The TAO of University Pandemic Forensics: How Good are COVID Simulations?

Online Supplementary Material

This online supplement employs the method described in the main paper to reconstruct the history of infectivity rates for:

- Northeastern University in Spring 2021,
- Purdue University in Fall 2020, and
- Case Western Reserve University in Fall 2020.

It also offers a detailed description of all calibration runs referred to in the main paper and in the online supplementary materials. To avoid confusion, Tables and Figures are numbered in sequence with tables figures in the main paper, beginning with Table 2 and Figure 12.

Northeastern University Spring 2021 Semester

We set initial model parameters to be the same as the calibrated solution for NU Fall 2020, using NU24, since there were no major policy changes between the Fall and Spring semesters. We varied only the number of active agents (20,000 in the Fall) to be 20,000, 25,000, or 30,000 out of 30,000 total agents (Figure 12).

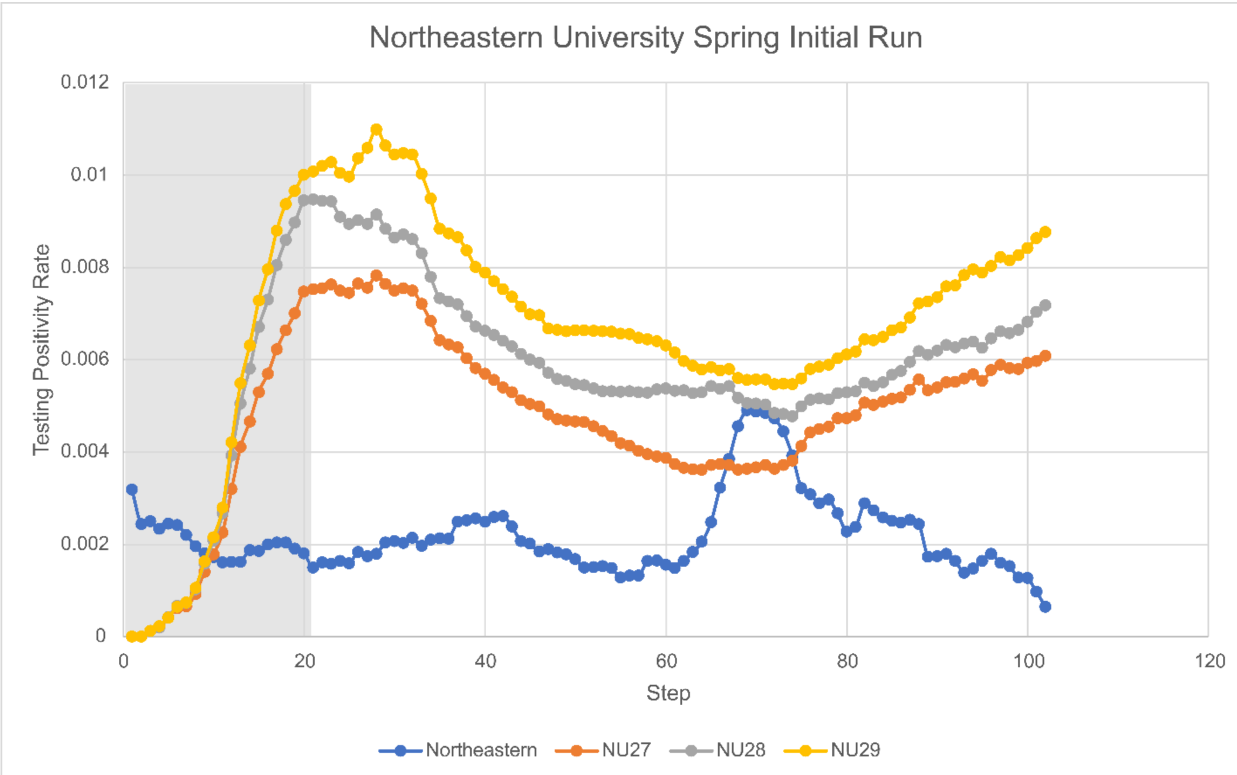

Figure 1. Calibration runs for Northeastern University, Spring 2021: NU27-NU29.

We expected discrepancies in three major areas: (1) the effect of vaccines (not present in the initial runs depicted in Figure 12), (2) differences in compliance in students in Fall vs Spring, and (3) kickstart error due to treating Fall and Spring as two entirely different systems, when in fact they are they are parts of a single system, and causally related, particularly given the continual presence of staff during the break between semesters when students leave campus. Dr. Yune told us that compliance in students was lower in the Spring compared to Fall. Therefore, we expected TAO output to have test positivity rates below the actual test positivity rates. However, our model output over-estimated the test positivity rates. We concluded that vaccination was probably an important consideration.

We added 450 daily vaccinations to the model starting on day 51. Vaccinated agents had a likelihood of immediate recovery when exposed. This implementation deliberately exaggerates the effect of vaccines with immediate impact and a high level of dissemination. Introducing vaccination did impact model runs but the impact did not eliminate the overestimated positivity rate (Figure 13). TAO was telling us that, under the same circumstances as Fall semester, NU should have had a much higher positivity rate in the Spring, but this was not what actually happened at NU.

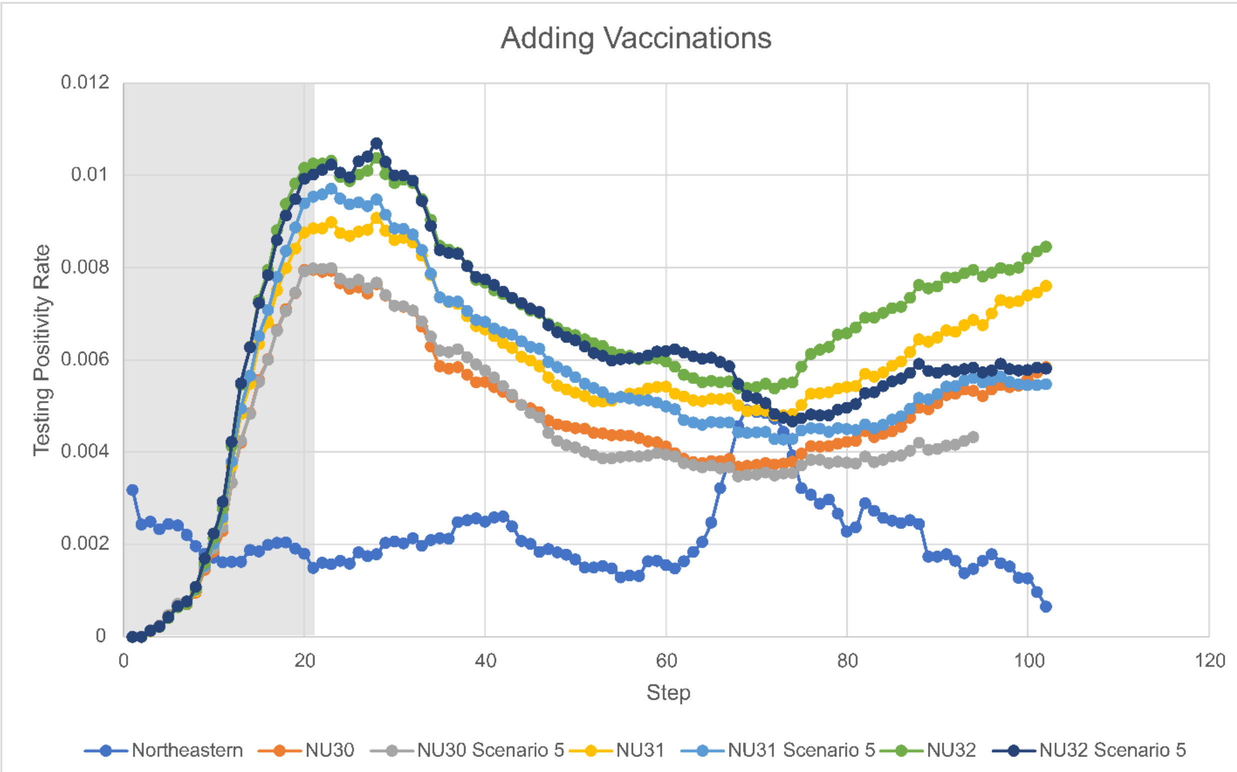

Figure 2. Calibration runs for Northeastern University, Spring 2021: NU30-NU32, with and without Scenario 5.

To penetrate this mystery, we spoke again with NU's Dr. Yune about the relative number of active on-campus people and the relative compliance compared to the Fall semester. Dr. Yune reiterated that compliance decreased from the Fall semester to the Spring semester and stated that there was no way to be sure about the number of on-campus people because that data was not systematically tracked. She did say that, because people's testing frequency was supposed to be determined by frequency of on-campus visits, the number of daily tests could indicate if fewer or more people were active on campus. Furthermore, she speculated that many students may have opted to attend classes primarily virtually, even while remaining in the Boston area.

Additionally, Dr. Yune explained that with entry testing, all people coming to campus should have been tested in the first week of the semester. Regarding vaccines, Dr. Yune stated that for most of the Spring semester, vaccines were not available to anyone except the most vulnerable, so she did not expect vaccines to be a strong factor. Lastly, we asked Dr. Yune for an estimate of the number of people who had contracted the virus prior to the Spring semester, so that we could tune our number of agents initially assigned to the "recovered" category.

We set the model parameters based on this new information from Dr. Yune. We set the number of initially recovered agents (agents who had contracted the virus already prior to the start of the Spring semester) between 5-10%. We reduced the compliance distribution from 65-95% to either 60-95%, 60-90%, or 50-90%. We pushed vaccinations to day 82. We ensured that all active agents were tested at least once in the first week. We set the number of active agents to a constant 20,000. Figure 14 displays the results.

