CovidU: A Model for COVID-19 Campus Decision Making*

Zagreb Mukerjee†

December 13, 2020

Abstract

This paper presents a model of COVID-19 on campuses, developed in summer 2020 to inform university decision-making. The discrete-time compartmental model expands a basic SIR-type model with representations of testing and isolation, contact between students, shared housing dynamics, social gatherings, and superspreader events. The paper demonstrates the utility of the model in three case studies. The first two cases are based on studies at Harvard (how many students to bring back, and the impact of social gatherings), finding that frequent mass testing makes the model campus relatively resilient to infection. In a third case, the model is used to estimate the realized R_0 at another university with an available dataset. The paper also presents interactive online software implementing this model for public use, with a novel automatic reporting feature (available at https://projects.iq.harvard.edu/covidu/app).

1 Introduction

In the summer of 2020, many universities confronted a pressing question: should students return to campus for the fall semester (Cai et al. 2020)? Campus environments, with shared living spaces and dense social networks, face particular risks of COVID-19 transmission (CDC 2020b). However, colleges and universities also have advantages in the fight against COVID-19: they can exercise a substantial degree of control over student populations, they often have dedicated medical facilities and staff, and they may have access to the specialized resources necessary for rapid or mass testing.

This paper grows out of an attempt to understand the conditions on campus, and applies that understanding to the problems faced by university decision-makers. In the summer of 2020, universities were balancing several priorities. The desire to provide a positive undergraduate experience, as well as the (often dire) need to maintain tuition revenue and keep enrollments high, were strong pressures in favor of reopening. Weighed against this were the obligations of the university to keep students, faculty, and staff safe, and to avoid spreading COVID-19 into their broader communities. This competing set of priorities was further complicated by a lack of knowledge. Some information from early outbreaks (such as those in Wuhan Province and on the Diamond Princess) was available, but the application to a campus environment was not straightforward.

*The paper, prepared in late Fall 2020, grows out of work I conducted staffing a summer 2020 Harvard committee tasked with amassing evidence for decisions about bringing Harvard's students back to campus for the fall semester. I worked with committee chair Gary King to design the model and accompanying software, with regular input from committee members Chris Avery, David Paltiel, James Stock, and Rochelle Walensky. With permission, I prepared this paper myself, to submit as a writing sample for graduate school applications; a more complete version, coauthored with committee members may appear in the future. The committee and I also appreciate the expert research and programming assistance provided by Olivia Fu and Davide Veronese.

†Research Data Scientist, Institute for Quantitative Social Science, 1737 Cambridge Street, Harvard University, Cambridge MA 02138; zagrebmukerjee.com, zagrebmukerjee@fas.harvard.edu

Our work was part of a Harvard effort to understand whether and how we could thread the needle and reopen safely for the fall. We built a model at Harvard University's Institute for Quantitative Social Science to assist Harvard administrators in deciding what the semester should look like. Our effort did not, and could not, encompass the entire utility function of the campus administrator - many priorities, such as the quality of the undergraduate experience, resist quantification. Rather, we sought to use epidemiological modeling and the best available data to clarify the mapping between policy decisions and an important subset of outcomes (such as case load, infection trajectory, and cost), so that the trade-offs involved could become clearer. The paper describes a generalization of the model for application to campuses and other settings. It also offers three specific case studies to illustrate the use of this model, and showcases interactive online software that allows any user to run this model for themselves.

The model in question is a discrete-time compartmental model, similar to SIR-type models frequently used by epidemiologists. It extends a similar model built by Paltiel, Zheng, and Walensky (2020). The model is designed to be as simple as possible, while capturing several important dynamics: mass testing, an isolation system, shared housing, social gatherings, superspreader events, and transmission from outside the ecosystem.

Of the three use cases presented, two of them re-create research done to support Harvard decision-making. First, as administrators decided how many students to bring back, we offered assistance in assessing the relationship between population size, testing cost, and spread of infection (particularly with respect to shared housing). We sought to address multiple competing ends, examining the epidemiological effects of increasing the population but also the financial burden of mass testing. Second, we submitted a report to the administration describing the potential effects of social gatherings on campus, arguing that even a substantial degree of student partying could be counteracted by a robust testing regime. We believed this to be a low-downside way to improve the qualitative student experience. The third case examines the trajectory of COVID-19 at the University of Illinois at Urbana-Champaign (UIUC). We use the observed outcomes, combined with the model, to estimate a range of realized values of the important R_0 parameter.

The interactive software detailed below allows others to build on our efforts and visualize the differing effects of a wide range of campus policy decisions. The CovidU app includes functionality for defining a custom set of parameters; visual and numerical summaries of results; it also contains causal effect functionality, which allows users to define treatment and control scenarios and examine the differences (with novel automatic text generation to summarize treatment and effect). We hope that this app, combined with the case studies above, will help interested parties to conduct their own research into COVID-19 transmission on college campuses, even without a background in modeling or epidemiology.

2 Model

The model describes the spread of COVID-19 to and between students on a college campus, as well as efforts to contain the spread through mass testing and isolation. The student population is divided into several compartments, as in a Susceptible/Exposed/Infected/Recovered model. Students move between the compartments at discrete time steps, as dictated by infection, progression of the disease, testing, and isolation.

The model considers the entire population without differentiating specific individuals. Individual-level probabilities of contact and infection are aggregated into by population-level movements between compartments (representing expected values). The model also assumes homogeneous mixing: with the exception of shared housing and social gatherings (as described below), it is assumed that every student interacts equally with every other student. COVID-19 tests in the model are conducted and returned within 8 hours, regardless of time of day.

For simplicity, the model assumes that reinfection is impossible, that symptomatic students can be immediately identified and isolated, that asymptomatic and symptomatic students recover at the same rate, and that isolation is 100% effective. These latter assumptions are made for simplicity, and are easily relaxable. The model is run for 80 days (D), with 3 time steps per day (θ).

A list of parameters, along with definitions, can be found in Appendix A. Some parameters used by the model are computed from more easily comprehensible inputs: for example, a reproduction number, R_0, and a probability of infection conditional on contact, p, are combined to compute an average number of contacts per student c_n. Many parameters are described in terms of days, but then converted to time-steps for the use of the model. These calculations can be found in Appendix A.

2.1 Overview

The model divides the population into eight compartments along two dimensions: whether individuals are isolated or not, and how far they have progressed in the disease. These compartments are described in Table 1, and their relationships are depicted in Figure 1:

	Not Isolated	**Isolated**
Uninfected	Susceptible (S_t)	False Positives (FP_t)
Infected/ Not Infectious	Exposed (E_t)	
Infectious/ No Symptoms	Non-Isolated asymptomatic (NA_t)	Isolated asymptomatic (IA_t)
Infectious/ Symptoms		Symptomatic (SY_t)
No Longer Infected	Recovered (ρ_t)/ Deceased (D_t)	

Table 1: Model Compartments

All students start as Susceptible, except for some number that represent the initial source of infection. Susceptible students are those who have not yet been infected with SARS-CoV-2. Exposed students have been infected, but the virus is incubating, meaning they are not yet infectious, cannot be detected by tests, and cannot develop symptoms. Asymptomatic students are infectious, and may develop symptoms. Some asymptomatic students are detected by testing and isolated. Symptomatic individuals are those who have developed symptoms. As mentioned above, it is assumed that all symptomatic students are immediately quarantined. Recovered students have had Covid-19 and are now neither infectious nor infectable.

Movement between these compartments happens in three ways: 1) Infection; 2) Disease Progression; 3) Testing and Isolation. Infection causes susceptible students to become exposed, and happens in several ways detailed below. Disease progression represents the development of COVID-19 through several stages, from initial exposure to asymptomatic/symptomatic cases, to recovery or death. Testing procedures isolate students from the population (correctly or incorrectly) to prevent further infection, and return them as appropriate.

