The Case for Altruism in Institutional Diagnostic Testing

Ivan Specht[1,2†⋆], Kian Sani[1,3†], Yolanda Botti-Lodovico[1,4], Michael Hughes[5], Kristin Heumann[5], Amy Bronson[5], John Marshall[5], Emily Baron[6], Eric Parrie[6], Olivia Glennon[7], Ben Fry[7], Andrés Colubri[1,8‡⋆], Pardis C. Sabeti[1,3,4,9,10‡⋆]

March 24, 2021

Abstract

Amid COVID-19, many institutions deployed vast resources to test their members regularly for safe reopening. This self-focused approach, however, not only overlooks surrounding communities but also remains blind to community transmission that could breach the institution. To test the relative merits of a more altruistic strategy, we built an epidemiological model that assesses the differential impact on case counts when institutions instead allocate a proportion of their tests to members' close contacts in the larger community. We found that testing outside the institution benefits the institution in all plausible circumstances, with the optimal proportion of tests to use externally landing at 45% under baseline model parameters. Our results were robust to local prevalence, secondary attack rate, testing capacity, and contact reporting level, yielding a range of optimal community testing proportions from 18% to 58%. The model performed best under the assumption that community contacts are known to the institution; however, it still demonstrated a significant benefit even without complete knowledge of the contact network.

1 Introduction

During any societal crisis, altruism has the potential to both satisfy moral duty and maximize "utility," leading to the best possible outcome for the greatest number of people [24]. It gains newfound urgency and utility during a pandemic, when important decisions must be made around allocating scarce resources, such as tests, therapies, and vaccines. In these instances more than ever, our own interests—our health, safety, and well-being—become highly interdependent on those of

[1]The Broad Institute of MIT and Harvard, Cambridge, MA, 02142, USA.
[2]Harvard College, Faculty of Arts and Sciences, Harvard University, Cambridge, MA, 02138, USA.
[3]FAS Center for Systems Biology, Department of Organismic and Evolutionary Biology, Faculty of Arts and Sciences, Harvard University, Cambridge, MA, 02138, USA.
[4]Howard Hughes Medical Institute, Chevy Chase, MD, 20815, USA.
[5]Colorado Mesa University, Grand Junction, CO, 81501, USA.
[6]COVIDCheck Colorado, Denver, CO, 80202, USA.
[7]Fathom Information Design, Boston, MA, 02114, USA.
[8]Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, MA 01655, USA.
[9]Massachusetts Consortium on Pathogen Readiness, Harvard Medical School, Harvard University, Boston, MA, 02115, USA.
[10]Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA, 02115, USA.
[†]These authors contributed equally.
[‡]These authors jointly supervised this project.
[⋆]Corresponding authors: ispecht@broadinstitute.org, andres.colubri@umassmed.edu, pardis@broadinstitute.org

others. Specifically for communicable diseases, testing is patently a public good because a positive result can reduce others' exposure and suffering by guiding isolation and quarantine practices.

Considerations of altruism and its efficacy have resurfaced in various COVID-19 response plans worldwide. As the disease began to spread in the U.S., it forced schools and businesses to cease in-person operations to mitigate its spread. To reopen, many of these institutions rushed to test their own members, enacting several-times-per-week or even daily testing protocols in hopes of preventing outbreaks [15]. Countless institutions spent millions on internal testing programs. Some universities, for instance, spent upwards of $1-2 million per week to test students and staff, often using clinical-grade diagnostics [3, 20]. Meanwhile, communities surrounding these institutions continued to struggle with ongoing clinical testing shortages and long delays for results. Even for institutional testing programs that considered supporting community testing, legal and regulatory barriers served as an additional deterrent from doing so.

To turn inward is a common and understandable approach during any crisis, but these expensive self-focused testing programs still left institutions blind to community cases that could potentially enter and spread like wildfire. For example, the NFL spent $100 million in total throughout the 2020-21 season on nearly one million tests for around 7,500 institutional members, administered daily [19]. Yet, the League still experienced several outbreaks [22]. They were not alone; outbreaks occurred within many similar testing programs, as the world witnessed prominently at the White House in Fall 2020.

We are now faced with the question of whether the confined use of significant resources to enable high-frequency testing within individual institutions alone is the most appropriate or effective way to contain a virus. We hypothesized that if institutions test altruistically—that is, designate a substantial portion of their testing capacity outside the institutions—it would not only be good for their communities, but also for them. That is, there would be lower case counts in these institutions themselves had their programs incorporated the testing of close contacts of its members into its testing strategy, thereby detecting potential COVID-19 encroachment.

This paper seeks to ascertain whether a self-focused or an altruistic testing approach is a more effective mitigation strategy. We provide a simple yet plausible epidemiological model to answer this question, comparing results under varying local community prevalence levels, social mitigation efforts, testing capacity, contact tracing adoption, and other parameters. We then discuss the significant real-world implications of our findings concerning how institutions might better allocate their available testing capacity

2 Epidemiological Model

To test our central claim, we built an agent-based epidemiological model of a hypothetical institution such as a workplace, school, or similar organization, accounting for interactions between within the institution as well as between the institution and its surrounding community. For a full, mathematically rigorous methods section, see Appendix A; here, we provide an intuitive, high-level description. We modeled two distinct groups: (1) institution members and (2) all of their first-degree close contacts outside the institution (hereafter referred to as the 'periphery'). We assumed that the periphery remains unchanged throughout the simulation. We then assessed the effect of redistributing some of the institution's testing capacity to the periphery, assuming for simplicity that diagnostic testing in the community was negligibly rare before the institution's intervention. The model provided critical insight into the optimal proportion of tests to redistribute, given several baseline assumptions about viral dynamics, prevalence, and more. Moreover, we assumed no

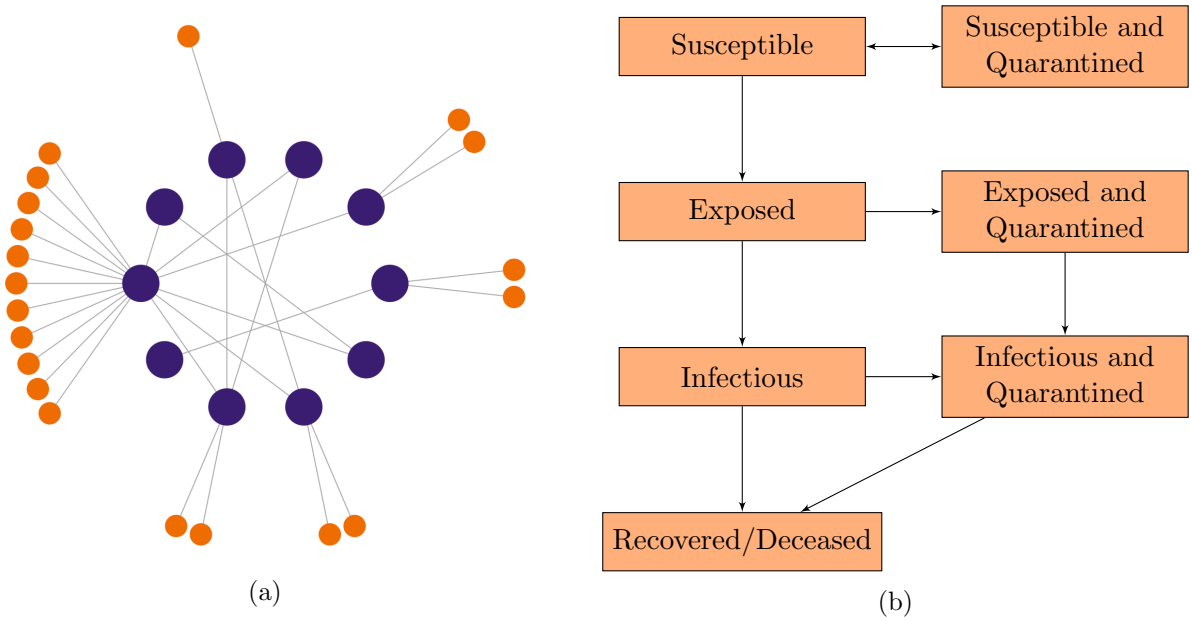


Figure 1: (a) Example of a contact network representing members of the institution (large, purple nodes) and their contacts in the periphery (small, orange nodes). Here we have 10 institution members who make an average of 2 contacts within the institution and 2 contacts outside the institution (variance = 3 for both distributions). (b) Flowchart of compartments and possible state transitions.

knowledge of the institution and the periphery beyond what institutional administrators/health officers might reasonably gather, such as the number of individuals, their frequent contacts, and the number of tests conducted. As such, our model gives a general framework by which institutions may assess possible testing protocols' effectiveness.

