updated models? Did the modifications/extensions the modelers made make sense and were they reasonable to you?
- Are you confident that the starting model was updated in ways that make sense? Is the final model structurally sound?
- As the model was update, was the parameter space being explored reasonable and broad enough/complete enough to support the questions required by the scenario?

2. What is your confidence in understanding model results and tradeoff between potential interventions? Select score on a 7-point scale.
    1. Very Low
    2. Low
    3. Somewhat Low
    4. Neutral
    5. Somewhat High
    6. High
    7. Very High

Explanation: Determine your confidence in your ability to do the following, based on the information presented to you by the modelers: assess model performance, assess effectiveness of all interventions considered in the scenario, and understand how uncertainty factors into all of this.

This score should be supported by the answers to the following questions:

- Did modelers communicate the impacts of interventions on trajectories? Was the effectiveness of interventions communicated?
- Did models help you to understand what would have happened had a different course of action been taken in the past?
- Where relevant to the question, was it clear how to interpret uncertainty in the results? Were key drivers of uncertainty in the results, communicated?

Scenario 3: Causal Analysis with Interventions

Estimated % of time: Baseline 30%; Workbench 20%

In this scenario, we are interested in determining the effects of masking and social distancing on Covid-19 infections using simulated data. The simulations use contact matrices and populations subdivided into three age groups. The data are generated from an SEIR model.

In these questions, we provide the contact matrices and population data as well as the outputs of the simulated SEIR model. We ask you to calibrate a model, compute β at different time intervals, and to estimate the causal effects of interventions.

The model can be described with the diagram described in Figure 2, and the following set of ordinary differential equations:

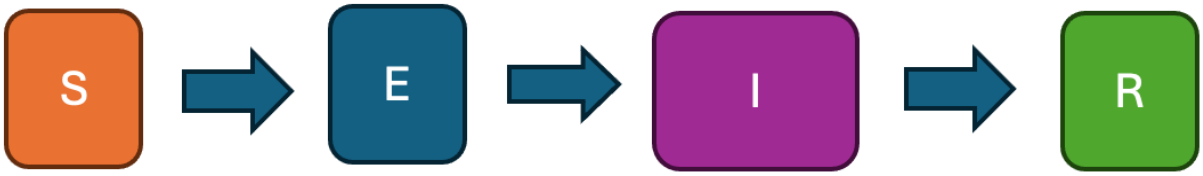

Figure 2. Model structure for Scenario 3: Causal Analysis with Interventions

$$\frac{dS_i}{dt} = -\beta * \frac{S_i}{N} * (1 - m_{ew} * m_{cw}) * \sum_{j=1} M_{ijw} * E_j$$

$$\frac{dE_i}{dt} = \beta * \frac{S_i}{N} * (1 - m_{ew} * m_{cw}) * \left(\sum_{j=1} M_{ijw} * E_j \right) - r_{E \to I} * E_i$$

$$\frac{dI_i}{dt} = r_{E \to I} * E_i - r_{I \to R} * I_i$$

$$\frac{dR_i}{dt} = r_{I \to R} * I_i$$

The above equations include the following constant parameters:

- $r_{E \to I}$, the rate of transition from compartment E to I = 0.08/day
- $r_{I \to R}$, the rate of transition from compartment I to R = 0.06/day
- β, which we ask you to estimate

And three parameters which change over time:

- m_{cw} is mask compliance over interval *w*
- m_{ew} is mask efficacy over interval *w*
- M_{ijw} is the value of the contact matrix for row *i* and column *j* (from age group *i* to age group *j*) during time interval *w*

Use the following initial conditions (all units are number of people):

S1	S2	S3	E1, E2, and E3	I1, I2, and I3	R1, R2, and R3
10305660	15281905	12154442	50	50	0

Supplementary files *'population.csv'* and *'ContactMatrix.csv'* contain data on the population counts and the contact matrix for each of the three age groups. The output data provided has counts for S, E, I, and R for each of the three age strata. The output files are called '*S3SimulationRuns.csv'* and '*S3SimulationRuns.RDS'*. These have the same information but in a slightly different format.

The contact matrix, M, is provided in the supplementary file, but is also written below in Table 1.

Table 1. Contact matrix for Scenario 3: Causal Analysis with Interventions. Units are average number of contacts per day.

