12-Month ASKEM Hackathon Scenarios

Nirupama Bhattacharya, Zachary Terner, Sam Malloy, Robyn Kozierok

To prepare for the 12-month evaluation scenarios, we have developed a series of hackathon scenarios that include analogous components of the evaluation scenarios. These questions are meant to help guide and prioritize critical development for success in the evaluation. Compared to the 6-month evaluation, scenarios in the 12-month evaluation will attempt to explore more of the complexity and nuances around the evolutionary nature of the Covid-19 pandemic. The goal is to have all questions, with the possible exception of questions related to agent-based modeling, answerable within the workbench.

Contents

Hackathon Scenario 1: Forecasting with NPIs

Background: In the early stages of an outbreak of a novel pathogen, vaccines, therapeutics, and diagnostics are typically unavailable. Policymakers must resort to nonpharmaceutical interventions like social distancing, masking, increasing hygiene practices, and quarantine/isolation. There are many existing approaches to incorporate nonpharmaceutical interventions into compartmental models.

Time Period: Early 2020

Location: Singapore

Questions: You want to implement a masking intervention in a simple compartmental model and simulate epidemic trajectories under different compliance scenarios. You found an existing model that incorporates masking as a time-dependent modification to the β parameter (**https://doi.org/10.3390/ijerph18179027** plus accompanying code), and you want to ensure that the model is working as expected by reproducing plots in the publication.

1) Replicate an analysis from the paper.
   a) *(TA1 Model Extraction Workflow, TA2 Model Representation)* Extract the SEIRD model (equations 1-5, with time-varying β as defined in equations 8-9) and load into the workbench. In equation 8, let $kappa = \gamma R_0$, $R_0 = 5$. In equation 9, let k = 5.
   b) *(TA3 Simulation Workflow, Unit Test)*: Replicate Figure 3 from the paper, which maps to the first scenario in the paper- implementing a masking intervention at several different timepoints (delays of 0 days, 50 days, 100 days, and control case, from the date of first infection) in the pandemic, with 100% compliance. Recreate the Fig. 3 curves (including peak infection times and levels), up to some reasonable margin of error.
   c) *(TA2 Model Modification Workflow, TA3 Simulation Workflow)*: Update the $\beta(t)$ function to be defined as equations 8 and 10, with k1 = 5, and k2 = 1. This reflects the paper's second scenario, gradual noncompliance with the masking policy over time. Rerun the simulation with several different delays in enforcing a mask policy (ranging from 0 to 140 days), and replicate Fig. 5, up to some reasonable margin of error.
   d) *(TA2 Model Modification Workflow, TA3 Simulation Workflow)*: Update the system of equations to include equations 6 and 7. This adds the potential for reinfection. Compare with the outcomes from 1c. What impact does immunity loss and potential for reinfection have?
   e) *(TA3 Simulation Workflow)* Repeat 1d, but now incorporate uncertainty into the simulations, by assessing a range of possible noncompliance levels (as defined by β_{nc} and $k2$) and a range of reinfection rates (as controlled by ϵ).

	Problem 1a-b	**Problem 1c**	**Problem 1d**	**Problem 1e**
Inputs	• Paper and code • Parameters: Defined in problem	• Extracted model from 1b	• Model from 1c	• Model from 1d
Task	• Extract SEIRD model	• Update $\beta(t)$	• Update system of equations	• Add uncertainty to noncompliance

			parameters and reinfection rates • Run probabilistic simulation
	• Set parameters, run simulation	• Run simulation	• Run simulation	
Outputs	• SEIRD outcomes over time • Find peak times and levels of infections • Vary time of mask policy, replicate paper Fig. 3	• SEIRD outcomes over time • Vary time of mask policy, replicate paper Fig. 5	• SEIRD outcomes over time • Compare outcomes against 1c	• SEIRD outcomes over time, with uncertainty ranges

2) **Highly Recommended Exploratory Work**

For additional preparation for the evaluation, the following exercises may be useful:

a) *(TA1 Search and Discovery for Papers/Models/Data, Model Extraction Workflows)* The use of wastewater-based data in analysis of COVID-19 trends is a promising area of active research. Incorporating such nontraditional data into existing modeling frameworks is an important next step to make these data more useful. Use the TA1 pipeline to conduct a search for papers that incorporate wastewater data into compartmental, ODE-based modeling frameworks. Ingest at least one of these models into TERArium, inspect for accuracy of the ingestion and groundings, and explore basic implementation of the models. Search for wastewater datasets, ingest into the workbench, transform according to common approaches in the literature, and plot against observational case count data.