2.2 Infection

Infection is the mechanism of movement from the Susceptible to the Exposed compartment, representing the spread of COVID-19. It can happen in one of several ways:

1. Through community transmission: A susceptible student contacts an asymptomatic student, either as

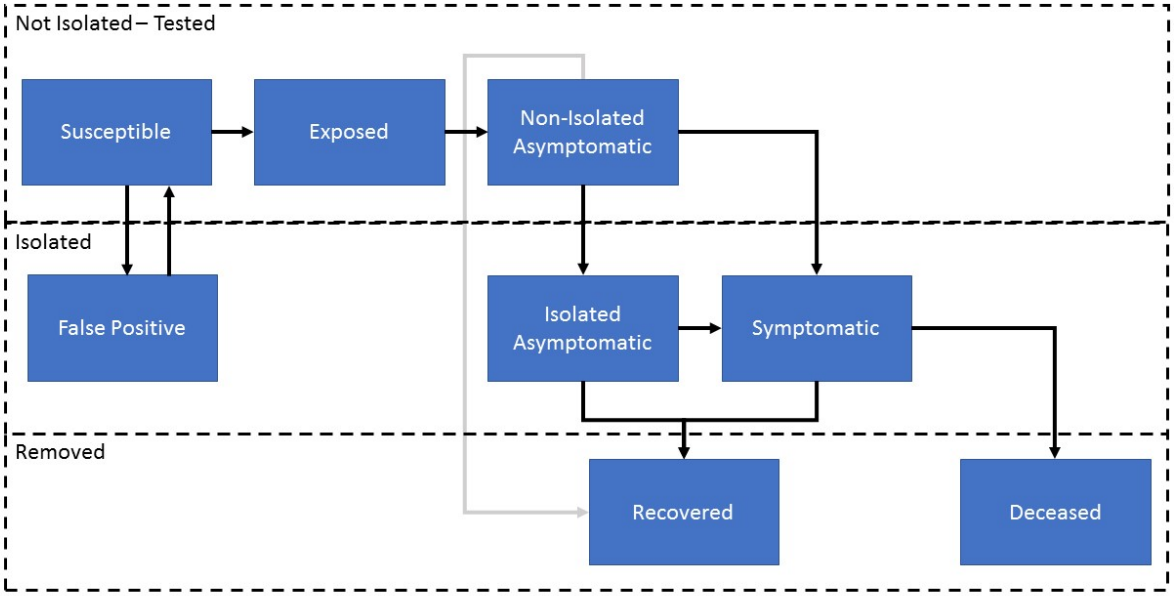


Figure 1: Model Flow

part of their day-to-day routine, in their dormitories, or at a social gathering, and receives a sufficient viral load to be infected.[1]

2. Through external transmission or so-called superspreader events, both of which are represented by infections created without a specific source.

Community transmission is represented by a contact matrix; this allows for the expression of different forms of interaction. As an example, consider a three-student population. All three students interact with each other normally, creating some degree of contact c_n; in addition, Students A and B are roommates, creating additional contact c_r. So students A and B have $c_r + c_n$ contacts, while C has c_n contacts with each of A and B. We use this to create a symmetric contact matrix, as seen in Table 2.

	Student A	Student B	Student C
Student A	–	$c_r + c_n$	c_n
Student B	$c_r + c_n$	–	c_n
Student C	c_n	c_n	–

Table 2: An example contact matrix

The degree of increased contact between students sharing rooms, bathrooms, or other housing, c_r, is based on a parameter, Probability of Shared Housing Infection (P_r), and applies to each of n_r students sharing housing. We think of P_r as analogous to a household secondary attack rate, and compute the degree of additional contact that would correspond to the given attack rate.

1. We think of "contacts" consistent with our parameter p, probability of infection given contact (Luo et al. 2020). The CDC uses a brief face-to-face interaction as a benchmark contact (CDC 2020a).

The baseline degree of contact c_n corresponds to normal interaction between students (e.g. in dining halls or student centers). It is entirely based on the R_0 parameter, combined with the time-invariant parameter p, the probability of infection conditional on contact with an infectious individual.[2] We assume students mix homogeneously, so we can divide that average across the entire population (for more details on these calculations, see Appendix A).

In addition, our model has a third source of student contact - social gatherings. A certain proportion of the population (n_{soc}) attends social gatherings at each time period. Each attendee experiences some additional degree of contact, c_{soc}.

For our current model, we are only interested in the average contacts per student c_{tot}. Given our assumption of homogeneous mixing we can compute this using the total population n_{pop}:

$$c_{tot} = (c_n n_{pop} + c_r n_r) + n_{soc} c_{soc} / n_{pop} \tag{1}$$

We use p to compute infection from aggregate contacts at each time step. If Student A is infectious in the contact matrix of Table 2, Student B has a $p(c_r + c_n)$ probability of having been infected, and Student C has a probability of pc_n. For simplicity, we suppose that infection probability accrues additively, so if both Student A and Student C are infectious, then Student B has a $p(c_r + 2c_n)$ probability of being infected.[3]

The first infectious student dropped into a fully susceptible population would thus contact c_{tot} other students and create pc_{tot} infections. Later infectious students will generate fewer infections, given that some of their contact is with students who are already infected. Thus the infections generated by each infectious student scale with S_t, the current susceptible population, as a share of the non-isolated population ($S_t + NA_t + E_t$). Our model assumes the symptomatic students are immediately isolated; thus, the only infectious students are the students in group NA_t. This lets us compute the total number of new infections from community transmission $I_{c,t}$:

$$I_{c,t} = \frac{S_t}{S_t + E_t + NA_t} NA_t \, p \, c_{tot} \tag{2}$$

The value in Equation 2 is augmented by adding a flat rate representing external infections $I_{e,t}$ (based on daily value $I_{e,d}$), as well as superspreader events I_s, occasions of brief but intense infectious activity, such as mass gatherings (Frieden and Lee 2020). We define an indicator function $F(t)$ as 1 if a superspreader event takes place at time t, 0 otherwise.

$$I_{tot,t} = I_{c,t} + I_{e,t} + F(t) I_s \tag{3}$$

2.3 Disease Progression

Our model captures the progression of COVID-19, from initial exposure, to asymptomatic and perhaps symptomatic infectiousness, to either recovery or death. The time spent in each stage is defined by a parameter, a number of days, which is then turned into a rate of movement between the populations. So if it takes 5 days for exposure to develop into asymptomatic infection, we represent that as $\frac{1}{5}$ of the exposed population becoming asymptomatic daily.

2. R_0, the basic reproduction number, is the number of cases generated by an infectious individual introduced to a susceptible population. It is a property of the virus and of social structure. If students were to be interacting as normal, R_0 would be higher; if they were to be under social distancing restrictions, that would correspond to lower R_0. This is discussed in more detail in Section 5.

3. For very large values of p or R_0, this may overstate the new infections generated by an infectious student.

The exposed population becomes asymptomatic at an incubation rate r_i, based on a time to incubation parameter t_i; it is at this point that they become infectious (as well as testable; more on that below).

$$NA_{new,t} = r_i E_t \tag{4}$$

Upon becoming asymptomatic a student either recovers, or develops symptoms. The cumulative proportion developing symptoms s is used to compute a symptom development rate r_s (see Appendix A).

$$SY_{new,t} = r_s(NA_t + IA_t) \tag{5}$$

The asymptomatic and symptomatic populations can recover; the symptomatic populations can die. The rates of recovery r_r (based on the time to recovery parameter t_r) and death r_δ (based on a conditional mortality parameter, δ) are inputs. For simplicity, the rate of recovery is the same regardless of the degree to which an individual experiences symptoms.

$$\rho_{new,t} = r_r(NA_t + IA_t + SY_t) \tag{6}$$

$$D_{new,t} = r_\delta SY_t \tag{7}$$

2.4 Testing and Isolation

Testing and isolation represent the tools of the campus administration to control the spread of COVID-19; the non-isolated population is tested, and positive tests result in isolation (as does symptom development).

The testing represents a PCR or other diagnostic test, with the following parameters: specificity Sp (probability of negative test if actually negative); sensitivity Se (probability of positive test if actually positive); delay between test administration and isolation of students with positive results d; testing cadence τ (based on τ_d, in daily terms). Students can test as "true positives" if and only if they are infectious.

The Susceptible and Exposed students are tested. Asymptomatic students are divided into Non-Isolated Asymptomatic and Isolated Asymptomatic; the former are tested. These groups (not differentiable to the administration) are tested at a rate determined by the cadence parameter; if the testing cadence τ_d is every 7 days, we represent this as $\frac{1}{7}$ of each group being tested each day. For simplicity, it is assumed that Exposed students cannot test positive. Recovered students are not tested.

There are two sets of positive tests:

$$IA_{new,t} = (1/\tau)Se\, NA_{t-d} \tag{8}$$

$$FP_{new,t} = (1/\tau)(1 - Sp)\, S_{t-d} \tag{9}$$

The first group represents "true positives", who are correctly isolated. The second group is false positives. False negatives are members of NA who remain in NA. Once isolated, students are administered a confirmatory test, which is assumed to be 100% specific. The false positives are released after a certain amount of time r_{fp} (computed based on a number of days t_{fp}). The true positives progress as per "Disease Progression" above, but are not sources of transmission in the infection step.

$$FP_{ret,t} = r_{fp} FP_t \tag{10}$$

The cost of each test is π. The confirmatory test costs π_c. The total testing cost in each period C_t is given by

$$C_t = \pi(1/\tau)(S_t + E_t + NA_t) + \pi_c(IA_{new,t} + FP_{new,t}) \tag{11}$$

The first term represents the regular testing of all students who are not isolated; the second term represents the confirmatory testing administered when entering isolation.

2.5 Model Summary

We can combine the three mechanisms above with the eight compartments to create the governing equations that determine the movement of the model at each step.