Modeling viral propagation between an institution and its periphery requires detailed information on how the agents involved interact with each other. A typical means to capture this information is a simple undirected graph, in which two nodes (i.e., people) share an edge if and only if those two people interact during the modeled period [1]. In a real-world context, we might construct this graph by, for example, surveying members of the institution and its surrounding community about their social interactions. For our model, however, we assumed knowledge only of the mean and variance in contact numbers inside and outside the institution, which is likely more feasible to estimate in most contexts.

We proceeded by modeling N agents who interact according to a random graph, in which node degrees follow an overdispersed Negative Binomial distribution fit to the observed mean and variance (see Appendix A for the random graph generation algorithm) [12]. To account for institution-periphery interactions, we assigned each agent a number of regular contacts made outside the institution, drawn from a different Negative Binomial distribution. See Figure 1a for an example of a contact network generated via this method. This contact network may or may not be fully known to the institution; we included a model parameter that captures the extent to which agents report their contacts.

Having established a graphical contact network, we experimented with different testing strategies to simulate the spread of COVID-19. As a first step, we assumed that the periphery exhibits

an epidemiological steady-state, computed based on the proportion of tests distributed there. By steady-state, we mean that the probability of an individual in the periphery being in a certain epidemiological state remains constant over the course of the simulation (see Appendix C for a more detailed treatment of this assumption). Within the institution, by contrast, we set the initial infection rate low in comparison to the periphery, reflective of the fact that many institutions returning to in-person activities have rigorous testing/quarantining protocols. For these agents, we modeled viral states probabilistically via an adapted N-intertwined mean-field approximation (NIMFA). In its original form, the NIMFA states that at a given time, the rate of transmission from agent j to adjacent agent i is proportional to the product of the marginal probabilities that i is susceptible and that j is infectious [17, 18]. This approach offers the flexibility to encompass a wide range of epidemiological compartments while capturing the granularity of graphical contact networks [13, 21, 25].

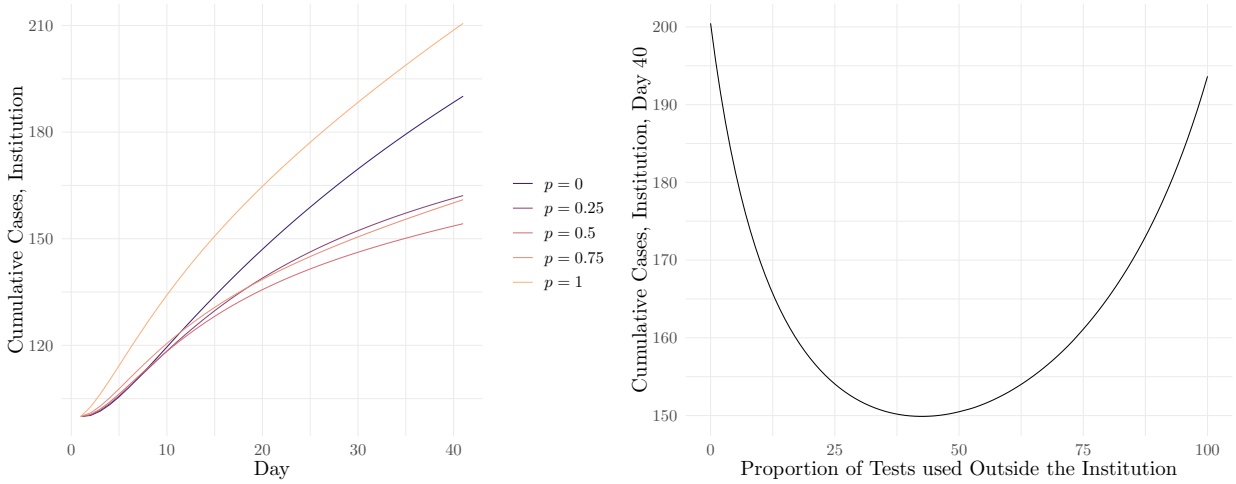
We then extended NIMFA to include testing-based interventions. First, we allowed the outgoing transmission rate to vary between agents as a partial means of modeling overdispersion (with the rest coming from the node degree distribution). To account for quarantine compartments, we set the time-dependent rate at which agent j enters quarantine proportional to the product of the marginal probabilities that j is infectious and that j receives a test. In turn, this latter probability depended on the test distribution strategy, the probability that j had not previously tested positive, and the number of adjacent nodes in j's contact network. Finally, accounting for the fact that COVID-19 cases exhibit an exposed (but not yet infectious) stage, we arrived at detailed compartmental model, in which agents transition between epidemiological states as depicted in Figure 1b.

3 Results

We first applied our model to a mid-sized university (N = 10,000), using real data we gathered at Colorado Mesa University (CMU). CMU established a testing program in summer 2020 initially focused on university students and staff, and began supporting testing in the greater Mesa County community later that year [4]. Contact tracers determined that the mean and variance of the number of close contacts within the institution were 2.3 and 2.4, respectively, and outside the institution were 0.2 and 1.9, respectively. They also found that the prevalence on campus at the beginning of the Spring 2020 semester was approximately 1%, and that they planned to conduct about 0.12 diagnostic tests per day per person. Supplementing our own data collection with that of the local public health authority, we compiled a complete set of parameters specific to CMU and ran the model accordingly (see Appendix D).

Our initial model results based on the CMU parameters provided strong evidence in support of an altruistic testing strategy. We observed that the projected number of cases 40 days after the beginning of the modeled period was lowest when CMU deployed 45% of its tests to the periphery (see Figure 2). This strategy reduced the institutional case count by 25% as compared to a self-focused testing strategy (i.e. 0% peripheral testing). However, our data from CMU—which informed our baseline parameters—were likely subject to several biases. While CMU administrators attested that they believed the data to represent the student body fairly well, students who contracted the virus (as every individual in our dataset did) were likely to have higher degrees of social interaction than those who did not, leading to a positive bias. On the other hand, CMU informed us that certain close contacts were likely not reported or otherwise not included in some cases, introducing a negative bias.

Because of the potential limitations and biases of our CMU data, and because our model relies



(a) (b)

Figure 2: (a) Modeled cumulative cases over time at CMU under 5 different proportions p of peripheral testing; (b) cumulative cases on day 40 as a function of the proportion of tests deployed to the periphery, with the minimum at 45% peripheral testing

on numerous parameters that vary widely between institutions, we proceeded by demonstrating robustness to and characterizing the influence of several factors on our results. These included four key factors: local community prevalence, social mitigation efforts, testing capacity, and contact tracing adoption (see Figure 3). Of course, there are more factors to assess, including distributions in numbers of contacts, variance in transmissibility, and initial prevalence within the institution. For an assessment of these factors, see Appendix B, or, for an interactive sensitivity analysis, visit `https://ispecht.shinyapps.io/covid19-altruistic-testing/`.

