November Monthly Epi Demo

Zachary Terner, Jack Mitcham, Nirupama Bhattacharya, Robyn Kozierok, Molly R. Creagar

Table of Contents

Scenario 1: Mpox

Mpox is a zoonotic disease endemic in parts of Africa caused by an orthopoxvirus (same genus as smallpox). There are two pathways to human infection: animal-to-human, and human-to-human. There are two clades, each with subclades, endemic in parts of western and central Africa. These subclades have different disease severity and infectiousness. Often, outbreaks of both clades occur simultaneously, and testing isn't always done to determine what percentage of the outbreak belongs to any particular clade. Sensitivity analysis can be used to handle the situation in which the clade is uncertain, or if a new subclade appears.

The following publication includes a compartmental Mpox model with dynamics for both rodents and humans: "Unfolding the Transmission Dynamics of Monkeypox Virus: An Epidemiological Modelling Analysis", *Mathematics* 2023, *11*(5), 1121; https://doi.org/10.3390/math11051121

1. (*TA1 Model Ingestion and Unit Test*) From the paper, ingest the model described in set of equations (1), and replicate the following graph in Figure 1 (use 0.03045 as the value for β_{rh} rather than 0.3045 shown in the paper), where the y axis represents the total infected population (I_1+I_2).

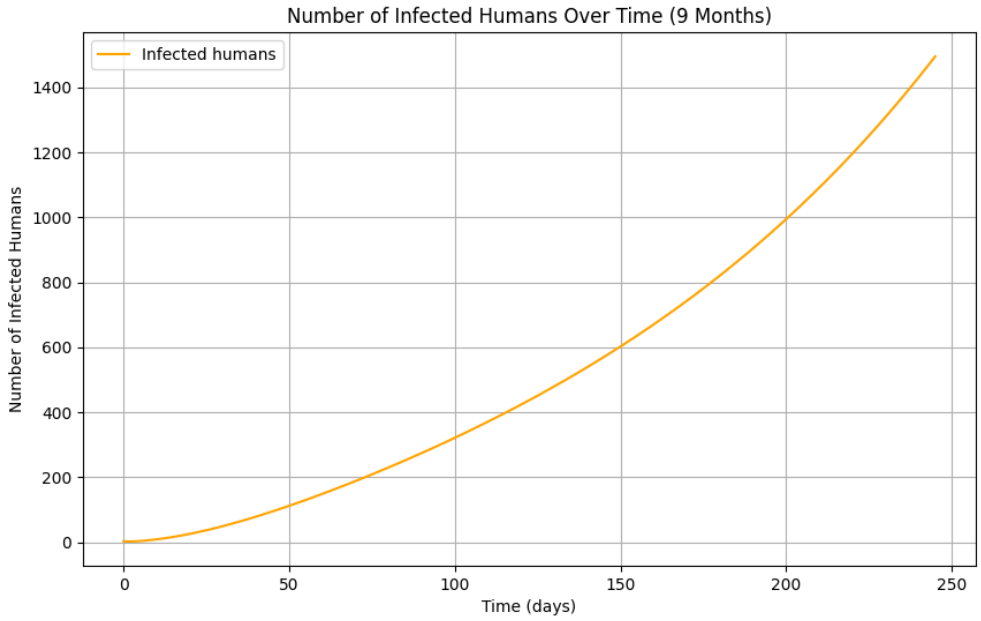


Figure 1. Unit test plot to replicate in Scenario 1, Q1

2. *(TA3 Model Configuration and Forecasting)* Configure the model with initial conditions defined in Section 4.1 and the parameter ranges described in the Table 2 of the paper, which reflect ranges of physically possible values. Perform a simulation of the range of plausible model outcomes for total infected population (I_1+I_2), and plot your results.

3. *(TA2 Model Editing and TA3 Updated Model Configuration and Simulation)* It is possible that pets could be infected by rodents, and then the pets could infect their owners. Create $S, E, I,$ and R compartments for pets, and allow pets to be infected by rodents, and for pets to infect humans. Pets can also infect one another, but at a lower rate than rodents infect one another. The transmission rate between pets and humans should be higher than between humans and rodents. Create a plot similar to Figure 1 showing the number of infected humans and infected pets.

