

# Hackathon Scenarios for ASKEM 6-Month Milestone

Epidemiology Use Case

*Updated January 11<sup>th</sup>, 2023*

Nirupama Bhattacharya, Zachary Turner, Sam Malloy, Robyn Kozierok, Lynette Hirschman

The views, opinions, and/or findings contained in this report are those of The MITRE Corporation and should not be construed as an official Government position, policy, or decision, unless designated by other documentation. This technical data deliverable was developed using contract funds under Basic Contract No. W56KGU-18-D-0004.

## Contents

Hackathon Scenarios.....	2
Scenario 1: Vaccination .....	2
Scenario 2: Limiting Hospitalizations.....	5
Scenario 3: Limiting Deaths .....	8
Scenario 4: Testing and Return to Campus .....	11
Challenge Scenario .....	14

# Hackathon Scenarios

## Scenario 1: Vaccination

### Scenario Ask:

While perusing publications on Covid-19 models, you come across a model from early 2020 that was developed to describe the first Covid wave in Lombardy, Italy. You're interested in updating this model for 2022, to include vaccinations.

1. Before updating the model, you want to make sure you have a good understanding of the original model, can execute it, and reproduce the results in the publication describing the model. The paper doesn't include code, but you think it's feasible to create an executable version of the model and reproduce the results based on the model descriptions in the paper alone. The paper DOI is: 10.3389/fpubh.2020.00230. There are three 'unit tests' to ensure the model representation that we want to execute, is correct: Reproduce the results in Figs. 2A, 3A, and 3B
  - a. [Challenge] Ingest model and pass unit tests from publication alone (do not start with any code as input)
  - b. Ingest model and pass unit tests from publication and corresponding Code Version A
  - c. Ingest model and pass unit tests from publication and corresponding Code Version B
2. Update the model from Question 1, to include vaccination. There are a number of ways to implement vaccination in an epidemiological model, but no matter the modeling approach, it should have an impact on one or more disease outcomes (e.g infections or deaths). Ensure your updated model is not the same as the model referenced in question 3a. Aside from these guidelines, there are no restrictions on modeling choices. If it is not clear how to update the model, do a small literature review/search, to understand how other published models account for vaccination.
3. Model Comparison:
  - a. In addition to your updated model from Question 2 you are aware of the following two specific models that include vaccination
    - i. You find a publication that adds vaccination to the original model from question 1, at <https://biomedres.us/pdfs/BJSTR.MS.ID.007413.pdf>. (Please note the formatting error on pg. 4, where the first term in the equation for  $\frac{dS}{dt}$  should be  $\mu N$ )
    - ii. You are also aware of the CHIME SVIIvR model (which adds vaccination to the original CHIME model, and was part of the starter kit)
  - b. Do a structural model comparison between the models in questions 2, 3.a.i, and 3.a.ii. The structural comparison should include a summary or diagram describing similarities and differences between the models, with respect to parameters, variables/states, pathways, etc.
  - c. Compare simulation outputs between the three models, for the following two scenarios. Assume initial values and parameter values are consistent (to the extent possible) with Table 1 in <https://biomedres.us/pdfs/BJSTR.MS.ID.007413.pdf>. For initial values that are not specified, choose reasonable values and ensure they are the same between the three models being compared.
    - i. Vaccine efficacy = 75%, population vaccinated = 10%
    - ii. Vaccine efficacy = 75%, population vaccinated = 80%

Distribution Statement A. Approved for public release: distribution is unlimited

MITRE Public Release Case Number 24-2089

© 2025 The MITRE Corporation.