We first assessed sensitivity to the prevalence of COVID-19 among members of the periphery (V_{0*}). While the 6% positivity rate in Mesa County in January 2020 was reflective of national statistics at the time, different institutions faced significantly higher or lower caseloads in their surrounding communities, representative of factors such as density and public health resources that varied from place to place. Beyond that, rates have naturally varied over the course of the outbreak. We observed that under higher values of V_{0*}, the effectiveness of redistributing tests to the periphery grew, with the optimal proportion increasing from 24% when $V_{0*} = 2\%$ to 58% when $V_{0*} = 16\%$ (see Figure 3a). This finding is unsurprising, given that the probability of any individual test detecting a case in the periphery grew with higher values of V_{0*}. In turn, the resulting quarantine (of both the peripheral member and their contact in the institution) minimized the probability of the virus breaching the institution.

Next, we turned to social mitigation efforts. While several model parameters capture these efforts, we focused on the secondary attack rate (SAR) for institution members—that is, the probability of transmission between an infectious individual and a susceptible one. Individuals may decrease the proportion of contacts they infect by, for example, wearing masks and socially distancing—thereby decreasing the observed SAR. Many institutions employ different mitigation measures, but altruism proved an effective strategy under a wide range of values for the observed SAR among institution members, denoted μ_ρ (see Figure 3b). For $\mu_\rho = 0.05$, the optimal proportion of peripheral testing lay at 54%; this proportion decreased to 25% when $\mu_\rho = 0.4$. This again fell in line with our expectations, as for higher μ_ρ values, a single case within the institution had much greater potential

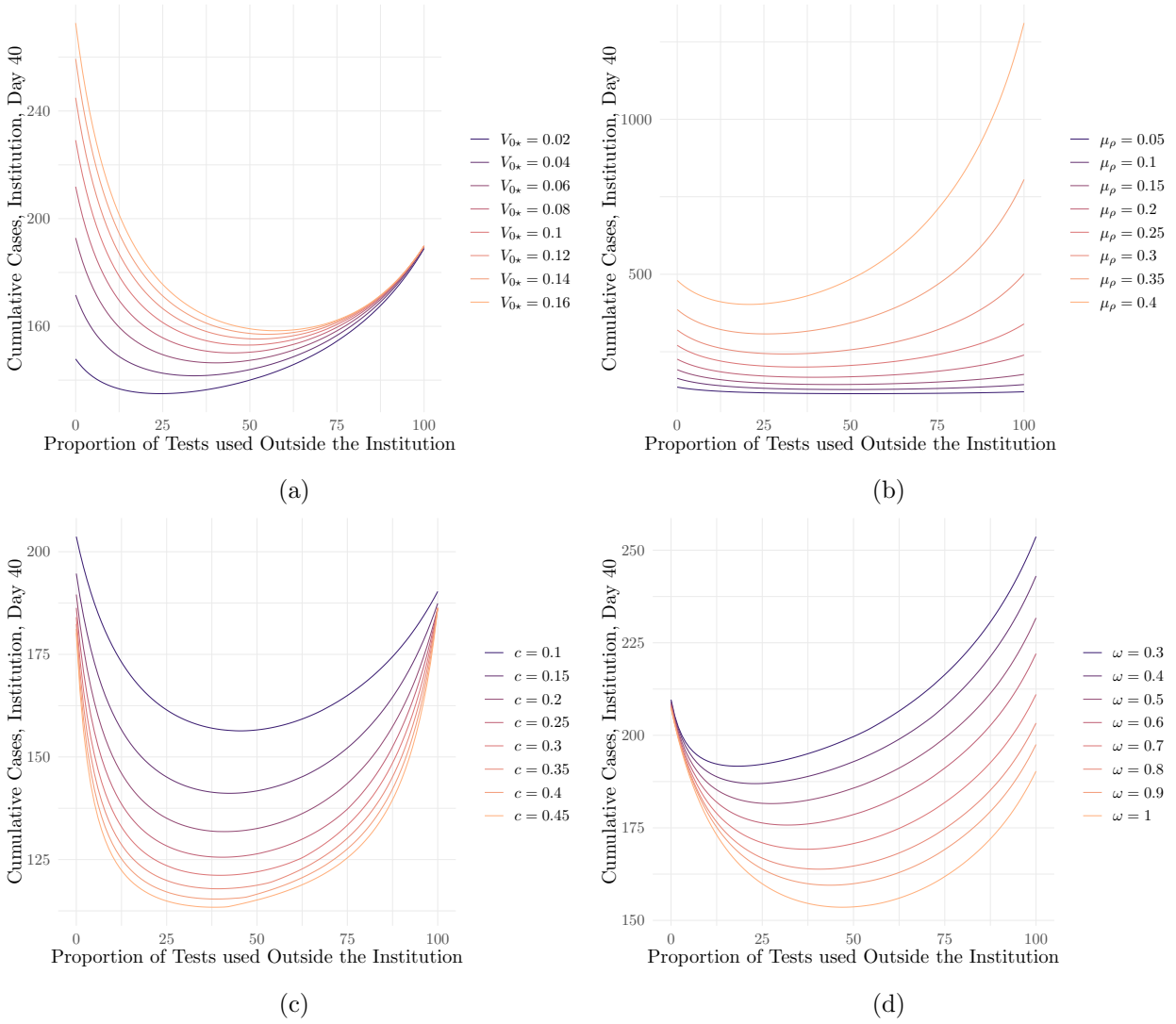


Figure 3: Cumulative cases on day 40 as a function of the proportion of tests deployed to the periphery under different values of (a) the initial prevalence in the periphery, V_{0*}; (b) the secondary attack rate among institution members, μ_ρ; (c) the tests-per-person-per-day ratio, c, and (d) the proportion of contacts traced ω

to spread, limiting the effectiveness of peripheral testing and lowering the optimal proportion of tests to be administered outside the institution. On the other hand, when μ_ρ was low, a single case within the institution likely did not give rise to an outbreak. This made institutional testing less essential, leaving more capacity to establish a 'barrier of defense' in the periphery to prevent cases from breaching the institution in the first place.

We then focused on c, the number of tests administered by the institution per day per person. An intuitive way to think about this parameter is that on average, an institution member receives a diagnostic test every $1/c$ days. We found that test redistribution to the periphery remained an effective strategy even for relatively low values of c, such as $c = 10\%$ (see Figure 3c). The optimal proportion of peripheral testing stayed relatively constant, ranging from about 38% to 48% for values of c between 45% to 10%. This result reflected the fact that under our baseline CMU parameters, institution members average about 10 times more contacts within the institution than outside it. As such, the size of the periphery was small, limiting the possible pathways for the virus to breach the institution. Our results tell us that even with limited testing resources, tests would best be used to prevent the virus from entering the institution via these pathways. Note that we also investigated different distributions of contacts within and outside the institution; for an analysis, see Appendix B, Figures S1a-S1d.