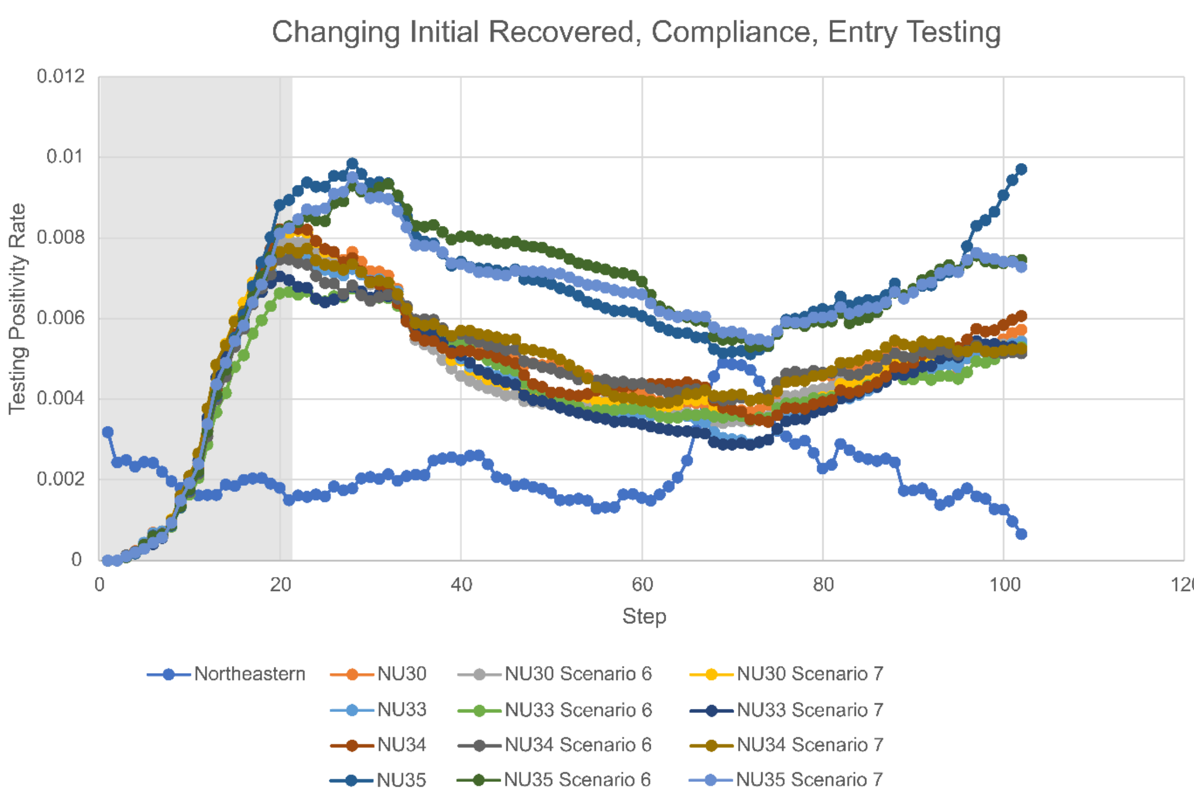

Figure 3. Calibration runs for Northeastern University, Spring 2021: NU30 and NU33-NU35 with Scenarios 6-7.

In these model runs, the test positivity rate was still overestimated for most of the simulation. We concluded that for our simulation to have a smaller discrepancy, we would have to reduce the number of active agents; we varied the number of active agents at 10,000, 15,000, and 20,000 out of 30,000. We included a variation where all active agents were tested on day 1, with the intention of determining what the maximal effect of entry testing could be. We also ran a variation decreasing the external infection rate by 75%. Figure 15 displays the results.

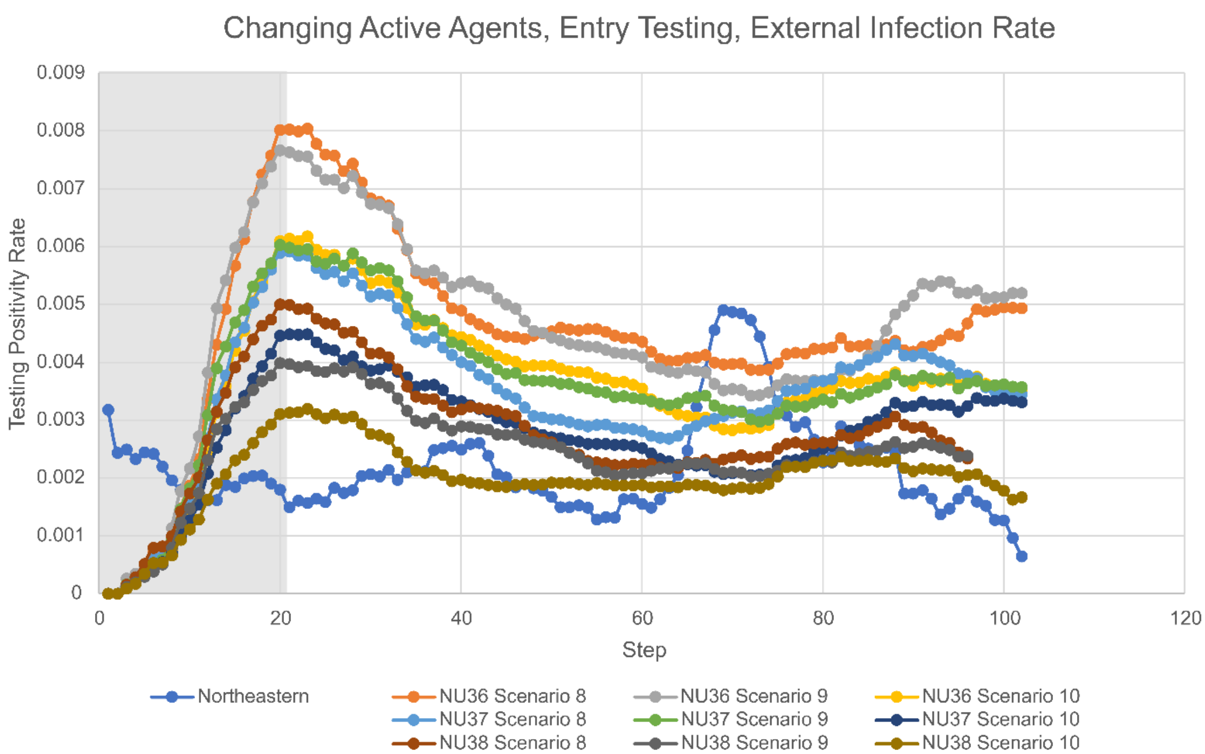

Figure 4. Calibration runs for Northeastern University, Spring 2021: NU36-NU39 with Scenarios 8-10.

We found that reducing the external infection rate and setting the number of active agents to 10,000 brought the positivity rate much closer to the actual positivity rate. However, there is a spike in the real positivity rate around day 70, and none of our runs were producing such spikes. We hypothesized that campus activity levels were relatively low during the cold Boston winter but may have increased with warmer temperatures. So, we reran the model with increasing the number of active agents from 10,000 to either 15,000 or 20,000, starting at day 60-62. We also tried increasing the external infection rate on those days. We chose these dates because from Day 62 and beyond, the temperature was consistently 60 degrees Fahrenheit or above. (https://www.accuweather.com/en/us/boston/02108/march-weather/348735?year=2021). Table 2 presents the parameter scenarios for the NU39 runs and the results are in Figure 16.

Table 1. Eight scenarios for parameter combination NU39.

Scenario 11	10,000 agents increasing to 15,000 agents on day 60
Scenario 12	Same as Scenario 11 with an increase in external infection rate
Scenario 13	10,000 agents increasing to 20,000 agents on day 60
Scenario 14	Same as Scenario 13 with an increase in external infection rate
Scenario 15	10,000 agents increasing to 15,000 agents on day 62
Scenario 16	Same as Scenario 15 with an increase in external infection rate
Scenario 17	10,000 agents increasing to 20,000 agents on day 62
Scenario 18	Same as Scenario 17 with an increase in external infection rate

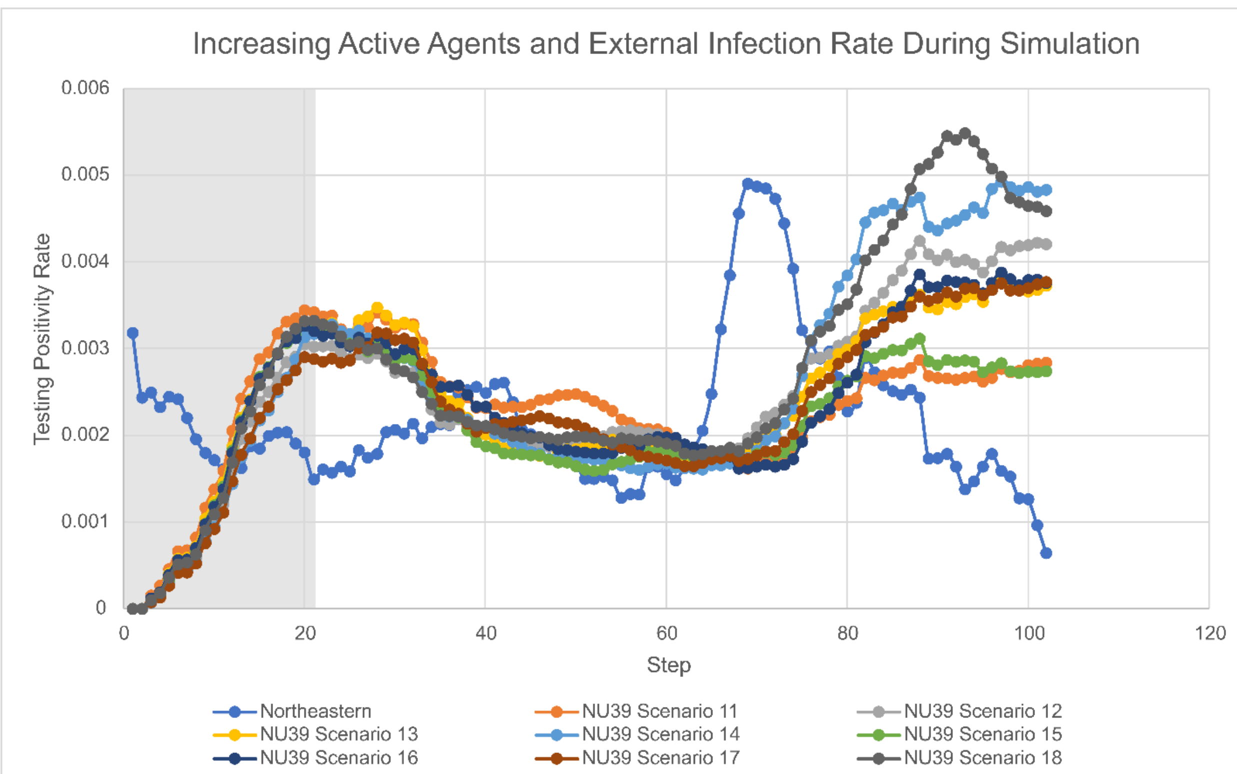

Figure 5. Final calibration runs for Northeastern University, Spring 2021: NU39 with Scenarios 11-18.

Running these inputs, we did see an increase in positivity rate, but the shape of the increase was not the same as the spike observed in the real-world data. At this point, we decided to terminate the validation process; although we did not produce a calibrated output that was as close as our other calibrated outputs, we felt that we could still reach some forensic conclusions.