b) *(Lower Priority, TA2 Structural Model Comparison)* Building on the search and discovery task above, conduct a structural model comparison between two or more SIR-type models that incorporate wastewater-based data as an input. Highlight key tradeoffs between models.

c) *(TA2 Model Modification Workflow)* Extend the core model extracted in Problem 1 of this scenario, by adding age stratification, new compartments (e.g., representing diagnosed/undiagnosed individuals), and vaccination, while maintaining the time-varying transmission aspects of the original model. Ensure that the functionality to extend an extracted model is available in the workbench.

d) *(TA3 Simulation Workflow)* Conduct a simple forecasting exercise using the model + real-world interventions. For example, the Oxford COVID-19 Government Response Tracker has catalogued several years of interventions that can be used to inform modeling decision making. Select a state from the 2020/2021 summary reports (https://www.bsg.ox.ac.uk/research/publications/us-state-level-oxcgrt-reports) and use this to construct a simple timeline of a masking intervention. Calibrate the model from Q1a using gold standard data from the Covid-19 ForecastHub for a period before the masking intervention, and construct a forecast that extends through the intervention period. Then demonstrate how the forecast varies based on compliance rates.

Hackathon Scenario 2: Multiple Vaccines and Variants

Models from early in the Covid-19 pandemic were relatively simple, as we had no vaccines available, only NPIs to mitigate the spread of the disease, and only one strain (the original wild type) to worry about. By mid-2021, the situation was much more complex, as we had access to multiple vaccines, each with multiple doses, and the emergence of multiple variants. You are interested in modeling this situation with SV2(AIR)3 (https://www.nature.com/articles/s41598-022-06159-x, https://doi.org/10.1038/s41598-022-06159-x), one of the first models to represent multiple Covid-19 variants, and the competition between them.

1. Extract the simplest version of the model with 1 vaccine and 1 variant (shown in Fig. 1A)
   a. *(TA1 Model Extraction Workflow)* Extract the model from Matlab code *'svair_simple.m'* and *'get_beta.m'* and with the publication equations and Fig. 1A (with the exception of one small difference between code and figure). Note that in this code, the units of all state variables are in terms of fraction of the total population.
   b. *(TA3 Simulation Workflow)* Simulate the first 14 months of the pandemic beginning on January 1st, 2020. Consider only the Wild Type variant, and no vaccination during this period. Use parameter values from the supplementary material, for the wild type. Use the population of Ontario in 2019: 14.57 million people, and let the natural death rate $\mu = 2.05e - 5$. For initial conditions, let $I(0) = 1e - 6, S(0) = 1 - sum(I(0) + A(0))$, and all other values be 0. When do I(t) and A(t) peak, and what is the value of these variables at their peaks? How do the I(t) and A(t) profiles compare with Fig. 3d,f (which considered 3 variants and 2 types of vaccines)? Given that this question assumes only the presence of the Wild Type variant, do the results seem reasonable?
   c. *(TA2 Domain Knowledge Grounding; ASKEM Workbench Only)* Demonstrate that the domain knowledge groundings accurately reflect new concepts with the extensions such as "variants".

Inputs	• Publication with equations, and simplest SVAIR version of the model described in Fig. 1A • Accompanying code
Tasks	• Model extraction • Set parameters according to 1b and supplementary materials • Simulate model
Outputs	• Extracted model grounded to the domain knowledge graph, and simulation outcomes (S, V, A, I, R)

2. Extend the model to incorporate multiple strains and 2 vaccine types, according to publication Fig. 1B and equations #2-12.
   a. *(TA 1 Model Extraction Workflow)* Get the extended model by extracting it directly from the publication, and code at https://github.com/MehrshadSD/SARS-CoV-2-variants-of-concern-in-Ontario.

b. *(TA2 Model Modification/Extension Workflow)* Alternately, use TA2 model extension/transformation tools, to extend the simpler model extracted from Question 1. How does this extended model compare with the extracted model from 2a?