Finally, we accounted for the fact that institution members may not report all of their contacts in the periphery, or may have contacts that the institution cannot test due to factors such as geographic disparity. We captured the proportion of reported, testable contacts with the parameter ω. As our results suggest, contact reporting needed not be perfect for peripheral testing to help curb viral spread (see Figure 3d). Even if institution members reported only 30% of their contacts, the optimal proportion of peripheral testing lay at 18%; this proportion increases to 45% as the fraction of reported contacts grows to 100%. This result makes sense because the most socially-active—and therefore riskiest—members of the institution had many peripheral contacts, at least some of which will likely be known to the institution even under imperfect contact tracing (e.g. familial contacts). A positive test from even just one of these contacts would send the original institution member into quarantine, allowing the 'barrier of defense' strategy to remain an effective means of protecting the institution.

We note that the model exhibited some slight variance between runs due to the stochastic nature of the contact network generation step. While such stochasticity slightly affected numerical values for case counts between model runs, the shapes of the curves in Figures 2-3 remained consistent.

4 Discussion

Our model supports our hypothesis that the altruistic approach—in which institutions test beyond their walls—is the most effective protection strategy. In every instantiation of the model, we observed that deploying some proportion of diagnostic tests to the periphery significantly reduces the cumulative caseload at the end of a 40-day period. The optimal proportion of peripheral testing was 45% under baseline parameters and ranged from 18% to 58% under different levels of local prevalence, social interaction, testing, and contact tracing.

Our methods serve as a general framework for modeling one specific population within the context of another, and we hope that further research may help refine the intricacies of such dynamics. We also hope our work provides justification for institutions to consider implementing an altruistic testing strategy, and for legal and regulatory bodies to create a path for them to do so. We encourage institutions to partner with local public health authorities to support testing or connect

members of the periphery with the appropriate testing provider, as Colorado Mesa University and the University of California Davis have done [4, 8].

Our results urge a fundamental rethinking of how institutions with substantial testing capacity approach safety amid outbreaks. Epidemics are one of those rare instances where a seemingly selfless approach is, in fact, the most self-serving: institutions must help test beyond their walls to stay safe within them.

Ethics Statement

The study received a "Not Human Subjects Research" determination by the Broad Institute's Office of Research Subject Protection.

Code Availability

Data from CMU and code used to implement the model and generate Figure 2a is available at `https://github.com/broadinstitute/covid19-altruistic-testing`.

Conflicts of Interest

Pardis C. Sabeti is a co-founder and shareholder of Sherlock Biosciences and is a non-executive board member and shareholder of Danaher Corporation. Andrés Colubri and Pardis C. Sabeti are inventors on patents related to diagnostics and Bluetooth-based contact tracing tools and technologies filed with the USPTO and other intellectual property bodies.

Acknowledgements

We would like to acknowledge the following individuals for their support and feedback for this work: Kristian G. Andersen, Tim Foster, Gabrielle Gionet, William Hanage, Daniel Larremore, David O'Connor, Roy R. Parker, and Megan Vodzak, as well as members of the Sabeti Lab, the Broad Institute, Colorado Mesa University, COVIDCheck Colorado, and Fathom Information Design. This work was made possible by the Gordon and Betty Moore Foundation, Howard Hughes Medical Institute, Flu Lab, and a cohort of generous donors through TED's Audacious Project, including the ELMA Foundation, MacKenzie Scott, the Skoll Foundation, and Open Philanthropy.

References

[1] Bansal S, Grenfell BT and Meyers LA (2007). When individual behaviour matters: homogeneous and network models in epidemiology. J. R. Soc. Interface. 4: 879–891. http://doi.org/10.1098/rsif.2007.1100.

[2] Billah MA, Miah MM, Khan MN (2020) Reproductive number of coronavirus: A systematic review and meta-analysis based on global level evidence. PLoS ONE 15(11): e0242128. doi:10.1371/journal.pone.0242128

[3] COVID-19 Information. (2021, February 24). Weekly COVID-19 tests. Retrieved March 15, 2021, from https://www.harvard.edu/coronavirus/testing-tracing/harvard-university-wide-covid-19-testing-dashboard/weekly-covid-19-tests/

[4] COVID-19 testing. (2021). Retrieved March 15, 2021, from https://www.coloradomesa.edu/covid-19/testing/index.html

[5] Endo A, Abbott S, Kucharski AJ, Funk S. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. Wellcome Open Res. 2020;5:67.

[6] He X, Lau EHY, Wu P, et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat Med 26, 672–675 (2020). https://doi.org/10.1038/s41591-020-0869-5

[7] Hinch R, Probert WJM, Nurtay A, et al., OpenABM-Covid19 - an agent-based model for non-pharmaceutical interventions against COVID-19 including contact tracing. *medRxiv*. 2020; 2020.09.16.20195925. 10.1101/2020.09.16.20195925. https://www.medrxiv.org/content/early/2020/09/22/2020.09.16.20195925

[8] Hubler, S. (2021, January 30). A California University tries to shield an entire city from coronavirus. Retrieved March 15, 2021, from https://www.nytimes.com/2021/01/30/us/college-coronavirus-california.html

[9] Koh WC, Naing L, Chaw L, Rosledzana MA, Alikhan MF, Jamaludin SA, et al. (2020) What do we know about SARS-CoV-2 transmission? A systematic review and meta-analysis of the secondary attack rate and associated risk factors. PLoS ONE 15(10): e0240205. https://doi.org/10.1371/journal.pone.0240205.

[10] Larremore DB, Wilder B, Lester E, et al. Test sensitivity is secondary to frequency and turnaround time for COVID-19 surveillance. September 8, 2020. https://www.medrxiv.org/content/10.1101/2020.06.22.20136309v2.

[11] Liu Y, Yan L, Wan L, et al. Viral dynamics in mild and severe cases of COVID-19. The Lancet Correspondence 20 (6), 656-657 (2020). https://doi.org/10.1016/S1473-3099(20)30232-2

[12] Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Med. 2008;5(3): e74.

[13] Alam M, Kabir K, and Tanimoto J. Based on mathematical epidemiology and evolutionary game theory, which is more effective: quarantine or isolation policy?. Journal of Statistical Mechanics: Theory and Experiment (2020). 033502. 10.1088/1742-5468/ab75ea.

[14] Nadworny E and McMinn S (2020, October 06). Even in COVID-19 hot Spots, many colleges aren't Aggressively testing students. Retrieved March 15, 2021, from https://www.npr.org/2020/10/06/919159473/even-in-covid-hot-spots-many-colleges-arent-aggressively-testing-students

[15] Paltiel AD, Zheng A, Walensky RP. Assessment of SARS-CoV-2 Screening Strategies to Permit the Safe Reopening of College Campuses in the United States. JAMA Netw Open. 2020;3(7):e2016818. doi:10.1001/jamanetworkopen.2020.16818

[16] Phucharoen C, Sangkaew N, Stosic K, The characteristics of COVID-19 transmission from case to high-risk contact, a statistical analysis from contact tracing data, EClinicalMedicine, Volume 27, 2020, 100543, ISSN 2589-5370, https://doi.org/10.1016/j.eclinm.2020.100543

[17] Prasse, B., Van Mieghem, P. Time-dependent solution of the NIMFA equations around the epidemic threshold. J. Math. Biol. 81, 1299–1355 (2020). https://doi.org/10.1007/s00285-020-01542-6

[18] Qu B, Wang H. (2017) The Accuracy of Mean-Field Approximation for Susceptible-Infected-Susceptible Epidemic Spreading with Heterogeneous Infection Rates. In: Cherifi H., Gaito S., Quattrociocchi W., Sala A. (eds) Complex Networks & Their Applications V. COMPLEX NETWORKS 2016. Studies in Computational Intelligence, vol 693. Springer, Cham. https://doi.org/10.1007/978-3-319-50901-3_40

