Baseline and Evaluation Scenarios for ASKEM 6-Month Milestone

Epidemiology Use Case

Updated January 29th, 2023

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

 This technical data deliverable was developed using contract funds under Basic Contract No. W56KGU-18-D-0004.

Contents

Baseline and Evaluation Scenarios

Scenario 1: Exercises with Age Stratification

Scenario Ask: In order to consider more nuanced interventions, we would like for models to account for different age groups and their contact dynamics. Start with a basic SIR model without vital dynamics, and stratify it according to the following questions.

1. Start with a simple stratification with three age groups: young, middle-aged, and old
    a. Begin with a situation where the population size across each age group is uniform: N_young = 2k, N_middle = 2k, N_old = 2k. Assume only one person in each age group is infectious at the beginning of the simulation. Let gamma = 1/14 days, and let R0 = 5. Assume gamma, beta, and R0 are the same for all age groups.
        i. Simulate this model for the case where the 3x3 contact matrix is uniform (all values in matrix are 0.33)
        ii. Simulate this model for the case where there is significant in-group contact preference – you may choose the numbers in the matrix to represent this in-group preference.
        iii. Simulate this model for the case where there is no contact between age groups. You may choose the numbers in the matrix, but ensure it meets the requirement of no contact between age groups.
        iv. Simulate social distancing by scaling down the uniform contact matrix by a factor (e.g. multiply by 0.5)
        v. Repeat 1.a.iv for the scenario where the young population has poor compliance with social distancing policies, but the old population is very compliant.
    b. Repeat 1.a for a younger-skewing population: N_young = 3k, N_middle = 2k, N_old = 1k
    c. Repeat 1.a for an older-skewing population: N_young = 1k, N_middle = 2k, N_old = 3k
    d. Compare simulation outputs from 1a-c, and describe any takeaways/conclusions.
2. Now find real contact matrix data and stratify the basic SIR model with the appropriate number of age groups to match the data found. To simulate the model with realistic initial values, find data on population distribution by age group. As in question 1, let gamma = 1/14 days, and let R0 = 5. Assume gamma, beta, and R0 are the same for all age groups.
    a. If the data you've found supports this, compare the situation for a country with significant multi-generational contact beyond two generations (as indicated by multiple contact matrix diagonal bandings), and for a country without.
    b. If the data supports this, try implementing interventions like: (1) School closures (2) Social distancing at work and other locations, but not at home.

Question	Tasks	TA Workflow Tested	Metrics
0	Model Discovery: find an appropriate model that represents a basic SIR model as specified	TA1: Search and Discovery (for models)	**Time**: How long does it take to find the appropriate model?
1	Model Transformation	TA2: Model Transformation	**Time**: How long does stratification take? **Quality (qualitative):** Does stratified model make sense given the scenario?

1	Simulation tasks, according to Question 1	TA3: Simulation Workflows	**Time**: How long does it take to set up initial and parameter values, and do forward simulation? **Quality (qualitative):** Does output seem reasonable given the scenario?
2	Search for data: Real-world contact matrix data and population distribution data.	TA1: Search and Discovery (for data)	**Time**: How long does it take to find data? How long does it take to get data into a usable form for modeling?
2	Model Transformation: Stratify model according to Question 2 and real-world data found	TA2: Model transformation	**Time**: How long does stratification take? **Quality (qualitative):** Does stratified model make sense given the data found?
2	Simulation tasks, according to Question 2	TA3: Simulation Workflows	**Time**: How long does it take to set up initial and parameter values, do forward simulation with and without interventions? **Quality (qualitative):** Does output seem reasonable given the scenario?
	[*Optional*] If at any point, you need to search for parameter values, do a literature review, or find datasets, please track time spent, approach taken, and sources/databases you searched across.	TA1: Search and Discovery	**Time:** How long does search for required information take?

Scenario 2: Reproducing SIDARTHE and SIDARTHE-V

In 2020 the SIDARTHE model was published to describe the first wave of the Covid-19 pandemic in Italy. In 2021, this model was updated to include vaccination (SIDARTHE-V).

