Controlling COVID-19 via test-trace-quarantine

Cliff C. Kerr[1*], Dina Mistry[1†], Robyn M. Stuart[2,3†], Katherine Rosenfeld[1], Gregory R. Hart[1], Rafael C. Núñez[1], Prashanth Selvaraj[1], Jamie A. Cohen[1], Romesh G. Abeysuriya[3], Lauren George[1], Brittany Hagedorn[1], Michał Jastrzębski[4], Meaghan Fagalde[5], Jeffrey Duchin[5], Michael Famulare[1], and Daniel J. Klein[1]

[1] Institute for Disease Modeling, Seattle, WA, USA
[2] Department of Mathematical Sciences, University of Copenhagen, Copenhagen, Denmark
[3] Burnet Institute, Melbourne, VIC, Australia
[4] GitHub, Inc., San Francisco, CA, USA
[5] Public Health - Seattle and King County, Seattle, WA, USA

** Correspondence to: ckerr@idmod.org*
† Contributed equally

COVID-19 containment efforts in the United States so far have largely focused on physical distancing, including school and workplace closures. However, these interventions have come at an enormous societal and economic cost. Here, we use an agent-based model, calibrated to detailed demographic, mobility, and epidemiological data for the Seattle region, to investigate the feasibility of alternative control strategies, focusing on "test-trace-quarantine": a combination of (a) routine testing of primarily symptomatic individuals, (b) tracing and testing their known contacts, and (c) placing their contacts in quarantine. We assess the requirements for implementing this strategy, including its robustness to low compliance, delays, and other factors such as variability in overall transmission rates. We find that for the Seattle setting, if mask compliance remains high and schools remain closed, realistic levels of testing and tracing are sufficient to maintain epidemic control under a return to full workplace and community mobility.

Introduction

Within six months of the world first becoming aware of COVID-19, the total number of confirmed cases exceeded 10 million, with the true number of infections likely much higher. As the pandemic has evolved, so too have global public health responses. Many of the initial efforts to contain the spread of new infections focused on border controls, but when these proved insufficient to prevent community transmission, the focus turned to broad-based lockdowns and physical distancing measures. Whilst effective (*1*), such measures come at enormous cost (*2*). Consequently, governments are increasingly relaxing lockdowns in favor of more targeted "test-and-trace" strategies, whereby only those most likely to have COVID-19 – for example, those who have symptoms, or who have been in contact with a confirmed case – are required to quarantine. Such strategies have the potential to offer the epidemiological benefits of a large-scale lockdown with smaller economic and societal costs.

Several studies have examined test-and-trace-based containment strategies of COVID-19 in different contexts. Modeling studies have provided evidence that the success of such strategies depends on the proportion of symptomatic cases as well as the speed and completeness of contact tracing (*3–5*). Given an estimated R_0 of 2.0–3.5, the number of effective contacts must be reduced by at least 50–70% to achieve epidemic control. Despite this stringent requirement, China successfully demonstrated the feasibility of epidemic control through mandatory home-based quarantine and isolation of those with confirmed infections. This strategy was successfully followed by South Korea, Singapore, and other countries (6). However, success has not been universal, and some countries have had to re-impose restrictions due to epidemic resurgence (7).

Thus, there are global precedents demonstrating the success of containment strategies based on (a) high rates of routine testing, (b) rapid return of test results, (c) high rates of contact tracing, and (d) social support for people who have been diagnosed or been in contact with a diagnosed case, allowing these people to effectively isolate or quarantine – a strategy we refer to as test-trace-quarantine (TTQ). However, the success of this strategy depends on how effectively each component is implemented. To date, the COVID-19 response in the United States (and other Western countries, such as the United Kingdom) has been marked by insufficient quantities of test kits and associated supplies, along with challenges in implementing contact tracing at scale (*8*). It is therefore crucial to ask what the minimum requirements of a TTQ strategy are such that epidemic control can be maintained as lockdowns are relaxed.

Here we investigate what would be required for a high-income, urbanized setting – using the example of Seattle, Washington, USA – to successfully transition from a policy of physical distancing and other non-pharmaceutical interventions towards TTQ-driven containment. In order to capture the nuances of TTQ, we developed a detailed, data-driven, agent-based model of COVID-19 ("Covasim") and applied it to King County, the local government area that includes Seattle and the surrounding metropolitan area. Covasim captures features important to epidemic spread, including age structure and population size; realistic transmission networks in different social layers, including households, schools, workplaces, communities, and long-term care facilities; age-specific disease outcomes; and viral dynamics capturing infectivity variations within and between hosts. Covasim also supports an extensive set of interventions, including symptomatic and asymptomatic testing, isolation, contact tracing, and quarantine, as well as other non-pharmaceutical interventions such as physical distancing, hygiene measures, and protective equipment such as masks. These interventions can incorporate the

effects of delays, loss-to-follow-up, micro-targeting, and other factors. The purpose of this study is to explore the minimum bounds on and relative importance of (a) routine testing, (b) contact tracing, and (c) the efficacy of isolation and quarantine.

Non-pharmaceutical interventions achieved initial epidemic control

The first case of SARS-CoV-2 in the USA was diagnosed on January 20 in the Seattle metropolitan area (*9*); the first documented COVID-19 death in the USA was on February 27, a resident of a long-term care facility (LTCF) also in the Seattle area (*10*). Local and state governments began issuing a series of measures to control the expanding epidemic, including school closures on March 12 and a shelter-in-place order ("Stay Home, Stay Healthy") on March 23 (*11*).

We fit the Covasim model to data on COVID-19 deaths and diagnoses in King County using Optuna, a parameter optimization library (*12*). Detailed demographic information (including population age structure and contact patterns, school enrolment, employment, and LTCF residency), mobility data (provided by SafeGraph; see safegraph.com), and COVID-19 testing data were used as inputs. As shown in Fig. 1, Covasim was able to accurately reproduce the detailed time trends of both diagnoses and deaths (Fig. 1A and B), including the age distribution of each (insets). We estimate that approximately 80,000 SARS-CoV-2 infections (95% confidence interval: 60,000–100,000 infections) occurred in King County between February 1 and June 9 (Fig. 1C), out of a total population of 2.25 million. A total of 8,595 cases had been diagnosed by June 8, for an overall diagnosis rate of 11% (95% CI: 9–14%). The basic reproduction number, R_0, is estimated to have been approximately 2.3 (95% CI: 2.1–2.6) prior to policy interventions, consistent with other estimates for the Seattle region (*13*). The effective reproduction number, R_e, is estimated to have dropped below 1 for the first time during the 10-day period between the announcements of school closures and the shelter-in-place order (Fig. 1D); this period also coincided with the peak number of active infections, nearly 15,000 (Fig. 1C).

Calibrated model parameters, which provide estimates of transmission dynamics and intervention effectiveness, are shown in Fig. 1E. The parameters used for calibration were: overall transmissibility β, defined as the probability of transmission between an infectious and susceptible adult on a single day in a typical household setting; transmission relative to baseline, which may change due to mask usage, hygiene, physical distancing, and other measures; and the odds ratio for people with COVID-19 symptoms being tested versus people without symptoms (i.e., uninfected, asymptomatic, or presymptomatic people). To determine the impact of mobility-related changes in transmission, we calibrated the model both using reductions in the number of work and community contacts based on SafeGraph mobility data (M, blue), and using no mobility data (N, red). Including the mobility changes, we found that relative transmissibility reduced to 78±4% of its initial value in mid March and 67±6% from late April onwards, reflecting the impact of other non-pharmaceutical interventions (NPI), including interpersonal distancing, hygiene, and mask use; this drop is consistent with increasing trends of protective health behaviors (*14*). To verify the calibration, we excluded mobility data and recalibrated the model, finding that relative transmissibility dropped rapidly to 29±3% of its initial value by mid March, and has remained relatively constant since, indicating that observed increases in mobility have been offset by the increase in health behaviors described above. All other parameters had consistent values between the two calibrations, including the change in transmission at LTCFs (estimated to have dropped to 15–20% of its initial value) and overall transmissibility (estimated at

4.2±0.1% per household contact per day). The symptomatic testing odds ratio increased significantly between January and late March, from 7±2 to 22±4, which may be due to a seasonal and NPI-mediated reductions in other respiratory infections, as well as refinements in testing criteria and targeting. Despite these increases in testing odds ratio, data on routine testing yield (the number of diagnoses divided by the number of tests) showed a declined from 10–15% in March to 1.5–2.5% in early June, due to the much lower number of active infections.

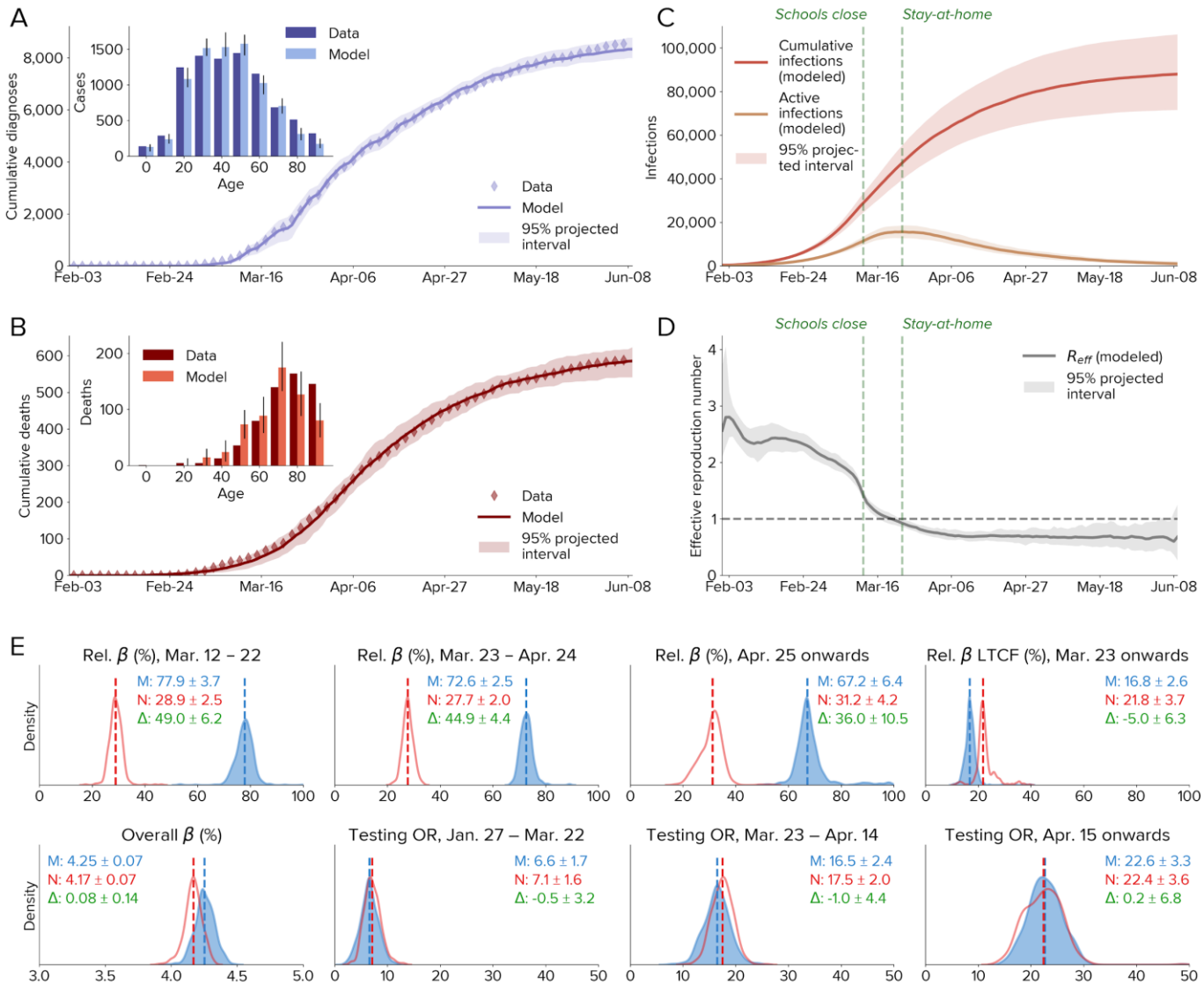

Fig. 1. Calibration of the model to data from Seattle-King County, Washington, from February 1 to June 9. A–B: The cumulative number of diagnosed cases and deaths, over time and by age. **C–D:** Estimated numbers of cumulative infections, active infections, and the effective reproduction number; dashed lines show policy interventions. **E:** Calibration of model parameters with SafeGraph mobility data (M, blue) and with no mobility data (N, red); differences (Δ, green) are only significant for relative transmission (β) parameters. LTCF, long-term care facility; OR, odds ratio.

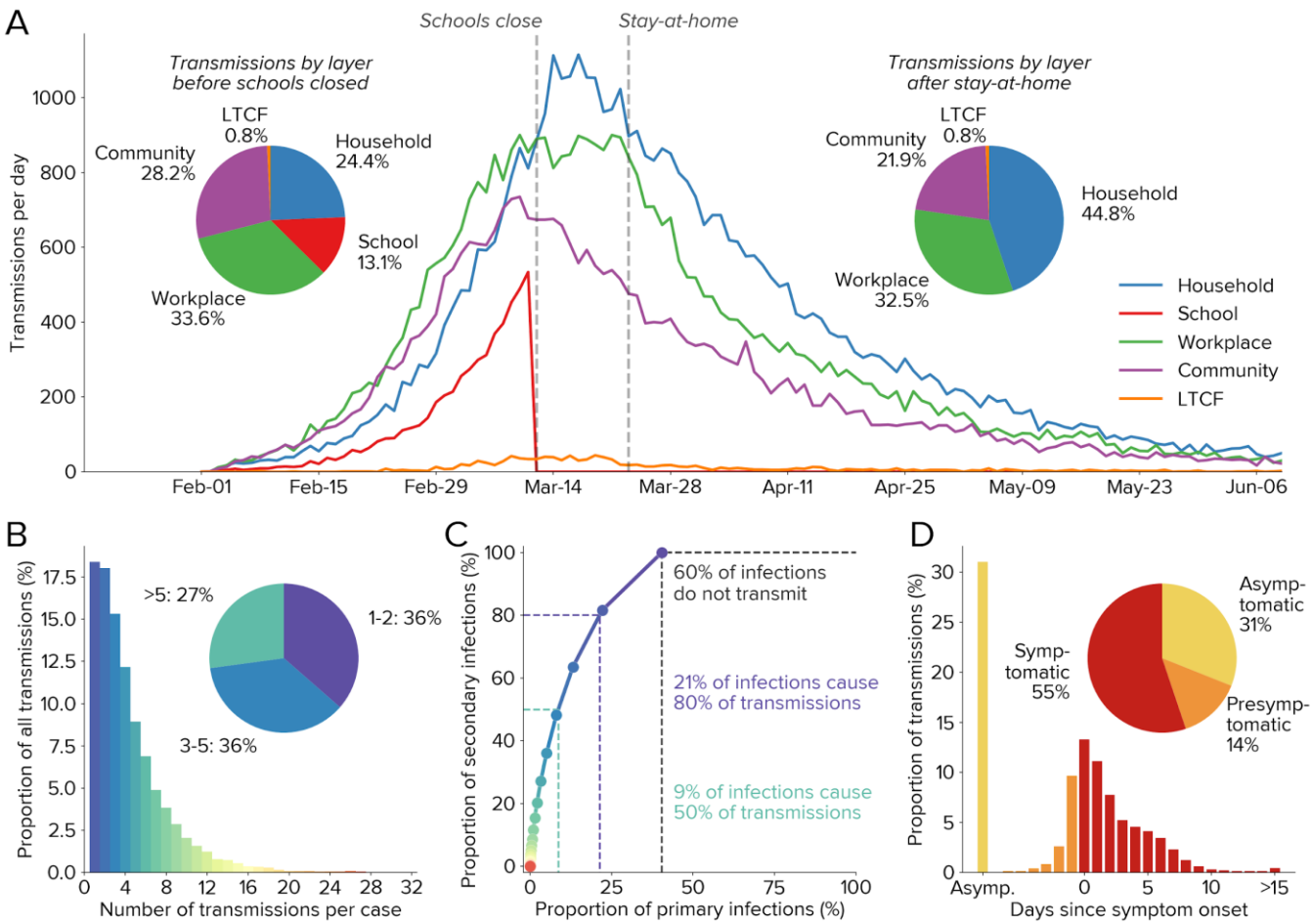

Fig. 2. Modeled transmission dynamics. **A:** Infections over time by contact layer. **B–C:** Overdispersion of infections, with roughly equal numbers of infections attributable to individuals who transmit to 1–2 others, 3–5 others, or more than 5 others, with fewer than 10% of primary infections responsible for more than 50% of transmissions. **D:** Infections as a function of symptom onset, showing that slightly over half of infections are transmitted by symptomatic individuals.

