

Patterns

Analyzing the impact of a real-life outbreak simulator on pandemic mitigation: An epidemiological modeling study

Highlights

- Outbreak simulation technology can help society mitigate and preempt viral outbreaks
- The technology provides social network statistics that power epidemiological models
- Those statistics can make interventions more efficient and more effective

Authors

Ivan Specht, Kian Sani, Bryn C. Loftness, ..., Todd Brown, Pardis C. Sabeti, Andrés Colubri

Correspondence

ispecht@broadinstitute.org (I.S.),
pardis@broadinstitute.org (P.C.S.),
andres@broadinstitute.org (A.C.)

In brief

Specht et al. demonstrate the effectiveness of using outbreak simulation technology to mitigate and preempt viral outbreaks. They first report the distribution, timing, duration, and connectedness of student social contacts based on outbreak simulations at two universities. Using these contact networks, they then construct epidemiological models to predict the transmission pathways of different pathogens. Finally, they show that the granularity of outbreak simulation data enables institutions to improve outbreak mitigation by proactively and strategically testing/vaccinating individuals based on their social interaction levels.



Article

Analyzing the impact of a real-life outbreak simulator on pandemic mitigation: An epidemiological modeling study

Ivan Specht,^{1,2,11,*} Kian Sani,^{1,3} Bryn C. Loftness,^{1,4} Curtis Hoffman,^{1,5} Gabrielle Gionet,¹ Amy Bronson,⁴ John Marshall,⁴ Craig Decker,⁵ Landen Bailey,⁵ Tomi Siyanbade,^{1,2} Molly Kembell,^{1,3} Brett E. Pickett,⁵ William P. Hanage,^{1,6} Todd Brown,¹ Pardis C. Sabeti,^{1,3,7,8,9,*} and Andrés Colubri^{1,10,*}

¹The Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

²Harvard College, Faculty of Arts and Sciences, Harvard University, Cambridge, MA 02138, USA

³FAS Center for Systems Biology, Department of Organismic and Evolutionary Biology, Faculty of Arts and Sciences, Harvard University, Cambridge, MA 02138, USA

⁴Colorado Mesa University, Grand Junction, CO 81501, USA

⁵Department of Microbiology and Molecular Biology, College of Life Sciences, Brigham Young University, Provo, UT 84606, USA

⁶Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA

⁷Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA

⁸Massachusetts Consortium on Pathogen Readiness, Harvard Medical School, Harvard University, Boston, MA 02115, USA

⁹Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA 02115, USA

¹⁰Program in Bioinformatics and Integrative Biology, University of Massachusetts Chan Medical School, Worcester, MA 01655, USA

¹¹Lead contact

*Correspondence: ispecht@broadinstitute.org (I.S.), pardis@broadinstitute.org (P.C.S.), andres@broadinstitute.org (A.C.)

<https://doi.org/10.1016/j.patter.2022.100572>

THE BIGGER PICTURE Outbreak simulation technology can greatly enhance individual and community pandemic preparedness while helping us understand and mitigate outbreak spread. Building on an existing platform called Operation Outbreak (OO), an app-based program that spreads a virtual pathogen via Bluetooth among participants' smartphones, we demonstrate the power of this approach. We investigate the first- and second-degree contacts of OO participants, analyzing the differential risk associated with various local contact network structures. We use OO data to construct an epidemiological model with which communities may predict the spread of infectious agents and assess the effectiveness of mitigation measures. Based on our findings, we advocate for wider adoption of outbreak simulation technology to study the implications of social mixing patterns on outbreaks in close-knit communities to aid pandemic preparedness and response.



Proof-of-Concept: Data science output has been formulated, implemented, and tested for one domain/problem

SUMMARY

An app-based educational outbreak simulator, Operation Outbreak (OO), seeks to engage and educate participants to better respond to outbreaks. Here, we examine the utility of OO for understanding epidemiological dynamics. The OO app enables experience-based learning about outbreaks, spreading a virtual pathogen via Bluetooth among participating smartphones. Deployed at many colleges and in other settings, OO collects anonymized spatiotemporal data, including the time and duration of the contacts among participants of the simulation. We report the distribution, timing, duration, and connectedness of student social contacts at two university deployments and uncover cryptic transmission pathways through individuals' second-degree contacts. We then construct epidemiological models based on the OO-generated contact networks to predict the transmission pathways of hypothetical pathogens with varying reproductive numbers. Finally, we demonstrate that the granularity of OO data enables institutions to mitigate outbreaks by proactively and strategically testing and/or vaccinating individuals based on individual social interaction levels.



INTRODUCTION

Infectious disease outbreaks have repeatedly emphasized the potential for detailed contact tracing data to improve public health.^{1–3} The coronavirus disease 2019 (COVID-19) pandemic in particular saw the rapid development and deployment of contact tracing technologies in an effort to curb the spread of the virus, accompanied by advances in network science to facilitate use of graphical contact data as a means of pandemic mitigation. Despite the theoretical benefits of such technologies, adoption rates were often low, stemming from numerous factors, including a lack of enforceability and privacy concerns. Without a critical mass of users, these technologies failed to capture the majority of transmission links, compromising their effectiveness.^{4–6} Many contact tracing platforms, such as those built on the Google-Apple Exposure Notification (GAEN) application programming interface (API), generally operated on the principle that contact network data would never be shared unless a user were to test positive.⁷ Although such a policy benefits the user from a privacy standpoint, it neglects the possible benefit of knowing the user's typical social patterns and, when needed, intervening accordingly. Finally, the pandemic consistently pointed to young adults in educational settings (e.g., college campuses) as being disproportionately likely to spread COVID-19.^{7–9} However, young adults generally expressed a particular lack of willingness to adopt digital contact tracing technologies.^{10,11}

To facilitate engagement of children and young adults in public health, we built an experiential education platform called Operation Outbreak (OO) that enables scenario planning for infectious disease outbreaks.¹² OO consists of a suite of tools for learners that includes a smartphone app, a textbook, and a multi-disciplinary curriculum. The smartphone app simulates the spread of a pathogen through a population by transmitting a “virtual pathogen” between participating phones within a threshold proximity. The app also collects anonymous data on the time, duration, and distance of all close contacts between users, as typical contact tracing apps do. OO then processes these simulated transmissions into summary statistics useful for students, teachers, and administrators alike, including levels of social interaction and risk of exposure broken down by participant as well as for the group at large. These statistics also feature as part of the OO curriculum, allowing participants to engage directly with epidemiology through experiential learning.