4. *(TA2 Model Stratification)* The model separates high and low risk human populations but assumes homogeneity within the compartments. Stratify the human Susceptible, Infected (each of I_1 and I_2), and Exposed compartments in the model by two age groups (18-50 years, over 50 years) and vaccination status (vaccinated, unvaccinated). Data to inform the stratification can be found here:
   Age: https://www.cdc.gov/mmwr/volumes/72/wr/mm7233a3.htm
   Vaccine effectiveness: https://jynneos.com/

Scenario 2: Ensemble Forecasting

Please redo the Forecasting Challenge from the 12-month milestone for a single timepoint (this is the same timepoint that was part of the 18-month program evaluation as well). Data for the following time period should already be readily available and processed, from the 18-month evaluation.

- **Time**: July 19th, 2021, during the upswing of the Covid wave caused by the arrival of the Delta variant. Vaccines were available at this time.
- **Location**: New York state
- **Task**: Take a single model, calibrate it using historical data prior to the given date, and predict cases, hospitalizations, and deaths over a 4-week timeframe starting from the given date. Do the same thing with an ensemble of at least 4 structurally different models (e.g. one model could include hospitalizations, another could focus on wastewater, one could be stratified, etc.). Each of the 4 models should have at least one compartment not present in the others.

You can compare your own outputs against other forecasters who submitted their results in the Covid-19 Forecast Hub for this time period, using the following metrics:

- **Weighted interval score (WIS)**: "Can be interpreted as a generalization of the absolute error to probabilistic forecasts and allows for a decomposition into a measure of sharpness [spread] and penalties for over- and underprediction." A lower score indicates better performance.
- **Absolute error**: is the absolute value of the difference between the actual value and the point forecast. The point forecast of a model when not provided explicitly is taken to be the 50% quantile of the forecast distribution.

For example scripts MITRE will use to evaluate the forecasts, please refer to this Github link.

Scenario 3: Scenario Templates

For Questions 1-3, use the SEVIRHD model described in the Model Definition below. For Question 4, you will be using a different model (defined in the question).

0. Model Definition

The compartments for this SEVIRHD model are Susceptible, Exposed, Vaccinated, Infected, Hospitalized, Recovered, and Dead.

The base model is defined as follows:

$$\frac{dS}{dt} = -\beta * \frac{S * I}{N} - (\nu * S)$$

$$\frac{dE}{dt} = \beta * \frac{S * I}{N} + \epsilon * \frac{V * I}{N} - \sigma * E$$

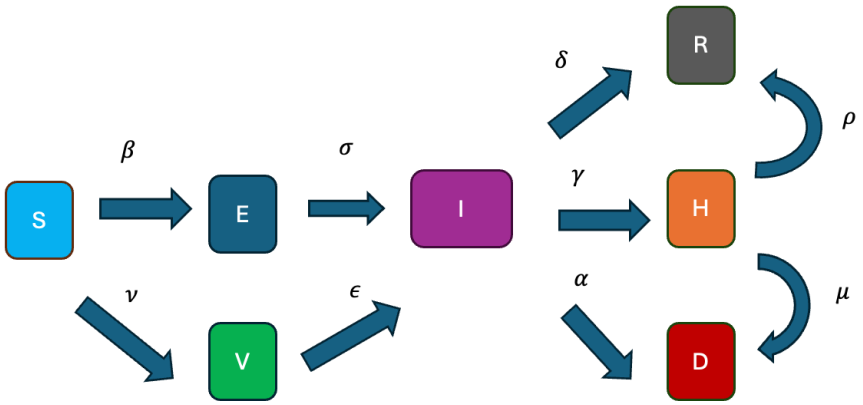
$$\frac{dV}{dt} = \nu * S - \epsilon * \frac{V * I}{N}$$

$$\frac{dI}{dt} = \sigma * E - (\gamma + \delta + \alpha) * I$$

$$\frac{dR}{dt} = \delta * I + \rho * H$$

$$\frac{dH}{dt} = \gamma * I - (\rho + \mu) * H$$

$$\frac{dD}{dt} = \alpha * I + \mu * H$$



Where:
$N = S + E + V + I + R + H = 100000$ people is the total population excluding deceased individuals,
$\beta = 0.60$ *new exposures per infected person/ day* is the transmission rate,
$\nu = 0.005$ / *day* is the vaccination rate,
$\sigma = 0.10$ / *day* is the rate exposed and unvaccinated individuals become infected,
$\epsilon = 0.05$ / *day* is the rate exposed and vaccinated individuals become infected,
$\delta = 0.05$ / *day* is the rate infected individuals recover,
$\gamma = 0.02$ / *day* is the rate infected individuals become hospitalized,
$\alpha = 0.01$ / *day* is the death rate for infected individuals who are not hospitalized,
$\rho = 0.10$ / *day* is the rate hospitalized individuals recover, and
$\mu = 0.02$ / *day* is the death rate for hospitalized individuals.

