Modeling for COVID-19 College Reopening Decisions: Cornell, A Case Study

Peter I. Frazier[a,1,3], **J. Massey Cashore**[a,2], **Ning Duan**[a,2], **Shane G. Henderson**[a,2], **Alyf Janmohamed**[a,2], **Brian Liu**[a,2], **David B. Shmoys**[a,2], **Jiayue Wan**[a,2], **and Yujia Zhang**[b,2]

[a]School of Operations Research & Information Engineering, Cornell University, Ithaca, NY 14850; [b]Center for Applied Mathematics, Cornell University, Ithaca, NY 14850

This manuscript was compiled on July 8, 2021

We consider the use of epidemiological models to support college reopening decisions during the COVID-19 pandemic: whether to reopen for in-person instruction and, if so, what interventions to implement. The central challenge in this use of mathematical modeling is sensitivity of predictions to input parameters coupled with uncertainty about these parameters. Substantial uncertainty remains today and was even larger when decisions were made for the fall 2020 semester. Moreover, universities' unique characteristics hinder translation of outcomes, models and parameters from the general population: a high fraction of young people, who have higher rates of asymptomatic disease and social contact, intermixing of these young people with more vulnerable individuals, and an enhanced ability to implement behavioral and testing interventions. We describe how epidemiological models supported Cornell University's decision to reopen for in-person instruction in fall 2020 and supported the design of an asymptomatic screening program instituted concurrently to prevent viral spread. We demonstrate how the structure of these decisions allowed risk to be minimized despite parameter uncertainty and how this generalizes to other university settings. We find, in particular, that twice-per-week asymptomatic screening of undergraduates provides robust protection against COVID-19 across a range of parameter settings, including parameters derived from a retrospective analysis based on data collected in Cornell's fall 2020 semester. In this retrospective analysis, parameter uncertainty did indeed cause substantial differences between predicted and realized outcomes but the quality of the decisions made was nonetheless high when compared to the best decisions that could have been made in hindsight.

COVID-19 | Epidemiological Modeling | Parameter Uncertainty | Asymptomatic Screening

Can universities safely reopen for in-person instruction during the COVID-19 pandemic? If so, then how? Universities across the globe faced this question in summer 2020 and face similar questions today as they contemplate in-person instruction with partially vaccinated student populations and more transmissible variants with the potential for immune escape in fall 2021. This question has significant consequences since virtual instruction degrades educational and mental health outcomes (1) but virus outbreaks in student populations threaten the health of students, more vulnerable staff and faculty that interact with them, and community members. This question was challenging to answer in summer 2020, in part because experiences at the city, state and national level do not easily generalize to university populations. Indeed, university populations are younger than the general population and thus have increased rates of contact (2) that may elevate virus transmission (3, 4). Mitigating this risk but adding to the question's difficulty, universities could choose from a broad collection of interventions, many of which would be difficult or impossible to implement for the general population: mandatory testing of students upon arrival to campus (5, 6); social distancing measures in and out of the classroom (7); behavioral contracts (8); travel restrictions; contact tracing; and asymptomatic screening (9, 10).

Universities have responded to this central question in dramatically different ways. Many went completely online in fall 2020 and spring 2021 while some opened with a few modest interventions (11). Others, like Cornell University's Ithaca campus (12), opened with an aggressive set of interventions including social distancing, asymptomatic surveillance and travel restrictions. This diversity in approaches reflects, in part, a diversity of circumstance, such as proximity and interaction with population centers, prevalence in those population centers, availability of housing to quarantine students, and the desires of the surrounding community, all of which should be considered (13). However, it also reflects a fundamental lack of knowledge about how policy translates into outcomes. Even today, with the opportunity to look back on the 2020-2021 academic year, the extent to which university outcomes are explained by interventions, circumstance, differences in under-reporting bias, or luck is not completely clear. Such uncertainty and diversity in approach among universities reflects the larger response to the pandemic, in which US states and national governments adopted dramatically different responses

Significance Statement

- The decision of whether to reopen universities directly impacts 7% of the US population (students, staff) and indirectly impacts tens of millions more (families, communities).
- After witnessing large COVID-19 outbreaks among students in the 2020-21 academic year, universities are planning for future semesters despite uncertainty about vaccine hesitancy, more transmissible variants with the potential for immune escape, and community prevalence. They must decide whether to bring students back for instruction and, if they do, what interventions to implement.
- While uncertainty in input parameters required by epidemiological models introduces substantial uncertainty in outcomes, we show that models can nonetheless provide insight into college reopening decisions that minimize risk.

Frazier led the study. Henderson and Shmoys provided additional modeling and research guidance. Cashore, Duan, Janmohamed, Liu, Wan and Zhang provided additional modeling and computation.

[1]Lead author. [2]All others contributed equally.

[3]To whom correspondence should be addressed. E-mail: pf98cornell.edu

to the pandemic despite apparently similar circumstances.

Epidemic models would seem to offer the power to resolve this uncertainty in support of high-quality decisions. They allow prediction, customized to the circumstances of a university, city, state, or nation. By varying the interventions *in silico* and observing predicted outcomes one can hope to choose the best course of action.

Ostensibly, this strategy requires predictions to be accurate. Unfortunately, epidemic models only approximate reality (14). Ever-present uncertainty in model input parameters coupled with the potential for exponential growth significantly limit accuracy. Small differences in behavioral and biological parameters can cause huge differences in predicted case counts. As a consequence, epidemic models have been maligned for producing inaccurate point estimates (14, 15).

In this article we demonstrate that, perhaps paradoxically, simulation models can support effective selection of COVID-19 interventions even when they are unable to provide accurate point estimates of epidemic outcomes. (The use of epidemic models in the presence of significant parameter uncertainty is also discussed in, e.g., (16). In such settings, clear communication of uncertainties is key; see, e.g., (17).) We demonstrate this in the specific context of deciding whether Cornell University's Ithaca campus should reopen for the fall 2020 semester. In this context, we conducted a simulation-based analysis in summer 2020 using a compartmental SEIR model with multiple subpopulations; see (7, 18, 19) for closely related models. Our work contributed heavily to the decision to reopen the campus for residential instruction (12) and was used in myriad other ways including the choice of screening frequencies within an asymptomatic screening program that was a critical part of Cornell's strategy, and sizing of quarantine capacity.

Our approach hinges on delineating those simulation model input parameters yielding epidemics that can be successfully controlled from those that cannot. If the set of plausible input parameters are contained within the set of safe parameters, then we can be highly confident, though never certain, that the epidemic can be controlled. At Cornell in summer 2020, we demonstrated this to the university administration for a suite of interventions available with in-person instruction: frequent asymptomatic screening, testing students on arrival to campus, contact tracing, social distancing on campus, limits on student and employee travel, masking requirements, and a behavioral compact curtailing student social gatherings. It was also possible that we would have found that plausible ranges of the input parameters overlapped the portion of parameter space where epidemics would grow out of control, in which case we would not have been able to recommend reopening.

We found that access to regular asymptomatic screening (9, 20), with an ability to increase testing frequency if needed, was critical. Asymptomatic surveillance was enabled at Cornell through a major effort to support large-scale sample collection, to retool a veterinary lab to enable high-volume PCR testing, and the use of pooled testing. Our analysis and interventions were customized to Cornell's environment, but the results would likely generalize to other universities and other settings. Indeed, those few universities employing a similar asymptomatic screening approach succeeded by and large in controlling campus outbreaks (21–24). See also (25–30) for explorations of the interaction of pooled testing and asymptomatic surveillance for controlling epidemics.

We also found it was critical to analyze epidemic growth if in-person instruction were not offered, to quantify the relative merits of the alternative to in-person instruction. Survey results (12, 31) suggested that a significant number of students would return to the Ithaca area even if in-person instruction were not offered. Without the benefits of the legal framework offered by in-person instruction, frequent asymptomatic screening would have been difficult to mandate for this population. Moreover, our analysis suggested that many of those parameter settings in which asymptomatic screening would not ensure safe in-person instruction would also be ones in which a significant outbreak would occur in the local student population under virtual instruction. This resulted in the decision to reopen Cornell's Ithaca campus (12).

The purpose of this paper is to explain how epidemic modeling, coupled with careful analysis of input parameters, can provide key insights to those deciding how to keep campuses and their surrounding communities safe. More broadly, these insights are of value to those using epidemic models to support decisions in other contexts during the COVID-19 pandemic and future epidemics. We view the key contributions of this paper to be as follows.

- We provide a modeling and decision-making framework for helping universities decide whether and how to open. Central is a careful accounting for uncertainty. A recommendation to reopen relies on plausible parameter ranges falling within a portion of the input parameter space corresponding to epidemics that can be controlled with the intended interventions. The same methods apply to modifying interventions midway through a semester.

- Central to this question is: what is the alternative? At Cornell, an analysis suggesting health risks would be higher if the campus opened for online-only instruction led to the decision to reopen for fall in-person instruction.

- We highlight the value of asymptomatic surveillance, which in addition to enabling early isolation of positive cases, also provides near-real-time information on infection levels. This permits recourse decisions such as increased testing frequency, stricter social-distancing, or in an extreme case, moving to all-online instruction.

- Our modeling structure can be, and was, employed to provide insight on a host of additional decisions, e.g., what frequency should be used to test each campus subpopulation, what is the needed quarantine capacity, and after the end of the semester should asymptomatic surveillance continue for the remaining campus population.

- We evaluate this framework by providing a retrospective analysis of what happened when Cornell reopened and revisit which aspects of our model were accurate, which aspects were not, and what consequences this had for the quality of the decisions that were made.

- We highlight the impact of sensitivity analyses that showed the efficacy of contact tracing. This led to the development and use at Cornell of *adaptive testing*, i.e., testing all individuals in similar social circles as a positive case, and not just those identified as close contacts.

- We explore those qualities of Cornell and the surrounding Ithaca community that enabled a successful fall semester.

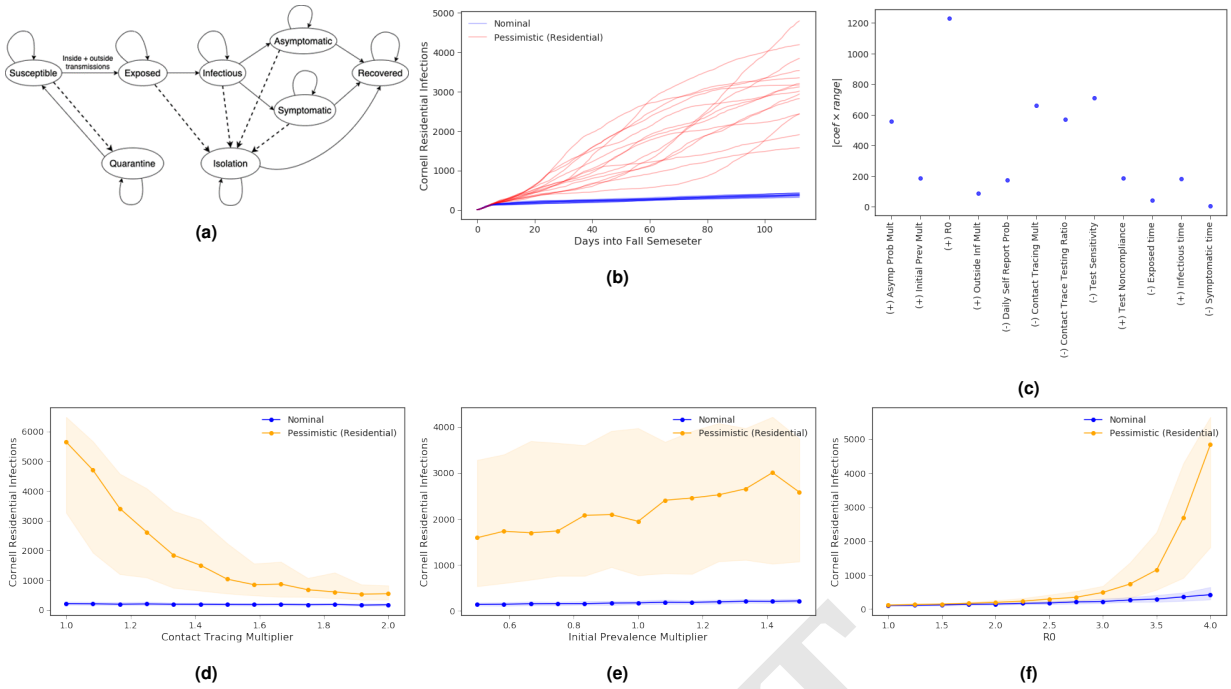


Fig. 1. (a) The dynamics of our compartmental simulation across compartment categories (ellipses). Population counts are maintained for each compartment on each day, and compartments comprise a category, e.g., "susceptible", a demographic group, e.g., undergraduates in high-density housing, and the elapsed time in that compartment category. Solid lines represent virus transmission, disease progression, and the end of quarantine/isolation. Dashed lines represent the effect of testing, self-reported symptoms and contact-tracing, which put individuals testing positive into isolation and their contacts into quarantine. (b) Sample cumulative infection counts over time, under the nominal and pessimistic (for residential infections) parameter scenarios. (c) The first-order effect of parameter uncertainty on predicted infections, using a linear model to estimate sensitivity of infections to each parameter and a range of plausible parameter values. Each dot shows the estimated effect of uncertainty on predicted infections (absolute value of the regression coefficient times the uncertainty range's width) with the sign of the regression coefficient indicated in the label. The uncertainty of this estimated effect (derived from the regression coefficient's 95% confidence interval) is approximately 60 for all parameters. (d-f) Plots depict the number of infections (lines provide the median; shading indicates the 10-90th percentile range across simulation replications) as three key parameters vary while holding the others fixed, under nominal and pessimistic scenarios.

Our work adds to the broader literature using epidemic modeling in the context of universities. See, e.g., (32) for a perspective on the challenges of reopening as informed by a variety of epidemic models, (33, 34) for the use of agent-based modeling to evaluate mitigation strategies to enable safe in-person instruction, (35) for probabilistic modeling of strategies to suppress virus spread in dorms and classrooms, and (36) for a study of interventions for generic small residential campuses.

1. Results

A. Model Predictions and Parameter Sensitivity. We focus on Cornell's main campus, located in Ithaca, New York (37). Ithaca is located in Tompkins County with a population of approximately 102,000. Approximately 12,000 undergraduates (of which 35% live on campus), 6,800 graduate students, and 10,000 employees study or work on the Ithaca campus. Ithaca is 4-5 hours by car from major city centers such as New York, Boston, Philadelphia, and Toronto.

In June 2020 we developed a compartmental simulation model to predict infections and hospitalizations for Cornell's fall 2020 semester (Fig 1a). The model contains compartments by subpopulation, stages of symptom development and infectivity, and quarantine / isolation status (Methods A-C). Fig 1b illustrates the model's output under two different sets of input parameters, simulating a surveillance testing program similar to the one Cornell used in fall 2020. In the simulation results pictured, all undergraduates and student-facing employees were tested twice a week, grad students and non-student facing employees once a week, and off-campus employees monthly. Infections in the Cornell population are roughly proportional to other outcomes of importance (SI 3.E) and therefore figures here focus on this outcome. Since our model is stochastic, there is variance in case trajectories with fixed parameters.

A core challenge was significant uncertainty about underlying input parameters, given that critical epidemiological quantities had not been measured accurately in college populations for SARS-CoV-2 in June 2020. To represent this uncertainty, we identified ranges containing plausible values for each simulation parameter (SI 3.A) based on information available in June 2020.

To understand sensitivity of model predictions to uncertain parameters, we then evaluated residential infections at 2000 parameter configurations chosen using a Latin Hypercube design (38) over the collective set of parameters defined by these plausible ranges (while also supporting a sensitivity analysis over four additional parameters used for predicting outcomes under virtual instruction, SI 3.D). We fit a linear model to quantify the first-order impact of each parameter on infections. We then multiply the estimated rate (from this linear model) at which residential infections change as we vary a parameter by the width of the range quantifying uncertainty for this parameter (Fig 1c and SI Table S17). To first order, this value is the change in predicted infections resulting from moving the parameter from the lower to the upper bound of its corresponding range. Parameters for which this value is largest are those that lead to the greatest uncertainty about infections,

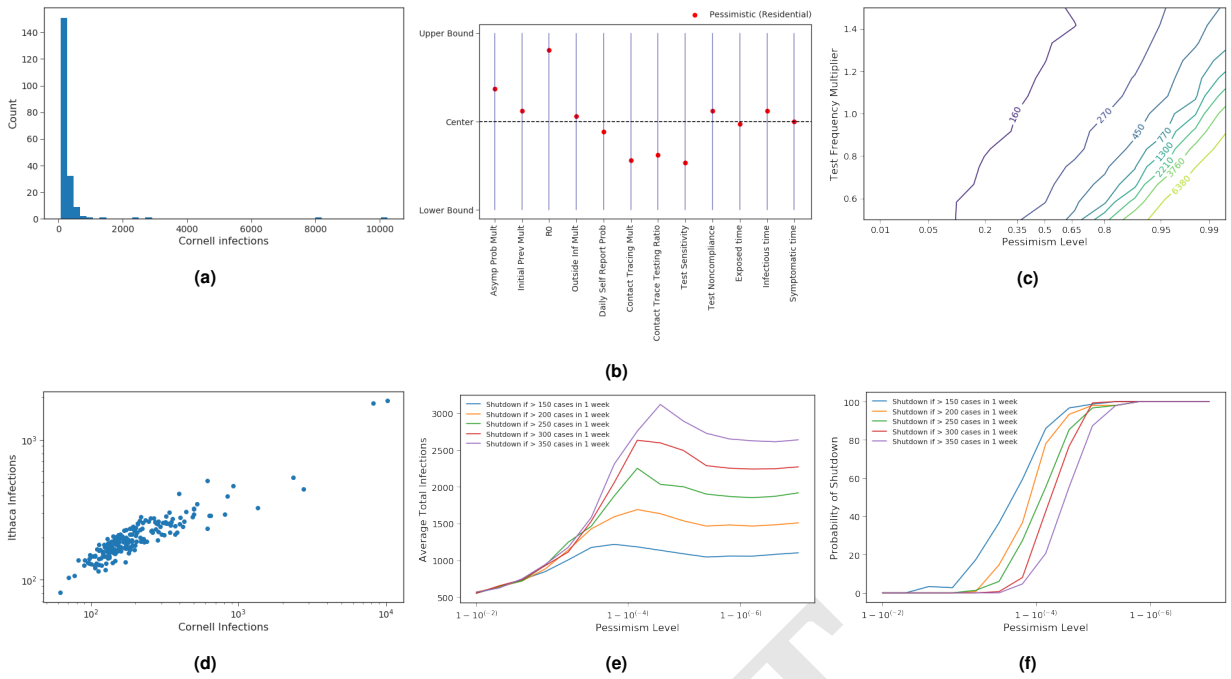


Fig. 2. (a) Histogram of median Cornell infections when parameters are sampled from the prior. (b) Parameter values of the pessimistic (for residential infections) scenario when the range of each parameter has been normalized. (c) Contour plot showing the number of Cornell infections as test frequency and the level of pessimism changes. (d) Scatter plot showing the median number of Cornell and Ithaca infections for each of the 200 points sampled from the prior. (e) The average number of infections and (f) probability of shutdown under each of a set of shutdown policies that model the more nuanced decision-making process available to leaders in practice.

whether because our uncertainty about the parameter is large or because outcomes are sensitive to it.

The effect of uncertainty is substantial, with uncertainty about several individual parameters creating uncertainties of more than 500 infections relative to a baseline of approximately 250 infections.The parameters that most drive uncertainty about infections are those that influence 1) transmission of the virus, especially R_0 as well as two contact tracing parameters and a parameter governing the likelihood an infected individual develops symptoms, and 2) our ability to control it through testing (test sensitivity). These parameters are described in detail in the SI's Section 1.

To understand nonlinear dependence on parameters and interactions across parameters, we additionally plotted outcomes as we vary each parameter individually while holding the other parameters fixed. Fig 1d-f shows three of the parameters for which uncertainty has the largest effect on outcomes, while others are shown in the SI. In each plot, the two lines correspond to two different parameter scenarios described in detail in the next section.

B. Coping with Parameter Uncertainty. Uncertainty about parameters presented the central challenge when deciding whether it would be safe to bring students back to campus. As demonstrated through sensitivity analysis (Results A), simulation outcomes are sensitive to parameters that were unknown in June 2020. This prevented accurate point estimates for the number of infections that would result from a particular set of interventions. A central theme of this article, however, is that accurate point estimates are not a prerequisite for supporting good decisions through modeling.

We show that asymptomatic screening is a powerful measure that can be brought to bear against this challenge of parameter uncertainty. Any fixed test frequency prevents widespread epidemic growth over a set of parameter settings. A larger test frequency results in a larger set of safe parameter settings.

To understand whether a candidate value of 2x / week asymptomatic screening would be enough to make residential instruction safe under plausible parameter values, we formed a Bayesian prior over parameters consisting of independent normal distributions for each parameter. The (marginal) mean and variance of the prior over each parameter was chosen so that the resulting symmetric 95% Bayesian credible interval corresponded to the previously-selected plausible ranges for each parameter (Results A, SI 3.A). The nominal scenario consists of setting each parameter to its mean.

We then drew random sets of parameters from this Bayesian prior and ran our simulation for each, to form a prior distribution over infections accounting for parameter uncertainty (Fig 2a). In most parameter settings, 2x / week testing is sufficient to achieve substantial infection control but there are some parameter settings where large outbreaks occur.

To better understand the impact of interventions like testing on robustness to parameter uncertainty, we developed a one-dimensional family of parameter configurations with varying levels of pessimism about the number of infections (Methods D, SI 3.A). This family of parameter configurations is indexed by a pessimism level between 0 and 1, with larger levels corresponding to parameter configurations with more infections. The parameter configuration at pessimism level q is the most likely configuration under the prior for which median infections is equal to the q-quantile of infections under the prior, assuming that infections for a given parameter configuration are given by the previously fitted linear model. The nominal scenario corresponds to pessimism level 0.5. We

additionally define the "pessimistic scenario" as the parameter configuration corresponding to pessimism level 0.99, indicating it as "pessimistic (residential)" in plots to distinguish it from a scenario that is pessimistic about a different outcome defined in Results C. Fig 2b summarizes how the pessimistic and nominal scenarios differ. Relative to the nominal scenario, the pessimistic scenario significantly increases the asymptomatic ratio and transmission rate while decreasing test sensitivity and contact tracing effectiveness. These parameters also have the largest absolute normalized effect on infections (Fig 1c).

We then plotted infections using 2x / week testing at parameter configurations across a range of pessimism levels (Fig 2c at test frequency multiplier = 1.0; Fig 3c). This level of testing is sufficient to keep the number of infections below 1000 in all but the most pessimistic parameter configurations. Still, there are some parameter configurations where more than 1000 infections arise, and infections grow rapidly as the parameter configuration grows more pessimistic.