[19] Seifert, K. (2021, February 12). How the NFL navigated COVID-19 this season: 959,860 tests, $100 million and zero cancellations. Retrieved March 15, 2021, from https://www.espn.com/nfl/story/_/id/30781978/how-nfl-navigated-covid-19-season-959860-tests-100-million-zero-cancellations

[20] Smola, J (2020, September 11). Ohio state spending $2 million per week in federal funds on COVID testing. Retrieved March 15, 2021, from https://www.dispatch.com/story/news/coronavirus/2020/09/11/ohio-state-spending-2-million-per-week-in-federal-funds-on-covid-testing/42583503/

[21] Stollenwerk N, van Noort S,Martins J, Aguiar M, Hilker F, Pinto A & Gomes G. A spatially stochastic epidemic model with partial immunization shows in mean field approximation the reinfection threshold. Journal of Biological Dynamics, 4(6), 634-649 (2010). 10.1080/17513758.2010.487159

[22] The Associated Press. The virus' toll: How 32 NFL teams have been hit by COVID-19. AP (2020). https://apnews.com/article/nfl-technology-tennessee-titans-minnesota-vikings-new-orleans-saints-dfd427d677d3b2e7a995adacda95c39f

[23] Watts, D., Strogatz, S. Collective dynamics of 'small-world' networks. Nature 393, 440–442 (1998). https://doi.org/10.1038/30918

[24] White, M. A Kantian-economic approach to altruism in the household. Palgrave Commun 1, 15005 (2015). https://doi.org/10.1057/palcomms.2015.5

[25] Yang J, Wang G, Zhang S. Impact of household quarantine on SARS-Cov-2 infection in mainland China: A mean-field modeling approach. Mechanical Biosciences and Engineering 17(5): 4500–4512 (2020). 10.3934/mbe.2020248

A Methods

Overview. Here we derive the full set of modeling equations we used to conduct our analysis. We construct an agent-based epidemiological model (ABM) with known contact network in order to compare the effectiveness of various testing regimes at controlling viral spread within an institutional context. Specifically, we model two groups of agents—the members of an "institution" (such as a school, business, etc.) and all of their close contacts outside the institution (henceforth referred to as the "periphery")—and assume the institution has access to a disproportionately-large tests-per-person-per-day ratio in comparison to the periphery. We then analyze the effect of redistributing some of the institution's testing capacity to the periphery. The model provides key insight into the optimal proportion of tests to be redistributed in such a way. More broadly, the set of equations we provide serve as a general framework by which institutions may assess the relative effectiveness of various testing protocols.

We simulate an institution with N agents that interact according to a random graph whose node degree distribution reflects known contact patterns of the institution's members. Let M_i be a random variable representing the number of edges connected to node i. To reflect the overdispersion typically associated with node degrees in social contact networks, we assume that the M_i's are i.i.d. negative binomial with mean μ and variance σ^2 [12]. We can express M_i as a function of the activity levels $\mathbf{a} = (a_1, \ldots, a_N)$ of the N individuals—parameters that represent sociability, whose distribution we will derive. Assuming proportionate mixing, we define ϵ_{ij}, the probability of nodes i and j sharing an edge, to be

$$\epsilon_{ij} = \frac{a_i a_j}{\sum_{k=1}^{N} a_k}$$

for $i \neq j$, and 0 for $i = j$. Assuming that ϵ_{ij} is small in general (an assumption we will justify below), we have that $X_i|\mathbf{A} = \mathbf{a} \dot{\sim} \text{Pois}(a_i)$, where $\dot{\sim}$ means "approximately distributed as." So, we now need to choose a distribution for the A_i's to ensure that the marginal distribution of X_i is Negative Binomial. Since the Negative Binomial distribution may be expressed as the Poisson distribution compounded with the Gamma distribution, we model the A_i's as i.i.d. Gamma(α, λ) random variables. To solve for the parameters, we have that

$$\mu = \mathrm{E}[X_i] = \mathrm{E}[\mathrm{E}[X_i|\mathbf{A}]] = \mathrm{E}[A_i] = \frac{\alpha}{\lambda}$$

$$\sigma^2 = \mathrm{Var}[X_i] = \mathrm{E}[\mathrm{Var}[X_i|\mathbf{A}]] + \mathrm{Var}[E[X_i|\mathbf{A}]] = \mathrm{E}[A_i] + \mathrm{Var}[A_i] = \frac{\alpha}{\lambda} + \frac{\alpha}{\lambda^2}.$$

Thus,

$$\alpha = \frac{\mu^2}{\sigma^2 - \mu} \quad \text{and} \quad \lambda = \frac{\mu}{\sigma^2 - \mu}.$$

Finally, the assumption that $a_i a_j \ll \sum_{k=1}^{N} a_k$ follows from the Cauchy-Schwarz inequality. We have that:

$$\mathrm{E}[\epsilon_{ij}] = \mathrm{E}\left[\frac{A_i A_j}{\sum_{k=1}^{N} A_k}\right] \leq \sqrt{\mathrm{E}[A_i^2 A_j^2]\mathrm{E}\left[\frac{1}{(\sum_{k=1}^{N} A_k)^2}\right]}.$$

For the first expectation within the radical, we have that

$$\mathrm{E}[A_i^2 A_j^2] = \mathrm{E}[A_i^2]\mathrm{E}[A_j^2] = \mathrm{E}[A_i^2]^2 = \frac{\alpha^2(1+\alpha)^2}{\lambda^4}.$$

For the second expectation, let $S = \sum_{k=1}^{N} A_k$. Then $S \sim \text{Gamma}(N\alpha, \lambda)$. Therefore

$$\text{E}\left[\frac{1}{S^2}\right] = \int_0^\infty \frac{f_S(s)}{s^2}\,\text{d}s = \frac{\lambda^2}{(N\alpha-1)(N\alpha-2)}.$$

so long as $N\alpha > 2$. Thus,

$$\text{E}[\epsilon_{ij}] \leq \sqrt{\frac{\alpha^2(1+\alpha)^2}{\lambda^2(N\alpha-1)(N\alpha-2)}} < \sqrt{\frac{\alpha^2(1+\alpha)^2}{\lambda^2(N\alpha-2)^2}} = \frac{\alpha(1+\alpha)}{\lambda(N\alpha-2)}$$

which approaches 0 as $N \to \infty$.

Thus, we may generate our contact network by drawing the a_i's from the Gamma(α, λ) distribution, then drawing a Bern(ϵ_{ij}) random variable for each unordered pair of nodes $\{i, j\}$ to generate the edges of the graph stochastically. Note that the probability of ϵ_{ij} exceeding 1 is small because $P(\epsilon_{ij} \geq 1) \leq \text{E}[\epsilon_{ij}]$ by Markov's inequality; in practice, this never occurred for all parameter combinations studied in this paper.

For contacts outside the insitution, we assume for simplicity that no two agents share a peripheral contact. Letting Y_i be the number of close contacts in the network of agent i outside the institution, we model the Y_i's as i.i.d. Negative Binomial random variables with mean μ_* and variance σ_*^2. Contact reporting outside the institution, however, is likely imperfect—and even among reported contacts, some may be unwilling to receive a diagnostic test. To address this, we introduce the parameter ω, which represents the probability of any given contact outside the institution being traced and testable by the institution. Then, letting Z_i be the number of traced and testable contacts made by agent i outside the institution, we model $Z_i|Y_i \sim \text{Bin}(Y_i, \omega)$. Moving forward, we will treat the contact network as fixed, and thus we will use the notations $\mathbf{x}, \mathbf{y}$, and $\mathbf{z}$ to represent the values taken on by random variables $\mathbf{X}, \mathbf{Y}$, and $\mathbf{Z}$. We refer to the set of all traced and testable people outside the institution as the "known periphery"

Viral propagation model. To capture the structure of a given graphical contact network, we take an agent-based approach for members of the institution; for the periphery, we implement a compartmental model that describes the population at large. Moreover, we assume that the prevalence in the periphery exhibits an epidemiological steady-state, a somewhat atypical assumption that we will explain further below and test as part of our sensitivity analysis (see Figure 3a and Appendix B). Finally, we include inter-compartmental "flux" terms in our modeling equations to account for the propagation of cases from the periphery into the institution.