4. Create an equally weighted ensemble model using the three models in 3b, and replicate the scenarios in 3.c.i and 3.c.ii. How does the ensemble model output compare to the output from the individual component models?
5. For any of the models in question 3, conduct a sensitivity analysis to determine which intervention parameters should be prioritized in the model, for having the greatest impact on deaths – NPIs, or vaccine-related interventions?
6. For any of the models in question 3, add age stratification to the model and leverage data from the provided contact matrix and following resources. You may ignore vital dynamics. Assume that vaccination status does not have an impact on contact rates between age groups. Assume age-specific vaccination, vaccine effectiveness, hospitalization, and mortality rates, if relevant to the model. For other parameters, you may find reasonable values from the literature (including any of the papers referenced in this scenario) and/or make simplifying assumptions about whether they have different values based on age group.
  - a. For age-specific vaccine effectiveness parameters – you can utilize data compiled by the US CDC, available here (<https://covid.cdc.gov/covid-data-tracker/#vaccine-effectiveness>). You can assume that only mRNA vaccines are used and that efficacy data in Italy would be similar to that of the United States. For a search and discovery challenge, you can try to identify vaccine utilization by manufacturer in the target area and align this data with vaccine-specific efficacy data across age groups for the time window in question. This task should not, however, be a limiting factor in making progress on subsequent downstream TA tasks.
  - b. You may find Italy population distribution data/information, or vaccination rates by age group, from any source, or make a simplifying assumption about similarities with data from the United States.
  - c. See provided contact matrix – “Italy\_contact\_matrix.csv”. Matrix values represent mean number of contacts that an individual from an age group represented by each row, would encounter with age groups represented by each column. There are 16 five-year age groups from 0-80 years, with X1 representing the youngest age group, and X16 representing the oldest age group.
  - d. With the age-stratified model, simulate the following situations. You may choose initial values that seem reasonable given the location and time, and you can reuse values from any of the publications referenced):
    - i. High vaccination rate among older populations 65 years and older (e.g. 80%+), and low vaccination rate among all other age groups (e.g. below 15%)
    - ii. High vaccination rate among all age groups
    - iii. Repeat d.i and d.ii, but now add a social distancing policy at schools, that decreases contact rates by 20% for school-aged children only.
    - iv. Compare and summarize simulation outputs for d.i-d.iii

Question	Task	TA Workflow Tested	Metrics
1	Model extraction and unit testing	TA1: Model Extraction, Model Execution/Unit Tests	<b>Time:</b> How long does knowledge extraction take? How long does it take to get model into executable form? This includes time to iterate on unit test(s) until you are confident the output is correct. <b>Accuracy:</b>

			<ul style="list-style-type: none"> <li>• Were you able to faithfully reproduce results of unit tests?</li> <li>• Qualitative score on metadata quality (correctness, relevance, completeness), based on human inspection of the equations, variables, parameters, etc.</li> </ul>
2	[Optional] Literature Review: to explore how others have accounted for vaccination	TA1: Search and Discovery	<b>Time:</b> How long does search/review take? <b>Relevance (qualitative):</b> How relevant are the findings of the search/lit review, to the question?
2	Model extension/transformation: update the model to include vaccination	TA2: Model Transformation	<b>Time:</b> To extend/modify model
3.a.i	Do model extraction and ensure model can be executed	TA1: Model Extraction; Model Execution	<b>Time:</b> How long does knowledge extraction take? How long does it take to get model into executable form? <b>Quality:</b> Qualitative score on metadata quality (correctness, relevance, completeness), based on human inspection of the equations, variables, parameters, etc.
3.a.ii	Search for model, and ensure model can be executed	TA1: Search and Discovery; Model Extraction; Model Execution	<b>Time:</b> How long does it take to find this model, and ensure it is executable with some set of default parameters?
3b	Model Comparison	TA2: Model Comparison	<b>Time:</b> To execute model comparison <b>Quality (qualitative):</b> Is model comparison output interpretable and does it capture major differences and similarities correctly?
3c	Model simulation according to scenario	TA3: Simulation Workflow	<b>Time:</b> How long does it take to put together simulation workflow and get output?
4	Create ensemble model and do simulations according to scenario	TA3: Simulation Workflow; Answers to Scenario Questions	<b>Time:</b> How long does it take to put together simulation workflows and get a final answer?
5	Sensitivity analysis	TA3: Simulation Workflows; Answers to Scenario Questions	<b>Time:</b> Assembly and execution of sensitivity analysis <b>Quality:</b> Are the sensitivity estimates reasonable given known structural features of the model?
6	Model extension/transformation: add stratification	TA2: Model Transformation	<b>Time:</b> To extend/modify model
6d	Simulations according to scenario	TA3: Simulation Workflow	<b>Time:</b> How long does it take to put together simulation workflow and get output?