	Age Group 1	**Age Group 2**	**Age Group 3**
Age Group 1	38.62	20.56	6.12
Age Group 2	20.56	28.22	11.60
Age Group 3	6.12	11.60	20.01

In the simulation, two interventions happen simultaneously:

Masking

- From t = 0 to t = 50 days, no masking occurs.
- From t = 50 to t = 100 days, some masking happens ($m_{cw} = 0.5, m_{ew} = 0.6$) and spread of Covid decreases.
- From t = 100 to t = 150 days, masking still happens ($m_{cw} = 0.4, m_{ew} = 0.2$) but with less intensity.

Social distancing

- From t = 0 to t = 20 days, no social distancing occurs.
- From t = 20 to t = 80 days, social distancing happens, reducing contact rates to 30% of their original values across the board.
- From t = 80 to t = 150 days, social distancing happens, reducing contact rates to 80% of their original values across the board.

This simulation is deterministic, but we draw β randomly from a distribution and run the simulation 25 times with slightly different values of β. This is intended to allow us to ask questions about uncertainty.

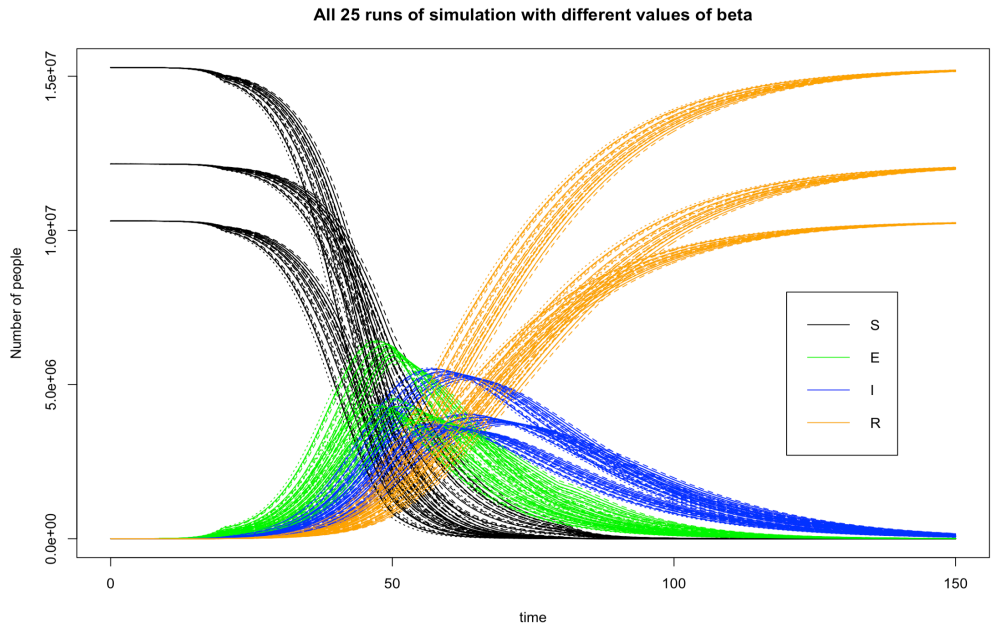

Figure 3. All twenty-five runs of the simulation stratified by age group

1. **Model Extraction (see S1Q1 for definition of model extraction)**: Extract the model and set default parameters and initial conditions. For now, use a dummy value for β. Note the time to extract the model and get it into an executable state that can run a simple test simulation and get sensible results. For workbench modelers, model extraction time may include human-in-the-loop curation, and for baseline modelers, this time may include debugging code. Provide simulation results from your test simulation.

2. **Model Calibration**:
    - Calibrate the model to estimate β in all 25 runs of data provided. Since all runs were generated using a different value of β, each value of β should be a little different. Save the values of β for use in Q5 and Q6.
    - Average all 25 runs together and calibrate a model to estimate β using the averaged data. Use this calibrated model and averaged data for Q3 and Q4.

3. **Causal Effects**: Estimate the average treatment effects on infections, for each of the last four unique intervals, using the following approach:
    - To estimate the average treatment effect (ATE) for the *nth* time interval, use your calibrated model from Q2b, parameterized only for the *(n-1)th* time interval, and generate a forecast of the model over the *nth* interval, where no change in interventions take place (you assume the interventions in place in the *(n-1)th* interval continue uninterrupted). Compute the root mean squared error (RMSE) between the model forecast of infections in the *nth* interval and the average of the supplementary data for the *nth* interval. This is the ATE for the *nth* interval.