Forensic Conclusions from NU for Spring 2021

Our model supports the conclusions that were was less on-campus activity and less extra-system contact in the NU Spring 2021 semester compared to the NU Fall 2020 semester. After reducing compliance, increasing the number of initially recovered agents, and implementing exaggerated entry testing, the only way for our model to closely match NU Spring 2021 positivity rates was reducing the starting number of active agents to 10,000 (compared to 20,000 in the Fall semester). We hypothesized that the spike in Spring semester around day 70 may be due to weather changes in the Boston area, such as increased temperature, encouraging more on-campus activity, but the model does not fully support that hypothesis; increasing the number of active agents after the temperature was consistently above 60 degrees Fahrenheit does cause a spike in the simulation positivity rate, but the spike does not closely resemble the spike in the real-world data. Another hypothesis is that the spike was caused by some significant outbreak events, or that some agents left the system for vacation and returned infected; we did not test these hypotheses in the model. Note that NU did not have an official Spring Break in the Spring 2021 semester, but students could have travelled while attending classes virtually.

Purdue University Fall 2020 Semester

We initialized the PU model with 51,000 total agents, based on online statistics and estimating who had in-person jobs (see https://www.purdue.edu/datadigest/). We employed TAO's default generated university contact network and agent schedule. We set a compliance distribution of 60-

95% for all agents. We set the external infection rate multiplier to 2.5% or 5% of regional case numbers for Tippecanoe County, IN. Regarding NPIs, it appeared that PU primary NPIs were physical distancing and mask-wearing. People in the university were active except in exceptional circumstances, so we set 97.5% of agents to be active. Given that the positivity rate for PU was around 3% in the first couple weeks, we set the percentage of initially infected agents to 0.72% (about 360 agents). PU's COVID-19 testing dashboard included weekly number of tests administered. We converted this to an average daily number of tests for each given week, and provided those daily number of tests as input to the model.

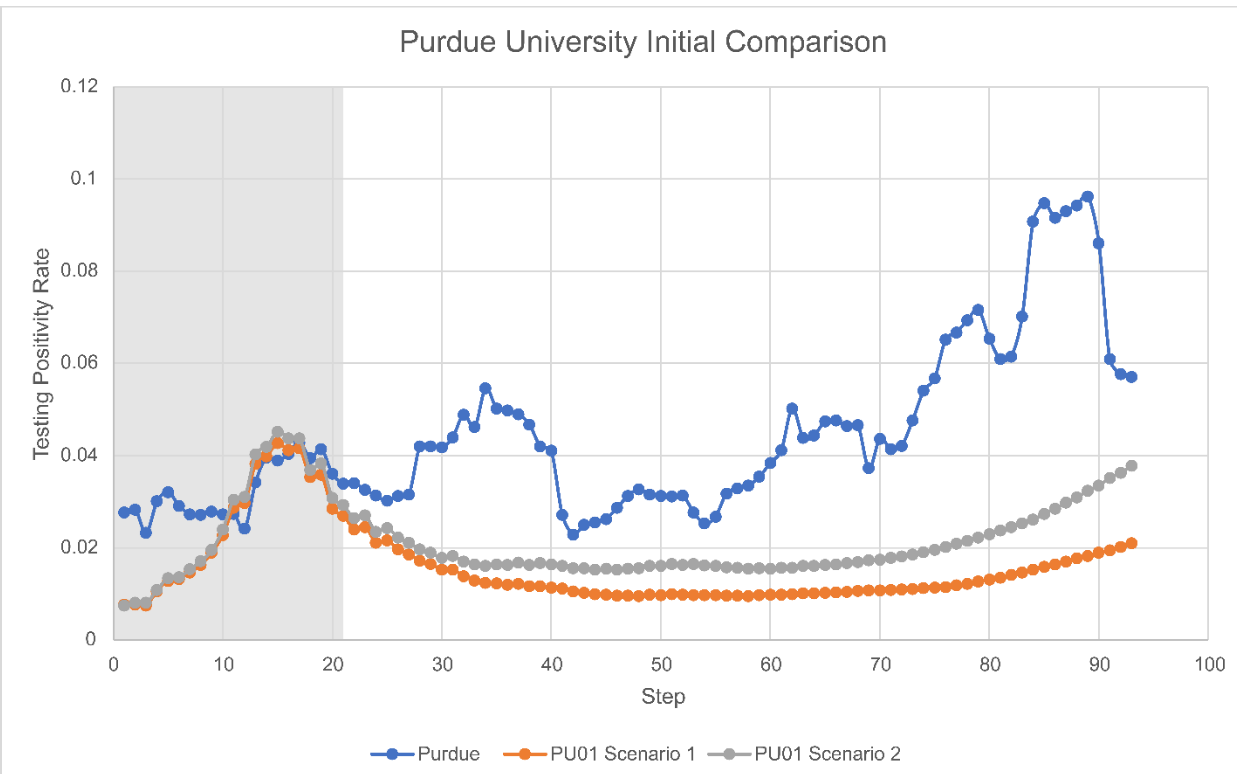

Figure 6. Calibration runs for Purdue University, Fall 2020: PU01 with Scenarios 1-2.

Figure 17 shows that these assumptions resulted in under-matching the real-world data. Because of the erratic nature of the PU observed positivity rates curve, we wanted to stimulate the model into having more outbreak events. We did this by increasing the likelihood of agents throwing parties from 1% to 5% and increasing the likelihood of an agent attending a party from

5% to 20%. We also decreased the compliance distribution from 60-95% to 60-85%. The results are in Figure 18.

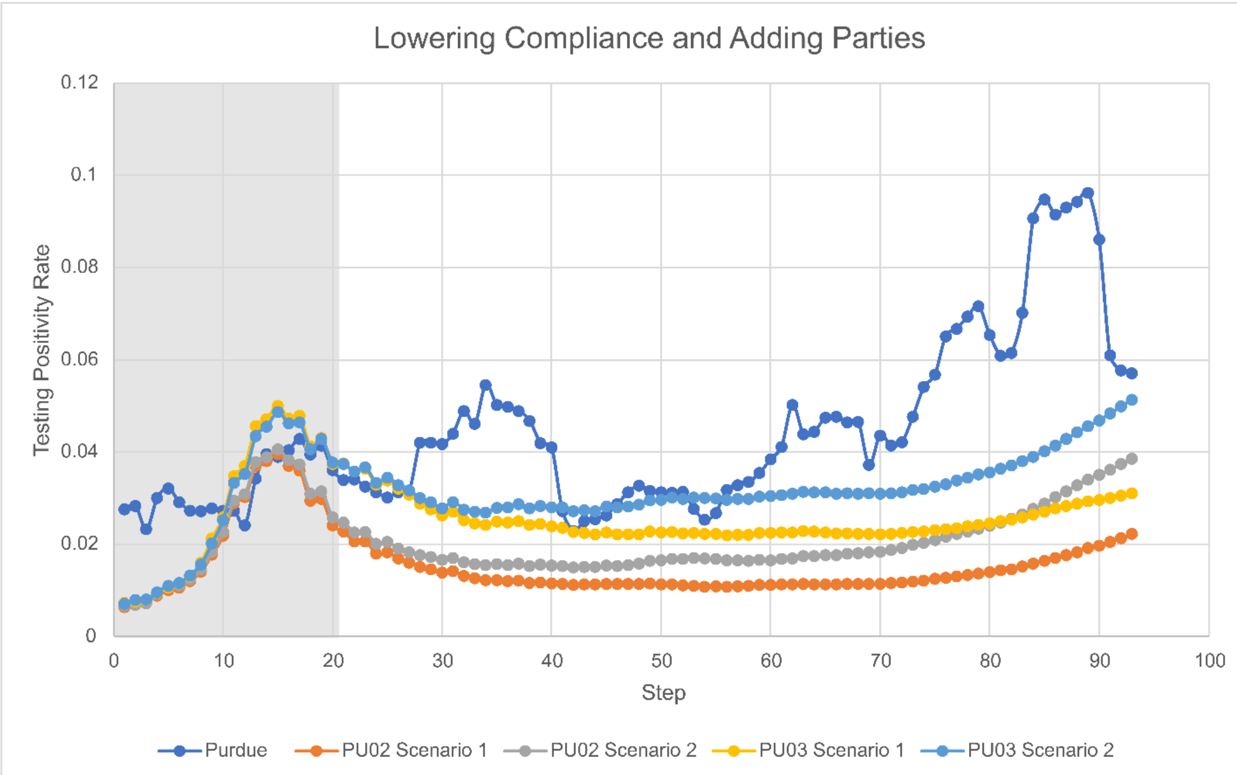

Figure 7. Calibration runs for Purdue University, Fall 2020: PU02-PU03 with Scenarios 1-2.

While lowering compliance and increasing parties brought the model output closer to PU positivity rate, there was still an erratic effect in the PU data that we were not capturing with the model. At this point in our parallel calibration with NU Fall 2020, we were trying different base infectivity rates between 2.5% and 5%, so we varied the base infection rate to be either 3% or 4%. To try and match the more erratic positivity rate, we increased the party parameters in some variations such that the likelihood of an agent throwing a party was 0-25% and the likelihood of attending a party was 40-90%. Figure 19 presents the results.

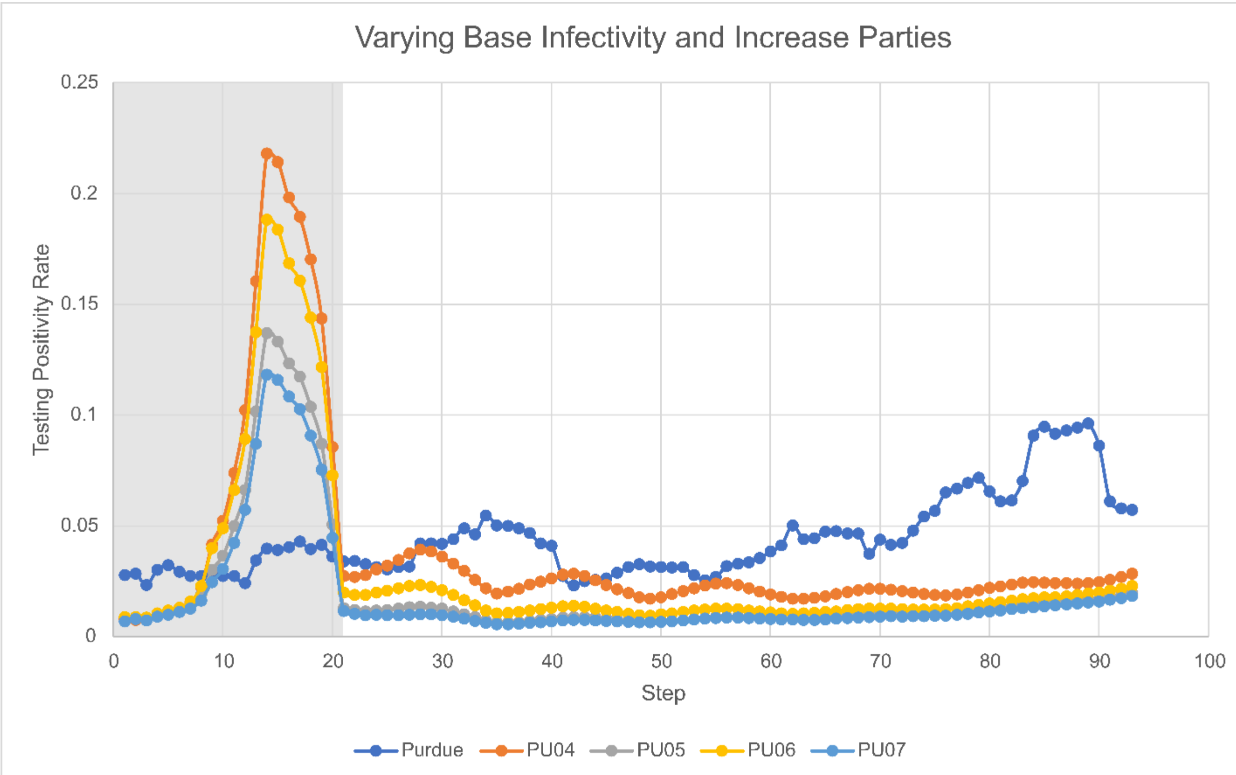

Figure 8. Calibration runs for Purdue University, Fall 2020: PU04-PU07.