1. Start with the original SIDARTHE model.
    a. First, you want to make sure you have a good understanding of the original model, can execute it, and reproduce the results found in the publication describing the model. The paper doesn't include code, but there is an SBML version of the model. Regardless of the starting point, 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. The paper DOI is: 10.1038/s41591-020-0883-7, pdf: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7175834/pdf/41591_2020_Article_883.pdf. The BioModels repository (where the SBML model can be found) is here: [https://www.ebi.ac.uk/biomodels/BIOMD0000000955.](https://www.ebi.ac.uk/biomodels/BIOMD0000000955) Please complete extraction and pass unit tests under the following conditions, to test the process when the source material is of various levels of quality or completeness. Please note that one of the code versions is an accurate representation of the model, and is organized and commented. The other code version has intentional mistakes in the model definition, and is not as well organized. Please consider the following conditions:
        i. *[Challenge]* Ingest model and pass unit tests from publication alone (do not start with any code as input)
        ii. Ingest model and pass unit tests from publication and corresponding Code Version A
        iii. Ingest model and pass unit tests from publication and corresponding Code Version B
    b. There are two 'unit tests' we want to pass, to ensure that we understood and can reproduce the published model:
        i. Unit Test #1: Set the initial values and parameters, as described in the Supplementary Methods section of the publication (pg. 9 of the pdf):
            1. Initial Values: I = 200/60e6, D = 20/60e6, A = 1/60e6, R = 2/60e6, T = 0, H = 0, E = 0; S = 1 – I – D – A – R – T – H – E. Let total population = 60e6.
            2. Parameters: α = 0.570, β = δ = 0.011, γ = 0.456, ε = 0.171, θ = 0.371, ζ = η = 0.125, μ = 0.017, ν = 0.027, τ = 0.01, λ = ρ = 0.034 and κ = ξ = σ = 0.017.
            3. Simulate for 100 days, and determine the day and level of peak total infections (sum over all the infected states I, D, A, R, T). *Expected output*: The peak should occur around day 47, when ~60% of the population is infected.
        ii. Unit Test #2: Now update the parameters to reflect various interventions that Italy implemented during the first wave, as described in detail on pg. 9. Simulate for 100 days, reproduce the trajectories in Fig. 2B, and determine the day and level of peak total infections (sum over all the infected states I, D, A, R, T). *Expected output*: Trajectories in Fig. 2B, peak occurs around day 50, with ~0.2% of the total population infected.
    c. The difference between 1.b.i and 1.b.ii are changes in some parameter values over time. Describe the difference in outcomes between b.i and b.ii. Perform a sensitivity analysis

to understand the sensitivity of the model to parameter variations and determine which parameter(s) were most responsible for the change in outcomes.

d. Now return to the situation in b.i (constant parameters that don't change over time). Let's say we want to increase testing, diagnostics, and contact tracing efforts (implemented by increasing the detection parameters ε and θ). Assume that θ >= 2* ε, because a symptomatic person is more likely to be tested. What minimum constant values do these parameters need to be over the course of a 100-day simulation, to ensure that the total infected population (sum over all the infected states I, D, A, R, T) never rises above 1/3 of the total population?

2. Next, we want to explore the updated model SIDARTHE-V, which is found at https://doi.org/10.1038/s41591-021-01334-5, pdf: https://www.nature.com/articles/s41591-021-01334-5

   a. Do a structural model comparison of the original SIDARTHE and SIDARTHE-V. The structural comparison work product should include a summary or diagram describing similarities and differences between the models, with respect to parameters, variables/states, pathways, etc.

