medRxiv preprint doi: https://doi.org/10.1101/2020.09.04.20188680; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

AN EVALUATION OF EDUCATIONAL INSTITUTIONS SAFE REOPENING STRATEGIES FOR IN-PERSON CLASSES AMID THE COVID-19 PANDEMIC

[1]Ujjal Kumar Mukherjee, [2]Subhonmesh Bose, [1]Anton Ivanov, [1]Sebastian Souyris, [1]Sridhar Seshadri, [3]Padmavati Sridhar, [1]Ronald Watkins, and *[1]Yuqian Xu*

[1]Gies College of Business, University of Illinois at Urbana-Champaign, [2]Electrical and Computer Engineering, Grainger College of Engineering, University of Illinois at Urbana-Champaign, [3](Student) University of California, Berkeley

**Corresponding Author. Email: ukm@illinois.edu*

ABSTRACT

Research Context. *Can educational institutions open up safely amid COVID-19?* We build an epidemiological model to investigate the strategies necessary for institutions to safely reopen. The four measures that are most relevant for in-person opening are: (i) wide-spread rapid testing, possibly saliva-based, (ii) enforcement of mask-wearing, (iii) social distancing, and (iv) contact tracing.

Research Design. Using an analytical setup, we theoretically demonstrate that institutions need to test at a relatively high level (e.g., at least once every week for all individuals) in the initial phases of reopening. Guided by the analytical setup, we derive insights from an agent-based simulation. Contact tracing is relatively more important when the positivity rate from random testing is relatively low, which is likely during the initial phases. An adaptive testing strategy based on positivity rates can help institutions optimally manage the costs and risks of reopening. Finally, to demonstrate the strategies in practice, we provide empirical estimates of some of the educational institutions opening up experience and comment on mitigation strategies. Empirically, we characterize the role of testing using data from the SHIELD program at the University of Illinois at Urbana Champaign (UIUC).

Results. We show that increasing the testing levels from 0.2 per capita per day to 0.3 per capita per day can reduce the infectivity from 0.25 to 0.01, with an average slope of the infectivity to the testing curve being 0.35 in this range. We also cross-validate the results with data from a large number of universities in the United States, and show that institutions with higher levels of testing are associated with lower infections. The estimated marginal effect of increasing testing levels by 1% per capita per day across universities can reduce the positivity by an average of 0.0228% with a 99% confidence interval of [0.0209%-0.0253%]. We also provide an estimate of the locational effects of institutions on mitigation strategies. We estimate from data on 228 different universities across the United States that an increase of infection rate at the county where a university is located by 1% has the potential to increase the institutional infection rate by an average of 0.14% with a 99% confidence interval of [0.032% – 0.248%] across all universities. This indicates that universities are not closed systems, rather they are open systems subject to external influence, and the extent of external influence potential is an important consideration for opening up of universities.

Contributions. This paper contributes to the nascent literature on combating the COVID-19 pandemic and is especially relevant for large organizations. This work is motivated and guided by the SHIELD program of UIUC. We provide important policy pointers for the reopening of universities.

Short Title. *Reopening Strategies Amid COVID-19.*

Acknowledgments. *We acknowledge the kind guidance and support of Ron Watkins, Managing Director at SHIELD Illinois & Associate Dean Gies College of Business. SHIELD Illinois is a saliva-based rapid testing program under which 10,000 staff and students are tested every day, from a population of approximately 50,000 on campus. The program is also expanded nationally (https://news.uillinois.edu/view/7815/190493450). We also acknowledge the funding provided by C3.ai for conducting the research, and the data support received from c3.ai Data Lake as part of the grant.*

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.04.20188680; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

INTRODUCTION

The COVID-19 pandemic has closed several organizations nation-wide, including educational institutions, with severe economic and social consequences[1]. While the importance of reopening is well acknowledged, there is less of a consensus on the strategies necessary for the safe opening of educational institutions in the United States and around the globe[2-9]. The Center for Disease Control (CDC) has issued reopening guidelines[10] that include extensive hand hygiene, cloth face coverings, disinfections, physical barriers and spacing of individuals inside enclosed surroundings, and frequent testing. The early experience of reopening has met with relatively unfavorable outcomes, such as the Cherokee County School District in the state of Georgia, USA, that had to send 250 of its staff and students to quarantine after reopening in August, 2020[11-12]. The University of North Carolina, Chapel Hill, USA, reversed its plans for in-person classes after more than 130 confirmed infections in the first week of reopening[17]. Motivated by these observations, we explore these questions: (i) *Can schools and educational institutions and other organizations open safely amid COVID-19?* (ii) *If so, what are the measures required to open educational institutions and other organizations to commence in-person operations while maintaining public-health safety from the spread of COVID-19?* We use a combination of a stylized analytical model and an agent-based simulation analysis to provide policy guidance toward answering these questions. We evaluate the following strategies: (i) *widespread rapid testing of individuals*, (ii) *mask-wearing and other safety measures such as handwashing and disinfecting*, (iii) *social distancing*, and (iv) *contact tracing* of detected individuals. The questions are motivated and guided by the SHIELD program of the University of Illinois at Urbana-Champaign, in which the university is testing 10,000 students and staff every day (testing per individual per week is 0.2) in order to hold in-person classes. We use data from the UIUC SHIELD program and similar data from other institutions to comment on the success of proposed strategies, and how institutions can adapt to emerging epidemics within institutions to ensure the health and safety of their members.

This work offers interesting insights into re-opening strategies that we delineate below. Methodologically, the compartmental diffusion models, without modification, are not suitable for the analysis of small populations required to evaluate re-opening strategies for institutions. Therefore, we propose an epidemic model that takes into account small population dynamics, suitable to model institutional settings. The large size of universities creates complexities of efforts and escalation of costs in ensuring that testing can be performed at numbers that are sufficient to dampen rates of infection. While doing less of one action can be compensated by adequately increasing the extent of the other, not all subsets of actions are feasible from a practical standpoint. For example, only testing without proper mask enforcement and social distancing will require testing almost every individual every day for safe reopening. Our analysis demonstrates that optimally allocating testing capacity between random testing and contact tracing is important. Interestingly, and somewhat counterintuitively, the value of contact tracing is higher when the positivity rate from random testing is relatively low. Positivity rates from random testing is an indicator of current and future infections. Low positivity rates imply that the probability of discovering infected individuals from random testing is low. Therefore, during the initial stages of reopening when the infection load is likely to be low, focusing greater efforts toward contact tracing is crucial. However, contact tracing needs to be optimally combined with random testing. We demonstrate that given a probability of infection transmission of 5% and a contact rate of 10 individuals per day, that describes a somewhat typical scenario, every individual need to be tested once every 5 days or more frequently for dampening infections in large educational institutions. Rather than adopting a fixed testing capacity, a flexible adaptive system based on Bayesian updating of estimated positivity rates of testing can be more cost-efficient. During the initial stages of reopening, it is important to test more. The testing levels can be reduced adaptively as the infection load (positivity rate) decays. Such an adaptive testing strategy is inherently forward-looking and considers the risk of ongoing and future transmission of the virus. We empirically demonstrate that the location of a university is an important consideration because of the possible external influx of infections from the environment. Specifically, we demonstrate with data that the higher the infection rate of the county where a university is located, the higher is the infection rate within the university. We also provide evidence that the relationship is, in fact, dyadic in that large universities with a significant influx of students from outside, contribute significantly towards the growth of infection in the surrounding region. Specifically, we show, for the two large universities we analyze, reopening significantly increases the infection

medRxiv preprint doi: https://doi.org/10.1101/2020.09.04.20188680; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

load of the county where the universities are located. Our final contribution is the synthesis of information from the opening experience of large universities (UIUC, ISU, and others) and the estimation of infection transmission parameters from the data collected in the process. Our research will likely permit other large institutions to gauge the risk of various re-opening strategies that will, in turn, shape policies they adopt.

THE SARS-CoV-2 INFECTION AND MITIGATION STRATEGIES

The SARS-CoV-2 is a novel strain of coronavirus that currently does not have an approved cure. For mitigation, countries implemented strategies that varied from complete lockdown of large geographical areas to partial restrictions on mobility and mask enforcements in public places. A particular challenge associated with this virus is its asymptomatic transmission in which many infected individuals remain asymptomatic from a few days to several weeks[13] and yet transmit the disease to susceptible people. An estimation of the COVID-19 infection in Italy[14] suggests that the infected population in Italy as of May 4, 2020, may have ranged between 2.2 and 3.5 million, where detected infections were approximately 200,000. Therefore, a strategy of random testing is required to detect and arrest the spread of further infections through systematic isolation and quarantining of those who test positive for infection. A study[15] indicates that mask-wearing reduces infection transmission by 51% from normal transmission rates. The FDA has recently approved saliva-based rapid testing based on the loop amplified isothermal amplification (LAMP), which costs significantly less than the usual PCR test[16]. This provides the opportunity to amplify testing rapid testing within institutions.

THE INFECTION PROCESS

We develop an analytical framework for the infection and testing process, ideally suited for institutions with a relatively smaller number of individuals as compared to the population of counties and cities. In the institutional setting, large sample-based compartmental epidemic models, which assume populations to be significantly larger than the number of infections, are not adequate and will likely provide erroneous estimates and wrong policy recommendations. Therefore, we consider small sample dynamics that, despite being stylized, reflect more accurately the realities within educational institutions, e.g., overlaps among contacts of infected individuals and external inflow of infections and the inherent randomness in the dynamics of the infection. The analytical model we develop guides our numerical experiments.