In these runs, we did see a more erratic test positivity output with noticeable ups and downs. The large spike in the first twenty days of the simulation test positivity rate is kickstart error due to contact tracing working unrealistically well in the beginning of the simulation. However, at the time, we had not yet introduced the contact tracing interview recall, and we thought we had too many initially infected agents, so we lowered the number of initially infected agents to about 180. We also noticed that the model was particularly inaccurate after day 60, failing to account for some transmissions. So, we injected 500 random infections on day 70. The results are in Figure 20.

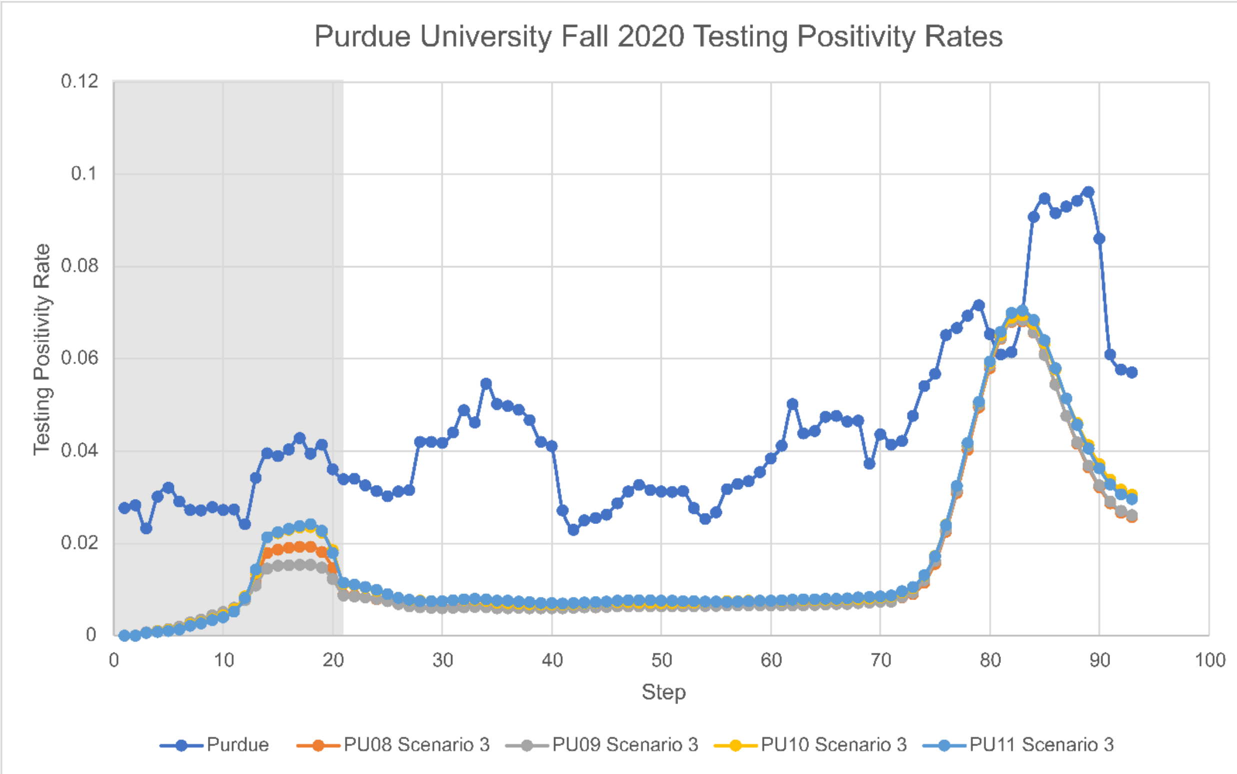

Figure 9. Calibration runs for Purdue University, Fall 2020: PU08-PU11 with Scenario 3.

Lowering the number of initial infections decreased the test positivity rate dramatically, and injecting infections increased the positivity rate. But the simulation output was still too low. At this point, we had introduced the contact tracing interview recall parameter. So, next, we tried new parameters for recall rates (40-50%), number initially infected (180-360), and compliance (60-95% or 20-60%). Figure 21 displays the results.

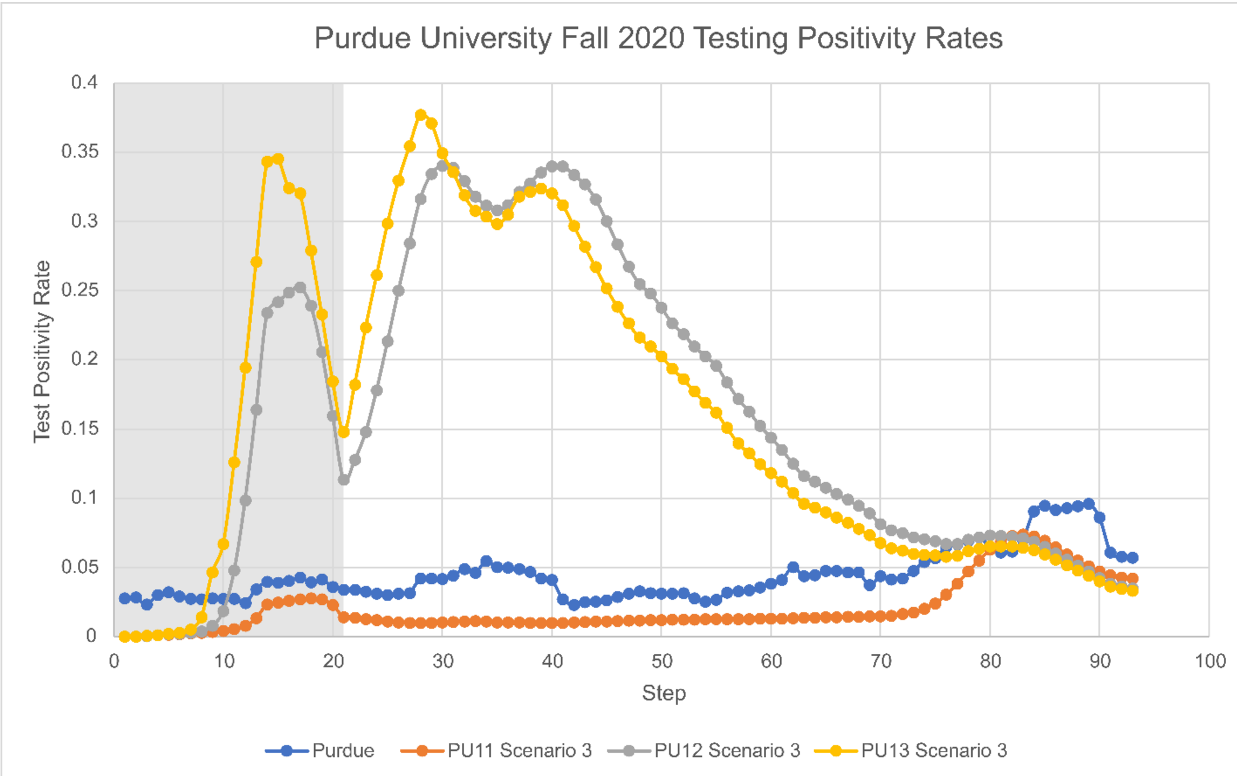

Figure 10. Calibration runs for Purdue University, Fall 2020: PU11-PU13 with Scenario 3.

Outputs were strongly over or strongly under PU's testing positivity rates. So we tried more moderate compliance distributions of 40-60% and 40-80%, while still varying interview recall between 40% and 50%. We still inject 500 infections on day 70. Figure 22 displays the results.

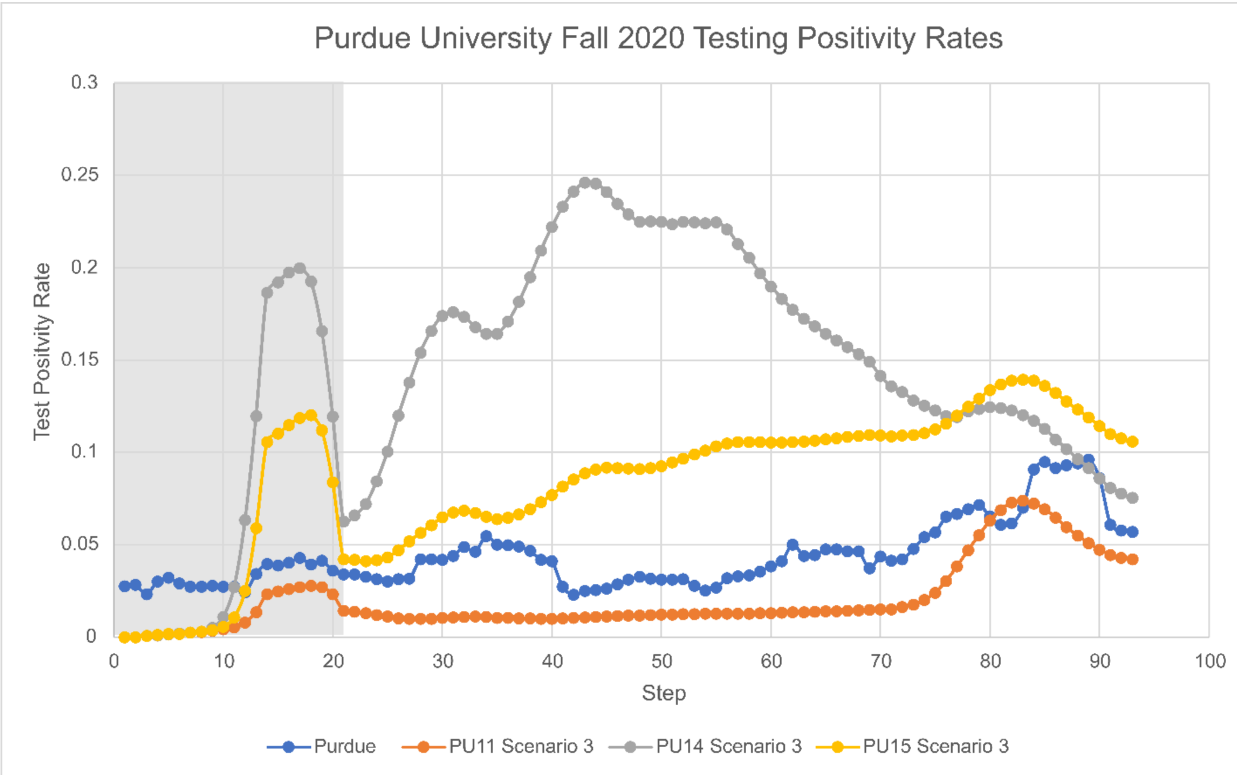

Figure 11. Calibration runs for Purdue University, Fall 2020: PU11 and PU14-15 with Scenario3.