The Susceptible population shrinks with new infections (Equation 3), and with new false positives (Equation 9). It grows as false positives are released (Equation 10).

$$S_{t+1} = S_t - I_{tot,t} - FP_{new,t} + FP_{ret,t} \tag{12}$$

The False Positive population grows and shrinks as members of the Susceptible population are incorrectly quarantined, and released after retesting (Equations 9 and 10).

$$FP_{t+1} = FP_t + FP_{new,t} - FP_{ret,t} \tag{13}$$

The Exposed population grows as new Susceptible students are infected (Equation 3), and shrinks as the Exposed students become Asymptomatic (Equation 4).

$$E_{t+1} = E_t + I_{tot,t} - r_i E_t \tag{14}$$

The population of Non-Isolated Asymptomatics - the main vectors of community transmission - grows as more Exposed students become contagious (Equation 4), and shrinks as testing detects asymptomatic carriers and sends them to isolation, with a delay (Equation 8).

$$NA_{t+1} = NA_t + r_i E_t - (1/\tau) Se\, NA_{t-d} \tag{15}$$

The Isolated Asymptomatics can recover (Equation 6) or develop symptoms (Equation 5):

$$IA_{t+1} = IA_t + (1/\tau) Se\, NA_t - (r_r + r_s) IA_t \tag{16}$$

The Symptomatic can recover (Equation 6) or die (Equation 7):

$$SY_{t+1} = SY_t + r_s(NA_t + IA_t) - (r_r + r_\delta) SY_t \tag{17}$$

and the Recovered (Equation 6) and Deceased (Equation 7) are terminal states.

$$\rho_{t+1} = \rho_t + r_r(NA_t + IA_t + SY_t) \tag{18}$$

$$D_{t+1} = D_t + r_\delta SY_t \tag{19}$$

3 Case 1: Deciding on the Fall 2020 Harvard Population

Harvard, along with many universities, closed abruptly in the spring of 2020, sending students home to avoid an outbreak of COVID-19. Over the next several months, the university was able to assemble substantial resources, including dedicated testing capacity through the Broad Institute and student isolation capabilities. Thusly armed, administrators decided to bring at least some students back for the fall semester - but how many?

The Subcommittee for Evidence-Based Decision Making used the above model to inform thinking on how many students should be brought back. As discussed above, the campus decision maker has a complex utility function, with conflicting pressures such as student safety, testing and isolation costs, and the quality of the student experience. Our model did not explicitly net these factors, but we sought to describe the range of potential outcomes given policy choices. The model was populated with plausible parameters based on guidance from the Centers for Disease Control and Prevention (CDC 2020c), as well as the best available literature at the time. The model then compared several permutations of student population and testing frequency, with conservative epidemiological assumptions. The subcommittee found that frequent testing made the campus relatively robust to outbreaks, even with the degree of shared housing necessitated by a larger student population.

The following is a reconstruction of that work; the model has evolved since then, with the addition of some features and the correction of some errors, but the substantive findings are the same as those generated by the Subcommittee.

3.1 Parameters

The crucial parameter R_0 (base infection rate), as well as s (proportion developing symptoms) and δ (mortality), were derived from moderate to pessimistic CDC planning scenarios (CDC 2020c). Expected properties of PCR testing were given to us by Harvard administrators and the Broad Institute (responsible for the testing).[4]

Table 3 describes the parameters used for the Harvard population, and sources for the parameters where applicable. The parameters are defined in Appendix A.

4. It is worth noting that the R_0 parameter describes, among other things, expected levels of student contact: a greater value of R_0 can describe a scenario in which social distancing rules are either not in place, or widely ignored

Disease Properties			
Parameter Name	Notation	Value(s)	Source
Basic Reproduction Number	R_0	2.5/4.0[5]	CDC (2020c)
Probability of Infection Given Contact	p	2.6%	Luo et al. (2020)
Prob. of Shared Housing Infection	P_r	15%	Avg of estimates in Bar-On et al. (2020)
Incubation Time	t_i	3	Li et al. (2020); He et al. (2020); Lauer et al. (2020)
Recovery Time	t_r	14	Lauer et al. (2020); CDC (2020d)
Proportion Developing Symptoms	s	30%	CDC (2020c)
Mortality Rate if Symptomatic	δ	.05%	CDC (2020c)
Testing			
Test Specificity	Sp	90.0%	Broad Institute[6]
Test Sensitivity	Se	80.0%	Broad Institute
Cost per Test	π	25$	Harvard Assumption
Confirmatory Test Cost	π_c	50$	Harvard Assumption
Testing Cadence	τ_d	2;4[7]	Quantity of Interest
Time to release false positives	t_{fp}	1	Harvard Assumption
Population			
Student Population	n_{pop}	1000/ 3400/ 6600	Quantity of Interest
Students Asymptomatic at Beginning	NA_0	3/ 10.2/ 19.8	Assumption; fixed proportion of n_{pop}
Avg. Number Sharing Housing	n_r	1/ 5/ 10	Computed from n_{pop}
External Infections per Day	I_{ed}	0.1/ 0.34/ 0.66	Based on Cambridge, MA cases from City of Cambridge (2020), combined with n_{pop}

Table 3: Parameters for Harvard

3.2 Results

We examined 12 scenarios that varied on 3 dimensions: number of students (1000, 3400, or 6600, given by the administration based on potential housing plans); testing cadence (2 or 5 days); and R_0 (2.5 or 4, per the

5. Used to compute number of student contacts.

6. Our understanding of these tests has evolved since the summer. The values in this table reflect a pessimistic estimate as of Summer 2020. In Section 5, we use more current values.

7. A τ_d of 2 means that tests occur every other day.

CDC baseline and worst-case scenarios).[8] We varied the number of starting asymptomatic students along with the population. A major implication of the population change was that more students needed to share living space (particularly bathrooms) if more students were permitted to return to campus.

For each scenario, we computed and presented several outputs of the model. In Table 4 we showcase some of the most important metrics. We also presented the results graphically, as can be seen in Figure 2. These findings were originally generated as a series of pairwise comparisons of detailed scenarios (using the reporting tool described in Section 6, but are shown here all at once in summary form. We suggested that scenarios be evaluated not only on summary statistics such as total students infected but also on trajectory: an upwardly-sloping trajectory suggested a less controlled spread of SARS-CoV-2, and potential risk of an exponential outbreak.

Scenario Population/R_0/testing cadence	Total Symptomatic Cases	Cases per 1000	Total Testing Cost
R_0 of 2.5 and testing every other day			
1000/ 2.5/ 2 days	6.3	6.3	\$1,100,375
3400/ 2.5/ 2 days	26.7	7.9	\$3,729,895
6600/ 2.5/ 2 days	73.5	11.1	\$7,197,959
R_0 of 4.0 and testing every other day			
1000/ 4.0/ 2 days	14.7	14.7	\$1,084,315[9]
3400/ 4.0/ 2 days	77.1	22.7	\$3,643,564
6600/ 4.0/ 2 days	301.9	45.7	\$6,870,971
R_0 of 2.5 and testing every 5th day			
1000/ 2.5/ 5 days	21.0	21.0	\$446,234
3400/ 2.5/ 5 days	172.8	50.8	\$1,480,554
6600/ 2.5/ 5 days	1030.3	156.1	\$2,656,707
R_0 of 4.0 and testing every 5th day			
1000/ 4.0/ 5 days	236.4	236.4	\$366,038
3400/ 4.0/ 5 days	954.2	280.6	\$1,079,836
6600/ 4.0/ 5 days	1934.3	293.1	\$1,809,628

Table 4: Model results for varying populations of Harvard students

The model campus was relatively robust to increased population with an every-other-day testing cadence; even with the pessimistic assumption $R_0 = 4.0$, and the full normal student body of 6600 students, we saw a relatively contained infection, with only 11.1 symptomatic cases per 1000 students. In contrast, testing every fifth day resulted in substantially more infections on campus, and potentially greater risk of uncontrolled spread. In the most extreme scenario, with 6600 students on campus, $R_0 = 4.0$, and testing every fifth day,

8. The choices of 2 and 5 days are meant to be illustrative - our work considered values of τ_d from 1 to 7.

9. In this and following cases, testing cost falls off with more infection, since isolated, recovered, and deceased students are not tested.

nearly all the students in our model were infected. These results were largely robust to the introduction of a superspreader event, as shown in Appendix B.

Based on our work and a number of other considerations, the Harvard administration eventually opted for a relatively conservative campus population of 1500, with three tests a week. The ability to suppress infection through mass testing is in many ways particular to Harvard and similar institutions with the financial and logistical capability to regularly administer such testing. However, we hope that these findings can still prove useful in a more general sense, especially as tests become cheaper, more accurate, and more widely available.