## Scenario 2: Limiting Hospitalizations

**Scenario Background:** You are a disease modeler supporting the Los Angeles County Department of Public Health, at the beginning of the original Omicron wave. The LA County Board of Supervisors is concerned about what the next few months will look like, and what level of intervention will be required to manage what is shaping up to be a large winter Covid-19 wave. Vaccines were broadly available during this time period and vaccination should be accounted for in the modeling.

### Scenario Setting/Situation:

Time = December 28th, 2021 (right around upswing of Omicron wave), Location = LA County

### Scenario Asks:

1. Find a model capable of forecasting Covid cases and hospitalizations (these don't need to be broken down by vaccination status, but the model should account for vaccination in some way). Parameterize model either using data from the previous two months (October 28<sup>th</sup> – December 28<sup>th</sup>, 2021), or with relevant parameter values from the literature. Forecast Covid cases and hospitalizations over the next 3 months under no interventions.
2. Based on the forecast, do we need interventions to keep total Covid hospitalizations under a threshold of 3000 on any given day? If there is uncertainty in the model parameters, express the answer probabilistically, i.e., what is the likelihood or probability that the number of Covid hospitalizations will stay under this threshold for the next 3 months without interventions?
3. Assume a consistent policy of social distancing/masking will be implemented, resulting in a 50% decrease from baseline transmission. Assume that we want to minimize the time that the policy is in place, and once it has been put in place and then ended, it can't be re-implemented. Looking forward from "today's" date of Dec. 28, 2021, what are the optimal start and end dates for this policy, to keep projections below the hospitalization threshold over the entire 3-month period? How many fewer hospitalizations and cases does this policy result in?
4. Independent from #3, assume there is a protocol to kick in mitigation policies when hospitalizations rise above 80% of the hospitalization threshold (i.e. 80% of 3000). When hospitalizations fall back below 80% of the threshold, these policies expire.
  - a. When do we expect these policies to first kick in?
  - b. What is the minimum impact on transmission rate these mitigation policies need to have the first time they kick in, to (1) ensure that we don't reach the hospitalization threshold at any time during the 3-month period, and (2) ensure that the policies only need to be implemented once, and potentially expired later, but never reimplemented? Express this in terms of change in baseline transmission levels (e.g. 10% decrease, 50% decrease, etc.).
5. Now assume that instead of NPIs, the Board wants to focus all their resources on an aggressive vaccination campaign to increase the fraction of the total population that is vaccinated. What is the minimum intervention with vaccinations required in order for this intervention to have the same impact on cases and hospitalizations, as your optimal answer from question 3? Depending on the model you use, this may be represented as an increase in total vaccinated population, or increase in daily vaccination rate (% of eligible people vaccinated each day), or some other representation.

Question	Tasks	TA Workflow Tested	Metrics
1	<p>Find a model that meets the scenario requirements (or can be updated to meet the requirements).</p> <p>Model requirements:</p> <ul style="list-style-type: none"> <li>Needs to output infections and hospitalizations (or can be modified to do so)</li> <li>Needs to support NPIs that effect a change in baseline transmission levels, including support for interventions implemented at different points in time, and for various lengths of time, which may depend on the values of output variables (hospitalizations in particular). The model should allow the user to choose the specific start date of the intervention, and length of time in place, etc., or set triggers for intervention start and end times.</li> <li>Needs to represent vaccination with some way to increase level of vaccination in population</li> </ul>	TA1: Model Search and Discovery	<p><b>Time:</b> How long does search for appropriate model take?</p> <p><b>Relevance (qualitative):</b> How suitable is the selected model for the scenario described?</p>
1	[Optional] If relevant model is not already in the ASKEM system, do model extraction from paper/code	TA1: Model Extraction	<p><b>Time:</b> How long does knowledge extraction take? How long does it take to get model into executable form?</p> <p><b>Quality (qualitative):</b> Qualitative score on metadata quality (correctness, relevance, completeness), based on human inspection of the equations, variables, parameters, etc.</p>
1	[Optional] If required, do necessary model extension/transformation to meet scenario requirements	TA2: Model Extension/Transformation	<b>Time:</b> to extend/modify model
1	Parameterize model according to time and location. This may require a literature search, or search for data to inform model parameters.	TA1: Search and Discovery (for parameters)	<b>Time:</b> How long does search for information required to fully parameterize the model?
1-5	Answer questions 1-5, with supporting output to justify the answers.	TA3: Simulation Workflows; Answers to Scenario Questions	<b>Time:</b> For each question, measure time to set up simulation workflows, get final answer and prepare supporting output.