c. *(TA3 Simulation Workflows/ Unit Tests)* As in the paper, simulate the spread of Covid-19 in Ontario, over a 2 year period from January 1st, 2020 – December 31st, 2021, incorporating 2 vaccine types (Pfizer and AstraZeneca), and 3 Covid-19 variants/strains (including Wild Type, Alpha, and Delta, ignoring Omicron at the end of 2021). Set parameter values according to publication supplementary material. Use the 2019 population of Ontario, 14.57 million people. For initial conditions, assume the entire population was susceptible at the beginning of the simulation timeframe, except for 15 infected symptomatic individuals. Reproduce Figs. 3a, b, d, and e.

	Problem 2a	Problem 2b	Problem 2c
Inputs	SV2AIR3 publication + accompanying code	Extracted model from Problem 1	• Either model from 2a or 2b • Parameters: as described in publication and supplementary material document
Task	Extract full SV2AIR3 model	Extend model to create full SV2AIR3 model	Run simulation for January 1st, 2020 – December 31st
Outputs	Extracted model	Extended model, compare against 2a	Reproduce Figs. 3a, b, d, and e.

3. *(TA3 Simulation Workflows)* With the full model, implement the 3 additional scenarios described in the paper and rerun the simulations from January 1st, 2020 – December 31st, 2021. When updating parameters you may choose the exact values to change them to, and qualitatively compare simulation outcomes with Fig. 7.

a. Assume that a smaller percentage of vaccinations were with Pfizer, but keep the total vaccinations the same as in 2c. Given that the Pfizer vaccine offered stronger protection against all variants, compared to AstraZeneca, what is the impact?

b. Assume a lower provincial vaccination rate compared to the baseline percentage of 74% of the population vaccinated by September 1st, 2021.

c. Assume less stringent NPIs than what actually took place.

	Problem 3
Inputs	Full SV2AIR3 model from Problem 2
Task	Implement 3 scenarios described in the paper and problem description
Outputs	Model output for 3 scenarios, and comparison of outcomes

4. Omicron Extension. The research in the paper was done before the Omicron variant, but the authors ran predictions for a fictional variant Neos, designed to be deadlier than previous variants, through increased infectivity, vaccine escape, and fraction of asymptomatic infections. There are some similarities between this fictional variant, and Omicron. Assume you are in the fall of 2021, at a time when Delta was on the downswing, and Omicron was emerging.

   a. *(TA2 Model Modification Workflow)* Simplify the model structure to still consider multiple vaccines, but only the Delta and Omicron variants

   b. *(TA3 Simulation Workflow)* Use the following sources on Omicron, to set relevant parameters in the model, run the model for the time period July – December 2021, and compare with the hypothetical outcomes with Neos, shown in Fig. 9.

      i. Epidemiological data for Ontario can be found through the Google COVID-19 Open Data portal: https://storage.googleapis.com/covid19-open-data/v3/location/CA_ON.csv

      ii. Summary data on Omicron-related characteristics can be found at https://www.cdc.gov/mmwr/volumes/71/wr/mm7137a4.htm

	Problem 4
Inputs	Full SV2AIR3 model from Problem 2
Task	• Simplify model structure to consider only 2 variants • Update model parameters with realistic values, to reflect presence of Omicron • Simulate model for July-December 2021, to include Omicron, and compare with paper Fig. 9 (outcomes for hypothetical variant Neos)
Outputs	Output of modified model with realistic parameters, to reflect emergence of Omicron in later 2021, and comparison with hypothetical variant from paper

5. Change the setting to the United States

   a. *(TA1 Search and Discovery, Data Workflows)* See the publication supplementary material and sources references, to identify which parameters should be updated for the new location. Where needed, search for US-relevant parameter values from the literature. Use US CDC data from https://data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-United-States-Jurisdi/unsk-b7fc, which includes vaccinations by type and dosage. For the US, consider the following two vaccine types: (1) Modern and Pfizer mRNA vaccines, and (2) Janssen /J&J viral vector vaccine.

   b. *(TA3 Simulation Workflows)* After updating parameters for the US context, rerun the simulation for the time period January 1st, 2020 – December 31st, 2021. How do outcomes compare with Question 2c?