We hypothesized that additional testing could help mitigate this risk. We performed simulation experiments varying the testing frequency and the parameter configuration's pessimism level (Fig 2c). When the test frequency is low, a pessimistic configuration results in many infections (e.g., more than 6K infections at 1x / week testing at a pessimism level of .95). When the test frequency is high enough, however, predicted infections remain low even under such pessimistic scenarios.

Plots of infections as a function of test frequency and simulation parameters were distributed in public reports (39), at several Faculty Senate meetings (40), and in other open fora. These reports also included nominal and pessimistic scenarios (SI 3.B) similar to the ones detailed here, though highlighting more concerning outcomes. They were an important component of deliberation at Cornell on whether testing could allow a safe residential reopening. They resulted in the decision, at Cornell, that if the campus were to open for residential instruction, we would use a test frequency that was as large as could be provided reliably, to maximize the range of parameter settings under which Cornell's strategy would provide effective infection control.

While we focus here on a single outcome, infections in the campus population, other outcomes are important: infections in the surrounding community created by clusters in the campus population; and hospitalization and deaths. As shown in Fig 2d and SI 3.E these outcomes tend to move together when varying parameters and the overall frequency of testing. Thus, for the purposes of understanding the overall level of risk and deciding whether to reopen, considering only on-campus infections tends to produce the same decisions as would more holistic consideration of on-campus and community outcomes examining infections, hospitalizations and deaths.

In addition to testing, which was instituted at the start of the semester, a second measure combating parameter uncertainty is *recourse*: the opportunity to change decisions based on data as it is observed. Among the decisions that can be made in this way, smaller interventions include changes to behavioral interventions and modifications to test frequency. A few large interventions are also available: shutting the campus down, taking strong temporary measures to reduce social contact, and even sending students home. The effectiveness of the above-mentioned temporary measures was of primary interest in supporting the decision of whether to reopen Cornell's campus. Indeed, a reopening decision that was *likely* to be safe but not guaranteed to be so under a fixed set of interventions (testing frequency) can be made more safe by developing a plan up front. Whether such safety options exist in the event that estimates are wrong is an important component of deciding whether to bring students back at all.

Modeling decision-makers' ability to halt in-person instruction and shut down the campus upon observing substantial transmission, Fig 2e-f shows outcomes under plans of the form "initiate campus shutdown if confirmed cases exceed more than Y cases in X weeks" as a function of the location on the nominal-pessimistic line. While simpler than the more nuanced decision-making process available to leaders in practice, these plans take a simple interpretable form to support broad understanding, and are reminiscent of CUSUM charts used for monitoring industrial processes for defects (41). They essentially "learn" whether reality is such that infection control is not being provided, and then respond when a threshold is reached. By varying X and Y, one achieves different tradeoffs between acting quickly and "false alarms" in which the campus would be shut down when infection control would have been achieved if it had stayed open. Fig 2e-f demonstrates that adding recourse through shutting down can reduce the risk associated with reopening: in the unlikely event that transmission is high enough that the chosen test frequency does not prevent spread, then a prompt shutdown can mitigate the negative health consequences.

C. Virtual vs. Residential Instruction. While substantial focus was given to what would happen if universities reopened for residential instruction in the late summer of 2020 (32–36), an equally important consideration is what would happen if they did *not*. If a university chooses to offer only virtual instruction in place of in-person classes, many students may still elect to return to the university's local area. This was the case at Cornell, where a significant number of students had signed leases with local landlords for the fall before the severity of the pandemic became clear, and where a survey of students revealed a significant fraction were planning to return to Ithaca even if residential instruction and on-campus housing were not offered (12, 31).

For universities like Cornell located in college towns, where the student population represents a significant fraction of the overall population during normal semesters, this influx of students may represent a significant increase in the number of young people in the area. This could be dangerous because these young people may be socially cohesive and, during normal times, young people have elevated rates of social contact (42, 43). Moreover, a university may have reduced ability to mandate and enforce behavioral restrictions and asymptomatic screening for students taking classes virtually even if they are currently in the university's local area.

Thus, when deciding whether to reopen, in addition to whether the number of infections can be kept reliably low during a residential semester, an additional key consideration is the risk of an outbreak among virtual instruction students and from there to the community. In other words, a key tradeoff is whether to invite back all students and have stronger behavioral and screening interventions, or to have a smaller number of students return but have weaker interventions.

To study this tradeoff, we extended our model to capture virtual instruction at Cornell's Ithaca campus. Under virtual

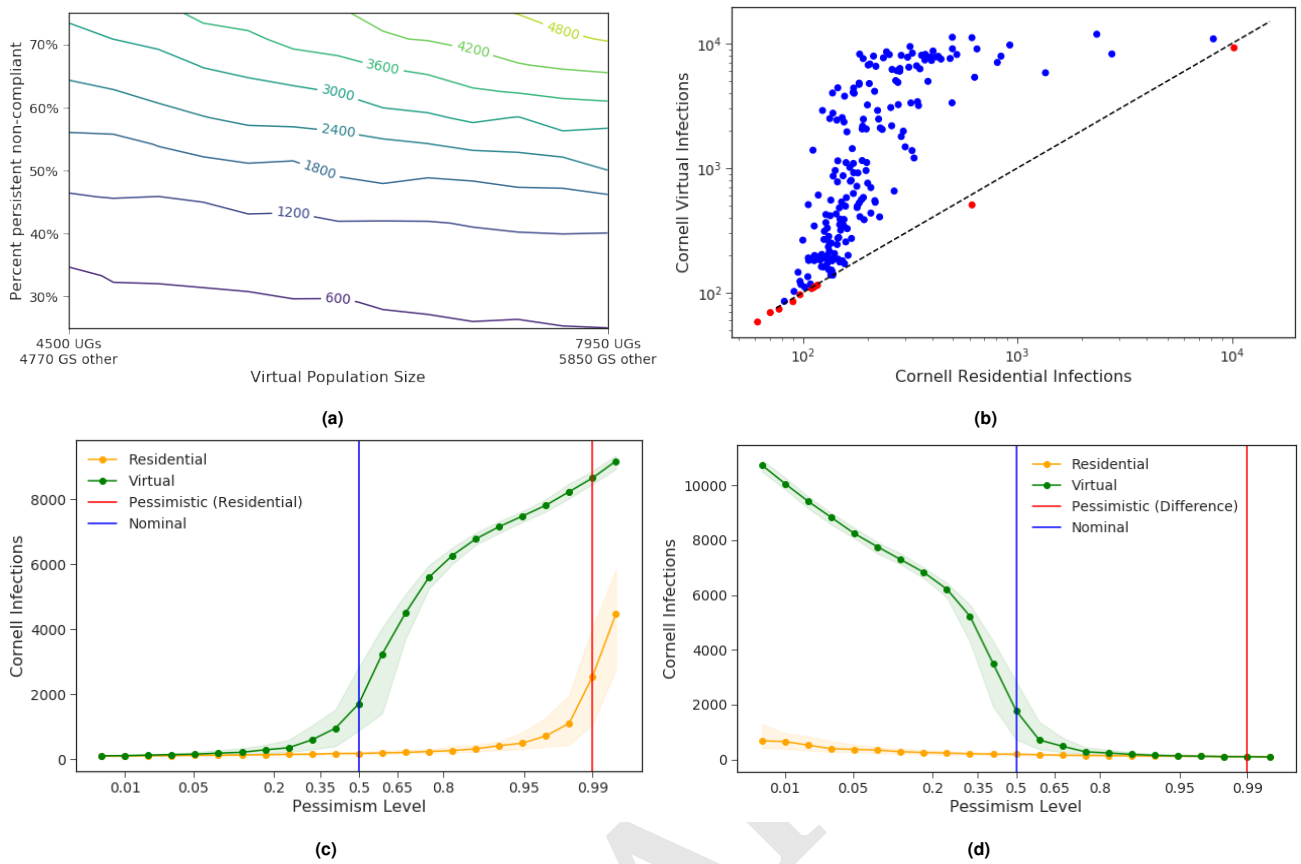


Fig. 3. (a) The number of infections in the Cornell community under virtual instruction under the nominal scenario, varying the fraction of off-campus students engaged in virtual instruction that do not use the offered testing ("persistent non-compliance") and the total number of off-campus students. If off-campus students' willingness to comply with 2x / week optional testing is not sufficiently high, a large number of infections result. (b) Under 200 parameter configurations drawn from the prior, the number of infections under virtual and residential instruction. Infections are smaller under residential instruction than under virtual instruction in most parameter configurations (blue dots), and when they are not (red dots) they are not substantially larger (c-d) Number of infections under virtual and residential instruction under a range of parameter configurations with different levels of pessimism. (c) uses a pessimistic configuration that maximizes residential infections while (d) maximizes residential - virtual infections.

instruction, staff and faculty along with some research-focused graduate students stay on campus. They are tested 2x / week and are subject to the same behavioral compact governing student behavior under residential instruction. Based on survey results and discussions with local landlords, we also model some other students returning to Ithaca to live while taking classes virtually, outside of the control of the university. We assume that Cornell offers twice-weekly testing to these students, but noncompliance is higher than in a residential semester.

This extended model included four additional parameters, about which we also had uncertainty (SI Table S16). We thus extended our parameter uncertainty framework to understand the range of outcomes possible under virtual instruction. We first generated ranges for these additional parameters based on information available in June 2020 (SI 3.A), extending our prior probability distribution to include independent normal priors for these new parameters. We also extend our nominal scenario for residential instruction to set each of these additional parameters to the center of its range.

In this extended model, reduced population density lowers transmission for unmonitored off-campus students relative to off-campus students during residential instruction, but lower compliance with social distancing and masks (which Cornell cannot mandate to these students) raises transmission. These effects can offset each other, or one can be more dominant, depending on parameters. Reduced use of now-optional testing reduces the benefits of this intervention. Fig 3a shows that reduced use of testing by the off-campus population can lead to a substantial number of infections, while outcomes are less sensitive to the number of off-campus students.

To understand the relative safety of residential and virtual instruction under plausible parameter configurations, we sampled 200 parameter configurations from the prior. Fig 3b then plots the number of infections under residential and virtual instruction for each of these configurations. We see that infections are fewer under residential instruction in almost all parameter configurations. In those parameter configurations where there are more infections under residential instruction, the number of additional infections is small. This suggests that residential instruction is a safer strategy than virtual instruction, given the information available in June 2020.

Exploring further, we extended each of the 2000 parameter configurations used for our residential infection sensitivity analysis (Results B) to include the four additional virtual-instruction parameters (SI 3.D) and used simulation to predict virtual instruction infections (in the on-campus students and employees and the off-campus virtual instruction students in Ithaca) under each parameter configuration, enabling a

comparison with predicted residential instruction infections for each configuration.

We then set out to identify a new collection of parameter configurations of varying pessimism about the *relative* safety of residential instruction compared to virtual instruction. To do so, we adopt the same approach used to identify configurations of varying pessimism about residential infections, but taking our primary outcome as the difference in infections between residential instruction and virtual instruction (a positive value indicates residential has more infections) (SI 3.A). The nominal scenario remains the same and corresponds to pessimism level 0.5. We obtain a new pessimistic scenario, corresponding to pessimism level 0.99. Unlike the pessimistic scenario for residential infections, this new pessimistic scenario (SI Table S17, Fig S7) decreases R_0 relative to nominal. This is because the most likely parameter configurations with small residential - virtual infections (according to the fitted linear model) are those in which transmission is small regardless of instruction method.

Fig 3c-d plots residential and virtual instruction infections for the two families of parameter configurations, one varying our pessimism about residential infections (Fig 3c), and the other varying our pessimism about residential - virtual infections (Fig 3d). In almost all scenarios, virtual infections are larger than residential infections, and are sometimes much larger. Those few scenarios where residential infections are larger have few infections in both modes of instruction.

Thus, modeling suggests that virtual instruction presented a substantial risk of high infection counts in the student population and the surrounding community, while residential instruction would result in lower infection counts under a broad range of the most reasonable parameter settings. This was a primary basis for Cornell's decision to reopen for residential instruction (12).

D. Design of Asymptomatic Testing Protocol. Our modeling approach was also an important tool for supporting detailed design of Cornell's surveillance testing strategy.

A first key question was the testing frequency for students and employees. Operational constraints limited the total number of tests that could be done per day. We hypothesized that targeting more frequent testing to those groups likely to have higher rates of transmission would provide more robust infection control within this constraint.

We enumerated testing policies consisting of a testing frequency (1x or 2x per week) for the 6 groups spending significant time on campus (e.g., undergraduates living on campus) under both the pessimistic (for the residential infections outcome) and nominal scenarios, producing 64 testing policies. We then discarded those policies not on the Pareto frontier under the pessimistic scenario. Fig 4a shows the resulting Pareto frontier and highlights the policy that was selected.

A second key question was the sampling methodology for surveillance testing. We considered anterior nares (AN) and nasopharyngeal (NP) sampling. While NP is more sensitive (44), it is also less comfortable and we hypothesized that this discomfort might lower test compliance. Fig 4b, generated using our simulation under the pessimistic (residential) scenario, evaluates this trade-off between test sensitivity and test compliance. AN was chosen for Cornell's surveillance sampling methodology in part based on this analysis, because the risk of a substantial loss in test compliance caused by NP sampling would not be known until after the launch of the program, while the test sensitivity of AN was measured before launch and was known to be sufficient for robust infection control.

E. A Retrospective View. Students returned to Cornell's Ithaca campus for residential instruction in the fall of 2020 based in large part on the results of the analysis above (12). An asymptomatic screening program using AN samples and pooled PCR testing was implemented as recommended with undergraduates being tested twice per week, graduate students once per week, and staff and faculty at a frequency between twice per week and once every two weeks depending on how often they were on campus. In addition, students were tested on arrival using NP swabs. Of the student population, 75% returned to the Ithaca campus (45).

This provides an opportunity to evaluate with hindsight the modeling described above. Using de-identified aggregated data obtained from fall surveillance, we build and calibrate a model with a revised set of groups based on risk separation observed during the semester.

In the revised model we consider employees and students to be two completely separated groups that do not infect each other, because transmission between students and non-student employees was observed through contact tracing to be extremely rare. We separate students into three groups: undergraduates who are in social Greek-life organizations or on varsity athletic teams; other undergraduates; and graduate / professional students. During the semester, observed cases and contact tracing suggested that the greatest transmission was associated with off-campus social activity and co-habitation, especially in the Greek-life and athletics community. Our modeling in the summer of 2020 separated students by the density of their housing, but we did not see substantial transmission associated with residence halls, suggesting that this distinction could be removed.

We then used observed data to estimate key model parameters: one governing the effectiveness of contact tracing for students and employees; the rate of transmission between the 3 student groups up to a proportionality constant; and the rate of infection from outside sources for students and employees.

After, one free parameter remained for employees (transmission rate) and one for students (a proportionality constant giving the transmission rate between each student group). We estimated these parameters, separately for employees and for students, by calibrating simulation results to observed data. More precisely, for each population we varied the free parameter and calculated the sum of squared differences between the mean of the model's predictions and the observed cumulative infection count (Fig 5a-b). Fig 5c compares observed total (student + employee) cases to simulation trajectories from the calibrated model. For details see SI Section 2.

Fig 5d-f shows results from our retrospective evaluation using this calibrated model, focusing on 2 measures of quality: consistency with the range of plausible scenarios identified earlier; and quality in hindsight of the decision made.

First, we evaluate whether the calibrated model is consistent with the range of likely scenarios identified in Results B. A comparison of calibrated parameters (SI Table S20) shows that the calibrated parameters are consistent with our prior. Fig 5d compares observed infections to a range of scenarios. Observed infections were close to the nominal scenario and well

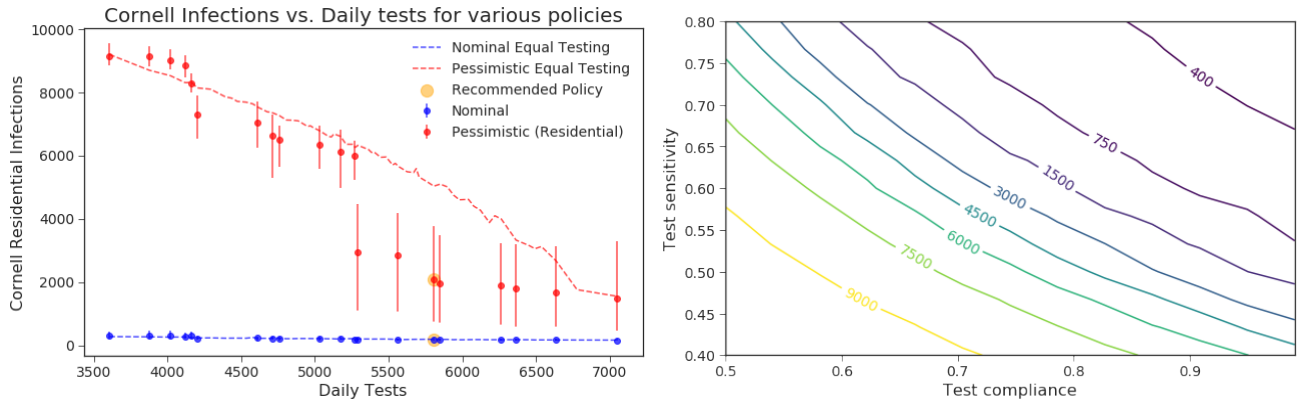


Fig. 4. (a) Select Pareto-efficient testing policies (frequencies for each group) according to median simulated infections. Vertical bars depict the distribution of infections, ranging from the 10% to the 90% quantile over simulation outcomes. The point highlighted in yellow corresponds to the testing frequencies Cornell selected for the fall semester. Policies shown are Pareto-efficient policies from the set of policies where each on-campus group is tested either once or twice a week. The dotted lines are simulation estimates of the expected number of infections if tests are split homogeneously among the on-campus population. (b) Contour plot showing the number of infections in the Cornell community as test compliance and test sensitivity vary under the pessimistic (residential) scenario.

under the pessimistic scenario for almost the entire semester, consistent with their design.

While the observed values are quite close to the nominal scenario's predictions, this was accidental. Indeed, it was clear *a priori* that model outputs are sensitive to inputs and these inputs were unknown, so the predictive accuracy of the range across scenarios, in the sense of whether it contained reality or not, is more important than whether a point prediction was close to the realized trajectory.

Fig 5d also includes predictions from the nominal scenario in our June 2020 report (46). This scenario was nearly identical to the nominal scenario reported here, except that three key parameters were set to conservative values given the urgent need to generate recommendations without enough time to identify a plausible range for these parameters (SI 3.B). This reduced the accuracy of this particular point prediction by making it much more conservative, emphasizing the importance of basing decisions based on a range of plausible scenarios rather than a single point prediction.

Second, we study the quality of the decisions made, relative to potentially better decisions that could have been made with the benefit of hindsight. We focus on two key decisions: the design of the testing policy and whether to reopen campus.

Fig 5e shows the expected number and range of Cornell infections under Pareto-optimal testing policies, testing each of the 3 student groups from the retrospective analysis at either 1x or 2x per week, and testing employees at a frequency averaged across those used in practice. Retrospectively, we selected one of many testing policies resulting in few infections, but we did not select the most efficient. The most efficient testing policies require segmenting the riskiest undergraduates (who are in Greek-life organizations or on varsity athletic teams), which we did not recognize at the beginning of the semester. The figure also shows that there would have been limited benefit in increasing the total testing capacity.

Turning to the question of whether to reopen campus, Fig 5f shows the expected Cornell infections for a *virtual* fall semester under our calibrated model as the number of returning undergraduate students and their test compliance varies. More than 80% of the returning students would need to remain test-compliant throughout the entire semester to achieve a number of infections comparable to reopening campus (where fewer than 250 occurred). As discussed previously, enforcement of test compliance would have been significantly more challenging in a virtual scenario and therefore, the decision to reopen campus was robust.

We conclude that while the benefit of hindsight would have allowed us a modest gain in the efficiency of testing, the decision we selected gave similar health outcomes to the best policies with the benefit of hindsight.

2. Discussion

A. Use at Cornell. The primary purpose of the epidemiological modeling effort at Cornell was to answer the question: is there a way to safely reopen the campus for in-person instruction? Paralleling this question was the accompanying concern: what are the comparative impacts of the "best" implementation of in-person instruction to the "best" implementation of virtual instruction? The previous section outlined the analysis derived from our models in answering these questions. Specifically, the model provided the basis for determining the overall testing frequency required to limit spread, as well as the refined policy of varying test frequency across university sub-populations.

Beyond this primary goal, the modeling framework has many secondary applications. Foremost among these is the ability to do recourse planning: having a baseline model and real-time information about the prevalence of the epidemic on campus enabled measures responding to current conditions. The most visible of these was the color-coded state of the campus, which ranged from the green ("new normal") to the red alert which called for careful closing of campus, i.e., ending in-person instruction and the emptying of the on-campus dormitories. One of the advantages of the model in its use at Cornell was that it enhanced communication with the community — as a tool for explaining the considerations in making decisions, and for explaining what might happen and why. For example, the limitations on campus behavior imposed by moving the campus alert code from green to yellow could

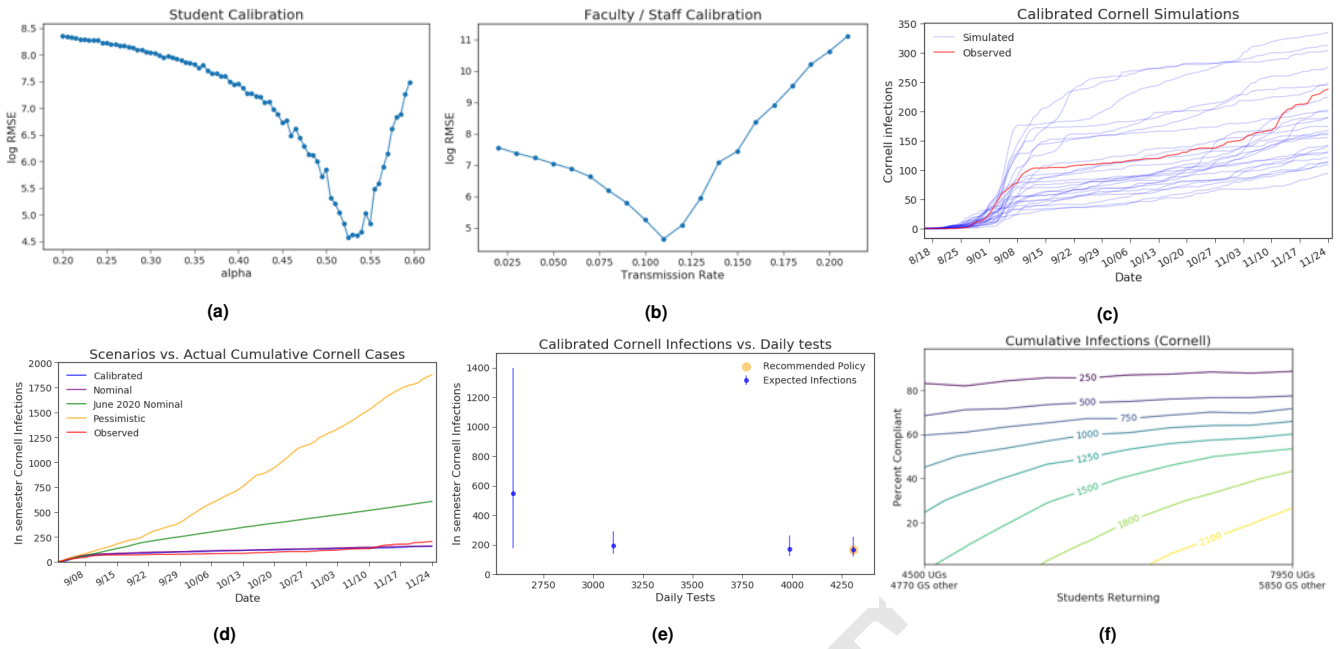


Fig. 5. (a) The log of the root mean-squared error (RMSE) of projected student infections versus α, the proportionality constant that multiplies the contact matrix to obtain the daily transmission rate. The fitted value of α is where RMSE is lowest, at $\alpha = 0.525$. (b) The log of the root mean-squared error (RMSE) of projected employee infections versus employee transmission rate. The fitted value of transmission rate that minimizes RMSE is 0.11. (c) Simulated cumulative case trajectories for all Cornell cases under calibrated parameters and observed fall 2020 cases, including the pre-semester period prior to the start of asymptomatic surveillance. (d) Cornell fall 2020 cases relative to the calibrated trajectory and to the nominal, pessimistic, and June 2020 nominal scenarios. (e) Re-creating the Pareto efficient testing frontier based on calibrated parameter values. (f) Using calibrated parameters to re-create the contour plot showing expected Cornell infections (including students) under a virtual instruction scenario as the number of returning students and their test compliance varies.

be modeled by changes in the parameters such as the number of daily contacts, and so the impact of this recourse approach could be demonstrated in a concrete way.