Use of mobile technology to collect proximity data for epidemiology modeling has been explored for some time.¹³ In 2010, the FluPhone project used Bluetooth connectivity in early smartphones to quantitatively measure societal mixing patterns, conducting virtual epidemics that inform models characterizing the spread of disease in the social network between participants.¹⁴ In 2018, as part of a documentary marking the centenary of the 1918 influenza pandemic, the British Broadcasting Corporation (BBC) released a separate mobile app in which United Kingdom citizens could contribute their movement and contact data for a day.¹⁵ These data were used to construct geographical models of population connectivity that were applied last year to evaluate the impact of COVID-19 control strategies.¹⁶ More recently, researchers transmitted virtual “viral strains” via Bluetooth in a college campus in New Zealand with the goal of making real-time forecasts of COVID-19 spread.¹⁷

OO differs from prior projects in several ways. The OO app has many more features to heighten the realism of the outbreak experience. Participants can visualize their level of illness and unlock quick response (QR) codes to receive a mask, diagnostic, or vaccine, and beacons can be used to represent fomites. The app is actively developed for iOS and Android to ensure that all smartphone users are equally supported and able to participate. It is also part of a larger platform that provides not only the experiential learning simulation but also curricular and professional development materials that contextualize the simulation in the broader context of outbreak science studies. The anonymized data collected by the app may be used for learning activities and epidemiological modeling, with learning as an incentive to generate data and data-driven models as a cornerstone of the learning process. We have been developing OO and running simulations continuously since 2016, which sets it apart from these more circumscribed experiments. The modular architecture of the OO platform will also allow us to incorporate new proximity sensing technologies, such as ultra-wideband,¹⁸ as they emerge and become available on consumer-level devices and to continually enhance the experience and data collection.

In this paper, we quantify and explore the social interaction patterns observed among 787 participants of two OO simulations conducted at two universities in the United States: Colorado Mesa University (CMU) and Brigham Young University (BYU) during the COVID-19 pandemic. We provide a graphical analysis of the contact networks, focusing in particular on first- and second-degree contacts and the relationship between known and unknown transmission pathways. We analyze the times and settings that pose the greatest risk for viral transmission. Finally, based on the OO data, we construct an epidemiological model to measure the efficacy of mitigation strategies informed by OO; in particular, diagnostic testing and vaccinations.

Methodology

Simulation methodology

The OO app, which gathered all data used in this study, is available to the general public in the Apple App Store and Google Play Store. Upon opening the app, users enter a simulation code provided by an OO administrator to join a simulation. During the simulation period, the OO app uses Bluetooth Low Energy (BLE) communication to record all proximate interactions between OO participants up to a distance of approximately 3 m and at a resolution of 1 s. Some of these interactions result in simulated viral transmission when one party is in the infectious state, with the probability of transmission per unit of time prespecified in the parameters of the simulation. Contact detection over Bluetooth was implemented using a cross-platform software library for iOS and Android called p2pkitt,¹⁹ which combined public Bluetooth APIs provided by each mobile platform with platform-specific technology, such as WiFi-direct, to maximize proximity sensing. Participants may engage in various “interventions” (e.g., receive virtual masks, personal protective equipment, or vaccines) by scanning physical QR codes distributed by the OO administrators throughout the simulation. All events over the course of an OO simulation—contacts, transmissions, use of interventions, recoveries, deaths, and more—are recorded in a backend database that houses the dataset used for this study.

Recruitment

At CMU, we primarily sought to recruit first- and second-year on-campus students, many of whom had high levels of involvement in on-campus activities and policymaking. This presumably led to some positive bias in their levels of interaction. Our main goal at CMU was to empower students with information on their close contacts and encourage them to consider the epidemiological impact of their social behavior. At BYU, we mainly advertised to individuals studying the life sciences with the goal of generating data about student behavior. Unlike at CMU, BYU OO participants received daily summary statistics for the simulation, including the total numbers of new contacts, infections, recoveries, and deaths. We recruited a total of 787 participants between CMU and BYU. At CMU, 327 students signed up to participate, of which 240 remained after filtering the data (Results). The CMU simulation lasted 6 days, from October 29 until November 4, 2020, which included Halloween weekend. At BYU, 460 participants signed up to participate, comprising students and BYU faculty, of which 402 remained after filtering. The BYU simulation lasted 9 days, from February 19 until March 1, 2021. For CMU and BYU, the simulation occurred during a period where pandemic mitigation measures were implemented at both universities, such as social distancing, event size restrictions, and hybrid class cohort splitting. Students were still living on campus and commuting from off-campus residences at both universities. For additional information, see the [supplemental experimental procedures](#).

Student engagement

CMU and BYU students expressed an overall willingness to share some of their personal data to engage in the outbreak simulation experience. We hypothesize that this willingness is largely due to the anonymous network information collected about a virtual virus. This differs from traditional contact tracing technologies that are related to the actual spread of COVID-19. Beyond contributing and analyzing their own data, many students took advantage of the opportunity to learn more about public health. In particular, CMU and BYU student participants exhibited strong interest in learning how a system for tracking close contacts during an outbreak can help mitigate outbreaks and how individual interactions can disproportionately impact campus-wide health. With the goal of incentivizing pandemic-mitigating behaviors, we would expect OO to actively or passively influence student interaction dynamics throughout the duration of the simulation. Across both simulations, however, we observed little change in students' behavior depending on their epidemiological state within the game (i.e., susceptible, infectious, vaccinated, etc.), which likely improved the reliability of the social network data but lessened the similarity to an actual outbreak. Overall, the educational focus of OO made it well positioned to gather data useful for epidemic mitigation without appearing as a threat to students' privacy.

RESULTS

We began by investigating OO contact data to better understand the differential risk among individuals associated with their contact patterns. First, we simply measured the raw number of contacts per OO participant at CMU and BYU, filtering out (1) duplicate contacts (multiple contacts between the same pair of

individuals), (2) contacts shorter than 1 min, and (3) contacts made by persons who did not participate in the entire OO simulation. We chose the threshold of 1 min as a proposed cutoff for what constitutes a social contact. Although contacts of (for example) just over 1 min are unlikely to result in transmission, these shorter contacts will hold far less weight than longer ones in determining an individual's risk of contracting the virus. For the BYU simulation, we only analyzed the first week of data to reduce weekday/weekend bias; for the CMU simulation, we were unable to do so because it lasted only 6 days. Both schools exhibited an overdispersed distribution in the number of contacts per individual, consistent with previous findings (Figure 1, blue distribution).²⁰ The mean number of contacts per person was 9.29 at CMU (SD = 11.48, range = 0–58) and 11.13 at BYU (SD = 14.32, range = 0–82). See Table S1 for the graphical properties of the two networks.

We then looked more closely at the network properties of the contacts. The clustering coefficients—the overall probability that any two contacts of a given person themselves had a contact (experimental procedures)—were equal to 0.280 at CMU and 0.243 at BYU. This result is consistent with the findings of Mayer et al. (2008)²¹ on undergraduate student social network dynamics, which reported a range of 0.17–0.27 for clustering coefficients across 10 American universities based on Facebook data. To characterize the likely physical environments for these contacts, we also analyzed the time of day/week when these contacts were most likely to occur, observing spikes during class time at BYU and evenings at CMU. We appreciate that CMU may have exhibited higher-than-normal and otherwise uncharacteristic interaction levels because of social gatherings on the night of Friday, October 30, one night prior to Halloween (Figure 2).