For example, calibrate a model using data from time interval (0, 20), and simulate the model over the interval (20, 50). Compare the simulated output over the interval (20, 50) to the average of the provided data for the interval (20, 50) to estimate the ATE of the set of interventions in the interval (20, 50) (as defined originally in the scenario background), on infections.

- For each interval you calculate ATE for, generate plots comparing the actual data (all compartments) to the forecasted output had there been no change in interventions.
- Include uncertainty in the estimated effects.

4. **Interventions**: Use your fitted model from Q2b to conduct an approximation of a sensitivity analysis.
    - Change the original reduction in contact matrix at t = 20 days to 40% of the original value. How does that affect infections at t = 50 days? Calculate ATE for this change in reduction.
    - Repeat Q4a but change the reduction in contact matrix to 20% of the original value. Show the change using plots and changes in calculated ATE.
    - (Optional) Change the reduction in contact matrix in other ways (e.g., instead of changing from a 30% decrease to 40% decrease, change the 30% decrease to 50% decrease), or by changing which age groups have a reduction in contact rate, to demonstrate how various types and levels of contact reduction can affect outcomes.

5. **Intervention Optimization**: In this question, we will ask you to find the minimum level of mask efficacy needed to ensure that the maximum number of infections in the most populous age group (I_2) is below 5,000,000 people, with 90% confidence.
Without knowledge of the exact distribution from which β is drawn from, but with simulated data provided, one way to approach this is with the following steps:
    a. Examine the values of β from Q2 and fit a distribution to these values. Use this approximate distribution to calculate a β you can use to represent the appropriate quantile for the confidence level.
    b. Using the value of β you calculated in Q5a, determine the minimum level of mask efficacy needed to ensure that the maximum number of infections in I_2 remains below 5,000,000 people. Demonstrate this with a plot of simulation outcomes.

6. **Intervention Optimization**: What is the latest time the first masking intervention (currently at t = 50 days) can start to keep total infections below 11,000,000 people at any point in time in the simulation, with 95% confidence? Assume nothing else in the original simulation specification changes. You can apply a similar procedure to the one in Q5 (find the right β from a fitted distribution, and then optimize over the parameter of interest) to solve this question. Demonstrate your answer with a plot of simulation outcomes.

7. (**Optional**) In this question, we use the original SEIR model defined in the scenario introduction, but with no masking or social distancing interventions. Instead, we provide data generated from an SEIR model where β varies at every time step over the course of the simulation. The data are in the file '*ChangingBeta.csv*'.
    a. Using the original social distancing matrix, configure the SEIR model in 3 different ways using the following values of β: 0.10, 0.13, and 0.16. Keep all other parameters the same (aside from intervention parameters, which are set to 0). Calibrate an ensemble model using the 3 model configurations and the provided data. Compute RMSE between your calibrated ensemble model (infections variable output) and the true infections output in the data provided.
    b. Similarly, calibrate a *single* SEIR model to the simulated data and compute RMSE. This model should have one constant value of β. Compare the calibrated ensemble output from Q7a to the single model calibrated output. Plot both against the true data to demonstrate goodness-of-fit of the calibration.

Scenario 3 Summary Table

Question	Inputs	Tasks	Outputs
Q1	Model description	• Extract equations • Extract parameter values • Iterate/curate extraction and execute model until a test simulation gives reasonable results	• Extracted models grounded with all variables and parameters defined, and with units • Test simulation plot • Time to do model extraction • Time to execute extracted model and plot results
Q2	Simulated data	Calibrate model to data	Calibrated β values, and single calibrated model using averaged data
Q3	Calibrated model from Q2b	• Estimate average treatment effects of interventions, with uncertainty • Plot data against counterfactuals	• Estimated ATE values with uncertainty • Plots showing counterfactual scenarios
Q4	Calibrated model from Q2b	Implement changes in contact matrix	• Plots showing how changing the contact matrix affects the output • Values for average treatment effect
Q5	Simulated data	Conduct optimization for the time of the first masking intervention	A plot showing that infections can be kept below 5 million on any given day for a particular value of mask efficacy
Q6	Simulated data	Conduct optimization for minimum mask efficacy	A plot showing when interventions need to start to keep total infections below 11 million on any given day.
Q7	Simulated data using a changing value of beta	• Create an ensemble model with 3 configurations of the same model • Calibrate ensemble model to the provided data • Compute accuracy (RMSE) of the single model and the ensemble as compared to the true simulated data	Plots and RMSE calculations showing how well the calibrated single model and the ensemble models fit the simulated data