With a closer positivity rate match, we could more easily fine-tune the model. Since kickstart error was still very significant, we reduced agent interview recall to 30% and set the number of initially infected agents to 180. We varied the compliance between 40-80% and 50-80%. Figure 23 displays the results.

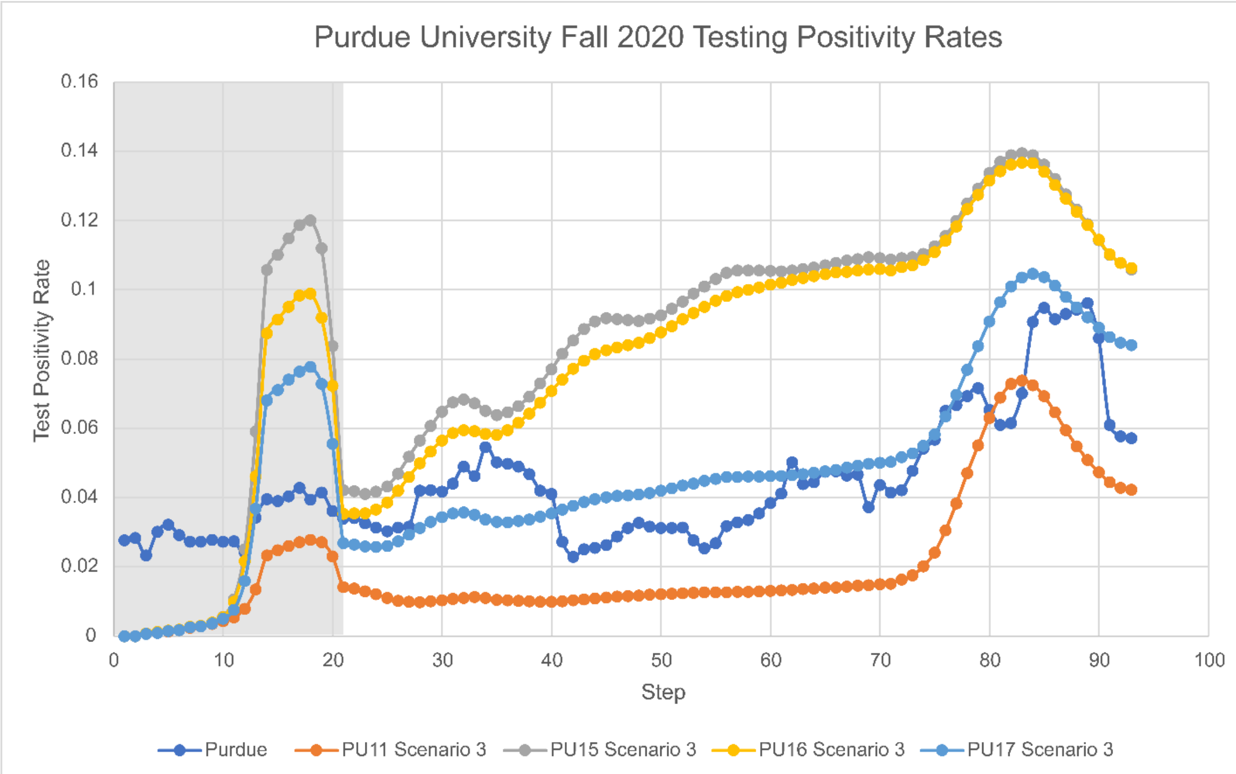

Figure 12. Calibration runs for Purdue University, Fall 2020: PU11 and PU-15-17 with Scenario 3.

This iteration resulted in our final calibrated parameter set, with the following notable characteristics: (1) a super-spreader event causing 500 infections at day 70, (2) low contact tracing interview recall, and (3) medium compliance (50-80%). The result is in Figure 24.

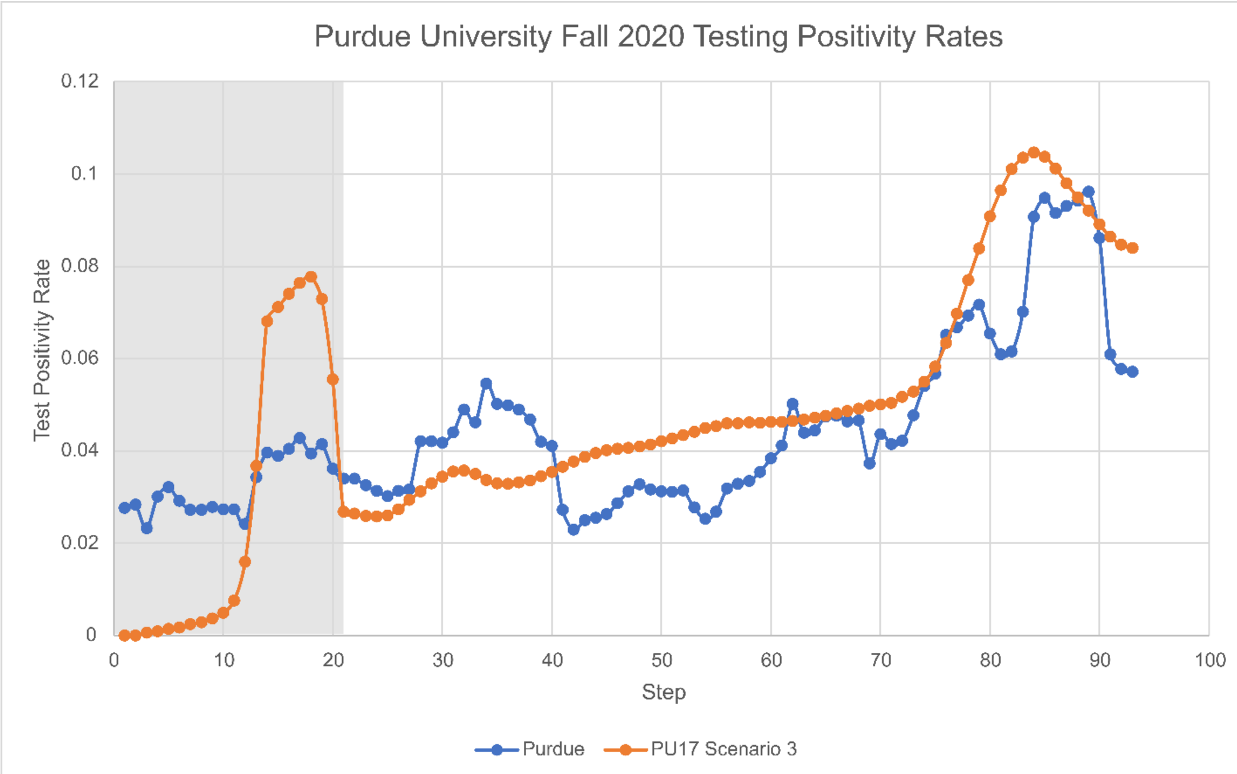

Figure 13. Final calibration run for Purdue University, Fall 2020: PU17 with Scenario 3.

Forensic Conclusions from PU for Fall 2020

Super-Spreader Event around Day 70: Our model could not adequately account for the rise in cases near the end of the semester without injecting infections into the system. From this, our model supports the conclusion that there was some spreader event in the late-middle part of the semester that does not fit within the regular abstractions of the model.

No-Transmission Comparison: We ran the calibrated model while disabling transmission to see how much of the positivity rate is due to interacting outside of the closed system. We found that the number of infections is due primarily to transmission from the internal system. Around day 30, internal transmissions make up about 70% of the positive test cases, and around day 70, internal transmissions make up about 77% of the positive test cases. Note that the spike after day 70 is due to the injected random infections, which we kept in the no-transmission model since we cannot conclude with confidence whether those infections came from campus transmission or from the external system. Figure 25 displays the results. PU is almost opposite the NU case: most infections were spread internally, not externally, indicating low compliance with the administration's stated NPIs.

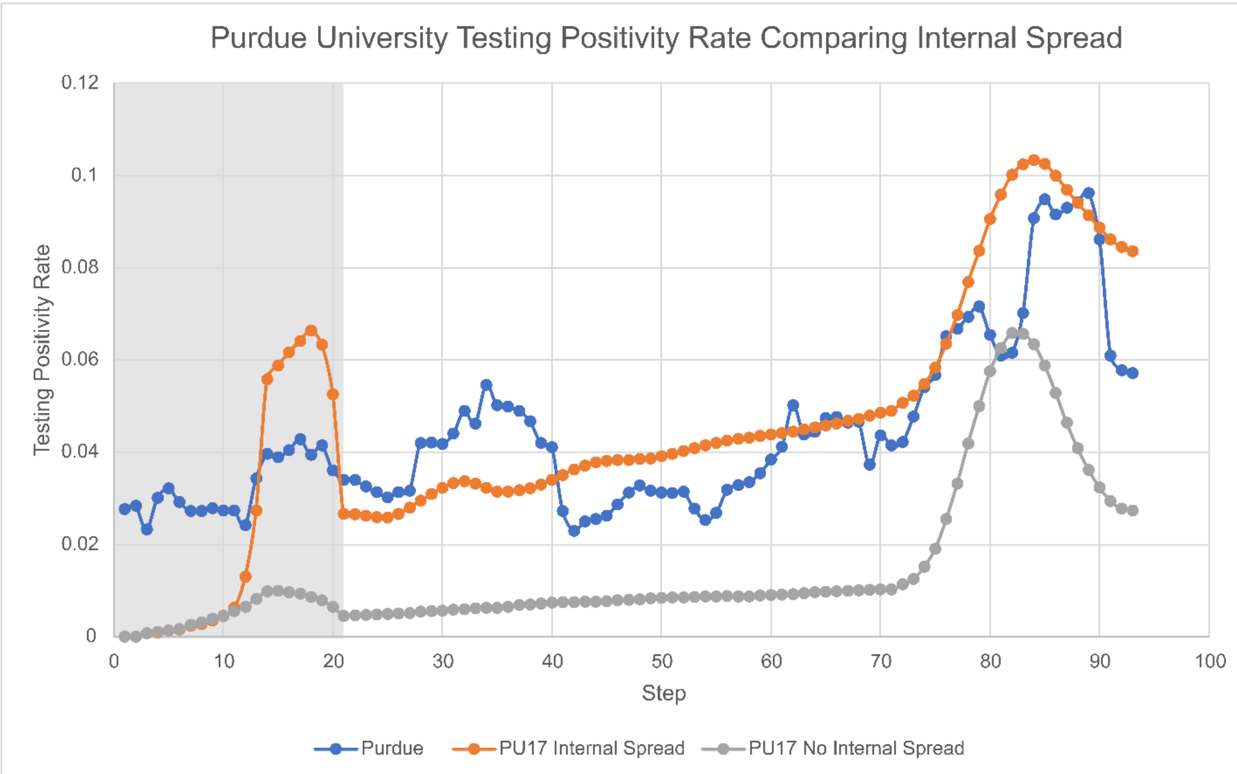

Figure 14. The calibrated model for Purdue University, Fall 2020 (PU17), run with and without internal spread.