Consider an institution with N individuals. Let x_t denote the number of undetected infections, n_t denote the number of uninfected individuals, and $r_t = N - n_t - x_t$ equals the number of detected and isolated individuals at time $t = 1, \dots, T$. Therefore, the number of infectious individuals is given by $x_t + n_t$. We calculate the expected number of individuals who newly get infected at time t. To that end, we calculate the probability that an uninfected individual i contracts the infection from an infected individual j at time t. This probability is the product of the probabilities that j is infected, i is not already infected, j comes in contact with i, and successful infection transmission occurs, assuming independence among these events. With a contact rate of $\mathcal{M}$ to denote the number of unique individuals that one person meets on average, this probability is $\mathbb{P}(i \leftarrow j \text{ at } t) = \frac{x_t}{x_t+n_t} \times \frac{n_t}{x_t+n_t} \times \frac{\mathcal{M}}{x_t+n_t} \times \pi_m$, where π_m is the probability of infection transmission with subscript m denoting the extent of mask enforcement, measured by the fraction of the organization that generally wear masks. The infection transmission probability is a non-increasing function of m. With this model, the expected number of new infections becomes $\Delta x_t = \pi_m \mathcal{M} \frac{x_t n_t}{x_t+n_t}$. No institution is a closed isolated system. Interaction with the outside world is inevitable. We incorporate interactions with external populations through a multiplicative factor $r > 1$ in the number of new infections. As an example to illustrate our model, a population of 5,000 with 100 infections will lead to 98 new infections on average in the next period given an infectivity rate of $\pi_m = 0.1$, contact rate of $\mathcal{M} = 10$, and external interaction of 1% captured via $r = 1.01$.

THE TESTING PROCESS

Testing offers the means to both estimates the prevalence of the disease and to control it through the quarantining of confirmed cases. In this work, we consider two kinds of testing—random testing and contact tracing. We now count the number of infected individuals identified as being positive through these two

medRxiv preprint doi: https://doi.org/10.1101/2020.09.04.20188680; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

channels of testing. Assume throughout that the testing process is rapid and the results are available the same day. Let the total testing capacity be T_t. The number of new detections and isolations via random testing T_{rt} people alone is $\Delta R_{rt} = \frac{x_t}{x_t+n_t} T_{rt}$. The total number of expected contacts of the these randomly tested individuals is $\mathcal{M}\Delta R_{rt}$. In a small population, many of these individuals may belong to more than one list of contacts from all who test positive. The likelihood that a contact is in some such individual's list is $\frac{\mathcal{M}}{x_t+n_t}$, yielding $1-\left(1-\frac{\mathcal{M}}{x_t+n_t}\right)^{\Delta R_{rt}} \approx \frac{\mathcal{M}\Delta R_{rt}}{x_t+n_t}$ as the probability of landing in at least one among such lists. Therefore, the expected number of unique contacts of ΔR_{rt} positive individuals become $(x_t+n_t-\Delta R_{rt})\frac{\mathcal{M}\Delta R_{rt}}{x_t+n_t} = \mathcal{M}\Delta R_{rt}\left(1-\frac{\Delta R_{rt}}{x_t+n_t}\right)$. Assume that contact tracing is not fully efficient, and we are only able to identify $\eta_I \in (0,1)$ the fraction of the contacts for testing. Then, one must test $C_t := \eta_I \mathcal{M}\Delta R_{rt}\left(1-\frac{\Delta R_{rt}}{x_t+n_t}\right)$ individuals via contact tracing, as long as this number does not exceed the total number of tests $\eta_a T_t$ allotted to contact tracing, where $\eta_a \in (0,1)$. In the event that it does, the number of tests through contact tracing becomes $\eta_a T_t$. Therefore, the total number of contacts who are traced and tested is $\min\{\eta_a T_t, C_t\}$. We assume that unused allocation of contact tracing is used for random testing. The number of random tests administered is given by $T_{rt} = (1-\eta_a)T_t + [\eta_a T_t - C_t\,]_+$, where $[A]_+$ the positive part of A. The likelihood of detection among the contact-traced individuals is the probability of infection transmission, given that one has come in contact with an infected individual. In our model, this probability is $\frac{\pi_m n_t}{x_t+n_t} + \frac{x_t}{x_t+n_t}$. Therefore, the total number of detections of the disease in period t can be written as $\Delta R_t = \frac{x_t T_{rt}}{x_t+n_t} + \min\{\eta_a T_t, C_t\} \times \frac{\pi_m\, n_t + x_t}{x_t+n_t}$.

THE POSITIVITY RATE: AN INDICATOR FOR RATE OF SPREAD

Absent capabilities to test the whole population in a single day, allocation decisions must be made based on the observable positivity rate among the population, that we define as the ratio of positive test outcomes and the number of tests. Mathematically, this ratio is $\hat{p}_t = P_t/T_t$, where T_t tests yield P_t positive outcomes. The positivity rate is a leading indicator for infection prevalence and future expected infections and serves as a proxy for the fraction $\frac{x_t}{x_t+n_t}$ of the total size of the population that is infected. Recall that the expected number of new infections at time t is given by,

$$\Delta x_t = \pi_m \mathcal{M}(x_t+n_t)\frac{x_t}{x_t+n_t}\left(1-\frac{x_t}{x_t+n_t}\right) \approx \pi_m \mathcal{M}(x_t+n_t)\hat{p}_t(1-\hat{p}_t).$$

When the organization is sufficiently large, we expect this approximation to be largely accurate. For small $\hat{p}_t \sim$ 1-2%, the slope of the new estimated infections with respect to the positivity rate is given by $r\,\pi_m \mathcal{M}(x_t+n_t)$, that provides an estimate of the infection transmission rate (infectivity). This number provides a glimpse into the efficacy of preventative measures such as mask-wearing and social distancing in containing the disease. For $\hat{p}_t$ up to 50%, this slope is also positive, meaning that it grows with the positivity rate, as expected. Proceeding very similarly, the expression for the new expected number of positive results ΔR_t yields $\frac{\partial \Delta R_t}{\partial \eta_a} = \pi_m T_t(1-\hat{p}_t)$ for small η_a, which is positive but decreasing in the positivity rate. Therefore, in the initial stages of the epidemic spread when the positivity rates are typically low, it is important to conduct contact tracing, preferably using a combination of electronic and manual approaches. A contact tracing strategy is more focused on individuals who had already been in contact with infected individuals. However, the marginal return from contact tracing reduces as the positivity rate grows. Our numerical results from an agent-based simulation capture a similar sentiment. As an illustrative example, for a positivity rate of 2% with an infection transmission probability of 5%, the yield in contact tracing is likely to be at least 3% higher than random testing. When the positivity rate is 8%, this advantage of contact tracing over random testing disappears. Similarly, when the positivity rate is close to 40%, only random testing is largely sufficient. When the positivity rate is low, contact tracing is significantly more important than when it is high.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.04.20188680; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

FIXED TESTING CAPACITY

Decision-makers estimate the likely range of the infection that they expect in an institutional setting by observing other similar institutions, the state of infection in the social environment in which an institution is situated, and the extent of external interaction; these factors determine the baseline risk of COVID-19. If the estimated positivity rate from prior testing is p_m, one can set the testing capacity to T_m such that $r\pi_m \mathcal{M} p_m(1-p_m)N \leq (p_m + \eta_a(1-p_m))T_m$, where the right-hand-side of the inequality is approximately the number of positive tests expected from a combination of contact tracing and random testing. In deriving this expression, we have assumed that the contact rate is at least as high as the contact testing capacity, i.e., $\eta_a T_t \leq C_t$. In so doing, we ensure that T_m satisfying this inequality will always lead to the infection dying out. We specifically seek the minimum testing capacity for which the yield of positive tests exceeds the new number of infections. In other words, this policy seeks to tackle new infections every period by testing enough at the same period to bring the reproduction ratio of the disease to less than unity—a condition that ensures eventual decay of the epidemic. Rearranging, we get $\frac{T_m}{N} \geq \frac{r\,\pi_m \mathcal{M}\, p_m(1-p_m)}{p_m+\eta_a(1-p_m)}$. Maximizing the ratio on the right-hand-side over $p_m \in [0,1]$, we infer that a fixed testing capacity of $r\,\pi_m \mathcal{M}/\left(1+\sqrt{\eta_a}\right)^2$ ensures safe reopening. We assume in this analysis that preventative measures such as mask-wearing and social distancing are held constant, i.e., $m(t) = m$ throughout. Therefore, for a transmission rate of $\pi_m = 0.05$, mobility of $\mathcal{M} = 10$, and $r = 1.1$, and a 30% allocation for contact testing, approximately 20% of the organization needs to be tested every day, indicating that every individual need to get tested at least once a week.