   c. **Challenge** *(TA2 Model Modification Workflow)* As noted in the publication, a major limitation of the model is the lack of age structure, as mortality rates and impacts of NPIs affect different age groups at varying levels. Stratify the model to match the number of age groups found in US CDC vaccination and dosage data: https://data.cdc.gov/Vaccinations/COVID-19-Vaccinations-in-the-United-States-

Jurisdi/unsk-b7fc. Then rerun the simulation for the time period January 1st, 2020 – December 31st, 2021. How do outcomes differ by age group? For United States contact matrices, refer to https://doi.org/10.1371/journal.pcbi.1005697; https://doi.org/10.1371/journal.pcbi.1009098

	Problem 5
Inputs	Updated SV2AIR3 model from Problem 4
Task	• Update model parameters to reflect US context • Simulate model for January 1st, 2020 – December 31st, 2021, now reflect evolution of pandemic in United States • Challenge: stratify model by age, and rerun simulation
Outputs	Simulation outcomes for January 1st, 2020 – December 31st, 2021, for the United States

(Challenge) Hackathon Scenario 3: Agent-based Modeling with COmplexVID-19

1) *(TA1/2 Model Extraction and Model Representation Workflow, <u>ASKEM Workbench Only</u>)* Represent COmplexVID-19 in the agreed upon ASKEM ABM representation, interoperable with the needs of all TAs

2) *(TA3 Simulation Workflow/ Unit Test)* Run COmplexVID-19 with the configuration for Brazil at the start of the Covid-19 pandemic, as described in the Results section of <u>Scabini et. al</u>. (<u>https://doi.org/10.1016/j.physa.2020.125498</u>), and recreate Fig. 4a (or a qualitatively similar scaled down version) for cases and deaths:

   a) Run for 100,000 nodes (agents), with a scaling factor of 57, and 100 repetitions. Feel free to adjust these values given simulation runtime limitations and time constraints.

   b) Use the age and family distributions indicated in Table 1 of the publication.

   c) Consider the start of the pandemic to be February 26th, 2020 (the date of the first confirmed case in Brazil), and simulate for 300 days after

   d) Set vary β from 0.2 to 0.35 in increments of 0.05

   e) Implement a moderate quarantine after 27 days (May 26th, 2020), representing isolation measures implemented by most Brazilian states (see paper for description of this intervention).

   f) When do cases and deaths peak, and what are levels of the peaks? Do these match Fig. 4a?

Inputs	• Model: ASKEM version of COmplexVID-19 model (with additional code comments and some minimal translations) • Parameters: Use parameter settings indicated in <u>Scabini et. al</u>
Tasks	• Set parameters • Simulate ABM
Outputs	• Daily cases and deaths, over 300 days of simulation • Dates and levels for when cases and deaths peak • Qualitative comparison of outputs against <u>Scabini et. al</u>, Fig. 4a

3) We now consider changing aspects of the Brazilian population distribution, which impact the agent networks created by the model.

   a) *(TA3 Simulation Workflow)* Adjust the age distribution of the country population, by increasing the population fraction in the 60+ age group by 10%, and decreasing the population fraction in the 0-13 yrs age group and 18-24 yrs age group, by 5%. Run the model for 300 days starting on February 26th, keeping all other parameter settings the same as in Q2. What is the impact of these changes to age distribution, on infections, hospitalizations, and deaths?

   b) *(TA3 Simulation Workflow)* Adjust the distribution of family sizes to the following values. Run the model for 300 days starting on February 26th, keeping all other parameter settings the same as in Q2. What is the impact of this change, on infections, hospitalizations, and deaths?

Family Size	1 person	2 people	3 people	4 people	5 people	6 people	7 people	8 people	9 people	10 people
% of Households	17	19	20	20	15	1	4	2	1.2	0.8

	Problem 3a	**Problem 3b**
Inputs	Model: COmplexVID-19 model from Problem 2	Model: COmplexVID-19 model from Problem 2
Tasks	Change **age** distribution	Change **family size** distribution
Outputs	• Daily cases and deaths, over 300 days of simulation • Comparison of outputs against Q2 outputs	• Daily cases and deaths, over 300 days of simulation • Comparison of outputs against Q2 outputs