Testing Adequacy: As with NU, so with PU: we compared the number of simulated infections in the model to the simulated test positivity outputs over time to gain some insight into the adequacy of test sampling. In Figure 26, the standard deviation range for the simulated daily test positivity rate is usually within one standard deviation of the simulated true percentage range (x 3.5). This means that in any given simulation, there is a reliable correlation between the test positivity rate and the true positivity rate.

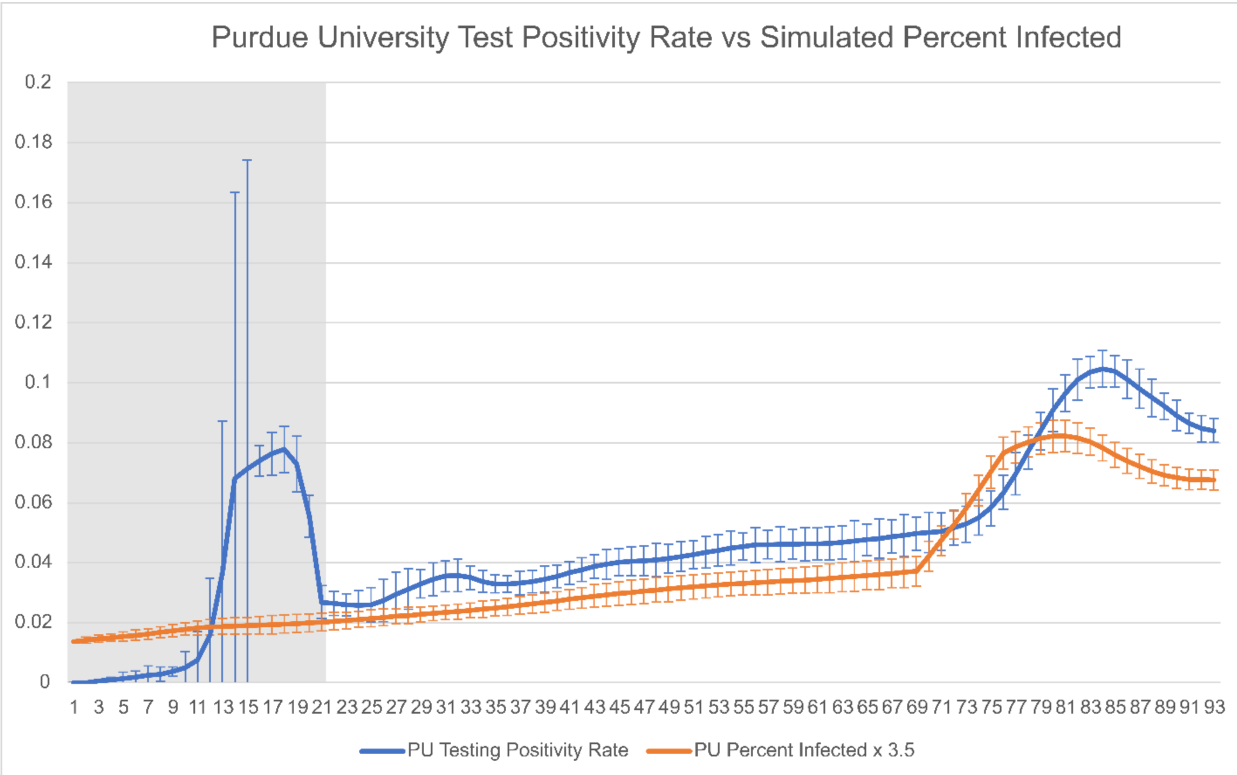

Figure 15. Purdue University, Fall 2020, simulated test positivity rate versus simulated percent actually infected.

Case Western Reserve University Fall 2020 Semester

We initialized the CW model in a similar way to the PU model, with the major differences being that we set the number of agents to 18,362 (based on online data at https://case.edu/about/), used CWRU COVID-19 test dashboard for number of tests administered weekly, and used CW regional data for Cuyahoga County, OH. Figure 27 displays these preliminary results.

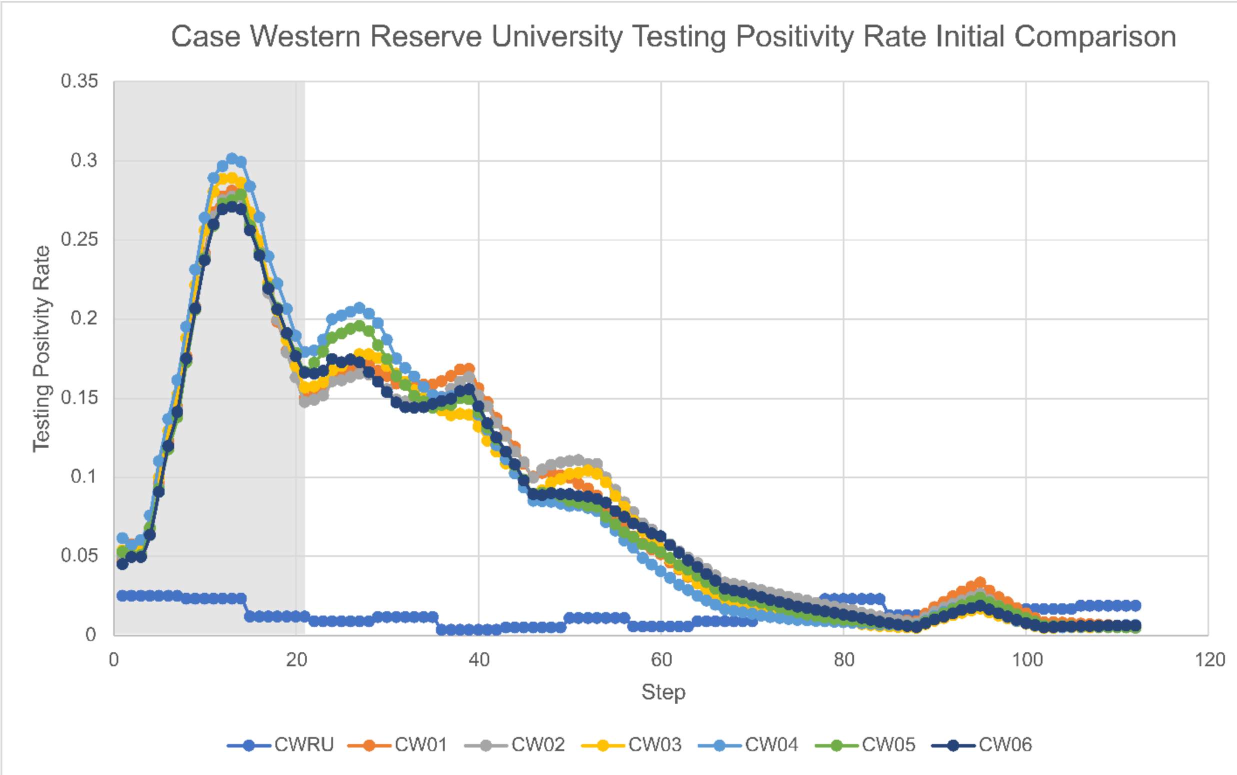

Figure 16. Calibration runs for Case Western Reserve University, Fall 2020: CW01-CW06.

It should be noted that CW only publicized weekly testing rates, and so their positivity rate is a step function. To correct for the initial flare-up of the positivity rates in the simulation runs, the multiplier for the external infection rate was varied from 0.125-0.25%, the initially infected was varied from 0.36-0.72%, and the percent of active agents was varied from 75-100%. Figure 28 displays the results of these actions.

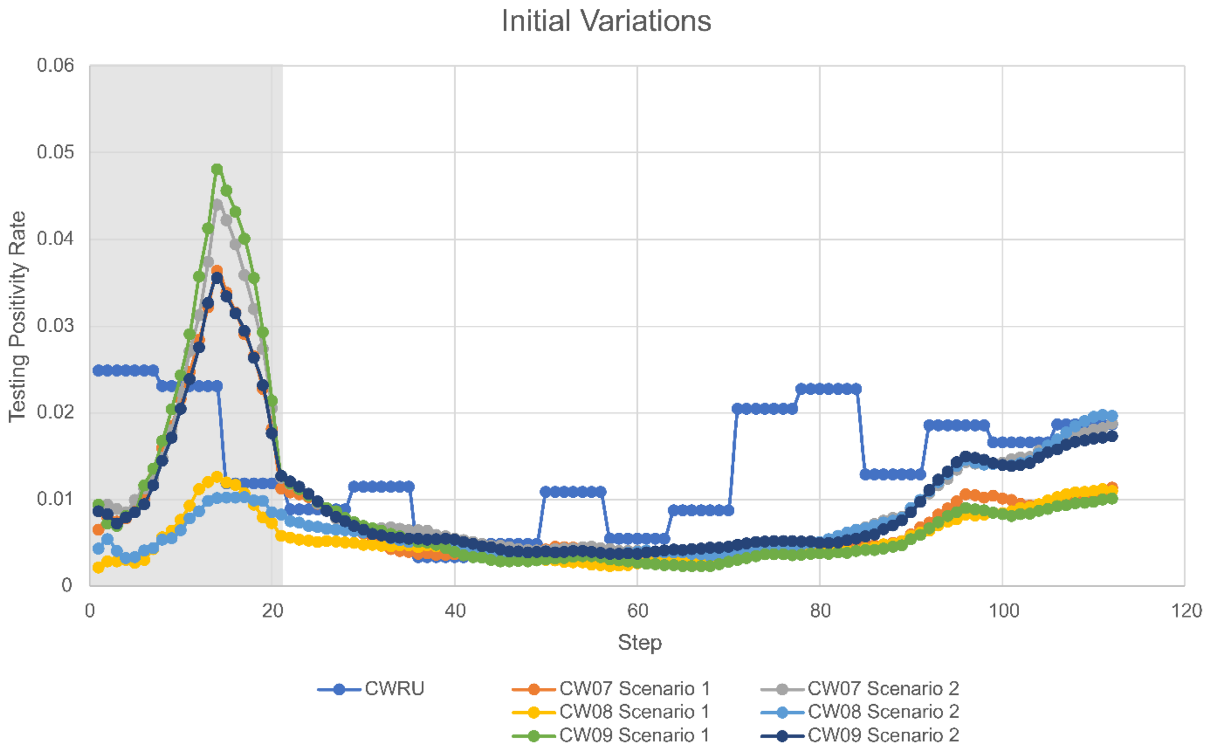

Figure 17. Calibration runs for Case Western Reserve University, Fall 2020: CW07-CW09 with Scenarios 1-2.