In addition to these strategic questions, the epidemiological modeling provided a framework for considering a broad range of tactical and operational policy questions. First, it was essential to estimate, and then plan for, quarantine and isolation capacity. In doing so, the model was used to understand both the requirements for the initial arrival period, and for the stochastic evolution of demand for these resources as the semester unfolded. Second, the model quantified the impact of shortening the time between collecting samples and isolating positive individuals, which supported the design of the testing infrastructure. Third, the Cornell academic calendar was restructured to limit impact of a potential "second wave": in-person education ended prior to Thanksgiving and the final three weeks of instruction were given virtually. Similar to the primary question of virtual vs. in-person education, the epidemiological modeling demonstrated the imperative of continued surveillance testing for the Cornell community residing in Ithaca beyond Thanksgiving.

B. Use at Other Universities. The *conclusions* of our Cornell study do not directly translate to other universities, but the key decisions informed by the model are broadly generalizable. Accordingly, our modeling framework can be readily adapted to support decision-making at other universities.

In translating this approach to other university campuses, a number of characteristics should be explicitly considered. For many of these, Cornell was particularly well-suited for in-person education. We highlight the following characteristics:

- the amount of interaction between the campus and surrounding communities, as well as the initial COVID-19 prevalence in the surrounding community; Ithaca had low prevalence and is several hours' drive from major cities;
- the extent to which campus culture provides "buy-in" for restrictions on student behavior; Cornell's administration worked closely with student leadership to create a student compact establishing behavioral norms;
- the extent to which restricting travel beyond the immediate environs is feasible; again, geography meant that travel beyond campus was generally not common, and restrictions were relatively easy to implement;
- the fraction of the campus community that will reside locally for both in-person and virtual modes; as discussed above, many students live in off-campus housing and survey data indicated a significant number would return even for virtual instruction;
- the ability to identify student cohorts with elevated social contact, and the social connectivity of those involved; case clusters at Cornell were often connected to off-campus social activity in a small segment of the student population and an understanding of this mechanism for viral spread supported contact tracing;
- the capacity of the institution to provide surveillance testing; the College of Veterinary Medicine at Cornell had experience using PCR to track viral infections in animal populations, which could be converted and rescaled to

provide the required capacity, especially since there was the expertise to further augment capacity through pooling;

- the ability to cope with the demands of arrival testing, both in terms of contact tracing and quarantine capacity (which depends on initial prevalence); in this aspect, Cornell students' willingness to self-quarantine in advance of their return appears to have limited initial incidence to levels below anticipated values.

In addition to these considerations, there are other local conditions relevant to the adaptation of our approach that are not specifically dependent on our modeling effort. For example, we believe that it is important to have appropriate enforceable quarantine and isolation space on campus.

Beyond universities, our modeling framework can be applied to other closed (or nearly closed) communities such as cruise ships, prisons, retirement homes, homeless shelters (47), military bases, or professional sports bubbles. While differences in the behavioral and health characteristics of the populations involved, as well as the interventions available, would require adjustments to model parameters, any of these relatively closed communities could be modeled and analyzed in a similar way, coping with parameter uncertainty as discussed here.

Like most decision-making based on epidemiological models, ours has a variety of limitations. Most importantly, our results are highly sensitive to input parameters for which there was significant uncertainty. Although our results showed that our conclusions were remarkably robust, our June 2020 nominal scenario was cautious in selecting conservative values for parameters based on literature and other relevant data. Cumulatively, this caused the predictions we released in the summer of 2020 to be conservative.

Beyond parameter uncertainty, our model itself has limitations due to its simple structure. For example, we assume that all members of a subgroup are homogeneous and therefore we do not model contact network structure. As a result, while it is similar to that used in other analysis (19), our model of contact tracing has limitations, e.g., we assume that all quarantined contacts are in the pre-infectious exposed state. Further, within a single compartment, we assume that infectiousness does not vary across individuals or over time, in contrast to, e.g., (9). Moreover, we model the impact of age and its effect on the distribution of infection severity in a discretized way. We do not model overdispersion in contacts (for example, one individual having many more contacts than another) because our prior over transmission parameters is based on empirically derived R_0 estimates. When infections are driven by spread among subpopulations with the highest transmission, empirical estimates of R_0 naturally correspond to transmission rates in these subpopulations.

3. Methods

Here we provide an overview of modeling methods. Full details appear in the SI.

A. Simulation Model: Overview. We model COVID-19 transmission using a compartmental model with Susceptible, Exposed, Infectious and Recovered compartments along with additional compartments to reflect specific characteristics of COVID-19 and the interventions applied (Fig 1a). To account for asymptomatic and pre-symptomatic transmission, the infectious phase is split into 3 compartments: Infectious, Asymptomatic and Symptomatic. Individuals formerly in the Exposed compartment enter the Infectious compartment before randomly being assigned to either the Asymptomatic or Symptomatic compartments. We also add compartments for quarantined non-infected individuals (Quarantine) and isolated infected individuals (Isolation). Due to significant age and social heterogeneity in a university community, we replicate the compartments described above for each of several university groups. The groups interact via cross-group contacts. In addition, since the severity of COVID-19 varies significantly with age, the probabilities of symptom severities are determined by a group's age distribution.

B. Simulation Model: Transmission. Modeled COVID-19 transmission is governed by two variables: contact rates between groups, represented by a contact matrix, and the probability of transmission during any interaction. The contact matrix is estimated using pre-pandemic measurements of age-based socialization patterns. The probability of transmission is calibrated to match external estimates of R_0 (2.5 for the nominal scenario). Transmission is modeled as stochastic, and the number of new infections in each group has a Poisson distribution with a mean determined by the product of the contact matrix and the transmission probability and the number of non-isolated infectious individuals in each group.

C. Simulation Model: Quarantine and Isolation. We model three mechanisms for identifying and quarantining / isolating infected individuals.

1. Testing: every day, a fraction of each group is selected uniformly at random to be tested and test results are available the same day. The number of individuals selected to be tested per group per day is determined by the group's testing frequency.

2. Symptomatic self-reporting: Every day, each symptomatic individual not in Isolation has a constant probability of self-reporting their symptoms, upon which they enter Isolation the same day.

3. Contact tracing: Each case found through testing or self-reporting is contact traced. (Cases found through contact tracing are not, themselves, modeled as contact traced. This is an approximation of reality that is necessary because our simulation tracks counts within groups, not individuals.) Each contact trace moves a Poisson-distributed number of people to Quarantine or Isolation after a 1-day delay. We assume that contact tracing only finds contacts within the same group as the source case, reflecting the social dynamics of college campuses.

D. Parameter Configurations of Varying Pessimism. To summarize the effect of parameter uncertainty on an outcome (infections in a residential semester, or the difference in infections between residential and virtual instruction), we developed a one-dimensional family of parameter configurations with varying levels of risk for each outcome. For each real number y we consider the set of parameter configurations $A(y)$ whose median outcome is equal to y according to the fitted linear model. (All configurations are in exactly one $A(y)$, so this partitions the parameter configuration space.) For each y, we select a

representative configuration from $A(y)$: the one that is most likely under the prior. As we increase y, outcomes under the representative configuration tend to degrade. We then graph the outcome under this configuration versus $P(\cup_{y' \leq y} A(y'))$, i.e., the probability under the prior of seeing a parameter configuration whose outcomes (under the linear model) are no worse than those in $A(y)$. We refer to this probability as the "pessimism level". It ranges from 0 to 1, with larger values corresponding to representative parameter configurations that are more pessimistic. For details see SI Section 3.A.

E. Retrospective Parameter Estimation and Model Calibration. We estimate most model parameters for the fall semester (initial prevalence, outside infection rate, a contact matrix with entries that are proportional to the inter-group transmission rates, and contact-tracing effectiveness) directly from fall semester data available in operationally-focused non-research public communication (48). These public health communications were based on de-identified positive-case, testing, and contact-tracing stored along with student life, housing, and employee data in a HIPAA-compliant database. This data was collected and analyzed pursuant to the urgent public health need presented by the pandemic. Additional aggregations of de-identified data for research purposes were provided by Cornell University and deemed by Cornell's IRB to not meet the regulatory definition of human subjects research.

Remaining parameters, i.e., the proportionality constant scaling the contact matrix for the student group and the rate of transmission for the employee group were estimated by calibrating the model output to the infection trajectories observed during the Fall 2020 semester. As already discussed, we consider employees and students to be two completely separate groups that do not infect each other. The calibration results are robust to the addition of a small but positive employee-student transmission term, although such a term can change outcomes more substantially in extreme regimes where student or employee prevalence is much higher than observed in the fall.

Acknowledgments

We gratefully acknowledge support from Cornell University. Additional support was provided by National Science Foundation grants TRIPODS+X DMS-1839346 and CMMI-2035086, and Air Force Office of Scientific Research FA9550-19-1-0283. We are also grateful to many individuals from Cornell University, the Tompkins County Health Department, and the Cayuga Health System who together kept our community safe.

1. E Black, R Ferdig, LA Thompson, K-12 virtual schooling, COVID-19, and student success. *JAMA Pediatr.* **175**, 119–120 (2021).
2. J Wallinga, P Teunis, M Kretzschmar, Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *Am. J. Epidemiol.* **164**, 936–944 (2006).
3. G Yamey, RP Walensky, Covid-19: re-opening universities is high risk. *BMJ* **370** (2020).
4. SM Danielle Ivory, Robert Gebeloff, Young people have less COVID-19 risk, but in college towns, deaths rose fast. *The New York Times* (2020) Accessed: 2021-03-12.
5. L Rennert, C Kalbaugh, C McMahan, L Shi, CC Colenda, The urgent need for phased university reopenings to mitigate the spread of COVID-19 and conserve institutional resources: A modeling study. *medRxiv* (2020).
6. L Rennert, CA Kalbaugh, L Shi, C McMahan, Modelling the impact of presemester testing on COVID-19 outbreaks in university campuses. *BMJ open* **10**, e042578 (2020).
7. E Brooks-Pollock, et al., High COVID-19 transmission potential associated with re-opening universities can be mitigated with layered interventions. *medRxiv* (2020).
8. A Hale, The behavioral compact, explained. *Cornell Dly. Sun* (2020) Accessed: 2021-04-11.
9. JT Chang, FW Crawford, EH Kaplan, Repeat SARS-CoV-2 testing models for residential college populations. *Heal. Care Manag. Sci.* **24**, 305–318 (2021).
10. CW Stubbs, M Springer, TS Thomas, The impacts of testing cadence, mode of instruction, and student density on fall 2020 COVID-19 rates on campus. *medRxiv* (2020).
11. College Crisis Initiative, The college crisis initiative. https://collegecrisis.shinyapps.io/dashboard/ (2020) Accessed: 2021-03-12.
12. MI Kotlikoff, ME Pollack, Why Cornell will reopen in the fall, https://www.wsj.com/articles/why-cornell-will-reopen-in-the-fall-11593535516. *Wall Str. J.* (2020) Accessed: 2021-04-04.
13. EH Bradley, MW An, E Fox, Reopening colleges during the coronavirus disease 2019 (covid-19) pandemic—one size does not fit all. *JAMA Netw. Open* **3**, e2017838–e2017838 (2020).
14. WC Roda, MB Varughese, D Han, MY Li, Why is it difficult to accurately predict the COVID-19 epidemic? *Infect. Dis. Model.* **5**, 271–281 (2020).
15. T Blackwell, COVID-19 fight has relied too much on uncertain math modelling, some scientists say, https://ottawacitizen.com/news/canada/covid-19-fight-has-relied-too-much-on-uncertain-math-modelling-some-scientists-say. *Ott. Citiz.* (2021) Accessed:2021-03-12.
16. I Holmdahl, C Buckee, Wrong but useful — what COVID-19 epidemiologic models can and cannot tell us. *New Engl. J. Medicine* **383**, 303–305 (2020).
17. S Eker, Validity and usefulness of COVID-19 models. *Humanit. Soc Sci Commun* **7** (2020).
18. EM Hill, BD Atkins, MJ Keeling, M Tildesley, L Dyson, Modelling SARS-CoV-2 transmission in a UK university setting. *medRxiv* (2020).
19. FC Motta, et al., Benefits of surveillance testing and quarantine in a sars-cov-2 vaccinated population of students on a university campus. *medRxiv* (2021).
20. AD Paltiel, A Zheng, RP Walensky, Assessment of SARS-CoV-2 screening strategies to permit the safe reopening of college campuses in the United States. *JAMA network open* **3**, e2016818–e2016818 (2020).
21. M Peplow, How one university built a COVID-19 screening system, https://cen.acs.org/biological-chemistry/infectious-disease/one-university-built-COVID-19/98/i42. *Chem. Eng. News* (2020) Accessed: 2021-04-04.
22. News @ Northeastern, Here's why Northeastern is testing everyone on the Boston campus for the coronavirus, https://news.northeastern.edu/coronavirus/university-messages/heres-why-northeastern-is-testing-everyone-on-the-boston-campus-for-the-coronavirus/ (2020) Accessed: 2021-04-04.
23. TN Denny, et al., 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. *Morb. Mortal. Wkly. Rep.* **69**, 1743 (2020).
24. MD Fox, DC Bailey, MD Seamon, ML Miranda, Response to a COVID-19 outbreak on a university campus—Indiana, August 2020. *Morb. Mortal. Wkly. Rep.* **70**, 118 (2021).
25. C Gollier, O Gossner, Group testing against COVID-19. *Covid Econ.* **2** (2020).
26. B Cleary, et al., Using viral load and epidemic dynamics to optimize pooled testing in resource constrained settings. *medRxiv* (2020).
27. Y Lin, et al., Group testing enables asymptomatic screening for COVID-19 mitigation: Feasibility and optimal pool size selection with dilution effects. *arXiv:2008.06642* (2020).
28. N Chen, M Hu, C Zhang, Capacitated SIR model with an application to COVID-19. *Available at SSRN 3692751* (2020).
29. M Müller, PM Derlet, C Mudry, G Aeppli, Testing of asymptomatic individuals for fast feedback-control of COVID-19 pandemic. *Phys. biology* **17**, 065007 (2020).
30. GD Lyng, NE Sheils, CJ Kennedy, D Griffin, EM Berke, Identifying optimal COVID-19 testing strategies for schools and businesses: Balancing testing frequency, individual test technology, and cost. *medRxiv* (2020).
31. Cornell University Committee on Teaching Reactivation Options, Committee on teaching reactivation options report, https://covid.cornell.edu/_assets/files/ctro-final-report.pdf (2020) Accessed: 2021-04-04.
32. N Ghaffarzadegan, LM Childs, UC TÃuber, Diverse computer simulation models provide unified lessons on university operation during a pandemic. *BioScience* (2020).
33. UK Mukherjee, et al., An evaluation of educational institutions reopening strategies for in-person classes safely amid the COVID-19 pandemic. *medRxiv* (2020).
34. PT Gressman, JR Peck, Simulating COVID-19 in a university environment. *Math. biosciences* **328**, 108436 (2020).
35. M Borowiak, et al., Controlling the spread of COVID-19 on college campuses. *arXiv:2008.07293* (2020).
36. R Bahl, et al., Modeling COVID-19 spread in small colleges. *arXiv:2008.09597* (2020).
37. Cornell University Real Estate, FAQ, https://realestate.fs.cornell.edu/about/faq.cfm#CP_JUMP_5682 (2021) Accessed: 2021-03-12.
38. MD McKay, RJ Beckman, WJ Conover, A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **42**, 55–61 (2000).
39. Cornell COVID-19 Modeling Team, Covid-19 Mathematical Modeling for Cornell's Fall Semester, https://covid.cornell.edu/_assets/files/covid_19_modeling_main_report.pdf (2020).
40. Cornell University Faculty Senate, June 24 Online Faculty Senate Meeting, https://theuniversityfaculty.cornell.edu/faculty-senate/archives-and-actions/archived-agenda-and-minutes/online-faculty-senate-june-24/ (2020).
41. SV Crowder, An introduction to the Bernoulli CUSUM in *2017 Annual Reliability and Maintainability Symposium (RAMS).* (IEEE), pp. 1–6 (2017).
42. J Wallinga, P Teunis, M Kretzschmar, Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *Am. journal epidemiology* **164**, 936–944 (2006).
43. J Mossong, et al., Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med* **5**, e74 (2008).
44. RA Lee, et al., Performance of saliva, oropharyngeal swabs, and nasal swabs for sars-cov-2 molecular detection: a systematic review and meta-analysis. *J. Clin. Microbiol.* **59**, e02881–20 (2021).
45. M Rosenberg, Three-quarters of enrolled students are zooming from Ithaca this fall, https://cornellsun.com/2020/11/08/three-quarters-of-enrolled-students-are-zooming-from-ithaca-this-fall/. *The Cornell*

Dly. Sun (2020) Accessed: 2021-04-04.
46. JM Cashore, et al., Covid-19 Mathematical Modeling for Cornell's Fall Semester (https://covid.cornell.edu/_assets/files/covid_19_modeling_main_report.pdf) (2020).
47. LA Chapman, et al., Comparison of infection control strategies to reduce COVID-19 outbreaks in homeless shelters in the United States: a simulation study. *medRxiv* (2020).
48. Cornell COVID-19 Modeling Team, Mathematical Modeling for Cornell's Spring Semester, https://covid.cornell.edu/_assets/files/general-audience-spring-modeling-20210216.pdf (2021).

PNAS
www.pnas.org

Supplementary Information for

Modeling for COVID-19 College Reopening Decisions: Cornell, A Case Study

P. I. Frazier, J. M. Cashore, N. Duan, S. G. Henderson, A. Janmohamed, B. Liu, D. B. Shmoys, J. Wan, and Y. Zhang

Corresponding Peter I. Frazier.
E-mail: pf98@cornell.edu

This PDF file includes:

Supporting Information Text

This appendix is split into three sections. In the first section, we describe our model and methodology for estimating its parameters. Papers we reference in this section are from the summer of 2020 since this is when we were estimating the parameters for the fall 2020 semester. The second section relates to calibration of our model in the retrospective study. The last section shows the sensitivity of our model to varying input parameters.

Portions of this appendix have been previously released as part of the communication of our public health work (1–4).

Code implementing the simulations described is available at https://github.com/peter-i-frazier/group-testing.

1. Model

Model Overview. We model the spread of COVID in the Cornell and surrounding Greater Ithaca community using a multi-group stochastic compartmental simulation model. Each group is modelled using a discrete-time Markov chain (DTMC) with the state described below. All these DTMCs are linked together by the transmission process.

- Number people in Susceptible
- Number people in Exposed with x days remaining until they become Infectious (ID) for x in $\{0, 1, ..., 7\}$
- Number people in Infectious with x days remaining until they become Symptomatic/Asymptomatic for x in $\{0, 1, ..., 8\}$
- Number people in Symptomatic with x days remaining until they recover for x in $\{0, 1, ..., 20\}$
- Number people in Asymptomatic with x days remaining until they recover for x in $\{0, 1, ..., 20\}$
- Number people in Recovered
- Number people in Quarantine
- Number people in Isolation
- Number people who will be contact traced in future days (allows us to account for contact tracing delay)

We only maintain counts of the aggregate number of people in each state, not the trajectories of each individual. We use the term 'free individuals' to refer to everyone not currently in Quarantine or Isolation. Similarly, we use 'free and infectious' individuals to refer to all free individuals that are Infectious, Symptomatic, or Asymptomatic.

Every day corresponds to 1 state transition of the DTMCs. The transition kernel reflects five key dynamics:

1. Natural disease progression of infected individuals
2. Surveillance testing
3. Symptomatic self-reporting
4. Contact tracing
5. Transmission and new infections

1. Natural Disease Progression. Figure S1 shows the compartments we use to model the progression of COVID. The probability that someone transitions from Infectious to Symptomatic depends on the age distribution of their group. Once someone has been infected, we assume that they cannot be re-infected.

The Isolation compartment is for isolated individuals who are infected and the Quarantine compartment is for quarantined individuals who are not infected. Once an infected person has been identified and isolated, they cannot create any new infections and leave quarantine/isolation after they are no longer contagious. Every day, each person in Quarantine or Isolation has a constant probability of being released (to Susceptible and Recovered respectively).

If a free individual is infected (Exposed, Infectious, Symptomatic or Asymptomatic) and not isolated, they transition from their current compartment with x days remaining to the same compartment with $x - 1$ days remaining. If there are no remaining days in their current compartment, they transition to the next compartment. At this time, the length of stay in their next compartment is realized and the state of the DTMC reflects this realization. Transitions from Susceptible to Exposed occur due to transmission events and at that time their length of stay in Exposed is realized.