We hypothesized that the raw number of first-degree contacts served as a reasonable proxy for risk of infection but could be improved by taking into account (1) durations of contacts and (2) second-degree contacts. Applying the same filtering processes for contacts as described above, we observed high variance in numbers of second-degree contacts for CMU and BYU participants (Figure 1, red distribution). The mean number of second-degree contacts per person was 60.73 at CMU (SD = 51.61, range = 0–151) and 100.76 at BYU (SD = 84.07, range = 0–264). This analysis gives us a sense of the distribution in the number of second-degree contacts but not the relationship between first- and second-degree contacts, which clearly have a strong correlation. Therefore, we fitted the functional $y = \alpha(1 - e^{-\beta x})$ to the number of second-degree contacts as a function of first-degree contacts using least squares. This functional form is a natural choice in that it passes through the origin with some positive slope (few first-degree contacts imply few second-degree contacts) and eventually plateaus (the number of second-degree contacts is bounded by the population size). Despite a relatively low root-mean-square error (RMSE) of 13.36 at CMU and 19.27 at BYU, indicating that the number second-degree contacts can be accurately predicted from the number of first-degree contacts (Figure 3), there were still some individuals whose second-degree contact counts were significantly higher or lower than the model would predict. Figure 4 presents illustrative examples of an individual who had 7 first-degree contacts but only 32 second-degree contacts

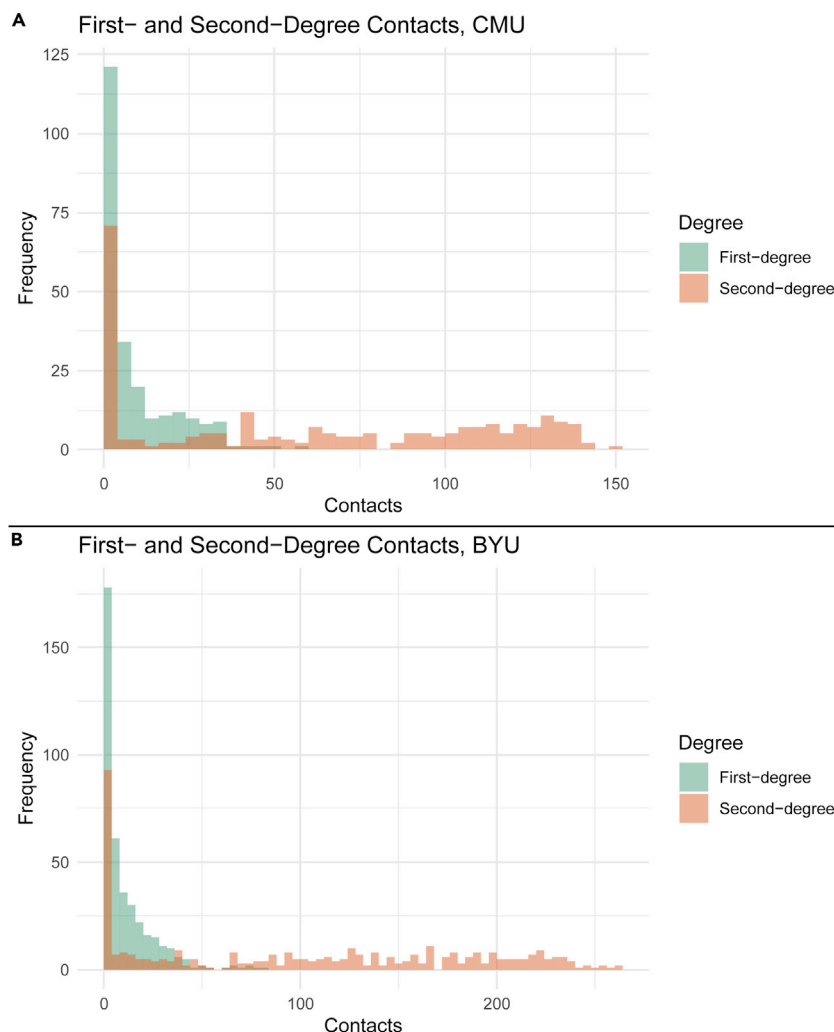


Figure 1. Histograms of contacts per student during CMU and BYU simulations

Histograms of contacts per student at CMU (A) and BYU (B) over the course of 1 week.

We found that second-degree contacts had a statistically significant impact on probability of infection, even beyond what could be captured by first-degree contacts alone. Taking into consideration our previous finding about contact duration, we constructed an additional predictor variable—duration-weighted second-degree contacts—by multiplying the total durations of the two contacts involved. For example, if persons A and B interact for a total duration of 60 s, and persons B and C interact for a total of 80 s, then the second-degree contact between persons A and C via person B contributes a factor of 4,800 to person A's duration-weighted second-degree contacts. To compute the total value of this statistic for person A, we simply sum over all possible second-degree contacts for person A, including second-degree contacts that are also first-degree contacts. Using duration-weighted second-degree contacts as an additional predictor variable in our regression analysis, we found high statistical significance as well as a slight increase in adjusted R^2 compared with duration-weighted first-degree contacts alone (Figure 5).

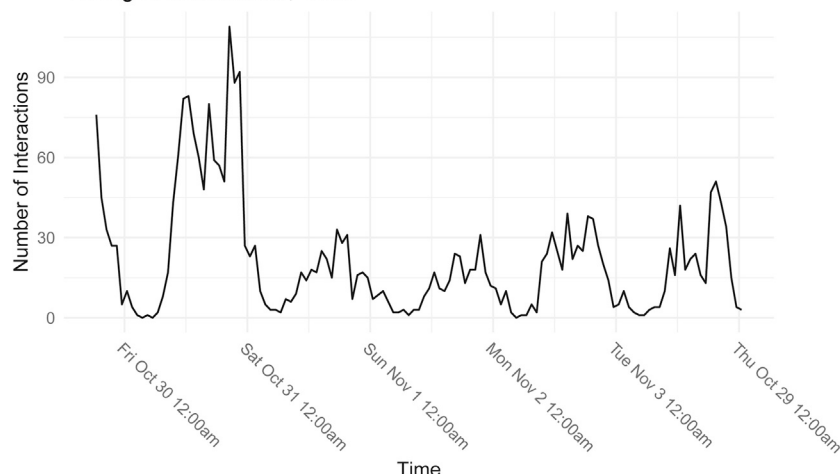
We first ran the epidemiological model varying only the basic reproductive number (R_0) to reflect differences in infectivity associated with different variants of COVID-19. Using kernel density estimation from the Monte-Carlo simulation, we determined the distribution of the cluster size after a 4-week period assuming one initial case. Under all simulations, this distribution was positively skewed—increasingly so with higher values of R_0 . See Figure 6 for a summary of the results.

We then measured the impact of diagnostic testing and vaccinations, which could be implemented according to a random strategy (i.e., equal probability of testing/vaccination for everyone) or an OO-based strategy (i.e., probability of testing/vaccination proportional to social activity level). Here, social activity level is simply defined as the number of first-degree contacts. Under four different levels of testing and vaccination, the OO-based strategy drastically reduced the reproductive number and case counts, with a smaller number of tests/proportion vaccinated corresponding to a more dramatic reduction (Figures 7 and 8). The large credible intervals in Figure 8 are largely due to the fact that the size of the outbreak correlates strongly with the number of transmissions made by the index case, and the probability of an index case making zero onward transmissions is non-negligible.

