Effectiveness of social distancing strategies for protecting a community from a pandemic with a data-driven contact network based on census and real-world mobility data

David Martín-Calvo[1,4*], Alberto Aleta[2,3,*], Alex Pentland[5], Yamir Moreno[2,3,&], Esteban Moro[1,5,&]

[1] Department of Mathematics and GISC, Universidad Carlos III de Madrid, Leganés, Spain.
[2] ISI Foundation, Turin, Italy
[3] Institute for Biocomputation and Physics of Complex Systems, University of Zaragoza, Zaragoza 50018, Spain
[4] Zensei Technologies S.L., Madrid, Spain.
[5] Connection Science, Institute for Data Science and society, MIT, Cambridge, US
[*] These authors contributed equally.
[&] Co-lead authors. To whom correspondence should be sent: yamir.moreno@gmail.com, esteban.moroegido@gmail.com

Key findings

1. School closures do not have a major impact on controlling the epidemic, despite closing them, infections keep occurring within the households and the community layers.
2. Passive social distance strategies are not enough to contain the epidemic, indicating that active strategies need to be established. For instance, large scale testing, remote symptoms monitoring, isolation and contact tracing.
3. School closures and self-distancing at 90% of adoption is a feasible strategy for minimizing the effects of the epidemic, but only if they are applied for a long period of time.
4. A full confinement is not feasible and will not solve the problem, without active measures in place after the confinement, since there would be a new outbreak.
5. If high resolution mobility data is available, our data-driven approach with real world data can be easily replicated for new cities or countries to measure the impact of social distance strategies and the epidemic.

Summary

The current situation of emergency is global. As of today, March 22nd 2020, there are more than 23 countries with more than 1.000 infected cases by COVID-19, in the exponential growth phase of the disease. Furthermore, there are different mitigation and suppression strategies in place worldwide, but many of them are based on enforcing, to a more or less extent, the so-called social distancing. The impact and outcomes of the adopted measures are yet to be contrasted and quantified. Therefore, realistic modeling approaches could provide important clues about what to expect and what could be the best course of actions. Such modeling efforts could potentially save thousands, if not millions of lives. Our report contains preliminary results that aim at answering the following questions in relation to the spread and control of the COVID-19 pandemic:

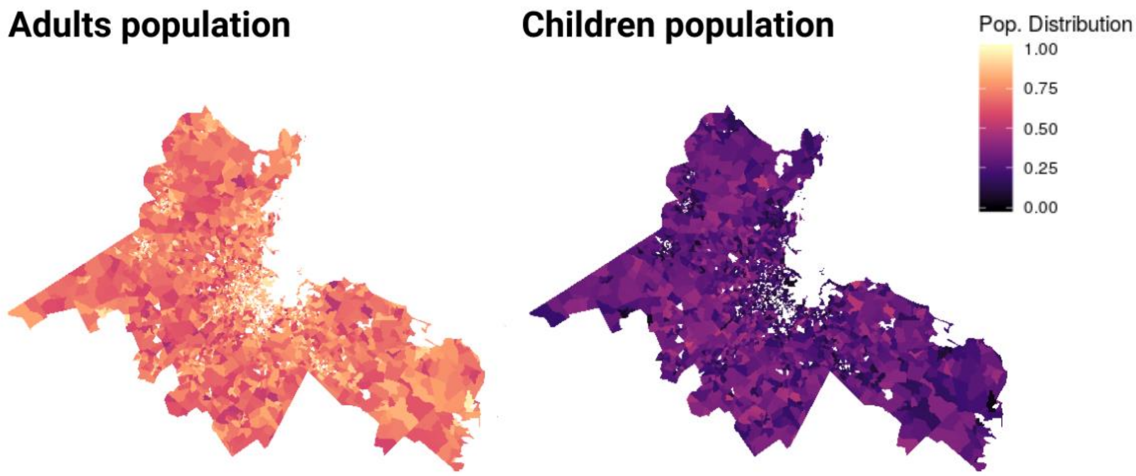
- What is the expected impact of current social distancing strategies?
- How long should such measures need to be in place?
- How many people will be infected and at which social level?
- How do R(t) and the epidemic dynamic change based on the adopted strategies?
- What is the probability of having a second outbreak, i.e., a reemergence?
- If there is a reemergence, how much time do we have to get ready?
- What is the best strategy to minimize the current epidemic and get ready for a second wave?