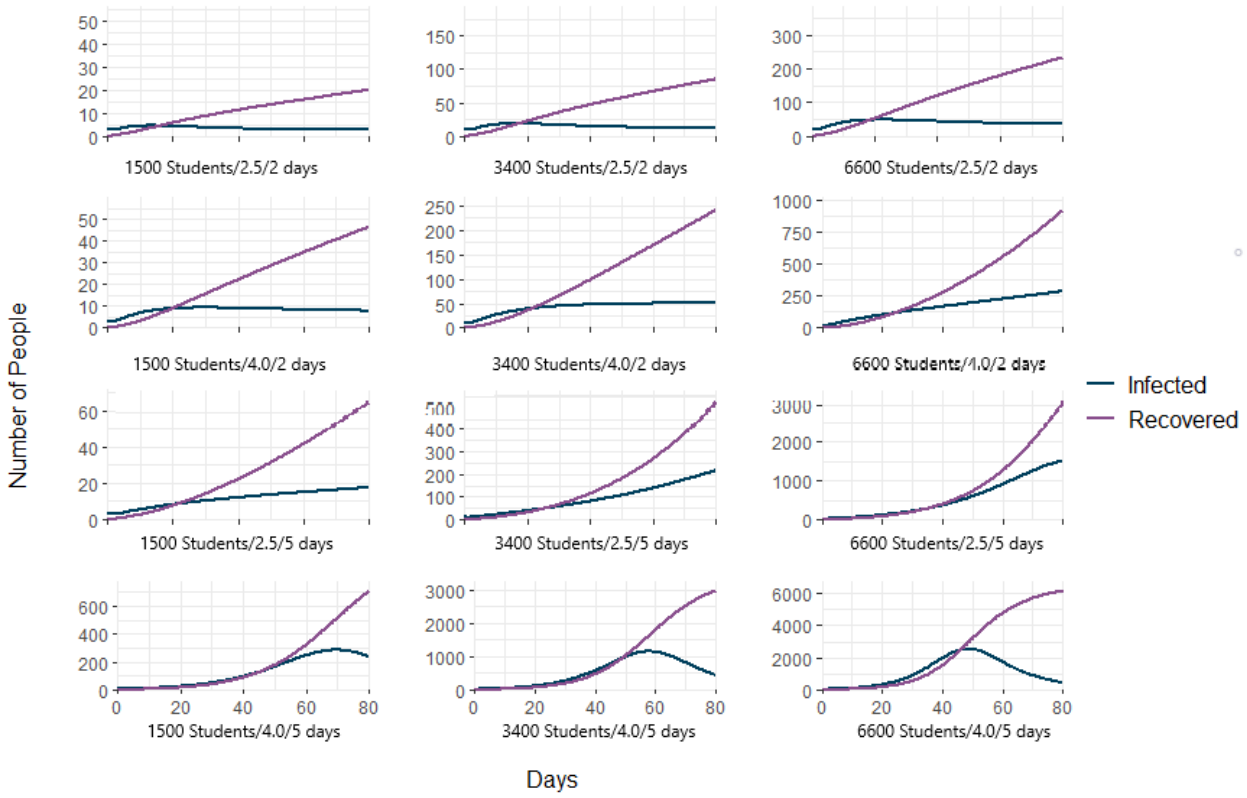


Figure 2: Charts of the number of infected and recovered students in 12 scenarios from our assessment of the optimal Harvard population.

4 Case 2: The Case for Social Gatherings

After the Harvard administration had decided how many students to bring back to campus (see Section 3), we moved on to examining specific questions about the nature of campus life. We undertook several extensions of our model (including a depiction of contact tracing, which may be incorporated into the main model in future). We present one such extension: our study of the effect of limited student social gatherings. After the initial set of plans was formulated and disseminated to students, the students reported widespread dissatisfaction, particularly with the planned restrictions on social gatherings (Bi and Srinivasan 2020).

Given the policy utility function discussed above, we saw several potential upsides to studying social gatherings. We believed that allowing some degree of gathering could be a substantial improvement in the

quality of the student experience. We were also influenced by public choice considerations - responding visibly to student concerns could increase compliance with the overall portfolio of measures. Finally, we suspected that gatherings would take place regardless of restrictions, and we thought it important to assess the potential effects.

We believed our model would show limited epidemiological downside to gatherings in the context of a conservative student population and frequent mass testing. Our results confirmed this hypothesis.

4.1 Parameters

Where not otherwise described, parameters are the same as specified in Table 3. As described above, social gatherings in the model are represented as a source of additional contacts. First, we created a control scenario, with no social gatherings. We then created three treatment scenarios with escalating degrees of gathering.

In our first scenario, the average student attended one gathering per week. The average gathering was assumed to create 10 additional contacts (recalling our definition of one contact as a brief face-to-face interaction, and considering that multiple shorter or indirect contacts can in aggregate be equivalent to one longer contact).[10] In our second scenario, the number of contacts per gathering created 20 contacts. The third scenario involved 20 contacts per gathering, and 2 gatherings per student per week. We believed this to be a pessimistic upper bound for student social activity, meant to correspond to unrestricted pre-pandemic behavior. These parameters can be seen in Table 5.

Parameter Name	Notation	Value(s)	Source
Basic Reproduction Number	R_0	2.5	CDC (2020c)
Student Population	n_{pop}	1500	Harvard decision
Students Asymptomatic at Beginning	NA_0	4.5	Assumption; fixed proportion of n_{pop}
Avg. Number Sharing Housing	n_r	1	Function of n_{pop}
External Infections per Day	I_{ed}	0.15	Proportion of n_{pop} consistent with previous case
Social Gathering Attendees/Period	c_{soc}	0/214/428	Quantity of interest. Everybody parties 1x/2x a week.
Contacts per Social Gathering	c_{soc}	10/20	Assumption

Table 5: Parameters for studying social gatherings at Harvard

4.2 Results

In all 3 social gathering scenarios, we found that the pandemic remained under control and spread to a very small proportion of the population, as can be seen in Table 6 and Figure 3. The most pessimistic scenario did see an upward-sloping case trajectory, suggesting more risk of uncontrolled outbreak. Nevertheless, these

10. As a frame of reference, we found that an R_0 of 2.5 and a p of 2.6% meant that the average student had a baseline of about 10 contacts on a normal day.

results suggested that, under conditions of regular testing, a reasonable degree of social gathering could be permitted without undue epidemiological risk.

Gatherings	Contacts per Gathering	Total Symptomatic Cases	Cases per 1000
None	-	17	11
1/student/wk	10	21	14
1/student/wk	20	26	18
2/student/wk	20	46	31

Table 6: Model results for introducing social gatherings (Results are cases per semester)

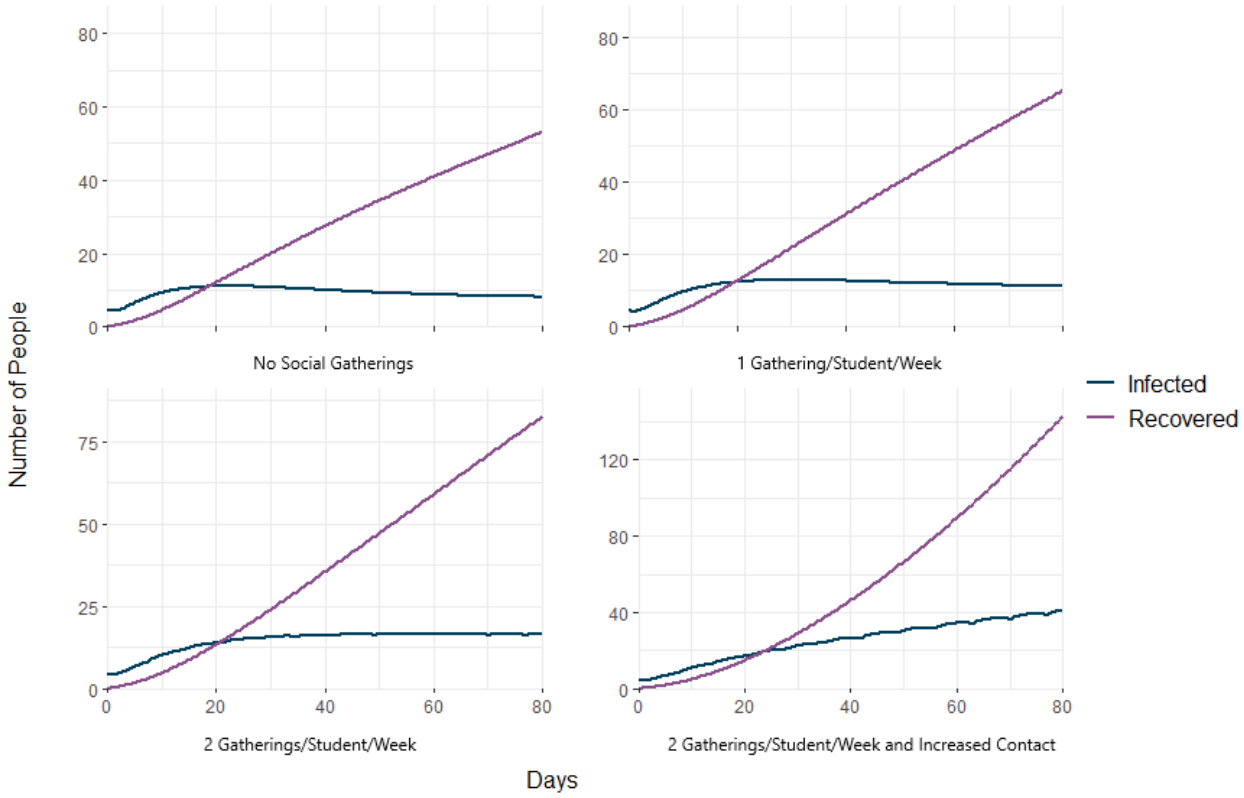


Figure 3: Scenarios with increasing degrees of social gathering.