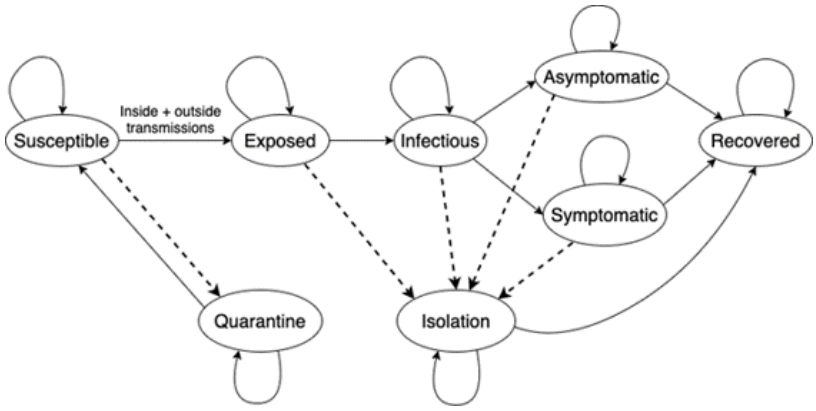


Fig. S1. Timeline of disease progression in an infected individual.

2. Surveillance testing. Every day a fraction of the group's free population is independently randomly selected. This fraction selected for testing can vary by group but is constant over the horizon of the simulation. We assume that people in compartments Infectious, Symptomatic and Asymptomatic are detectable by testing. Each test has a constant independent probability of producing an incorrect result (false positive or negative). False positives move people from Susceptible to Quarantine while false negatives do not change the state of the individual. True positives move people from an infectious state (Exposed, Infectious, Symptomatic or Asymptomatic) to Isolation. Test results are assumed to be available the same day. Each positive case identified through surveillance testing produces a contact trace.

3. Symptomatic self-reporting. Every day, each symptomatic individual has an independent, constant probability of self-reporting symptoms. Upon self-reporting, they are moved to Isolation and generate a contact trace. The probability of self-reporting every day is calibrated to data provided by the CDC.

4. Contact tracing. Each contact trace removes a random number of people from the free and infectious population and from the susceptible population. Symptomatic self-reports remove more susceptible and free and infectious people since these cases have likely been in the community longer than people identified via surveillance testing. We also assume a deterministic (1 day) delay between initiating a contact trace and isolating the contacts. The number of people and infectious cases removed is calibrated to Tompkins County contact-tracing data.

We do not contact trace positive cases found via contact tracing. Contact tracing only removes individuals in the same group as the source.

5. Transmission and new infections. We model two sources of new infections. The first is outside infections which refers to infections imported from interactions outside of Tompkins County. This is a daily rate per person estimated from travel-related Tompkins County cases.

The second source of new infections is local transmission due to free and infectious individuals. The rate of local spread is governed by two parameters: the contact rates between groups and the probability of transmission during an interaction. The contact rates are estimated using pre-pandemic contact surveys and account for age-varying compliance with wearing a mask and social distancing. The probability of transmission is calibrated to match the R0 of the disease.

Model Details. We first discuss the intra-group dynamics (disease progression, symptom severity, contact tracing, surveillance testing) followed by inter-group dynamics (transmission).

Intra-group Dynamics.

A. Individual Disease Progression. Our simulation assumes that the disease progresses through several stages in each infected individual, represented in Figure S1.

Parameters for the length of time in each state are given in Table S1.

Table S1. Parameters for disease progression in an individual.

Parameter Description	Nominal Parameter Value(s)	Sources
Time from exposure to infectious	Poisson(2) days	(5); (6); (7); (8)
Time from infectious to symptom onset	Poisson(3) days	
Time in symptomatic state	Poisson(12) days	(9)
P(self-report each day | asymptomatic)	0	Conservative assumption
P(self-report each day | symptomatic)	0.22	CDC planning scenario (10)

To justify the choice of time in the Exposed and Infectious states: (5) does a pooled analysis and finds the median incubation period to be 5.1 days, with a confidence interval of 4.5 to 5.8 days. (8) and (6) find that transmissions can occur 2-3 days before symptom onset. Thus we set the time in the Infectious state to be Poisson(3), and subtract its mean (3 days) from the incubation period mean to get a mean of 2 days for the exposed state.

In the simulation we model the time to self-report symptoms (for symptomatic patients) as being geometrically distributed with a single parameter that is the probability of self-reporting each day. This was chosen to match the average time from symptom onset to hospitalization for influenza-like illness (ILI) according to the CDC (10), which is based on (11). The latter paper reports that

- 35% of symptomatic individuals seek care in ≤ 2 days,
- 47% of symptomatic individuals seek care in $3-7$ days,
- 18% of symptomatic individuals seek care in ≥ 8 days.

We model this as a random number of days that is conditionally uniform(0,2) with probability 35%, conditionally uniform(3,7) with probability 47%, and conditionally uniform(8,12) with probability 18%. The resulting mean of this distribution is $.35 \times 1 + .47 \times 5 + .18 \times 10 = 4.5$ days. The daily probability of self-reporting for symptomatic individuals is then chosen to be $1/4.5 \approx 0.22$ so that the mean time to self-report, $1/0.22 = 4.545$, approximately matches this value.

B. Severity of Symptoms. Our simulation model separates symptomatic from asymptomatic individuals. Over the course of the simulation, symptomatic individuals self-report each day with some probability, while asymptomatic individuals do not self-report. Symptomatic infections can be of different levels of severity, ranging from mild pneumonia symptoms to critical life-threatening conditions. Thus we divide the symptomatic individuals into three different severity levels. In total, we consider four different severity levels, defined as follows:

- Severity level 1: patient is asymptomatic.
- Severity level 2: patient shows mild symptoms, but does not require hospitalization.
- Severity level 3: patient needs to be hospitalized, but does not require intensive care.
- Severity level 4: patient requires intensive care.

At the end of each simulated period, we allocate the symptomatic individuals to severity levels 2-4 with certain proportions. These proportions are estimated from data as explained below. Once an individual is assigned to a severity level they remain there; further transitions between severity levels are not modeled.

Let $P(\text{sev } i)$ be the probability that, as a result of a single contact with an infected person, an individual becomes infected and falls within severity level i. Thus the sum of these probabilities over $i = 1, 2, 3, 4$ is the probability of infection as a result of a single contact. Then, the probabilities that as a result of a single contact an individual becomes infected and asymptomatic, respectively infected and symptomatic, are

$$P(\text{asymptomatic}) = P(\text{sev } 1), \text{ and}$$
$$P(\text{symptomatic}) = P(\text{sev } 2) + P(\text{sev } 3) + P(\text{sev } 4).$$

We want to find $P(\text{sev } i)$ for the population while considering age-based factors. Specifically, we model how the severity of the disease varies with age, and that older age groups are more likely to become infected after an interaction with an infectious person. To that end,

$$P(\text{sev } i) = \sum_{\text{age}} P(\text{sev } i|\text{infected, age})P(\text{infected}|\text{age})P(\text{age}), \text{ where}$$
$$P(\text{infected}|\text{age}) = P(\text{infected}|\text{contact, age})P(\text{contact}|\text{age}) \propto P(\text{infected}|\text{contact, age}).$$

The proportionality in the second equation comes from the assumption of a homogeneous well-mixed population within each group. Therefore, the distribution of the age of contacts is the distribution of the age of the population in the group.

Severity Calculation Part 1: Severity and Infection given Age We obtain values for the probability of infection as a function of age from (12), which reports the probability of infection through a close contact for different age groups among 4941 close contacts traced from early cases in Guangzhou, China. These estimates are given in the first row of Table S2.

Later, we will estimate the age distribution ($P(\text{age})$) for Cornell's fall semester.

Table S2. Parameters for age-stratified infection probability and severity level distribution. Sources: (12–16).

	Age grp 1 (0-17)	Age grp 2 (18-44)	Age grp 3 (45-64)	Age grp 4 (65-74)	Age grp 5 (75+)
P(infection | age)	1.8%	2.2%	2.9%	4.2%	4.2%
P(sev 1 | infected,age)	17.0%	52.0%	31.0%	13.0%	13.0%
P(sev 2 | infected,age)	81.6%	47.2%	65.9%	80.6%	80.6%
P(sev 3 | infected,age)	1.1%	0.6%	2.2%	4.7%	4.7%
P(sev 4 | infected,age)	0.3%	0.2%	0.9%	1.7%	1.7%

The severity level distribution for each age stratum is estimated from a combination of data sources.

We first estimate $P(\text{sev 1}|\text{infected, age})$, the asymptomatic rate for each age group, as follows.

1. Fix the asymptomatic rate for the 75+ age group, $P(\text{sev 1}|\text{infected, age grp 5})$ to 13%. The 13% figure comes from (17), where a nursing home in Seattle had 3 asymptomatic cases out of 23 confirmed cases.

2. To estimate the asymptomatic rate of the remaining four age groups, we attempt to match the following data points by minimizing the sum of squared errors, subject to the (assumed) constraint that the asymptomatic rates decrease over age groups 2 through 5.

   (a) The CDC estimated that the population asymptomatic rate in the USA was 35% (Source: (10)). Weighting our age-stratified asymptomatic rates by the age distribution for the US population we should obtain a value close to 35%. (Sources for age demographics: (18) and (19).)

   (b) The Diamond Princess cruise ship had an estimated 17.9% asymptomatic rate (Source: (20)). Exactly as we did for the CDC US-population rate, we use age strata for the infected passengers on the Diamond Princess to attempt to match the 17.9% rate.

   (c) A study of 78 infected patients from Wuhan had the following age profile for the 33 asymptomatic patients: 25^{th} percentile: 26 yrs, 50^{th} percentile: 37 yrs, 75^{th} percentile: 45 yrs (Source: (21)). We attempted to match these percentiles. We use the age demographics of China for this purpose. (Source: (22).)

To this point then, we have estimated the asymptomatic rate for each of the 5 age groups, $P(\text{sev 1}|\text{infected, age})$. We next divide the remaining probability within each age group into severity levels 2, 3 and 4 using CDC numbers for hospitalization rates and ICU rates in the nominal planning scenario (10). By our definition, hospitalization includes both severity levels 3 and 4, and ICU corresponds to severity level 4. The three equations we need for the three unknowns (probability of each of severity levels 2, 3 and 4) are

1. $P(\text{symptomatic}|\text{infected,age}) = P(\text{sev } 2,3,4|\text{infected,age}) = 1 - P(\text{sev 1}|\text{infected,age})$.

2. Given that a patient is symptomatic, the probability they will be hospitalized is

$$P(\text{sev } 3,4|\text{infected,age})/P(\text{symptomatic}|\text{infected,age}).$$

3. Given that a patient is hospitalized, the probability that they will be admitted to the ICU is

$$P(\text{sev } 4|\text{infected,age})/P(\text{sev } 3,4|\text{infected,age}).$$

The CDC (10) estimates the symptomatic case hospitalization ratio to be 1.7% for age 0-49, 4.5% for age 50-64, and 7.5% for ages 65+. The percent admitted to ICU among those hospitalized is 21.9% for age 0-49, 29.2% for age 50-64, and 29.8% for ages 65+. We recognize that the age cutoffs are slightly different to ours. We match the CDC's estimates for age 0-49 to our first two age groups, those for age 50-64 to our second age group, and those for 65+ to our fourth and fifth age groups. The probabilities of severity levels 2, 3, 4 are calculated accordingly to fit these estimates.

Severity Calculation Part 2: Age Distribution To complete our severity calculation, we first identify different groups on Cornell's campus and estimate their distribution over the five age groups. The parameter values are given in Table S3.

Table S3. Information for different population groups on Cornell's campus. The size of each group as well as the faculty age distribution are provided by (23); the age distribution for academic professionals, staff, and students are assumed.

	Group size	Age group 1 (0-17)	Age group 2 (18-44)	Age group 3 (45-64)	Age group 4 (65-74)	Age group 5 (75+)
Faculty	1684	0%	33.1%	46.1%	17.9%	2.9%
Academic professionals	1114	0%	90%	10%	0%	0%
Staff	7485	0%	50%	50%	0%	0%
Students	24027	0%	100%	0%	0%	0%

For the Fall reopen, each of the 7 Cornell groups has an age distribution based on the table above. This age distribution dictates the severity distribution for each group. We assume that the remaining group (Greater Ithaca) has the same age distribution as the US population.

C. Contact Tracing. In our model, each positive case identified through self-reporting and a fraction of cases identified through asymptomatic surveillance initiates a contact trace. Contact tracing is not recursive, in that we do not model contact tracing of cases identified in a contact trace. This is for simplicity, but also because the number of contacts of those identified in a contact trace are likely to have had fewer contacts than those identified by self reporting or asymptomatic surveillance, since their detection was not triggered by one of these two mechanisms. (Here we use the term "contact" in the sense of potentially leading to infection, rather than a more restrictive sense used by the Tompkins County Health Department (TCHD).) Our model of contact tracing is necessarily simplistic, since we do not model individuals and their contact networks in our compartmental simulation.

Every positive case identified through self-reporting initiates a contact trace. Each contact trace results in some number of individuals isolated and quarantined. We take the number of isolations per contact trace to be a Poisson random variable and the number of quarantines per contact trace to be a constant. We assume that the contacts of each positive case do not overlap, so in generating the total number of individuals isolated or quarantined based on, e.g., n new positive cases identified through self-reporting, we can simply generate a single Poisson random variable with a mean that is n times that for a single case. It remains to specify the mean of the Poisson random variable for the number of isolations per initiated contact trace, and the constant number of quarantines per initiated contact trace. We assume that the positive case has had, on average, c contacts per day for t days, for a total of ct contacts. Contacts are infected independently of one another with probability p. Contacts, whether infected or not, are assumed to be remembered by the positive case with probability r. The value of p is estimated to be 1.8% in Section H below. The value of c is on the order of 12 or 13, depending on the group, as discussed in Section H below. Given that the positive case self-reported, they must be symptomatic, and so t is taken to be the sum of the means of the times in the Infectious and Symptomatic states. Under our nominal parameters, this gives $t = 3 + 1/0.22 = 7.55$ days. The value of r is taken to be 0.5, in line with anecdotal evidence from the TCHD. Accordingly, the expected number of contact-traced infected contacts is $ctpr = 0.85$. It is reasonable to expect the expected number of contact-traced non-infected contacts to be $ct(1-p)r = 46.3$, but this number reflects a great deal of double counting of individuals. Anecdotal evidence from TCHD suggests that on the order of 7 individuals are identified through contact tracing on average, suggesting that the number of contact-traced non-infected contacts should be taken to be $7 - ctpr = 6.15$ under nominal parameters. We adopt this figure instead.

Positive cases identified through asymptomatic surveillance are modeled in the same manner, except that cases identified in this manner would typically be identified earlier in the course of their disease, at which point they would have infected fewer people. We model this by only initiating contact traces for a fraction of the positive cases identified through asymptomatic surveillance. We take the number of contact traces initiated on each day to be Poisson with mean $N/2$, where N is the number of surveillance positives from the relevant day.

All infected cases identified through contact tracing are pulled from the Exposed, Infectious, Symptomatic and Asymptomatic states, in that order of precedence, and enter the Isolation state. All non-infected cases identified through contact tracing are pulled from the Susceptible state and enter Quarantine.

Table S4. Parameters for contact tracing.

Parameter Description	Nominal Parameter Value(s)	Sources
Fraction of contacts recalled, r	0.5	
Contact tracing delay	1 day	(24)
Contact traces initiated per screening positive	0.5	
Contact traces initiated per self-report positive	1	
(Implied) New isolations per initiated contact trace	0.85	Calculation in text
(Implied) New quarantines per initiated contact trace	6.15	(24)

D. Outside Infections. We estimate the probability of outside infection per person per day, which arises from infections imported from outside the modeled groups due predominantly to travel outside Tompkins County. TCHD data reports 13.2 travel-related COVID cases per month from March 2020 to July 2020. The asymptomatic rate at that time was estimated to be approximately 50%, so the actual number of cases is estimated to be twice this number, or 26.4 cases per month. Assuming that during this period there were 75,000 people in Tompkins County, we arrive at a figure of $26.4/30/75,000 = 1.2 \times 10^{-5}$ for the probability of outside infection per person per day.

An additional source of outside infections comes from students returning at the start of the fall semester, which we model next.

E. Students Returning and Initial Prevalence. In advance of the fall 2020 semester, New York state required all travellers from high-prevalence states to self-quarantine for two weeks upon arrival. The list of high-prevalence states changed throughout August 2020, in advance of the Fall Semester. Our analysis is based on New York State's list of High Prevalence states on August 7, 2020. We model the return of students to campus in two phases: (1) a 14-day period when students from high-prevalence states arrive and self-quarantine, followed by (2) move-in weekend when other students arrive.

The modeled student arrival process is summarized below.

- Some students get tested remotely and are isolated if positive. Others come without being tested. Students coming from high-prevalence states are less likely to have test access at home.

- Students traveling to campus risk additional infection after being tested at home prior to departure (if they are tested) and during travel.
- Students are required to be tested upon arrival as a condition for enrollment. Students are strongly encouraged to use the first available testing date, though some will instead choose to be tested later. Positives are isolated, including some false positives. If a student comes from a high-prevalence state, then the student is required to self-quarantine for 14 days.
- Some positive cases already exist on campus due to infections from the greater Ithaca area.
- Some positive cases among incoming students are missed because of false negatives and because some students are early enough in their infection to not be PCR-detectable.
- These two sources of cases (existing and new) combine to create an on-campus prevalence.
- This on-campus prevalence creates additional cases on campus. Some additional cases are also created on campus due to outside infections from the greater Ithaca area.
- During the two-week period before the move-in weekend, regular surveillance testing had not begun, but contact tracing was underway.

E.1. 14-day self quarantine. Here we discuss the model for the arrival of students from high-prevalence states for which New York State requires a mandatory 14-day self-quarantine. The students among these that have access to housing in which they can self-quarantine are modeled as arriving in Ithaca two weeks before classes start. Other students in this group without such housing are modeled as either choosing to start classes virtually or, in a few cases, coming to Ithaca without complying with the required quarantine period in violation of state law.

Incoming Student Population Sizes: Student data suggested that roughly 33% of the undergraduate students and 23% of the graduate / professional students have homes in states designated by New York State as "high prevalence" requiring mandatory quarantine.

We assume that many such students with off-campus housing will spend the mandatory quarantine period in Ithaca in that housing. For students that originally planned to be in on-campus housing, we assume that the majority will not come to Ithaca at the start of the semester but rather will begin the semester online; a small fraction will quarantine somewhere outside Ithaca and return during the move-in weekend; while another small fraction will fail to comply with the law, either using non-compliant quarantine in shared housing in Ithaca, or by arriving during move-in weekend without having quarantined. Assuming that 10% of continuing undergraduates and 75% of continuing graduate / professional students have stayed in Ithaca, the total number of students arriving 2 weeks in advance from high prevalence states is 3750, including 2500 undergraduate students and 1250 graduate / professional students.

Compliance: Despite the mandatory self-quarantine order, we do not assume full compliance. We estimate the daily transmission rate to be reduced by 40% compared with the nominal setting. We do this to model several kinds of non-compliance with quarantine. First, some students required to quarantine may do so in non-compliant locations shared with others. Second, some students may break quarantine and have social interaction. Third, although students were asked to test on arrival (so that positives can be isolated and monitored, reducing the danger of transmission), testing was offered only three times a week so there may be a delay between arrival and the first available test date.

Testing Before Departure: Cornell students were asked to test before departing to come to campus, but this was not mandated due to a lack of test access for some students. We assume that $\frac{1}{3}$ of students from high-prevalence states were tested at home, and $\frac{2}{3}$ from low-prevalence states, both using nasopharyngeal (NP) sampling with 90% sensitivity (25).

Testing on Arrival: As discussed above, we assume that students are tested once on arrival. We assume NP sampling with 100% compliance. Because the semester had not begun, and mandatory asymptomatic screening had not started, we assume that no other testing is done.

Prevalence Estimation for High-Prevalence States: Prevalence at the origin of students from high-prevalence states is assumed to be 4%. This estimate was obtained by multiplying daily new positive cases, an underreporting factor (assumed to be 10, i.e. for each reported positive case there are 9 positive cases not reported), and the average number of days an infected individual is active (assumed to be 20).

Population Already in Ithaca: The total number of students that either stay in Ithaca during the summer or come to Ithaca early from other "low prevalence" states is estimated to be 4090 (including 1130 undergraduate students, 2960 graduate / professional students). All 10280 employees are assumed to remain in Ithaca throughout the summer. The prevalence among the group of unquarantined students and the group of employees is assumed to be 0.1%, which is consistent with the estimated persistent prevalence level in the greater Ithaca area during summer 2020. (See below)

Assuming 31 confirmed cases, which is what was observed over the first 21 days of July 2020, that cases last 20 days, and 2x-4x underreporting in Tompkins County (less than elsewhere due to excellent testing access), gives 60 - 120 active cases, or 0.075% - 0.15% prevalence.

Interactions: During the two-week period before classes start, we assume no interaction between students and employees. We use a multi-group simulation consisting of four groups – self-quarantined students, unquarantined students, employees, and the greater Ithaca community – to model different behaviors (reflected by daily transmission rate) within and across the

groups. As noted elsewhere, we assume 40% compliance with quarantine requirements amongst self-quarantining students. The transmission matrix for the self-quarantine period is summarized in Table S5.

Table S5. Inter- and intra-group transmissions per day during the self-quarantine period, based on the multi-group simulation, which use contacts from the literature, choose an infectivity calibrated to an estimate of R0, and then multiply to get transmission. Each entry gives the expected number of transmissions per day from one infected member of the row group to each of the column groups.

Group (pop. size)	Self-quarantined Students	Unquarantined Students	Faculty / Staff	Greater Ithaca Community
Self-quarantined Students (3748)	0.031	0.010	0	0.018
Unquarantined Students (4087)	0.0087	0.053	0	0.031
Faculty / Staff (10283)	0	0	0.031	0.027
Greater Ithaca Community (62000)	0.0011	0.0020	0.0044	0.060

Simulation results give us that the initial prevalence among Cornell students in Ithaca immediately prior to move-in weekend is 0.17% and 0.087% for faculty and staff.

E.2. Move-in weekend and low-prevalence states. *Prevalence Estimation for Low-Prevalence States:* NY state designated a state as "high prevalence" if its daily reported number of new positive cases exceeded 10 per 100,000 population. Assuming an under-reporting factor of 10 and an average active period of 20 days, this daily new positive case threshold translates to a prevalence of 10 / 100,000 * 10 * 20 = 2%. Hence, the overall prevalence in student origins that are not designated as "high prevalence" is at most 2%. This prevalence is prior to any testing at the origin prior to departure for Cornell.

Incoming Student Population Sizes: As discussed previously, in addition to students from low prevalence states we assume that a small fraction of the students (300) from high-prevalence states that plan to live on-campus will return during the move-in weekend. Although these students will have presumably self-quarantined for 14 days elsewhere, we pessimistically assume non-compliance and consider their prevalence upon entering Ithaca to be 4%. Given it is a small population compared to students from low-prevalence states (with prevalence $< 2\%$), and the assumed under-reporting factor of 10 is large given the access to testing in low-prevalence states at the time, we assume that the overall prevalence among students returning during the move-in weekend is exactly 2%. We estimate the total number of students returning during the move-in weekend to be 10770, including 8180 undergraduate students and 2590 graduate / professional students.