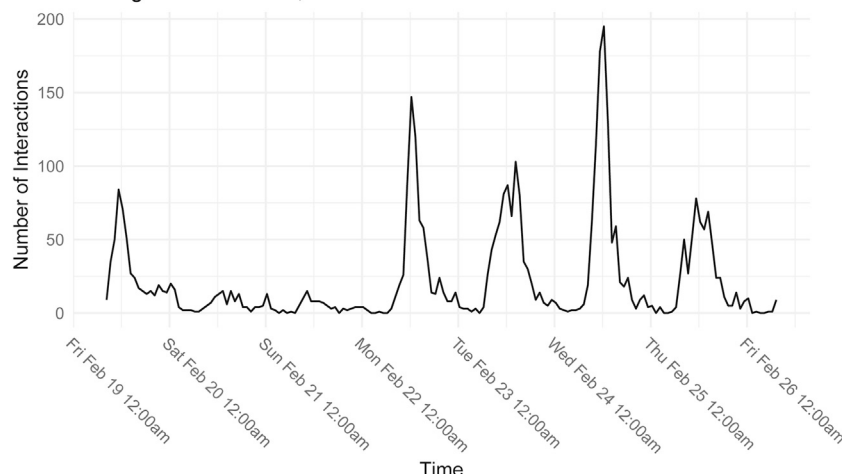
and another with only 3 first-degree contacts but 126 second-degree contacts.

We hypothesized that the relationship between first- and second-degree contacts, as well as the durations of such interactions, would leave certain individuals more or less prone to infection than their first-degree contacts alone would suggest. To test this hypothesis, we first simulated the spread of COVID-19 through the real OO contact networks using mean-field approximation, a computationally efficient method for estimating the probabilities of each person being in each epidemiological state (susceptible, exposed, infectious, recovered) at a given time.^{22,23} We then regressed the probability that each individual had been infected against various statistics describing social contacts. We began with two extremely simple statistics: equal-weighted and duration-weighted numbers of contacts. “Equal-weighted” means the number of contacts for an individual; “duration-weighted” means the sum of durations of all contacts for an individual. Assuming no intercept term, the regressions yielded an adjusted coefficient of determination (R^2) of 0.566 and 0.929, respectively, for CMU and 0.430 and 0.886, respectively, for BYU (Figure 5). These results emphasized the impact of including contact duration in risk assessment.

A Timing of Interactions, CMU



B Timing of Interactions, BYU



DISCUSSION

The OO data we gathered offer a number of substantive conclusions about how close-knit communities such as schools and universities should factor social interaction patterns into their pandemic response. Beyond providing a distribution of the volume, duration, and timing of social contacts, a deeper look at the OO contact network structure reveals the added risk of cryptic transmission pathways; that is, pathways largely unbeknownst to the infectee as a result of the variance in the distribution of second-degree contacts. As revealed by our regression analysis, these second-degree contacts significantly impact individual-level risk and, therefore, may help public health authorities best identify individuals who are most liable to contracting or transmitting the virus.

We then propose a framework by which OO-participating institutions may construct an epidemiological model based on OO network data. These models rely on statistical inference techniques that allow them to be constructed even when only a fraction of institution members participate in OO. Based on these models, institutions may view how various pathogens

Figure 2. Number of interactions recorded during each hour of simulation at CMU and BYU

(A and B) These data reflect the 240 participants at CMU (A) and 402 participants at BYU (B) for whom we have complete contact information. The data start on Thursday, October 29, 2020 at 6:30 p.m. Mountain Daylight Time for CMU and on Friday, February 19, 2021 at 8:18 a.m. Mountain Standard Time for BYU. Times at CMU do not account for daylight savings time, which ended on November 1, 2020.

with different epidemiological parameters will likely propagate through the population.

We demonstrate the potential benefit of using OO social activity data as a means of strategically testing and/or vaccinating individuals in a population. Although such a strategy hinges on a high OO participation rate relative to the population (which we observed neither at CMU nor at BYU), the theoretical reduction in cumulative cases is drastic, even under relatively low levels of testing and/or vaccination. Any such proactive risk-based measures, however, would have to be implemented thoughtfully to not incentivize riskier behaviors. This is a place where such an educational outbreak simulation can be useful as an opportunity for communities to think through varying behavioral responses and outcomes in a low-stakes setting.

The data generated by the OO app may be used for further epidemiological ana-

lyses of various severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) transmission characteristics, such as the high overdispersion that results in most introductions going extinct. We plan to use the OO networks to investigate the effects of overdispersion on the outbreak dynamics by introducing pathogens with varying levels of this parameter. Our data can also be helpful in distinguishing between virological and behavioral superspreading. In the former, a subset of infections is more infectious per contact, and in the latter, all individuals are highly infectious at some point, but the superspreaders happen to make more contacts. We hope such further research will stress the importance of analyzing individual behaviors in the context of infectious disease outbreaks.

Limitations of the study

OO comes with some limitations in terms of its reliability of modeling individual-level risk and outbreaks more generally. For example, student interaction patterns may change dramatically between the time of the OO simulation and the time of an actual epidemic. In these particular cases, we ran the simulations during the ongoing COVID-19 pandemic, so the recorded OO data may not reflect typical student behavior but may better

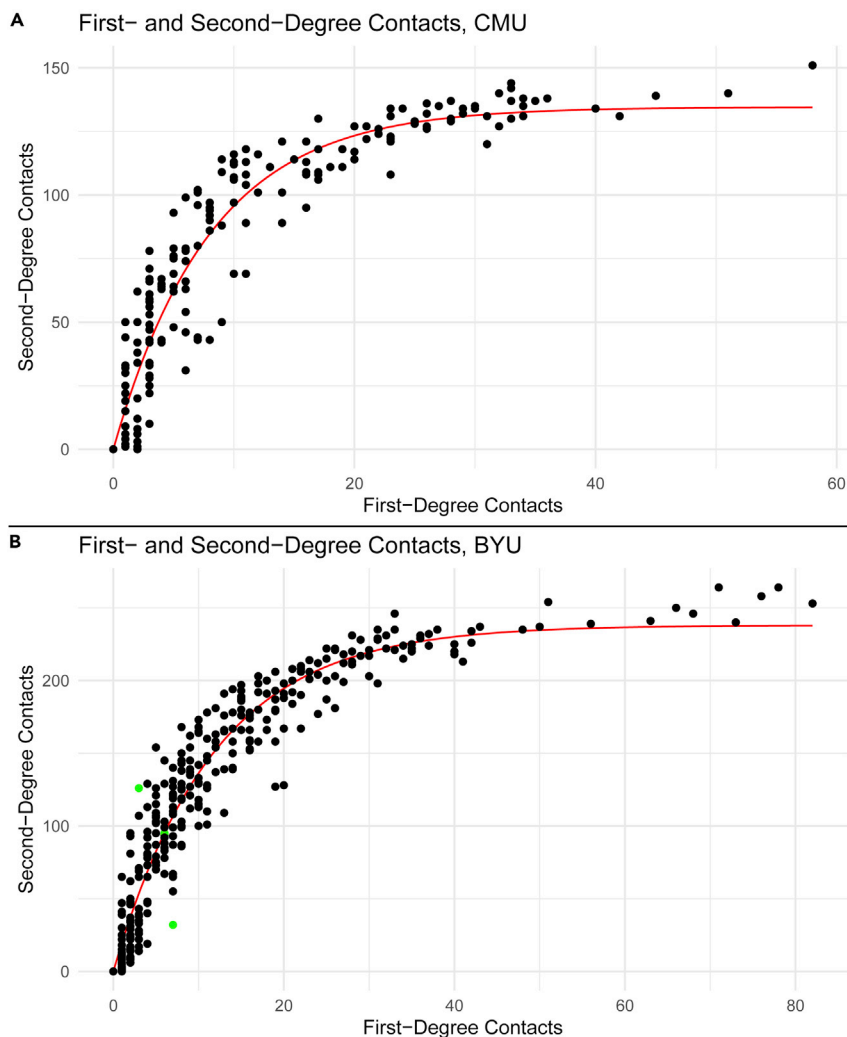


Figure 3. Scatterplot of first-degree contacts and second-degree contacts for each participant at CMU and BYU

(A and B) For each group, we fitted the fitted functional form $Y = \alpha(1 - E^{-\beta X})$. Least-squares estimates: for CMU (A), $\alpha = 238$; $\beta = 0.0850$ for BYU (B). BYU nodes with subgraphs featured in Figure 4 are highlighted in green.

over an extended period of time, as is the case with OO. Earlier experiments like FluPhone, BBC Contagion, and SafeBlues, although supporting the value of this research, also highlight the difficulty associated with developing and maintaining such platforms as technology evolves over time.