It may appear at first that the introduction of social gatherings is pure cost without benefit. However, one must beware the sin of valuing only that which is directly measurable. Permitting gatherings may have many important but intangible benefits: allowing students to socialize might increase compliance with best practices across the board, as well as increasing morale and preserving a key element of the undergraduate experience. We emphasized, in our communication of our findings, the importance of these qualitative evaluation in conjunction with our quantitative results.

5 Case 3: Computing R_0 for the University of Illinois

The University of Illinois at Urbana-Champaign (UIUC) saw a notable outbreak of COVID-19 as students came on campus, primarily in the undergraduate population (Good and Butler 2020). The administration implemented a large scale testing program to control the outbreak; at one point, UIUC accounted for nearly 20% of all tests in Illinois (Deliso and Bhatt 2020). We use a snapshot of data made available by UIUC in this case.[11]

By populating our model with parameters reflecting the particular circumstances on the UIUC campus, we can compute a range of plausible values for the important parameter R_0. This parameter reflects the number of cases a single case will generate in a fully susceptible population. It is a function of the SARS-CoV-2 virus, but also of baseline social structure; the same pathogen might have an R_0 of 4 in a densely populated city, but only 1.5 in a sparse rural area (or an area with extensive social distancing practices).[12] It is important to note that R_0 is not a function of testing or isolation protocols. We hope that the social structure of UIUC is similar enough to that of other campuses that this will provide a baseline for understanding past and future outbreaks on college campuses.

We extracted a time series of realized positivity from UIUC's fall semester dashboard (UIUC 2020d). We use this to estimate parameters for UIUC, using ranges where reasonable values could not be computed. This allows us to solve for the values of R_0 such that the positivity in our model would most closely approximate the realized values at UIUC. The data were truncated to start on August 17th, the undergraduate move-in date, and run up to November 4th (UIUC 2020a).

5.1 Parameter estimation

The derivation of parameters can be seen in Table 7. Where not specified, parameters are the same as in Table 3.

5.2 Results

The model was tested for various values of the crucial parameter NA_0, the number of asymptomatic students at the start of the semester. This gives a range of realized values for R_0. Values of R_0 were obtained by fixing the parameters as above and then using a hill-climbing algorithm to find the value of R_0 that minimized root mean squared distance between model positivity and observed positivity.

The estimated values of R_0 depended substantially on the assumed number of students arriving with COVID-19. As a benchmark, if the incidence of COVID-19 in UIUC entering students was comparable to that in the general population in August 2020 (based on data from the New York Times), 0.25% of the students (130), would have had COVID-19 at the start, which means our model estimates an R_0 of 2.3 (The New York Times 2020). The complete results can be seen in Table 8.

We await further data collection from diverse college campuses to develop a better understanding of R_0 and other important parameters. If further research confirms our estimate of around 2.3, that means the current base infection rate on campuses, is similar to that of the rest of the current United States, per the best estimate of the CDC (2020c). Put another way, suppose both UIUC and the entire U.S., with all current

11. While a Harvard dataset was available, at the time this paper was prepared the incidence of COVID-19 on Harvard campus was quite low, limiting the utility of this dataset for estimating parameter values.

12. This parameter is distinct from the also-frequently-cited R_t. R_t reflects a point in time depiction of how fast the disease is reproducing itself - hence the focus on whether R_t is greater or less than 1. R_0 represents a more time-invariant (if environment-specific) perspective on how fast the disease would spread when introduced to a heretofore uninfected population

13. We found that the results were mostly robust to different values of this parameter

Parameter Name	Notation	Value(s)	Source
External Infections per Day	I_{ed}	0.2	Based on data from Champaign-Urbana Public Health District (2020)[13]
Test Specificity	Sp	100.0%	Ranoa et al. (2020)
Test Sensitivity	Se	100.0%	Ranoa et al. (2020)
Testing Cadence	τ_d	3.5	UIUC (2020c)
Student Population	n_{pop}	52,000	Deliso and Bhatt (2020)
Students Asymptomatic at Beginning	NA_0	130/260/520/1040	Range of plausible values
Avg. Number Sharing Housing	n_r	15	Based on housing documents at UIUC (2020b)

Table 7: Parameters for University of Illinois at Urbana-Champaign

Starting Asymptomatic	Percent of population	Estimated R_0
26	0.05%	2.8
52	0.10%	2.6
130	0.25%	2.3
260	0.50%	2.1
520	1.00%	1.8
1040	2.00%	1.4

Table 8: Observed values of R_0 at UIUC

restrictions in place, were entirely free of coronavirus. Then suppose that both were exposed to the same degree of initial infection. The equivalence of R_0 means that the infection would spread at similar rates.

If true, this finding has important epidemiological implications, but also changes the policy calculus. Many discussions of colleges under COVID-19 have assumed that a reopened campus poses an inherently high risk of transmission. If the R_0 on campus is equivalent to the off-campus R_0, that would mean the baseline infectivity on campuses is comparable to elsewhere - combined with the mass testing and contact tracing capabilities available to universities, this may well mean that students are actually better off on campus.

6 Interactive Software

We have prepared an interactive software tool, available at https://projects.iq.harvard.edu/covidu/app, that implements this model for general use. The software is easy to use, and comes with built-in guidance. A user can define values of every parameter described in this paper on the tool, and see a standardized output for the given scenario, with the option to view interactive charts or generate a long form PDF report. A picture of this software's interface and output can be seen in Figure 4.

The model can also be used to estimate causal effects: after defining two separate scenarios, the user

can generate a standardized comparison of the two scenarios, with charts and tables of the differences in parameters and in key outcomes. The PDF reports of the comparisons include a novel auto-summarization feature, which generates text that describes the treatment and summarizes the estimated causal effect. We hope this automatic text will be helpful in immediately conveying the "big picture" of a causal effect comparison, which can then be augmented by more detailed information. An example report is attached as Appendix C.

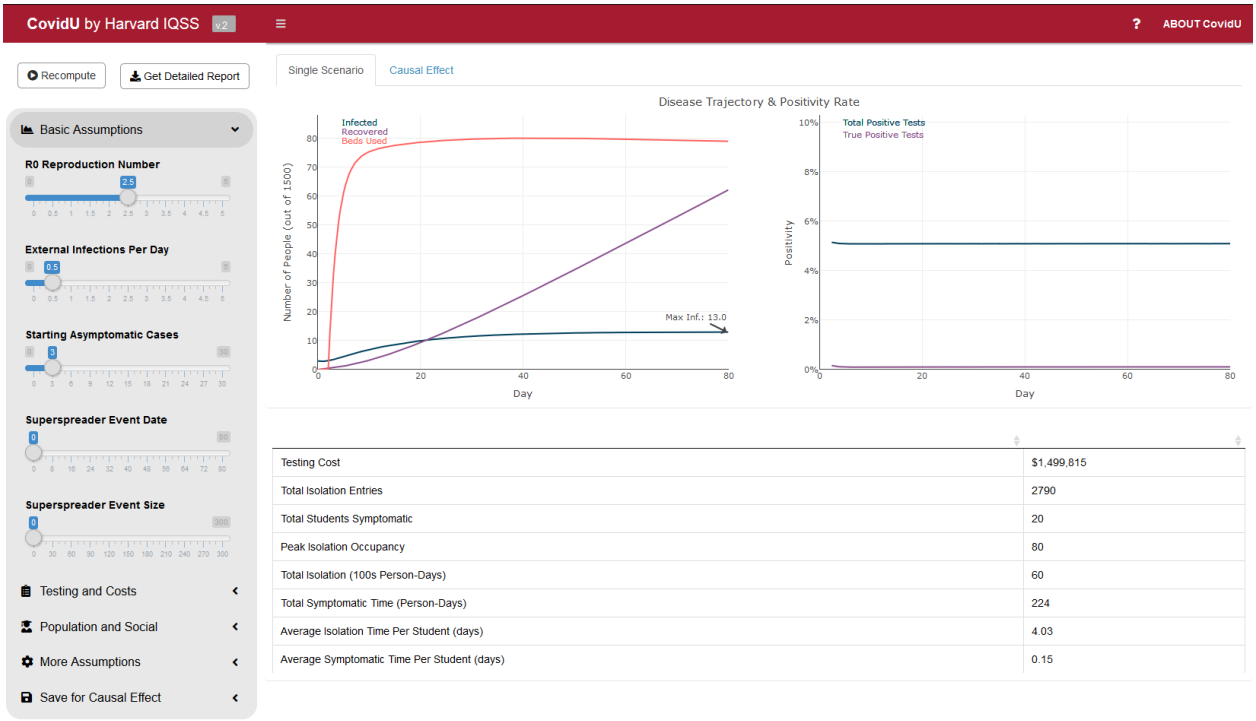


Figure 4: The interactive software implementing this model

A Appendix: Parameter Definitions

This table lists all parameters in the model, along with definitions and relevant computations.