ADAPTIVE TESTING CAPACITY

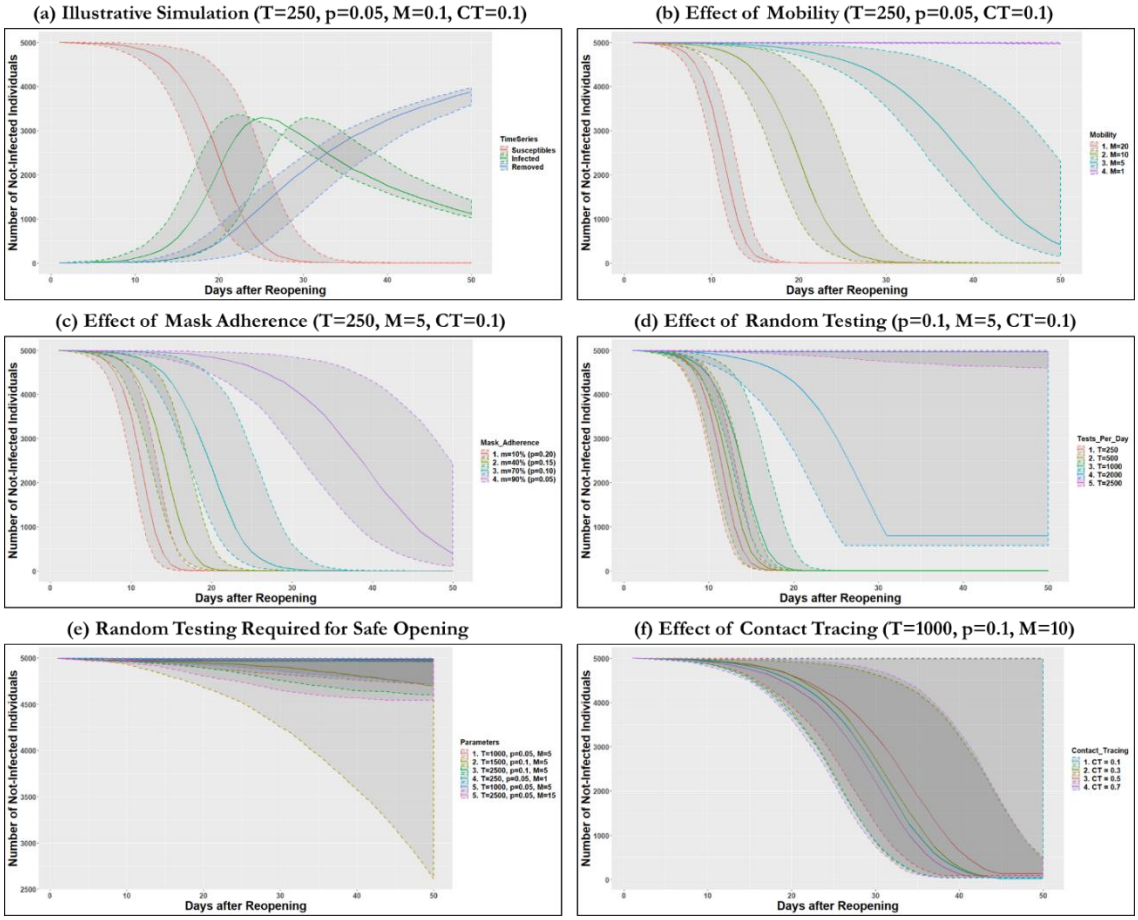
The fixed testing strategy is rather restrictive and conservative in that it plans for the maximum capacity that may be needed over the course of the disease and does not allow altering that capacity with observed positivity rate. Here, we alter that paradigm and adopt an adaptive testing strategy that allocates a capacity with one day look ahead to ensure dampening of the infection. The adaptive testing can be enacted by observing the positivity in testing from day t and incorporating the estimated positivity in deciding the testing level for day $(t+1)$ so as to dampen the positivity rate. However, such a mechanism does not account for uncertainty, and any discrepancy that may exist between the observed positivity and the real positivity. To account for the risk of uncertainty in estimating the real positivity from observed positivity, we add a risk buffer and consider that the decision maker ascertains the required number of tests with some level of confidence η_b, i.e., a decision risk of $(1-\eta_b)$. Such an allocation policy again seeks to test enough to tackle new infections every period with an adequate amount of testing in that period. We consider a small sample parametric setup for the institutional reopening strategy. The testing dynamics in a parametric small sample setup is assumed to follow the hypergeometric process where the random testing is analogous to the sampling without replacement, as opposed to with replacement, which would follow a binomial process. While in most cases, particularly when the sampling size is small compared to the population size, the Binomial process is a good approximation of the Hypergeometric process. However, in the institutional testing scenario, the sampling size can be as high as 20% to 30% of the population per day. Particularly, in the institutional testing scenario, the hypergeometric parameters are $\{(x_t+n_t), p_t(x_t+n_t)$ and the number of draws is the daily number of tests. Therefore, if the estimated positivity is $\hat{p}(t)$ from random testing in day t, then the decision-maker allocates an adaptive testing capacity T_{t+1} for day $t+1$. With T_{t+1} tests the expected number of positive results for the Hypergeometric testing process is $T_{t+1}\,\hat{p}_t$. For a confidence level of η_b, the administrator considers the $(1-\eta_b)$ quantile of the Hypergeometric process which indicates that the testing process detects the necessary number of positive cases to dampen the disease with a probability of η_b. The number of positive cases ΔR_{t+1} that the administration hopes to detect in the next period is then given by a Hypergeometric distribution such that $\mathbb{P}\left(\Delta R_{t+1} \leq \mathbb{E}(\Delta X_t)\right) = 1-\eta_b$, where the expected increase in new infections from t to $t+1$ is $\mathbb{E}(\Delta X_t) = r\,\pi_m \mathcal{M}\,\hat{p}_t(1-\hat{p}_t)\,(x_t+n_t-\Delta R_t)$. The probability $1-\eta_b$ acts as a risk buffer which ensures that the risk of decisions in testing capacity is accounted for safety.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.04.20188680; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

AGENT-BASED SIMULATION OF INSTITUTIONAL REOPENING

As indicated earlier, general analytical solutions of the standard epidemiological processes are insufficient for relatively smaller populations, such as those in various institutions and universities. We use an agent-based simulation to evaluate the possible combination of strategies that can enable the safe reopening of large institutions. In particular, we test the relative efficacy of random and adaptive testing under different scenarios of transmission of COVID-19. The infection transmission and the detection dynamics follow the analytical setup that we described earlier. Agent-based simulations have been extensively used in the context of epidemic spreads and transmissions[18]. In agent-based simulations, a collection of independent agents is allowed to interact with each other following probabilistic but simple rules. The Agent-Based Simulation is an alternative to the compartment-based modeling (e.g., SIR) and offers several advantages[19]. The complex interaction of a large number of agents over a time horizon (modeled as cycles of interaction) makes the agent-based simulation rich enough to generate complex dynamic behavior. These simulation models can account for randomization at the individual level and generate a distribution over possible scenarios, as shown in Figure 1a, where we show an illustrative example of the output of the agent-based simulation. We highlight that the parameters of our experiments remain unchanged within each simulation lasting 50 rounds. We have run the simulation 1,000 times for each set of parameters, and have plotted the median, the 95-th quantile, and the 5-th quantile of the distribution of the areas under the susceptible curves for each set of parameters from the 1000 runs. The area below the susceptible curve for a fixed population provides an estimate of the performance of the preventative strategies and testing. As an illustration, a sharp drop in the number of non-infected individuals will lead to a lower area below the susceptible curve, which indicates a high rate of infection. We summarize the strategies in terms of the normalized area under the susceptible curve.

Figure. 1. Agent-Based Simulation Output for Combinations of Strategies.



medRxiv preprint doi: https://doi.org/10.1101/2020.09.04.20188680; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

In the agent-based simulation, we create a collection of 5,000 agents, distributed over a space. Each agent moves independently in a random manner. The mobility is controlled by $\mathcal{M}$, which determines the average number of contacts per agent per day with other agents. We operationalize the contact rate for each individual in the simulation by randomly drawing $\mathcal{M}$ other individuals from the population who come in contact with the focal individual. Some of these $\mathcal{M}$ individuals may or may not be infected based on the infection diffusion that occurs across rounds for each experiment. We incorporate the impact of mask enforcement on the infectivity parameter p. Recent studies[15] have characterized the effect of mask-wearing on infection transmission in COVID-19. They have estimated the transmission rate to be 0.2 between two individuals who come in contact when neither of the individuals wear a mask. Masks reduce the infectivity by somewhere between 25% and 50%, with a minimum lower infectivity of 0.05. Furthermore, some studies[18] on mask wearing indicate that partial mask wearing is not effective. In our simulation, we consider that no mask wearing leads to an infectivity of 0.2 and vary it as $p(1-m^2)$, where m measures the percentage of the population wearing mask. The decrease is assumed proportional to m^2 to reflect the importance of mask wearing of both parties in an interaction, where the probability of each wearing a mask is m. For example, when mask enforcement is 75%, the infectivity is modeled as $0.2 \times (1-0.75^2) \approx 0.1$. We introduce 5 infected individuals at the beginning of the simulation. Further, we introduce a low probability of random external infection (0.005%) in every period. Since the initial infection and the external infections are not controllable strategies, we have kept these two parameters constant over cycles of simulation. As agents come in contact with other agents, they can transmit infection if one of the agents is already infected. The infection transmission occurs with a probability that is dependent on mask-wearing behavior. Finally, the testing is incorporated at two levels. A part of the testing is dedicated to random testing, where random individuals are tested. If random individuals are already infected, they are detected and quarantined. The second part of the testing is contact tracing. The detected individuals reveal their contacts with a probability equal to the informational efficiency of contact tracing. While in reality, the informational efficiency can be modified with institutional efforts, the informational efficiency is inelastic given a certain state of technology for contact tracing. We modify the allocation efficiency and the allocation of test capacity for contact tracing. The outcome measure is the number of infected individuals during the first 50 days. The summary outcomes from the agent-based simulation study are provided in Figure 1. Below, we summarize observations from the agent-based simulation.

