Simulation Sleuthing: Using Model Calibration to Forensically Analyze Pandemic Management Strategies

Gabe McDonnell Maayan[1,2][0000-0001-5725-7622]*, Wesley J. Wildman[1,2][0000-0002-7571-1259]*, George Hodulik[2][0000-0003-0623-4564], Mark Lane[3][0000-0002-9402-2264] and G. Edward Powell, Jr.[3][0000-0002-5748-2632]

[1] Boston University, Boston MA 02215, USA
[2] Center for Mind and Culture, Boston MA 02215, USA
[3] TensorX, Nashville TN 37215, USA
gmaayan@mindandculture.org
wwildman@justhorizons.org
ghodulik@mindandculture.org
mlane@tensor-x.com
epowell@tensor-x.com

Abstract. Computational simulations designed to supply decision support for institutional managers and policy professionals need to be evaluated. Careful evaluation can validate (or invalidate) models, generate forensic insights into pandemic management strategies, and alert us to the risks involved in using artificial societies for decision support. Here, we evaluate The Artificial Organization (TAO), a decision-support tool for pandemic management, against detailed COVID-19 testing data from three universities in the United States in interestingly different physical environments: Northeastern University is urban and merged into the local community, Case Western Reserve University is urban with a relatively self-contained campus organization, and Purdue University has a rural setting. Findings shed light on the adequacy of testing regimens, evidence for super-spreader events, and the critical role of compliance with masking and social distancing. Our findings not only confirm TAO's utility for forensic analysis of pandemic-management strategies but also demonstrate the indispensable impact of integrating human factors into decision-support simulations, as TAO does.

Keywords: COVID-19, pandemic, forensics, agent-based model, validation, calibration, university, human behavior

1 Introduction

The magnitude of the COVID-19 pandemic became clear only in stages, partly due to ineffective pandemic surveillance. All along the way, there was uncertainty and controversy surrounding the pandemic and its impact. Modeling and simulation help us

* Joint first authors

understand a situation and predict future outcomes [1]. Early on, there were models from various institutions that predicted millions of deaths in the US alone [2]. These models suffered from a lack of calibration, and they did not adequately model non-pharmaceutical interventions (NPIs), accounting for the vital role within pandemic management of human behaviors and beliefs. In response, the CDC announced a $26M program in Advanced Outbreak Forecasting and Analytics [3].

Enter human simulation, which refers to computational simulations that take account of a wide range of human factors in modelling complex social phenomena. Human simulation is an extension of the established domain of computational social simulation that focuses on building agent-based models (ABM) with psychologically plausible and behaviorally believable AI agents [4], which fit well with the multi-agent AI modelling (MAAI) paradigm. Though human behavior is notoriously difficult to predict in detail, human-simulation MAAIs raise the right questions and thus generate insights of the kind that decision makers most need in pandemic management [5].

And indeed, there was a proliferation of MAAIs during the height of the pandemic. An early 2021 review [6] found 126 articles that describe an agent-based simulation of the spread of COVID-19, most of which were created for the purpose of examining pharmaceutical and non-pharmaceutical interventions and for predicting future outcomes. Despite the large quantity of models, there was a clear lack of initialization and validation using real-world data. The authors found 43.7% of the models used census data to initialize the simulated population, 22.2% used mobility data (e.g. cellphone data) to capture realistic movement patterns, and only 25% of the models described any form of validation at all. This lack of validation and use of real-world data was often caused simply because the validation data did not exist or was not publicly available in the very early days of the pandemic [6]. However, a more recent 2023 survey [7] examined 340 models (258 ABMs) related to COVID-19 and still found only 50.3% used real-data for initialization or validation.

Nearly all the models examined in the two reviews were used to generate future predictions to guide pandemic-management decisions, but human-simulation MAAIs are also useful for forensic analysis of what NPIs worked and what didn't, and under what circumstances. By forensic analysis, we mean analysis after the fact, to find out what happened and why. Scientists often move quickly on to the next challenge, but it is valuable to pause and ask how well we managed something as momentous as a pandemic. To that end, we used TAO to forensically examine pandemic management strategies in three universities in the United States.

To date, there have been several published studies that take a retrospective look at pandemic management strategies [8–13]. These studies use both differential equation-based approaches [8] as well as agent-based modeling approaches [9–13]. All use some level of demographic data to initialize the simulation, testing or healthcare data to measure outcomes. Some use other forms of data such as mobile phone mobility data [8, 9], GIS data to recreate realistic spatial landscapes [13], and WIFI-use data to estimate building occupancy [13]. Most are used to investigate NPIs though one is only focused on pharmaceutical interventions [11]. One of the studies adapts an earlier prominent agent-based model, Covasim [14], to model the pandemic in Spain [12]. Of these earlier models, Chen et al.'s study focuses specifically on pandemic policy in universities, similar to this study [13].

Universities present a unique challenge for pandemic management strategies. Communal living and a high frequency of social gatherings contribute to high virus transmission rates [15, 16]. On the other hand, university populations tend to be much younger than the general population, leading to more asymptomatic cases and fewer deaths [17]. Universities can also generally exert far tighter control over student and, to a lesser extent, staff behavior than a typical workplace, and therefore implement especially effective anti-pandemic interventions [18–20].