Since Covasim includes detailed models of demographics and intra-host models, it can be used to investigate mechanisms of transmission, as shown in Fig. 2. We find that infections were primarily driven by transmission in workplace and community "contact layers" (accounting for approximately 62% of the total) prior to interventions. Surprisingly, following distancing interventions, the total proportion of infections due to workplace and community transmission only reduced slightly, to 53% (Fig. 2A). This is in part due to the high overdispersion of SARS-CoV-2 infections (Fig. 2B): a majority of people infected do not transmit, while 50% of infections are caused by just 9% of people infected (Fig. 2C). These results hold even though "superspreading" events (>20 infections per index case) appear to be rare and are unlikely to be major epidemic drivers. Nonetheless, a relatively small proportion of highly infectious individuals are likely responsible for a majority of ongoing COVID-19 spread. Preliminary data from contact tracing of roughly 2,000 household contacts in King County provides further evidence for this: of the roughly 40% who received a COVID-19 test, 40% of them tested positive (i.e., roughly 15% of traced household contacts were positive). High-risk index cases and contacts were preferentially both traced and tested, so these estimates represent an upper bound on the attack rate. This value is consistent with international estimates, which have ranged from 5% (*15*) to 19% (*16*). To be consistent with our estimated value of R_e, these relatively low household attack rates

requires high dispersion and significant non-household transmission. We also estimate that roughly 55% of transmissions are from symptomatic individuals, similar to previous estimates (Fig. 2D) (*17*, *18*).

Idealized test-trace-quarantine results in self-limiting epidemic dynamics

Before investigating TTQ in the context of King County, we first consider how TTQ impacts SARS-CoV-2 transmission in a hypothetical population. In an idealized TTQ scenario, where all contacts are traced, all traced contacts enter into quarantine, and delays are less than the duration of infectiousness, epidemic control can be achieved even for high transmission scenarios. This is because as a cluster of infections grows, the probability increases that someone from that cluster will be diagnosed, and when this occurs, idealized contact tracing would identify everyone in the cluster, including both "upstream" and "downstream" infections, removing that cluster from the infectious pool.

Figure 3 shows an illustrative example of idealized TTQ resulting in epidemic control. In a hypothetical population of 100 people without interventions, infections continue until herd immunity prevents further spread (Fig. 3A). If a high level of testing and isolation is introduced (15% probability of testing per day for people with symptoms, coupled with 80% effective isolation), the number of infections is only modestly decreased despite nearly half of cases being diagnosed, since a large proportion of transmission occurs before cases are diagnosed. Adding a moderate level of tracing (70% of household contacts, 10% of workplace and school contacts) significantly reduces the number of infections (Fig. 3C), due to the rapid diagnosis of traced contacts and the preventative effect of quarantine.

Crucially, the effectiveness of contact tracing is proportional to incidence, and thus it results in self-limiting epidemic dynamics. Fig. 3D shows a hypothetical population of 30,000 people in a medium transmission scenario (R_0 = 2.5), where borderline epidemic control ($R_e \approx 1$) can be achieved through either moderate physical distancing alone (i.e., 60% reduction in β), high levels of routine testing and isolation alone (75% daily probability of people with symptoms testing and isolating), or TTQ (8% daily probability of people with symptoms testing, 90% of contacts of diagnosed individuals being traced and quarantined, and 75% probability of testing on entering quarantine). In a low transmission setting (Fig. 3E, R_0 = 2.0), both physical distancing and testing lead to rapid epidemic extinction, while TTQ maintains $R_e \approx 1$. Conversely, in a high transmission setting (Fig. 3F, R_0 = 3.0), TTQ again maintains $R_e \approx 1$, while physical distancing and testing do not achieve epidemic control. This is because distancing and routine testing act like constant multipliers on transmission; they will achieve epidemic control if and only if they bring $R_e < 1$. In contrast, in a TTQ scenario with no program capacity constraints, an increase in the number of infections will result in more diagnoses, more contacts being traced, more people placed in quarantine, and more people tested in quarantine. This mechanism limits the size of individual clusters of infections in a TTQ setting, as well as placing an upper bound on epidemic growth even with extremely high transmission scenarios (e.g., R_0 = 5). However, this phenomenon only occurs with sufficient levels of contact tracing: analogous to R_e, self-limiting dynamics only occur if at least one new case is detected on average for each new index case whose contacts are traced. Whether or not this occurs is determined by the probability of contact tracing, the probability of testing in quarantine, the population network structure, and the rate of transmission.

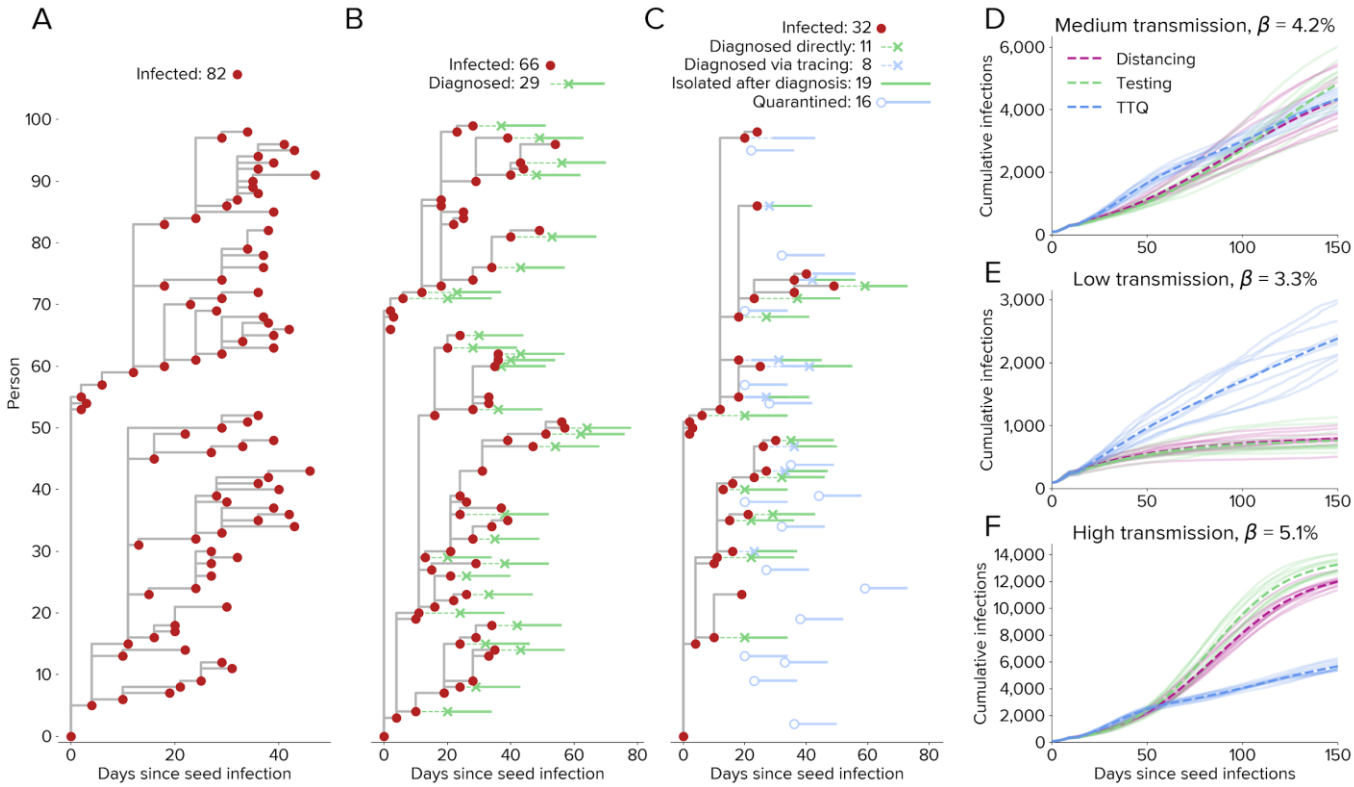

Fig. 3. Epidemic dynamics differ depending on the intervention. **A–C:** Transmission trees for a cluster of 100 people under three scenarios: (**A**) no interventions, (**B**) testing and isolation only (starting on day 20), and (**C**) test-trace-quarantine. **D–F:** Comparison of interventions for different levels of transmissibility. For medium baseline transmission (**D**), moderate distancing, high testing, or high tracing each result in $R_e \approx 1$. For low transmission (**E**), the same distancing and testing interventions both result in $R_e < 1$, while the same tracing intervention maintains $R_e \approx 1$. For high transmission (**F**), the same distancing and testing interventions both result in $R_e > 1$, while the same tracing intervention continues to maintain $R_e \approx 1$.

Realistic test-trace-quarantine scenarios allow high mobility

Implementing a successful TTQ strategy requires solving a challenging prioritization problem: whom to test, whom to trace, how to ensure people safely isolate and quarantine, how to quick return test results, and how to quickly trace contacts. Barriers to high performance in each of these areas are varied, including limited budgets, shortages of equipment and staff, racial and economic inequalities, and behavioral compliance.

Figure 4A shows how six aspects of TTQ in the Seattle context impact the estimated numbers of infections: (1) effectiveness of isolation and quarantine (i.e., relative reduction in transmission), (2) the probability of a person's household and workplace contacts being traced following diagnosis, (3) the number of contacts who are tested following tracing, (4) the number of routine tests per 1000 people per day, (5) the delay between when a person is tested and when they receive their result, and (6) the delay between when a person receives a positive result and when their contacts are notified. Using the model calibrated to the Seattle region up until June 10 and projecting forward for a 12 week period, we consider a hypothetical scenario of high mobility (100%), high testing (~4,000 routine tests per day, or 1.8 per 1,000 people per day, compared to ~2,000 routine tests per day as of June 10), and high tracing (50% of all household and workplace contacts traced within 2 days, compared to roughly 50% tracing of household and close 0% tracing of workplace contacts as of June 10). This scenario corresponds to an R_e

value of just below 1 given a return to 100% mobility (Fig. 4B), and an overall attack rate (cumulative infections divided by population size) over this period of approximately 0.5%.

All six aspects of the TTQ intervention had a significant impact on epidemic outcomes. The most important aspects (in terms of overall attack rate) are isolation/quarantine effectiveness and routine testing probability; low performance on either of these could result in a 10-fold increase in the attack rate. The next most important are contact tracing probability and test delay; low performance on these could lead to a three-fold increase in infections. Contact tracing delay and the probability of testing people in quarantine had the smallest impact, but low performance could still almost double the attack rate. Even with short testing and tracing delays (e.g., 1 day each), a significant proportion of secondary transmission will have occurred before contacts can be notified, due to asymptomatic, presymptomatic, and early symptomatic transmission, combined with delays between symptom onset and testing (which has a median of approximately 3–4 days). However, as long as the process does not exceed the incubation period (4–5 days), contact tracing can still effectively prevent tertiary spread.

While quarantine testing had the smallest overall impact on the attack rate, it identifies cases efficiently. The test yield for routine testing in early June was roughly 1.5%; in comparison, yield from testing of traced contacts was approximately 40% (due in large part to the preferential testing of the highest risk contacts). Asymptomatic testing has limited benefit in routine testing due to its low yield, but asymptomatic testing is an efficient strategy for traced contacts. If the total number of tests is constrained, the optimal policy is to test as many contacts as possible – including those without symptoms – unless the yield drops below that of routine testing. Tracing and testing non-household contacts is one of the most challenging aspects of TTQ due to its dependence on behavioral compliance. However, this aspect is critical since transmission in schools, workplaces, or the community is necessary for transmission chains to be sustained. Thus, when an index case's contacts are traced, there is considerable benefit in testing them in addition to placing them in quarantine.

To explore the practical implications of these results for Seattle, we simulated three reopening scenarios (Fig. 4B): 70%, 85%, and 100% mobility in the workplace and community contact layers, relative to baseline. We assume these mobility changes occur in the context of continued use of masks, physical distancing, and other NPI, which together have reduced transmission to roughly two-thirds of its initial value (see Fig. 1E). Together with school closures, a return to 100% workplace and community mobility in the absence of other interventions would increase R_e to 1.5 (compared to its estimated initial value of 2.3), while the "high mobility, high intervention" scenario described above would reduce it to just below 1. Another scenario that would result in $R_e \approx 1$ for 100% mobility is a tripling of routine testing (from 0.9 to 3.0 tests per 1,000 people per day), with no additional contact tracing. While epidemic control without contact tracing may be achievable via very high levels of testing and isolation, it is unclear whether these levels of testing could be sustained given the low rates of active infections at the start of the scenario. Furthermore, this would require there being a sufficient population of people who have COVID-like symptoms who are not currently testing, which may not be the case if there is a larger than expected proportion of asymptomatic infections, or if increased testing is primarily driven by increased numbers of symptomatic infections.

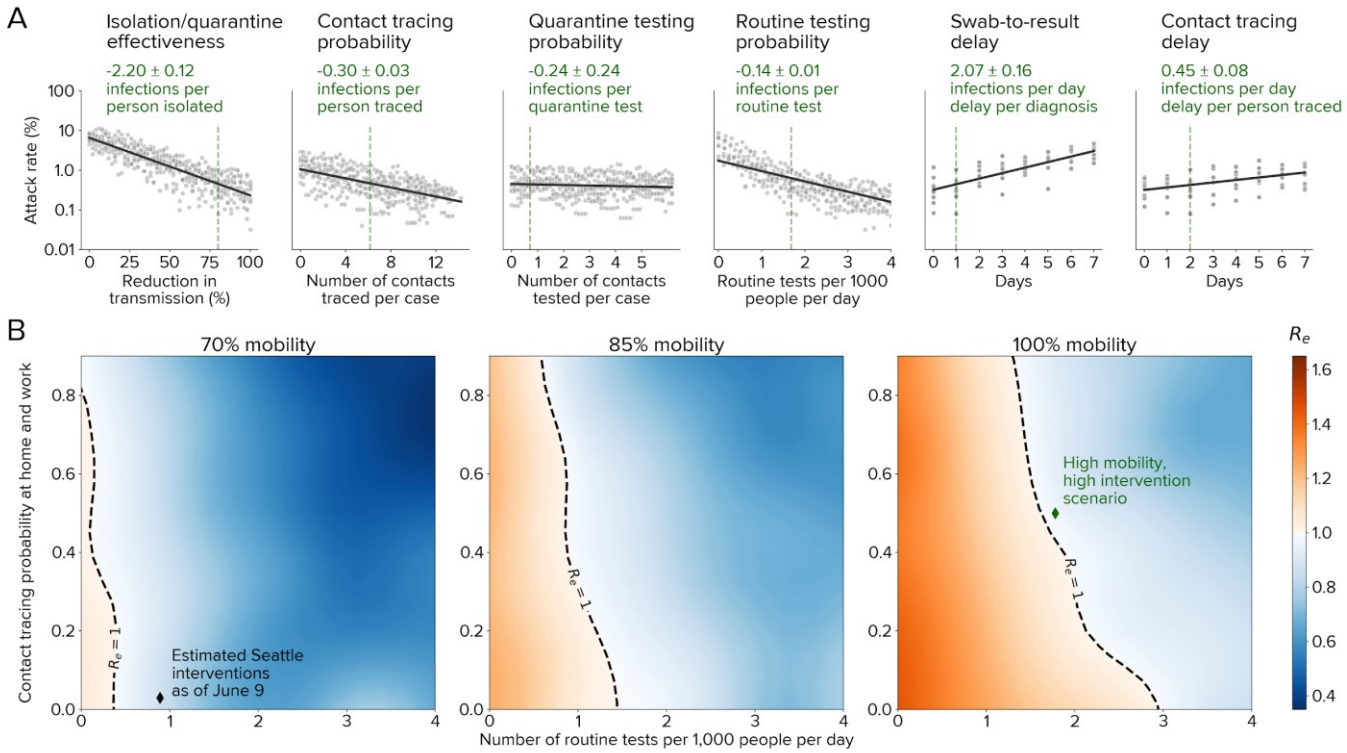

Fig. 4. Impact of testing, tracing, and quarantine. A: Relative importance of different aspects of the TTQ strategy for a scenario of high mobility (full return to baseline workplace and community movement patterns), high testing, and high tracing in Seattle (dashed green lines). Each dot shows a simulation, with other parameters held constant. Isolation/quarantine effectiveness has the greatest impact, with 2.2 infections averted for each person fully isolated, although all parameters have a significant impact on epidemic outcomes. **B:** Countering the effects of increased mobility via testing, tracing, and quarantine. Current interventions (black diamond) are estimated to keep $R_e < 1$ for 70% of baseline mobility levels (left). For a return to full mobility (right), high levels of both testing and tracing are required to maintain epidemic control (green diamond, corresponding to the dashed lines in panel A).

Discussion

Seattle achieved epidemic control between mid March and early June as a result of (a) reduced mobility (the "Stay Home, Stay Healthy" campaign), (b) adoption of other non-pharmaceutical interventions to reduce transmission risk, including interpersonal distancing, hand washing, and face masks, and (c) modest rates of routine testing, plus low but increasing rates of contact tracing. We separated the effects of the first two components using detailed mobility data, and were therefore able to assess the impact of reopening whilst assuming that distancing, hand washing, and face mask use will continue. We find that with mobility at 70% of its pre-COVID levels, epidemic control could be attained despite relatively low testing and tracing. However, returning to higher mobility levels would require identifying and isolating more cases, either through significantly increasing the number of routine tests conducted, or by using contact tracing to target tests toward those more likely to be infected. Our results illustrate the complex interdependencies between different non-pharmaceutical interventions, with the success of each strategy conditional on the interventions already in place.