Prevalence of returning students: Students were asked to test before departure, but this was not mandated due to a lack of test access. We assume that $\frac{2}{3}$ of the students from low-prevalence states were tested at home, using nasopharyngeal (NP) sampling (90% sensitivity). Hence, the fraction of returning students that are infectious is estimated to be $2\% * (1 - \frac{2}{3} * 90\%) = 0.8\%$. In addition, we also assume a small per-day infection probability during travel. The travel duration and the likelihood that students use public transportation (with an associated elevated daily infection probability) depends on the geographic origin of students. Weighting these probabilities by geographic origin of students, we estimate that an additional 0.1% of the returning students are infected during travel to campus. Among them, 45% are estimated to be in the Infectious state upon arrival (which can be detected with probability 90%), and 55% are estimated to be in the Exposed state upon arrival (which cannot be detected by arrival testing). Assuming arrival testing with NP sampling and 100% compliance, the fraction of returning students that are infected and not identified by arrival testing is $(0.8\% + 45\% * 0.1\%) * 10\% + 55\% * 0.1\% = 0.14\%$.

The initial prevalence estimates for the student groups combine the initial prevalence estimates from the 14-day simulation (local students and self-quarantine of high-prevalence states) and move-in weekend (low prevalence state students) to reflect the composition of each group. The initial prevalence of all the groups after arrival and immediately prior to the semester is summarized in Table S6.

Table S6. Initial prevalence estimates for modelling of Cornell Fall semester.

	UG high	UG low	GS research	GS class	FS student	FS not student	FS off	Greater Ithaca
Initial Prevalence	0.156%	0.161%	0.166%	0.1628%	0.087%	0.087%	0.087%	0.08%

F. Testing Details. For asymptomatic surveillance we assume a sensitivity of 60% for PCR testing from observed self-collected anterior nares (AN) sampling, using the same test sensitivity for both pooled and individual testing. This is based on preliminary results from a validation effort at Cornell in which paired AN and nasopharyngeal (NP) swabs were collected and tested from the same individuals. Testing of AN samples identified 75% of the positives found via NP. As before, we assume a sensitivity of 90% for NP (25), that all of the positives missed by NP (10% of all positives) are also missed by AN (since these individuals would likely have low viral loads), and that an additional 25% of the 90% of the positives found by NP are missed by AN (or $0.25 \times 0.90 = 22.5\%$ of positives). This results in a sensitivity of $1 - 0.1 - 0.225 = 67.5\%$. Since AN samples are self-collected in surveillance testing, which is subject to the risk of improper sample collection, we adopt a pessimistic estimate of 0.6 for the sensitivity of surveillance tests using AN.

This estimate may be somewhat pessimistic, since some studies suggest that NP's sensitivity is higher than 90% (26), and some positives may be missed by NP sampling because of improper sampling technique (27).

On the other hand, this calculation does not explicitly account for the loss in sensitivity due to pooling. Cornell uses pools of size 5 in surveillance testing and retests the original sample when a pool tests positive. Based on existing mathematical

models for pooled testing, this procedure should diagnose the same set of positives as does unpooled surveillance, unless the sample has a Ct value within $\log_2(5) = 2.3$ cycles of the limit of detection. Because SARS-CoV-2 viral loads vary by several orders of magnitude (28), the fraction of samples with a viral load in this range is small.

Inter-group Dynamics.

G. Group Details. We model the spread of COVID by splitting the campus into 8 groups and considering the interactions between groups and among themselves. We also track infections and hospitalizations in each group. The abbreviation for each group is in brackets after its name.

1. Undergraduates living in high-density housing (dorms, fraternity and sorority houses) [UG high]
2. Undergraduates living in low-density housing [UG low]
3. Graduate students primarily engaged in research [GS research]
4. Graduate and professional students primarily engaged in classroom instruction [GS class]
5. Faculty / staff working on campus who are student facing [FS student]
6. Faculty / staff working on campus who are not student facing [FS not student]
7. Faculty / staff working off campus [FS off]
8. Greater Ithaca community [Greater Ithaca]

Table S7. Group sizes for modelling of Cornell Fall semester.

	Groups	UG high	UG low	GS research	GS class	FS student	FS not student	FS off	Greater Ithaca
Group Size	6920	8123	3645	4921	3598	3598	1907	4778	62000

H. Transmission. Transmission within and between each group is governed by the "transmission rate matrix." This is estimated first by estimating a rate of contacts within and between each group, and then calibrating the transmission probability per contact to a value of R0. There is a transmission rate matrix for summer of 2020 to model the pre-semester period and a transmission matrix for fall of 2020.

The term "contact" is used consistently with the literature, where a contact is defined as a two-way conversation or a physical interaction (e.g., a kiss or handshake) (29). Thus, it includes those contacts that are more brief than the CDC's definition of a close contact (6 feet or less and 15 minutes or more).

We now describe our estimation methodology for both the summer 2020 and fall 2020 matrices in an algorithmic manner.

1. Choose a nominal value of R0 in the general US population under normal circumstances. We used 2.5 as per CDC Planning Scenarios (10).
2. Choose a number of contacts per day for each age group based on the literature. We use contacts per day from (29).
3. Choose a transmission probability per contact that matches R0 to get transmissions/day as computed from (contacts / day) * (transmission / contact) for individuals, broken down by age. Based on the contact rate matrix from Step 2 and the age distribution within the US, the average number of contacts per day within the US population is 12.7. Given an R0 of 2.5 and the expected infectious period of the disease, the transmission probability is estimated to be 1.8%.
4. For each of the groups UG student, Graduate/Professional student, staff/faculty, non-Cornell Tompkins County resident, use the age distribution to calculate transmissions / day for each group, under pre-social-distancing conditions. We will subsequently adjust for social distancing.

   Transmissions per day for each group under pre-social-distancing conditions, based on the age-stratified contact rates in (29)

   - Undergraduate Student: assuming age group 15-19 in (29)
     - 17.58 contacts / day * 1.8% infectivity rate = 0.32 transmissions per day
   - Graduate Students: age group 20-29 in (29)
     - 13.57 contacts / day * 1.8% infectivity rate = 0.24 transmissions per day
   - Faculty / Staff: using the age distributions from Table S3
     - 12.9 contacts * 1.8% infectivity rate = 0.23 transmissions per day

- Non-Cornell Greater Ithaca residents: assuming the same age demographics as reported by US census (19)
  - 12.7 contacts * 1.8% infectivity rate = 0.23 transmissions per day

5. Calculate the rate of transmission between groups using summer case count observations in Tompkins County as well as the pre-social-distancing contact rates assumed above.

   - Calibrate impact of social distancing among the Cornell summer-population (staff/faculty + summer-resident graduate/professional and UG students) and the Greater Ithaca population. Set R0 in this population to 0.75 based on the Ithaca Summer 2020 R0 argument below. This means transmissions per day is reduced 70% from our pre-social-distancing calculation (which is calibrated to R0 = 2.5).
   - Literature also suggests that younger people are less likely to abide by social distancing regulations (30). Therefore we will assume that the impact (multiplier) of social distancing is 50% less effective for students during the summer. A 70% reduction for this group becomes a 70%/1.5 = 47% reduction.
   - Using this estimate and the following additional assumptions, we can create an estimate of the summer transmission matrix. Assumptions:
     - Undergraduates and course-based graduate students all leave Ithaca over the summer.
     - 75% of research-based graduate students remain in Ithaca.
       * Transmissions per day during summer: 0.24 (1 - 0.47) = 0.127
     - All faculty/staff remained in the Ithaca area during the summer and worked remotely.
       * Transmissions per day during summer: 0.23 (1 - 0.7) = 0.069
     - The non-Cornell Ithaca community observed 70% social distancing.
       * Transmissions per day during summer: 0.23 (1 - 0.7) = 0.069
     - Breakdown of contacts by group:
       * Percent of contacts with outside community. From Figure 2A in (29), about 60% of contacts are from home, work, school, or multiple. About 20% are leisure. We will assume that social distancing scaled down transmissions proportionately, and will model 60% of transmissions for faculty/staff as Cornell-related. For faculty/staff Cornell transmissions, the majority of the contacts are within their own group (student facing, not student facing, off campus).
       * Graduate students will have 75% of contacts, and thus transmissions, be Cornell-related. About 25% of these Cornell contacts are with faculty/staff and all others are with grad students. The majority of the contact with faculty/staff is with people that will be on campus and student-facing in the fall.
     - Symmetry condition for daily transmissions: The expected daily transmissions between group 1 and group 2 is the expected daily transmissions between group 2 and group 1. Therefore, selecting the daily transmission rate per person in group 1 with group 2 determines the daily transmission rate of someone in group 2 with group 1.

6. This results in the summer transmission rate matrix in Table S8. The overall average transmission rate per day (within the Cornell community) for summer is 0.0828.

Table S8. Summer 2020 transmission rate matrix for Cornell.

Groups	GS research	FS student	FS not student	FS off	Greater Ithaca	Expected Transmissions Per Day
GS research	0.072	0.021	0.0009	0.0036	0.0324	0.127
FS student	0.0169	0.018	0.0028	0.0054	0.029	0.071
FS not student	0.0013	0.0051	0.033	0.0036	0.029	0.071
FS off	0.0020	0.0041	0.0015	0.033	0.029	0.068
Greater Ithaca	0.0014	0.0016	0.00087	0.0021	0.064	0.069

7. To derive the transmission matrix for Fall 2020, we assume that the pairwise rates of interaction between grad students, faculty/staff and the Ithaca community remain the same as during the summer, but there will be an increase in overall transmission due to an influx of students arriving to campus.

   - Younger people are less likely to wear masks and socially distance (30). We assume that students (undergraduates, graduate students (course-based)) reduce their pre-social-distancing transmissions by 30%, about half as effective social distancing as in Ithaca during the summer. This is more pessimistic than our previous assumption regarding graduate research students who reduced their transmissions by 47%. We do not assume an increase in transmissions per day of graduate research students with faculty/staff or the Ithaca community.
   - Undergraduates (off campus): Edmunds 2006 (31) surveys undergraduate students and finds that 15.2% of their contacts are with people over the age of 30. This represents the percent of their contacts with faculty/staff and the Ithaca community. We reduce this number to 10% to reflect the reduced staff on campus. Almost all of these contacts are with student-facing staff and there is some contact with the Ithaca community.

- Undergraduates (high-density housing) have more transmissions per day with other people in high-density housing, half the transmissions per day with the Ithaca community, and the same transmissions to faculty/staff and grad students as undergraduates (off campus).
- Graduate students (course-based) have the same transmissions per day to graduate students (research), faculty/staff, and Ithaca as undergraduates (off campus). Inter-group transmissions are selected to approximate expected transmissions per day for the group.
- Grad student (research): we assume 100% of graduate research students are in Ithaca in the fall semester, while this number is assumed to be 75% during the summer.
- All rates between grad student (research), faculty/staff and Ithaca community remain the same as in the summer transition matrix.

8. This leads to the Fall 2020 transmission rate matrix in Table S9. The average transmission rate per day within the Cornell community is 0.198.

Table S9. Fall 2020 transmission rate matrix for Cornell.

Groups	UG high	UG low	GS research	GS class	FS student	FS not student	FS off	Greater Ithaca
UG high	0.22	0.072	0.0018	0.0018	0.018	0.0009	0.0009	0.0018
UG low	0.061	0.15	0.0018	0.0018	0.018	0.0009	0.0009	0.0036
GS research	0.0034	0.0039	0.072	0.0018	0.021	0.0009	0.0036	0.033
GS class	0.0025	0.0031	0.0013	0.16	0.018	0.0009	0.0009	0.0036
FS student	0.035	0.040	0.022	0.024	0.018	0.0028	0.0054	0.029
FS not student	0.0033	0.0038	0.0018	0.0023	0.0051	0.033	0.0036	0.029
FS off	0.0013	0.0016	0.0028	0.0009	0.0041	0.0015	0.033	0.029
Greater Ithaca	0.0002	0.00047	0.0019	0.00029	0.0016	0.00087	0.0021	0.064

Ithaca Summer 2020 R0 Case counts in Tompkins County in the summer of 2020 are consistent with R0 < 1 among the non-Cornell Tompkins County and summer-resident Cornell population. However, the R0 was large enough that importing new cases created a not insubstantial number of additional cases. For the purposes of estimating the R0 of the non-Cornell community, we focus on July 2020 data.

First, according to the Tompkins County Health Department (TCHD), the number of new cases per day rose at the beginning of July when prevalence nationwide rose, but gradually declined after. If R0 were bigger than 1 in Tompkins County, then we would expect that new cases would grow exponentially. The fact that this did not happen suggests that R0 < 1.

Second, the TCHD reports that 16 out of 31 cases between July 1 and July 21 had relevant travel to a high-prevalence region. Let us make the following assumptions:

- Assume reporting bias is the same for both individuals infected locally and infected due to travel.
- Assume that all of these cases between July 1 and July 21 resulted from clusters initiated by external travel that happened in July, predominantly July 4, and not from clusters that were present in Tompkins County before July. This is based on the observation that prevalence in June in Tompkins County was very low. Also, if one assumes that some local July cases began due to pre-existing clusters then this will cause our R0 estimate to decrease further.
- Let us momentarily assume that all clusters initiated by July travel concluded by July 21. This assumption is too optimistic, and will create an R0 estimate that is too low — we will adjust for this in a moment.

In general, in a large fully susceptible population with $R0 < 1$, each new case creates a cluster that infects $1 + R0 + (R0)^2 + (R0)^3 + \ldots = 1/(1 - R0)$ individuals, including the original case. (This ignores the effect of immunity and is accurate for R0 sufficiently below 1.)

Then, under these assumptions, to find R0 in Tompkins County in July, we need to find a number such that $16/(1-R0) = 31$. Solving for R0 we get $R0 = 1 - (16/31) = 0.48$.

Finally, our third assumption above was too optimistic. In fact, some clusters that started in July due to known travel likely still had not finished infecting new people. In light of this, we increase our estimate of R0 to 0.75.

I. Virtual Instruction. This section looks at the scenario of *virtual instruction*, where research-based graduate students are on campus and subject to mandatory testing and asymptomatic screening and other students are asked not to return. In this scenario, some of these students choose to return to Ithaca despite this request. Cornell has reduced ability to enforce behavior changes and regular asymptomatic screening as compared with the residential instruction setting.

This section describes the methodology for selecting parameters for this virtual instruction scenario. In addition to the change in undergraduate and class-focused graduate student test compliance, which reflects Cornell's reduced ability to enforce behavior changes among the returning undergraduate population, two sets of additional parameters are changed relative to the Cornell re-open scenario: the group sizes (Table S7) and the transmission rate matrix (Table S9).

1. Group Sizes

Table S10 gives the population size for each group for virtual instruction. We assume that the last three columns — Faculty/Staff not student-facing, Faculty/Staff off-campus, and Greater Ithaca — are independent of the policy change since the people in those groups are very likely to obey the same routines regardless of the scenario.

Table S10. Group sizes for virtual instruction scenario.

Groups	UG high	UG low	GS research	GS class	FS student	FS not student	FS off	Greater Ithaca
Group Size	0	3468	1594	1434	3598	1907	4778	62000

To estimate the population sizes for the student groups, we used results from a survey sent out on May 29, 2020 to all students enrolled at the time, while attending to two concerns:

(a) Not all of the students who received the survey responded.

(b) The survey result does not include students who would enroll in the fall of 2020 for the first time, namely rising undergraduate freshmen and new graduate students.

For the first concern, since 71% of the undergraduates and 48% of the graduates responded, we assume these percentages generalize to the whole population. For the second concern, we will explain group by group how we handle it.

- Undergraduate students:
  - For the UG high-density housing ("UG high") group: we set the group size to be 0, since on-campus dorms would be closed.
  - For the UG low-density housing ("UG low") group: the number is calculated from 11186*0.31=3486 where 11186 is the number of undergrads surveyed and 31% is the percentage who responded "very likely" to return for a virtual semester. The number of survey recipients, 11186, does not include any of the incoming first year students. Using this number, we are implicitly assuming that no freshman students come to Ithaca under a virtual instruction scenario, which is conservative in the sense that it under-estimates unsurveilled students in this scenario.
- Graduate students:
  - There are two graduate student groups, GS class and GS research. In the residential scenario, these groups have population sizes 4921 and 3645, respectively.
  - From the May 29 survey results, we estimate that 53% of the graduate student population would return under a virtual instruction scenario. We assume this percentage applies evenly across both class-based and research-based graduate students.
  - We assume that 25% of research graduate students are first years, and 50% of class-based graduate students are first years. We assume that non-first-year students in each group are subject to the 53% return percentage, from which we obtain $4921 \times 0.5 \times 0.53 = 1304$, and $3645 \times 0.75 \times 0.53 = 1449$, corresponding to the number of non-first-year students who return to Ithaca from each of the GS class and GS research groups.
  - For the first-year graduate students in each group, we assume that the 53% likely-to-return proportion is reduced by a further 90% in the case of class-based students, and 70% in the case of research-based grad students. This gives a total of $4921 \times 0.53 \times 0.5 \times 0.1 = 130$ and $3645 \times 0.25 \times 0.53 \times 0.3 = 149$ first year graduate students returning to Ithaca in each of the groups.
  - Combining the above, we get 1434 class-based graduate students and 1594 research-based graduate students.
- Faculty and Staff

  As we stated above, we assume faculty and staff behaviors are somewhat independent of the scenarios. Thus, we keep the faculty populations the same as an in-person semester in each group.

2. Transmission Rate Matrix

The transmission rates for virtual instruction are based on the transmission rates for residential instruction with some adjustments. As a reminder, transmission rate = contacts / day * 1.8% infectivity rate, and we assume that the interaction between faculty/staff within themselves and with the Greater Ithaca community does not depend on scenarios. The main idea for estimating transmission rates for virtual instruction is that class-based students would interact less with faculty and staff, but more with the Greater Ithaca community. Student interactions among themselves depend on their compliance with the behavioral compact (e.g., mask-wearing and social distancing) and housing density in Collegetown. We explain each of the transmission rates we have re-calculated below.

- UG high

  - Since we assume no one in "UG high" will return, there is no transmission from this group to others.
- UG low / GS class within-group
  - The virtual scenario has two competing effects: reduced density of transmissions due to fewer people on campus, and potential increase in transmissions due to Cornell's reduced ability to enforce mask wearing, social gathering restrictions, and abundant asymptomatic testing.
  - First, we discuss the effect of social gathering and mask wearing. In the residential instruction scenario, we assumed that Cornell's ability to legally mandate mask wearing and social gathering restrictions with a behavioral compact resulted in a 30% reduction in transmission between pre-social-distancing periods and a residential fall semester. Under virtual instruction, since Cornell will not be able to enforce a behavioral compact, we assume that this reduction in transmission (between the summer and a virtual fall semester) will be less than between the summer and residential instruction. While one might imagine that there would be no reduction in transmission between the summer and a virtual instruction fall given Cornell's reduced ability to enforce a behavioral compact, we optimistically assume a 15% reduction. This has the effect of increasing the within-group transmission rates of "UG low" and "GS class" by a factor of (1-15%)/(1-30%) from the residential setting.
  - Second, Section 3.1 and Figure 4 of (32) suggest that the mortality rate of infectious diseases rises with population density up until population density reaches 200 people per square mile and then levels off. Below, we estimate that virtual instruction reduces the population density to roughly 2000 / square mile from roughly 6000 / square mile under residential instruction. Although the literature thus suggests that there will be no reduction in transmissions due to virtual instruction relative to residential instruction, we conservatively assume that virtual instruction will result in a reduction of transmissions by 20%.
  - Population-density calculation: For the people who live in Ithaca, according to the percentage in Section A5, roughly 30% of the juniors, seniors and class-based graduate students who live in Collegetown are returning this fall. Moreover, we estimate that roughly 20% of Collegetown residents are not undergraduates and not class-based graduate students. Thus, in total the density in Collegetown is around (0.8*0.3+0.2*0.5)=0.34 of Collegetown residents are returning. Since the City of Ithaca has a living density of 5893 people per square mile, Collegetown has 5893*0.34=2004 people per square mile for the virtual instruction scenario.
  - Combining the two effects described above, we multiply the residential within-group transmission rate for "UG low" and "GS class" by a factor of (1-15%)/(1-30%) * (1-20%) = 0.9712
- UG low / GS class with faculty, staff and graduate students:
  - "UG low" and "GS class" will have much less interaction with "FS student" and "FS not student" since they do not need to see any professors in person. Thus, we assume the transmissions from any "UG low" and "GS class" person to on-campus faculty will drop to minimal to be the same as transmissions to any off-campus faculty.
- UG low / GS class with Greater Ithaca:
  - A virtual semester that shuts down the campus including the dining halls would increase undergraduate interaction with Greater Ithaca for reasons like groceries and other necessary activities. However, "UG low" and "GS class" are unlikely to leave the Collegetown area very frequently. Therefore, a number larger than the transmission rate for UG low / GS class with Greater Ithaca in the residential instruction scenario but less than that of GS research would be a reasonable estimate. Thus, we set the transmission rate for UG low / GS class with Greater Ithaca to be a little over half of that of GS research with Greater Ithaca, the figures of which do not change from scenario to scenario.

In summary, Table S11 gives the virtual instruction transmission matrix.

Table S11. Virtual instruction transmission rate matrix for Cornell.

Groups	UG high	UG low	GS research	GS class	FS student	FS not student	FS off	Greater Ithaca
UG high	0	0	0	0	0	0	0	0
UG low	0	0.20	0.0018	0.0018	0.0009	0.0009	0.0009	0.0018
GS research	0	0.0039	0.072	0.0018	0.021	0.0009	0.0036	0.033
GS class	0	0.0043	0.0020	0.16	0.0009	0.0009	0.0009	0.0018
FS student	0	0.00087	0.0095	0.00036	0.018	0.0028	0.0054	0.029
FS not student	0	0.0017	0.00075	0.00068	0.0051	0.033	0.0036	0.029
FS off	0	0.00066	0.0012	0.00028	0.0040	0.0015	0.033	0.029
Greater Ithaca	0	0.0001	0.0008	0.0004	0.0016	0.00087	0.0021	0.064

J. Matrix input for simulation. We have previously described how we estimated transmission matrices for the Fall semester (Tables S9 and S11). These matrices represent the average number of new infections per day in the column group from each free and infectious person in the row group. Unfortunately, our code is not structured to directly take the transmission matrix as an input.

Instead, it takes the so-called "interaction matrix" as an input, where the mean number of new infections in group i from group j in a day is given by

$$p * \texttt{free\_susceptible}[i] * \texttt{interactions}[i,j] * \texttt{free\_infectious}[j]/\texttt{free\_total}[j]. \quad [1]$$

Here, p is the probability of transmission per interaction, $\texttt{interactions}[i,j]$ is the value of the matrix inputted to the simulation at row i and column j, $\texttt{free\_susceptible}[i]$ is the number of free and susceptible individuals in group i, $\texttt{free\_infectious}[j]$ is the number of free and infectious individuals in group j, $\texttt{free\_total}[j]$ is the total number of free individuals in group j.