In this study, we present an iterative calibration process of the TAO MAAI, conducted in conjunction with university policymakers, to forensically examine the pandemic management strategies of three universities that did the general public the great favor of publishing detailed testing data for Fall 2020 and Spring 2021: Northeastern University, Purdue University, and Case Western Reserve University. In what follows, we first briefly describe TAO and spell out the verification and validation process we employed. We then present the results of our forensic analysis before concluding with a discussion of those results.

2 Methods

TAO is built on top of a MAAI produced as part of The Artificial University (TAU) project from the Center for Mind and Culture and the Virginia Modeling, Analysis, and Simulation Center at Old Dominion University. The TAU project produced two publications [21, 22], a dashboard [23], and an open-source MAAI exploring COVID-19 transmission in an abstract university setting [24]. TAU was validated by subject-matter experts and the open-source model was made available to colleges and universities with web-scraped sample data to facilitate decision support.

TAO extends TAU to more general organizations – particularly organizations that have a clear structures and schedules, including large corporations, military bases, cruise ships, etc. TAO uses Simudyne's proprietary modelling platform, but the code is available by negotiation with Simudyne, so the calibration procedure and forensic analysis presented here is replicable by other researchers. Much of TAO's model architecture is already present in TAU, which is open source, has been validated, and already establishes the value of human-simulation techniques for pandemic management. Thus, the aim of this paper is not merely to present another human-simulation model; rather, it is to show how TAO can be stress-tested with demanding real-world data and used for forensic analysis of pandemic responses, thereby revealing which strategies worked and indirectly demonstrating the value of rigorous human-simulation models.

2.1 Model Design

For a full description of the TAO model, see the supplemental materials† as well as the TAU paper [21]. TAO is made up of people agents, place agents, and a central agent, representing the central authority of university policymakers. People agents go to places where infected individuals may spread the infection to other individuals. The central agent dictates interventions such as contact tracing and testing.

At initialization, people agents are assigned a schedule of places to which they go over a two-week period, and this two-week schedule repeats itself for the duration of the simulation. There is one time-step per day in the analyses below, but TAO supports more fine-grained time steps. At the beginning of each time step, each agent consults the schedule and goes to the designated places. Agents may host an unscheduled event such as a party, and agents decide whether to attend the unscheduled events.

Interventions include masking, quarantines, hybrid classes, closing fitness centers, canceling events, etc. Infections are spread based on contact between infected and susceptible agents who are in the same place, as well as spontaneously, representing infections due to contact outside of the closed system of the organization. The effectiveness of an intervention is dependent on individual agents' compliance levels.

In addition to interventions, the central agent performs randomized testing. Test samples are collected by a randomly selected group of people, and test results are available after a delay period. The test-positive agents report their contacts over the last 14 days to the central agent; the contact list may be incomplete due to the positive agent randomly "forgetting" who they contacted. Agents on the contact list are told to isolate for 14 days after contact and may be tested depending on the contact tracing protocol. Test results are the primary output of the model, since we do not have true infectivity rates for comparison.

2.2 Technical Architecture

The model is written using the proprietary Simudyne SDK in Java and compiled into an executable JAR file that runs on a server accepting model parameters via a REST API. The server is run on a high-powered machine (100 GByte RAM, dual X5675 chips @ 3.07GHz, Linux OS) via a Python script, which reads model parameters from a csv file and sends the parameter to the model server via Simudyne's REST API. The model runs produce custom csv and parquet outputs.

We ran the model with the number of agents varying between 400 and 85,000. The runtime complexity appeared to be Order (n2.5), with n = number of agents. A model with 85,000 agents running for 120 time steps (days) took approximately 200 minutes to complete.

† Online supplemental materials are available at https://github.com/centerformindandculture/TAO.

2.3 Methodology

The point of this paper is to stress-test TAO using real-world data, thereby generating forensic insight into pandemic management strategies and to evaluate the quality of the COVID simulation that most richly incorporates human factors. Our methodology had two parts: (1) verify that model outputs are reasonably aligned with expectation, and (2) iteratively calibrate the model to match real test positivity rate data and to evaluate pandemic response policies.

Through verification, we gain some level of trust in the simulated data by making sure outputs such as infection curves under controlled circumstances look like they would for comparable epidemiological models. We also measure the sensitivity of different parameters, confirming that they have expected outputs in controlled circumstances – that is, all else being equal, high levels of compliance for physical distancing and mask wearing should lead to fewer transmissions. The sensitivity analysis has the benefit of drawing attention to the most important parameters to explore in the calibration process.

In calibration, we search for a combination of TAO parameters that yields output aligning most closely with real-world data. Recall that our goal is not to predict future outcomes or to produce highly specific quantitative insight; our goal is to gain qualitative forensic insight about the past, which also happens to amount to an extremely stringent test of TAO's quality. In short, TAO augments our forensic-ideation capability.