The handful of countries that achieved control of their COVID-19 epidemics without relying on extensive lockdowns have done so using a diversity of approaches, but those that have used a broader set of interventions appear to have achieved more stable epidemic control. Taiwan and South Korea

used high rates of testing, contact tracing, mask compliance, and other interventions to quickly bring their epidemics under control (*19*). Japan has had high mask compliance and a relatively high rate of contact tracing, but relatively low testing rates; after early control, reported new cases increased from late May through early July (*20*). Australia and New Zealand achieved early epidemic control via strong travel restrictions and high rates of testing and contact tracing; however, mask use has been rare, and a rapid increase in reported cases in Australia occurred in early July. This observational evidence for a diverse range of interventions is consistent with our finding that each aspect of the response is roughly equally important: shortfalls in one area (e.g., low rates of routine testing or mask usage) may be partially compensated for by high performance in another (e.g., high rates of contact tracing). However, these examples suggest that epidemic control may be only fleeting in the presence of continued importation events unless performance is high in all three areas (testing, tracing, and either mask usage or distancing).

There are several limitations of this study. First, we do not consider geographical clustering of contact networks, and therefore cannot model hotspots or outbreaks in particular areas or neighborhoods. Although this may not be crucial to capture given that interventions over the first phases of the response were set at a county-wide level, it is possible that subsequent phases of the response may require more localized policy actions. Second, we do not account for environmental transmission. Third, there is continued debate around how susceptibility and transmissibility vary by age and with comorbidities; our choice of model parameters reflects what we consider to be the best available evidence, but this may change as new evidence comes to light. Finally, there is considerable uncertainty around other crucial characteristics of both SARS-CoV-2 transmission (including the extent to which it is seasonal, and the proportion of asymptomatic and presymptomatic transmission) and the impact of interventions (such as mask efficacy). Such uncertainties are common in mathematical models, especially for novel diseases. We have handled these uncertainties by sampling parameters from appropriate distributions; however, additional data on transmission characteristics would help refine our understanding of the most important transmission pathways, as well as how to disrupt them.

Author contributions: Covasim model development was led by Cliff Kerr, Robyn Stuart, Romesh Abeysuriya, and Daniel Klein, with additional support by Gregory Hart, Katherine Rosenfeld, Prashanth Selvaraj, Rafael Núñez, Jamie Cohen, Lauren George, and Michał Jastrzębski. The SynthPops model was developed by Dina Mistry, with additional support by Cliff Kerr, Romesh Abeysuriya, Daniel Klein, and Lauren George. The health systems component was based on a model developed by Brittany Hagedorn. Data were provided and curated by Meaghan Fagalde and Jeffrey Duchin. Analyses were performed by Cliff Kerr, Dina Mistry, Robyn Stuart, Katherine Rosenfeld, Gregory Hart, Rafael Núñez, Prashanth Selvaraj, and Jamie Cohen. Supervision was provided by Jeffrey Duchin, Michael Famulare, and Daniel Klein. The manuscript was written by Cliff Kerr, Dina Mistry, Robyn Stuart, Rafael Núñez, and Daniel Klein. Manuscript review and editing was performed by Katherine Rosenfeld, Jamie Cohen, and Michael Famulare.

Acknowledgments: Review of data and results was provided by Matthew Golden, Cathy Wasserman, and Ian Painter. Literature and code reviews were performed by Anna Palmer, Dominic Delport, Carrie Bennette, Bradley Wagner, Stewart Chang, Jasmina Panovska-Griffiths, and Edward Wenger. Schematics were produced by Amanda Izzo. Additional contributors to the Covasim model include: from GitHub, William Fitzgerald, Hamel Husain, Cory Gwin, Julian Nadeau, Rasmus Wriedt Larsen, Aditya Sharad, and Oege de Moor; from Microsoft, William Chen, Scott Ayers, and Rolf Harms; from the Institute for Disease Modeling, Mary Fisher, Dennis Chao, Christian Wiswell, Jennifer Schripsema, Christopher Lorton, Clinton Collins, Christopher Jones, Charles Eliot, Svetlana Titova, Dejan Lukacevic, Jeffrey Steinkraus, John Sheppard, Robert Hart, Guillaume Chabot-Couture, Caitlin Bever, Helen Olsen, Greer Fowler, and Natalia Corona; from the Burnet Institute, Nick Scott and Sherrie Kelly; and from the Jet Propulsion Laboratory, Casey Handmer. We also wish to thank the participants of the Covasim Users Group, including Julie Maher, Dean Sidelinger, and Erik Everson from the Oregon Health Authority; André Lin Ouédraogo from the Institute for Disease Modeling; and David P. Wilson from the Bill and Melinda Gates Foundation.

Funding: Bill and Melinda Gates through the Global Good Fund. Institutional support, including high-performance computing resources and library access, was provided by the Burnet Institute and the University of Sydney School of Physics.

Data and materials availability: The Covasim model code is fully open source and available via GitHub. Most data are available from the King County Data Dashboard; data not publicly released are available upon request with permission from Public Health Seattle & King County. Analysis scripts are available from the authors upon request.

Appendix: Methods

We used Covasim (covasim.org), an agent-based model of COVID-19 transmission and interventions developed by the Institute for Disease Modeling, to estimate the extent to which testing, tracing, and quarantine would enable the relaxation of physical distancing measures as part of reopening of the economy. The Covasim model code is fully open source and available online; data and analysis scripts are available upon request. All simulations were run using Covasim version 1.5.0. Section 1 describes the data, analyses, and customizations specific to this study; full details of the Covasim model are provided in Sections 2–4.

1 Data and analyses

1.1 Data and sources

As an agent-based model, Covasim can make use of rich data sources. Mechanistic representations of individuals, contacts, and infections enables physical parameter values to be input directly or used as priors during model fitting. Default values for most inputs come from publicly available data and literature, and are provided below.

To model the Seattle-King County region for this analysis, we used epidemiological data provided by the Washington State Department of Health (WA-DoH) under a use agreement. WA-DoH maintains all COVID-19 data as a line-list in the Washington Disease Reporting System (WDRS), and has provided weekly exports to the study team for the purpose of conducting this and other analyses in support of model-based decision making. We aggregated line-list entries to daily totals by 10-year age bins to produce target data for model calibration. The resulting dataset includes the number of positive and negative tests (by date of sample collection) as well as the number of deaths (by date of death) in King County. The WDRS records also enabled us to characterize the distribution of delays between symptom onset on diagnostic swab; we used these data to validate the implementation of the testing intervention.

The partnership with Public Health Seattle-King County (PHSKC) has provided additional context to ensure the model captures transmission, testing, care, and contact tracing in this setting. Public health co-authors and other county officials have provided data, under use agreement, and insights on testing campaigns, focal outbreaks, schools, hospitalization, and congregate settings such as long-term care facilities (LTCF). Much of these data are available publicly on daily summary and LTCF dashboards. Insights on how the testing program has evolved over the course of the epidemic has been used to set change points for testing parameters, which are identified during the calibration process.

Washington is a home-rule state, meaning that laws can be set at the local level, and as such the contact tracing programs are led by each local health jurisdiction. Seattle-King County was one of the first jurisdictions in the state to pilot contact tracing using local health resources. Today, the county has sufficient human resources to trace approximately half of the cases, while the other half are handled by the Washington State contact tracing program. Data for this analysis were provided by PHSKC.

We used data on weekly foot traffic patterns obtained through SafeGraph to model the degree of mobility in the workplace and community layers of the model. This publicly available dataset is based on anonymized cell phone data, which connects foot traffic counts with points of interest visits for over 5 million unique locations in the United States. This rich dataset enables a detailed view of hourly visits to specific locations. Using an aggregation of visits in King County across all industries starting at the end of January, we classified visits with a dwell-time less than four hours as community-associated mobility, and those visits with a dwell-time of more than four hours as workplace-associated mobility. Using the last week of January as a baseline value for "pre-COVID mobility", we assigned a weekly mobility level for community and work relative to the baseline value.

1.2 Population and network model

To model King County with detailed information on the demographics and network structures of the King County population, we used SynthPops, an open-source data-driven model for generating realistic synthetic contact networks for populations. Further details are provided below. For the population of King County, we used a combination of data sources at the county, state, and country resolution with the SynthPops model (version 0.7.2). Specifically, data from the 2018 American Community Survey (ACS) at the county resolution (*21*) was used to estimate age and household size distributions. The US Census Bureau (*22*) provided data for the age of reference individuals by household size at the country resolution, and age mixing contact matrices for the US are drawn from (*23*). For the network layer of schools, we used 2018 ACS 1-year estimates for county enrollment rates by age (*21*), municipal records on school enrollment numbers (*24*), student-teacher ratios, and the average class size for schools (*25*). For the network layer of workplaces, we used 2018 ACS 1-year estimates for county employment rates by age (*26*) and 2015 county estimates for workplace sizes (*27*).

In King County, a significant percent of COVID-19 cases and deaths have occurred to date within the LTCF population. To capture the dynamics of transmission in this setting, the SynthPops model was extended to include the contact layer of those living and working in long term care facilities to reflect the initial outbreaks that occurred in these facilities (*28*). Data on the demographics of LTCF residents for Washington state (*29*) was used to estimate the number of residents and their ages for Seattle area facilities. From these data, we estimate approximately 15,000 individuals aged 60 and older are residents of long term care facilities within the Seattle area. Additional King County data on the number of residents per facility and resident-to-staff ratios were used to sample facility sizes and populate the facilities with both residents and staff members. There were an average of 123 residents per facility and 132 staff per facility, but there is wide variation in both total numbers of residents and staff and in the resident-to-staff ratios. For the purposes of this model, residents considered to be living in these facilities were not assigned any additional outside contacts (household, school, workplace, or community). Staff members were drawn from the labor force of the population under 60 years of age. With large facility sizes, we modeled close contacts in facilities by sampling for each resident and staff member a subset of 20 contacts, ensuring that each resident is in contact with at least one staff member.

Five synthetic populations of 225,000 modeled agents (representing 10% of King County's population) were generated using SynthPops, with dynamic scaling used to rescale this population to represent the full 2.25 million population of King County (see Section 2.6.3). A typical synthetic population had roughly 500,000 household contacts, 1 million school contacts, 1.9 million workplace contacts, 4.5

million community contacts, and 31,000 LTCF contacts. To reflect the relative amount of time spent with each contact across different layers, relative transmission weights per layer were set to be 100% for households (as a reference value), 50% for LTCFs, 20% for schools and workplaces, and 10% for community contacts. A subset of the simulations were also run with 2.25 million agents to verify that results were consistent with and without rescaling.

1.3 Calibration methodology

We calibrated model parameters using Optuna, a Python-based optimization library (*12*), using the tree-structured Parzen estimator (TPE) sampler (*30*). This sampler trains models of $p(\theta|y)$ and $p(y)$, where θ is a set of parameters and y is a (scalar) output of a loss (objective) function, to find the region of the parameter space that minimizes y. We defined the loss function to be the sum of the absolute differences between observed data and the corresponding model predictions for (a) cumulative diagnoses per day, (b) cumulative deaths per day, (c) total diagnoses by age (using 10-year age bins), (d) average test yield by age, and (e) total deaths by age. To equalize the weight given to each point in each of these five different data types, each data type was normalized by the maximum value in the data.

We used 57,600 simulation runs during the calibration process to ensure broad exploration of parameter space. To determine parametric uncertainty, we used a cutoff value for the loss function of 10, which corresponded to no more than roughly 2% average relative error per point in the diagnoses and deaths time series, and roughly 10% average relative error for diagnoses, yield, and deaths age distributions; this cutoff was also roughly a factor of 2 larger than the single best-fitting simulation (which had a total loss of 4.4). Using this cutoff, the posterior distribution consisted of the 466 best-fitting parameter sets for the calibrations that used SafeGraph mobility data, and the 408 best-fitting parameter sets for calibrations that did not. Median values and 95% confidence intervals for epidemic projections and parameter distributions (Fig. 1) were produced using these parameter sets. Detailed transmission characteristics (Fig. 2) were based on the single best fit with mobility data. For scenario analyses (Fig. 4), the top 10 best-fitting parameter sets with mobility data were used, as described below.

The parameters determined during the calibration process can be grouped into three components, namely: (1) epidemic initial conditions; (2) testing process; and (3) effect of NPIs on disease transmission. Calibrated parameters are shown in Table S1; uniform priors were used. Simulations were initialized with 400 seed infections, distributed at random throughout the population, on February 1. This initialization was chosen by calibrating the number of seed infections and overall transmission rate (β) to the subset of data prior to major policy or mobility changes (i.e., February 27), and for consistency with other estimates of the initial reproduction number in King County. In the calibration, larger numbers of seed infections were compensated for by smaller transmission rates; 400 seed infections was the fewest that could be used (reflecting the highest baseline transmission rate) that provided a reasonable match to the data. We also use SafeGraph data to determine the proportion of network edges in workplace and community layers that should be removed or restored over time based on observed changes in the mobility. Other model parameters were set to use Covasim defaults, as described below.

Table S1. Model parameters and calibrated values determined via fitting model outputs to King County data.

Type	Parameter	Calibrated value with SafeGraph data: median (95% CI)	Calibrated value without SafeGraph data: median (95% CI)	Final search interval
Baseline epidemic properties	Probability of transmission per contact per day (β, %)	4.2 (4.1, 4.4)	4.2 (4.0, 4.3)	[3.3, 4.8]
Testing process	Odds ratio of testing symptomatics before March 22	6.7 (3.3, 9.9)	7.1 (4.6, 10.7)	[1.0, 50.0]
	Odds ratio of testing symptomatics between March 23 and April 14	16.5 (12.0, 21.0)	17.5 (13.7, 21.1)	
	Odds ratio of testing for symptomatics after April 15	22.6 (16.9, 25.6)	22.4 (16.2, 28.4)	
Effect of NPIs on disease transmission	Relative transmission rate in work and community layers from March 12 to March 22 (%)	78 (71, 83)	29 (25, 33)	With SafeGraph: [50, 100]
	Relative transmission rate in work and community layers from March 23 to April 24 (%)	73 (69, 78)	28 (24, 32)	Without SafeGraph: [10, 60]
	Relative transmission rate in work and community layers after April 25 (%)	67 (60, 90)	31 (22, 37)	
	Relative transmission rate in LTCFs after March 23 (%)	17 (13, 21)	22 (15, 34)	[10, 40]

1.4 Idealized test-trace-quarantine scenarios

To explore the theoretical properties of test-trace-quarantine, we created a hypothetical population of 30,000 people with 100 seed infections starting on day 0 and simulated for 150 days. Population demographics were based on Seattle, Washington, USA, but contact networks were generated using a "hybrid" algorithm rather than SynthPops; this algorithm is described in Section 2.4.3. Simulations were run with 10 different random seeds, for three different transmission levels: medium transmission (β = 4.2% per household contact per day, consistent with estimated β for Seattle), low transmission (β = 3.3%), and high transmission (β = 5.1%). Parameters for each of the three intervention scenarios (physical distancing, testing, and testing plus tracing) were chosen to bring $R_e \approx 1$ for the medium transmission scenario. These intervention parameters were held constant for the low and high transmission scenarios.

1.5 Realistic test-trace-quarantine scenarios

We used the 10 parameter configurations with the best fit to the data over the period February 1 to June 10 as the basis for the test-trace-quarantine (TTQ) scenarios shown in Fig. 4A. Scenarios were also run with other sets of calibrations (including the top 100, and using the same goodness-of-fit threshold used for the distributions shown in Fig. 1E). These results did not differ qualitatively and only modest quantitative differences were observed; the top 10 calibrations were chosen to ensure the best fit to data while still capturing both parametric and stochastic uncertainty. To explore the relative importance of different intervention parameters, we ran a sweep of 50 points for each of the six

parameters (described below), for each of the 10 parameter configurations, for a total of 3,000 simulations. Each scenario began on June 10 and ended on August 30, which was chosen as the period prior to the potential reopening of schools. Scenarios began with an immediate return to 100% mobility in work and community (from a baseline value of 43% from the last reported SafeGraph data on June 1) as well as immediate implementation of the testing and contact tracing interventions with the parameter values described below. Relative β (compared to baseline), reflecting mask use and other non-pharmaceutical interventions, remained constant throughout the scenarios at its last calibrated value, which varied from 73% to 82%.

For each scenario, only one parameter at a time was varied. Note that the parameters interact nonlinearly; for example, the impact of the contact tracing delay depends on the amount of contact tracing. Additionally, the impact of interventions is dependent on the epidemic dynamics: with very low mobility and hence baseline transmission, the impact of interventions will be reduced. Thus, the baseline scenario was chosen to reflect (a) a situation where $R_e \approx 1$, which is the point most sensitive to small differences in intervention effectiveness; and (b) a balance between testing and contact tracing that is intended to reflect a realistic scale-up of both current programs. While other baseline points would be possible, this scenario aims to reflect a potentially achievable point by which Seattle-King County could maintain $R_e \approx 1$ with full reopening.