We begin with the compartmental model for the periphery. We model five compartments: Susceptible (S_*), Exposed (E_*), Infectious (V_*), Exposed/Infectious and Quarantined (Q_*), and Recovered/Deceased (R_*). Note that from now on, we will use the lower star notation for variables/parameters specific to the periphery. For our purposes we take exposed to mean "contracted the virus, but not yet infectious," and assume that every agent enters the exposed stage before entering the infectious stage [6]. Let γ and δ be the recovery rate and exposed-to-infectious transition rates, respectively. Let c be the tests-per-person-per-day ratio, of which a proportion p are designated for members of the known periphery. Assuming for simplicity that each member of the known periphery is equally likely to be tested on any given day, we may set η_*, the daily probability of an agent in the known periphery being tested, as follows:

$$\eta_* = \frac{pcN}{\sum_{i=1}^{N} z_i}. \tag{1}$$

We only need to compute the steady-state probabilities of being exposed and infectious in the periphery, since the other states do not substantively impact our model. We assume that *prior to interventions*, i.e. testing and quarantine, the steady-state probability of being infectious is V_{0*}. Given the transition rates, we may then assume that E_{0*}, the steady-state probability of being exposed prior to interventions, is $\gamma V_{0*}/\delta$. Finally, since γV_{0*} recover per day, it follows that γV_{0*} must transition from susceptible to infected each day to maintain the steady state.

Next, we include the effects of interventions, which affect only the known periphery. COVID-19 diagnostic tests have been shown to be highly reliable during the infectious stage of illness, but often fail to detect the extremely low viral loads present during the exposed staged [6]. As such, we let ψ_E be the test sensitivity during the exposed stage and ψ_V be the test sensitivity during the infectious stage. Modeling the exposed-to-quarantined and infectious-to-quarantined rates as $\eta_*\psi_E$ and $\eta_*\psi_V$, respectively, we arrive at the following system of differential equations:

$$\frac{\mathrm{d}E_*(t)}{\mathrm{d}t} = -\delta E_*(t) - \eta_*\psi_E E_*(t) + \gamma V_{0*} \tag{2}$$

$$\frac{\mathrm{d}V_*(t)}{\mathrm{d}t} = -\gamma V_*(t) - \eta_*\psi_V V_*(t) + \delta E_*(t) \tag{3}$$

where $E_*(t)$ and $V_*(t)$ are the probabilities of an individual in the known periphery being exposed and infected at time t, respectively. Setting both derivatives equal to 0 and solving in terms of V_{0*}, we obtain the steady states, which we denote $E_{\infty*}$ and $V_{\infty*}$, respectively:

$$E_{\infty*} = \frac{\gamma V_{0*}}{\eta_*\psi_E + \delta} \tag{4}$$

$$V_{\infty*} = \frac{\gamma\delta V_{0*}}{(\eta_*\psi_V + \gamma)(\eta_*\psi_E + \delta)} \tag{5}$$

In the "unknown periphery" (i.e. members of the periphery who institution members do not report), we may assume that the the probabilities of being exposed and infected are E_{0*} and V_{0*}, respectively.

We now turn our focus to the institution. Given the level of granularity provided by contact networks, along with the difficulty in capturing agent-specific conditions through more traditional epidemiological methods (i.e. compartmental modeling for the population at large), we model the state of each agent in the institution probabilistically using N-Intertwined Mean-Field Approximation (NIMFA) [17]. A widely-used and computationally-efficient approximation for the true stochastic process, the NIMFA approximates the joint probability of some agent i being susceptible and some other agent j being infectious as the product of the marginals. In its original form, the NIMFA states that:

$$\frac{\mathrm{d}v_i(t)}{\mathrm{d}t} = -\gamma v_i(t) + s_i(t)\sum_{j=1}^{N} \mathbb{1}_{ij}\beta_j v_j(t) \tag{6}$$

where $v_i(t)$ and $s_i(t)$ are the probabilities of agent i being infectious and susceptible at time t, respectively; $\mathbb{1}_{ij}$ is the indicator of an edge between i and j; γ is the recovery rate (as before), and β_j is the probability of transmission between infectious agent j and a susceptible agent per unit time.

We allow for heterogeneity in $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_M)$, where M is the total number of agents in both the institution and the periphery, to account for part of the overdispersion in COVID-19 transmission (the other part coming the node degree distribution in our graphical model). To derive $\boldsymbol{\beta}$, we begin by drawing $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_M)$, the (heterogeneous) secondary attack rate specific each agent, from a Beta distribution with mean μ_ρ and variance σ_ρ^2. Since we defined ρ_i to be the probability of transmission over the course of agent i's infection, and since the probability density of recovery at time t is equal to $\gamma \exp(-\gamma t)$, we may write the total probability of transmission ρ_i in terms of the daily probability of transmission β_i:

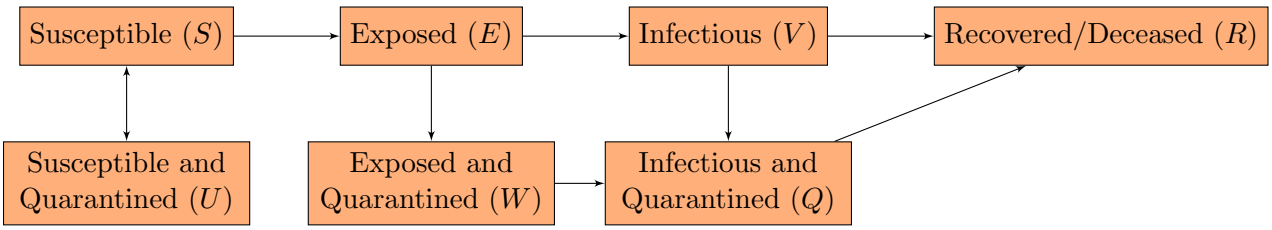
$$\rho_i = \int_0^\infty (1 - (1 - \beta_i)^t)\gamma \exp(-\gamma t)\, \mathrm{d}t. \tag{7}$$

And, solving for β_i, we obtain:

$$\beta_i = 1 - \exp\left(-\frac{\gamma \rho_i}{1 - \rho_i}\right). \tag{8}$$

Note that in the case of COVID-19, limited literature exists on the variation in the secondary attack rate between agents, and the extent to which this phenomenon influences overdispersion is not known. As such, our distribution of the ρ_i's will be weakly informed, so we assess sensitivity to μ_ρ and σ_ρ^2 in Appendix B.