Verification with Infection Curves, Parameter Sweeps, and Network Sensitivity Analysis. Through the verification process, we aim to gain some level of trust in the model as well as a deep understanding of model parameters and their relative effects. Because the theoretical basis for TAO derives from earlier work on the open-source TAU, we start from a point of some trust in TAO regarding basic mechanics. But we verify anyway, in this case by seeking emergent behavior from TAO that matches infection curve outputs for specific situations routinely considered within epidemiological models.

Next, we performed model verification of specific parameters. To do this, we ran parameter sweeps of the model with a small-university scenario of 1000 students, 100 faculty, and 200 staff. We sampled the nearly 100,000 combinations of parameter settings using a grid approach, running 5808 parameter combinations, each for 120 days (one full semester), and each 30 times to ensure confidence in model outputs.

From these simulation runs, we collected outputs of cumulative infections, peak number of infections, and total deaths. We then performed network sensitivity analyses on these parameter sets and the corresponding outputs, which resulted in a normalized representation of the magnitude and direction of the impact of each parameter. Normalized output means that parameters oscillating near one have less influence on the model, and those that vary significantly in amplitude demonstrate more influence. Second-order effects may appear, but these may be difficult to interpret because of potential parameter interactions.

Calibration using Real-World Experience Data. In the calibration process, we aligned TAO's test positivity rate with real-world experience using daily or weekly positivity rates and count of tests administered in universities. We then iteratively adjusted model parameters to match output with the university data as closely as possible. This calibration journey generated forensic insight into COVID-19 transmission and university protocols during the modelled semesters.

We contacted the selected universities and spoke with the experts best placed to answer key questions about protocols, timing, and compliance. Unfortunately, university data-gathering efforts were such that these experts did not always know even basic parameters such as the number of active agents due to uncertainty about in-person attendance with hybrid classes.

Data Selection and Scraping. In the absence of complete clarity about the real-world system, the next best thing is data on fine-grained testing administration and results. We considered a dozen universities with COVID-19 dashboards and excluded universities whose dashboards did not include time-series data dating back to Fall 2020. We also excluded universities whose COVID-19 protocol explicitly involved most of the university population going primarily remote. Once we settled on three universities, we scraped the testing information from the COVID-19 dashboards.

The selected universities were Northeastern University in Boston, Massachusetts (NU), Case Western Reserve University in Cleveland, Ohio (CW), and Purdue University in West Lafayette, Indiana (PU). These three universities have diverse characteristics that make them attractive for model calibration, and therefore both for generating forensic insight and for stress-testing TAO. NU is an urban school embedded in a local community and relies on a significant portion of off-campus activity. CW is also an inner-city school, but more on-campus focused. PU is a large state-funded school and more isolated than the other two. Calibration for NU in Fall 2020 is presented in the main paper; calibration procedures for NU in Spring 2021, CW in Fall 2020, and PU in Fall 2020 are presented in the supplemental material.

Initial Model Parameters. We selected initial model parameters for each university based on public information from university websites. This involved determining the size of the student body, faculty, and staff populations, as well as checking what NPIs were in place. We used the testing administration information as input to the model, which allowed us to evaluate assumptions about the numbers of agents on campus each day or week – a point about which these universities reported considerable uncertainty. For modelling exogenous infection (agent infection from interacting outside of the university system), we used New York Times county-level COVID-19 testing results [25].

Iterative Calibration. Starting from the initialized model, we ran 30 simulation runs for each combination of parameter settings. We ran parameter variations to compare the effect of different parameter selections in the presence of possible interaction effects. After simulation runs were complete, we compared the aggregated time-series output for the test-positivity rate against real-world data. We also reviewed the aggregated time-series output for the number of active infections in the simulation. From the output of multiple simulation runs, we hypothesized reasons for discrepancies between model data and real-world data, and thereby generated new parameter settings, repeating the same process with those new settings.

At times, we sought guidance from university officials to obtain on-the-ground insight about protocols, timing, and compliance behaviors.

3 Results

3.1 Verification

For various scenarios often considered within epidemiological models, TAO produces the expected outputs. These scenarios investigate the outcomes of diseases with varied R_0 levels. In TAO, R_0 is an emergent characteristic and not an input – we create a scenario with R_0 less than one by reducing transmissibility. Network sensitivity analyses for key parameters revealed the following have the largest impacts: mask mandates, student social distancing, and percent initially infected. The impacts of the first two rely heavily on compliance.

For both lines of validation, TAO's behavior matches what we expect to see. This supports the basic confidence we have in TAO from validation of its open-source precursor, TAU: the extension from TAU to TAO did not introduce dynamics that invalidate the model. See the supplemental materials for a full analysis and figures from the verification process.

3.2 Calibration

We used the iterative calibration process for the Fall 2020 and Spring 2021 semesters for Northeastern University (NU), and the Fall 2020 semesters for both Purdue University (PU) and Case Western Reserve University (CW). We describe initial model parameters for each calibration first, and then our reasoning for each iteration. We finish each section with a summary of the forensic conclusions we can draw from the calibration process. The supplemental materials list detailed parameter settings for each calibration run, as well as closeness-of-fit statistics (the square root of the sum of the squared differences in daily test positivity).

When comparing simulation runs, we ignore the first three weeks of output because the simulation is starting with a clean slate even though real individuals have some relevant history before the beginning of the semester. Those first three weeks are shaded in the figures to indicate their relative unimportance.