Noting that the simulation runs were much smoother than the observed data from CW, the interview recall rate was varied from 35-70%, and intermittent parties were added (Figure 29).

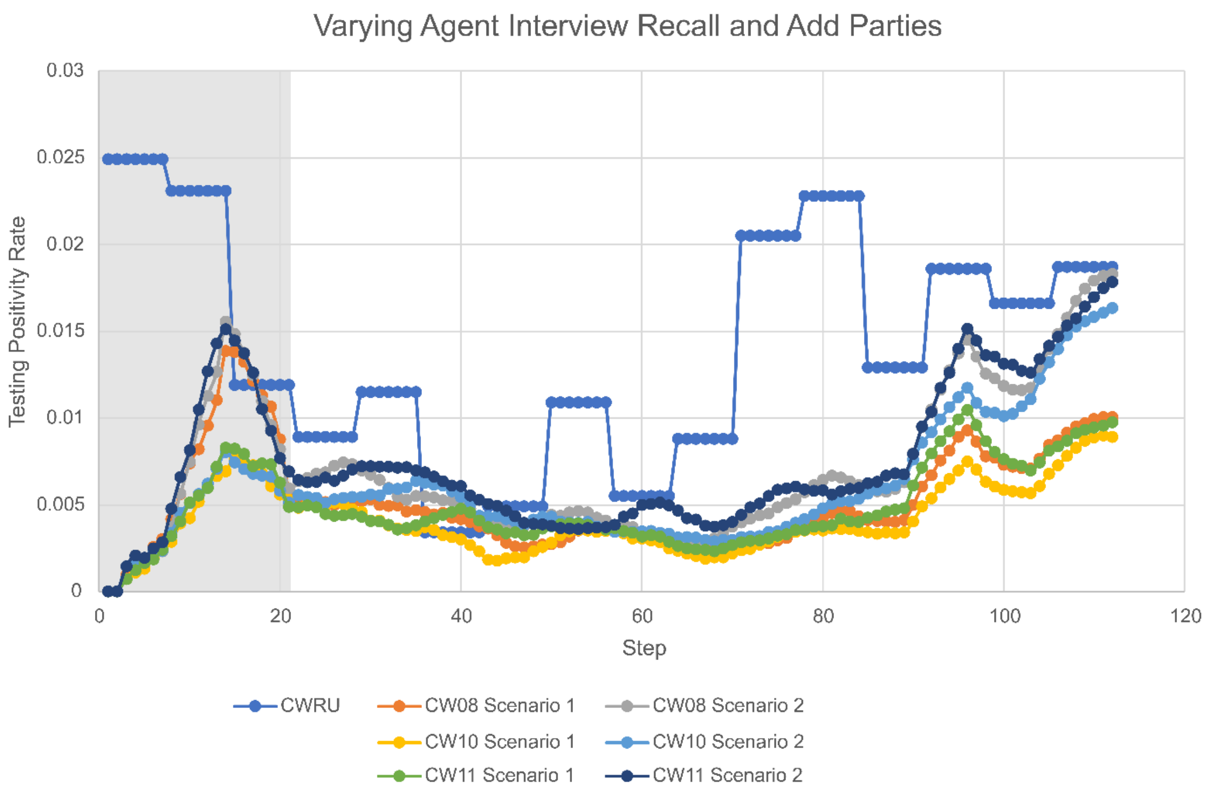

Figure 18. Calibration runs for Case Western Reserve University, Fall 2020: CW08 and CW10-CW11 with Scenarios 1-2

The final results for the CW Fall 2020 semester compared to the simulation runs involved adding random infections, adding infections at a different step than before, adding infections gradually over each week, and running the simulation a full week before testing was started to build up a contact history. Final key parameters are as follows: the interview recall rate was set at 50%, the base infection rate was 2.5%, 100% active agents were employed, and 0.36% initially infected agents were assumed (about 66 agents). Compliance values varied between 60-95%. Figure 30 displays the results.

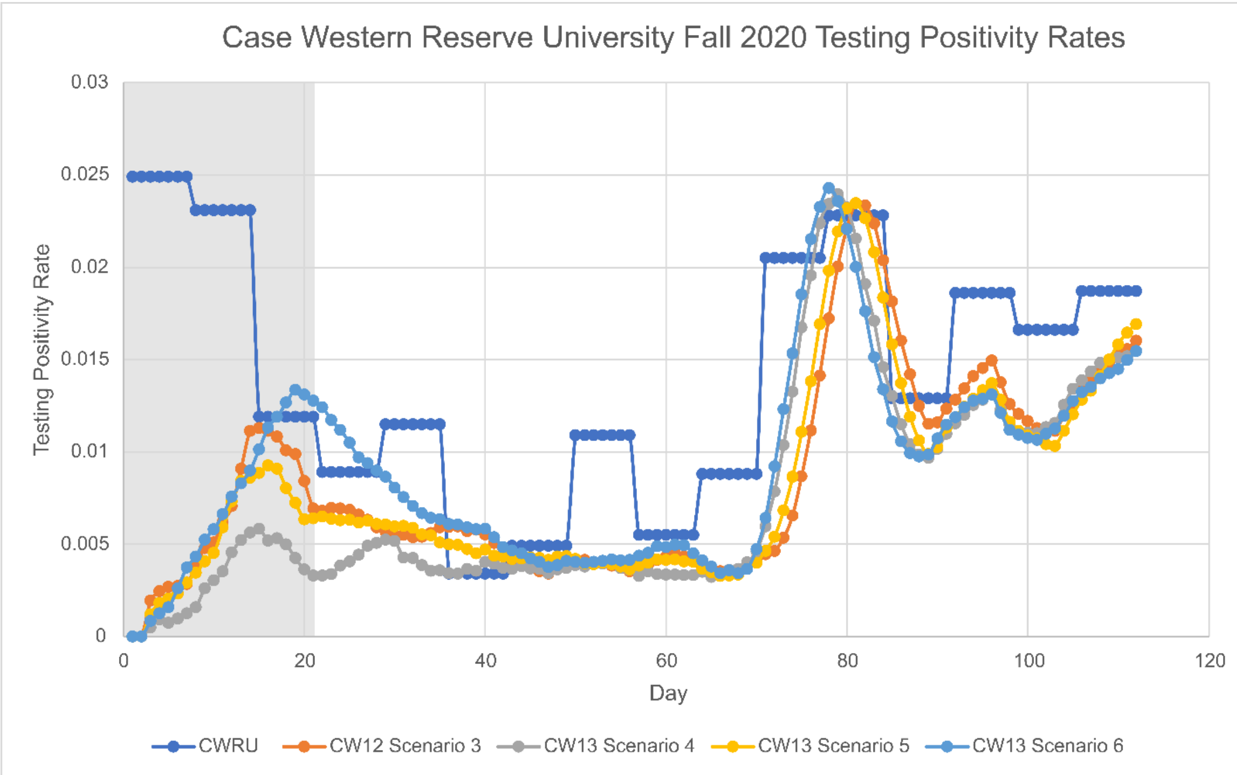

Figure 19. Final calibration runs for Case Western Reserve University, Fall 2020: CW12 and CW13 with Scenarios 4-6.

The real-world data indicates a super-spreader event about two thirds into the semester; in the aftermath of that event, CW did not closely match the external infection rate derived from the surrounding county. This suggests either that the university did not have enough testing or that people were insufficiently compliant with the university's health guidelines.

Forensic analysis: Testing Adequacy

Although we were able to calibrate TAO to match RWE test-positivity rates closely, we determined that our calibration journey was inconclusive due to inadequate testing for the CWRU Fall 2020 semester. We define the number of tests administered to be adequate if the standard deviation range of our simulated positivity rates mostly overlap with the actual simulated percent infected, adjusted with some scalar across the whole simulation output. Figure 31 shows that the standard deviation range for the simulated daily test positivity rate is always several times greater than the simulated actual percentage range (x 3.5). This means that, in an average single simulation run, the simulated test positivity rate may be off by multiple standard deviations, in a way that is not predictable.