Having solved for $\boldsymbol{\beta}$, we now adapt the NIMFA to include additional transmission pathways and compartments for each individual agent while staying true to the original idea. Within the institution, we model seven compartments: Susceptible (S), Exposed (E), Infected (V), Recovered/Deceased (R), Susceptible and Quarantined (U), Exposed and Quarantined (W), and Infected and Quarantined (Q). See below for a diagram of all possible state transitions:



For any of the above states $Z \in \{S, E, V, R, U, W, Q\}$, we use the notation $z_i(t)$ to mean the probability that individual i is in state Z at time t. Since our agent-based model applies only to members of the institution, we allow i to range over integers only from 1 to N, inclusive. Revising (6) to account for the "exposed" compartment and for each agent's outside-the-institution contacts, and then applying a Poisson approximation (to avoid the fact that the sum of the transmission probabilities may exceed 1), we obtain:

$$\frac{\mathrm{d}e_i(t)}{\mathrm{d}t} = -\delta e_i(t) + s_i(t)\left(1 - \exp\left(-\sum_{j=1}^{M} \mathbb{1}_{ij}\beta_j v_j(t)\right)\right) \tag{9}$$

$$\frac{\mathrm{d}v_i(t)}{\mathrm{d}t} = -\gamma v_i(t) + \delta e_i(t). \tag{10}$$

Note that for $j > N$, we take $v_j(t)$ to equal $V_{\infty *}$ if j is a member of the known periphery and V_{0*}, otherwise. For ease of notation, from now on we let

$$\mathrm{B}_i(t) = \left(1 - \exp\left(-\sum_{j=1}^{M} \mathbb{1}_{ij}\beta_j v_j(t)\right)\right).$$

Next, we revise (9) and (10) to include the impact of quarantine due to a positive test result. To do so, we first need to decide on a prioritization for diagnostic tests within the institution. Let $\eta_i(t)$ be the probability that agent i is tested at time t. It is natural that we model this probability as roughly proportional to the number of known close contacts agent i has (both within and outside the institution) times the probability that agent i has not already been identified as a positive case. However, we also need to satisfy that the number of tests administered within the institution per unit time always sums to $(1-p)cN$, and choosing a constant of proportionality to satisfy this condition may cause $\eta_i(t)$ to be greater than 1 for certain i. As such, we instead model $\eta_i(t)$ as

$$\eta_i(t) = 1 - \exp\Big(-\alpha(t)(x_i + z_i)(s_i(t) + e_i(t) + v_i(t))\Big) \tag{11}$$

with $\alpha(t)$ chosen to satisfy

$$\sum_{i=1}^{N} \eta_i(t) = (1-p)cN. \tag{12}$$

Note that no closed-form solution for $\alpha(t)$ exists; in practice, we compute it numerically for each discrete timestep Δt. Just as in the model for the periphery, we then set the exposed-to-quarantined and infectious-to-quarantined rates (due to a positive test result) equal to $\eta_i(t)\psi_E$ and $\eta_i(t)\psi_V$, respectively. Now accounting for the quarantine rate due to a positive test result, we obtain:

$$\frac{\mathrm{d}e_i(t)}{\mathrm{d}t} = -\delta e_i(t) - \eta_i(t)\psi_E e_i(t) + s_i(t)\mathrm{B}_i(t) \tag{13}$$

$$\frac{\mathrm{d}v_i(t)}{\mathrm{d}t} = -\gamma v_i(t) - \eta_i(t)\psi_V v_i(t) + \delta e_i(t) \tag{14}$$

$$\frac{\mathrm{d}w_i(t)}{\mathrm{d}t} = -\delta w_i(t) + \eta_i(t)\psi_E e_i(t) \tag{15}$$

$$\frac{\mathrm{d}q_i(t)}{\mathrm{d}t} = -\gamma q_i(t) + \eta_i(t)\psi_V v_i(t) + \delta w_i(t). \tag{16}$$

Finally, we include the effect of quarantine due to a positive test result from an agent's first-degree contact. Again applying Poisson approximation to the events of each institutional agent's first-degree contacts testing positive, we may write the time-dependent rate $\Gamma_i(t)$ at which agent i enters quarantine due to a known neighbor testing positive as:

$$\Gamma_i(t) = \left(1 - \exp\left(-\eta_* z_i(\psi_E E_{\infty *} + \psi_V V_{\infty *}) - \sum_{j=1}^{N} \mathbb{1}_{ij}\eta_j(t)(\psi_E e_j(t) + \psi_V v_j(t))\right)\right).$$

Revising (13-16) once again and adding in the remaining states, we arrive at our complete set of differential equations:

$$\frac{\mathrm{d}s_i(t)}{\mathrm{d}t} = -s_i(t)\Gamma_i(t) - s_i(t)\mathrm{B}_i(t) + \eta_i(t)u_i(t) \tag{17}$$

$$\frac{\mathrm{d}e_i(t)}{\mathrm{d}t} = -e_i(t)\Gamma_i(t) - \delta e_i(t) - \eta_i(t)\psi_E e_i(t) + s_i(t)\mathrm{B}_i(t) \tag{18}$$

$$\frac{\mathrm{d}v_i(t)}{\mathrm{d}t} = -v_i(t)\Gamma_i(t) - \gamma v_i(t) - \eta_i(t)\psi_V v_i(t) + \delta e_i(t) \tag{19}$$

$$\frac{\mathrm{d}u_i(t)}{\mathrm{d}t} = -\eta_i(t)u_i(t) + s_i(t)\Gamma_i(t) \tag{20}$$

$$\frac{\mathrm{d}w_i(t)}{\mathrm{d}t} = -\delta w_i(t) + e_i(t)\Gamma_i(t) + \eta_i(t)\psi_E e_i(t) \tag{21}$$

$$\frac{\mathrm{d}q_i(t)}{\mathrm{d}t} = -\gamma q_i(t) + v_i(t)\Gamma_i(t) + \eta_i(t)\psi_V v_i(t) + \delta w_i(t) \tag{22}$$

$$\frac{\mathrm{d}r_i(t)}{\mathrm{d}t} = \gamma v_i(t) + \gamma q_i(t). \tag{23}$$

For the analyses conducted in this paper, we solved this system of differential equations using a discrete timestep of $\Delta t = 1$ day.

B Further Sensitivity Analysis

Here we build on the analysis in the Results section by assessing sensitivity to μ, σ^2, μ_*, σ_*^2, σ_ρ^2, and $\mathbf{v}_0$, using a one-at-a-time approach. As in the Results section, for each parameter, we selected a range of possible values under which to plot cumulative cases on day 40 as a function of p (the proportion of tests deployed outside the institution). We found that the strategy of deploying at least some proportion of tests to the periphery is robust to a wide variety of parameter combinations (see Figure S1). Note that we observed some noise in our analyses of μ_* and σ_*^2 due to the skewness in the distribution of peripheral contacts; nonetheless, we still observed a clear trend in the day-40 cumulative case count as a function of peripheral testing under each values of μ_* and σ_*^2 (see Figures S1c, S1d).

C Peripheral Steady State

In our epidemiological model, we made the critical assumption that the periphery exhibits an epidemiological steady state. The rationale behind this assumption is that over the course of the COVID-19 pandemic, overall trends in cases (at the county, state, and national levels alike) have fluctuated based on a number of factors outside the control of individual institutions. As such, we chose not to model any change in the extra-institutional prevalence, but instead assumed it to be constant and then assess sensitivity to this factor (see Figure 3a). Here we provide additional analysis of how long after the implementation of testing in the periphery it takes in order to achieve such a steady state. Numerically approximating $E_*(t)$ and $V_*(t)$ based on equations (2-3), we obtained the curves shown in Figure S2, which provide estimates for the probabilities of being exposed and infectious in the periphery as a function of the duration of testing. We set all other relevant parameters to their baseline values (see Appendix D).