			<b>Quality (qualitative):</b> Does output seem reasonable given the scenario?
	[ <i>Optional</i> ] If at any point, you need to search for parameter values, do a literature review, or find datasets, please track time spent, approach taken (e.g. what were the keywords or key concepts you searched by), and sources/databases you searched across.	TA1: Search and Discovery	<b>Time:</b> How long does search for required information take?

## Scenario 3: Limiting Deaths

### Scenario Setting:

- Timeframe = May 1, 2020 to June 15, 2020
- Location = State of Massachusetts

You are part of a modeling team supporting the State of Massachusetts, that has used a relatively simple deterministic model to provide projections at the start of the pandemic in early 2020. You now wish to provide more rigorous and insightful recommendations to policymakers by including probabilistic concepts in your projections. You have identified a Bayesian model called [MechBayes](#) that forecasts COVID-19 cases and deaths. You want to explore some stochastic aspects of the model and quantify the uncertainty in your forecasts, as well as determine if we need interventions to reduce deaths below a given threshold.

By default, the model outputs forecasts for daily cases, daily deaths, cumulative cases, and cumulative deaths. The questions below will focus on different approaches to minimize the number of cumulative deaths and cumulative cases.

For the questions below, use the regular SEIRD model provided in MechBayes.

### Scenario Ask:

1. Provide a forecast of cumulative Covid-19 cases and deaths over the 6-week period from May 1 – June 15, 2020 under no interventions, including 90% prediction intervals in your forecasts. Compare the accuracy of the forecasts with true data over the six-week timespan.
2. Based on the forecasts, do we need additional interventions to keep cumulative Covid deaths under 6000 total? Provide a probability that the cumulative number of Covid deaths will stay under 6000 for the next 6 weeks without any additional interventions.
3. We are interested in determining how effective it would be to institute a mandatory mask mandate for the duration of the next six weeks. What is the probability of staying below 6000 cumulative deaths if we institute an indefinite mask mandate starting May 1, 2020?
  - a. Consult with TA1 to identify literature or pre-existing data that could be used to translate the implementation of the intervention into the model via the transmission parameter (or by other means that you identify). The translation of the intervention into the model would ideally include uncertainty related to compliance levels with the policy, which should also be informed by the literature or other sources.
4. This model includes a detection rate parameter. As the [paper](#) says,

*"We allow the detection rate to vary over time following a Gaussian random walk on the log-odds scale, as shown above. This is meant to loosely model changes in diagnostic testing over time; in practice, it provides flexibility in the model that likely captures other changes in the relationship between reported cases and deaths over time, such as changes in the fatality ratio of the population infected at a given time."*



- a. We are interested in determining how detection rate can affect the accuracy and uncertainty in our forecasts. In particular, suppose we can improve the baseline detection rate by 20%, and the detection rate stays constant throughout the duration of the forecast. Assuming no additional interventions (ignoring Question 3), does that increase the amount of cumulative forecasted cases and deaths after six weeks? How does an increase in the detection rate affect the uncertainty in our estimates? Can you characterize the relationship between detection rate and our forecasts and their uncertainties, and comment on whether improving detection rates would provide decision-makers with better information (i.e., more accurate forecasts and/or narrower prediction intervals)?
  - b. Compute the accuracy of the forecast *assuming no mask mandate* (ignoring Question 3) in the same way as you did in Question 1 and determine if improving the detection rate improves forecast accuracy.
5. Model modification, model space exploration and model selection.
  - a. Convert the MechBayes SEIRHD model to an SIRHD model by removing the E compartment. Compute the same six-week forecast that you had done in Question 1a and compare the accuracy of the six-week forecasts with the forecasts done in Question 1a.
  - b. Further modify the MechBayes SEIRHD model and do a model space exploration and model selection from the following models, based on comparing forecasts of cases and deaths to actual data: SEIRD, SEIRHD, and SIRHD models. Use data from April 1, 2020 – April 30, 2020 from the scenario location (Massachusetts) for fitting these models. Then make out-of-sample forecasts from the same 6-week period from May 1 – June 15, 2020, and compare with actual data. Comment on the quality of the fit for each of these models.
  - c. Do a 3-way structural model comparison between the SEIRD, SEIRHD, and SIRHD models.
6. Challenge question: MechBayes comes with several modeling types, including SEIRD and SEIRD Renewal. The MechBayes paper does not describe the renewal model in much detail – it only cites a different [paper](#) that implemented a renewal model. (In fact, the renewal model was added to the GitHub repository months after the MechBayes paper went on MedRxiv).