We ran numerous calibration runs and designate each set of parameter values as XXyy, where XX is the abbreviation for the university and yy is a run number. For example, NU19 is the 19th set of parameter values for Northeastern University. The calibration journey was long and intensive in every case, so we report only on highlights where we were able to draw a critical conclusion that moved the process forward.

While the simulations ran for 120 days, graphs display results for the number of days for which university data is relevant – normally 100 days or a little more.

Northeastern University Fall 2020 Semester. We set initial model parameters to be 30,000 total agents, based on online statistics [26, 27]. We used TAO's default generated university contact network and agent schedule. NU's primary NPIs were physical distancing, mask-wearing, and hybrid classes (allowing students the option to attend classes virtually or in person). All agents who were on campus for consecutive days did testing. The number of tests per day was determined by the real-world number of tests administered by NU each day.

We started the calibration process by varying the external infected rate and compliance rates. In all variations, the simulated test positivity rate was too high. Even with the lowest external infection rate (6e-5) and the highest compliance range (60-95%), we still saw positivity rates significantly higher than NU. To reduce positivity rates, we used the high compliance range, reduced the number of active agents, and varied the external infection rate (note: an external infection rate of p% means that agents have a p% daily chance of spontaneous infection from outside the closed university system). This yielded results closer to NU, but simulation test positivity rates were flat compared to NU, which had a noticeable slope after day 60.

We concluded that a flat daily external infection rate was causing the test positivity rates to be flat. Thus, we allowed the daily external infection rate to fluctuate based on the number of COVID-19 cases in the region. We generated an infection rate time series based on census population data and New York Times COVID data (for NU, we used Suffolk County data).[‡] Parallel calibration efforts for PU also suggested the base infectivity rate to be too high so we reduced it in the NU model. Our next simulation outputs were much closer to the NU test results.

At this time, we met with Dr. Sehyo Yune, Assistant Vice Chancellor for COVID-19 Wellness at NU, to gather feedback and check model assumptions. Dr. Yune estimated that two-thirds of people (20,000) were actively coming to campus, which was higher than our inputs. Dr. Yune agreed about our agent compliance distribution (60%-80%) and did not express any major concerns about other parameter settings. With this feedback, we changed the model parameters to include 20,000 active agents throughout the semester.

Further iterations included modifications to the number of initially infected agents, agent recall in the contact tracing process. Key values in the final iteration of the NU Fall 2020 calibration were an interview recall of 50%, masking and distancing compliance of 60-95%, 20,000 active agents, 0.00375% initially infected agents (or about 1 agent), and test delays of 2 days. The base infectivity rate varied between 3-4%. Fig. 1 displays the results.

While the final set of parameters does not yield a perfect match, it is close. We see a gradual rise that is similar to NU up to day 80. We then see a spike of similar

---

[‡] With P as the population for a county from US Census data, C_n as the number of current COVID-19 cases in that county on day n, and S as a scalar value, we calculate the external infection rate (EI_n) for a county on day n as: $EI_n = S \times (C_n - C_{n-14}) / P$. That is, the external infection rate is the number of new infections from two weeks previous divided by the population and multiplied by a scalar. Through experimentation we found the best value for S to be between 0.025-0.05 (between 2.5% and 5%).

magnitudes in both the simulation and NU rates, with the simulation a few days earlier, and both the simulation and NU finish with an upward trend in positivity rates.

Forensic Conclusions from NU for Fall 2020 Semester. From the NU Fall 2020 calibration procedure, we reached several conclusions. First, we found that using a variable external infection rate based on regional case numbers was much more effective than a flat external infection rate. Second, with parallel PU Fall 2020 calibration, we determined that the calibrated value for our base infection rate was likely between 2.5% and 5%. Third, we discovered the necessity of a contact tracing interview recall parameter to make TAO's simulated contact tracing process match its unreliable real-world counterpart more adequately.

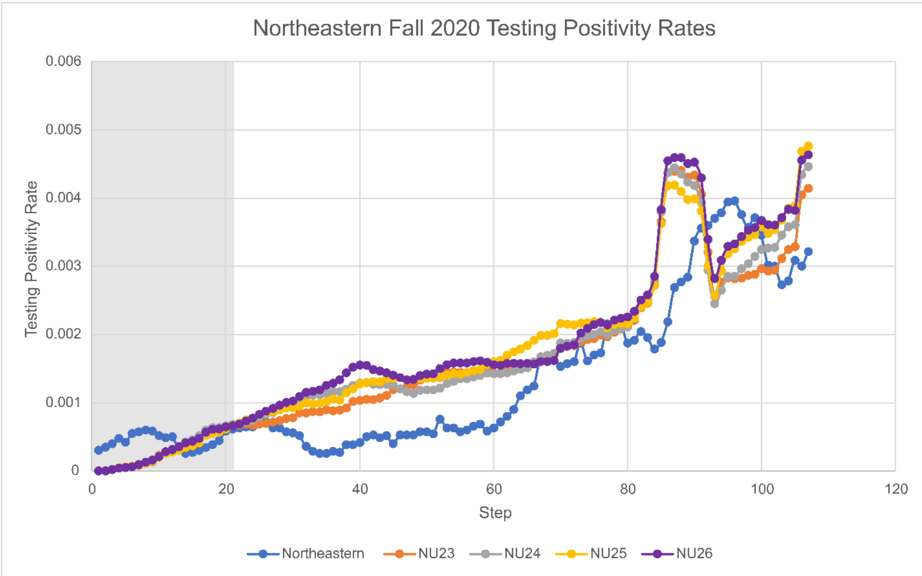

Fig. 1. Final calibration runs for Northeastern University, Fall 2020.