For initial conditions, assume $S(0) = 99900$ people, $I(0) = 100$ people, and all other states begin with 0 people.

1. **Sensitivity Analysis**: For this question we do a sensitivity analysis to understand how model outcomes are impacted by sources of uncertainty (which are not under the control of the decisionmaker) – see Figure 2 describing a sensitivity analysis example.

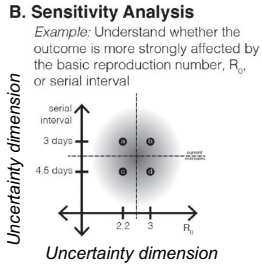


Figure 2. Sensitivity analysis example plotted against two uncertainty axes.

For this question, use the model defined in Q0 and assume there is uncertainty in the following parameters:

- $\gamma = 0.02$ /*day*, the rate at which infected individuals become hospitalized
- $\epsilon = 0.05/day$, the rate exposed and vaccinated individuals become infected

Vary each of these parameters from ½ of their given value in Q0, to 1.5 times their given value. In other words, let γ vary from $0.01/day$ to $0.03/day$ and let ϵ vary from $0.025/day$ to $0.075/day$.

Explore the sensitivity of the model output to these two parameters.

2. **Decisionmaker Question**: For this question we construct a decisionmaker question looking at multiple intervention options (which are under control of the decisionmaker) with no sources of uncertainty in the parameters. See Figure 3 describing a decisionmaker example.

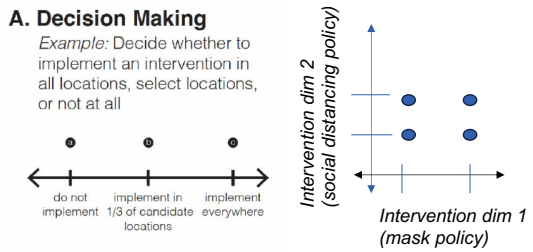


Figure 3. Decisionmaker examples in 1D and 2D, mapped on intervention axes.

Consider the following intervention options:

- Baseline (change nothing from Q0 model configuration)
- Option 1
  - *Limited mask intervention*: assume 15% decrease in transmission rates
  - *Limited vaccination campaign*: increase vaccination rate to 1% of Susceptible population being vaccinated daily
- Option 2:
  - *Stringent mask intervention*: assume 50% decrease in transmission rates
  - *Limited vaccination campaign*: increase vaccination rate to 1% of Susceptible population being vaccinated daily
- Option 3:
  - *Limited mask intervention*: assume 15% decrease in transmission rates
  - *Stronger vaccination campaign*: increase vaccination rate to 1.5% of Susceptible population being vaccinated daily
- Option 4:
  - *Stringent mask intervention*: assume 50% decrease in transmission rates
  - *Stronger vaccination campaign*: increase vaccination rate to 1.5% of Susceptible population being vaccinated daily

These options are summarized in the following parameter table:

Q2 Parameter Table

	Masking	
Vaccination	*Option 1* • Limited masking • Limited vaccination $\beta = 0.51\ new\ exposures\ per\ infected\ person/\ day$ $\nu = 0.01/day$	*Option 2* • More stringent masking • Limited vaccination $\beta = 0.30\ new\ exposures\ per\ infected\ person/\ day$ $\nu = 0.01/day$
	Option 3 • Limited masking • Higher vaccination rate $\beta = 0.51\ new\ exposures\ per\ infected\ person/\ day$ $\nu = 0.015/day$	*Option 4* • More stringent masking • Higher vaccination rate $\beta = 0.30\ new\ exposures\ per\ infected\ person/\ day$ $\nu = 0.015/day$

What are the expected model outputs (cumulative infections, cumulative hospitalizations, total deaths) after 200 days, for each of these 4 intervention options, and which option shows the greatest improvement over the baseline?

3. **Value of Information**: For this question, we build upon Q2, and explore how choice of intervention (which is under control of the decisionmaker) is impacted by sources of uncertainty (which are not under control of the decisionmaker)? See Figure 4 describing a value of information example.