Basic		
Notation	Parameter Name	Definition
D	Number of Days	How many days the model is run. In this paper, we use 80 days
θ	Time steps per day	How many time steps in a day. 24 hours / θ is the number of hours in a time step. In this paper, θ is 3.
Disease Properties		
Notation	Parameter Name	Definition
R_0	Basic Reproduction Number	Number of cases generated by an infectious individual introduced to a susceptible population. Used with p to compute c_n, normal degree of contact between students. $\beta = (R_0/(r_r + r_s))$ $c_n = \beta/p$
p	Probability of Infection given Contact	Probability that a contact between an infectious individual and a susceptible individual results in an infection, where contact is a face-to-face interaction (Luo et al. 2020)
P_r	Probability of Shared Housing Infection	Cumulative probability that an infectious individual infects a susceptible individual if they share housing (roommates, shared bathroom, suitemates). Used to compute c_r, additional contact from shared housing: $c_r = (r_r/p) * (P_r/(1 - P_r))$
t_i	Incubation Time	Time it takes between being infected and becoming infectious (and potentially developing symptoms). Used to compute r_i, incubation rate: $r_i = 1/(t_i * \theta)$
t_r	Recovery Time	Average time it takes to recover after incubation is completed, where recovery means no longer being infectious, no longer having symptoms, and no longer potentially developing symptoms. Used to compute r_r, recovery rate: $r_r = 1/(t_r * \theta)$
s	Cumulative Proportion Developing Symptoms	Given a population of asymptomatically infected individuals, this is the proportion that will develop symptoms. Used to compute r_s, the rate of symptom development in the asymptomatic population: $r_s = r_r * s/(1 - s)$

δ	Conditional Mortality Rate if Symptomatic	The cumulative probability that a symptomatic individual will die. Based on an undergraduate population without comorbidities. Used to compute r_δ, death rate; let $k = \delta/(r_s/(r_s + r_r))$; then $r_\delta = r_r * k/(1-k)$.

Testing

Notation	Parameter Name	Definition
Sp	Test Specificity	Probability that a test will detect no COVID-19 where it does not exist. Equal to 1 - false positive rate.
Se	Test Sensitivity	Probability that a test will detect COVID-19 where it does exist. Equal to 1 - false negative rate.
π	Cost per Test	Dollar cost of a test
π_c	Cost per Confirmatory Test	Cost of a test administered to students entering isolation, to confirm results of the earlier testing.
τ_d	Testing Cadence	Determines rate of testing. If τ_d is 2 days, that corresponds to testing every other day; if τ_d is 7 days, that means weekly testing. $\tau = \tau_d * \theta$
t_{fp}	Time to Release False Positives	How long it takes to confirm that a positive test is a false positive, and re-release the isolated individual. Used to compute r_{fp}, the rate of false positive release: $r_{fp} = 1/(t_{fp} * \theta)$

Population and Social

Notation	Parameter Name	Definition
n_{pop}	Student Population	Number of students
NA_0	Students Asymptomatic at Beginning	Number of students starting as asymptomatic and infectious. $S_0 = n_{pop} - NA_0$
n_r	Average Number Sharing Housing	Average Number of individuals in each shared housing group; either roommates, suitemates, or sharing bathroom
$I_{e,d}$	External Infections per Day	External infections are infections coming from outside the campus community - e.g. students going out to lunch and being served by an infectious individual
n_{soc}	Number of Students Attending Social Gatherings Per Period	How many students go to social gatherings
c_{soc}	Contacts per Social Gathering	The additional degree of contact created by social gatherings

$F(t)$	Superspreader Indicator Function	Function determining whether or not a superspreader event occurs at a given time (e.g. a concert, a rally, a large party)
I_s	Superspreader Event Size	The number of additional infections created by a superspreader event

B Appendix: Further Results for Case 1

The charts in Figure 5 depict further results for Case 1 (see Section 3), with the introduction of a small (2.5% of population) superspreader event on Day 25. The conclusions are largely similar: testing every other day keeps COVID-19 contained in our model.

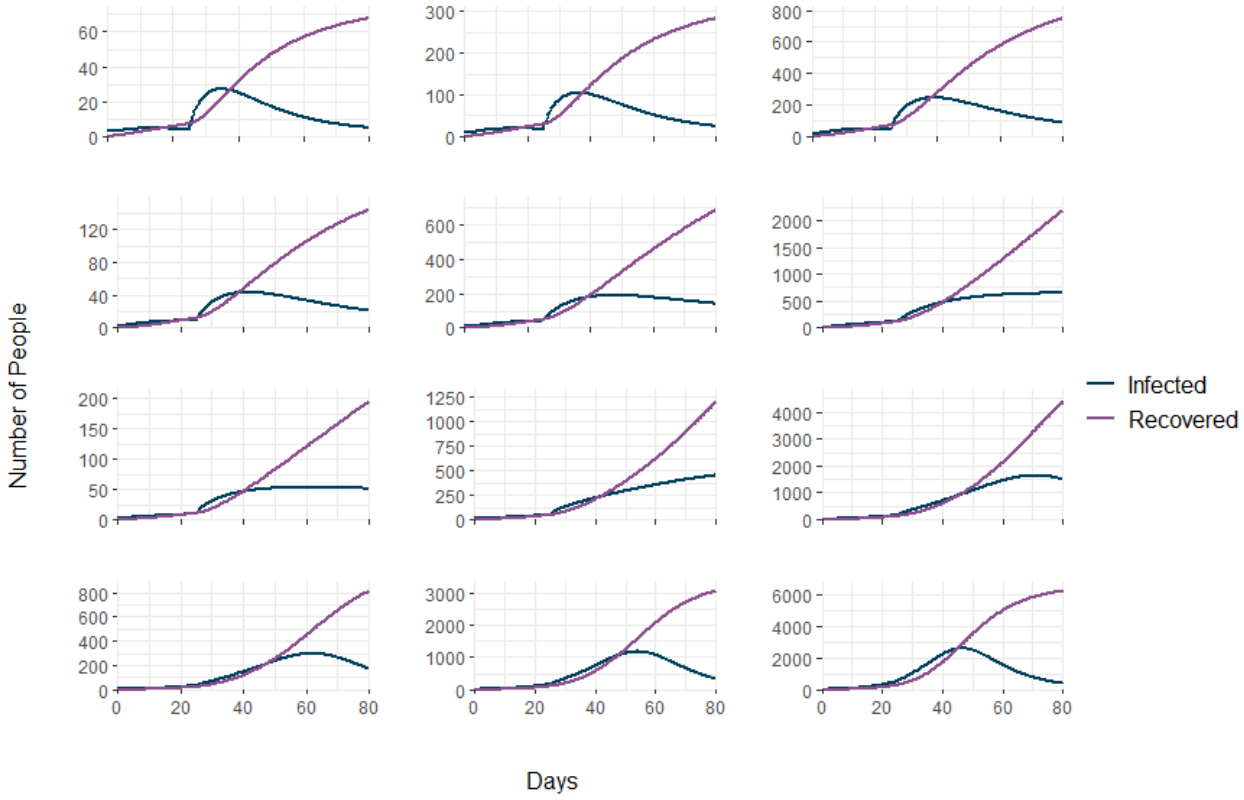


Figure 5: Charts of the number of infected, recovered, and isolated students in 12 scenarios from our assessment of the optimal Harvard population, with superspreader event. Order is the same as in Figure 2

C Appendix: Example Report

CovidU: Comparison Report [1]

Scenarios Evaluated:

Treatment: housing group size 15, 25.0% intra-pod infection risk. (Details on Page 2)

Control: housing group size 1, 15.0% intra-pod infection risk. (Details on Page 3)

Parameters not listed on this page are held constant at values listed on pages 2 and 3.