Takeaways

We are more connected than we may think. Social contact patterns observed in the university setting revealed significant variation in local contact networks between individuals, leaving some overexposed or underexposed to risk in ways the individual may not recognize. Knowledge of individual-level risk can have a drastic impact on the ability of an institution to mitigate an epidemic. To prepare for the next pandemic, it is essential that we gather social contact data in times of health to prepare for times of sickness. A platform such as OO that integrates pandemic education with preparation and mitigation can engage at-risk populations, such as students, and incentivize

reflect behavior in times of public health crises. The degree to which participants actively engaged with their OO health statuses (i.e., quarantining when infected with the virtual pathogen) differs from an actual pandemic, in which there are real consequences associated with contracting the virus. We have previously conducted research on OO data collected before the pandemic.¹²

Our statistical analyses are also limited by the network data being completely anonymized. Therefore, no additional meta-data are available to provide an increased understanding of the narratives behind different interaction patterns. Assuming a participation rate of less than 100%, there will always be individuals whose social activity levels cannot be computed, so any strategic testing/vaccination plan cannot be tailored to that missing fraction of the population. From a technological standpoint, although Bluetooth-based proximity sensing is widely available on most smartphones, mobile operating systems often pose restrictions for the use of such capabilities. The recent availability of open-source Bluetooth libraries such as Herald, which contains the basis for the contact tracing app TraceTogether, offers an ongoing solution to research on proximity-sensing technologies that aims to be conducted

them to comply with public health interventions by allowing them to be active and informed participants in pandemic response.

EXPERIMENTAL PROCEDURES

Resource availability

Lead contact

Further information and requests for data and code should be directed to and will be fulfilled by the lead contact, Ivan Specht (ispecht@broadinstitute.org).

Materials availability

The OO smartphone application is publicly available in the Apple App Store and the Google Play Store.

Data and code availability

All datasets generated by the OO backend, as well as all original code, have been deposited to Zenodo Data: <https://doi.org/10.5281/zenodo.6584459> and are publicly available as of the date of publication. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Epidemiological model

Leveraging the anonymous contact networks generated by OO, we propose a method by which such data may be used to construct an epidemiological model that simulates the spread of pathogens and measures the impact of mitigation measures. Although the model constructed here does not

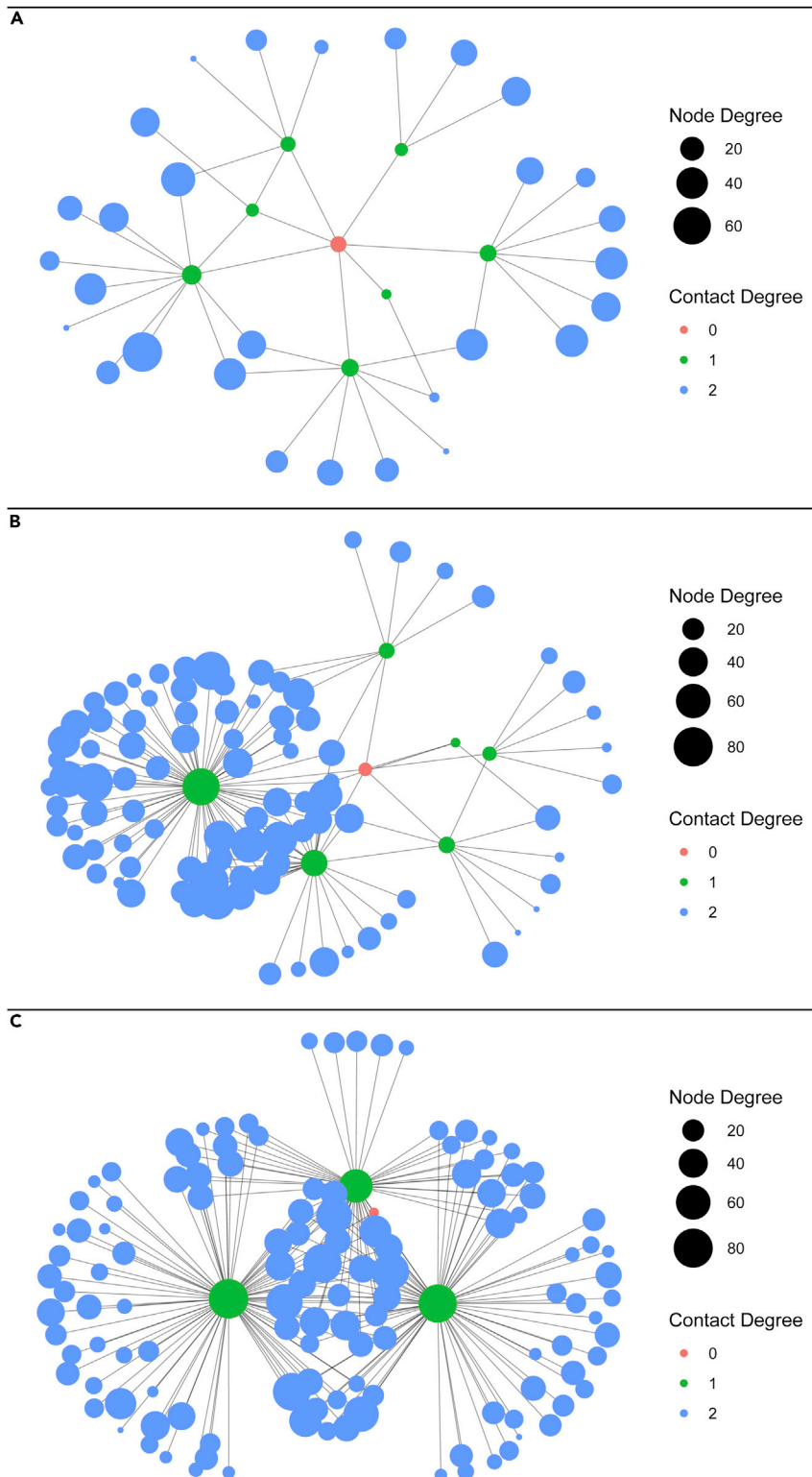


Figure 4. Representative subgraphs of the BYU contact network

(A–C) Across each of these three subgraphs, the number of secondary contacts (blue) is (A) lower, (B) equal, and (C) higher than the model would predict based on the number of first-degree contacts (green). In (A), the red node has 7 first-degree contacts but only 32 second-degree contacts. In (B), the number of second-degree contacts aligns with what we would predict based on the number of first-degree contacts. In (C), the red node has only 3 first-degree contacts but 126 second-degree contacts. Edges between second-degree contacts are omitted for visual clarity.