Note that $\texttt{interactions}[i,j]$ was intended to represent the number of contacts within group j by a single person in group i on a single day and $\texttt{free\_infectious}[j]/\texttt{free\_total}[j]$ is the fraction of the free population in j that is infectious. Thus, the expected number of contacts that a free susceptible person in group i would have with a free and infectious person in group j would be $\texttt{interactions}[i,j] * \texttt{free\_infectious}[j]/\texttt{free\_total}[j]$. We then multiply by the number of free susceptible individuals in group i and the probability of transmission upon contact to get the total number of contacts with infectious people in group j by free and susceptible people in group i. This recovers Equation 1.

To convert the transmissions matrix (Tables S9 and S11) to the interaction matrix used as an input to our simulation, we will count in two ways the number of interactions between infectious people in group j and susceptible people in group i, and set them equal to each other.

First, consider the infectious people in group j and count their interactions with people in group i. There are a total of $\texttt{free\_infectious}[j] \cdot \texttt{transmissions}[j,i]$ transmissions from group j to group i. This implies $\texttt{free\_infectious}[j] \cdot \texttt{transmissions}[j,i]/p$ total interactions with susceptible people in group i.

The second way to count the number of interactions is starting with the susceptible population in group i which has a total of $\texttt{free\_susceptible}[i] \cdot \texttt{interactions}[i,j]$ contacts with members of group j. Of these contacts the following fraction are with infectious people in group j, $\texttt{free\_infectious}[j]/\texttt{free\_total}[j]$. Therefore, there are a total of $\texttt{free\_susceptible}[i] \cdot \texttt{interactions}[i,j] \cdot \texttt{free\_infectious}[j]/\texttt{free\_total}[j]$ interactions between infectious members of group j and susceptible members of group i.

Setting these two expressions equal to each other and cancelling $\texttt{free\_infectious}[j]$ gives us $\texttt{transmissions}[j,i] = p \cdot \texttt{interactions}[i,j] \cdot \texttt{free\_susceptible}[i]/\texttt{free\_total}[j]$. Given low prevalence, we then assume that the susceptible and total free populations of each group are approximately their respective population sizes. This yields $\texttt{transmissions}[j,i] = p \cdot \texttt{interactions}[i,j] \cdot \texttt{population}[i]/\texttt{population}[j]$.

2. Model Calibration

This section describes parameter estimation and model calibration *after* the fall 2020 semester ended. Sections A and B describe model calibrations for students and employees, respectively.

Parameter estimation relies on data from the following sources:

- Aggregated de-identified positive-case, testing, and contact-tracing data collected during the semester and stored along with student life, housing, and employee data in a HIPAA-compliant database. This data was collected by Cornell to meet an urgent public health need while fighting the pandemic. This data was then aggregated and shared by the institution with the authors for research purposes. A determination was made by Cornell's Institutional Review Board (IRB) that the use of this previously collected aggregated data for research does not constitute human subjects research.

- Data in a publicly available report pursuant to the urgent public health need presented by the pandemic (33).

The data sources for all parameters are summarized in Table S12.

Table S12. Data sources for parameter estimates/calibration. "V" indicates that the data is obtained from the HIPAA-compliant database; "P" indicates that data is obtained from the publicly available report.

	Parameter name	Source
Student calibration	Population size	P
	Observed trajectories	P
	Arrival schedule	V
	Testing frequency	V
	Test compliance	V
	Outside infection rate	P
	Contact matrix	P
	Contact tracing effectiveness parameters	V, P
	Initial prevalence	V
Employee calibration	Observed trajectory	P
	Testing frequency	V
	Outside infection rate	V
	Contact tracing effectiveness parameters	V

A. Model Calibration for Students. We use a multi-group dynamic population simulation model for the student population, which consists of three subgroups:

- Group 1: undergraduates, with Greek or athlete affiliation;
- Group 2: undergraduates, neither Greek nor athlete;
- Group 3: graduate or professional students.

We employ this population breakdown because we observe substantial differences in infections and contacts for these specific subgroups. We set August 16, 2020 - November 24, 2020 to be the time period for our calibration, as the majority of undergraduates left the greater Ithaca area at the time of the Thanksgiving holiday. We divide the time horizon into two non-overlapping periods: the pre-semester period (8/16/2020 - 9/2/2020) and the in-semester period (9/3/2020 - 11/24/2020). Here we describe the parameters estimated directly from fall 2020 data.

Population size We use students' degree program information, Greek and athlete rosters, and daily check-in data to divide students residing in Ithaca into 3 subgroups, obtaining the population sizes given by the Table S13.

Table S13. Sizes of the three student groups used in student calibration and projection.

Group	Population size
1 (UG with Greek or Athletics Affiliation)	3533
2 (Other UG)	8434
3 (Graduate and Professional Students)	6202

Arrival Schedule Arriving schedules for groups 1, 2 and 3 are determined based on the arrival dates indicated by students in their Fall semester checklist, and the move-in schedule for students living on campus.

Testing Frequency The model does not track individuals and their test schedules. Rather, each member of a population is assumed to test on a given day with a given probability.

- Pre-semester period:
  - Groups 1-2: We divide the total number of non-arrival tests performed (3255) during the period by the total number of person days during the pre-semester period (127466) to estimate the testing frequency for the undergraduate students in the pre-semester period, to get 0.0255 per day per person, i.e., each person has one test on average every 39 days.
  - Group 3: 0.
- In-semester period:
  - Groups 1 and 2: 2/7 per day, corresponding to being tested 2x / week.
  - Group 3: 1/7 per day, corresponding to being tested 1x / week.

  The testing frequency during the in-semester period is consistent with the testing frequency for students assumed in the main simulation model.

Test Compliance We estimate student test compliance to be 97.4%. This value is calculated based on the fraction of scheduled student surveillance tests completed over the course of the fall semester (including both on-time tests and those tests that were delayed but completed).

Outside Infection Rate We consider a positive student case to be an outside infection if they satisfied both of the following conditions:

- they did not test positive in an adaptive test, nor were they in the contact trace of other positive cases;
- they had recent travel history.

Table S14 summarizes the number of outside infections in each group during the semester and the corresponding outside infection rate, which is the number of outside infections divided by (population size of the group $\times$ time horizon in days).

Table S14. The number of outside infections in each group during the semester and the corresponding outside infection rate.

Group	Outside infection case count	Outside infection rate (per day)
1	5	1.42E-5
2	6	7.11E-6
3	4	6.45E-6

Note that the period considered does not include the post-Thanksgiving period. During the post-Thanksgiving period, graduate students tested positive at a higher rate due to travel.

Contact Matrix We define the *daily transmission matrix* T such that the value $T(i,j)$ gives, for each infectious non-isolated non-quarantined positive in group i, the expected number of additional positives created in group j on a given day. It is difficult to estimate the daily transmission matrix directly from data because we do not observe for how many days an individual was positive. Instead, we aim to estimate the *contact matrix* M. The value $M(i,j)$ in the contact matrix is the expected number of positive cases that an infectious individual in group i creates in group j over the course of his or her infection. We then assume that the average length of time an infectious individual in a given group spends circulating (i.e., not isolated or quarantined) during the fall semester does not depend on their group. Under this assumption, M is proportional to T. Below, in our calibration to observed infection counts during the fall semester, we estimate the proportionality constant, α, and then $T = \alpha M$.

To estimate the contact matrix, we make the following additional assumptions:

- Each case identified through adaptive testing was generated by a source case in the same group.
- All positives in the student population created by an infectious student are identified as a close contact of that student (even if they were originally identified and tested because of surveillance testing, symptoms, or adaptive testing).

We first classify the student positives in the in-person semester (184 cases between 8/16/2020 and 11/24/2020) into source cases and secondary cases. Here, "secondary cases" include those identified via contact tracing or adaptive testing. The remaining cases, identified through surveillance testing, symptomatic self-reporting, arrival testing, or testing positive after returning from travel, are classified as source cases.

Based on these assumptions, we estimate the contact matrix using the following methodology:

- We begin by identifying all positive cases in each group i. Let this be $N(i)$.
- For each group j, we count the number of positives in group j that were identified as a close contact of a person in group i. Let this be $L(i,j)$. A positive who is a close contact of people in multiple groups is counted proportionally to the groups of the people that identified them as contacts. For example, a positive person in group 2 who is identified as a close contact of one person in group 1 and two people in group 2 would contribute $\frac{1}{3}$ to $L(1,2)$ and $\frac{2}{3}$ to $L(2,2)$.
- For each group j, we additionally count the number of positive people that were identified through adaptive testing but were not identified as a close contact. In an abuse of notation, let this be $L(j)$.
- The value of $M(i,j)$ is then $(L(i,j) + \mathbb{1}\{j = i\}L(j))/N(i)$.

We use identified contacts in producing these estimates. When contacts are not identified, this could distort the estimates. Assuming that contact tracing is equally effective for all source groups and "destination" groups, thus resulting in a constant fraction of contacts missed, the fact that we only use the matrix up to a multiplicative proportionality constant should ensure that the resulting error is controlled. The resulting contact matrix M is shown in Table S15.

Table S15. The contact matrix M. Cell $M(i,j)$ is the average number of positive cases in group j that an infectious individual in group i creates over the course of his or her own infection.

Source cases group (counts)	Average # positive contacts in Group 1	Average # positive contacts in Group 2	Average # positive contacts in Group 3
Group 1 (125)	$(81 + 11)/125 = 0.736$	$3.5/125 = 0.028$	0
Group 2 (44)	$1/44 = 0.023$	$(4.5 + 2)/44 = 0.148$	$1/44 = 0.023$
Group 3 (15)	0	0	$1/15 = 0.067$

Contact Tracing Effectiveness Parameters Our stochastic compartmental model does not track individuals. Instead, it tracks the number of individuals in a collection of different states. This makes it difficult to simulate contact tracing at an individual level. Instead, our model relies on the following two parameters:

1. `cases_isolated_per_cluster`: The number of positive cases isolated for each contact trace (which models both contact tracing and adaptive testing) initiated by a self-reporting symptomatic individual or one identified through surveillance testing.
2. `cases_quarantined_per_cluster`: The number of negative cases quarantined for each contact trace initiated by a self-reporting symptomatic individual or one identified through surveillance testing.

In the simulation, all infected individuals are considered to be isolated, even if we would not have known in reality that the individual was positive and would have initially placed them into quarantine.

`cases_isolated_per_cluster` corresponds to the average number of secondary cases identified through an initiated trace from a positive case in real life. This can be estimated from the ratio of the number of secondary cases (105) to the number of source cases (79), which gives 1.329. In comparison, the effective `cases_isolated_per_cluster` assumed in the projections for the fall is $0.85/2 = 0.43$, which is approximately $1/3$ of the calibrated value. This in part explains the conservative projections for the fall.

`cases_quarantine_per_cluster` can be estimated from the ratio of the number of negative cases identified in contact tracing (378) to the number of sources cases (79), which gives 4.785. Individuals identified in adaptive testing are not quarantined.

In summary, our estimated parameters are

- `cases_isolated_per_cluster`: 1.329;
- `cases_quarantined_per_cluster`: 4.785.

Initial Prevalence The model relies on an initial prevalence of free and infectious cases. The calibrated values are

- Group 1: 5.77 average initial cases;
- Group 2: 3.37 average initial cases;
- Group 3: 0.

For groups 1 and 2, we consider the initial free and infectious cases at the beginning of the simulation to be the union of those imported positive cases missed by the arrival test, and those secondary cases infected by arrival positives due to the lag between arrival and taking arrival tests.

This produces 11 cases, out of which 5 cases are in group 1, and 6 are in group 2.

Then, we estimate the number of imported positive cases missed by the arrival tests based on the number of arrival positives, the sensitivity of the arrival testing (assumed to be 90% for nasopharyngeal sampling PCR test) for individuals in the post-exposure pre-convalescence infectious period and the probability that an infected person is in the exposed state and thus not identifiable by a PCR test (estimated to be 0.2 based on state occupancy times in our model). Hence, for any positive case arriving in Ithaca, the probability that it is not identified by the arrival test is $P(\text{exposed state}) + P(\text{not in exposed state}) \cdot (1 - \text{sensitivity}) = 0.2 + 0.8 \times 0.1 = 0.28$. This implies that for every arrival positive case, there are $0.28/(1 - 0.28) = 0.39$ free and infected cases acting as the initial cases in the simulation. In more detail, $(\#\text{ observed cases}) = (1 - 0.28) \cdot (\#\text{ cases})$, and $(\#\text{ free and infectious cases}) = 0.28 \cdot (\#\text{ cases})$, so $(\#\text{ free and infectious cases}) = 0.28 \cdot (\#\text{ observed cases})/(1 - 0.28) = 0.39(\#\text{ observed cases})$.

Thus, the number of free and infectious cases created immediately are:

- Group 1: $0.39 \times 5 = 1.95$;
- Group 2: $0.39 \times 6 = 2.34$.

Third, we estimate the number of secondary cases resulting from the arrival positives, due to the fact that students did not take their arrival test right upon arrival and hence could infect other students during the testing delay. This is obtained based on the contact matrix M (as described above), assuming that each arrival positive in group j infects $M(i, j)$ individuals in group i.

We summarize the number of secondary cases in each group below:

- Group 1: $5 \times 92/125 + 6 \times 1/44 = 3.82$;
- Group 2: $6 \times 6.5/44 + 5 \times 3.5/125 = 1.03$.

In summary, the average number of initial cases in groups 1 and 2 are given below:

- Group 1: $1.95 + 3.82 = 5.77$;
- Group 2: $2.34 + 1.03 = 3.37$.

For group 3, since we did not observe its first positive case after 8/16/2020 until 9/12/2020, we set the initial prevalence to be zero.

Calibration Results We calibrate our model's projected infections to the actual trajectory within 3 subgroups from 8/16/2020 - 1/10/2021, as shown below. The total number of positive cases observed within the time period is described below and the trajectories are described in Figure S2.

- Group 1: 120, with 5 arrival positives excluded;
- Group 2: 38, with 6 arrival positives excluded;
- Group 3: 15.

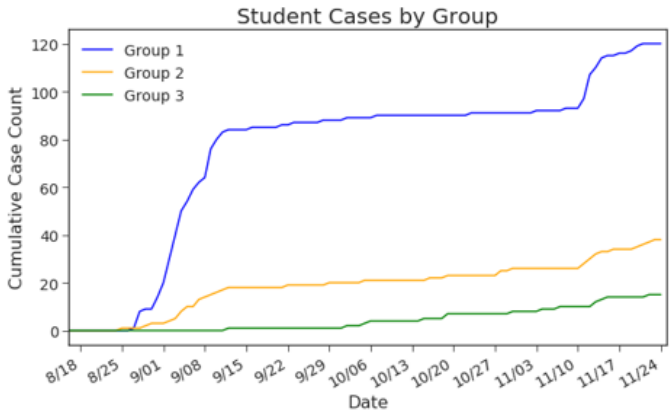


Fig. S2. Observed infections (excluding arrival positives) among students during the fall semester, shown for each of the three student groups.

Here we tune the parameter α in the simulation, i.e., the proportionality constant described in the contact matrix section above. For each value of α, we compute the mean squared error of the simulated results described as follows:

- Let $\text{sim}(t, i, j)$ denote the number of infections on day t in replication i for group j according to the simulation.
- Let $\text{actual}(t, j)$ denote the number of infections observed on day t for group j.
- Then, the error score associated with α is given by

$$\text{err}(\alpha) = \sum_{j \in \{1,2,3\}} \sum_{t=1}^{T} \left(\frac{1}{N} \sum_{i=1}^{N} \text{sim}(t, i, j) - \text{actual}(t, j) \right)^2 / T,$$

where N is the number of simulation replications and T is the simulation horizon.

Figure 5a in the main paper shows the log root mean-squared error of our model predictions versus α. We see that when $\alpha = 0.525$, the lowest score is obtained. Figure S3 shows the simulated trajectories (25 in each group) when $\alpha = 0.525$, in comparison to the actual trajectories for students cases in different subgroups.

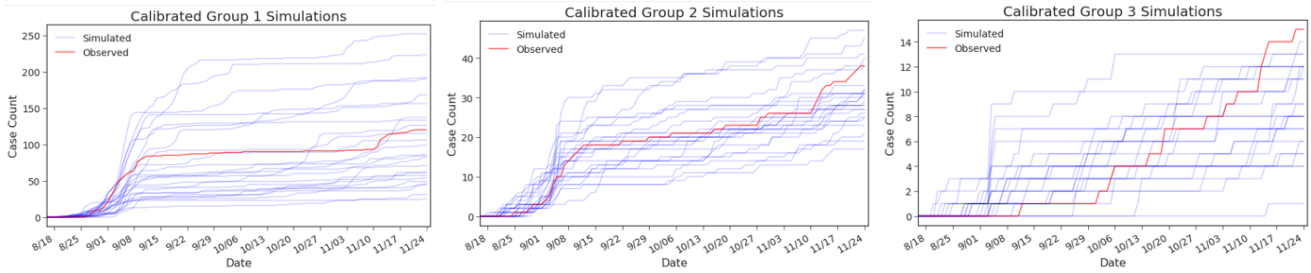


Fig. S3. Observed infection trajectories for each student group, over the course of the fall semester, plotted along with stochastic sample trajectories from the simulation under the estimated parameters.

B. Model Calibration for Employees. To calibrate our model for faculty and staff we use a single-group simulation model consisting of all faculty and staff with population size 10283, and include all infections that occurred between August 16, 2020 and January 10, 2021.

We have access to less detailed data about employees compared with students. In particular, we do not have access to contact tracing data for the fall semester. Understanding the difficulties of estimating inter-group transmission rates given a lack of contact tracing data, we elect not to partition the employee group (partitions considered included those based on county of residence or job type).

Observing rising infection counts among faculty and staff after Thanksgiving, we decide to include December and early January in the period of interest. We divide the time horizon into two non-overlapping periods: the pre-semester period (8/16/2020 - 9/2/2020) and the period after (9/3/2020 - 1/10/2021).

In place of contact tracing data, we leverage "`cluster_ids`" that were generated from manual review of employee cases. An employee case is assigned a `cluster_id` if that case is believed to be linked to at least one other case at Cornell, with all linked cases being assigned the same `cluster_id`. The use of the term "cluster" is perhaps misleading, since even pairs of positive cases that are linked through off-campus contact (often, two employees living together) are given a cluster id. These

cluster ids allow us to estimate outside infections and `cases_isolated_per_cluster`. In most cases, evidence suggests that those individuals without a cluster id were infected through non-Cornell interaction. This evidence, when it exists, consists of information obtained from contact tracing (e.g., that there is known close contact with a positive non-Cornell-affiliated individual) or the lack of other cases at Cornell at similar times in parts of the employee population that would interact with the positive individual on campus.

Here we describe the parameters estimated directly from fall 2020 data in the model calibration for employee group.

Testing Frequency 0 during pre-semester period; 0.098 per day after. (The latter value is an average across those tested once per week and those tested once every two weeks.)

Outside Infection Rate We classify a case as an "outside infection" if they did not contract the virus through interactions with other Cornell cases. (Transmission from one Cornell case to another is not considered an outside infection, even if the transmission occurred away from Cornell's campus.) To estimate the outside infection rate for Cornell employees (faculty/staff), we assume that

- Cases without `cluster_id`s are outside infections;
- Exactly one case in each identified cluster is an outside infection, while the remaining cases in the cluster are not outside infections.

Based on these two assumptions, we have a simple formula for calculating the number of outside infections: (# cases without a `cluster_id`) + (# clusters). Below we summarize the outside infection counts during the specified time period.

- 246 employee cases in total in the date range 8/16/2020 - 1/10/2021; 159 without a `cluster_id`; 25 distinct clusters.
- # outside infections = 159 + 25=184 (74.8%); # non-outside infections = 62 (25.2%).
- Average Daily outside infection rate: 184 / (# faculty and staff × 148 days) = 1.21E-4, i.e., in a population of 10,000 people, we should expect to see 1.2 infections per day due to travel and interaction with the outside community.

To address the rising trend in the number of employee cases, in the simulation we used a time-varying outside infection rate (measured in infections per day), which is computed by weekly faculty/staff outside infections divided by (# faculty and staff × 7 days). We assume that the outside infection associated with each `cluster_id` occurred during the week of the first identified case associated with that `cluster_id`.

Contact Tracing Effectiveness Parameters Recall that our simulation quantifies the effectiveness of contact tracing through a parameter, `cases_isolated_per_cluster`, which is the number of cases isolated for each cluster traced. Cluster traces are initiated by the discovery of a self-reporting symptomatic individual or by a case found via surveillance testing.

The number of positive cases isolated per contact trace is lower bounded by 0 and upper bounded by the average number of secondary positive cases per cluster. This is because it is only those cases in a cluster that can be linked through contact tracing. Here, we think of solo cases without a `cluster_id` as clusters of size 1.

To estimate this upper bound, we average (cluster size - 1) across all clusters. There are 25 identified clusters with size > 1, containing 87 cases in total. There are 159 cases without a `cluster_id`. Therefore, the average cluster size is (87 + 159) / (25 + 159) = 1.34, and avg (cluster size - 1) is 0.34.

We choose to use $0.34 \times 0.75 = 0.255$, assuming that 75% of the people found in clusters among Cornell employees were found via contact tracing or adaptive testing, with another 25% found via symptomatic self-reporting or surveillance testing. This is based, in part, on the observation that a large fraction of Cornell employee clusters are among family members and these would almost always be found via contact tracing. We assume that it is rare for positives in the Cornell community to be found via contact tracing of people who are not part of the Cornell community.

Thus, in summary, `cases_isolated_per_cluster` = 0.255.

Outcomes are insensitive to the parameter `cases_quarantined_per_cluster`, which determines the number of negative individuals quarantined, because its only effect on infections is to reduce the number of susceptible people that can be infected. Given that the fraction of the population quarantined is small, it has little effect on outcomes over several orders of magnitude. Because information about employee quarantines was unavailable, we set it to 2.5, a value close to half of the value observed for students, since employees were observed to have fewer contacts than students.

Calibration Results We calibrate our model's projected infections to the actual trajectory from 8/16/2020 - 1/10/2021, which is shown in Figure S4.