Decision-maker Panel Questions

1. What is your confidence in understanding model results and tradeoff between potential interventions? Select score on a 7-point scale.
    1. Very Low
    2. Low
    3. Somewhat Low
    4. Neutral
    5. Somewhat High
    6. High
    7. Very High

Explanation: Determine your confidence in being able to assess effectiveness of all interventions considered in the scenario and understand how uncertainty factors into results.

The decision-maker confidence score should be supported by the answers to the following questions:

- Do you understand the effects of interventions on trajectories? Was the effectiveness of interventions communicated?
- Is it clear how to interpret uncertainty in the results? Do you understand the key drivers of uncertainty in the results?
- Did models help you to understand what would have happened had a different course of action been taken in the past? How confident are you that the counterfactual analysis correctly explained what would have happened had a different course of action been taken?
- How confident are you that the analysis correctly identified and attributed responsibility to causal drivers in the scenario?

Scenario 4: Stock and Flow (Workbench Only) *

Estimated % of time: Workbench 10%

This scenario is meant to be conducted by workbench modelers only. There is no baseline for this scenario. This scenario will not be presented in the Decision-maker Panel.

This question evaluates the ability of the workbench to ingest, visualize/inspect, and extend a simple **stock and flow** model. This question will not require simulation of the model.

1. Ingest the *stock_flow_evaluation.mdl* file into Terarium.
2. Visually inspect the model and take a screen capture when you have a working visualization of the full model structure.
3. Which factors (variables or parameters) affect the *transmission rate* parameter in this model? List all that you can identify from model inspection here and how you identified the factor (visual inspection, model code, etc.). Add rows as needed.

Factor	**Source**

4. Do any other factors, besides the *transmission rate* parameter, affect the transition from "susceptible" to "exposed" compartments? List all factors below along with your sources.

Factor	**Source**

5. List the equation governing the term *fatality rate* in this model.
6. List the equation governing the term *isolation effectiveness* in this model.
7. Modify the model to branch the infected "stock" into a symptomatic infected stock and an asymptomatic infected stock. Ensure that the flows into and out of the new stocks are appropriate. Take a screen capture of the final result.
8. *Optional Bonus:* Modify the model to incorporate testing. Use examples from the literature to guide your modification and record your sources as you go.

Scenario 5: The Lac Operon (Workbench Only) *

Estimated % of time: Workbench 15%

This scenario is meant to be conducted by workbench modelers only. There is no baseline for this scenario. This scenario will not be presented in the Decision-maker Panel.

In E. Coli, lactose metabolism is regulated by a group of genes called the *lac operon*. This group is a sequence of six genes that do the following (and are described in Figure 4):

- The last three genes (*z, y, a*) in the sequence are transcribed and translated into three enzymes that are involved in the metabolism of lactose (which allows bacteria to use the sugar lactose as an energy source):
    i. *β-galactosidase* (synthesized from gene *z*), which degrades lactose by splitting it into *glucose* and *galactose*
    ii. *lactose permease* (synthesized from gene *y*), a protein that transports lactose into the cell
    iii. *transacetylase* (synthesized from gene *a*), which has a smaller, supporting role in the metabolism process
- The first three genes (*i, p, o*) regulate enzyme production:
    i. Gene *i* is transcribed and translated into a protein called the *lac repressor*
    ii. When lactose is not present in the environment, the *lac repressor* binds to gene *o* (the operator gene), and blocks *RNA polymerase (RNAP)* from binding to gene *p* (the promoter gene) to start the transcription of genes *z*, *y*, and *a*, mentioned above.
    iii. When lactose is present, it binds to the lac repressor and removes it from gene *o*, which then allows genes *z*, *y*, and *a* to be transcribed, and the enzymes *β-galactosidase, lactose permease, and transacetylase*, to be synthesized.
- Note that this is a somewhat simplified model, and in reality, the absence and presence of glucose also plays a role in regulating these processes.