No-Transmission Comparison. We ran the calibrated model while disabling within-university transmission to see how much of the positivity rate was due to interacting outside of the closed system. Positive infections are primarily made up of transmission from the external system. Around day 55, internal transmissions make up about 33% of the positive test cases, and around day 86, internal transmissions make up about 25% of the positive test cases. This indicates relatively successful suppression of on-campus transmissions, which is particularly important given NU's urban, embedded location.

Testing Adequacy. TAO tracks the actual number of infections within the artificial university at any given time. We can compare this number to the simulated outputs for test positivity over time to gain insight into the adequacy of test sampling for detecting the number of infections. We define the number of tests administered to be adequate if the standard-deviation range of simulated positivity rates mostly overlap with the simulated percent infected, adjusted with some scalar across the whole simulation period. We found that the standard deviation range for the simulated daily test positivity rate is almost always within one standard deviation of the infectivity rate derived from simulated actual infections (x 3.5). This means that in any given simulation, there is a reliable correlation between the test positivity rate and the actual positivity rate.

Purdue University Fall 2020 Semester. The full Purdue University calibration exercise is described in the supplemental materials, but we include the final calibration run here for comparison with the Northeastern University calibration (Fig. 2). Notable characteristics of the Purdue University model includes (1) a super-spreader event causing 500 infections at day 70, (2) low contact tracing interview recall, and (3) medium compliance.

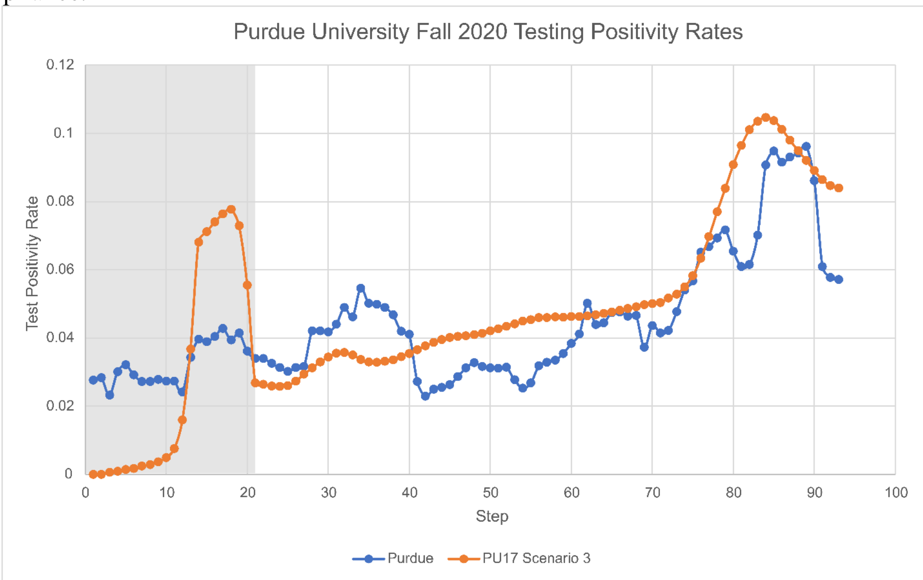

Fig. 2. Final calibration runs for Purdue University, Fall 2020.

The first is most notable because our model could not adequately account for the rise in cases near the end of the semester without injecting infections into the system, leading to the conclusion that there was some spreader event in the late-middle part of the semester.

4 Discussion

It is said that all simulation models are wrong but that some can be useful. In practice, research teams developing simulations of the COVID-19 pandemic have not ventured to find out how wrong their simulation models have been. Here we have reported calibration results for TAO – the COVID simulation that takes extensive account of human factors – to see how wrong it is. We calibrated our model against multiple real-world data sets in multiple time periods (Fall semester data from three universities, Spring semester data from one). Our forensic analysis has shed light on the following topics: Test Frequency Adequacy, Level of Intra-Organizational Spread, Outbreak Information, Time-Relative Person Activity Levels, and Importance of Modelling Human Factors and Evaluating Models.

Test Frequency Adequacy. In any situation where a sample of a population is being surveyed or tested, we must ask if the sample is representative of the population. In the domain of SARS-CoV-2 testing, specifically, we should ask if the test positivity rate of a sample of the population is representative of the true positivity rate within the population. In a more static system, we can easily calculate a sample size that will represent the population with a low confidence interval and high power. However, when monitoring the highly dynamic system of disease transmission, the sample size calculation is typically less clear. This is especially true when the true rate of infection is low from a statistical perspective (< 5%), and the tests themselves may be inaccurate.