Effect of Social Distancing and Institutional Restrictions on Mobility

In Figure 1b, N is maintained at 5000, T at 250, CT at 0.1, and p at 0.05, which is equivalent to 95% of the population wearing masks. In Figure 1b, the red curve represents $\mathcal{M} = 20$, indicating that each individual has a high contact rate (20 per day on average) with other individuals. In this scenario, the whole organization gets infected within 2 weeks of reopening. The scaled (max area set to 1) median area under the susceptible curve is 0.225. The maximum positivity rate in the organization is 60%, which is achieved on the 15th day. Also, Figure 1a is an illustrative outcome from the simulation that shows the median case in solid lines and the upper/lower 90th confidence intervals in dotted lines. The red line indicates the number of infected individuals in the system who are infected but not yet detected or removed. The black line indicates the number of individuals who are infected but detected and removed from the system by quarantining or through other mobility restrictions. Finally, the blue line indicates the number of individuals who are not yet infected. Furthermore, from Figure 1b we observe that when the mobility reduces to $\mathcal{M} = 10$ corresponding to the green curve, then the number of days to full infection increases to 3 weeks instead of 2 weeks. The area under the susceptible curve is 0.377. It is to be noted that these figures are for a low level of testing $\left(\frac{T}{N} = 0.05\right)$, where every individual is tested only once a month. For $\mathcal{M} = 5$, we find that in the 5th quantile of the distribution of the area under the simulated susceptible curves, while the whole organization gets infected in 2 months (approximately 50 days), more than 50% of the organization remains uninfected for the 95th quantile of the area under the susceptible curves. The median of the distribution of the area on the susceptible curves is 0.763. Finally, for $\mathcal{M} = 1$, i.e., each individual only meets one other individual, then the organization largely remains uninfected even at a low level of testing. The area under the susceptible curve is 0.998. However, at $\mathcal{M} = 1$ the idea of in-person operations is severely restrictive, and a combination of other measures may be required.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.04.20188680; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Effect of Mask Wearing and Sanitization Strategy

Mask wearing determines the probability of transmission π_m denoted as p. Mask enforcement has been implemented as an interpolated proportion of the percentage of the population that adheres to mask-wearing. As indicated in recent studies, In Figure 1c, for ($p = 0.20$), the whole organization gets infected within 15 days with a median area under the susceptible curve of 0.224; in Figure 1c ($p = 0.15$), the whole organization gets infected within 20 days and the median area under the susceptible curve is 0.281; in Figure 1c ($p = 0.1$), the whole organization gets infected within 30 days with a median area under the susceptible curve equal to 0.355; and finally, for ($p = 0.05$), the infection rate is reduced significantly and the majority of the organization remains uninfected, with a median area under the susceptible curve of 0.735. These figures demonstrate that mask enforcement is an essential part of the reopening of institutions. In Figure 1c, the level of testing is still low when each individual gets tested only once a week. The figures demonstrate that just by mask enforcement, organizations cannot ensure a safe reopening. Even when the probability of transmission is only 5%, a significant section of the organization still gets infected.

Effect of Testing Capacity

Figures 1d and 1e, respectively, indicate the simulation outcome for different levels of testing for fixed values of infectivity, and the different levels of testing that the organization would need to institute at different levels of infectivity for the organization to reopen safely. Figure 1d reveals that when testing levels are low, then, even with high levels of mask enforcement and restrictions on contact rates, significant transmission risk remains. In Figure 1d, we vary the level of testing from 250 tests per period (red curve) for a population of 5000 individuals, 500 tests per period (green curve), 1000 tests (magenta curve), 2000 tests (blue curve), and 2500 tests (pink curve). Maintaining all other parameters invariant, the testing levels have a significant impact on the expected spread of the disease after reopening. For testing levels of $\frac{T}{N} = 0.05, 0.1, 0.2, 0.4$ and 0.5, the simulated areas under the susceptible curves are 0.248, 0.252, 0.281, 0.521 and 0.989, respectively. Only when $\frac{T}{N} \geq 0.4$ we observe favorable outcomes with moderate levels of mask enforcement given by $p = 0.1$, which corresponds to 75% mask enforcement and contact rate of $\mathcal{M} = 10$. A 75% mask enforcement indicates that around 75% of the population is wearing masks throughout the simulation. The mask enforcement level remains constant across cycles (representing time) within a simulation. Therefore, at these levels of mask-wearing and mobility-related measures, every individual needs to get tested twice a week to have a safe opening. Indeed, the theoretical minimum number of tests per day for the parameters in Figure 1d is 0.32.

Figure 1e demonstrates the level of testing required for a safe opening for varying levels of mask enforcement when mobility and contact frequency are kept fixed at $\mathcal{M} = 5$. Corresponding to $p = 0.05, 0.1$ and 0.15 testing levels required to dampen the disease and maintain the area under the susceptible curve close to 1 are 1000 $\left(\frac{T}{N} = 0.2\right)$, 1500 $\left(\frac{T}{N} = 0.3\right)$ and 2500 $\left(\frac{T}{N} = 0.5\right)$ per day. A similar pattern can be observed for the other three scenarios, where $p = 0.05$ for all three scenarios, but the mobility and contact rate changes from 1 to 15, and the level of testing required changes from 250 to 2500. Recall that the minimum level of testing required for safe opening is $\frac{T_m}{N} \geq \frac{r\,\pi_m \mathcal{M}\, p_m(1-p_m)}{p_m+\eta_a(1-p_m)}$, where p_m is the positivity rate. Therefore, for the above parameters ($p = 0.05, r = 1.1, M = 5$), the minimum level of testing required is $\frac{T_m}{N} \geq 0.2295$, which amounts to testing about 1150 individuals per day for a 5,000-member organization. Therefore, our numerical experiment corroborates our estimates from theory.

Contact Tracing and Focused Testing

We already concluded from Figure 1e that large-scale random testing is important. Figure 1f illustrates the complementary role that contact tracing and random testing play in mitigating the infection spread. We denote the level of contact tracing by CT. With $CT = 0.3$, the 95th quantile of the curve of the number of susceptible (higher the better) indicates a much-reduced rate of infection as compared to the case with $CT = 0.1$. This observation continues to hold even for a contact tracing rate of 0.5. The impact of contact tracing, however, reduces when CT is driven north of 0.7. For such high values of CT, random testing is significantly reduced.

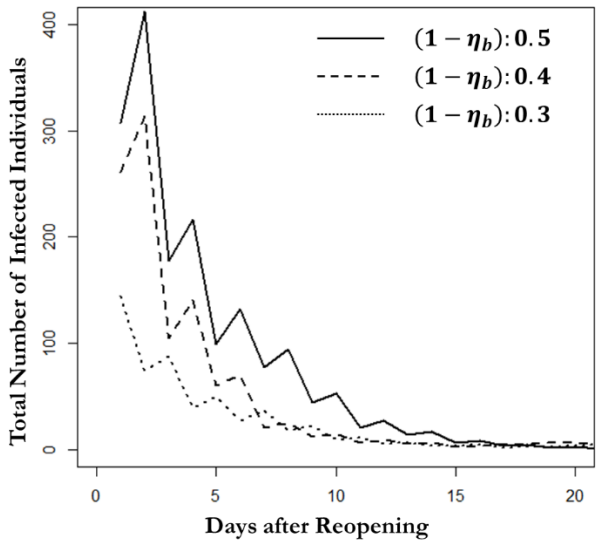
medRxiv preprint doi: https://doi.org/10.1101/2020.09.04.20188680; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

The influx of infections in an open system precludes the possibility of catching an adequate number of infections through contact tracing alone. Thus, arbitrarily increasing contact tracing at the expense of random testing becomes counterproductive. One needs to optimally select the mix.

Adaptive Testing Strategy

The discussion so far has been focused on fixed strategies, where the testing levels remain fixed at every period. Due to the variable nature of the infection, a fixed capacity may lead to unnecessary testing and expenses. In adaptive testing, we consider the case where the decision-maker estimates the positivity rate and decides the number of tests based on that estimate. We assume that she does so in a way that seeks to dampen the infection in the next time period with a given confidence (called uncertainty risk-based buffer). We show the infection dynamics for a population of 50,000 individuals with an initial infection load of 500 individuals, and an infectivity of 0.05. Figure 2 shows three levels of confidence (risk buffer), e.g., 0.5 (median), 0.6, and 0.7. At all levels of risk buffer, the disease significantly dampens. For higher-risk buffer, the disease dampens relatively more sharply. The maximum testing capacity however increases as the risk buffer increases. At 0.5, the initial testing capacity required is 10,000, whereas for 0.7 the maximum testing capacity required is 12,000. The adaptive testing capacity required reduces with the reduction in infection levels.

Figure 2. Fixed versus Adaptive Testing (N=50,000, p=0.05, M=5, Max T=10000, Starting positivity = 0.001)



CASE ANALYSIS: REOPENING EXPERIENCE OF UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN AND ILLINOIS STATE UNIVERSITY

We have so far described a theoretical model and an agent-based simulation that can provide policy pointers for reopening strategies. We tune these models with data on the reopening experience of two large public universities in the state of Illinois: (i) the University of Illinois at Urbana-Champaign, Champaign County, IL, and (ii) Illinois State University, Normal, McLean County, IL. Both universities closed down for in-person classes mid-March 2020 in the wake of the rising cases of COVID-19 infection in the United States and the State of Illinois. However, both universities decided to reopen for in-person classes in the Fall of 2020. In-person classes commenced on August 21, 2020, after closing the campuses for more than 5 months prior to the reopening. Both universities rolled out extensive plans for student awareness, mandating mask-wearing in university buildings, and extensive testing of students under the SHIELD program with rapid saliva-based tests. The University of Illinois at Urbana-Champaign planned to test all its 50,000 students, faculty, and staff twice a week using an electronic mobile application for managing the testing process and contact testing[24]. As expressed by an article in New York Times[24] these were the most extensive reopening plans that were put in

medRxiv preprint doi: https://doi.org/10.1101/2020.09.04.20188680; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

place. Despite the extensive plans, the spread of COVID-19 infection after re-opening within and from outside the campus was far from expectation. In UIUC, despite the extensive rapid testing of 20% of the population every day, the number of infected individuals sharply rose to around 1000 within 2 weeks of reopening. Similarly, the Illinois State University had 9000 students on campus in the fall, however, soon after opening up the positive rate of the COVID-19 infection increased significantly.