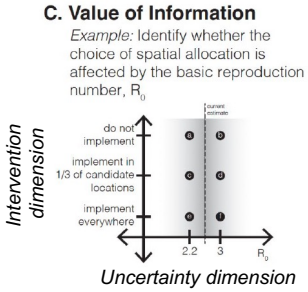


Figure 4. Value of information example mapped to uncertainty and intervention axes

In this question we keep the same decisionmaker intervention options from Q2, and we additionally consider uncertainty in the following two parameters:

- $\gamma = 0.02/day$, the rate at which infected individuals become hospitalized
- $\epsilon = 0.05/day$, the rate exposed and vaccinated individuals become infected

Vary each of these parameters from ½ of their given value in Q0, to 1.5 times their given value. In other words, let γ vary from $0.01/day$ to $0.03/day$ and let ϵ vary from $0.025/day$ to $0.075/day$.

With this parameter uncertainty, how do model outputs (cumulative infections, cumulative hospitalizations, total deaths) after 50 days, for each of the 4 decisionmaker intervention options, compare with Q2 (which did not have parameter uncertainty)? How is your choice of the best intervention option affected by the uncertainty in γ and ϵ?

4. **Horizon Scanning**: For this question, we consider a new model, and there are no dependencies on Q0-Q3. Assume there is an emerging pathogen with unknown transmissibility and unknown severity. We are interested in conducting a horizon scan to get a sense of how lethal the pathogen could be.

   Use the following SEID model:

$$\frac{dS}{dt} = -\beta * \frac{S * I}{N}$$

$$\frac{dE}{dt} = \beta * \frac{S * I}{N} - \sigma * E$$

$$\frac{dI}{dt} = \sigma * E - \mu * I$$

$$\frac{dD}{dt} = \mu * I$$

   Where:
   $N = S + E + I + D = 80000$ people is the total population,
   $\beta = 0.50$ *new exposures per infected person/ day* is the transmission rate,
   $\sigma = 0.15$ $/ day$ is the rate exposed individuals become infected, and
   $\mu = 0.075 / day$ is the rate infected individuals die.

   Use the following initial conditions:
   - $S(0) = 79900$ people
   - $E(0) = 0$ people
   - $I(0) = 100$ people
   - $D(0) = 0$ people

   Given the huge amount of uncertainty around the emerging pathogen, conduct a horizon scan by simulating model outcomes when the lower and upper bounds of transmission rate β are $0.2/day$ and $0.8/day$, and simultaneously, the lower and upper bounds of death rate μ are $0.05/day$ and $0.10/day$. Compare total deaths and total infections at the end of the simulation as the outcomes of interest.

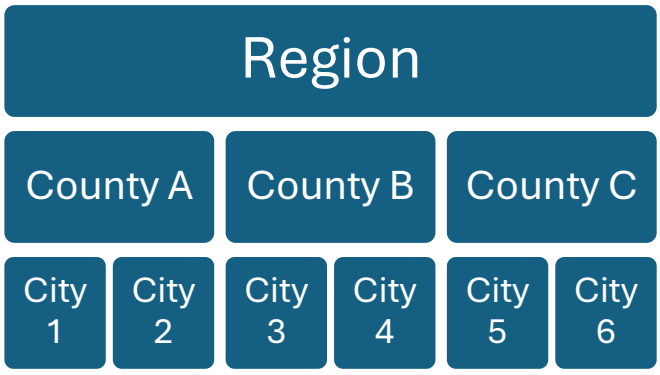
Q4 Parameter Table

	Transmissibility	
Death rate	*Outcome 1:* • Low transmissibility • Low death rate $\beta = 0.20$ *new exposures per infected person/ day* $\mu = 0.05/day$	*Outcome 2:* • High transmissibility • Low death rate $\beta = 0.80$ *new exposures per infected person/ day* $\mu = 0.05/day$
	Outcome 3: • Low transmissibility • High death rate $\beta = 0.20$ *new exposures per infected person/ day* $\mu = 0.10/day$	*Outcome 4:* • High transmissibility • High death rate $\beta = 0.80$ *new exposures per infected person/ day* $\mu = 0.10/day$

Scenario 4: Hierarchical Modeling

In this scenario, we have simulated data from a geographic region. This region is made up of three counties, A, B, and C, and each county consists of two cities. The cities are numbered 1-6. Cities in the same county have similar transmission rates (drawn from the same distribution), but transmission rates vary stochastically over time. Counties do not share similar transmission rates with other counties.