TAO permits us to set the number of tests per day and compare simulated test-positivity rates to the simulated percentage of infections (both of which are among the model outputs). This comparison shows how well the simulated testing rate samples the entire university population. Our analysis (above and in the supplemental materials) suggests that NU and PU had testing rates that allowed a sample to be procured that matched the general trends in the tested population. On the other hand, CW had testing rates that led to volatile outputs in our simulation, suggesting that test-positivity rates for CW in Fall 2020 were too low to assess reliably the number of infections within the university population.

Level of Intra-Organizational Spread. In our model, SARS-Cov2 infection can occur in two ways: between an infected agent and a susceptible agent (intra-organizational), or between a susceptible agent and the outer system (extra-organizational). Intra-organizational spread is determined by agent-to-agent contact. Extra-organizational spread is estimated by looking up the regional test positivity rates for the simulated day, and estimating how often a simulated person would interact outside the system (e.g. use public transit).

By running the best-calibrated model with the parameter setting to disable intra-system spread, and comparing the output to the model with intra-system spread enabled, we found that intra-organizational spread was more significant in some universities and semesters than in others. In the TAO simulations for NU Fall 2020, intra-system spread played only a small role in affecting the test positivity rate output. By contrast, in the same semester for PU (see the supplement), intra-system spread plays the dominant role in affecting the test positivity rate output. This indicates strongly contrasting patterns of human behavior, with NU showing far greater compliance than PU.

Outbreak Information. During calibration, we noted that for CW and PU in Fall 2020, we could not initially reproduce some spikes in test positivity. We attribute this to the epistemic uncertainty and stochasticity of the system. For instance, a student may throw a party that becomes a super-spreader event. It is out of the scope of this model to include the cognitive processes by which the student chooses a date, venue, guests, etc., and some of those processes may be stochastic.

What we could do is note that our simulated output was not spiking when there were spikes in the real-world data, and ask "what kind of outbreak event would cause that spike?" In our PU and CW Fall 2020 calibration, we noted spikes in positivity rates around days 70 and 60 respectively that TAO did not reliably simulate. By introducing an outbreak event (injecting infections randomly into the population), the model's test-positivity rates more closely matched the data. Thus, the model suggests there was an outbreak event, or a series of outbreak events at those times. Note that for CW, test frequency inadequacy (see above) does affect this interpretation, as the spike could also be significantly affected by the random chance of test sample selection.

Time-Relative Person Activity Levels. When calibrating for NU Spring 2021, we found that our model significantly overestimated the test-positivity rate. COVID-fatigue, winter weather, and viable vaccines were all factors that led us to believe that the Spring would be a very different system than the Fall. Additionally, we were modelling the Spring semester as a separate system, when the Spring 2021 semester might be viewed as a continuation of the Fall 2020 system, particularly in view of university staff who continue working while students depart for the break between semesters.

Three main candidate explanations for the model's lower test-positivity rate are increased compliance, vaccine availability, and lower levels of in-person activity. Our collaborating NU expert, Dr. Yune, estimated that compliance with COVID-19 protocols was lower in the Spring 2021 than in Fall 2020, eliminating increased compliance as a likely explanation. Dr. Yune also said that vaccines were not widely available beyond people with special vulnerabilities during most of Spring 2021, so there is no likely explanation in that direction, either. Dr. Yune believed that the number of people present on campus probably increased from the Fall to the Spring, but noted that there is no reliable way to determine the amount of in-person activity because there may have been an increase in the number of students who opted to attend classes virtually even though living in the local area.

Thus, the NU Spring 2021 model may support the conclusion that person activity levels were lower in NU Spring 2021 than NU Fall 2020 due to some factor that the model does not consider. One hypothesis is that NU Students were less active in the winter, allowing for fewer intra-system and extra-system transmissions. This matches human behavior in cold climates and makes lots of sense in a pandemic context (Hong 2016, Kimura et al. 2015, Lee 2020, Turrisi et al. 2021).

Importance of Modelling Human Factors and Evaluating Models. Our forensic analysis of the pandemic-management strategies and performance in three universities during Fall 2020 and Spring 2021 using The Artificial Organization (TAO) has generated several specific insights (see above). But it has also supported two general conclusions.

First, a design point: it is critical to take account of human factors in computational simulations of disease transmission. Human factors determine the effectiveness of non-pharmaceutical interventions and taking account of human factors greatly impacts the accuracy of model projections.

Second, an ethical point: it is only fair that models offered for public consumption and use by organizational leaders and policy professionals should be evaluated as soon as circumstances permit. We have done this with TAO, demonstrating that TAO's modelling of pandemic-management policies and human behaviors is close enough to reality to generate useful insights into pandemic management.