We obtained data on the number of tests conducted per day at both universities and the number of positive cases that were identified at both campuses from the public COVID-19 dashboards for both universities. Using the data, we estimated the real positivity rate in the population from the observed positivity rate as follows. In a population of size N, let the real positivity rate of the population on day t be p_t. Suppose $T \leq N$ tests are conducted on that day. For the estimation of the real positivity of the population, we assume parametric distributions that can model the number of detected individuals given the number of tests T. If the real positivity of the population is p_t, then for a population of N individuals, the number of COVID-19 positive individuals is Np_t. Since we assume that a detected individual is removed from the population, the probability distribution of obtaining K detections from random testing follows a random sampling distribution without replacement. We assume that K. the number of confirmed COVID-19 cases follows a $\text{Hypergeometric}(N, Np_t, T)$ distribution where the ambiguity in the true prevalence of the infection Np_t is represented through the $\text{Beta} - \text{Binomial}\ (N, \alpha, \beta)$ distribution. These parametric distributions are ideally suited to estimation from measured positivity rates. We performed an MCMC estimation of the actual positivity rate in the population of each university as follows. We use the data from a week to estimate the positivity rate of the population in consideration for the week. Then, we slide the time window by a day and re-estimate the average positivity rates for that week. Such a technique provides a moving weekly-average of the daily positivity rates of the population. We further approximate the infection transmission factor (infectivity) using the estimated positivity rate and a compound factor $\beta_I = r\ \pi_m \mathcal{M}$ that accounts for the base infectivity, the mask-wearing behavior, and the degree of socialization within the population. In the university setting, the slope of the real positivity rates provides us with a good estimation of the infectivity within the population of the universities. In Figure 3, we provide the data and the estimation of the positivity rates for the University of Illinois at Urbana-Champaign. In Figure 4, we provide the estimates for Illinois State University. The data for UIUC comes from the public dashboard of the UIUC SHIELD program (one of the co-authors is a director of the program) at the website: https://covid19.illinois.edu/on-campus-covid-19-testing-data-dashboard/ (last accessed September 30, 2020). The data for ISU comes from the public dashboard of ISU at https://coronavirus.illinoisstate.edu/dashboard/ (last accessed September 30, 2020).

Upon re-opening, the number of tests per day was increased significantly at UIUC, as Figure 3a illustrates. Figure 3b shows that confirmed positives also increased significantly around the same time. A part of the increased confirmed cases is due to the increased tests, which led to higher rates of quarantine and isolation of the positive cases. While higher rates of detection with rapid saliva-based testing at high volumes help reduce future infections, the number of positive cases kept increasing, which indicates that social distancing and mask-wearing were not adequate for a safe reopening[24]. Figure 3c presents the scatterplot of the observed positivity rates from the tests directly with the estimated real positivity rates using a seven-day sliding window. Initially, the positivity rates exhibited a dip from 0.5% (opening up) to 0.4% after the first week. However, after classes resumed and the campus resumed in-person classes (*note that all students were not on campus, some were remotely located*) the positivity rate increased significantly to a median estimate of 1.23% with a 95% confidence interval of [0.82%-1.67%]. Figure 3d presents the estimates of the infection transmission rate (infectivity). The infectivity estimates in Figure 3d are the daily infectivity estimates from the raw data without considering time-series effects and the effect of testing, which we present subsequently. The estimated infectivity steadily increased after opening up, demonstrating that the strategies were not sufficient to dampen the disease. The infectivity escalated to a mean level of 0.11 with a 95% confidence interval of [0.079-0.155]. Based on the estimates presented in Figure 1 on the likely strategies of opening up, for infectivity of 0.15 (considering the worst case), every student needs to get tested every second day for the disease to decay over time. We believe this is a key insight that can help the focal university UIUC and provide an estimate for other universities on the likely infectivity with similar-sized campuses and student body sizes. Another complementary strategy is

medRxiv preprint doi: https://doi.org/10.1101/2020.09.04.20188680; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

one of stricter enforcement of preventative measures including social distancing and mask-wearing. We acknowledge that implementing such a strategy in a public university setting is a daunting task, especially given that a significant student population lives in off-campus private residence halls. For Illinois State University, with a relatively smaller campus and a relatively smaller student body compared to UIUC, the infectivity rates as shown in Figure 3d initially increased, but later stabilized and even reduced. However, the positivity estimates are significantly more than that of UIUC. One of the contributing factors for the high positivity is the relatively lower number of tests done per day; while, the aggregate $T/N = 0.15$ for UIUC, the same number is 0.07 for ISU. We believe that the relatively, lower levels of tests at ISU led to a significant rise in the positivity rates rapidly to more than 0.3 at the peak, which reduced only after the campus initiated stricter control on social distancing and mask-wearing.

Figure 3. COVID-19 Infection of University of Illinois at Urbana-Champaign, IL

Fig. 3a. Daily Tests at UIUC

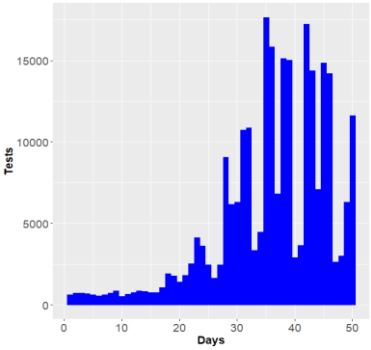


Fig. 3b. Total Confirmed Positive Cases

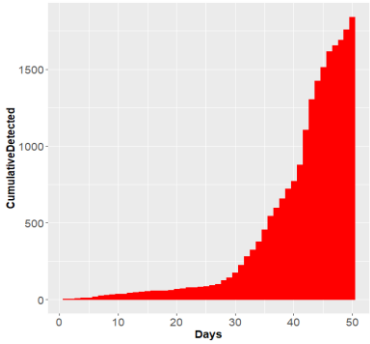


Fig. 3c. The Observed and Estimated Positivity Rates

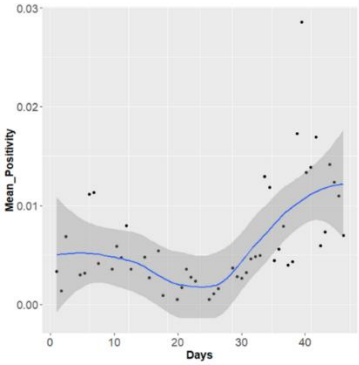
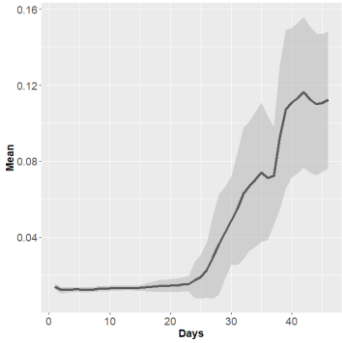


Fig. 3d. The Estimated Infectivity Rates



Note. *Days indicate days from the start of campus move-in, which is August 10, 2020*

Figure 4. Estimates for Illinois State University, Normal, IL

Fig. 4a. Daily Tests at ISU

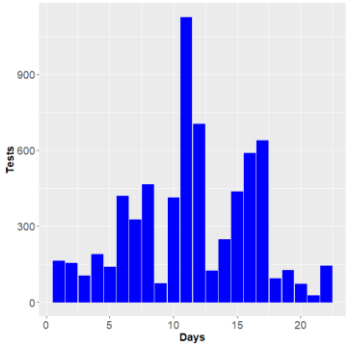
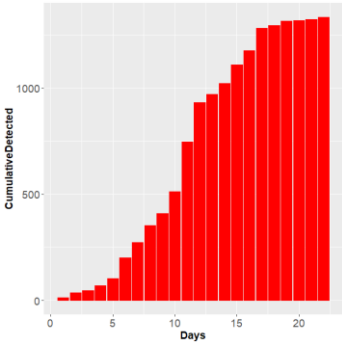


Fig. 4b. Total Confirmed Positive Cases



medRxiv preprint doi: https://doi.org/10.1101/2020.09.04.20188680; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Fig. 4c. The Observed and Estimated Positivity Rates

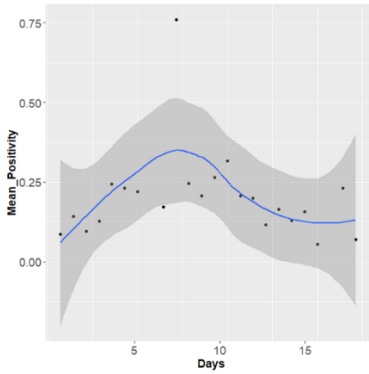
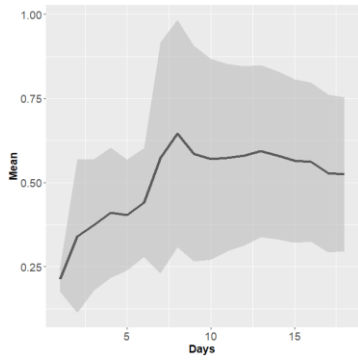


Fig. 4d. The Estimated Infectivity Rates



To estimate the relationship of infectivity at UIUC and ISU with respect to testing per capita, we consider the continuous-time version of the infection model that was developed for the simulation study. A linear regression setup for estimating the effect of infectivity and positivity within universities is not consistent due to the non-linear nature of the infection dynamics over time and the nature of the dependency of testing with positivity. This dependency is interesting in that lesser testing leads to growth in infections that in turn require amplifying testing needs later to mitigate the disease. Since the data is observational, and not experimental, the data generation process is non-random and involves the discretion of administrators and policymakers, such that the administrators chose the capacity of testing based on observed positivity rates. This is understood from Figure 3a that the testing per day increased in line with the increase in the number of positive cases shown in Figure 3b. While the positivity rate depends on the underlying epidemic process that we do not observe, assume that we take the observed positivity rate as a proxy for the true fraction of the infected population. This non-random data generation process in the observed data series at both UIUC and ISU makes a straightforward linear regression-based analysis to become inconsistent. To estimate parameters from data, we make an additional simplification and assume $x_t + n_t$ to roughly remain constant. Then, the positivity rate (with such simplifying assumptions) changes with time as

$$\frac{dp_t}{dt} = \pi_m p_t(1-p_t) - p_t\left(\frac{T_t}{N}\right) \quad \Rightarrow \quad p_t = -\frac{\left(\pi_m - \left(\frac{T_t}{N}\right)\right)\exp\left\{\pi_m t + b_0\left(\frac{T_t}{N}\right)\right\}}{\exp\left\{\pi_m b_0 + \left(\frac{T_t}{N}\right)t\right\} - b_0\exp\left\{\pi_m + b_0\left(\frac{T_t}{N}\right)\right\}}.$$