   b. Set the same initial values and parameter settings in 1.b.i. Let V(t=0) = 0, τ (in SIDARTHE) = $\tau 2$ (in SIDDARTHE-V), and $\tau 1$ = (1/3)*$\tau 2$ (reflecting the fact that the mortality rate for critical conditions (state T), will always be larger than for other infected states). Assume that the vaccination rate *psi* is 0 to start with. The SIDARTHE-V model allows for three main types of interventions: (1) Those that impact the transmission parameters (α, β, γ and δ) – social distancing, masking, lockdown; (2) Those that impact the detection parameters (ε, θ) – testing and contact tracing; (3) Those that impact the vaccination rate psi – vaccination campaigns. Assume previously stated constraints: θ >= 2* ε, and $\tau 1$ = (1/3)*$\tau 2$.

      i. Let's say our goal is to ensure that the total infected population (sum over all the infected states I, D, A, R, T) never rises above 1/3 of the total population, over the course of the next 100 days. If you could choose only a single intervention (affecting only one parameter), which intervention would let us meet our goal, with minimal change to the intervention parameter? Assume that the intervention will be implemented after one month (t = day 30), and will stay constant after that, over the remaining time period (i.e. the following 70 days). What are equivalent interventions of the other two intervention types, that would have the same impact on total infections?

      ii. **[Changed 2/1]** Let's say our goal is to get the reproduction number Rt below 1.0, within the next 60 days. Which interventions will allow us to meet our goal, while minimizing total cumulative infections (over all infected states I, D, A, R, T)? If there are multiple options, show the tradeoff between change in parameter and infected populations – show the space of possible solutions. Which single intervention would have the greatest impact on Rt and let us meet our goal with minimal change to the intervention parameter, while minimizing total cumulative infections? Assume that the intervention will be implemented immediately. Use Rt as defined in the SIDDARTHE-V publication. No intervention and increasing the infected population, are not valid solutions for this problem.

Question	Task	Equivalent TA Workflow	Metrics
1a,b	• Model Extraction • Unit Testing	TA1: Model Extraction; TA1: Model Execution/Unit Testing	**Time**: 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 confident output is correct. **Accuracy**: • Were you able to faithfully reproduce results of unit tests? • Qualitative score on metadata quality (correctness, relevance, completeness), based on human inspection of the equations, variables, parameters, etc. • (TA1 only) Qualitative score on correctness of groundings/alignment
1c,d	• Simulations • Produce answers to scenario questions	TA3: Simulation Workflows (incl. sensitivity analysis, interventions); TA3: Answers to Scenario Questions	**Time:** How long does it take to set up and execute simulations and come up with answers to each part? **Quality (qualitative):** Does the answer address the scenario question adequately, and does it seem reasonable?
2a	Model Comparison	TA2: Model Comparison	**Time**: to execute model comparison **Quality (qualitative)**: Is model comparison output interpretable and does it capture major differences and similarities correctly?
2b	• Simulation • Provide answers to scenario questions	TA3: Simulation workflows (incl. sensitivity analysis, intervention optimization); TA3: Answers to Scenario Questions	**Quality (qualitative):** Does the answer address the scenario question adequately, and does it seem reasonable?

Scenario 3: Progressively Updating Model

In this scenario, we will be starting with a simple compartmental model, calibrating parameters and comparing with historical data, and progressively adding complexity to the model, to see how the fit improves. For all data, we will be using US national-level data. For calibration of parameters, you have flexibility to decide which parameters you would like to set using values found in the literature, and which will be estimated using fitting algorithms with real data.

Time range of data: June 1st 2021 – June 1st, 2022

- For questions 1-4, the 'training period' over which calibration will be done with data, is June 1, 2021 – September 30, 2021 (covering the predominant period of the Delta variant in the United States). The out-of-sample 'test period' over which fitted models can be used to 'forecast' and compare against historical data, is October 1, 2021 – January 1, 2022 (covering the period leading up to the Omicron wave).
- For questions 5-6, we want to consider multiple Covid waves. Let the 'training period' over which calibration will be done with data, be June 1st, 2021 – December 31st , 2021 (covering the Covid-19 Delta wave and part of the Omicron wave). The out-of-sample 'test period' over which fitted models can be used to 'forecast' and compare against historical data, is January 1st, 2022 – June 1st , 2022