Supplemental Materials. Online supplemental materials along with model code can be found at https://github.com/centerformindandculture/TAO.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Law AM (2024) Simulation modeling and analysis, Sixth edition. McGraw-Hill education, New York
2. Ferguson N, Laydon D, Nedjati Gilani G, Imai N, Ainslie K, Baguelin M, Bhatia S, Boonyasiri A, Cucunuba Perez Z, Cuomo-Dannenburg G, Dighe A, Dorigatti I, Fu H, Gaythorpe K, Green W, Hamlet A, Hinsley W, Okell L, Van Elsland S, Thompson H, Verity R, Volz E, Wang H, Wang Y, Walker P, Winskill P, Whittaker C, Donnelly C, Riley S, Ghani A (2020) Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. Imperial College London
3. Centers for Disease Control and Preventions (2021) CDC To Invest $26 Million in Advanced Outbreak Forecasting and Analytics. In: CDC. https://archive.cdc.gov/www_cdc_gov/media/releases/2021/p-1026-mill-funding.html. Accessed 18 Apr 2025
4. Diallo SY, Wildman WJ, Shults FL, Tolk A (2019) Human Simulation: Perspectives, Insights, and Applications. Springer International Publishing, Cham
5. Squazzoni F, Polhill JG, Edmonds B, Ahrweiler P, Antosz P, Scholz G, Chappin É, Borit M, Verhagen H, Giardini F, Gilbert N (2020) Computational Models That Matter During a Global Pandemic Outbreak: A Call to Action. J Artif Soc Soc Simul 23:10
6. Lorig F, Johansson E, Davidsson P (2021) Agent-Based Social Simulation of the Covid-19 Pandemic: A Systematic Review. J Artif Soc Soc Simul 24:5
7. Sun Z, Bai R, Bai Z (2023) The application of simulation methods during the COVID-19 pandemic: A scoping review. J Biomed Inform 148:104543. https://doi.org/10.1016/j.jbi.2023.104543
8. Fair KR, Karatayev VA, Anand M, Bauch CT (2022) Estimating COVID-19 cases and deaths prevented by non-pharmaceutical interventions, and the impact of individual actions: A retrospective model-based analysis. Epidemics 39:100557. https://doi.org/10.1016/j.epidem.2022.100557

9. Bosman M, Cordon Y, Duran-Sala M, Gabbanelli L, García-Pérez C, Jordan X, Manera M, Masjuan P, Medina A, Mir LM, Oròs A, Vitagliano V (2024) An agent based simulation of COVID-19 history in Catalonia using extensive real datasets. Sci Rep 14:31858. https://doi.org/10.1038/s41598-024-83238-1

10. Græsbøll K, Eriksen RS, Kirkeby C, Christiansen LE (2024) Digital twin simulation modelling shows that mass testing and local lockdowns effectively controlled COVID-19 in Denmark. Commun Med 4:1–9. https://doi.org/10.1038/s43856-024-00621-9

11. Cattaneo A, Vitali A, Mazzoleni M, Previdi F (2022) An agent-based model to assess large-scale COVID-19 vaccination campaigns for the Italian territory: The case study of Lombardy region. Comput Methods Programs Biomed 224:107029. https://doi.org/10.1016/j.cmpb.2022.107029

12. Rodríguez JP, Aleta A, Moreno Y (2023) Digital cities and the spread of COVID-19: Characterizing the impact of non-pharmaceutical interventions in five cities in Spain. Front Public Health 11:. https://doi.org/10.3389/fpubh.2023.1122230

13. Chen Y, Islam MT, Jain S, Barua Chowdhury BD, Son Y-J (2025) A Retrospective Evaluation of Pandemic Policy Impact on University Campus: An Agent-based Modeling Approach for Mobility, Disease Propagation, and Testing During COVID-19. Expert Syst Appl 266:126124. https://doi.org/10.1016/j.eswa.2024.126124

14. Kerr CC, Stuart RM, Mistry D, Abeysuriya RG, Rosenfeld K, Hart GR, Núñez RC, Cohen JA, Selvaraj P, Hagedorn B, George L, Jastrzębski M, Izzo AS, Fowler G, Palmer A, Delport D, Scott N, Kelly SL, Bennette CS, Wagner BG, Chang ST, Oron AP, Wenger EA, Panovska-Griffiths J, Famulare M, Klein DJ (2021) Covasim: An agent-based model of COVID-19 dynamics and interventions. PLOS Comput Biol 17:e1009149. https://doi.org/10.1371/journal.pcbi.1009149

15. Vang KE, Krow-Lucal ER, James AE, Cima MJ, Kothari A, Zohoori N, Porter A, Campbell EM (2021) Participation in Fraternity and Sorority Activities and the Spread of COVID-19 Among Residential University Communities — Arkansas, August 21–September 5, 2020. MMWR Morb Mortal Wkly Rep 70:20–23. https://doi.org/10.15585/mmwr.mm7001a5

16. Zivich PN, Eisenberg MC, Monto AS, Uzicanin A, Baric RS, Sheahan TP, Rainey JJ, Gao H, Aiello AE (2020) Transmission of viral pathogens in a social network of university students: the eX-FLU study. Epidemiol Infect 148:e267. https://doi.org/10.1017/S0950268820001806

17. Davies NG, Klepac P, Liu Y, Prem K, Jit M, Eggo RM (2020) Age-dependent effects in the transmission and control of COVID-19 epidemics. Nat Med 26:1205–1211. https://doi.org/10.1038/s41591-020-0962-9