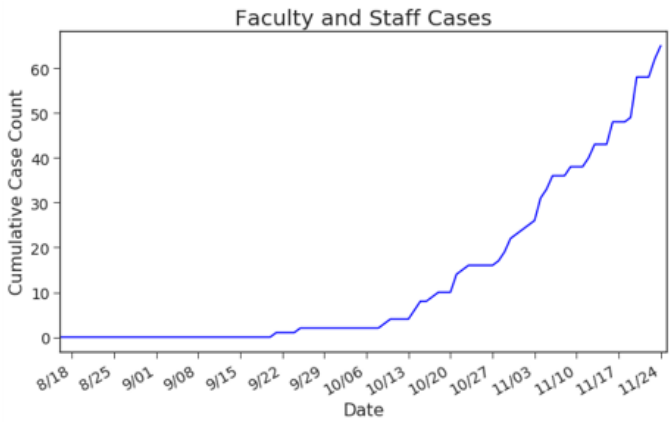


Fig. S4. Trajectory of employee infections from 8/16/2020 to 1/10/2021, the period used for calibration.

We then plot the root mean squared error (RMSE) between the observed trajectory and the average output of the simulation, versus the parameter we wish to calibrate, which is the daily transmission rate (# of other Cornell employees infected per day by a positive Cornell employee). Here, analogous to the error function used in the calibration for the student groups, the mean squared error is given by

$$\text{err}(\alpha) = \sum_{t=1}^{T} \left(\frac{1}{N} \sum_{i=1}^{N} \text{sim}(t,i) - \text{actual}(t) \right)^2 /T,$$

where $\text{sim}(t,i)$ is the number of infections on day t in replication i according to the simulation, $\text{actual}(t)$ is the number of infections observed on day t, N is the number of simulation replications, and T is the simulation horizon. A lower RMSE indicates a better fit. Note that many of these infections occurred between family members who are both Cornell employees but infected each other at home.

Figure 5b in the main paper shows the log RMSE versus employee transmission rate. We see in this figure that when the daily transmission rate is 0.11, the lowest RMSE is obtained. Thus, according to our calibrated model, each infectious employee infects 0.11 other employees on each day they are infectious.

Figure S5 shows 25 simulated trajectories when the daily transmission rate is 0.11, in comparison to the actual trajectory for faculty and staff cases. We observe that the observed case counts are reasonably well-represented by the simulation. Growth in cases during the semester is driven by an increase in outside infection rate rather than transmission within the Cornell population.

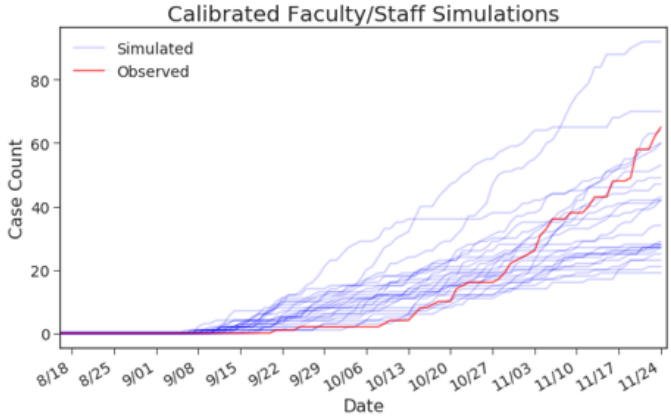


Fig. S5. Projections from our model (blue lines) using the calibrated daily transmission rate, compared with the observed infection trajectory (red line).

3. Parameter Uncertainty

This section presents our methodology for quantifying the effects of uncertainty in model parameters and additional results from applying this methodology not presented in the main paper.

We are specifically interested in the effect of parameter uncertainty on two outcomes: the safety of a residential semester as measured by the number of cases; and the relative safety of a residential semester compared to a virtual one, as measured by the difference in infections between these two instruction modes (residential infections - virtual infections). For both outcomes, a larger value is worse.

To quantify these effects, we

1. Identify a set of key parameters and their associated uncertainty to define a (joint) prior distribution. There are 12 key parameters that govern the number of residential infections and an additional 4 parameters that govern the number of virtual infections. These 16 parameters and their corresponding 95% credible intervals are summarized in Table S16.

2. Construct linear approximations of functions relating the input parameters to 1) the median number of residential infections, and 2) the difference in the median number of infections between residential and virtual instruction.

3. Using the geometry of the prior distribution and the linear approximations constructed in Step 2, identify a family of 1-dimensional parameter configurations with varying levels of pessimism. For each level of pessimism q, and each of the two outcomes (residential infections, residential - virtual infections) identify a set of parameter configurations whose median outcome is equal to the q-quantile of this outcome under the prior, as predicted by the linear approximation. Then, for each q, select as representative the configuration in this set with the largest density under the prior.

A. Parameter Scenarios. We adapt ideas from *robust optimization* (34) to address parameter uncertainty, with the goal of identifying and understanding the worst possible outcome over the parameter configurations.

We begin by constructing an *uncertainty set* derived from reasonable *ranges* for each parameter (see the "lower bound" and "upper bound" columns in Table S16). These ranges induce a natural central point in the space of parameter configurations, where each parameter takes the value at the midpoint of its range. We place a joint multivariate normal prior with independent marginals on the parameters with mean at the central point. We assume the range for each parameter given in Table S16 is a symmetric 95% credible interval, i.e., the true parameter value lies in this interval with 95% probability. More specifically, we define the following notation:

- $x \in \mathbb{R}^n$: vector of parameters; $n = 12$ for the residential case and $n = 16$ for the residential-virtual case.

- LB_i, UB_i: lower and upper bound of parameter i, as specified in Table S16. By assumption, (LB_i, UB_i) is a symmetric 95% credible interval for parameter i and parameters are mutually independent under the prior.

- $\Sigma = [\sigma_{ij}]$: the covariance matrix used in our multivariate normal prior. The components are specified by

$$\sigma_{ij} = \begin{cases} \sigma_i^2 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases}$$

  Each standard deviation σ_i is derived from the range (LB_i, UB_i) that we assume for parameter i. By virtue of assuming this range defines a 95% credible interval and assuming a normal prior, the range is related to the standard deviation by the equality $\frac{1}{2}(UB_i - LB_i) = 1.96\sigma_i$.

- $\mu = [\mu_i]$: the mean of our multivariate normal prior, as well as the central point of our parameter ranges. $\mu_i = \frac{1}{2}(LB_i + UB_i)$ for $i = 1, \ldots, 12$.

Next we consider the development of the linear approximations. As described above, we are interested in two outcomes. The first outcome is the number of cases in a residential semester. The second outcome is the number of residential infections minus the number of virtual instructions. In both settings, the outcome is worse at larger values.

To estimate the outcome over the parameter space, we sample 2000 parameter vectors using Latin hypercube sampling over the hypercube defined by all 16 ranges. For each parameter vector, we run 50 residential and virtual semester simulations and calculate the median value of the outcome of interest. We then construct a linear approximation, $c_0 + c^T x$, of the median, using linear regression on the corresponding 12 or 16 parameters of interest (x). The coefficients and standard error for each parameter in the linear regressions are presented in Tables S17 and S18.

To summarize uncertainty, we develop a one-dimensional family of parameter configurations associated with increasingly pessimistic outcomes. For each $y \in \mathbb{R}$, we consider the set $A(y)$ of parameter configurations whose expected outcome under the fitted linear model is equal to y. By construction, such sets $\{A(y), y \in \mathbb{R}\}$ are hyperplanes normal to c and partition the parameter space into two half-spaces. We find y^* such that the associated half-space, over which the expected outcome under the linear model is less than or equal to y^*, contains a prior probability mass of 0.99. We then determine the pessimistic configuration by selecting the representative point in $A(y^*)$ with the highest prior density. Figure S6 provides a visualization of this setup that may prove helpful in interpreting the following more precise explanation of how we summarize uncertainty.

For any $y \in \mathbb{R}$, the set of parameter configurations with expected outcome equal to y under the linear model is the hyperplane

$$A(y) = \{x : c_0 + c^T x = y\}.$$

Consider the half-space defined by this hyperplane over which the expected outcome under the linear model is less than or equal to y, $\{x : c_0 + c^T x \leq y\}$. Let $q(y) = P(c_0 + c^T X \leq y)$ be the prior probability mass in this half space, where $X \sim \mathcal{N}(\mu, \Sigma)$.

Parameter	Meaning	Lower Bound (LB)	Upper Bound (UB)	Justification for choice of range
Asymptomatic probability multiplier	Multiplier applied to $P(\text{asymptomatic})$ for each group, other severity levels scaled accordingly	24/47	70/47	CDC planning scenarios range: (15%, 70%), we use upper bound from here and our estimate of 47% for US population to define range.
Initial prevalence multiplier	Multiplier applied to initial prevalence	0.5	1.5	Base estimate uses 10x under-reporting rate. Estimate at the time for most states was 6-10x, with a max of 23x. Our reasonable aggregate estimate is 5-20x under-reporting.
R0	The baseline transmission rate of the disease is calibrated to an estimate of R0	1	4	CDC planning scenarios indicated the best guess was 2.5 (center) and pessimistic estimate was 4 (upper bound).
Outside infection multiplier	Multiplier applied to outside infection rate	0.5	1.5	Reasonable range representing our uncertainty.
Daily self-report probability	Daily probability that symptomatic person will self-report	0.22	0.5	Lower bound estimate from CDC time for seeking care for flu (11). This is likely pessimistic due to public awareness. A reasonable upper bound is people reporting symptoms within 2 days on average.
Contact Tracing multiplier	Multiplier on the effectiveness of contact tracing	1	2	Base estimate is from contact tracing effectiveness in Ithaca in the summer 2020. During the semester, we expect contact tracing to be slightly more effective (e.g. adaptive testing).
Contact Tracing testing ratio	Number of individuals identified per contact trace from surveillance testing positive relative to a symptomatic self-report	0.5	1.5	Our baseline estimate is that contact tracing should be as effective in both scenarios. We construct a reasonable range with center 1.
Test sensitivity	Sensitivity of surveillance tests	0.4	0.8	Reasonable range based on our group testing protocol.
Test non-compliance	Probability a surveillance test will be skipped	0.05	0.15	All students signed a behavioral compact, giving the university the ability to enforce compliance.
Exposed time (days)	Expected time in Exposed state	1	3	Reasonable range based on disease progression.
Infectious time (days)	Expected time in Infectious state	2	4	Reasonable range based on disease progression.
Symptomatic time (days)	Expected time in Symptomatic / Asymptomatic state	11	13	Reasonable range based on disease progression.
Persistent non-compliance	Percent of students in a virtual scenario who would not enroll in the surveillance testing program	0.25	0.75	Since Cornell would have limited enforcement options, many students may not inform the university that they have returned to Ithaca.
Intermittent non-compliance	In a virtual scenario, percent of scheduled surveillance tests that will be skipped	0.25	0.75	The percent of surveillance tests being skipped will be higher than under the residential scenario since the university has limited enforcement.
Virtual Transmissions per Day	Ratio of the transmission rate of students for the virtual scenario relative to the residential scenario	0.97	1.5	Due to the university's limited enforcement of masking, social distancing and testing, there will likely be an increase in transmission among students. The lower bound of 0.97 is explained in the virtual instruction section.
Virtual Population Size	The number of returning undergraduate and class based graduate students	0 (4500 UGs, 4770 GS other)	1 (7950 UGs, 5850 GS other)	Based on survey results of students. Lower bound is from proportions that replied "very or somewhat likely to return", upper bound students who answered "it depends / undecided".

Table S16. Parameter ranges. The first twelve are for the residential investigation; the last four are additional parameters for the virtual case. The last parameter, "virtual population size", is standardized to [0,1] which linearly interpolates between the lower and upper bounds.

Parameter	Linear coefficient	Std. Err	$P > \|t\|$	Coef × range	Pessimistic value
Regression Const	1014.7	429.2	0.018		
Asymptomatic prob multiplier	570.5	61.0	0.000	558.3	1.18
Initial prevalence multiplier	184.4	59.8	0.002	184.4	1.06
R0	409.1	19.9	0.000	1227.3	3.72
Outside infection multiplier	86.7	59.5	0.146	86.6	1.03
Daily self-report probability	-623.0	213.0	0.003	-174.4	0.34
Contact Tracing multiplier	-659.7	59.6	0.000	-659.7	1.28
Contact Tracing testing ratio	-571.2	59.6	0.000	-571.2	0.81
Test sensitivity	-1771.7	149.2	0.000	-708.7	0.51
Test non-compliance	1855.8	596.4	0.002	185.6	0.11
Exposed time (days)	-19.7	29.8	0.510	-39.3	1.97
Infectious time (days)	89.9	29.8	0.003	179.8	3.12
Symptomatic time (days)	-2.3	29.8	0.939	-4.6	12.0

Table S17. Fitted linear coefficient and computed pessimistic value for the residential instruction scenario.

Parameter	Linear coefficient	Std. Err	$P > \|t\|$	Coef × range	Pessimistic value
Regression Const	19870	827.5	0.000		
Asymptomatic prob multiplier	-4341.5	109.3	0.000	-4249.2	0.74
Initial prevalence multiplier	-119.8	106.9	0.263	-119.8	0.99
R0	-2493.7	35.6	0.000	-7481.1	1.11
Outside infection multiplier	6.1	106.7	0.954	6.1	1.00
Daily self-report probability	2645.8	380.6	0.000	740.8	0.37
Contact Tracing multiplier	951.1	106.6	0.000	951.1	1.56
Contact Tracing testing ratio	92.4	106.8	0.387	92.4	1.00
Test sensitivity	184.9	267.0	0.489	74.0	0.60
Test non-compliance	2769.7	1066.2	0.009	277.0	0.10
Exposed time (days)	-24.5	53.3	0.645	-49.1	1.99
Infectious time (days)	-538.2	53.2	0.000	-1076.4	2.87
Symptomatic time (days)	-261.6	53.3	0.000	-523.2	11.94
Persistent Non-compliance	-3474.3	213.4	0.000	-1737.1	0.45
Intermittent non-compliance	-2218.5	213.0	0.000	-1109.2	0.47
Virtual Transmissions per Day	-5522.9	201.0	0.000	-2927.1	1.14
Virtual Population Size	-1207.2	106.9	0.000	-1207.2	0.43

Table S18. Fitted linear coefficient and computed pessimistic value for residential - virtual infections.

Let y^* be such that $q(y^*) = 0.99$, so that y^* is such that the median outcome is less than y^* with prior probability 0.99. To find y^*, recall that $X \sim \mathcal{N}(\mu, \Sigma)$, so $c_0 + c^T X \sim \mathcal{N}(c_0 + c^T\mu, c^T\Sigma c)$, and then

$$\begin{aligned}
&P(c_0 + c^T X \le y^*) = 0.99 \\
\iff &P(\frac{c^T X - c^T\mu - c_0}{\sqrt{c^T\Sigma c}} \le \frac{y^* - c^T\mu - c_0}{\sqrt{c^T\Sigma c}}) = 0.99 \\
\iff &\frac{y^* - c^T\mu - c_0}{\sqrt{c^T\Sigma c}} = \Phi^{-1}(0.99) \\
\iff &y^* = \Phi^{-1}(0.99)\sqrt{c^T\Sigma c} + c^T\mu + c_0.
\end{aligned}$$

Let $x(y^*)$ be the point with the largest prior density in the hyperplane $A(y^*)$. We claim that $x(y^*)$ is the unique point in $A(y^*)$ lying on the line through μ in the direction Σc, that is $x(y^*) \in \{\mu + \lambda\Sigma c : \lambda \in \mathbb{R}\}$. Why? Maximizing the density over $A(y^*)$ is equivalent to minimizing the quantity $(x-\mu)^T\Sigma^{-1}(x-\mu)$ over all $x \in A(y^*)$, i.e., over all points x satisfying $c_0 + c^T x = y^*$. To find the optimum, define the Lagrangian $L(x;\eta) = (x-\mu)^T\Sigma^{-1}(x-\mu) - \eta(c_0 + c^T x - y^*)$; the optimum is characterized by the equation $\nabla_x L(x;\eta) = 0$, for some Lagrange multiplier $\eta \in \mathbb{R}$. The gradient of the Lagrangian is $\nabla_x L(x;\eta) = 2\Sigma^{-1}(x-\mu) - \eta c$, so the optimal point is given by $x(y^*) = \mu + \frac{\eta}{2}\Sigma c$, which is on the line through μ in the direction Σc as originally claimed.

We can thus find $x(y^*)$ as the unique point in the intersection of the hyperplane $A(y^*)$ and the ray $\{\mu + \lambda\Sigma c : \lambda \in \mathbb{R}\}$. We find that $\lambda = \Phi^{-1}(0.99)/\sqrt{c^T\Sigma c}$, and so the pessimistic point is given by

$$x(y^*) = \mu + \frac{\Phi^{-1}(0.99)}{\sqrt{c^T\Sigma c}}\Sigma c.$$

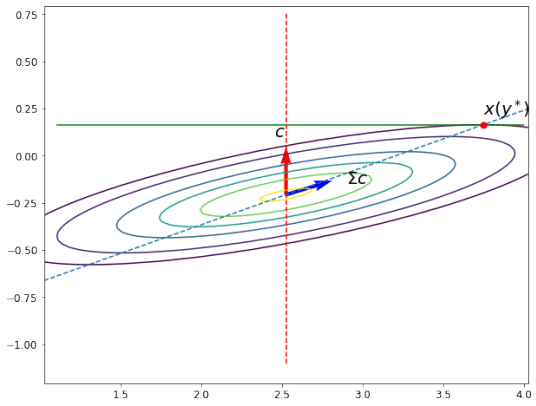


Fig. S6. Contours of the prior of the 12 parameters for the residential setting projected onto the space spanned by c (red arrow) and Σc (blue arrow). Without loss of generality, we align the vertical axis with the direction of c. The green line represents the hyperplane $A(y^*) = \{x : c_0 + c^T x = y^*\}$, which is perpendicular to c. The red dot represents $x(y^*)$, the unique point in $A(y^*)$ that lies on the line through μ in the direction Σc.

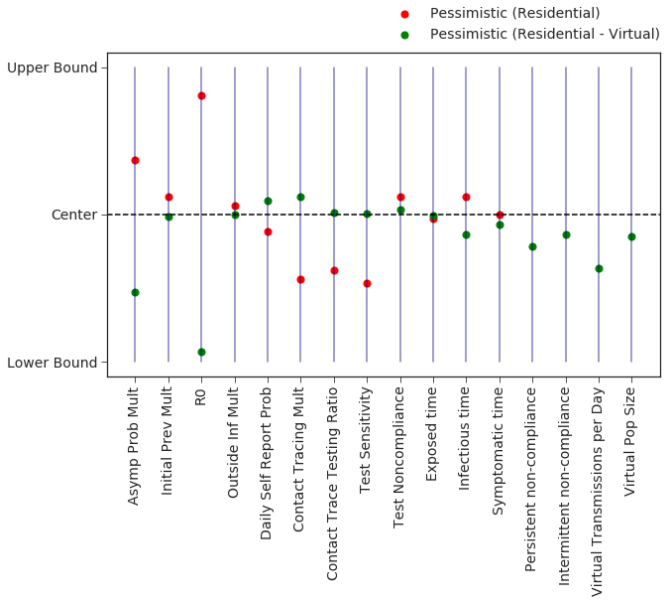


Fig. S7. Plot depicts the relative parameter values of both pessimistic scenarios.

Parameter	2020 Optimistic	2020 Nominal	2020 Pessimistic
Asymptomatic prob multiplier	1	1	1
Initial prevalence multiplier	1	1	1
R0	1.75	2.5	3.25
Outside infection multiplier	1	1	1
Daily self-report probability	0.18	0.18	0.18
Contact Tracing multiplier	1	1	1
Contact Tracing testing ratio	0.5	0.5	0.5
Test sensitivity	0.7	0.6	0.5
Test non-compliance	0.1	0.1	0.1
Exposed time (days)	2	2	2
Infectious time (days)	3	3	3
Symptomatic time (days)	12	12	12

Table S19. Parameter values for scenarios used in summer 2020 analysis. In this analysis the daily self-report probability should have been 0.22, but we used 0.18 due to a calculation error.

B. Scenarios from June 2020 report. As noted in Sections 1B and 1E of the paper, the nominal scenario reported here differs from the one reported in our June 2020 report (1). The 2020 nominal scenario was developed under time pressure and was intended to play a central role in the thinking of decision makers. It was therefore chosen to be somewhat conservative (meaning erring on the side of increased infections) with regard to a number of key parameters, especially contact-tracing parameters, as opposed to the nominal scenario presented here that is instead meant to represent our best estimate of the parameter values. Except for those key parameters, the 2020 nominal scenario resembles the nominal scenario reported here. The 2020 report also defined optimistic and pessimistic scenarios that, likewise, do not coincide with scenarios presented here. Table S19 lists the parameters for the scenarios explored in the 2020 report. See, also, Table S20.

C. Comparison of Prior to Calibrated Outcomes. Table S20 summarizes key parameter differences between nominal, pessimistic, Summer 2020 nominal and calibrated scenarios. The calibrated scenario includes parameter values that were directly estimated according to data from Fall 2020 or calibrated based on both our simulation model and data. Below we summarize how the 5 calibrated values compared to our prior range.

- Transmissions per day: The students with the highest transmission rate (Greek + Athletes) were within our prior range for transmissions. However, we overestimated the transmission rate for the remaining students.
- Cases found per contact trace: the effectiveness of contact tracing was very close to our nominal estimate.
- Initial prevalence: The students with the highest initial prevalence (Greek + Athletes) were within our prior range for initial prevalence. However, we overestimated the initial prevalence for the remaining groups.
- Outside infection rate: In the calibrated model, our definition for outside infection rate changed since we no longer explicitly modeled an Ithaca sub-population. Therefore, in the calibrated model an outside infection corresponds to any infection that originates outside the Cornell community. In all other scenarios, an outside infection refers to an infection from outside the Cornell or Ithaca community. Therefore, our prior range does not map conveniently to the calibrated definition.
- Test compliance for students: We underestimated the test compliance among students.