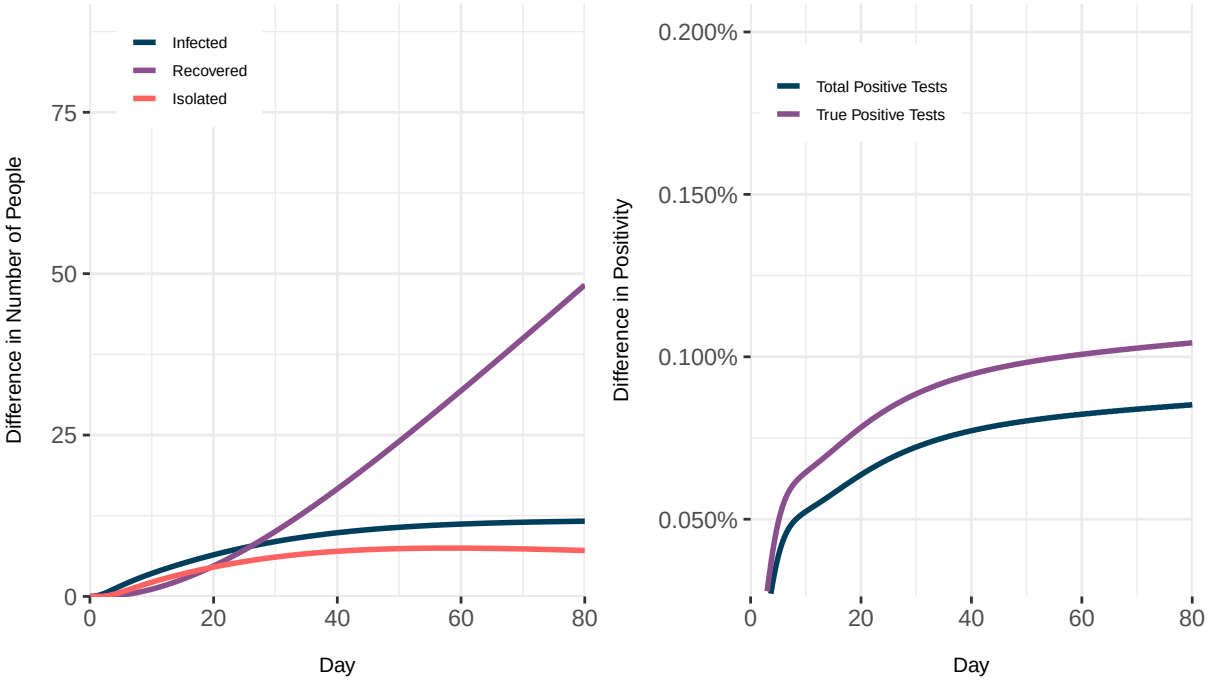
Conclusion: Changing from Control to Treatment shows: 24345 less $ in total testing cost (an decrease from 1515149 to 1490804), 5 more students placed in isolation (an increase from 2788 to 2793), 16 more students developing symptoms (an increase from 11 to 26)

Changes in Population and Social Assumptions

Average Size of Shared Housing	14
Probability of Roommate Infection	10.0%

Changes in Results

Testing Cost	-24,345$
Total Isolation Entries	5
Total Students Symptomatic	16
Peak Isolation Occupancy	7
Total Isolation (100s Person-Days)	5
Total Symptomatic Time (Person-Days)	170
Average Isolation Time Per Student (days)	0.30
Average Symptomatic Time Per Student (days)	0.11



[1]This report was generated using the CovidU app, from Harvard IQSS. For more information, see iq.harvard.edu/covidu

CovidU: Detailed Report [1]

Summary

R0 Reproduction Number	2.5	Infections the first case can be expected to generate (see Dictionary)
External Infections per Day	0.5	Size of external infection events
Starting Asymptomatic Cases	3	
Superspreader Event Date	0	Day on which a superspreader event occurs (if any)
Superspreader Event Size	0	Size of superspreader event(if any)

Testing and Costs

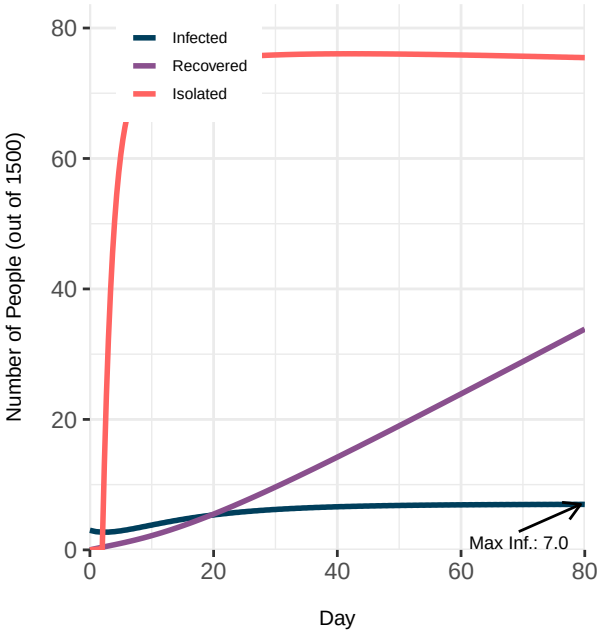
Test Sensitivity (%)	95.0%	Chance of detecting Covid if there is Covid
Test Specificity (%)	85.0%	Chance of detecting no Covid if there is no Covid
Testing Cadence	1	Interval between tests
Cost Per Test ($)	25$	
Confirmatory Test Cost ($)	50$	Currently assuming 100% sens/spec
False Positive Release Time	2	Days for False Positive to be released

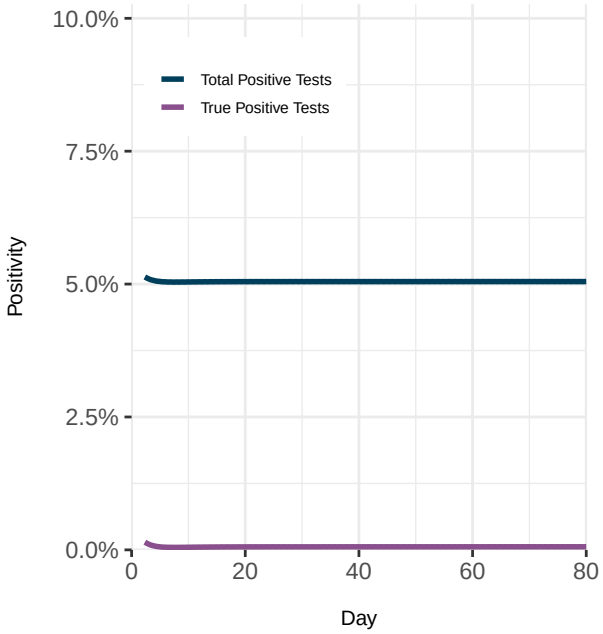
Population and Social

Student Population	1500	
Average Size of Shared Housing	1	Number of Students Living Together
Probability of Roommate Infection	15.0%	Chance one infected pod member infects another

Disease Assumptions

Probability of Infection Given Contact	2.5%	
Recovery Time (days)	14	Days to recover from illness
Incubation Time (days)	4	Time between getting infected & being infectious
Conditional Mortality	0.05%	Chance of death if hospitalized (for student age group)
Proportion Developing Symptoms	30.0%	Chance of symptoms developing if infected





[1] This report was generated using the CovidU app, from Harvard IQSS. For more information, see iq.harvard.edu/covidu

CovidU: Detailed Report [1]

Summary

R0 Reproduction Number	2.5	Infections the first case can be expected to generate (see Dictionary)
External Infections per Day	0.5	Size of external infection events
Starting Asymptomatic Cases	3	
Superspreader Event Date	0	Day on which a superspreader event occurs (if any)
Superspreader Event Size	0	Size of superspreader event(if any)

Testing and Costs

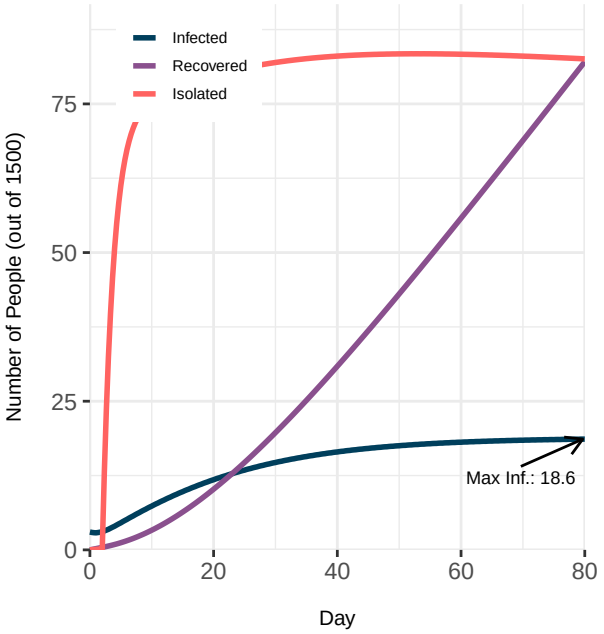
Test Sensitivity (%)	95.0%	Chance of detecting Covid if there is Covid
Test Specificity (%)	85.0%	Chance of detecting no Covid if there is no Covid
Testing Cadence	1	Interval between tests
Cost Per Test ($)	25$	
Confirmatory Test Cost ($)	50$	Currently assuming 100% sens/spec
False Positive Release Time	2	Days for False Positive to be released