Use TA1/TA2 tools to elucidate the differences between the SEIRD model and the SEIRD Renewal model. Can you explain and visually describe how these models differ?

7. Optional (if time): What is the latest date we can impose a mandatory mask mandate over the next six weeks to ensure, with 90% probability, that cumulative deaths do not exceed 6000? Can you characterize the following relationship: for every day that we delay implementing a mask mandate, we expect cumulative deaths (over the six-week timeframe) to go up by X?

Question	Tasks	TA Workflow Tested	Metrics
3	Search for information on how masking affects transmission rate (or would change the SEIRD model more generally)	TA1: Search and discovery for model, parameters	<b>Time:</b> How long does the search take? <b>Quality (qualitative):</b> Is the information about changing the transmission parameter, or a change to the model at large, reasonable?
5	Model transformation/extension as specified by scenario (for example, to add mask mandate, to modify model structure, etc.), model space exploration	TA2: Model transformation/extension; Model space exploration	<b>Time:</b> How long do transformations take? How long does model space exploration take? <b>Quality:</b> Do transformation and model selection results make general sense?
1-7	Simulations and answer to scenario questions	TA3: Simulation workflows	<b>Time:</b> How long does this take compared against a manual workflow? <b>Quality/completeness:</b> do the forecasts change in expected or reasonable ways? <ul style="list-style-type: none"> <li>- Does the mask mandate reduce cases/deaths?</li> <li>- Do the changes in forecasts or uncertainty we see when detection rate is increased make sense?</li> </ul>
5.c, 6	<ul style="list-style-type: none"> <li>• Model comparison of models in model space exploration</li> <li>• Challenge: Comparing SEIRD and SEIRD Renewal Model <ul style="list-style-type: none"> <li>- Ingest additional information</li> <li>- Come up with a comparison of SEIRD model and SEIRD renewal model</li> </ul> </li> </ul>	TA1: Search and discovery; model ingestion TA2: Model Comparison	<b>Time:</b> How long does it take to ingest the additional information (the paper and its artifacts) into the ASKEM system? How long does model comparison and visualization take? <b>Quality (qualitative):</b> Is model comparison output interpretable and does it capture major differences and similarities correctly?

## Scenario 4: Testing and Return to Campus

### Scenario Background

Universities face unique challenges when evaluating return-to-campus strategies at the beginning of each term. Testing programs are a key tool in maintaining outbreak control by identifying and interrupting transmission chains, but these programs can be prohibitively expensive when executed at-scale. At large institutions, testing programs are often linked with quarantine and isolation policies. Optimization strategies often rely on disease models to understand courses of action and their respective tradeoffs and seek to balance continuity of operations, limiting negative public health outcomes, and the financial cost of imposing interventions such as testing campaigns and maintaining isolation rooms.

Put yourself in the shoes of a modeling team supporting the COVID task force leader at a large university responding to short-turn requests from the President's office regarding the return-to-campus strategy.

**Specifically, you have been asked to provide guidance regarding the campus testing strategy.** How often should undergraduate students, graduate students, and faculty/staff be tested, and via which test type? What minimum weekly testing frequency (and modality) across cohorts would be required to maintain infections below isolation capacity?