The parameters π_m, b_0 are unknown. We estimate them using non-linear curve fitting. We estimate the parameters for both UIUC and ISU as shown in Table 1a. From Table 1a we find that the effect of testing on infections within universities is significant and the effects are considerably different for the two universities. While for these two universities we have daily data, we have collected weekly testing and infection data for a total of 228 other universities. The data is collected from publicly available dashboards of individual universities regularly by the research team and collated into a web-tool for visualization of the data. We combine this data with data from the New York Times tracking dashboard for universities and schools (https://www.nytimes.com/interactive/2020/us/covid-college-cases-tracker.html). The research team has developed the following website (*www.covidedutrends.com*) to accompany the paper to enable visualization of the data across universities in the United States. Since the weekly data is a small panel, the non-linear curve fitting approach is not possible. Therefore, in Table 1b we show the Poisson regression estimates for the effect of testing on positivity for all major universities in the Universities in the United States. As expected, we find that increased testing significantly reduced positivity for all universities taken together. Finally, in Figure 5 we show the marginal curves of positivity with respect to testing intensities for (i) UIUC in Figure 5a, and (ii) ISU in Figure 5b.

medRxiv preprint doi: https://doi.org/10.1101/2020.09.04.20188680; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Table 1. Parameter Estimates for the Infectivity Estimates with respect to Testing Intensity

Table 1a. Non-linear Curve Estimates for UIUC and ISU

Parameter	UIUC		ISU	
	Est. (SE)	p-Value	Est. (SE)	p-Value
π_m	0.06 (0.00)	<0.01	0.14 (0.03)	<0.001
b_0	-5.03 (0.61)	<0.001	2.62 (2.01)	0.21
Error SE	0.001		0.031	
Sample N	74		48	

Table 1b. Beta-reg. for all Universities

Parameter	Est. (SE)	p-Value
(Intercept)	-2.86 (0.06)	<0.001
T/N	-0.49 (0.16)	<0.01
$Sq(T/N)$	0.06 (0.03)	<0.05
N	228	
Wald χ^2	602.5 (4)	<0.001

Figure 5. The Non-linear Curve Fitting Predictions for Infectivity with respect to the Testing Intensity

Figure 5a. Predicted for UIUC

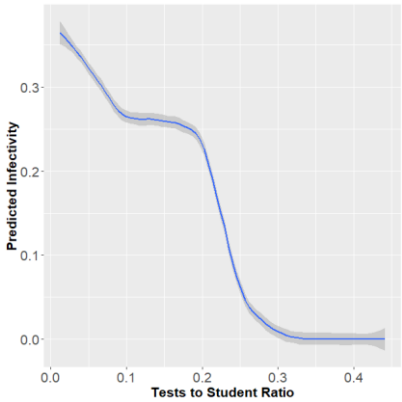
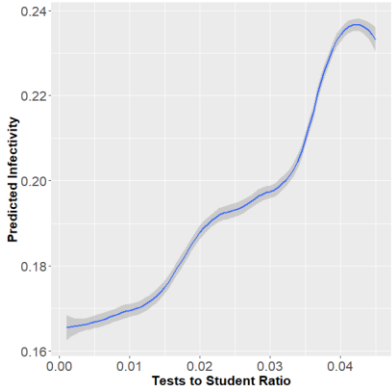


Figure 5b. Predicted for ISU



We observe from the figures that the higher intensity of testing at UIUC as part of the SHIELD program has led to a significant reduction in infections. Also, we observe from the data estimates that at a testing rate of 30%-40% of population daily at UIUC, the university can expect to control the COVID-19 infection and effectively dampen it down to negligible levels at the current levels of population mobility and exposure. However, from Figure 5b we observe that the test to student levels is much lower as compared to UIUC (approximately 10 times lower) and the predicted infectivity is higher than that of UIUC. This indicates that the number of rapid testing to the student ratio is an important policy lever for the reopening strategies. In the context of UIUC (Table 1a), the average reduction in infectivity is estimated to be from 0.35 to 0.25, when the testing per capita changes from 0.0 to 0.1 (10%). However, interestingly between the testing capacity of 0.1 to 0.2 per capita per day testing the infectivity does not change significantly, and remains stable. However, beyond the threshold of 0.2 testings per capita the reduction is infectivity is significantly sharper, with an average slope of >0.3, i.e., between testing of 0.2 to 0.3 per capita per day (once a week to twice a week) the infectivity estimates reduce from 0.25 to 0.01. This indicates that at lower levels of testing, the marginal benefit is much lower than at relatively higher levels beyond a threshold. These results, although estimates from the UIUC shield program, are applicable generally to other similar universities, and it seems that testing of at least 0.2 per capita per day is required for safe reopening. With considerations of the uncertainty of the disease propagation, a safer risk-averse threshold from the data is 0.3 per capita, when the infectivity is reduced to negligible levels.

Finally, in Table 1b, we present the results of estimation for 228 universities. The estimation results demonstrate that testing is essential for the safer reopening of universities. This is evident from the significantly negative slope parameter estimate (-0.49, p-value<0.01) of Table 1b, which indicates that universities that had relatively higher rates of testing to student ratio experienced lower levels of aggregate level infection within the universities. However, the squared term is positive and significant indicating a diminishing marginal return, however, the slope estimates of the linear and the squared term estimates of the testing per capita parameter indicate that the point of a turn-around for positive benefits of increasing testing is greater than 1, indicating that in the range of values for possible testing (0, 1) the marginal benefits in increasing testing at universities

medRxiv preprint doi: https://doi.org/10.1101/2020.09.04.20188680; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

remain significant and positive. To understand the effect of testing better we computed the marginal effect size of testing. Increasing testing per capita per day by 1% at the universities reduce the passivity rates by an average of 0.0228% with a 99% confidence interval of [0.0209%-0.0253%] when the average positivity across universities is 0.0493 [0.0451-0.0539]. From Figure 5, we reconfirm these findings. While for UIUC the positivity rates are predicted to be reduced with increased testing (refer Figure 3c showing infectivity range of 0-0.015), for ISU the story has been significantly different (refer Figure 4c showing infectivity range of 0-0.3, which is significantly higher than UIUC). Also, it has to be noted that the required testing rate needs to be around 0.3-0.4 for the infections to reduce to a negligible level.

LOCATION EFFECT ON REOPENING STRATEGY OF INSTITUTIONS

We have included a parameter to account for the infections coming from outside an institution. In the analytical model, the effect of the parameter is included in the infection rate. The possible interactions of institutions with the outside environment has the potential to significantly increase the infection rate within institutions, particularly when the infection load in the surrounding region is high. Before we delve into the effect of environment infection on the same within universities, we provide an estimate of the effect of institutions opening up on the surrounding regions due to the high influx of students and employees from other regions. Many of the large universities act as a transmission platform if the infection rates inside these institutions are not well controlled. In Figure 6, we present the infection rates of Champaign County in IL where UIUC is located and McLean County in IL where ISU is located. In figure 6a (for UIUC) and 6b (for ISU), the blue curves show the infection curves for the counties (without the institution infections) and the red curve shows the corresponding infection curve for the institutions. As seen from Figure 6, the reopening of institutions has significantly increased the infection load in the environment, particularly when infections within universities increased. We believe that the environmental infections and within institution infections (particularly for large institutions such as universities) have a reverse causal relationship with one reinforcing the other.

These graphs provide evidence that the location (environment) and the institutional infections are not independent but are rather intimately related to each other. One can profoundly impact the other. To investigate further, we collected data on infection rates within 228 large public and private universities after initial opening up (within 1-2 weeks) from university dashboards (collated at covidedutrends.com) and the New York Times database on tracking COVID-19 at Universities (https://www.nytimes.com/interactive/2020/us/covid-college-cases-tracker.html, last accessed September 30, 2020). Figure 7 provides a graphical view of the location of universities and infections. From Figure 7a, we observe that the higher rates of infections in universities are concentrated around the regions with generally higher rates of infections in the population. From Figure 7b, we observe that there is a positive correlation between institutional infectivity and county infectivity where the institutions are located.

To empirically characterize the effect of county infection load on institutional infections, we constructed the following variables: (i) University infection rate, which is the infection rate (%) after 2 weeks of reopening, (ii) Total infections, is the total number of detected COVID-19 infections in institutions, (iii) Enrollment, is the total fall enrollment from the US higher education news source, (iv) Cases is the total number of COVID-19 cases in the county of location, (v) County Infection Rate (%) is the infection rate at the county of location. In Table 2, we provide the descriptive statistic and correlation coefficients of the variables. In Table 2, we present the regression estimates with two different dependent variables. In column 1 of Table 2, we present the beta GLM of the University infection rate and in column 2, we present the quasi-Poisson (variance-inflated Poisson) GLM of the total number of infections.