The six intervention parameters are defined as follows:

1. **Isolation/quarantine effectiveness:** The relative change in transmission following either diagnosis (isolation) or after being notified as a known contact (quarantine). While in practice (and for the assumptions used during the calibration period) the effectiveness would differ between isolation and quarantine (with isolation expected to have higher effectiveness), as well as between contact layers (with a greater reduction in workplace and community transmission compared to household), here we used a single weighted average value to ensure that the slope has meaningful units (i.e., infections averted per person fully isolated or quarantined). The default value chosen was 80% efficacy, which is a weighted average between workplace and community contacts (where isolation efficacy is likely to be higher, e.g. 90–95% effectiveness) and household contacts (where isolation efficacy is likely to be lower, e.g. 40–70% effectiveness). This parameter was varied from 0% (no impact of isolation/quarantine) to 100% (zero transmission during isolation/quarantine).
2. **Contact tracing probability:** The proportion of household, workplace, and LTCF contacts of a person who has been diagnosed who are reached by contact tracers. (The proportion of community contacts reached is assumed to be zero for this analysis; schools are closed for the scenario period so there are no school contacts to trace.) The default value chosen was 50%, which again reflects a weighted average between household and LTCF contact tracing (where probabilities well above 80% are achievable) and workplace contact tracing (where probabilities in Seattle-King County are currently low). This parameter was varied from 0% (no contacts traced) to 100% (all household and workplace contacts traced).
3. **Quarantine testing probability:** The probability that a known contact, once traced, will be tested for COVID-19. The default values used were 80% for people with symptoms and 8% for people without symptoms, to approximately reflect current preliminary estimates from the contact tracing program, although these parameters are not well constrained by current data. In all cases, asymptomatic testing probabilities were assumed to be a factor of 10 lower than symptomatic testing probabilities. This parameter was varied from 0% (no testing of people in

quarantine) to 100% testing of people with symptoms and 10% testing of people who do not have symptoms (including uninfected, asymptomatic, and presymptomatic contacts).

4. **Routine testing probability:** The probability per person per day of a person receiving a test for COVID-19. The default values chosen were 16% per day for a person with active symptoms, and 0.16% for people who are uninfected or who do not have symptoms. These values correspond to an approximate doubling of the number of daily tests relative to June 10; the ratio of probabilities for people with and without symptoms was set to be 100, which was chosen to be consistent with the observed testing yield in the data (approximately 1.5–2.5%). This parameter was varied from 0% (no routine testing) to 50% daily symptomatic testing and 0.5% daily non-symptomatic testing, corresponding to a roughly 4-fold increase in testing rates compared to June 10.
5. **Swab-to-result delay:** The average number of days between when a person receives a COVID-19 swab to when they are notified of their result. The default value chosen was 1 day, reflecting a slight improvement on practice in Seattle-King County as of June 10 (approximately 1.5 days). This parameter was varied from 0 days (immediate return of test results) to 7 days.
6. **Contact tracing delay:** The average number of days between when a person receives a positive result from a COVID-19 test and when their contacts are traced and notified. The default value chosen was 2 days, which is somewhat shorter than estimates as of June 10 (3-5 days). This parameter was varied from 0 days (immediate notification of all contacts, although the swab-to-result delay is still present) to 7 days.

Because epidemic growth is an exponential process, the attack rate varied widely between scenarios, from less than 0.1% to 10%. The attack rate had nonlinear dependence on all intervention parameters. Thus, the attack rate was log-transformed prior to fitting. The ordinary least squares method from the Python package *statsmodels* was used for the fit. The uncertainty interval shown is the 95% confidence interval from the ordinary least squares fit. Because of the log transform, the slope of the line depends on the point of evaluation; in all cases, it was evaluated at the default value for each parameter. Since the dependent variable in the regression is attack rate, the slope is also dependent on the period of integration (here, 81 days); a longer integration period, for example, would lead to a higher cumulative attack rate and thus larger slopes.

For the reopening sweeps (Fig. 4B), all parameters except for routine testing probability and contact tracing probability were fixed at the default values described above. These two parameters were swept over the ranges described above, although simulations for which contact tracing probability was above 90% probability or more than 4 tests per 1,000 people per day were conducted were deemed unrealistic and are not shown in the figure. We simulated eight different reopening levels (65% to 100%, in 5% increments), and show 70%, 85%, and 100% to represent low, medium, and high transmission scenarios. Each sweep consists of 8000 simulations, with each simulation drawn from a uniform random distribution for (a) routine testing probability and (b) contact tracing probability, with each simulation also drawing from one of the 10 best calibrations as described above.

2 Covasim methodology

Models for examining COVID-19 transmission and control measures can be broadly divided into two main types: compartmental models and agent-based (or microsimulation) models, with the former generally being simpler and faster, while the latter are generally more complex, detailed, and computationally expensive. Numerous compartmental models have been developed or repurposed for COVID-19: Walker et al. (*31*) used an age-structured stochastic "susceptible, exposed, infectious, recovered" (SEIR) model to determine the global impact of COVID-19 and the effect of various social distancing interventions to control transmission and reduce health system burden; Read et al. (*32*) developed an SEIR model to estimate the basic reproduction number in Wuhan; Keeling et al. (*33*) use one to look at the efficacy of contact tracing as a containment measure; and Dehning et al. (3) used an SIR model to quantify the impact of intervention measures in Germany. In models such as those by Giordano et al. (*34*) and Zhao and Chen (*35*), compartments are further divided to provide more nuance in simulating progression through different disease states, and have been deployed to study the effects of various population-wide interventions such as social distancing and testing on COVID-19 transmission.

For microsimulation models, several agent-based influenza pandemic models have been repurposed to simulate the spread of COVID-19 transmission and the impact of social distancing measures in Australia (*36*), Singapore (*37*), the United States (*38*), and the United Kingdom (*39*). Additionally, agent-based models have been developed to evaluate the impact of social distancing and contact tracing (*3*, *5*, *40*). Features of these models include accounting for the number of household and non-household contacts (*3*, *5*, *38*); the age and clustering of contacts within households (*3*, *38*, *40*); and the microstructure in schools and workplace settings informed by census and time-use data (*40*). Branching process models have also been used to investigate the impact of non-pharmaceutical intervention strategies (*4*, *41*).

In developing Covasim, our aim was to capture the benefits of agent-based modeling (in particular, the ability of such models to simulate the kinds of microscale policies being used to respond to the COVID-19 pandemic), whilst making use of recent advances in software tools and computational methods to minimize the complexity and computational time typically associated with such models.

2.1 Overview

Covasim simulates the state of individual people, known as agents, over a number of discrete time steps. Conceptually, the model is largely focused on a single type of calculation: the probability that a given agent on a given time step will change from one state to another, such as from susceptible to infected, or from critically ill to dead. Once these probabilities have been calculated, a pseudorandom number generator with a user-specified seed is used to determine whether the transition actually takes place for a given model run.

The logical flow of a single Covasim run is as follows. First, the simulation object is created, then the parameters are loaded and validated for internal consistency, and any specified data files are loaded (described in Section 2.6.5). Second, a population is created, including age, sex, and comorbidities for each agent, drawing from location-specific data distributions where available. Third, agents are connected into contact networks based on their age and other statistical properties (Section 2.4). Next, the integration loop begins. On each timestep (which corresponds to a single day by default), the order

of operations is: dynamic rescaling (Section 2.6.3); applying health system constraints (Section 2.6.1); updating the state of each agent, including disease progression (Section 2.2); importation events (Section 2.6.2); applying interventions (Section 2.5); calculating disease transmission events across each infectious agent's contact network (Section 2.3); and the collation of outputs into results arrays (Section 2.6.6). The following sections describe each step in more detail.

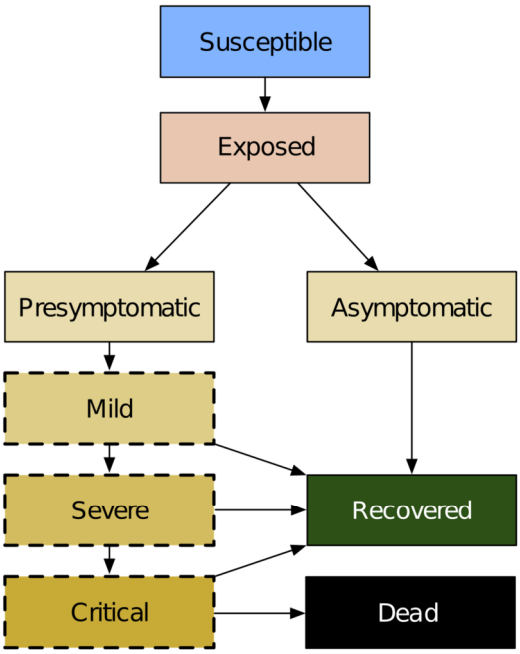

Figure S1. Covasim model structure, including infection (exposure), disease progression, and final outcomes. Yellow shading indicates that an individual is infectious and can transmit the disease to other susceptible agents. States with a dashed border are considered symptomatic with respect to symptomatic versus asymptomatic testing.

2.2 Disease progression

In Covasim, each individual is characterized as either susceptible, exposed (i.e., infected but not yet infectious), infectious, recovered, or dead, with infectious individuals additionally categorized according to their symptoms: asymptomatic, presymptomatic, mild, severe, or critical. A schematic diagram of the model structure is shown in Fig. S1.

The length of time after exposure before an individual becomes infectious is assumed to follow a log-normal distribution with a mean of 4.6 days, which is within the range of values reported across the literature (Table S2). The length of time between the start of viral shedding and symptom onset is assumed to follow a log-normal distribution with a mean of 1 day (Table S2). Exposed individuals may develop symptoms or may remain asymptomatic. Individuals with symptoms are disaggregated into either mild, severe, or critical cases, with the probability of developing a more acute case increasing with age (Table S3). Covasim can also model the effect of comorbidities, which act by modifying an individual's probability of developing severe symptoms (and hence critical symptoms and death). By default, comorbidity multipliers are set to 1 since they are already factored into the marginal age-dependent disease progression rates.

Estimates of the duration of COVID-19 symptoms and the length of time that viral shedding occurs are highly variable, but durations are generally reported to increase according to acuity (*42*, *43*). We reflect

this in our model with different recovery times for asymptomatic individuals, those with mild symptoms, and those with severe symptoms, as summarized in Table S2. All non-critical cases are assumed to recover, while critical cases either recover or die, with the probability of death increasing with age (Table S3).

Table S2. Default duration parameters, in days, used in the Covasim model.

Parameter	Description	Distribution (mean, std)	Source
s	Length of time after exposure before an individual is infectious (i.e., has begun viral shedding)	$s \sim lognormal(4.6,4.8)$	From Lauer et al. (*44*); additional sources Du et al., Nishiura et al., and Pung et al. (*45–47*).
i	Length of time after viral shedding has begun before an individual has symptoms	$i \sim lognormal(1,0.9)$	Linton et al. (*48*) report the incubation period as 5.6 days (95% CI: 5.0–6.3 days). Using the period of exposure before becoming infectious, we infer the period of viral shedding before symptomatic. However, other studies have estimated longer periods, e.g. (*17*).
r_a	Recovery time for asymptomatic cases	$r_a \sim lognormal(8, 2)$	Wölfel et al. (*49*)
r_m	Recovery time for mild cases	$r_m \sim lognormal(8, 2)$	Wölfel et al. (*49*)
r_s	Recovery time for severe cases	$r_s \sim lognormal(14, 2.4)$	Verity et al. (*50*)
r_c	Recovery time for critical cases	$r_c \sim lognormal(14, 2.4)$	Verity et al. (*50*)

Table S3. Age-linked disease susceptibility, progression, and mortality probabilities. Key: r_{sus}: relative susceptibility to infection; p_{sym}: probability of developing symptoms; p_{sev}: probability of developing severe symptoms (i.e., sufficient to justify hospitalization); p_{cri}: probability of developing into a critical case (i.e., sufficient to require ICU); p_{dea}: probability of death. Relative susceptibility values are derived from odds ratios presented in Zhang et al. (*51*); all other values are derived from Verity et al. (*50*) and Ferguson et al. (*39*), with validation from model fits to data on numbers of cases, numbers of people hospitalized and in intensive care, and numbers of deaths from Washington state and Oregon. *Overall values depend on the age structure of the population being modeled. For a population like the US or UK, the symptomatic proportion is roughly 0.7, while for populations skewed towards younger ages, the symptomatic proportion is lower.

	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80+	Overall*
r_{sus}	0.34	0.67	1.00	1.00	1.00	1.00	1.00	1.24	1.47	1.00
p_{sym}	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.5–0.75
p_{sev}	0.00050	0.00165	0.00720	0.02080	0.03430	0.07650	0.13280	0.20655	0.24570	0.1–0.2
p_{cri}	0.00003	0.00008	0.00036	0.00104	0.00216	0.00933	0.03639	0.08923	0.17420	0.05–0.1
p_{dea}	0.00002	0.00006	0.00030	0.00080	0.00150	0.00600	0.02200	0.05100	0.09300	0.005–0.015

2.3 Transmission and within-host viral dynamics

Whenever a susceptible individual comes into contact with an infectious individual on a given day, transmission of the virus occurs according to probability β. For a well-mixed population where each individual has an average of 20 contacts per day, a value of $\beta = 0.016$ corresponds to a doubling time of roughly 4–6 days and an R_0 of approximately 2.2–2.7, where the exact values depend on the population size, age structure, and other factors. The value of $\beta = 0.016$ is used as the default in Covasim; however, this value is typically calibrated by the user to best match local epidemic data.

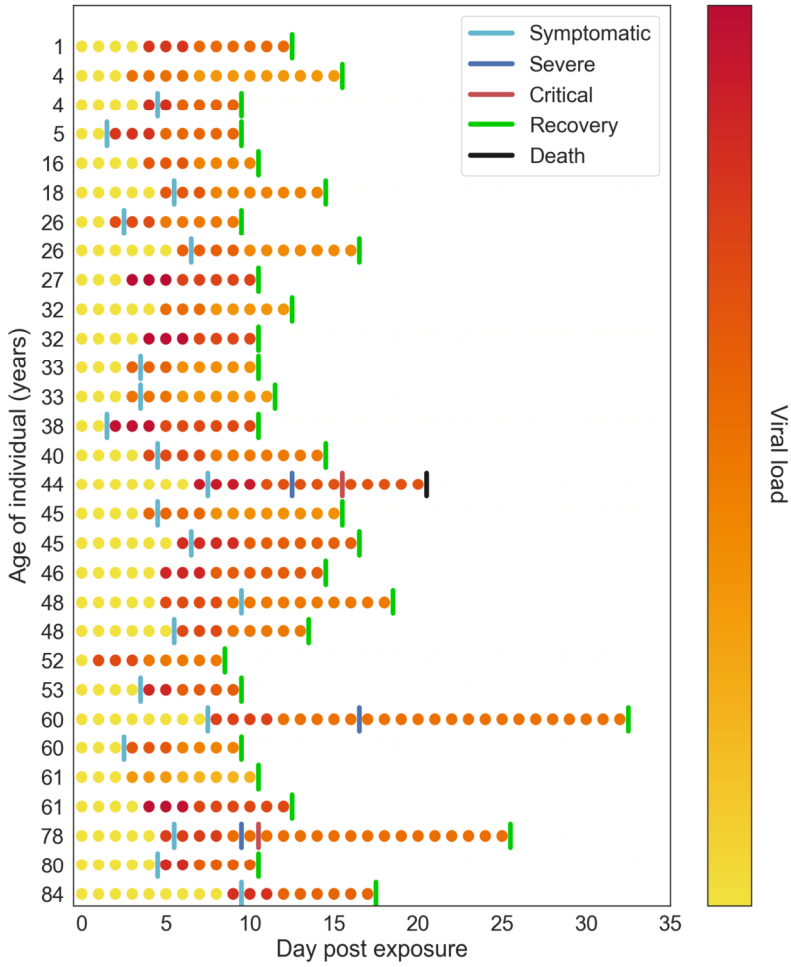

Figure S2. Example of within-host viral load dynamics in Covasim. Each row shows a different agent in the model. Color indicates viral load, which typically peaks the day before or the day of symptom onset, before declining slowly.

If realistic network structure (i.e., households, schools, workplaces, and community contacts) is included, the value of β depends on the contact type. Default transmission probabilities are roughly 0.05 per contact per day for households, 0.025 for workplaces and schools, and 0.005 for the community. These values correspond to relative weightings of 10:2:2:1, with a weighted mean close to the default β value of 0.016 for a well-mixed population (i.e., if different network layers are not used). When combined with the default number of contacts in each layer, age-based susceptibility, and other factors, for a typical (unmitigated) transmission scenario, the proportions of transmission events that occur in each contact layer are approximately 25% via households, 35% via workplaces, 15% via schools, and 25% via the community. The value of β can also be modified by interventions, such as physical distancing, as described below.

In addition to allowing individuals to vary across disease severity and time spent in each disease state, we allow individual infectiousness to vary between people and over time. We use individual viral load to model these differences in infectivity. Several groups have found that viral load is highest around or slightly before symptom onset, and then falls monotonically (*52–55*). As a simple approximation to this viral time course, we model two stages of viral load: an early high stage followed by a longer low stage. By default, we set the viral load of the high stage to be twice as high as the low stage and to last for either 30% of the infectious duration or 4 days, whichever is shorter. The default viral load for each

agent is drawn from a negative binomial distribution with mean 1.0 and shape parameter 0.45, which was the value most consistent with both international estimates (*56*, *57*) and fits to data in Washington state and Oregon. The daily viral load is used to adjust the per-contact transmission probability (β) for an agent on a given day (Fig. S2).

Evidence is mixed as to whether transmissibility is lower if the infectious individual does not have symptoms (*52*). We take a default assumption that it is not, but include a parameter that can be modified as needed depending on the modeling application or context, noting that some studies have used much lower rates of infectiousness for asymptomatic individuals (*18*).