### Setting/Assumptions/Details:

- Decision-maker being supported: University of Michigan president
- Time/Setting: It is late 2021 and you are planning for the Spring 2022 term at the University of Michigan (Ann Arbor campus) beginning in early January 2022. For the purpose of this scenario, consider a four-month period that begins January 1st and ends May 1st.
- Assume the following numbers of true infections at the onset of the term by population type:
  - Undergraduate=750
  - Graduate/professional=250
  - Faculty/staff=100
- Students can be sent to isolation dorms, but isolation dorms are expensive and have a capacity ceiling (430, in this case at the beginning of the term).
- Available interventions: The two primary testing mechanisms to consider for this scenario are rapid antigen tests and nucleic acid amplification tests (PCR type testing). Rapid antigen tests are less expensive and offer faster results but have been less reliable throughout the COVID-19 pandemic in the sense that the sensitivity of the tests is often lower (more false negatives). Reporting compliance is also challenging with rapid tests. For this exercise, you may either assume that tests are observed, or you may consider the impact of underreporting results (positive or negative). Specific data regarding sensitivity, turnaround time, and cost should be sourced from the literature when needed to address the questions below.
- Cohorts of interest: (1) undergraduate students; (2) graduate/professional students (single combined group for purpose of this exercise); and (3) employees (faculty + staff as a single combined group for purpose of this exercise). Contact patterns between groups should be drawn from the literature, though you may assume generalities for this exercise (i.e., general

population patterns for employees versus age and setting specific patterns for students). These contact patterns should be used to inform cohort-specific (within and across) mixing estimates.

#### Scenario Questions:

1. Define a return-to-campus strategy that minimizes total testing while maintaining infections below the initial isolation bed capacity of 430. The testing scheme can include an arrival testing strategy in addition to unique testing approaches within time periods of the simulation. Cohorts can have unique testing strategies defined by test type and number per week.
2. Identify an appropriate model and assign parameter values from a literature search. The model should support testing interventions or be able to be modified to include testing, and the model should be able to support different cohorts/strata, or be transformed/updated to do so. It may be prudent to start with an already ingested model as a starting point. TA3 may return to TA1 for reach-back at a later point to identify and ingest an alternative model that readily facilitates the concept of a testing campaign in a more nuanced way.
  - a. Delineated by test type (antigen and PCR, respectively), both manual users and ASKEM users may need to extract estimates for the following: Test sensitivity; test turnaround time; test compliance. Both will also need to identify transmission estimates for the time period and hospitalization rates given infection.
3. The model will need to include cohort stratification and appropriate treatment of the testing campaign as an intervention. The user will need to produce and maintain distinct cohorts for undergraduate student population, graduate/professional student population, and employee population.
  - a. The testing campaign needs to be implemented on a weekly cadence with the ability to modulate the number and type of tests applied across cohort. For a simplification, we can start with just one test type.
4. **Challenge** question: assume that antigen tests are one fifth the cost of PCR tests but also much less (~half) as sensitive. Incorporate the cost of the testing program into your recommendations.

Tasks	TA Workflow Tested	Metrics
Search for appropriate model and parameter values	TA1: Search and Discovery for model, parameters	<b>Time:</b> How long does the search take? <b>Quality (qualitative):</b> Is the model selection reasonable? Are key parameter values for this scenario sensible?
Model transformation to include cohort stratification	TA2: Model Transformation	<b>Time:</b> How long does it take to do the stratification? <b>Quality (qualitative):</b> Are the cohorts created/maintained through the simulation appropriately? Is the intervention (testing campaign) appropriately captured in the model transformation?

Simulations and answer to scenario question work products.	TA3: Simulation workflows	<b>Time:</b> How long does this take compared against a manual workflow? <b>Quality/completeness:</b> how rigorously is the question examined? Are the results sensible?
--	---------------------------	---

## Challenge Scenario

The original Bucky model is structured to handle population data stratified in 16 5-year bins, as described [in the documentation](#). You've recently found a publication about an age-structured SIR model describing the spread of Covid in the state of Washington, USA, which was 'ground zero' of the Covid-19 pandemic in the United States, with the first confirmed case and first confirmed death in the country. (<https://doi.org/10.1038/s41598-021-94609-3>). Modify the Bucky model to use data stratified in 9 10-year bins, as shown in the age-contact matrix in Figure 1. Simulate the first 3 months of the Covid-19 pandemic in Washington, using the modified Bucky model.