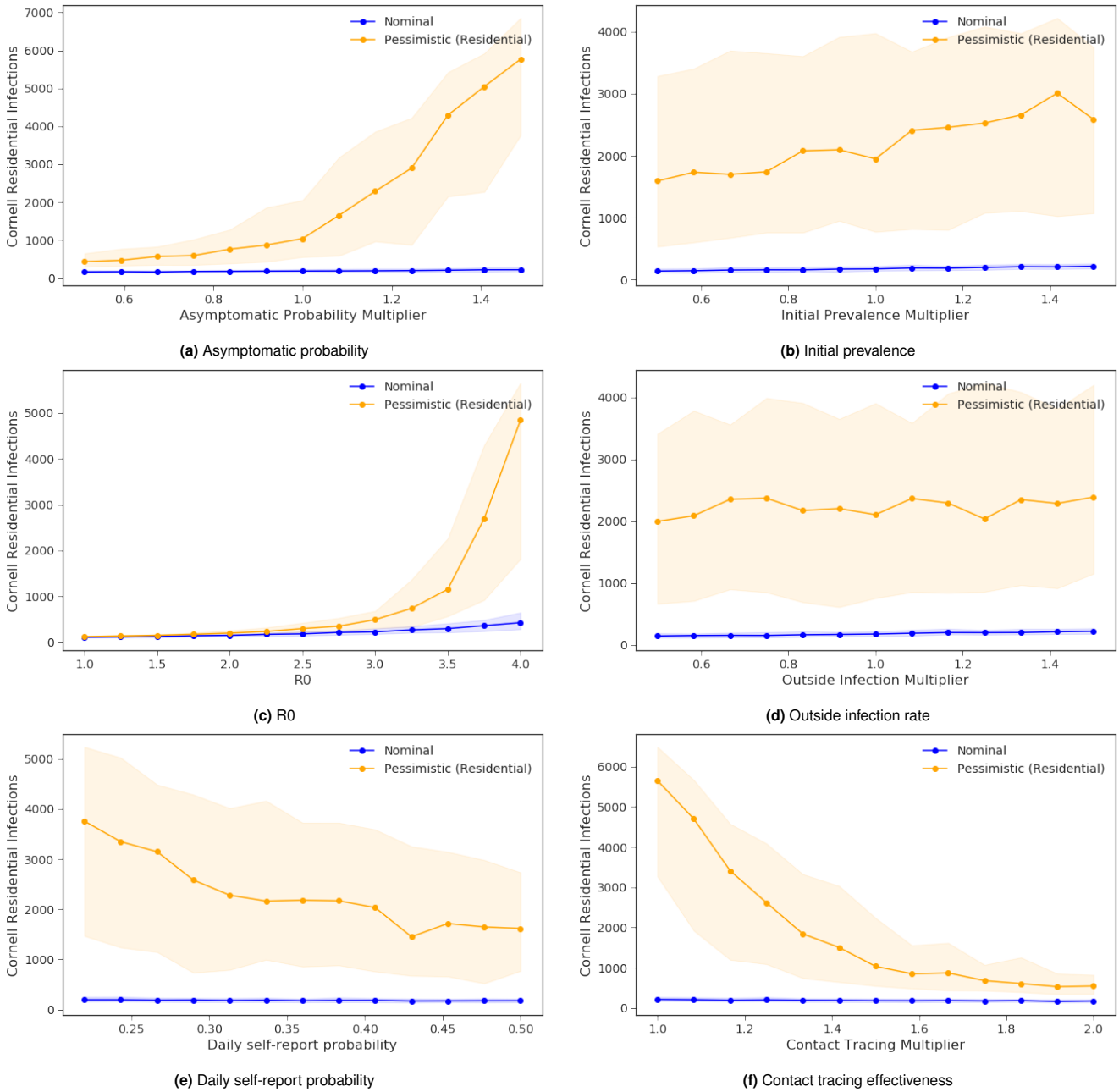
Since the groups changed between the uncertainty analysis and calibrated scenarios, some of the original 12 parameters in the uncertainty analysis are not appropriate for describing the calibrated scenario. For example, we used an outside infection multiplier to adjust all outside infection rates together in our uncertainty analysis. However, during our calibration, we arrived at group-specific rates which could not be mapped back to a single multiplier value. Therefore, we have replaced some of the 12 uncertainty parameters with new parameters that describe the same quantity (typically in different units).

As articulated in the faculty and staff calibration section, we assume that test compliance among this group is 1. This is because in the calibration for this group the testing frequency was directly estimated from data, which implies perfect compliance in the calibration simulations. Lastly, we used 0.18 as the daily self-report probability in summer 2020 scenarios because of a calibration error.

Parameter	Calibrated	Prior Range	Nominal	Pessimistic (Residential)	Summer 2020 Nominal
Transmissions per day	0.3742 (Greek + Athlete) 0.0867 (UG other) 0.0441 (GS) 0.11 (Faculty / Staff)	0.1217-0.4869 (UG Dorm) 0.0878-0.3512 (UG Off) 0.0528-0.2110 (GS research) 0.0726-0.2906 (GS class) 0.0705-0.2819 (FS student) 0.0328-0.1310 (FS not student) 0.0297-0.1187 (FS off)	0.3043 (UG Dorm) 0.2195 (UG Off) 0.1319 (GS research) 0.1816 (GS class) 0.1762 (FS student) 0.0819 (FS not student) 0.0742 (FS off)	0.4291 (UG Dorm) 0.3095 (UG Off) 0.1860 (GS research) 0.2561 (GS class) 0.2622 (FS student) 0.1219 (FS not student) 0.1104 (FS off)	0.3043 (UG Dorm) 0.2195 (UG Off) 0.1319 (GS research) 0.1816 (GS class) 0.1762 (FS student) 0.0819 (FS not student) 0.0742 (FS off)
Cases found per Contact Trace	1.329	0.92 - 1.84	1.38	1.214	0.92
Contact Tracing Testing ratio	1	0.5-1.5	1	0.84	0.5
Initial prevalence	0.163% (Greek + Athlete) 0.040% (UG other) 0 (GS + Faculty / Staff)	0.095% - 0.285% (UG + GS class) 0.0575% - 0.1725% (GS research) 0.04% - 0.12% (Faculty / Staff)	0.19% (UG + GS class) 0.115% (GS research) 0.08% (Faculty / Staff)	0.20% (UG + GS class) 0.121% (GS research) 0.084% (Faculty / Staff)	0.19% (UG + GS class) 0.115% (GS research) 0.08% (Faculty / Staff)
Asymptomatic probability multiplier	1	24/47-70/47	1	1.18	1
Outside infection rate	1.42×10^{-5} (Greek + Athlete) 7.11×10^{-6} (UG other) 6.45×10^{-6} (GS) Depends on local weekly Covid cases (Faculty / Staff)	$0.6 \times 10^{-5} - 1.8 \times 10^{-5}$	1.2×10^{-5}	1.27×10^{-5}	1.2×10^{-5}
Daily self-report probability	0.36	0.22-0.5	0.36	0.34	0.18
Test sensitivity	0.6	0.4-0.8	0.6	0.51	0.6
Test compliance	0.974 (Students) 1 (Faculty / Staff)	0.85-0.95	0.9	0.89	0.9
Exposed Time	2	1-3	2	1.97	2
Infectious Time	3	2-4	3	3.12	3
Symptomatic Time	12	11-13	12	12.0	12

Table S20. Summary of key parameter differences between calibrated, nominal, pessimistic, and Summer 2020 nominal scenarios. Blue indicates values calibrated directly to data and purple shows values calibrated via simulation. All remaining values are determined by assumption.

D. Sensitivity analysis for individual parameters. This section includes sensitivity analysis for model inputs. For the first 12 parameters, we show the sensitivity of residential infections and for the final 4 we show the sensitivity for virtual infections.



(a) Asymptomatic probability

(b) Initial prevalence

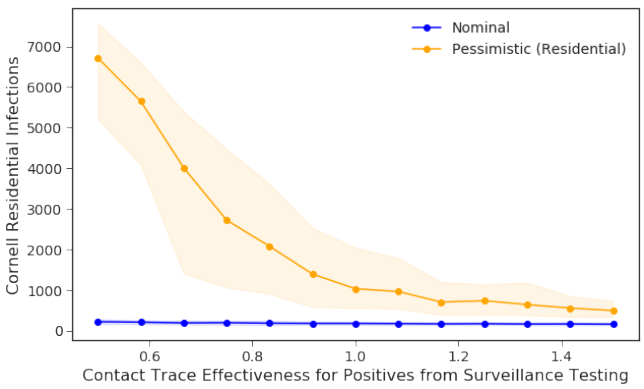
(c) R0

(d) Outside infection rate

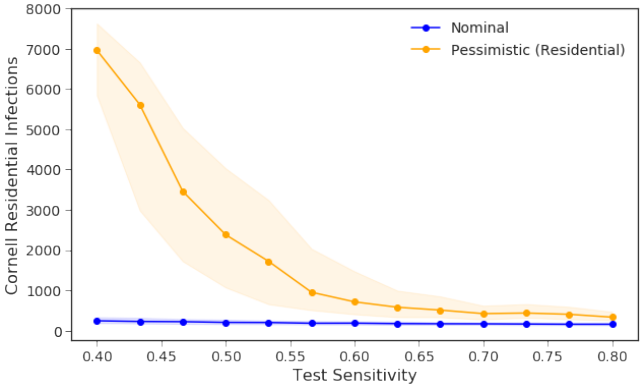
(e) Daily self-report probability

(f) Contact tracing effectiveness

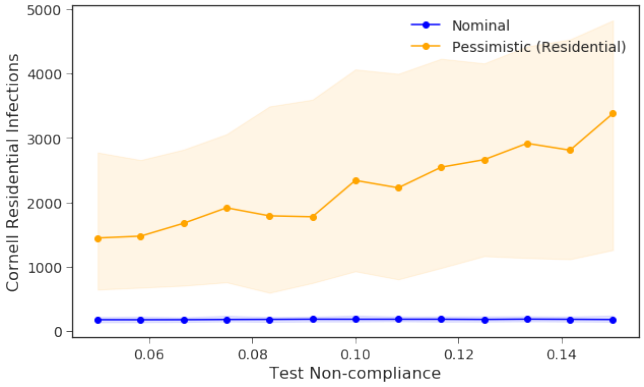
Fig. S8. Each plot depicts the 50th percentile of infections, with a wider range corresponding to the 10-90th percentile range, as the stated parameter varies, for both the nominal and pessimistic (residential) scenarios.



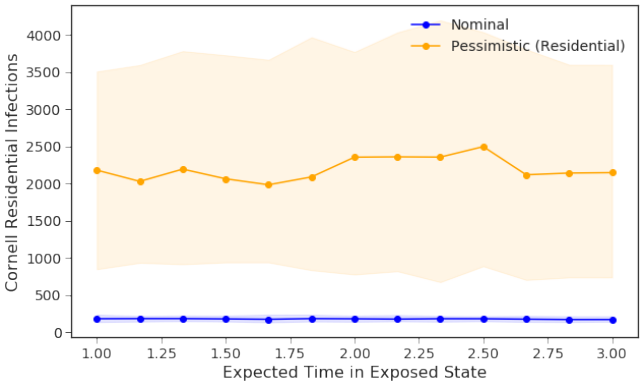
(a) Contact tracing effectiveness for surveillance testing positives



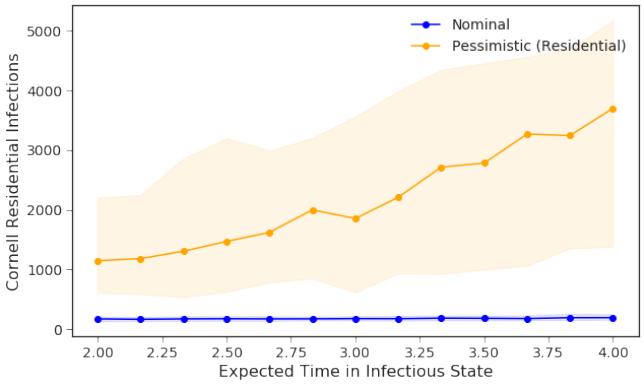
(b) Test sensitivity



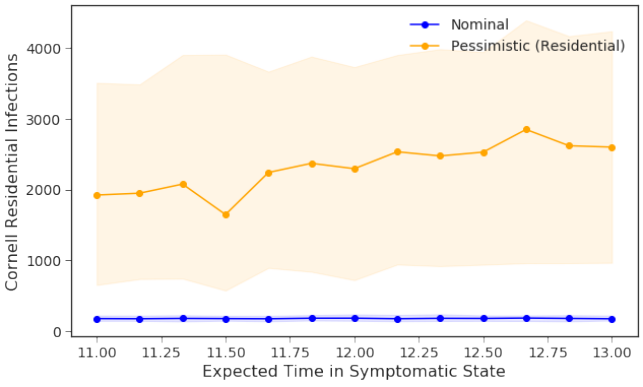
(c) Test non-compliance rate



(d) Expected time in Exposed state

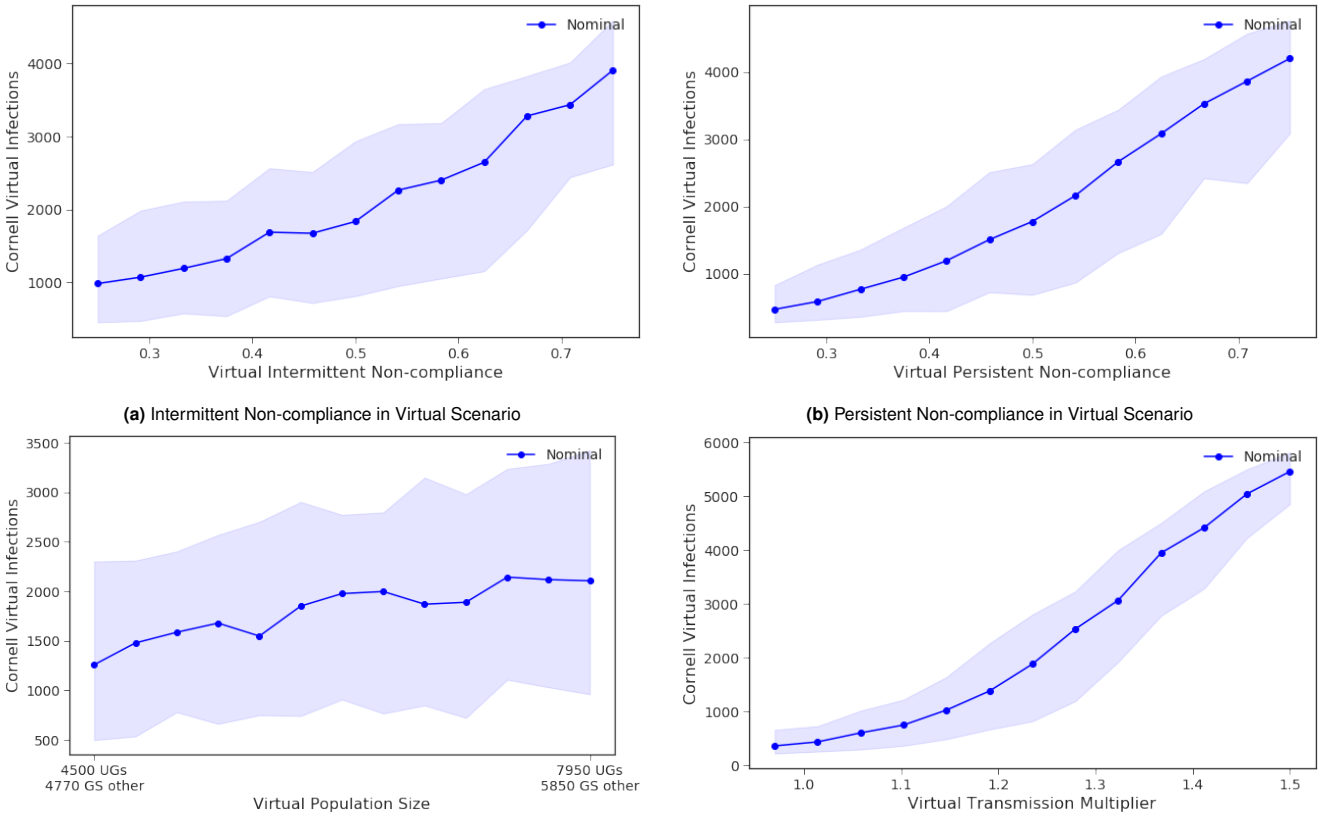


(e) Expected time in Infectious state



(f) Expected time in Symptomatic and Asymptomatic states

Fig. S9. Each plot depicts the 50th percentile of infections, with a wider range corresponding to the 10-90th percentile range, as the stated parameter varies, for both the nominal and pessimistic (residential) scenarios.



(a) Intermittent Non-compliance in Virtual Scenario

(b) Persistent Non-compliance in Virtual Scenario

(c) Population Size in Virtual Scenario

(d) Virtual Scenario Transmission Multiplier

Fig. S10. Each plot depicts the 50th percentile of virtual instruction infections, with a wider range corresponding to the 10-90th percentile range, as the stated parameter varies for the nominal scenario. Non-monotonicity is due to simulation error.

E. Correlation of Infection and Hospitalization metrics. In this section, we present graphs that demonstrate that the simulated number of Cornell infections is positively correlated with the number of Ithaca infections and Cornell and Ithaca hospitalizations. Due to this correlation, we use the number of Cornell infections as our primary metric.

In Figure S11, each point corresponds to a parameter vector sampled from the prior described earlier in this section.

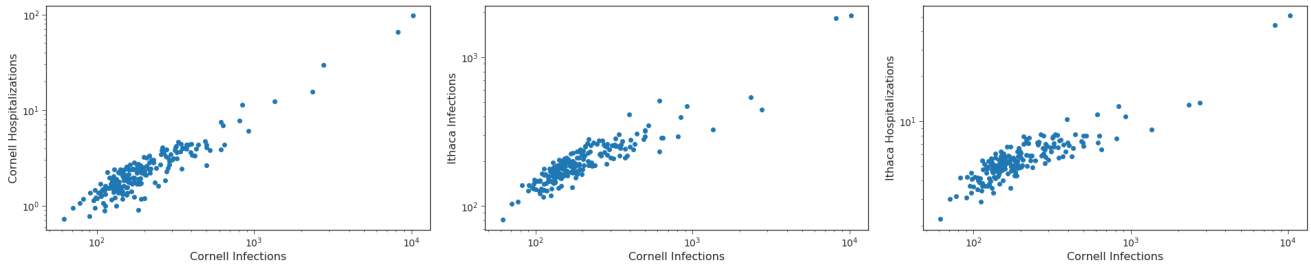


Fig. S11. Plot depicts the correlation of alternative metrics (Cornell hospitalizations, Ithaca hospitalizations, Ithaca infections) with the number of Cornell infections for parameter vectors sampled from the prior distribution.

References

1. JM Cashore, et al., Covid-19 Mathematical Modeling for Cornell's Fall Semester (https://covid.cornell.edu/_assets/files/covid_19_modeling_main_report.pdf) (2020).
2. JM Cashore, et al., Addendum: Covid-19 Mathematical Modeling for Cornell's Fall Semester (https://covid.cornell.edu/_assets/files/covid_19_modeling_addendum.pdf) (2020).
3. A Janmohamed, et al., Modeling Subpopulations in the Cornell Community and Greater Ithaca (https://docs.google.com/document/d/1gBoRecBRAbyONU7CgDV_rzkCVmrJJbTkSP2TpVBNl8g/edit#) (2020).
4. JM Cashore, et al., Mathematical Modeling for Cornell's Spring Semester (https://covid.cornell.edu/_assets/files/general-audience-spring-modeling-20210216.pdf) (2021).

5. SA Lauer, et al., The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: estimation and application. *Annals internal medicine* **172**, 577–582 (2020).
6. L Tindale, et al., Transmission interval estimates suggest pre-symptomatic spread of covid-19. *MedRxiv* (2020).
7. MM Arons, et al., Presymptomatic sars-cov-2 infections and transmission in a skilled nursing facility. *New Engl. J. Medicine* (2020).
8. WHO, Coronavirus disease 2019 (covid-19) situation report - 73 (https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200402-sitrep-73-covid-19.pdf) (2020).
9. WHO, Report of the who-china joint mission on coronavirus disease 2019 (covid-19) (https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf) (2020).
10. Centers for Disease Control and Prevention, Covid-19 pandemic planning scenarios (2020).
11. M Biggerstaff, et al., Influenza-like illness, the time to seek healthcare, and influenza antiviral receipt during the 2010–2011 influenza season—united states. *The J. infectious diseases* **210**, 535–544 (2014).
12. L Luo, et al., Modes of contact and risk of transmission in covid-19 among close contacts. *medRxiv* (2020).
13. China CDC, The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (covid-19)—china, 2020. *China CDC Wkly.* **2**, 113–122 (2020).
14. Y Dong, et al., Epidemiological characteristics of 2143 pediatric patients with 2019 coronavirus disease in china. *Pediatrics* (2020).
15. CDC COVID-19 Response Team, Severe outcomes among patients with coronavirus disease 2019 (covid-19)—united states, february 12–march 16, 2020. *MMWR Morb Mortal Wkly Rep* **69**, 343–346 (2020).
16. R Li, et al., Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (sars-cov-2). *Science* **368**, 489–493 (2020).
17. A Kimball, Asymptomatic and presymptomatic sars-cov-2 infections in residents of a long-term care skilled nursing facility—king county, washington, march 2020. *MMWR. Morb. mortality weekly report* **69** (2020).
18. Central Intelligence Agency, The world factbook - united states (2020).
19. Kaiser Family Foundation, Us population distribution by age (2018).
20. K Mizumoto, K Kagaya, A Zarebski, G Chowell, Estimating the asymptomatic proportion of coronavirus disease 2019 (covid-19) cases on board the diamond princess cruise ship, yokohama, japan, 2020. *Eurosurveillance* **25**, 2000180 (2020).
21. R Yang, X Gui, Y Xiong, Comparison of clinical characteristics of patients with asymptomatic vs symptomatic coronavirus disease 2019 in wuhan, china. *JAMA Netw. Open* **3**, e2010182–e2010182 (2020).
22. PopulationPyramid.net, 2019 china population (2020).
23. Cornell Institute for Research and Planning, Cornell university factbook (http://irp.dpb.cornell.edu/university-factbook/employees) (2019).
24. S McMullen, Covid-19 isolation and quarantine discussion with tompkins county health department (2020).
25. S Mallett, et al., At what times during infection is sars-cov-2 detectable and no longer detectable using rt-pcr-based tests? a systematic review of individual participant data. *BMC medicine* **18**, 1–17 (2020).
26. B Giri, et al., Review of analytical performance of covid-19 detection methods. *Anal. bioanalytical chemistry*, 1–14 (2020).
27. A Piras, et al., Inappropriate nasopharyngeal sampling for sars-cov-2 detection is a relevant cause of false-negative reports. *Otolaryngol. Neck Surg.* **163**, 459–461 (2020).
28. TC Jones, et al., An analysis of sars-cov-2 viral load by patient age. *MedRxiv* (2020).
29. J Mossong, et al., Social contacts and mixing patterns relevant to the spread of infectious diseases. *Plos Medicine* **5**, e74 (2008).
30. V Capraro, H Barcelo, The effect of messaging and gender on intentions to wear a face covering to slow down covid-19 transmission. *arXiv preprint arXiv:2005.05467* (2020).
31. W Edmunds, G Kafatos, J Wallinga, J Mossong, Mixing patterns and the spread of close-contact infectious diseases. *Emerg. themes epidemiology* **3**, 10 (2006).
32. H Hu, K Nigmatulina, P Eckhoff, The scaling of contact rates with population density for the infectious disease models. *Math. biosciences* **244**, 125–134 (2013).
33. Cornell COVID-19 Modeling Team, Mathematical Modeling for Cornell's Spring Semester, https://covid.cornell.edu/_assets/files/general-audience-spring-modeling-20210216.pdf (2021).
34. A Ben-Tal, A Nemirovski, Robust convex optimization. *Math. Oper. Res.* **23**, 769–805 (1998).