The simulated dataset has 100 days of data for each city. Each day contains 10 timesteps, so there are 1000 observations for each city.



Epidemiological data can be imperfect, and you will find that to be the case within these counties. Specifically:

- In County A, City 1, a large chunk of the data is missing due to a month-long issue with the city's data collection system.
- In County B, City 3, data was not collected regularly, and thus there are gaps missing in the data. About 30% of the timepoints have no data collected for them.
- In County C, City 6, the mayor of the city implemented an intervention 30% of the way through the period to reduce transmission rates even further. The same intervention was not applied in any other county or city.

The governor of the region is interested in answering a series of questions about the cities and counties. The data is recorded in an SEIR format, which follows these equations:

$$\frac{dS}{dt} = -\beta * S * \frac{I}{N}$$
$$\frac{dE}{dt} = \beta * S * \frac{I}{N} - \alpha * E$$
$$\frac{dI}{dt} = \alpha * E - \gamma * I$$
$$\frac{dR}{dt} = \gamma * I$$

Where:

S, E, I and R are the four states in the model: Susceptible, Exposed, Infectious, and Recovered (units: people),
$N = S + E + I + R$ is the total population (units: people),
β is the transmission rate (units: $new\ exposures\ per\ infected\ person/\ day$),
α is the rate of progression from exposed to infectious (units: $1/day$), and
γ is the rate of recovery for infectious individuals (units: $1/day$).

The initial conditions of the cities and the data for each city are provided in the *S4_data.csv* file. The initial conditions are also listed below.

Table 1. Initial conditions for hierarchical modeling scenario. All units are in terms of population (number of people).

	City 1	**City 2**	**City 3**	**City 4**	**City 5**	**City 6**
$\boldsymbol{N}$	1,000	1,200	500	1,500	2,000	10,000
$\boldsymbol{S(0)}$	999	1,199	499	1,499	1,950	9,900
$\boldsymbol{E(0)}$	1	1	1	1	50	100
$\boldsymbol{I(0)}$	0	0	0	0	0	0
$\boldsymbol{R(0)}$	0	0	0	0	0	0

1. Estimate city-level transmission rates.
    a. Use the data provided to estimate the transmission rate, β, individually for Cities 1-5 *without* incorporating information from any other cities.
    b. For County C, City 6, note that the intervention was implemented starting at $t =$ 30 days. Estimate the transmission rate before and after the intervention took place (without incorporating information from other cities). Also estimate an overall transmission rate for the whole time period, assuming that the intervention had no effect on transmission. Compare the three estimated transmission rates (before, after, overall) to assess the effectiveness of the intervention.

2. Using the data provided, estimate both county- and region- level transmission rates. Additionally, estimate transmission rates for each city, this time by incorporating information from other cities. Note any differences from the result in Q1.

3. For County A, City 1, impute the missing chunk of data using (1) the model that incorporated information from other cities and (2) the model which did not incorporate or include information from other cities. Plot both estimated SEIR curves and compare the differences between the two versions.

4. Repeat the same exercise in Q3, for County B, City 3.

5. (*Counterfactuals*) Imagine that the intervention applied in County C, City 6 had also been applied to County B, City 3 and City 4, at the same time. How would this have reduced the total number of infections in County B across the duration of time that the data was collected?

6. (*Optimization – choose a geography*) The governor of the region is aware of a new variant that has started to spread in a different region, and the governor fears that it may cause potential damage to the cities and counties they oversee. The best estimate of the transmission rate for this new variant is $1.2x$ as transmissible as previous variants.

   The governor has funding to apply an intervention in 2 of the 6 cities and is interested in minimizing the total number of infections in each city. The intervention is expected to reduce the transmission rate of the new variant to 80% of the most recent transmissibility rate before the intervention is applied. The intervention would start at t=100 timesteps (day 10) and can run for the rest of the period (up to day 100). For this question, initial conditions are the same as Table 1.

   In which cities should the governor implement the intervention? By implementing the intervention, what are the total number of infections the governor would expect over a 1000-timestep (100 day) period, and what are the total number of infections the governor would expect if they chose to do nothing? Provide 90% intervals on all projections.