Predictor(s)	Adjusted R-Squared	t Statistic(s)	Significance Code(s)
Equal-Weighted Contacts	0.566	17.703	***
Duration-Weighted Contacts	0.929	55.967	***
Duration-Weighted 1st- and 2nd-Degree Contacts	0.943	16.987, 7.848	***, ***
Predictor(s)	Adjusted R-Squared	t Statistic(s)	Significance Code(s)
Equal-Weighted Contacts	0.43	17.452	***
Duration-Weighted Contacts	0.886	55.972	***
Duration-Weighted 1st- and 2nd-Degree Contacts	0.896	20.825, 6.124	***, ***

Figure 5. Regression analyses for CMU and BYU

(A and B) We modeled the probability of infection at the end of the OO simulation as a linear combination of various factors, including equal-weighted contacts, time-weighted contacts, and time-weighted second-degree contacts.

necessarily reflect any individual institution, we show that the critical assumptions made reflect observations at CMU and BYU, and, therefore, the methodology may be applied to either university and likely many others.

Construction of an epidemiological model based on OO data is based on two key inference steps: network-based inference and time-based inference. By network-based inference, we seek to propose a reasonable model for how members of an entire institution interact with one another, given that the data gathered by OO only represent the interaction patterns of OO participants, a mere fraction of the institutional population. The simulations at CMU and BYU lasted 6 days and 9 days, respectively; epidemiological models for infectious diseases typically require longer time periods to derive meaningful results. This is what we call time-based inference; i.e., deriving a model for how people interact over longer periods of time, given only or 9 days' worth of data.

For the network-based inference step, we assumed that the true number of contacts, C , made by an individual who participated in OO over the simulated period follows a negative binomial distribution.²⁰ We further assumed that, given C , the proportion of contacts who also participated in OO follows a binomial distribution with size parameter C and probability parameter p . We then solved for the distribution of C via maximum likelihood estimation (MLE), given the observed number of contacts per OO participant.

The above framework allows us to generate node degrees for the university contact network but does not provide a characterization of the connectivity between nodes. Based on the CMU and BYU OO simulations, we found strong

evidence of proportionate mixing, meaning that the probability of two nodes sharing an edge is proportional to the product of their degrees. To substantiate this claim, we regressed the (binary) existence of an edge between two nodes against the product of their degrees and found a relatively high R^2 at both universities: 0.248 at CMU and 0.204 at BYU. Proportionately mixed contact networks based on the OO node degrees mimicked the OO network remarkably well at both universities. In terms of network properties, we focused in particular on the clustering coefficient, which is the overall probability that any two contacts of a given person themselves had a contact, and the average shortest path length, which is the shortest path between a pair of nodes, averaged over all such pairs. At CMU, the modeled clustering coefficient was 0.238 on average (95% CrI: 0.219–0.257) versus 0.280 in the actual network; the modeled average shortest path length was 2.40 on average (95% CrI: 2.34–2.46) versus 2.61 in the actual network. At BYU, the modeled clustering coefficient was 0.184 on average (95% CrI: 0.172–0.195) versus 0.243 in the actual network; the modeled average shortest path length was 2.50 on average (95% CrI: 2.45–2.55) versus 2.69 in the actual network. Based on this finding and a presumed lack of other available information about non-OO participants, we applied a proportionate mixing assumption to the model of the full student body, allowing us to stochastically generate contact networks by assigning each node an expected number of contacts and setting the probability of an edge accordingly.

For the time-based inference step, we implemented a bootstrap method, assuming for simplicity that interactions between any given pair of people

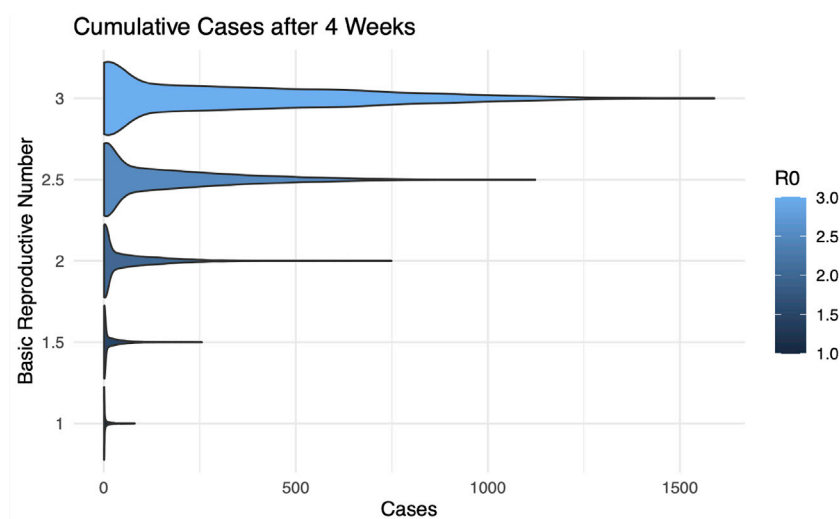


Figure 6. Results of the epidemiological model under five different possible values of R_0

Top: results expressed as a density estimate (top). Bottom: summary statistics from each model run.

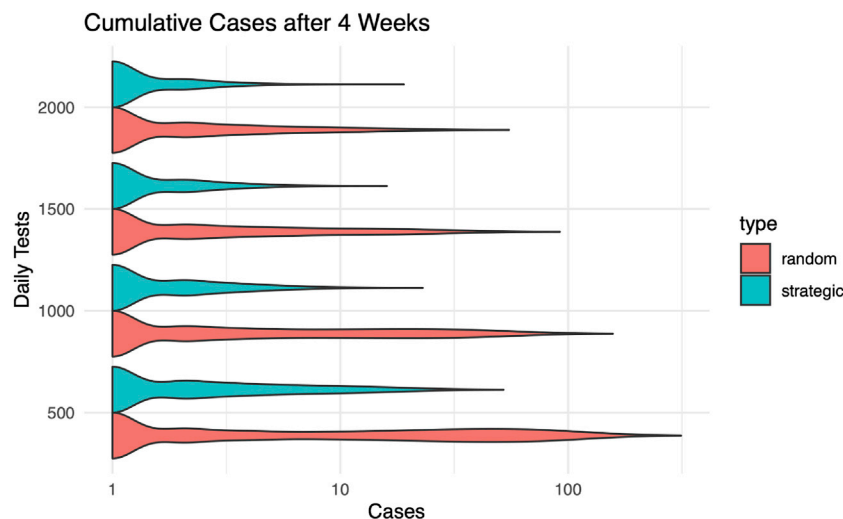


Figure 7. Results of the epidemiological model under four different possible testing rates

These rates range from 500–2,000 per day, administered randomly or based on activity level. Top: results expressed as a density estimate. Center: summary statistics from each model run under random testing. Bottom: summary statistics from each model run under strategic testing.