2.4 Contact network models

Covasim is capable of generating and using three alternative types of contact networks: random networks, SynthPops networks, and hybrid networks. Each of these may be useful in different settings, and in addition users have the option of defining their own networks. A comparison of the main features and use cases of each of Covasim's default network model options is shown in Table S4 and schematically in Fig. S3; details on each are provided in the following sections. In addition, to facilitate easy adaptation to different contexts, Covasim comes pre-loaded with data on country age distributions and household sizes as reported by the UN Population Division 2019 (population.un.org).

2.4.1 Random networks

Covasim generates random networks by assuming that each person in the modeled population can come into contact with anyone else in the population. Each person is assigned a number of daily contacts, which is drawn from a Poisson distribution whose mean value can be specified by the user depending on the modeling context (with a default value of 20). The user can also decide whether these contacts should remain the same throughout the simulation, or whether they should be sampled randomly from the population each day.

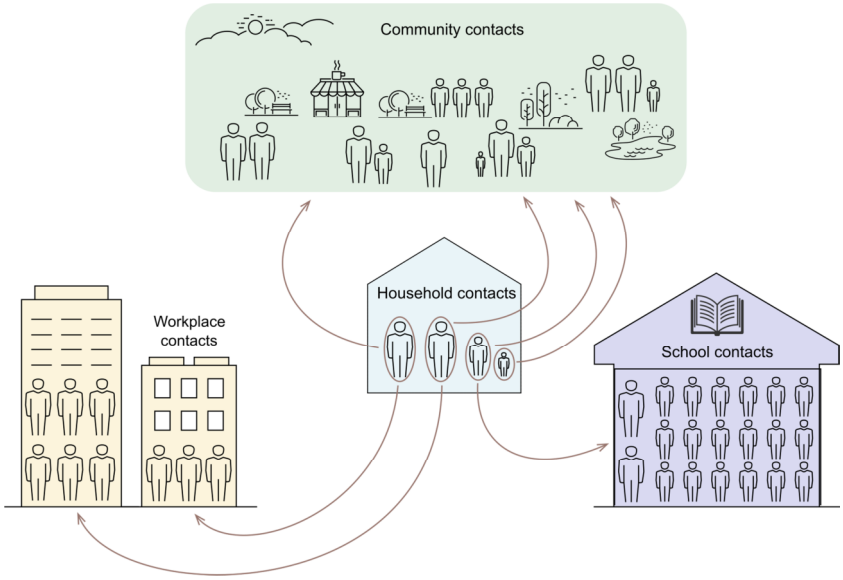

Figure S3. Schematic diagram of contact networks with multiple layers in Covasim.

2.4.2 SynthPops networks

Covasim is integrated with SynthPops, an open-source data-driven model capable of generating realistic synthetic contact networks for populations; further information, including documentation and

source code, is available from synthpops.org. Briefly, the method draws upon previously published models and empirical studies to infer high-resolution age-specific contact patterns in key settings (e.g., households, schools, workplaces, and the general community) relevant to the transmission of infectious diseases (*58–60*). Census and municipal data are used by SynthPops to inform demographic characteristics (e.g., age, household size, school enrollment, employment rates). Age-specific contact matrices (such as those in Prem et al. (*23*)) are then used to generate individuals and their expected contacts in a multilayer network framework. By default, SynthPops generates household, school, and work contact networks; community connections are generated using the random approach described above. An example synthetic network as generated by SynthPops is shown in Fig. S4.

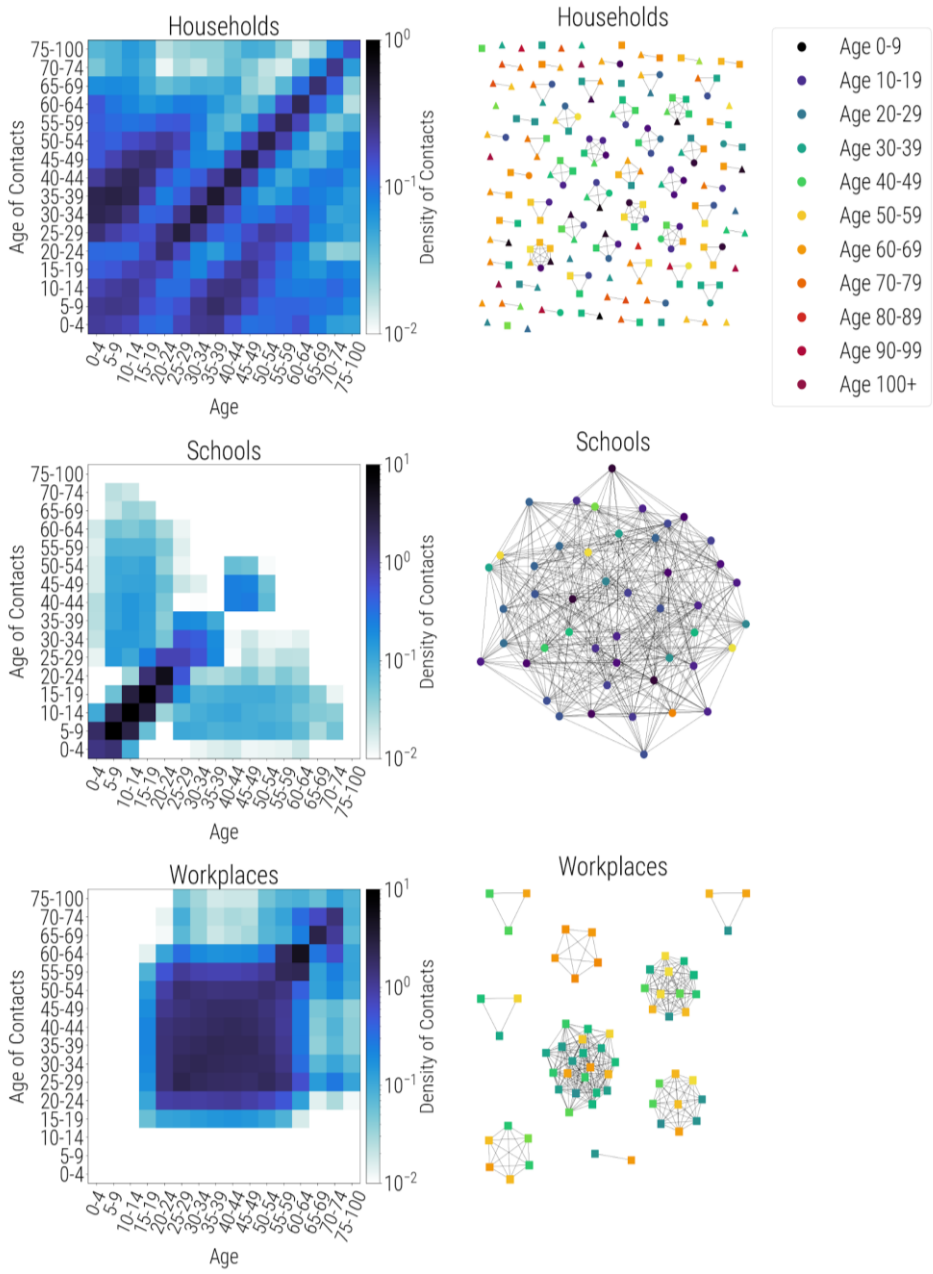

Figure S4. Synthetic population networks for households (top), schools (middle), and workplaces (bottom). Age-specific contact matrices are shown on the left, while actual connectivity patterns for a 127-person subsample of a population of 10,000 individuals are shown on the right. All individuals are present in the household network, including some with no household connections. A subset of these individuals, including teachers, are present in the school network (circles); another subset is present in workplace networks (squares); some individuals are in neither school nor work networks (triangles).

Table S4: Comparison of population options in Covasim.

Population type	Data requirements	Best suited for	Not well suited for
Random networks	None	Models of transmission in special settings such as prisons or cruise ships	Large or complex populations
Hybrid networks	Data on the age distribution and household sizes for each country are pre-loaded No additional data is required, but users can optionally specify the daily number of school, workplace, and community contacts	Population network models in data-rich settings; adaptable and suited to most modeling contexts	Populations with high heterogeneity in contact patterns or size distributions
SynthPops networks	Household, school, workplace, and community age mixing patterns School size distributions, enrollment rates by age, student-teacher ratios Workplace size distributions, employment rates by age Number of households, size distribution, and age/sex distribution	Complex populations in data-rich settings	Settings where the data requirements cannot be met, or where other social settings are critical contexts for disease transmission

2.4.2.1 Households

SynthPops uses data on the distribution of ages, household sizes, and the age of reference individuals conditional on household size for a given population, to generate individuals within households. The algorithm first generates household sizes from the household size distribution, and then assigns a reference individual (for example, the head of the household) with their sampled age conditional on the household size. To construct the other household members, age mixing contact matrices are used to infer the likely ages according to the age of the reference person and the population age distribution adjusted for non-reference ages.

2.4.2.2 Schools

A similar approach is used to construct schools. Census and municipal school data, or survey data such as from Demographic and Health Surveys (*61*), can be used to inform enrollment rates by age, school sizes, and student-teacher ratios. The SynthPops algorithm first chooses a reference student for the school conditional on enrollment rates to infer the school type, and then uses the age mixing contact matrix in the school setting to infer the likely ages of the other students in the school. Students are drawn from an ordered list of households, such that they reproduce an approximation of the neighborhood dynamics of children attending area schools together. Teachers are drawn from the adult population comprising the labor force and assigned to schools as needed, reflecting average student-teacher ratio data. With large schools, it is unlikely for each student and teacher to be in close contact with all other individuals. Instead, for each individual in the school layer we model their close

and effective contacts as a subset of contacts from their school who can infect them by sampling a random set of *n* other individuals in their school, where *n* is drawn from a Poisson distribution with rate parameter λ_s equal to the average class size ($\lambda_s = 20$ as a default).

2.4.2.3 Workplaces and community

The labor force is drawn using employment rates by age, and non-teachers are assigned to workplaces using data on establishment sizes. Workers are assigned to workplaces using a similar method with an initial reference worker sampled from the labor force and their co-workers inferred from age mixing patterns within the workforce. All workers (teachers included) are drawn at random from the population, to reflect the general mixing of adults from different neighborhoods at work. Similar to the school layer, large workplaces are unlikely to be fully connected graphs of contacts. Instead, for each worker, we model their close contacts as a subset of *n* contacts from other individuals in their workplace, where *n* is drawn from a Poisson distribution with rate parameter λ_w equal to the estimated maximum number of close contacts in the workplace ($\lambda_w = 20$ as a default).

For contacts in the general community, we draw *n* random contacts for each individual from other individuals in the population, where *n* is drawn from a Poisson distribution with rate parameter λ_c equal to the expected number of contacts in the general community (with $\lambda_c = 20$ as a default, as above). Connections in this layer reflect the nature of contacts in shared public spaces like parks and recreational spaces, shopping centers, community centers, and public transportation. All links between individuals are considered undirected to reflect the ability of either individual in the pair to infect each other.

The generated multilayer network of household, school, work, and general community network layers presents a population with realistic microstructure. This framework can also be extended to consider more detailed interactions in key additional settings, such as hospitals, encampments, shelters for those experiencing homelessness, and long-term care facilities.

2.4.3 Hybrid networks

Covasim contains a third option for generating contact networks, which captures some of the realism of the SynthPops approach but does not require as much input data, and is more readily adaptable to other settings. As such, it is a "hybrid" approach between a fully random network and a fully data-derived network. As with SynthPops, each person in the population has contacts in their household, school (for children), workplace (for adults), and community. A population of individuals is generated according to a location-specific age distribution, and each individual is randomly assigned to a household using location-specific data on household sizes.

Unlike SynthPops, the hybrid algorithm does not account for the distribution of ages within a household. Children are assigned to schools and adults to workplaces, each with a user-specified number of fixed daily contacts (by default, Poisson-distributed with means of 20 for schools and 16 for workplaces, chosen to match the mean values for SynthPops networks). Individuals additionally have contacts with others in the community (by default, Poisson distributed with a mean of 20). All children and young adults aged between 6 and 22 are assigned to schools and universities, and all adults between 22 and 65 are assigned to workplaces. This distinguishes it from SynthPops where enrollment or employment varies depending on the given data. A comparison of the different population structure options available in Covasim is listed in Table S4.

2.5 Interventions

A core function of Covasim is modeling the effect of interventions on disease transmission or health outcomes, to understand the impact that different policy options may have in a specific setting. In general, interventions are modeled as changes to parameter values. Covasim has built-in implementations of the common interventions described below, as well as the ability for users to create their own interventions.

2.5.1 Physical distancing, masks, and hygiene

The most basic intervention in Covasim is to reduce transmissibility (β) starting on a given day. This can be used to reflect both (a) reductions in transmissibility per contact, such as through mask wearing, personal protective equipment, hand-washing, and maintaining physical distance; and (b) reductions in the number of contacts at home, school, work, or in the community. However, Covasim also includes an "edge-clipping" intervention (considering a contact between two agents as a weighted "edge" between two "nodes"), where β remains unchanged but the number of contacts that person has is reduced. Complete school and workplace closures, for example, can be modeled either by setting β to 0, or by removing all edges in those contact layers; partial closures can be modeled by smaller reductions in either β or the number of contacts.

In general, both types of interventions have similar impact – for example, halving the number of contacts and keeping β constant will produce very similar epidemic trajectories as halving β and keeping the number of contacts constant. However, the distinction becomes important when considering the interaction between physical distancing and other interventions. For example, in a contact tracing scenario, the number of contacts who require tracing, number of tests performed, and number of people placed in quarantine are all strongly affected by whether physical distancing is implemented as a reduction in β of a specific edge, or removing that edge entirely.

2.5.2 Testing and diagnosis

Testing can be modeled in two different ways within Covasim, depending on the format of testing data and purpose of the analysis. The first method allows the user to specify the probabilities that people with different risk factors and levels of symptoms will receive a test on each day. Separate daily testing probabilities can be entered for those with/without symptoms, and those in/out of quarantine. The model will then estimate the number of tests performed on each day. The second method allows the user to enter the number of tests performed on each day directly, including multipliers on the probability of a person receiving a test if they have symptoms, are in quarantine, or are over a certain age. This method will then allocate the tests among the population.

Once a person is tested, the model contains a delay parameter that indicates how long people need to wait for their results, as well as a loss-to-follow-up parameter that indicates the probability that people will not receive their results. Additional parameters control the sensitivity and specificity of the tests.

2.5.3 Contact tracing

Contact tracing corresponds to notifying individuals that they have had contact with a confirmed case, so that they can be quarantined, tested, or otherwise change their behavior. Contact tracing in Covasim is parameterized by the probability that a contact can be traced, and by the time taken to identify and notify contacts. Both parameters can vary by type of contact, and can be controlled by the

user. For example, it may be reasonable to assume that people can trace members of their household immediately and with 100% probability, while tracing work colleagues may take several days and may be incomplete.

2.5.4 Isolation of positives and contact quarantine

Isolation (referring to behavior changes after a person is diagnosed with COVID-19) and quarantine (referring to behavior changes after a person is identified as a known contact of someone with confirmed or suspected COVID-19) are the primary means by which testing interventions reduce transmission. In Covasim, people diagnosed with COVID-19 can be isolated. Their contacts who have been traced can be placed in quarantine with a specified level of compliance; people in quarantine may also have an increased probability of being tested. People in isolation or quarantine typically have a lower probability of infecting others (if infectious) or of acquiring COVID-19 (if quarantined and susceptible). The default reductions for isolation are 70% in the household and 90% in school, work, and community layers, while quarantine is assumed to have lower compliance (40% reduction in the household and 80% in other layers). However, if psychosocial support is not provided to people in home isolation or quarantine, there may be an increased risk of passing on infection to, or acquiring infection from, other household members.

2.5.5 Pharmaceutical and user-defined interventions

Pharmaceutical interventions, including antiviral treatments and vaccines, are not explicitly implemented in Covasim due to the large considerable uncertainties regarding their eventual characteristics and availability. However, they can be defined by adapting existing Covasim interventions. For example, each agent in the model has a relative susceptibility parameter, which is a multiplicative factor on their risk of infection per exposure event. A vaccine of a given efficacy (which could include waning efficacy or increased efficacy from multiple doses) could be implemented by reducing an agent's relative susceptibility after receiving the vaccine. Similarly, antiviral treatments could be modeled by modifying an individual's probabilities of progression to severe disease, critical disease, and death, and by modifying their relative transmissibility.

Each intervention has full access to the simulation object at each timestep, which means that user-defined interventions can dynamically modulate any aspect of the simulation. This can be used to create interventions more specific than those included by default in Covasim; for example, age-specific physical distancing or quarantine. In addition, the same framework can be adapted for other purposes: for example, it is possible to define an "analyzer" function which, at each timestep, records additional details about the internal state of the model that are not included as standard outputs.

2.6 Additional features

2.6.1 Health system capacity

Individuals in the model who have severe and critical symptoms are assumed to require regular and intensive care unit (ICU) hospital beds, respectively, including ventilation in the latter case. The number of available hospital beds (ICU and otherwise) beds are input parameters. If the model estimates that the number of severe/critical cases is greater than the number of available non-ICU/ICU beds, then the health system capacity is exceeded. This means that severely ill individuals have an

increased probability of progressing to critical, and critically ill individuals who are unable to access treatment have an increased mortality rate (by default, both by a factor of 2).