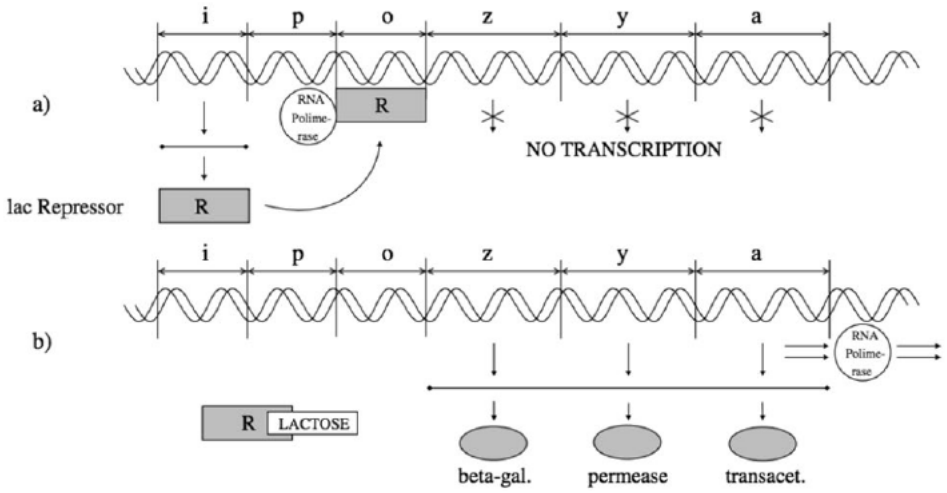

Figure 4. Regulation of lactose metabolism

A simplified representation of lac operon dynamics can be described by the following chemical reactions and kinetic constants.

*Table 2. Reactions and Kinetic Constants for Scenario 5: The Lac Operon (Workbench Only) **

#	Chemical Reaction	Reaction Description	Kinetic Constants (1/seconds)
1	$i \xrightarrow{k_1} i + r_I$	Gene i is transcribed into r_I (the mRNA for *lac repressor* protein, represented as I)	$k_1 = 0.02$
2	$r_I \xrightarrow{k_2} r_I + I$	r_I is transcribed into *lac repressor* protein I	$k_2 = 0.1$
3	$I + Lactose \underset{k_4}{\overset{k_2}{\rightleftharpoons}} I \cdot Lactose$	I interacts with *lactose* and becomes bound to it (represented by $I \cdot Lactose$), in equilibrium with the reverse process	$k_3 = 0.005$ $k_4 = 0.1$
4	$I + Op \underset{k_6}{\overset{k_5}{\rightleftharpoons}} I \cdot Op$	I binds with the *lac operon* (represented as a single entity *Op* in this context) where the bound complex is represented by $I \cdot Op$; in equilibrium the reverse process also occurs, where I becomes unbound from the *lac operon*	$k_5 = 1$ $k_6 = 0.01$
5	$RNAP + Op \underset{k_8}{\overset{k_7}{\rightleftharpoons}} RNAP \cdot Op$	*RNA polymerase (RNAP)* binds with the *lac operon*, (represented in this context by *Op*) where the bound complex is represented by $RNAP \cdot Op$; in equilibrium the reverse process also occurs, where *RNAP* becomes unbound from the *lac operon*	$k_7 = 0.1$ $k_8 = 0.01$
6	$RNAP \cdot Op \xrightarrow{k_9} Op + RNAP + r_{lac}$	*RNAP* bound to the *lac operon,* transcribes genes *z, y, and a* into a single mRNA r_{lac}	$k_9 = 0.03$
7	$r_{lac} \xrightarrow{k_{10}} r_{lac} + Z$	mRNA r_{lac} is translated into β-*galactosidase* (Z)	$k_{10} = 0.1$
8	$Lactose + Z \xrightarrow{k_{11}} Z$	*Lactose* is degraded by β*-galactosidase;* the details of how it is split into *glucose* and *galactose* are omitted	$k_{11} = 1e-5$
9	$r_I \xrightarrow{k_{12}} \emptyset$	mRNA r_I is degraded	$k_{12} = 0.01$
10	$I \xrightarrow{k_{13}} \emptyset$	Protein I is degraded	$k_{13} = 0.002$
11	$I \cdot Lactose \xrightarrow{k_{14}} Lactose$	Protein I bound to lactose is degraded	$k_{14} = 0.002$
12	$r_{lac} \xrightarrow{k_{15}} \emptyset$	mRNA r_{lac} is degraded	$k_{15} = 0.01$
13	$Z \xrightarrow{k_{16}} \emptyset$	β*-galactosidase* is degraded	$k_{16} = 0.01$