The regression estimates of Table 3 suggest that the county infection rates are significantly associated with the university infection rates (column 1, Par. Est. 1.40e-01, SE. 3.62e-02, p-value < 0.001) and the total number of infections (column 2, Par. Est. 5.12e-01, SE. 1.04e-01, p-value < 0.001). An increase in county infection rate by 1% has the potential to increase the institutional infection rate by an average of 0.14% with a 99% confidence interval of [0.032% – 0.248%] across all universities. This is due to the fact that universities are not closed systems, rather they are open systems with significant environmental influence on the internal

medRxiv preprint doi: https://doi.org/10.1101/2020.09.04.20188680; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

infection rates. Furthermore, large institutions with higher enrollment have a higher risk of infections. This indicates that large institutions located in counties or zip codes with higher rates of population-level COVID-19 infections face relatively higher levels of risk after reopening. This is an important consideration that institutions need to account for when they develop their reopening strategies. In some counties, where infection levels are very high such as Cook County in IL and New York City, institutions may choose to adopt purely online operations until an approved cure or vaccine for COVID-19 emerges. Only testing may not be sufficient unless commensurate measures are adopted for ensuring general mask-wearing behavior and social-distancing behavior.

Figure 6. Infection Curves of Counties and Large Public Universities Located in those Counties

Fig. 6a. Champaign County and UIUC

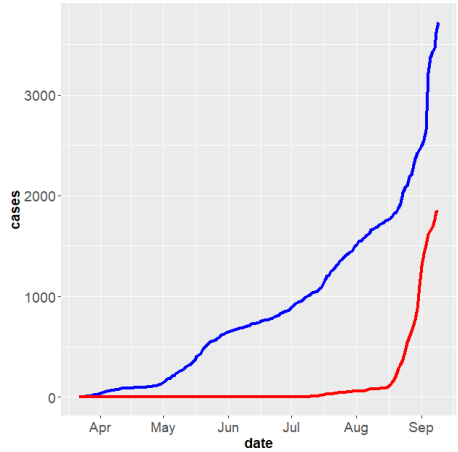


Fig. 6b. McLean County and ISU

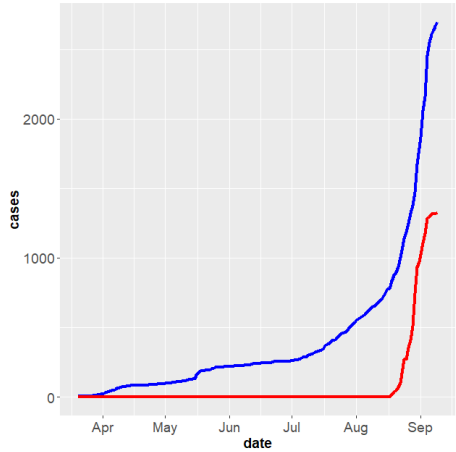


Figure 7. COVID-19 Infection Tracking within Large Universities in the US after Reopening

Fig. 7a. Infection Map of Universities after Reopening ($N = 228$)

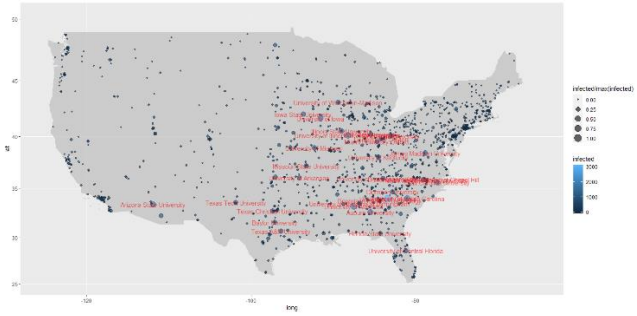


Fig. 7b. Infection Rate within Universities versus Infection Rate in the County of Location

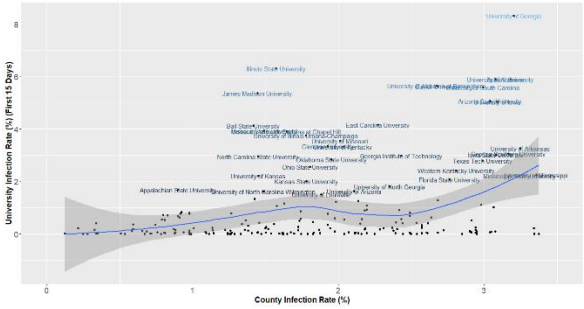


Table 2. Descriptive Statistics and Correlation Coefficients (Columns 1-5) of Variables

Sl	Variables	Mean	Std.Dev	(1)	(2)	(3)	(4)	(5)
(1)	University infection rate (%)	1.8	0.01	1.00				
(2)	Total infections (Nos.)	235.2	4365.8	0.95	1.00			
(3)	Enrollment (Nos.)	27,700.0	11,787.2	0.19	0.05	1.00		
(4)	Cases (Nos.)	22.923	43,641.3	-0.08	-0.10	0.14	1.00	
(5)	County Infection Rate (%)	1.98	1.03	0.21	0.21	0.07	0.40	1.00

medRxiv preprint doi: https://doi.org/10.1101/2020.09.04.20188680; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Table 3. Regression Estimates of (1) University infection rates and (2) Total infections.

Variables	(1) Dependent Variable: University infection rate (%) (Beta GLM)	(2) Dependent Variable: Total infections (Nos.) (Quasi-Poisson GLM)
Enrollment (Nos.)	3.49e-06 (2.99e-06)	3.10e-05 (7.68e-06) ***
Cases (Nos.)	-3.19e-06 (1.00e-06) **	-1.48e-05 (4.34e-06) ***
County Infection Rate (%)	1.40e-01 (3.62e-02) ***	5.12e-01 (1.04e-01) ***
(Intercept)	-4.30e00 (1.39e-01) ***	3.70e00 (3.74e-01) ***
Phi Parameter / Dispersion	154.12 (14.89) ***	641.46
Sample Size	228	228
Log-likelihood	747.3	847.9
F-Statistic	21.1 (df: 3, 3) *	14.9 (df: 3, 227) ***

*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

CONCLUSIONS

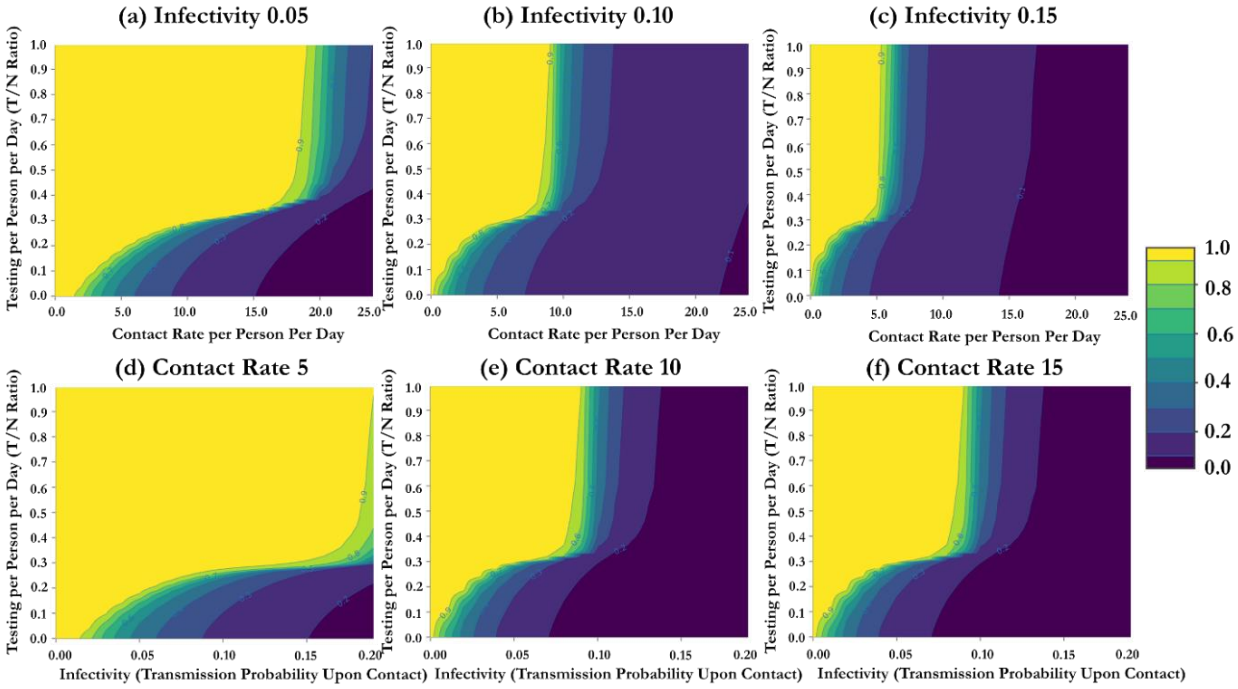
The reopening of institutions during the COVID-19 pandemic is challenging. Given reasonable levels of mask enforcement (5% chance of infection transmission given contact) and social distancing (5 contacts per person per day), for large institutions such as universities and colleges, a testing level of $\frac{T}{N} \geq 0.3$ can be sufficient to dampen the spread of the disease. This level of testing translates into testing every individual twice a week. However, if this level of testing is not feasible, then the shortfall can be compensated by ensuring higher stringency in mask enforcement and social distancing. If the testing level is around $\frac{T}{N} \geq 0.1$, then the average contact rates need to drop to at most 1 contact per day. These results are subject to the mathematical abstractions of simulation; however, these results provide a directional understanding of the combination of strategies that are important to consider while reopening institutions. We summarize the findings and the suggested strategy in Figure 8. Figure 8 provides a heat-map for safe reopening strategies and demonstrates the interaction of mask-wearing, which determines the infectivity upon contact with infected individuals, social distancing, which determines the contact rate of individuals, and the testing per person per day or the test capacity to population ratio under adaptive testing, subject to maximum capacity as shown. The metric of performance is the area under the susceptible curve, which is a function of the number of persons not infected at any point in time. The area under the susceptible curve is determined by the average basic reproduction number of an epidemic. While the basic reproduction number is an instantaneous measure, the area under the susceptible curve is a cumulative measure. From Figure 8, we observe that an institution needs to adapt to the estimates of infectivity and contact rates and adapt to changes in infectivity and contact rates. We have included several scenarios that provide a fairly comprehensive estimate of the rate of testing required. Many organizations are testing at a significantly high level; for instance, the University of Illinois at Urbana-Champaign has been testing at a rate of 10,000 $\left(\frac{T}{N} = 0.2\right)$ individuals every day for a population of approximately 50,000 individuals on campus under the SHIELD program, using a saliva-based rapid testing methodology. Some of the initial reopening experience confirms the value of a combination of strategies. Indiana University suspended all in-person activities in certain student housings after a rapid rise in COVID-19 cases after reopening[21]. Per a recent media report[22], several universities have more than 500 cases, such as the University of Alabama at Birmingham (972 cases), the University of North Carolina at Chapel Hill (835 cases), University of Central Florida (727 cases), Auburn University in Alabama (557 cases), Texas A&M University (500 cases), University of Notre Dame (473 cases), and the University of Illinois at Urbana-Champaign (448 cases), within days and weeks of reopening. Another study[23] indicated that colleges and universities would need to test every student once every two days to reopen safely. These outcomes and studies support the insights from our paper. Finally, we summarize an indicative and suggested strategy profile based on the study conclusions in Figure 8. Figure 8 shows the possible implications and strategy profiles for different levels of preventative adherence and contact rates. Figure 8 provides indicative direction for strategies for the reopening of universities. Generally, as discussed earlier, we observe that increases infectivity due to mobility and contact rates, and non-conformance