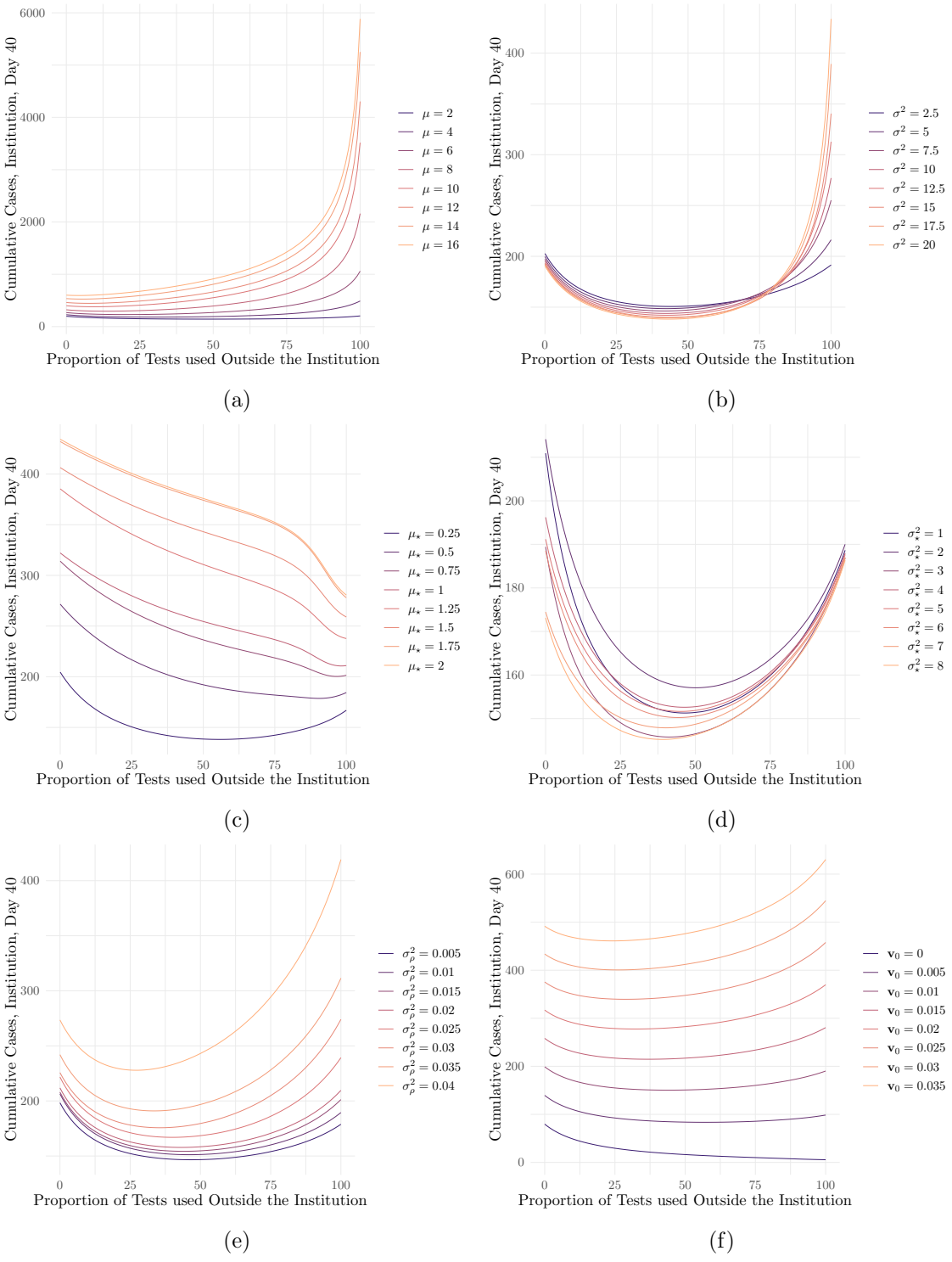
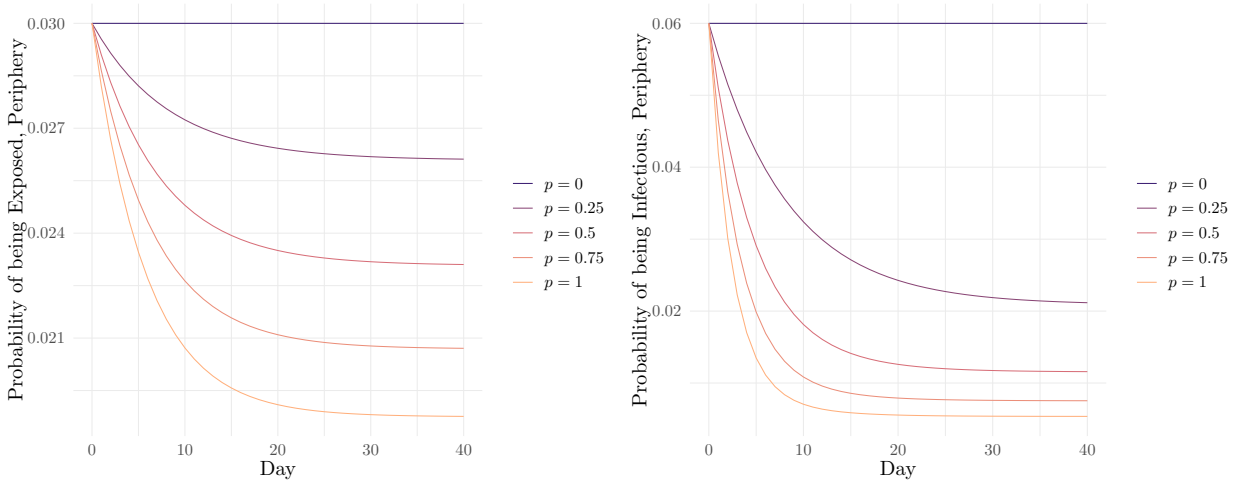


Figure S1: Cumulative cases on day 40 as a function of the proportion of tests deployed to the periphery under different values of (a) the mean number of institutional contacts, μ; (b) the variance in institutional contacts, σ^2; (c) the mean number of peripheral contacts, μ_*; (b) the variance in peripheral contacts, σ^2_*; (e) the variance in the secondary attack rate, σ^2_ρ; and (f) the initial prevalence at the institution, $\mathbf{v}_0$ (assumed to be uniform among agents)



(a) (b)

Figure S2: Modeled probabilities of an agent in the periphery being (a) exposed and (b) infectious over time starting with the implementation of community testing at time 0, under 5 different proportions p of peripheral testing.

D Model Parameters

See below for a table of model parameters, explanations, values, and citations. Parameters with no citation are estimated based on data gathered at CMU. Parameters marked $\star$ are weakly-informed.

Symbol	Description	Value	Citation
N	Number of people in the institution	10,000	
μ	Mean number of contacts per agent at the institution	2.3	
σ^2	Variance in number of contacts per agent at the institution	2.4	
μ_*	Mean number of contacts per agent outside the institution	0.23	
σ_*^2	Variance in number of contacts per agent outside the institution	1.8	
ω	Proportion of contacts made outside the institution known to testing authorities	1	
μ_ρ	Mean secondary attack rate	0.16	[16]
σ_ρ^2	Variance in secondary attack rate among agents	0.01	$\star$
γ	Recovery rate	0.1	[11]
δ	Exposed-to-infectious transition rate	0.2	[6]
$\mathbf{v}_0$	Initial probability of infection within the institution, for each agent	0.01 (uniform)	
V_{0*}	Initial probability of infection outside the institution	0.06	
c	Number of tests per day per person	0.12	
ψ_E	Test sensitivity during the exposed stage	0.02	[10]
ψ_V	Test sensitivity during the infectious stage	1	[10]