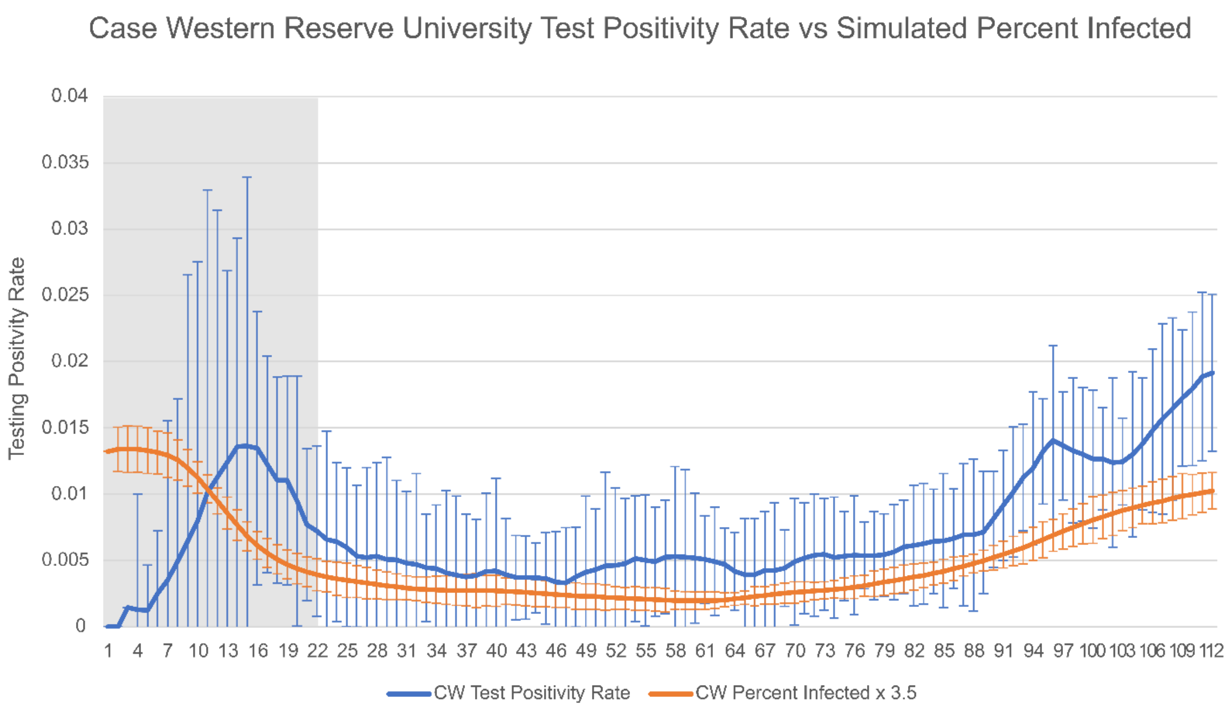

Figure 20. Case Western Reserve University, Fall 2020, simulated test positivity rate versus simulated percent actually infected.

Detailed Description of Calibration Runs

The following table describes detailed settings for calibration runs, with each run described in terms of changes relative to the previous run.

Northeastern University (NU) – name of run and description of changes between runs

NU01 Base run: 5.84E-5 external infection rate, 0.6-0.8 compliance, 15000 active agents, 0.05 base infection rate

NU02 1E-4 external infection rate

NU03 5.84E-5 external infection rate, 0.65-0.85 compliance

NU04 5.84E-5 external infection rate, 0.6-0.95 compliance

NU05 7500 active agents

NU06 2.92E-5 external infection rate, 15000 active agents

NU07 2.92E-5 external infection rate, 7500 active agents

NU08 1.46E-5 external infection rate, 15000 active agents

NU09 1.46E-5 external infection rate, 7500 active agents

NU10 Use a time series external infection rate, 15000 active agents

NU11 Use a time series external infection rate, 7500 active agents

NU12 Use a varying number of active agents over the semester

- NU Scenario 1: Vary the active agents: 30% semester start, 15% mid-semester, 25% semester end
- NU Scenario 2: Same as Scenario 1 with half the external infection rate
- NU Scenario 3: Vary the active agents: 30% semester start, 20% mid-semester, 30% semester end
- NU Scenario 4: Same as Scenario 3 with half the external infection rate

NU13 Use 0.075% initial infected agents, 20000 active agents

NU14 0.025 base infection rate

NU15 0.0125 base infection rate

NU16 0.05 base infection rate, 15000 active agents

NU17 0.025 base infection rate

NU18 0.0125 base infection rate

NU19 0.025 base infection rate, 20000 active agents, 0.00375% initial infected agents (0-1 initial infected)

NU20 0.0125 base infection rate

NU21 0.025 base infection rate, 50% agent interview recall

NU22 0.0125 base infection rate, 50% agent interview recall

NU23 0.025 base infection rate, 100% agent interview recall

NU24 Add in parties that reduce the compliance rate by 50%, 50% agent interview recall, 0.03 base infection rate

NU25 0.035 base infection rate

NU26 0.04 base infection rate

NU27 Initial spring run, same values as NU24 but using spring values for the tests per day and external infection rate time series

NU28 25000 active agents

NU29 30000 active agents

NU30 Add in vaccinations see scenario, 20000 active agents

- NU Scenario 5: Add 450 daily vaccinations starting on step 51
- NU Scenario 6: Add 450 daily vaccinations starting on step 82
- NU Scenario 7: Same as Scenario 6 and do 8,000 tests a day for the first week

NU31 25000 active agents

- NU Scenario 5: Add 450 daily vaccinations starting on step 51

NU32 30000 active agents

- NU Scenario 5: Add 450 daily vaccinations starting on step 51

NU33 20000 active agents, 10% initially recovered (up from 5% in all other runs)

- NU Scenario 6: Add 450 daily vaccinations starting on step 82
- NU Scenario 7: Same as Scenario 6 and do 8,000 tests a day for the first week

NU34 0.6-0.9 compliance (instead of 0.6-0.95)

- NU Scenario 6: Add 450 daily vaccinations starting on step 82
- NU Scenario 7: Same as Scenario 6 and do 8,000 tests a day for the first week

NU35 0.5-0.9 compliance

- NU Scenario 6: Add 450 daily vaccinations starting on step 82
- NU Scenario 7: Same as Scenario 6 and do 8,000 tests a day for the first week

NU36 5% initially recovered, 0.6-0.95 compliance

- NU Scenario 8: Add 450 daily vaccinations starting on step 75
- NU Scenario 9: Same as Scenario 8, do 20,000 tests on the first day, do 8,000 tests for the rest of the week in the first week
- NU Scenario 10: Same as Scenario 8 and reduce external infection rate to 75% of the normal rate

NU37 15000 active agents

- NU Scenario 8: Add 450 daily vaccinations starting on step 75
- NU Scenario 9: Same as Scenario 8, do 20,000 tests on the first day, do 8,000 tests for the rest of the week in the first week
- NU Scenario 10: Same as Scenario 8 and reduce external infection rate to 75% of the normal rate

NU38 10000 active agents

- NU Scenario 8: Add 450 daily vaccinations starting on step 75
- NU Scenario 9: Same as Scenario 8, do 20,000 tests on the first day, do 8,000 tests for the rest of the week in the first week
- NU Scenario 10: Same as Scenario 8 and reduce external infection rate to 75% of the normal rate

NU39 Same as NU38 but with variable active agents over the semester

- NU Scenario 11: Increase to 15,000 agents on day 60
- NU Scenario 12: Same as Scenario 11 with an increase in external infection rate (multiply by 1.25)
- NU Scenario 13: Increase to 20,000 agents on day 60
- NU Scenario 14: Same as Scenario 13 with an increase in external infection rate (multiply by 1.25)
- NU Scenario 15: Increase to 15,000 agents on day 62
- NU Scenario 16: Same as Scenario 15 with an increase in external infection rate (multiply by 1.25)
- NU Scenario 17: Increase to 20,000 agents on day 62
- NU Scenario 18: Same as Scenario 17 with an increase in external infection rate (multiply by 1.25)

Purdue University (PU) – name of run and description of changes between runs

PU01 Initial Purdue runs: variable external infection rate, 51000 agents, 49700 active agents, and 0.6-0.95 compliance

- PU Scenario 1: Use external infection scalar of 0.025
- PU Scenario 2: Use external infection scalar of 0.05

PU02 No hybrid classes, 0-20% chance student attends parties (up from 0-5%), 0-5% chance student hosts a party (up from 0-1%)

- PU Scenario 1: Use external infection scalar of 0.025
- PU Scenario 2: Use external infection scalar of 0.05

PU03 0.6-0.85 compliance

- PU Scenario 1: Use external infection scalar of 0.025
- PU Scenario 2: Use external infection scalar of 0.05

PU04 0.025 base infection rate, 0.4-0.85 compliance

PU05 0.0125 base infection rate

PU06 0.025 base infection rate, 0.5-0.85 compliance

PU07 0.0125 base infection rate

PU08 Reduce compliance at parties by 50%, 50% agent interview recall, 0.03 base infection rate, 0.8-0.95 faculty/staff compliance (up from 0.6-0.95), 0.36% initially infected (down from 0.72%), 0.6-0.95 student compliance

- PU Scenario 3: Add 500 random infections on day 70

PU09 40-90% chance student attends party, 0-25% chance student throws party

- PU Scenario 3: Add 500 random infections on day 70

PU10 0.04 base infection rate, 0-20% chance student attends party, 0-5% chance student throws party

- PU Scenario 3: Add 500 random infections on day 70

PU11 40-90% chance student attends party, 0-25% chance student throws party, add vaccines

- PU Scenario 3: Add 500 random infections on day 70

PU12 1-10% chance student attends party, 0.2-0.6 compliance, 0-5% chance student throws party

- PU Scenario 3: Add 500 random infections on day 70

PU13 0.72% initial infected, 40% agent interview recall

- PU Scenario 3: Add 500 random infections on day 70

PU14 50% agent interview recall, 0.36% initial infected, 0.4-0.6 compliance

- PU Scenario 3: Add 500 random infections on day 70

PU15 40% agent interview recall, 0.4-0.8 compliance

- PU Scenario 3: Add 500 random infections on day 70

PU16 30% agent interview recall

- PU Scenario 3: Add 500 random infections on day 70

PU17 0.5-0.8 compliance

- PU Scenario 3: Add 500 random infections on day 70

Case Western Reserve University (CW) – name of run and description of changes between runs

CW01 Initial CWRU runs: variable external infection rate and variable active agents, 0.6-0.95 compliance

CW02 0.7-0.95 compliance

CW03 0.8-0.95 compliance

CW04 Half the external infection rate (time series input), 0.6-0.95 compliance

CW05 0.7-0.95 compliance

CW06 0.8-0.95 compliance

CW07 0.6-0.95 compliance

- CW Scenario 1: Use external infection scalar of 0.0125
- CW Scenario 2: Use external infection scalar of 0.025

CW08 0.36% initial infected (down from 0.72%)

- CW Scenario 1: Use external infection scalar of 0.0125
- CW Scenario 2: Use external infection scalar of 0.025

CW09 0.72% initial infected

- CW Scenario 1: Use external infection scalar of 0.0125
- CW Scenario 2: Use external infection scalar of 0.025

CW10 0.36% initial infected, 35% agent interview recall

- CW Scenario 1: Use external infection scalar of 0.0125
- CW Scenario 2: Use external infection scalar of 0.025

CW11 70% agent interview recall, 40-90% chance student attends party, 0-25% chance student throws party

- CW Scenario 1: Use external infection scalar of 0.0125
- CW Scenario 2: Use external infection scalar of 0.025

CW12 Reduce compliance at parties by 50%, 50% agent interview recall, 0-20% chance student attends party, 0-5% chance student throws party

- CW Scenario 3: Add 50 random infections on step 68

CW13 Same main values as CW12 with multiple different scenarios

- CW Scenario 4: Add 50 random infections on step 65
- CW Scenario 5: Add 10 random infections a day for five days starting on step 65
- CW Scenario 6: Run the simulation for a week with zero tests to create contact networks. Add 50 random infections at step 72, which is still Oct 28 since we ran 7 steps before the real simulation started