2.6.2 Importations

The spatial movement of agents is not currently modeled explicitly in Covasim, and the population size for a given simulation is fixed. Thus, importations are implemented as spontaneous infections among the susceptible population. This corresponds to agents who become infected elsewhere and then return to the population.

2.6.3 Dynamic scaling

One of the major challenges with agent-based models is simulating a sufficient number of agents to capture an epidemic at early, middle, and late stages, without requiring cumbersome levels of memory or processor usage. Whereas compartmental SEIR models require the same amount of computation time regardless of the population size being modeled, the performance of agent-based models typically scales linearly or supralinearly with population size (see Section 2.7.1). As a consequence, many agent-based models, including Covasim, include an optional "scaling factor", where a single agent in the model is assumed to represent multiple people in the real world. A scaling factor of 10, for example, corresponds to the assumption that the epidemic dynamics in a city of 2 million people can be considered as the sum of the epidemic dynamics of 10 identical subregions of 200,000 people each.

However, the limitation of this approach is that it introduces a discretization of results: model outputs can only be produced in increments of the scaling factor, so relatively rare events, such as deaths, may not have sufficient granularity to reflect the epidemic behavior at a small scale. In addition, using too few agents in the model introduces stochastic variability patterns that do not reflect real-world processes in the entire population.

To circumvent this, Covasim includes an option for dynamic scaling. Initially, when the epidemic is small, there is no scaling performed: one agent corresponds to one person. Once a certain threshold is reached, however (by default, 5% of the population is non-susceptible), the non-susceptible agents in the model are downsampled and a corresponding scaling factor is introduced (by default, a factor of 1.2 is used). For example, in a simulation of 100,000 agents representing a true population of 1 million with a threshold of 10% and a rescaling factor of 2, dynamic scaling would be triggered when cumulative infections surpass 10,000, leaving 90,000 susceptible agents; dynamic rescaling would resample the non-susceptible population to 5,000 (representing 10,000 people) and increase the number of susceptible agents to 95,000 (representing 190,000 people), with every agent now counting as two. If the epidemic expands further, this process will repeat iteratively until the scale factor reaches its upper limit (which in this example is 10, and which would be reached after 100,000 cumulative infections). Through this process, arbitrarily large populations can be modeled, even starting from a single infection, maintaining a constant level of precision and computation time throughout.

2.6.4 Model outputs

By default, Covasim outputs three main types of result: "stocks" (e.g., the number of people with currently active infections on a given day), "flows" (e.g., the number of new infections on a given day), and "cumulative flows" (e.g., the cumulative number of infections up to a given day). For states that cannot be transitioned out of (e.g. death or recovery), the stock is equal to the cumulative flow. Flows

that are calculated in the model include: the number of new infections and the number of people who become infectious on that timestep; the number of tests performed, new positive diagnoses, and number of people placed in quarantine; the number of people who develop mild, severe, and critical symptoms; and the number of people who recover or die. The date of each transition (e.g., from critically ill to dead) is also recorded. By default, these results are summed over the entire population on each day; results for subpopulations can be obtained by defining custom analyzers, as described in Section 2.5.5.

In addition to these core outputs, Covasim includes several outputs for additional analysis. For example, several methods are implemented to compute both doubling time and the effective reproduction number R_e. By default, the epidemic doubling time is computed using the "rule of 70" (*62*), specifically:

$$T = \frac{w \log(2)}{\log(\, n_i(t)/(n_i(t-w))}$$

where T is the doubling time, w is the window length over which to compute the doubling time (3 days by default), and $n_i(t)$ is the cumulative number of infections at time t.

Numerous definitions of the effective reproduction number exist. In standard SIR modeling, the most common definition ("method 1") is (*63*):

$$R_e \; = R_0 S/N$$

where R_0 is the basic reproduction number, S is the number of susceptibles, and N is the total population size. However, with respect to COVID-19, many authors instead define R_e to include the effects of interventions, due to the implications that $R_e = 1$ has for epidemic control. A second common definition of R_e ("method 2") is to first determine the total number of people who became infectious on day t, then count the total number of people these people went on to infect, and then divide the latter by the former. "Method 3" is the same as method 2, except it counts the number of people who stopped being infectious on day t (i.e., recovered or died), and then counts the number of people those people infected. While methods 2 and 3 are implemented in Covasim, they have the disadvantage that they introduce significant temporal blurring, due to the potentially long infectious period (and, for method 3, the long recovery period). To avoid this limitation, the default method Covasim uses for computing R_e is to divide the number of new infections on day t by the number of actively infectious people on day t, multiplied by the average duration of infectiousness.

2.6.5 Data inputs

In addition to the demographic and contact network data available via SynthPops, Covasim includes interfaces to automatically load COVID-19 epidemiology data, such as time series data on deaths and diagnosed cases, from several publicly available databases. These databases include the Corona Data Scraper (coronadatascraper.com), the European Centre for Disease Prevention and Control (ecdc.europa.eu), and the COVID Tracking Project (covidtracking.com). At the time of writing, these data are available for over 4,000 unique locations, including most countries in the world (administrative

level 0), all US states and many administrative level 1 (i.e., subnational) regions in Europe, and some administrative level 2 regions in Europe and the US (i.e., US counties).

2.6.6 Calibration

The process of calibration involves finding parameter values that minimize the difference between observed data (which typically includes daily confirmed cases, hospitalizations, deaths, and number of tests conducted) and the model predictions. In practice, minimizing the difference between the model and data equates to maximizing a log-likelihood function. Since most data being calibrated to are time series count data, this function is defined as:

$$L = \sum_s \sum_t w_s \, f(c_d^{s,t}, c_m^{s,t})$$

where s is a time series of observations (such cumulative confirmed cases or number of deaths); t is the time index; w_s is the weight associated with s; $c_d^{s,t}$ and $c_m^{s,t}$ are the counts from the data and model, respectively, for this time series at this time index; and f is the loss, objective, or goodness-of-fit function (e.g., normalized absolute error, mean absolute error, mean squared error, or the Poisson test statistic). By default, if data are loaded into a simulation, Covasim calculates the loss using this method.

Calibrating any model to the COVID-19 epidemic is an inherently difficult task: not only is there significant uncertainty around the reported data, but there are also many possible combinations of parameter values that could give rise to these data. Thus, in a typical calibration workflow, most parameters are fixed at the best available values from the literature, and only essential parameters (for example, β) are allowed to vary.

Currently, calibration must be performed externally to Covasim. However, since a single model run returns a scalar loss value, these runs can be easily integrated into standardized calibration frameworks. An example implementation using Weights & Biases (wandb.com) is included in the codebase, but any standard optimization library – such as the optimization module of SciPy – can be easily adapted, as can more advanced methods such as the adaptive stochastic descend method of the Sciris library (*64*), or Bayesian approaches such as history matching (*65*) or Optuna (*12*).

2.7 Software architecture

Covasim was developed for Python 3.7 using the SciPy (scipy.org) ecosystem (*66*). It uses NumPy (numpy.org), Pandas (pandas.pydata.org), and Numba (numba.pydata.org) for fast numerical computing; Matplotlib (matplotlib.org) and Plotly (plotly.com) for plotting; and Sciris (sciris.org) for data structures, parallelization, and other utilities.

The source code for Covasim is available via both the Python Package Index (via `pip install covasim`) and GitHub (github.com/institutefordiseasemodeling/covasim). Covasim is fully open-source, released under the Creative Commons Attribution-ShareAlike 4.0 International Public License. More information is available at covasim.org, with full documentation at docs.covasim.org.

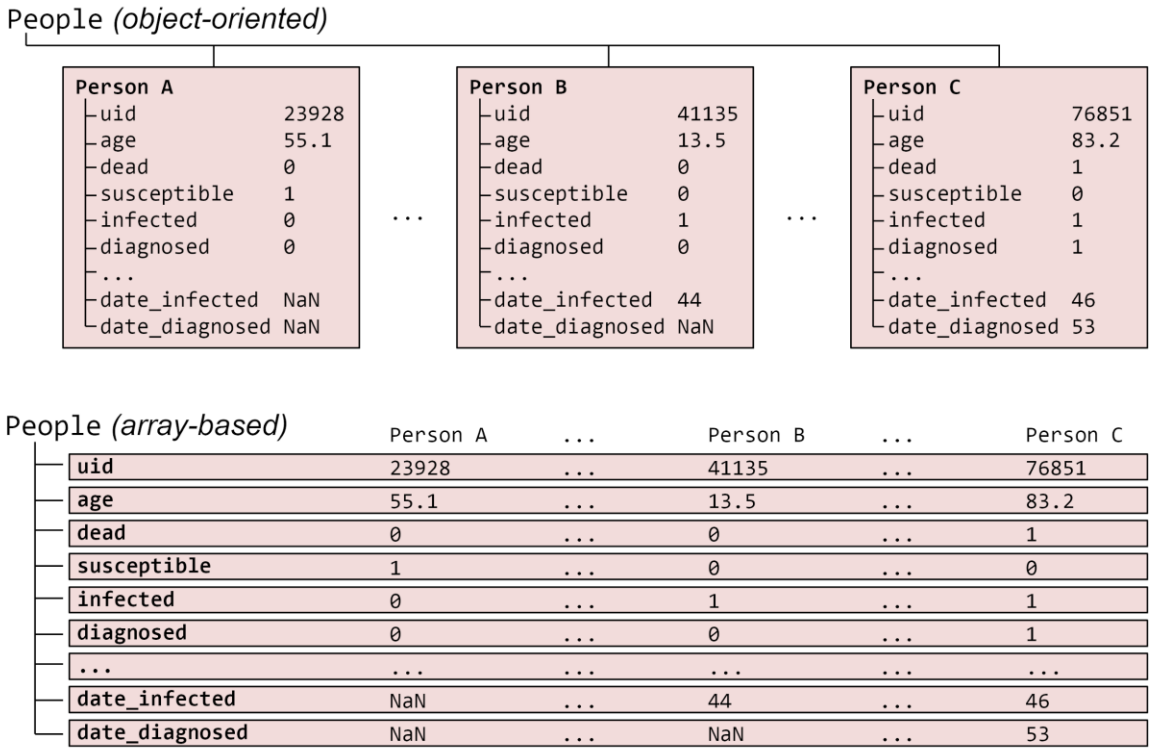

Figure S5. Illustration of the standard object-oriented approach for implementing agent-based models (top), where each agent is a separate object, compared with the approach used in Covasim (bottom), where agents are represented as slices through a set of state arrays. Dots (...) represent omitted entries; in practice, each agent has 39 states, and there are typically tens or hundreds of thousands of agents.

2.7.1 Performance

All core numerical algorithms in the Covasim integration loop – specifically, calculating intra-host viral load, per-person susceptibility and transmissibility, and which contacts of an infected person become infected themselves – are implemented as highly optimized 32-bit array operations in Numba. For further efficiency, agents are not represented as individual objects, but rather as indices of one-dimensional state arrays (Fig. S5). This approach avoids the need to use an explicit for-loop over each agent on every integration timestep. Similarly, contacts between all agents in the model are stored as a single array of "edges" per contact layer.

As shown in Fig. S6, these software optimizations allow Covasim to achieve performance comparable to C++, despite being implemented purely in Python. Performance scales linearly with population size over multiple orders of magnitude: memory scales at a rate of roughly 1 KB per agent, while compute time (benchmarked on an Intel i9-8950HK laptop processor) scales at a rate of roughly 2 million simulated person-days per second of CPU time. Thus, it is feasible to run realistic scenarios, such as tens of thousands of infections among a susceptible population of hundreds of thousands of people for a duration of a year, in under a minute on a personal laptop.

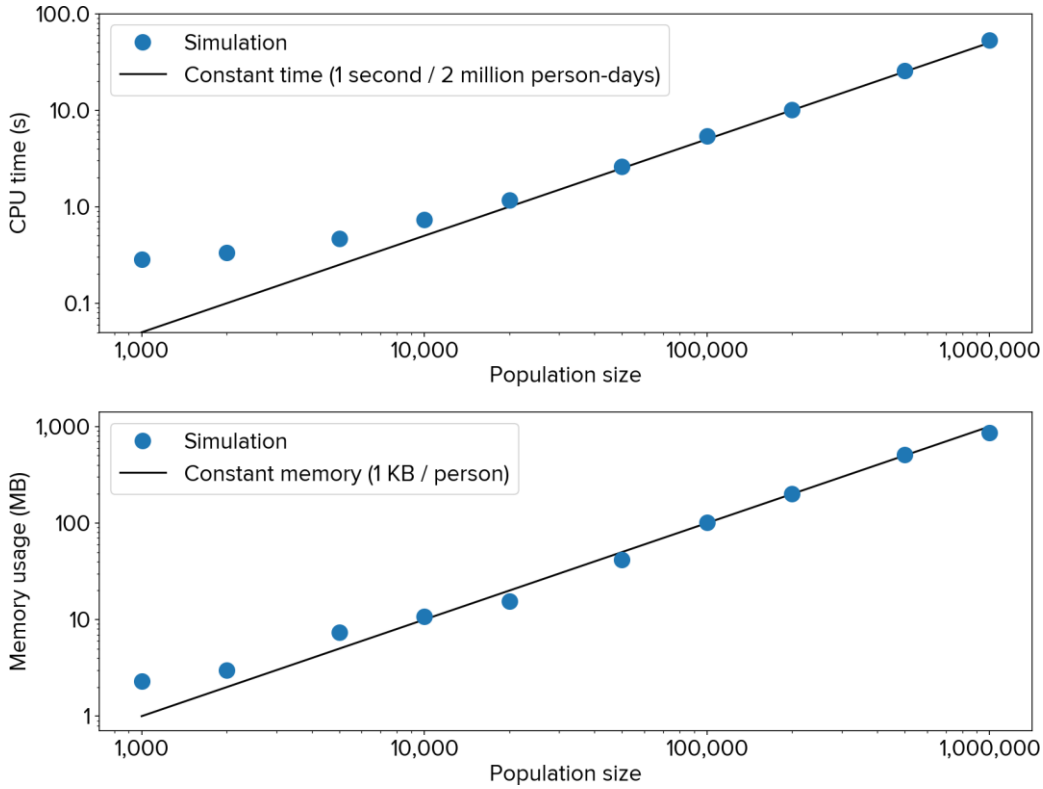

Figure S6. Covasim performance in terms of processor usage (top) and memory usage (bottom), showing linear scaling over almost three orders of magnitude of population size.

2.7.1 Deployment

For ease of use, a simple webapp for Covasim has been developed, based on Vue.js (for the frontend), ScirisWeb (for communicating between the frontend and the backend), Flask (for running the backend), and Gunicorn/NGINX (for running the server); this webapp is available at app.covasim.org. A screenshot of the user interface is shown in Fig. S7. A pre-built version of Covasim, including the webapp, is also available on Docker Hub (hub.docker.com).

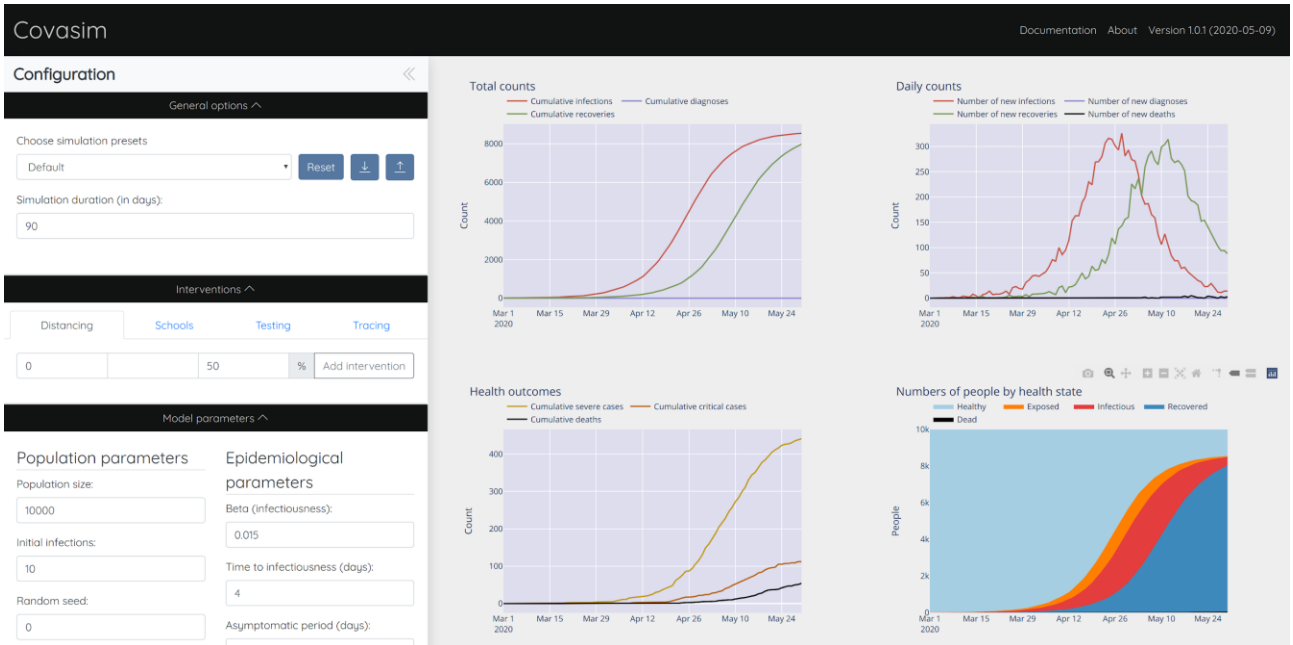

Figure S7. Covasim webapp user interface; screenshot taken from app.covasim.org.