18. Hamer DH, White LF, Jenkins HE, Gill CJ, Landsberg HE, Klapperich C, Bulekova K, Platt J, Decarie L, Gilmore W, Pilkington M, MacDowell TL, Faria MA, Densmore D, Landaverde L, Li W, Rose T, Burgay SP, Miller C, Doucette-Stamm L, Lockard K, Elmore K, Schroeder T, Zaia AM, Kolaczyk ED, Waters G, Brown RA (2021) Assessment of a COVID-19 Control Plan on an Urban University Campus During a Second Wave of the Pandemic. JAMA Netw Open 4:e2116425. https://doi.org/10.1001/jamanetworkopen.2021.16425

19. Denny TN, Andrews L, Bonsignori M, Cavanaugh K, Datto MB, Deckard A, DeMarco CT, DeNaeyer N, Epling CA, Gurley T, Haase SB, Hallberg C, Harer J, Kneifel CL, Lee MJ, Louzao R, Moody MA, Moore Z, Polage CR, Puglin J, Spotts PH, Vaughn JA, Wolfe CR (2020) Implementation of a Pooled Surveillance Testing Program for Asymptomatic SARS-CoV-2

Infections on a College Campus — Duke University, Durham, North Carolina, August 2–October 11, 2020. MMWR Morb Mortal Wkly Rep 69:1743–1747. https://doi.org/10.15585/mmwr.mm6946e1

20. Ranoa DRE, Holland RL, Alnaji FG, Green KJ, Wang L, Fredrickson RL, Wang T, Wong GN, Uelmen J, Maslov S, Weiner ZJ, Tkachenko AV, Zhang H, Liu Z, Ibrahim A, Patel SJ, Paul JM, Vance NP, Gulick JG, Satheesan SP, Galvan IJ, Miller A, Grohens J, Nelson TJ, Stevens MP, Hennessy PM, Parker RC, Santos E, Brackett C, Steinman JD, Fenner MR, Dohrer K, DeLorenzo M, Wilhelm-Barr L, Brauer BR, Best-Popescu C, Durack G, Wetter N, Kranz DM, Breitbarth J, Simpson C, Pryde JA, Kaler RN, Harris C, Vance AC, Silotto JL, Johnson M, Valera EA, Anton PK, Mwilambwe L, Bryan SP, Stone DS, Young DB, Ward WE, Lantz J, Vozenilek JA, Bashir R, Moore JS, Garg M, Cooper JC, Snyder G, Lore MH, Yocum DL, Cohen NJ, Novakofski JE, Loots MJ, Ballard RL, Band M, Banks KM, Barnes JD, Bentea I, Black J, Busch J, Conte A, Conte M, Curry M, Eardley J, Edwards A, Eggett T, Fleurimont J, Foster D, Fouke BW, Gallagher N, Gastala N, Genung SA, Glueck D, Gray B, Greta A, Healy RM, Hetrick A, Holterman AA, Ismail N, Jasenof I, Kelly P, Kielbasa A, Kiesel T, Kindle LM, Lipking RL, Manabe YC, Mayes J́, McGuffin R, McHenry KG, Mirza A, Moseley J, Mostafa HH, Mumford M, Munoz K, Murray AD, Nolan M, Parikh NA, Pekosz A, Pflugmacher J, Phillips JM, Pitts C, Potter MC, Quisenberry J, Rear J, Robinson ML, Rosillo E, Rye LN, Sherwood M, Simon A, Singson JM, Skadden C, Skelton TH, Smith C, Stech M, Thomas R, Tomaszewski MA, Tyburski EA, Vanwingerden S, Vlach E, Watkins RS, Watson K, White KC, Killeen TL, Jones RJ, Cangellaris AC, Martinis SA, Vaid A, Brooke CB, Walsh JT, Elbanna A, Sullivan WC, Smith RL, Goldenfeld N, Fan TM, Hergenrother PJ, Burke MD (2022) Mitigation of SARS-CoV-2 transmission at a large public university. Nat Commun 13:3207. https://doi.org/10.1038/s41467-022-30833-3

21. Wildman WJ, Diallo SY, Hodulik G, Page A, Tolk A, Gondal N (2020) The Artificial University: Decision Support for Universities in the COVID-19 Era. Complexity 2020:5910209. https://doi.org/10.1155/2020/5910209

22. Page A, Diallo SY, Wildman WJ, Hodulik G, Weisel EW, Gondal N, Voas D (2022) Computational Simulation Is a Vital Resource for Navigating the COVID-19 Pandemic. Simul Healthc 17:e141. https://doi.org/10.1097/SIH.0000000000000572

23. Center for Mind and Culture (2020) The Artificial University: Campus Pandemic Management. https://mindandculture.org/projects/past-projects/vivid-dashboard/. Accessed 22 Apr 2025

24. Center for Mind and Culture (2020) The Artificial University: Documentation and Code

25. The New York Times (2020) Coronavirus in the U.S.: Latest Map and Case Count. N. Y. Times

26. Northeastern University (2021) 2020-2021: Common Data Set. In: Univ. Decis. Support Northeast. Univ. https://uds.northeastern.edu/cds/2020-2021/. Accessed 22 Apr 2025

27. Northeastern University (2021) 2021 Facts and Figures. https://facts.northeastern.edu/2021. Accessed 22 Apr 2025