Daily Tests	Reproductive Number	Mean Cases, 4 Weeks	Median Cases, 4 Weeks	Credible Interval
500	1.5021	22.3094	4	1–127
1000	1.1954	9.4514	2	1–59
1500	0.9088	4.8157	2	1–30
2000	0.7312	3.0745	1	1–17

Daily Tests	Reproductive Number	Mean Cases, 4 Weeks	Median Cases, 4 Weeks	Credible Interval
500	0.8051	3.6758	2	1–19
1000	0.4898	1.8582	1	1–7
1500	0.3558	1.524	1	1–5
2000	0.2831	1.3801	1	1–4

are cyclical with a period of 1 week. The model in this paper uses 7 days' worth of BYU contact data as the bootstrap sampling set; for CMU and other simulations lasting less than 1 week, weeklong bootstrap samples could be generated by amalgamating 1-day bootstrap samples, separating by weekday and weekend. We assumed independence between the total duration of interactions between a pair of nodes and the degrees of those nodes, which was justified by the relatively low observed correlation between these factors of 0.067 (CMU) and 0.051 (BYU).

Under each randomly generated contact network and bootstrap sample of interaction times, we simulated the spread of a pathogen *in silico* on a network with 6,000 agents sampled from the BYU data. We assumed a single index case, sampled based on node degree, who entered the infectious stage at time 0. Letting f be the density function of the generation interval for the virus and letting I_{ij} be an indicator function of an interaction between infectious individual i and susceptible individual j , we set the probability of transmission from i to j equal to

$$\lambda \int_{v_0}^{\infty} f(t - v_0) I_{ij}(t) dt$$

where λ is a constant chosen to reflect the R_e (effective reproductive number) of the virus, and v_0 is the time when individual i contracts the virus (see Newman²⁴ for computation of λ ; see Hinch et al.²⁵ for a comparable methodology). In the event that a transmission occurred (drawn as a Bernoulli trial with probability of success as given in the above equation), we sampled the time of transmission from the density function given by $f(t - v_0) I_{ij}(t)$ up to a constant of proportionality. Because our primary focus in this paper is cumulative cases, and because $f(t)$ approaches 0 as $t \rightarrow \infty$, we did not take into account the recovery rate and assumed reinfection to be negligibly rare.

Finally, we modeled two possible interventions: testing and vaccination. For each intervention, we experimented with a “random” version (in which interventions were administered randomly) and a “strategic” version (in which

interventions were administered based on the level of social interaction exhibited per person). We assumed that tests had a constant turnaround time and sensitivity and that vaccines had already reached a constant and maximum effectiveness level by the start of the simulated period. We further assumed that individuals who test positive would isolate and therefore have no social interactions after the time of receiving the positive result.

We replicated this stochastic model 10,000 times, each time regenerating the node degrees, connectivity matrix, and bootstrap time series samples. We set the model to put out the total number of cases at the end of a 4-week period. The model was implemented in R v.4.0.4²⁶ with packages *igraph*,²⁷ *lubridate*,²⁸ *Rfast*,²⁹ *mixdist*,³⁰ and *ggplot2*.³¹ For a complete list of model parameters, see Table S2, and for a description of any of the aforementioned epidemiological terms, see Table S3.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.patter.2022.100572>.

ACKNOWLEDGMENTS

We acknowledge Fathom Information Design for providing data visualization support and Fuzz Productions for developing the version of the OO app used in the CMU and BYU simulations. We thank our interns at the Broad Institute for contributions to the OO curriculum, student leaders at CMU and BYU for coordination of the simulations, and all of our participants. This work was made possible by Gordon and Betty Moore Foundation grants 9125 and 9125.01. The BYU institutional review board determined this work to be non-human subjects research (protocol IRB2021-207). The Harvard Longwood Campus institutional review board of Harvard University determined this work to be non-human subjects research (as is typical for non-human subjects

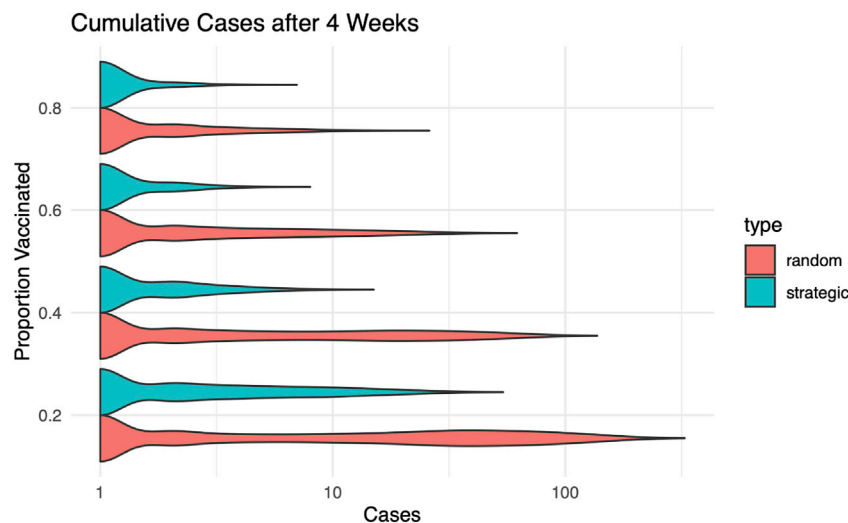


Figure 8. Results of the epidemiological model under four different possible vaccination rates

These rates range from 20%–80%, administered randomly or based on activity level. Top: results expressed as a density estimate. Center: summary statistics from each model run under random vaccination. Bottom: summary statistics from each model run under strategic vaccination.

Vaccination Rate	Reproductive Number	Mean Cases, 4 Weeks	Median Cases, 4 Weeks	Credible Interval
20%	1.6159	25.3323	5	1–138
40%	1.2329	10.0859	2	1–59
60%	0.8676	3.9482	1	1–22
80%	0.4693	1.8613	1	1–7

Vaccination Rate	Reproductive Number	Mean Cases, 4 Weeks	Median Cases, 4 Weeks	Credible Interval
20%	1.0896	4.0917	2	1–20
40%	0.5171	1.6964	1	1–5
60%	0.2348	1.265	1	1–3
80%	0.1173	1.1304	1	1–2

research projects at Harvard, no protocol number was issued). The University of Massachusetts Chan Medical School institutional review board determined this work to be non-human subjects research (protocol H00024280). The CMU institutional review board determined this work to be non-human subjects research (protocol 21–44). The Broad Institute of MIT and Harvard’s Office of Research Subject Protections determined this work to be exempt (protocol EX-6991).

AUTHOR CONTRIBUTIONS

Conceptualization, I.S., K.S., B.C.L., C.H., P.C.S., and A.C.; data curation, I.S., K.S., B.C.L., and A.C.; formal analysis, I.S., K.S., B.C.L., and A.C.; funding acquisition, K.S., T.B., P.C.S., and A.C.; investigation, I.S., K.S., B.C.L., and A.C.; methodology, I.S., K.S., B.C.L., and A.C.; project administration, I.S., K.S., B.C.L., P.C.S., and A.C.; resources, B.C.L., C.H., G.G., A.B., J.M., C.D., L.B., T.S., M.K., B.E.P., T.B., P.C.S., and A.C.; software, I.S. and A.C.; supervision, K.S., B.E.P., W.P.H., P.C.S., and A.C.; visualization, I.S., K.S., B.C.L., P.C.S., and A.C.; writing – original draft, I.S., K.S., B.C.L., C.H., P.C.S., and A.C.; writing – review & editing, I.S., K.S., B.C.L., P.C.S., and A.C.