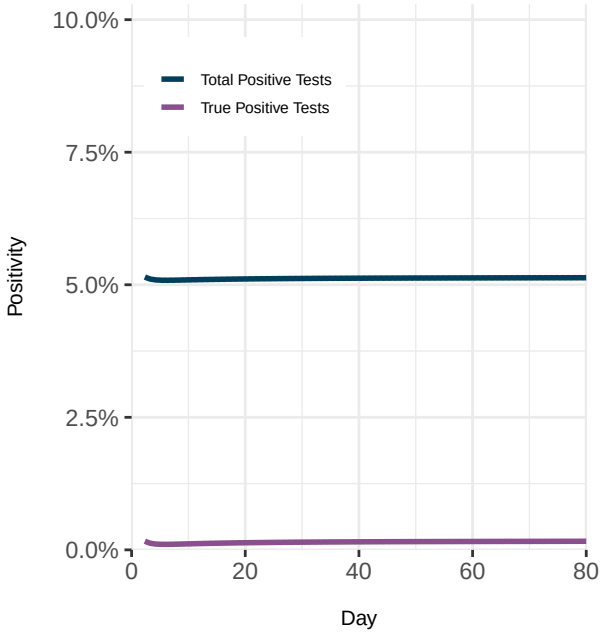
Population and Social

Student Population	1500	
Average Size of Shared Housing	15	Number of Students Living Together
Probability of Roommate Infection	25.0%	Chance one infected pod member infects another

Disease Assumptions

Probability of Infection Given Contact	2.5%	
Recovery Time (days)	14	Days to recover from illness
Incubation Time (days)	4	Time between getting infected & being infectious
Conditional Mortality	0.05%	Chance of death if hospitalized (for student age group)
Proportion Developing Symptoms	30.0%	Chance of symptoms developing if infected





[1] This report was generated using the CovidU app, from Harvard IQSS. For more information, see iq.harvard.edu/covidu

References

Bar-On, Yinon M., Ron Sender, Avi I. Flamholz, Rob Phillips, and Ron Milo. 2020. "A quantitative compendium of COVID-19 epidemiology." Preprint Article. *arXiv* (July 9, 2020). https://arxiv.org/abs/2006.01283.

Bi, Kevin, and Sahana Srinivasan. 2020. "What will Harvard students do this Fall?," July 12, 2020. Accessed November 19, 2020. https://medium.com/harvard-open-data-project/what-will-harvard-students-do-this-fall-cdccb698eb61.

Cai, Weiyi, Danielle Ivory, Mitch Smith, Alex Lemonides, and Lauryn Higgin. 2020. "What We Know About Coronavirus Cases on Campus." *The New York Times,* June 28, 2020. https://www.nytimes.com/interactive/2020/07/28/us/covid-19-colleges-universities.html.

Champaign-Urbana Public Health District. 2020. "Champaign-Urbana COVID-19 Coronavirus Information," November 17, 2020. Accessed November 17, 2020. https://www.c-uphd.org/champaign-urbana-illinois-coronavirus-information.html.

City of Cambridge. 2020. "Case Count Reporting in Cambridge," November 12, 2020. Accessed November 12, 2020. https://www.cambridgema.gov/covid19/casecount.

Deliso, Meredith, and Jay Bhatt. 2020. "Inside University of Illinois' massive COVID-19 testing operation." *ABC News,* September 10, 2020. Accessed November 5, 2020. https://abcnews.go.com/US/inside-university-illinois-massive-covid-19-testing-operation/story?id=72686799.

Frieden, Thomas R., and Christopher T. Lee. 2020. "Identifying and Interrupting Superspreading Events—Implications for Control of Severe Acute Respiratory Syndrome Coronavirus 2." *Emerging Infectious Diseases* 26 (6): 1059–1066. https://doi.org/10.3201/eid2606.200495.

Good, Gavin, and Raleigh Butler. 2020. "University of Illinois sees spike in faculty, staff, and grad student COVID-19 cases as county battles 'uncontrolled spread'." *CU-CitizenAccess,* November 16, 2020. Accessed November 17, 2020. https://www.cu-citizenaccess.org/2020/11/16/university-of-illinois-sees-spike-in-faculty-staff-and-grad-student-covid-19-cases-as-county-battles-uncontrolled-spread/.

He, Xi, Eric H. Y. Lau, Peng Wu, Xilong Deng, Jian Wang, Xinxin Hao, Yiu Chung Lau, et al. 2020. "Temporal dynamics in viral shedding and transmissibility of COVID-19." *Nature Medicine* 26:672–675. https://doi.org/10.1038/s41591-020-0869-5.

Lauer, Stephen A., Kyra H. Grantz, Qifang Bi, Forrest K. Jones, Qulu Zheng, Hannah R. Meredith, Andrew S. Azman, Nicholas G. Reich, and Justin Lessler. 2020. "The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application." *Annals of Internal Medicine* 172 (9): 577–582. https://doi.org/10.7326/M20-0504.

Li, Ruiyun, Sen Pei, Bin Chen, Yimeng Song, Tao Zhang, Wan Yang, and Jeffrey Shaman. 2020. "Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2)." *Science* 368 (6490): 489–493. ISSN: 0036-8075. https://doi.org/10.1126/science.abb3221. eprint: https://science.sciencemag.org/content/368/6490/489.full.pdf. https://science.sciencemag.org/content/368/6490/489.

Luo, Lei, Dan Liu, Xin-long Liao, Xian-bo Wu, Qin-long Jing, Jia-zhen Zheng, Fang-hua Liu, et al. 2020. "Modes of contact and risk of transmission in COVID-19 among close contacts." Preprint Article. *medRxiv* (March 26, 2020). https://doi.org/10.1101/2020.03.24.20042606.

Paltiel, A. David, Amy Zheng, and Rochelle P. Walensky. 2020. "Assessment of SARS-CoV-2 Screening Strategies to Permit the Safe Reopening of College Campuses in the United States." *JAMA Network Open* (July 31, 2020). https://doi.org/10.1001/jamanetworkopen.2020.16818.

Ranoa, Diana Rose E., Robin L. Holland, Fadi G. Alnaji, Kelsie J. Green, Leyi Wang, Christopher B. Brooke, Martin D. Burke, Timothy M. Fan, and Paul J. Hergenrother. 2020. "Saliva-Based Molecular Testing for SARS-CoV-2 that Bypasses RNA Extraction." *bioRxiv,* https://doi.org/10.1101/2020.06.18.159434. eprint: https://www.biorxiv.org/content/early/2020/06/18/2020.06.18.159434.full.pdf. https://www.biorxiv.org/content/early/2020/06/18/2020.06.18.159434.

The New York Times. 2020. "An ongoing repository of data on coronavirus cases and deaths in the U.S.," November 17, 2020. Accessed November 17, 2020. https://github.com/nytimes/covid-19-data.

University of Illinois at Urbana-Champaign. 2020a. "Fall 2020 Early Arrival," August 17, 2020. Accessed November 5, 2020. https://housing.illinois.edu/Tools/move-in/early-arrival.

———. 2020b. "Illinois University Housing." Accessed November 5, 2020. https://housing.illinois.edu/Living-Options/Residence-Halls/Undergraduate-Halls.

———. 2020c. "On-Campus COVID-19 Testing," September 10, 2020. Accessed November 5, 2020. https://covid19.illinois.edu/health-and-support/on-campus-covid-19-testing/#tested.

———. 2020d. "Shield Testing Data," November 5, 2020. Accessed November 5, 2020. https://go.illinois.edu/COVIDTestingData.

US Centers for Disease Control and Prevention. 2020a. "Case Investigation and Contact Tracing Guidance: Appendices," October 21, 2020. Accessed November 1, 2020. https://www.cdc.gov/coronavirus/2019-ncov/php/contact-tracing/contact-tracing-plan/appendix.html#contact.

———. 2020b. "Considerations for Institutions of Higher Education," October 29, 2020. Accessed November 1, 2020. https://www.cdc.gov/coronavirus/2019-ncov/community/colleges-universities/considerations.html.

———. 2020c. "COVID-19 Pandemic Planning Scenarios," September 10, 2020. Accessed October 20, 2020. https://www.cdc.gov/coronavirus/2019-ncov/community/colleges-universities/considerations.html.

———. 2020d. "When to Quarantine," October 27, 2020. Accessed November 12, 2020. https://www.cdc.gov/coronavirus/2019-ncov/if-you-are-sick/quarantine.html.