Molecular interactions are discrete in nature and are ideally represented by stochastic simulations. However, for the purpose of this scenario, we ignore this aspect and estimate the chemical reactions by the following system of deterministic ODEs:

$$\frac{d[i]}{dt} = 0$$
$$\frac{d[r_I]}{dt} = k_1[i] - k_{12}[r_I]$$
$$\frac{d[I]}{dt} = k_2[r_I] - k_3[I][Lactose] + k_4[I \cdot Lactose] - k_5[I][Op] + k_6[I \cdot Op] - k_{13}[I]$$
$$\frac{d[Lactose]}{dt} = (k_4 + k_{14})[I \cdot Lactose] - k_3[I][Lactose] - k_{11}[Z][Lactose]$$
$$\frac{d[I \cdot Lactose]}{dt} = k_3[I][Lactose] - (k_4 + k_{14})[I \cdot Lactose]$$
$$\frac{d[Op]}{dt} = k_6[I \cdot Op] - k_5[I][Op] - k_7[Op][RNAP] + (k_8 + k_9)[RNAP \cdot Op]$$
$$\frac{d[I \cdot Op]}{dt} = k_5[I][Op] - k_6[I \cdot Op]$$
$$\frac{d[RNAP]}{dt} = (k_8 + k_9)[RNAP \cdot Op] - k_7[Op][RNAP]$$
$$\frac{d[RNAP \cdot Op]}{dt} = k_7[Op][RNAP] - (k_8 + k_9)[RNAP \cdot Op]$$
$$\frac{d[r_{lac}]}{dt} = k_9[RNAP \cdot Op] - k_{15}[r_{lac}]$$
$$\frac{d[Z]}{dt} = k_{10}[r_{lac}] - k_{11}[Z][Lactose] - k_{16}[Z]$$

Where:

- i is lac operon gene i
- r_I is the mRNA for the lac repressor protein I
- $I \cdot Lactose$ is the protein I bound to *lactose*
- Op is the *lac operon* and in this model, is treated as a single entity in the context of binding reactions
- $I \cdot Op$ is the protein I bound to the *lac operon*
- $RNAP$ is RNA Polymerase (the enzyme that transcribes DNA into RNA)
- $RNAP \cdot Op$ is RNA Polymerase bound to the *lac operon*
- r_{lac} is the mRNA created from the transcription of the *lac operon* genes
- Z is *β-galactosidase*, one of the enzymes translated from the mRNA r_{lac}
- [ ] represents concentration of a species, in units of number of molecules per cell

1. Implement and simulate the dynamics of the lac operon as a regulatory network in a single cell, using parameter values from Table 2, and initial conditions from Table 3, for a time period of 1000 seconds, and with the following conditions:

a. Simulate the system when there is no lactose present in the environment (let $[Lactose]_0$ = 0 molecules/cell). Plot the mRNA and protein trajectories. When does Z reach peak concentration, and what is the peak concentration amount?
b. Simulate when there is lactose present in the environment (let $[Lactose]_0$ = 500 molecules/cell). Plot the mRNA and protein trajectories. When does Z reach peak concentration, and what is the peak concentration amount?
c. How does the Z trajectory in the presence of lactose, compare to its trajectory in the absence of lactose?

*Table 3. Initial Conditions for Scenario 5: The Lac Operon (Workbench Only) *, Question 1. All concentrations are in units of number of molecules per cell*

$[i]_0$	$[r_I]_0$	$[I]_0$	$[Op]_0$	$[I \cdot Op]_0$	$[RNAP]_0$	$[RNAP \cdot Op]_0$	$[r_{lac}]_0$	$[Z]_0$	$[I \cdot Lactose]_0$
1	0	50	1	0	100	0	0	0	0

2. Now simulate the model for 1000 seconds, where lactose is not present for the first 500 seconds, and 2000 molecules of lactose are added to the cell at t = 500 seconds. Plot and comment on the behavior of the mRNA and protein trajectories over the course of the simulation.