DECLARATION OF INTERESTS

P.C.S. is a co-founder and shareholder of Sherlock Biosciences and a non-executive board member and shareholder of Danaher Corporation. A.C., T.B., and P.C.S. are inventors on patents related to diagnostics and Bluetooth-based contact tracing tools and technologies filed with the USPTO and other intellectual property bodies.

Received: February 20, 2022

Revised: April 25, 2022

Accepted: July 18, 2022

Published: August 12, 2022

REFERENCES

- Wymant, C., Ferretti, L., Tsallis, D., Charalambides, M., Abeler-Dörner, L., Bonsall, D., Hinch, R., Kendall, M., Milsom, L., Ayres, M., et al. (2021). The epidemiological impact of the NHS COVID-19 app. *Nature* 594, 408–412.
- Sun, K., and Viboud, C. (2020). Impact of contact tracing on SARS-CoV-2 transmission. *Lancet Infect. Dis.* 20, 876–877.
- Willem, L., Abrams, S., Libin, P.J.K., Coletti, P., Kuylen, E., Petrof, O., Mogelmose, S., Wambua, J., Herzog, S.A., Faes, C., et al. (2021). The impact of contact tracing and household bubbles on deconfinement strategies for COVID-19. *Nat. Commun.* 12, 1524.
- Chan, E.Y., and Saqib, N.U. (2021). Privacy concerns can explain unwillingness to download and use contact tracing apps when COVID-19 concerns are high. *Comput. Human Behav.* 119, 106718.
- Kolasa, K., Mazzi, F., Leszczuk-Czubkowska, E., Zrubka, Z., and Péntek, M. (2021). State of the art in adoption of contact tracing apps and recommendations regarding privacy protection and public health: systematic review. *JMIR Mhealth Uhealth* 9, e23250.
- Moreno López, J.A., Arregui García, B., Bentkowski, P., Bioglio, L., Pinotti, F., Boëlle, P.Y., Barrat, A., Colizza, V., and Poletto, C. (2021). Anatomy of digital contact tracing: Role of age, transmission setting, adoption, and case detection. *Sci. Adv.* 7, eabd8750.
- Apple. Privacy-Preserving Contact Tracing. <https://www.apple.com/covid19/contacttracing>.
- Vang, K.E., Krow-Lucal, E.R., James, A.E., Cima, M.J., Kothari, A., Zohoori, N., Porter, A., and Campbell, E.M. (2021). Participation in fraternity and sorority activities and the spread of covid-19 among residential university communities - arkansas, August 21–September 5, 2020. *MMWR Morb. Mortal. Wkly. Rep.* 70, 20–23.

9. Wilson, E., Donovan, C.V., Campbell, M., Chai, T., Pittman, K., Seña, A.C., Pettifor, A., Weber, D.J., Mallick, A., Cope, A., et al. (2020). Multiple COVID-19 clusters on a university campus - North Carolina, August 2020. *MMWR Morb. Mortal. Wkly. Rep.* 69, 1416–1418.
10. Maytin, L., Maytin, J., Agarwal, P., Krenitsky, A., Krenitsky, J., and Epstein, R.S. (2021). Attitudes and perceptions toward COVID-19 digital surveillance: survey of young adults in the United States. *JMIR Form. Res.* 5, e23000.
11. Montagni, I., Roussel, N., Thiébaud, R., and Tzourio, C. (2021). Health care students' knowledge of and attitudes, beliefs, and practices toward the French COVID-19 app: cross-sectional questionnaire study. *J. Med. Internet Res.* 23, e26399.
12. Colubri, A., Kembell, M., Sani, K., Boehm, C., Mutch-Jones, K., Fry, B., Brown, T., and Sabeti, P.C. (2020). Preventing outbreaks through interactive, experiential real-life simulations. *Cell* 182, 1366–1371.
13. Cebrian, M. (2021). The past, present and future of digital contact tracing. *Nat. Electron.* 4, 2–4.
14. Yoneki, E., and Crowcroft, J. (2014). EpiMap: towards quantifying contact networks for understanding epidemiology in developing countries. *Ad Hoc Netw.* 13, 83–93.
15. Klepac, P., Kissler, S., and Gog, J. (2018). Contagion! The BBC four pandemic - the model behind the documentary. *Epidemics* 24, 49–59.
16. Firth, J.A., Hellewell, J., Klepac, P., Kissler, S., CMMID COVID-19 Working Group, Kucharski, A.J., and Spurgin, L.G. (2020). Using a real-world network to model localized COVID-19 control strategies. *Nat. Med.* 26, 1616–1622.
17. Asanjarani, A., Shausan, A., Chew, K., Graham, T., Henderson, S.G., Jansen, H.M., et al. (2022). Emulation of Epidemics via Bluetooth-Based Virtual Safe Virus Spread: Experimental Setup, Software, and Data. *medRxiv*. <https://doi.org/10.1101/2022.03.31.22273262>.
18. Etzlinger, B., Nusbaummuller, B., Peterseil, P., and Hummel, K.A. (2021). Distance Estimation for BLE-Based Contact Tracing – A Measurement Study. 2021 *Wireless Days (WD)* (IEEE). <https://doi.org/10.1109/wd52248.2021.9508280>.
19. Uepaa. p2pkit. <http://p2pkit.io/>.
20. Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Tomba, G.S., Wallinga, J., et al. (2008). Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med.* 5, e74.
21. Mayer, A., and Puller, S.L. (2008). The Old Boy (and Girl) Network: Social Network Formation on University Campuses. *Journal of Public Economics* 92, 329–347. <https://doi.org/10.1016/j.jpubeco.2007.09.001>.
22. Prasse, B., and Van Mieghem, P. (2020). Time-dependent solution of the NIMFA equations around the epidemic threshold. *J. Math. Biol.* 81, 1299–1355.
23. Qu, B., and Wang, H. (2017). The accuracy of mean-field approximation for susceptible-infected-susceptible epidemic spreading with Heterogeneous infection rates. *Stud. Comput. Intell.* 499–510. https://doi.org/10.1007/978-3-319-50901-3_40.
24. Newman, M.E.J. (2002). Spread of epidemic disease on networks. *Phys. Rev. E* 66, 016128.
25. Hinch, R., Probert, W.J., Nurtay, A., Kendall, M., Wymatt, C., Hall, M., et al. (2020). OpenABM-Covid19 - an agent-based model for non-pharmaceutical interventions against COVID-19 including contact tracing. *medRxiv*. <https://doi.org/10.1101/2020.09.16.20195925>.
26. R Core Team (2013). R: A Language and Environment for Statistical Computing.
27. Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, complex systems* 1695, 1–9.
28. Golemund, G., and Wickham, H. (2011). Dates and times made easy with lubridate. *J. Stat. Softw.* 40, 1–25.
29. Papadakis, M., and Tsagris, M. (2018). Rfast: a Collection of Efficient and Extremely Fast R Functions. <https://cran.r-project.org/web/packages/Rfast/index.html>.
30. Macdonald, P., and Du, J. (2012). mixdist: Finite mixture distribution models. <https://cran.r-project.org/web/packages/mixdist/index.html>.
31. Wickham, H., Chang, W., Henry, L., Pederson, T.L., Takahasi, K., Wilke, C., Woo, K., Yutani, H., and Dunnington, D.; R Studio. ggplot2: Elegant Graphics for Data Analysis. <https://ggplot2.tidyverse.org>.