1. Begin with a basic SIR model without vital dynamics. Calibrate the model parameters using data on cases during the 'training period'. Then evaluate the model during the out-of-sample 'test period'.
2. One issue with using case data as the reference against which models should be fit, is that case data tends to be noisy, and also undercounts actual infection numbers. Not everybody who was infected got tested or had access to tests during this time period. (Side note: in 2022, the issue with using case data is different, as tests are much more widely available, but home tests are usually not reported to any central authoritative agency that aggregates and releases the 'official' case numbers). Usually data on deaths or hospitalizations is more accurate and dependable. We would like to update our model to include deaths/and or hospitalizations, in order to incorporate data on those outcomes. Explore the space of closely related models (structurally speaking) that incorporate either deaths, hospitalizations, or both. For each model, calibrate parameters using data on hospitalizations or deaths, evaluate performance in the 'test period' (compare model output against data), and do model selection based on how well the fitted model output compares with data, for both the 'training' and 'test' periods. Do not consider vaccination or age stratification (these will be considerations in the following tasks).
3. Now update the model to include vaccination, and calibrate and comparison of model output, with data on deaths and/or hospitalizations broken down by vaccination status.
4. **[Challenge]** Add age stratification to the model. Repeat calibration and comparison of model output against data, using data on deaths and/or hospitalizations broken down by vaccination status and age group. The number of strata will depend on the age breakdowns in available data.
5. Early in the pandemic, there were naïve modeling choices made about the unlikeliness of reinfection. Now that we know reinfection is a reality, we want to update our model to incorporate this. Choose any of the models you worked with in #1-4, and add in mechanisms to represent reinfection. Repeat calibration and comparison of model output against data. Remember to update the data range as indicated above in the scenario definition.

6. **[Challenge]** For this question, define the 'training period' as June 1st, 2021 – December 31st, 2021. Define the out-of-sample 'test period' as January 1st, 2022 – June 1st, 2022. Using the models you developed in questions #1-5, can you create a weighted ensemble that outperforms all of the component models, for the 'test period'?
7. For each of #1-6, summarize your conclusions about the following:
    a. Do parameters fit from data seem reasonable and fall within typical ranges you might see in the broader literature? Provide references to support your conclusions.
    b. Describe how well the fitted model compares against historical data, both for the 'training' and 'test' periods.
8. [*Optional*] For any of #1-6, if guidance is needed on ways to update the models, do a literature search and incorporate aspects of published models.

Question	**Task**	**Equivalent TA Workflow**	**Metrics**
1-5	Model extension/transformation	TA2: Model Extension/ Transformation (incl. inserting compartments, stratification); TA2: Model Space Exploration	**Time**: How long does extension or transformation task take? **Quality (qualitative):** Does transformed model give plausible outputs?
1-6	• Model calibration with data • Forecasting with calibrated models and comparing against data	TA3: Simulation Workflows (incl. calibration, forecasting)	**Time:** How long does it take to set up and execute simulation workflows? **Quality (qualitative):** Does output of calibrated models seem reasonable?
6	Create model ensemble	TA3: Simulation Workflows (creating ensembles)	**Time:** How long does it take to set up and execute simulation workflow? **Quality (qualitative):** Does output of calibrated ensemble model outperform component models?
1-6	Search for relevant data	TA1: Search and Discovery (for data)	**Time:** How long does search for relevant data take? How long does it take to get data into usable form in the system. **Quality (qualitative):** How relevant are the results found, to the scenario context?
7a	Search for relevant parameter values in literature	TA1: Search and Discovery (for parameters)	**Time:** How long does search for relevant information take? **Quality (qualitative):** How relevant are the results found, to the scenario context?
7b	Describe how well model output compares against historical data	TA3/4: Visualizations (of outputs, comparisons of historical data to forecast outputs)	**Quality (qualitative):** Do assessments of how well models performed against historical data, seem reasonable given the structure and level of complexity of the models being considered?
8	[*Optional*] Literature review to see how other published models handle certain aspects	TA1: Search and Discovery	**Time:** How long does search for required information take?