2.7.2 Software tests

Covasim includes an extensive suite of both integration tests and unit tests; code coverage for version 1.0 is 94% (including compiled Numba functions), with much of the remaining 6% consisting of exceptions that are not raised by standard usage. In addition, outputs from the default simulations for each version are compared against cached values in the repository; since random seeds are stored, results are exactly reproducible despite the stochasticity in the model. When new data become available and parameter values are updated, previous parameters are stored, ensuring that any changes affecting the model outputs are intentional, and that previous versions can be easily retrieved and compared against.

3 Example usage

Several of Covasim's standard features are illustrated in Fig. S8. It represents a "calibrated" simulation (in terms of using a customized value of β) of 200,000 people, from February 10th until June 29th, starting with 75 seed infections. After an initial 45 days of uncontrolled epidemic spread, the following interventions are applied: March 26th, close schools and reduce work and community contacts to 70% of their original values; April 10th, reduce work and community to 30% of their original values; May 5th, reopen work and community to 80% of their original values; May 20th, begin testing 10% of people with COVID-like illness each day, and trace the contacts of people who test positive.

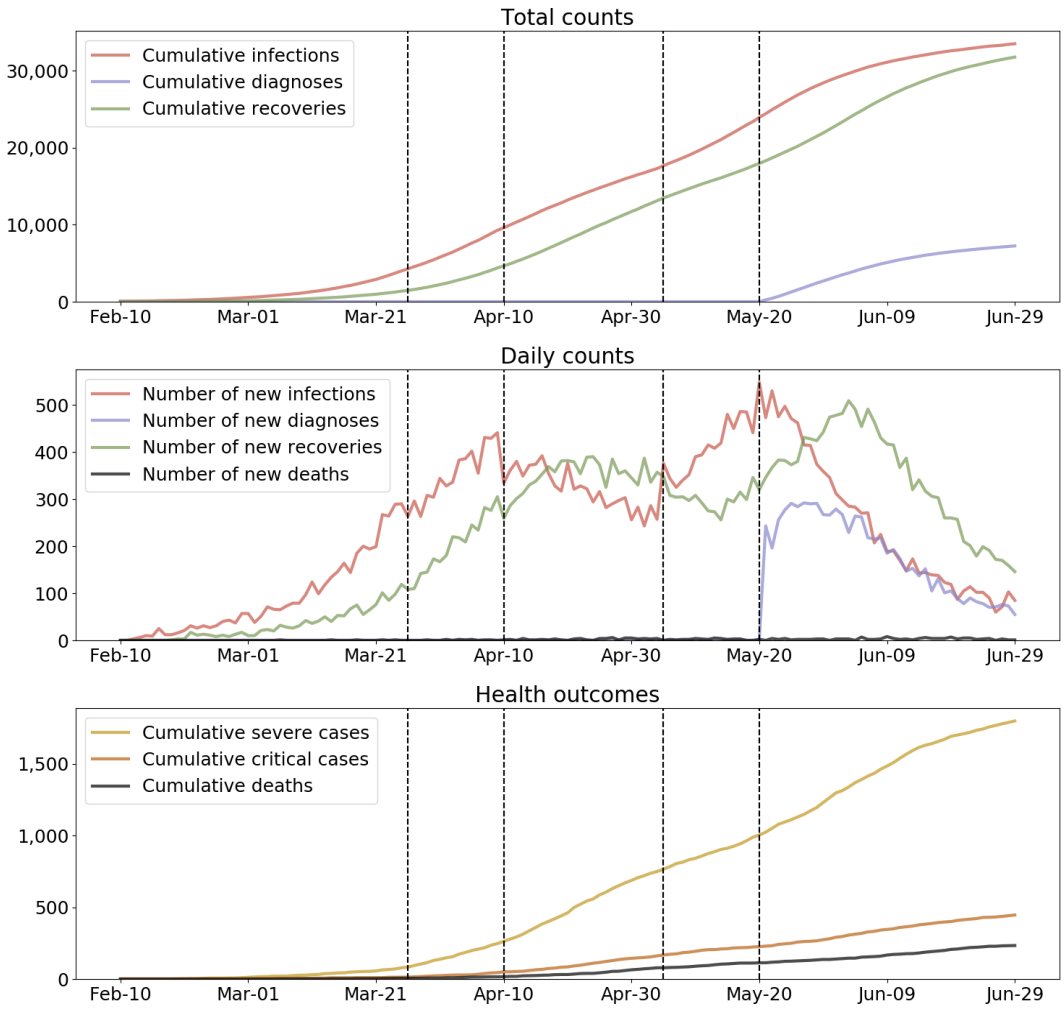

Figure S8. Illustrative example of a single run of a Covasim simulation. Interventions (described in the text) are shown as dashed vertical lines.

```python
1 import covasim as cv
2
3 # Define the simulation parameters
4 pars = dict(
5     pop_size     = 200e3,        # Define a population size of 200,000 people
6     pop_infected = 75,           # Start with 75 infected individuals
7     beta         = 0.012,        # Calibrate overall transmission to this setting
8     pop_type     = 'hybrid',     # Use realistic household, school, and work contacts
9     start_day    = '2020-02-10', # First day of the simulation
10    end_day      = '2020-06-29', # Last day of the simulation
11 )
12
13 # Define the interventions
14 trace_probs = dict(h=0.9, s=0.7, w=0.7, c=0.3) # Probability that a contact in each layer will be traced
15 trace_time = dict(h=0, s=1, w=1, c=3)          # Time required to trace contacts in each layer
16 interventions = [
17     cv.clip_edges(start_day='2020-03-26', end_day=None,         change={'s':0.0}),             # Close schools
18     cv.clip_edges(start_day='2020-03-26', end_day='2020-04-10', change={'w':0.7, 'c':0.7}),    # Reduce work and community
19     cv.clip_edges(start_day='2020-04-10', end_day='2020-05-05', change={'w':0.3, 'c':0.3}),    # Reduce both further
20     cv.clip_edges(start_day='2020-05-05', end_day=None,         change={'w':0.8, 'c':0.8}),    # Partially reopen
21     cv.test_prob(start_day='2020-05-20', symp_prob=0.10, symp_quar_prob=0.8, test_delay=2),    # Testing
22     cv.contact_tracing(start_day='2020-04-20', trace_probs=trace_probs, trace_time=trace_time) # Contact tracing
23 ]
24
25 # Create and run the simulation
26 sim = cv.Sim(pars=pars, interventions=interventions)
27 sim.run()
28 sim.plot()
```

Figure S9. Full listing of the code used to produce Fig. S8, including defining the parameters of the simulation (lines 4-11); defining the interventions (lines 14-23); and creating, running, and plotting the simulation (lines 26-28).

By default, Covasim shows time series for key cumulative counts, daily counts, and health outcomes (including deaths). All plotting outputs are configurable, and results can also be saved in Excel, JSON, or NumPy formats for further processing. While a full Covasim application would likely include additional complexity regarding calibration and plotting, other aspects of the example shown in Fig. 8 are comparable to a real-world exploratory policy analysis. Despite this, the Python script used to generate Fig. S8 is only 28 lines; this code is listed in Fig. S9.

In addition to running single simulations, Covasim also allows the user to run multiple simulations, which can be averaged over to determine forecast intervals. By default, the forecast intervals used correspond to the 10th and 90th percentiles of the simulated trajectories. In most contexts, these forecast intervals can be treated interchangeably with confidence intervals; however, since they are typically produced through a combination of stochastic variability and parameter uncertainty, they have a somewhat different statistical interpretation.

The preceding examples illustrate some aspects of Covasim's core functionality that are used in most applications. More in-depth analyses are also possible, leveraging either the default outputs, or the fact that the full state of the model is accessible to the user at every timestep via custom analysis functions. For example, detailed information about the transmission tree is stored for each simulation. This information can be used to determine the detailed microstructure of the infection patterns in a given simulation. Complete transmission trees for a small network under three different intervention scenarios are shown in Fig. S10, visualized via the ETE Toolkit (*67*). For realistically sized networks, it is not feasible to visualize entire transmission trees. However, their statistical properties can be analyzed to determine transmission routes and potential intervention targets. For example, such information can be used to determine the net contribution of schools (or even teachers at schools) to the overall epidemic trajectory.

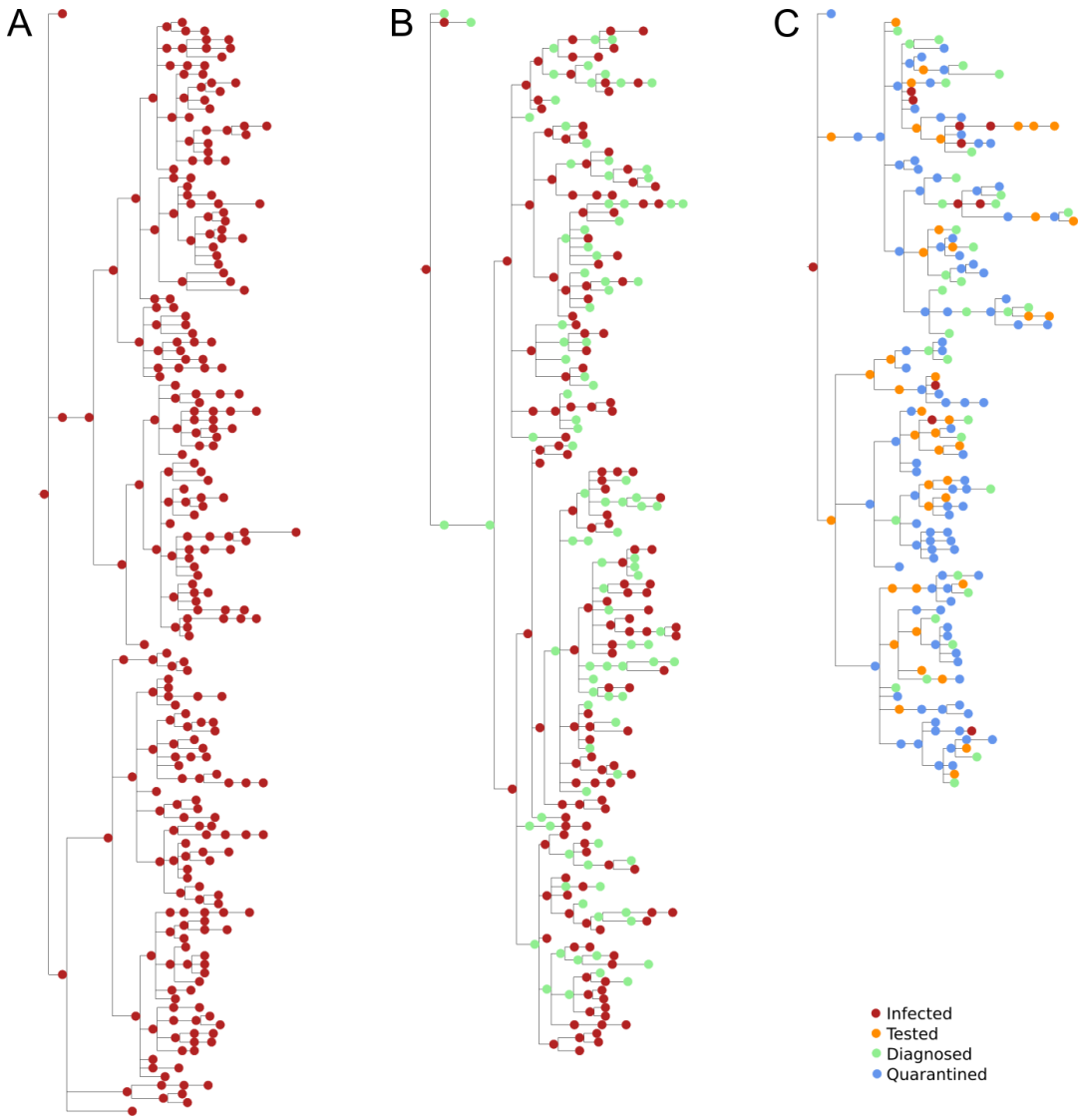

Figure S10. Example transmission trees for a hypothetical population of 300 individuals with a single seed infection on day 1, with (A) no interventions, (B) testing only, and (C) testing plus contact tracing. Time is shown on the horizontal axis, with each tree representing approximately 90 days. The vertical size of each tree is proportional to the total number of infections.

4 Limitations

Covasim is subject to the usual limitations of mathematical models, most notably constraints around the degree of realism that can be captured. For example, human contact patterns are intractably complex, and the algorithms that Covasim uses to approximate these are necessarily quite simplified.

Like all models, the quality of the outputs depends on the quality of the inputs, and many of the parameters on which Covasim relies are still subject to large uncertainties. Most critically, the proportion of asymptomatics and their relative transmission intensity, and the proportion of presymptomatic transmission, strongly affect the number of tests required in order to achieve workable COVID-19 suppression via testing-based interventions.

Dynamical models are commonly validated by comparing their projections against data on what actually happened. However, there are several challenges in using this approach for COVID-19, including (a) data quality issues (such as low case detection rates and under-reporting of deaths), (b) the difficulty of predicting future social and political responses that would significantly impact model projections (such as the timing of school and workplace reopening), and (c) the fact that model-based projections themselves have the potential to influence policy decisions, e.g., optimistic model projections may lead to relaxed policies, which in turn will lead to worse outcomes than predicted; pessimistic model projections may lead to stricter policies, which in turn will lead to better outcomes than predicted.

References

1. B. F. Maier, D. Brockmann, Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. *Science*. **368**, 742–746 (2020).
2. O. Coibion, Y. Gorodnichenko, M. Weber, "The Cost of the Covid-19 Crisis: Lockdowns, Macroeconomic Expectations, and Consumer Spending," *Working Paper Series* (Working Paper 27141, National Bureau of Economic Research, 2020), doi:10.3386/w27141.
3. A. J. Kucharski, P. Klepac, A. J. K. Conlan, S. M. Kissler, M. L. Tang, H. Fry, J. R. Gog, W. J. Edmunds, J. C. Emery, G. Medley, J. D. Munday, T. W. Russell, Q. J. Leclerc, C. Diamond, S. R. Procter, A. Gimma, F. Y. Sun, H. P. Gibbs, A. Rosello, K. van Zandvoort, S. Hué, S. R. Meakin, A. K. Deol, G. Knight, T. Jombart, A. M. Foss, N. I. Bosse, K. E. Atkins, B. J. Quilty, R. Lowe, K. Prem, S. Flasche, C. A. B. Pearson, R. M. G. J. Houben, E. S. Nightingale, A. Endo, D. C. Tully, Y. Liu, J. Villabona-Arenas, K. O'Reilly, S. Funk, R. M. Eggo, M. Jit, E. M. Rees, J. Hellewell, S. Clifford, C. I. Jarvis, S. Abbott, M. Auzenbergs, N. G. Davies, D. Simons, Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of SARS-CoV-2 in different settings: a mathematical modelling study. *The Lancet Infectious Diseases*. **0** (2020), doi:10.1016/S1473-3099(20)30457-6.
4. J. Hellewell, S. Abbott, A. Gimma, N. I. Bosse, C. I. Jarvis, T. W. Russell, J. D. Munday, A. J. Kucharski, W. J. Edmunds, F. Sun, S. Flasche, B. J. Quilty, N. Davies, Y. Liu, S. Clifford, P. Klepac, M. Jit, C. Diamond, H. Gibbs, K. van Zandvoort, S. Funk, R. M. Eggo, Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *The Lancet Global Health*. **8**, e488–e496 (2020).
5. M. E. Kretzschmar, G. Rozhnova, M. E. van Boven, Isolation and contact tracing can tip the scale to containment of COVID-19 in populations with social distancing. (2020), doi:10.2139/ssrn.3551343.
6. E. Gibney, Whose coronavirus strategy worked best? Scientists hunt most effective policies. *Nature*. **581**, 15–16 (2020).
7. S. Knapton, At least 11 countries have reimposed restrictions amid fears of coronavirus second wave. *The Telegraph* (2020), (available at https://www.telegraph.co.uk/news/2020/06/04/least-11-countries-have-reimposed-restrictions-amid-fears-coronavirus/).
8. S. Griffin, Covid-19: Lack of test and trace data is frustrating government scrutiny. *BMJ*. **369** (2020), doi:10.1136/bmj.m2239.
9. M. L. Holshue, C. DeBolt, S. Lindquist, K. H. Lofy, J. Wiesman, H. Bruce, C. Spitters, K. Ericson, S. Wilkerson, A. Tural, G. Diaz, A. Cohn, L. Fox, A. Patel, S. I. Gerber, L. Kim, S. Tong, X. Lu, S. Lindstrom, M. A. Pallansch, W. C. Weldon, H. M. Biggs, T. M. Uyeki, S. K. Pillai, First Case of 2019 Novel Coronavirus in the United States. *New England Journal of Medicine*. **382**, 929–936 (2020).