medRxiv preprint doi: https://doi.org/10.1101/2020.09.04.20188680; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

to mask-wearing, the testing intensity needs to increase non-linearly with respect to the contact rates. Therefore, a combination of strategies is warranted for the reopening of institutions.

In closing, we submit that the investigation into reopening strategies is subject to some limitations. The simulations demonstrated here are hypothetical and do not represent the practical complexities of a real organization. Furthermore, the actual implementation will entail additional organizational, social, or political constraints that have not been considered in the paper. Nevertheless, we believe that the results presented here provide significant practical insights, at least in a conceptual and directional sense, that can effectively enable institutions to reopen while controlling the risk of COVID-19 spread within the organization. Finally, we believe that different universities and institutions would need to customize the right combination of strategies based on the realities of reopening and the practicality of social distancing and other preventative measure adoption. Therefore, one size does not fit all, and adaptive customization of strategies is essential for the safe reopening of institutions.

Figure. 8. Summary of Strategies for Safe Reopening (Heat-Map of Scaled Area Under the Susceptible Curve – Higher the Better)



Data and Source Codes.

All data and source codes used in the paper are available at: https://github.com/Ujjal-Mukherjee/COVID-19-Reopening

A preprint link to the paper (earlier version) is available at:
https://www.medrxiv.org/content/10.1101/2020.09.04.20188680v2

The university tracking data dashboard created by the authors is available at:
https://public.tableau.com/profile/anton.ivanov3554#!/vizhome/Covid_Dashboard_v2/Dashboardv2?publish=yes (Owner. Anton Ivanov (co-author))

medRxiv preprint doi: https://doi.org/10.1101/2020.09.04.20188680; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

REFERENCES

1. Bonaccorsi, G., Pierri, F., Cinelli, M., Flori, A., Galeazzi, A., Porcelli, F., Schmidt, A.L., Valensise, C.M., Scala, A., Quattrociocchi, W. and Pammolli, F., 2020. Economic and social consequences of human mobility restrictions under COVID-19. *Proceedings of the National Academy of Sciences*, 117(27), pp.15530-15535.
2. Darville, S., July 23, 2020. Reopening Schools Is Way Harder Than It Should Be. N.Y. Times (https://www.nytimes.com/2020/07/23/sunday-review/reopening-schools-coronavirus.html) (last accessed. August 09, 2020)
3. Tingley, K., July 29, 2020. Why Is There No Consensus About Reopening Schools? N. Y. Times (https://www.nytimes.com/2020/07/29/magazine/schools-reopening-covid.html) (last accessed. August 09, 2020)
4. Goodnough, A., July 24, 2020. C.D.C. Calls on Schools to Reopen, Downplaying Health Risks. N. Y. Times (https://www.nytimes.com/2020/07/24/health/cdc-schools-coronavirus.html) (last accessed. August 09, 2020)
5. Roy, A., August 07, 2020. Why It's (Mostly) Safe to Reopen the Schools. The Wall Street Journal (https://www.wsj.com/articles/why-its-mostly-safe-to-reopen-the-schools-11596812466) (last accessed. August 09, 2020)
6. Madhusoodanan, J., July 20, 2020. 'Ethically troubling.' University reopening plans put professors, students on edge. Science Magazine (https://www.sciencemag.org/careers/2020/07/ethically-troubling-university-reopening-plans-put-professors-students-edge) (last accessed. August 09, 2020)
7. Nadworny, W., July 22, 2020. Colleges Spent Months Planning For Fall, But A COVID-19 Surge Is Changing Everything. NPR.ORG (https://www.npr.org/2020/07/22/893525083/colleges-spent-months-planning-for-fall-but-a-covid-19-surge-is-changing-everyth) (last accessed. August 09, 2020)
8. Strauss, V., July 14, 2020. This college president won't open campus this fall, to avoid 'inevitable spread' of coronavirus. Washington Post (https://www.washingtonpost.com/education/2020/07/14/this-college-president-wont-open-campus-this-fall-avoid-inevitable-spread-covid-19/)
9. Hartocollis, A., July 03, 2020. Colleges Face Rising Revolt by Professors. N.Y. Times (https://www.nytimes.com/2020/07/03/us/coronavirus-college-professors.html)
10. Center for Disease Control (CDC), May 30, 2020. Considerations for Institutions of Higher Education. (https://www.cdc.gov/coronavirus/2019-ncov/community/colleges-universities/considerations.html)
11. Stanglin, D., and Shannon, J., August 08, 2020. 250 students and staff asked to quarantine in Georgia district after one week of school. USA Today (https://www.usatoday.com/story/news/education/2020/08/08/hundreds-students-staff-quarantine-georgia-school-district-cherokee/3327173001/)
12. Hawkins, D., August 09, 2020. Nine people test positive for coronavirus at Georgia school where photos of packed hallways went viral. Washington Post (https://www.washingtonpost.com/education/2020/08/09/nine-people-test-positive-coronavirus-georgia-school-where-photos-packed-hallways-went-viral/)
13. Furukawa, N.W., Brooks, J.T. and Sobel, J., 2020. Evidence supporting transmission of severe acute respiratory syndrome coronavirus 2 while pre-symptomatic or asymptomatic. *Emerging Infectious Diseases*, 26(7).
14. Flaxman, S., Mishra, S., Gandy, A., Unwin, H.J.T., Mellan, T.A., Coupland, H., Whittaker, C., Zhu, H., Berah, T., Eaton, J.W. and Monod, M., 2020. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature*, pp.1-5.
15. Heidt, A., Aug 17, 2020. FDA Approves New Saliva-Based COVID-19 Test (SalivaDirect, an open-source protocol, avoids many of the supply bottlenecks of other tools and could be offered for as little as $10 a test). *The Scientist*. (https://www.the-scientist.com/news-opinion/fda-approves-new-saliva-based-covid-19-test-67829)

medRxiv preprint doi: https://doi.org/10.1101/2020.09.04.20188680; this version posted October 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

16. Levenson, E. August 18, 2020. UNC-Chapel Hill reverses plans for in-person classes after 130 students test positive for Covid-19. CNN Online (https://www.cnn.com/2020/08/17/us/coronavirus-college-university/index.html)
17. Rockett, R.J., Arnott, A., Lam, C., Sadsad, R., Timms, V., Gray, K.A., Eden, J.S., Chang, S., Gall, M., Draper, J. and Sim, E.M., 2020. Revealing COVID-19 transmission in Australia by SARS-CoV-2 genome sequencing and agent-based modeling. *Nature Medicine*, pp.1-7.
18. Kai, D., Goldstein, G.P., Morgunov, A., Nangalia, V. and Rotkirch, A., 2020. Universal masking is urgent in the covid-19 pandemic: Seir and agent based models, empirical validation, policy recommendations. *arXiv preprint* arXiv:2004.13553.
19. Wellendorf, K., August 18, 2020. 175 Iowa State students tested positive for COVID-19 over on-campus move-in period (https://www.amestrib.com/story/news/education/2020/08/18/covid-19-iowa-coronavirus-175-iowa-state-ames-students-tested-positive-during-move/3394667001/)
20. Lalvani, A., Pathan, A.A., Durkan, H., Wilkinson, K.A., Whelan, A., Deeks, J.J., Reece, W.H., Latif, M., Pasvol, G. and Hill, A.V., 2001. Enhanced contact tracing and spatial tracking of Mycobacterium tuberculosis infection by enumeration of antigen-specific T cells. *The Lancet*, 357(9273), pp.2017-2021.
21. Indiana University, August 27, 2020. IU takes action to stop COVID spread at Greek houses. (https://news.iu.edu/stories/2020/08/iub/releases/27-greek-houses-covid-quarantine-mitigation.html)
22. New York Times, August 26, 2020. Tracking Coronavirus Cases at U.S. Colleges and Universities. (https://www.nytimes.com/interactive/2020/us/covid-college-cases-tracker.html)
23. Bradley, E.H., An, M.W. and Fox, E., 2020. Reopening Colleges During the Coronavirus Disease 2019 (COVID-19) Pandemic—One Size Does Not Fit All. *JAMA Network Open*, 3(7), pp.e2017838-e2017838.
24. Chang, K., September 10, 2020. A University Had a Great Coronavirus Plan, but Students Partied On: An unexpected upswing in positive tests at the University of Illinois at Urbana-Champaign showed how even the most comprehensive approaches to limiting the virus's spread can break down. Nytimes (https://www.nytimes.com/2020/09/10/health/university-illinois-covid.html)