10. CDC, Washington State Report First COVID-19 Death | CDC Online Newsroom | CDC (2020), (available at https://www.cdc.gov/media/releases/2020/s0229-COVID-19-first-death.html).
11. A. Lasry, Timing of Community Mitigation and Changes in Reported COVID-19 and Community Mobility — Four U.S. Metropolitan Areas, February 26–April 1, 2020. *MMWR Morb Mortal Wkly Rep.* **69** (2020), doi:10.15585/mmwr.mm6915e2.
12. T. Akiba, S. Sano, T. Yanase, T. Ohta, M. Koyama, Optuna: A Next-generation Hyperparameter Optimization Framework, in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (Association for Computing Machinery, Anchorage, AK, USA, 2019; https://doi.org/10.1145/3292500.3330701), *KDD '19*, pp. 2623–2631.
13. N. Thakkar, G. Huynh, R. Etzioni, I. Painter, J. Lavista Ferres, M. Famulare, Situation Report 7: COVID-19 transmission across Washington State. *Institute for Disease Modeling. https://covid.idmod.org/data/WA_Situation_Report_7_COVID-19_transmission_across_Washington_State.pdf* (2020).
14. A. Kapteyn, M. Angrisani, D. Bennett, W. B. de Bruin, J. Darling, T. Gutsche, Y. Liu, E. Meijer, F. Perez-Arce, S. Schaner, K. Thomas, B. Weerman, Tracking the Effect of the COVID-19 Pandemic on the Lives of American Households. *Survey Research Methods.* **14**, 179–186 (2020).
15. H.-Y. Cheng, S.-W. Jian, D.-P. Liu, T.-C. Ng, W.-T. Huang, H.-H. Lin, Contact Tracing Assessment of COVID-19 Transmission Dynamics in Taiwan and Risk at Different Exposure Periods Before and After Symptom Onset. *JAMA Intern Med* (2020), doi:10.1001/jamainternmed.2020.2020.
16. Q.-L. Jing, M.-J. Liu, Z.-B. Zhang, L.-Q. Fang, J. Yuan, A.-R. Zhang, N. E. Dean, L. Luo, M.-M. Ma, I. Longini, E. Kenah, Y. Lu, Y. Ma, N. Jalali, Z.-C. Yang, Y. Yang, Household secondary attack rate of COVID-19 and associated determinants in Guangzhou, China: a retrospective cohort study. *The Lancet Infectious Diseases.* (2020), doi:10.1016/S1473-3099(20)30471-0.
17. X. He, E. H. Y. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan, X. Tan, X. Mo, Y. Chen, B. Liao, W. Chen, F. Hu, Q. Zhang, M. Zhong, Y. Wu, L. Zhao, F. Zhang, B. J. Cowling, F. Li, G. M. Leung, Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine.* **26**, 672–675 (2020).
18. L. Ferretti, C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, D. Bonsall, C. Fraser, Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science.* **368** (2020), doi:10.1126/science.abb6936.
19. C. J. Wang, C. Y. Ng, R. H. Brook, Response to COVID-19 in Taiwan: Big Data Analytics, New Technology, and Proactive Testing. *JAMA.* **323**, 1341–1342 (2020).
20. T. Hale, A. Petherick, T. Phillips, S. Webster, Variation in government responses to COVID-19. *Blavatnik school of government working paper.* **31** (2020).
21. Census profile: Seattle-Tacoma-Bellevue, WA Metro Area. *Census Reporter* (available at http://censusreporter.org/profiles/31000US42660-seattle-tacoma-bellevue-wa-metro-area/).
22. U. C. Bureau, America's Families and Living Arrangements: 2019. *The United States Census Bureau*, (available at https://www.census.gov/data/tables/2019/demo/families/cps-2019.html).
23. K. Prem, A. R. Cook, M. Jit, Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLoS Comput. Biol.* **13**, e1005697 (2017).
24. Data Portal OSPI (available at https://www.k12.wa.us/data-reporting/data-portal).
25. Home - Washington State Report Card, (available at washingtonstatereportcard.ospi.k12.wa.us).
26. Census - Table Results (available at https://data.census.gov/cedsci/table).
27. U. C. Bureau, Custom Tabulations. *The United States Census Bureau* (available at https://www.census.gov/programs-surveys/susb/data/custom-tabulations.html).

28. T. M. McMichael, D. W. Currie, S. Clark, S. Pogosjans, M. Kay, N. G. Schwartz, J. Lewis, A. Baer, V. Kawakami, M. D. Lukoff, J. Ferro, C. Brostrom-Smith, T. D. Rea, M. R. Sayre, F. X. Riedo, D. Russell, B. Hiatt, P. Montgomery, A. K. Rao, E. J. Chow, F. Tobolowsky, M. J. Hughes, A. C. Bardossy, L. P. Oakley, J. R. Jacobs, N. D. Stone, S. C. Reddy, J. A. Jernigan, M. A. Honein, T. A. Clark, J. S. Duchin, Epidemiology of Covid-19 in a Long-Term Care Facility in King County, Washington. *New England Journal of Medicine* (2020), doi:10.1056/NEJMoa2005412.
29. Long-Term Care Providers and Services Users in the United States--State Estimates Supplement: National Study of Long-Term Care Providers, 2015-2016, 155.
30. J. S. Bergstra, R. Bardenet, Y. Bengio, B. Kégl, Algorithms for Hyper-Parameter Optimization, in *Advances in neural information processing systems* (2011), pp. 2546–2554.
31. P. G. T. Walker, C. Whittaker, O. Watson, M. Baguelin, K. E. C. Ainslie, S. Bhatia, The Global Impact of COVID-19 and Strategies for Mitigation and Suppression. *Science* (2020), doi: 10.1126/science.abc0035.
32. J. M. Read, J. R. Bridgen, D. A. Cummings, A. Ho, C. P. Jewell, Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. *medRxiv*, in press, doi:10.1101/2020.01.23.20018549.
33. M. J. Keeling, T. D. Hollingsworth, J. M. Read, Efficacy of contact tracing for the containment of the 2019 novel coronavirus (COVID-19). *J Epidemiol Community Health* (2020), doi:10.1136/jech-2020-214051.
34. G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo, M. Colaneri, Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature Medicine*, 1–6 (2020).
35. S. Zhao, H. Chen, Modeling the epidemic dynamics and control of COVID-19 outbreak in China. *Quant Biol*, 1–9 (2020).
36. S. L. Chang, N. Harding, C. Zachreson, O. M. Cliff, M. Prokopenko, Modelling transmission and control of the COVID-19 pandemic in Australia. *arXiv:2003.10218* (2020) (available at http://arxiv.org/abs/2003.10218).
37. J. R. Koo, A. R. Cook, M. Park, Y. Sun, H. Sun, J. T. Lim, C. Tam, B. L. Dickens, Interventions to mitigate early spread of SARS-CoV-2 in Singapore: a modelling study. *The Lancet Infectious Diseases.* **0** (2020), doi:10.1016/S1473-3099(20)30162-6.
38. D. L. Chao, A. P. Oron, D. Srikrishna, M. Famulare, Modeling layered non-pharmaceutical interventions against SARS-CoV-2 in the United States with Corvid. *medRxiv*, in press, doi:10.1101/2020.04.08.20058487.
39. N. M. Ferguson, D. Laydon, G. Nedjati-Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunubá, G. Cuomo-Dannenburg, Others, Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. *London: Imperial College COVID-19 Response Team, March.* **16** (2020).
40. A. Aleta, D. Martin-Corral, A. P. y Piontti, M. Ajelli, M. Litvinova, M. Chinazzi, N. E. Dean, M. Elizabeth Halloran, I. M. Longini, S. Merler, A. Pentland, A. Vespignani, E. Moro, Y. Moreno, Modeling the impact of social distancing, testing, contact tracing and household quarantine on second-wave scenarios of the COVID-19 epidemic. *medRxiv* (2020), doi: 10.1101/2020.05.06.20092841.
41. C. M. Peak, L. M. Childs, Y. H. Grad, C. O. Buckee, Comparing nonpharmaceutical interventions for containing emerging epidemics. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 4023–4028 (2017).
42. Q. Bi, Y. Wu, S. Mei, C. Ye, X. Zou, Z. Zhang, X. Liu, L. Wei, S. A. Truelove, T. Zhang, W. Gao, C. Cheng, X. Tang, X. Wu, Y. Wu, B. Sun, S. Huang, Y. Sun, J. Zhang, T. Ma, J. Lessler, T. Feng,

Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *The Lancet Infectious Diseases* (2020), doi:10.1016/S1473-3099(20)30287-5.

43. Y. Yang, M. Yang, C. Shen, F. Wang, J. Yuan, J. Li, M. Zhang, Z. Wang, L. Xing, J. Wei, L. Peng, G. Wong, H. Zheng, M. Liao, K. Feng, J. Li, Q. Yang, J. Zhao, Z. Zhang, L. Liu, Y. Liu, Evaluating the accuracy of different respiratory specimens in the laboratory diagnosis and monitoring the viral shedding of 2019-nCoV infections. *medRxiv*, in press, doi:10.1101/2020.02.11.20021493.
44. S. A. Lauer, K. H. Grantz, Q. Bi, F. K. Jones, Q. Zheng, H. R. Meredith, A. S. Azman, N. G. Reich, J. Lessler, The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Ann Intern Med* (2020), doi:10.7326/m20-0504.
45. Z. Du, X. Xu, Y. Wu, L. Wang, B. J. Cowling, L. A. Meyers, Serial Interval of COVID-19 among Publicly Reported Confirmed Cases. *Emerg Infect Dis.* **26** (2020), doi:10.3201/eid2606.200357.
46. H. Nishiura, N. M. Linton, A. R. Akhmetzhanov, Serial interval of novel coronavirus (COVID-19) infections. *Int J Infect Dis.* **93** (2020), pp. 284–286.
47. R. Pung, C. J. Chiew, B. E. Young, S. Chin, M. I. C. Chen, H. E. Clapham, A. R. Cook, S. Maurer-Stroh, M. P. H. S. Toh, C. Poh, M. Low, J. Lum, V. T. J. Koh, T. M. Mak, L. Cui, R. V. T. P. Lin, D. Heng, Y.-S. Leo, D. C. Lye, V. J. M. Lee, K. Kam, S. Kalimuddin, S. Y. Tan, J. Loh, K. C. Thoon, S. Vasoo, W. X. Khong, N.-A. Suhaimi, S. J. H. Chan, E. Zhang, O. Oh, A. Ty, C. Tow, Y. X. Chua, W. L. Chaw, Y. Ng, F. Abdul-Rahman, S. Sahib, Z. Zhao, C. Tang, C. Low, E. H. Goh, G. Lim, Y. Hou, I. Roshan, J. Tan, K. Foo, K. Nandar, L. Kurupatham, P. P. Chan, P. Raj, Y. Lin, Z. Said, A. Lee, C. See, J. Markose, J. Tan, G. Chan, W. See, X. Peh, V. Cai, W. K. Chen, Z. Li, R. Soo, A. L. P. Chow, W. Wei, A. Farwin, L. W. Ang, Investigation of three clusters of COVID-19 in Singapore: implications for surveillance and response measures. *The Lancet.* **395** (2020), pp. 1039–1046.
48. N. M. Linton, T. Kobayashi, Y. Yang, K. Hayashi, A. R. Akhmetzhanov, S.-M. Jung, B. Yuan, R. Kinoshita, H. Nishiura, Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data. *Journal of Clinical Medicine* **9** (2020), p. 538.
49. R. Wölfel, V. M. Corman, W. Guggemos, M. Seilmaier, S. Zange, M. A. Müller, D. Niemeyer, T. C. Jones, P. Vollmar, C. Rothe, M. Hoelscher, T. Bleicker, S. Brünink, J. Schneider, R. Ehmann, K. Zwirglmaier, C. Drosten, C. Wendtner, Virological assessment of hospitalized patients with COVID-2019. *Nature* (2020), doi:10.1038/s41586-020-2196-x.
50. R. Verity, L. C. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai, G. Cuomo-Dannenburg, H. Thompson, P. G. T. Walker, H. Fu, A. Dighe, J. T. Griffin, M. Baguelin, S. Bhatia, A. Boonyasiri, A. Cori, Z. Cucunubá, R. FitzJohn, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, D. Laydon, G. Nedjati-Gilani, S. Riley, S. van Elsland, E. Volz, H. Wang, Y. Wang, X. Xi, C. A. Donnelly, A. C. Ghani, N. M. Ferguson, Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet Infectious Diseases* (2020), , doi:10.1016/S1473-3099(20)30243-7.
51. J. Zhang, M. Litvinova, Y. Liang, Y. Wang, W. Wang, S. Zhao, Q. Wu, S. Merler, C. Viboud, A. Vespignani, M. Ajelli, H. Yu, Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science* **368** (2020), pp. 1481–1486, doi: 10.1126/science.abb8001.
52. D. He, S. Zhao, Q. Lin, Z. Zhuang, P. Cao, M. H. Wang, L. Yang, The relative transmissibility of asymptomatic COVID-19 infections among close contacts. *International Journal of Infectious Diseases.* **94** (2020), pp. 145–147.
53. F.-X. Lescure, L. Bouadma, D. Nguyen, M. Parisey, P.-H. Wicky, S. Behillil, A. Gaymard, M. Bouscambert-Duchamp, F. Donati, Q. Le Hingrat, V. Enouf, N. Houhou-Fidouh, M. Valette, A. Mailles, J.-C. Lucet, F. Mentre, X. Duval, D. Descamps, D. Malvy, J.-F. Timsit, B. Lina, S. van-der-

Werf, Y. Yazdanpanah, Clinical and virological data of the first cases of COVID-19 in Europe: a case series. *Lancet Infect. Dis.* (2020).

54. K. K.-W. To, O. T.-Y. Tsang, W.-S. Leung, A. R. Tam, T.-C. Wu, D. C. Lung, C. C.-Y. Yip, J.-P. Cai, J. M.-C. Chan, T. S.-H. Chik, D. P.-L. Lau, C. Y.-C. Choi, L.-L. Chen, W.-M. Chan, K.-H. Chan, J. D. Ip, A. C.-K. Ng, R. W.-S. Poon, C.-T. Luo, V. C.-C. Cheng, J. F.-W. Chan, I. F.-N. Hung, Z. Chen, H. Chen, K.-Y. Yuen, Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study. *Lancet Infect. Dis.* **20**, 565–574 (2020).
55. L. Zou, F. Ruan, M. Huang, L. Liang, H. Huang, Z. Hong, J. Yu, M. Kang, Y. Song, J. Xia, Q. Guo, T. Song, J. He, H.-L. Yen, M. Peiris, J. Wu, SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients. *N. Engl. J. Med.* **382**, 1177–1179 (2020).
56. A. Endo, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, S. Abbott, A. J. Kucharski, S. Funk, Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Res.* **5**, 67 (2020).
57. D. Miller, M. A. Martin, N. Harel, T. Kustin, O. Tirosh, M. Meir, N. Sorek, S. Gefen-Halevi, S. Amit, O. Vorontsov, D. Wolf, A. Peretz, Y. Shemer-Avni, D. Roif-Kaminsky, N. Kopelman, A. Huppert, K. Koelle, A. Stern, Full genome viral sequences inform patterns of SARS-CoV-2 spread into and within Israel. *medRxiv*, in press, doi:10.1101/2020.05.21.20104521.
58. L. Fumanelli, M. Ajelli, P. Manfredi, A. Vespignani, S. Merler, Inferring the Structure of Social Contacts from Demographic Data in the Analysis of Infectious Diseases Spread. *PLOS Computational Biology*. **8**, e1002673 (2012).
59. D. Mistry, M. Litvinova, A. P. y Piontti, M. Chinazzi, L. Fumanelli, M. F. C. Gomes, S. A. Haque, Q.-H. Liu, K. Mu, X. Xiong, M. E. Halloran, I. M. Longini Jr., S. Merler, M. Ajelli, A. Vespignani, Inferring high-resolution human mixing patterns for disease modeling. *arXiv:2003.01214* (2020) (available at http://arxiv.org/abs/2003.01214).
60. T. Smieszek, V. C. Barclay, I. Seeni, J. J. Rainey, H. Gao, A. Uzicanin, M. Salathé, How should social mixing be measured: comparing web-based survey and sensor-based methods. *BMC Infect. Dis.* **14**, 136 (2014).
61. J. Huisman, J. Smits, Effects of Household- and District-Level Factors on Primary School Enrollment in 30 Developing Countries. *World Development.* **37**, 179–193 (2009).
62. S. T. Bakir, Compound Interest Doubling Time Rule: Extensions and Examples from Antiquities. *Communications in Mathematical Finance* **5** (2016).
63. H. Barratt, M. Kirwan, Public Health Textbook. *Health Knowledge* (2010), (available at https://www.healthknowledge.org.uk/public-health-textbook).
64. C. C. Kerr, S. Dura-Bernal, T. G. Smolinski, G. L. Chadderdon, D. P. Wilson, Optimization by adaptive stochastic descent. *PLoS ONE* **13**, e0192944 (2018), doi:10.1371/journal.pone.0192944.
65. I. Andrianakis, I. R. Vernon, N. McCreesh, T. J. McKinley, J. E. Oakley, R. N. Nsubuga, M. Goldstein, R. G. White, Bayesian History Matching of Complex Infectious Disease Models Using Emulation: A Tutorial and a Case Study on HIV in Uganda. *PLoS Comput Biol* **11**, e1003968 (2015).
66. P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. et al., SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods.* **17**, 352 (2020).
67. J. Huerta-Cepas, F. Serra, P. Bork, ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. Mol Biol Evol. 33, 1635–1638 (2016).