In this report, we provide details of the data analyzed, the methodology (and its limitations) employed as well as a quantitative and qualitative assessment of strategies based on social distancing and corresponding what-if-scenarios for control and mitigation. We use real world mobility and census data of the Boston area to build a co-location network at three different layers (community, households and schools), and a data-driven SEIR model that allows testing six different social distancing strategies, namely, (i) school closures, (ii) self-distancing and teleworking, (iii) self-distancing and teleworking plus School closure (iv) Restaurants, nightlife and cultural closures, (v) non-essential workplace closures and (vi) total confinement. We test the impact of establishing these strategies at different stages of the epidemic evolution and for different periods of time.

Data & Methods

Mobility and synthetic data for building the contact network

We used detailed mobility data and sociodemographic data from the Boston area, from Cuebiq and US Census, respectively. We use it to generate one networked subpopulation that describes the contact patterns of about 100.000 agents in the Boston area during a period of four months and three layers (Community, Households and Schools). The community layer is based on mobility data and measures exposure of individuals happening at different points of interests, from restaurants to work places; households are built using census data and, lastly, children belonging to the same geographical location are linked together to create the school layer. In the community layer, links are built using co-location data, with weights proportional to the time two individuals have been in the same place. Note that the data does not allow to disentangle if the contacts take place in the context of the workplace or in the general community, therefore, we have merged both layers.



The figure shows the distribution of adults and children in the population by US Census block groups in the Boston area. The Adult population comes from real world mobility data and children are synthetically generated with the help of the US Census data about Household type by size and Family type by presence of children under 18 years.

Calibrating layer weights using historical ILI data and a SIR model

Initially, we need to calibrate layer weights, as they refer to a different quantity depending on the layer. Links in households and schools have a weight equal to 1. Conversely, links in the community layer are a measure of time, since it is the product of the fraction of time that individuals i and j have been in the same place. To obtain the relative weights of these layers, we implement a SIR model to simulate the propagation of a typical ILI [1]. In particular:

- An infected individual *i* can transmit the disease to *j* with probability
$$\rho w_{ij} w_{\text{layer}}$$
- An infected individual recovers with probability 1/3

Then, we have four free parameters ($\rho, w_{\text{school}}, w_{\text{household}}, w_{\text{community}}$). We fit the model to a typical ILI scenario, namely, R0=1.3 and the fractions of infections that should take place in each setting are set to 0.18, 0.30 and 0.52 for schools, households and the community+workplaces, respectively. Note that this can be done because we can assume that ILI and COVID-19 transmit similarly via human to human contacts.

SEIR model for modelling the COVID-19 epidemic

Once the links of the different layers are calibrated, the subpopulation network is complete. Next , we adopt a SEIR model, to include an extra compartment that better suits the natural history of COVID-19:

- A susceptible individual *j* can be infected by an infected individual *i* if there is a link between them. This will happen with probability
$$\rho w_{ij} w_{\text{layer}}$$
Note that in this case w_{layer} is no longer a free parameter.
- If an individual is infected, it will move to the exposed compartment. Following [2] we consider that individuals in this state cannot infect, nor be infected again, up to 24h before their symptoms' onset, since several studies have shown that the serial interval is slightly smaller than the incubation period. The individual will leave the compartment after a certain number of days, sampled from a log-normal distribution with mean 5.2 [3].
- The probability of moving to the recovered compartment is set so that the generation time is 7.5 [4,5].
- After recovery individuals are assumed to be immune to becoming re-infected by the virus.

Once the model is set, we fit the only free parameter, rho. We choose its value such that R0=2.5 [4,5] (simulations with other values does not change the conclusions qualitatively).

Social distancing strategies tested

We explore against the baseline six different strategies:

I. **School closures** are simulated by removing all the schools from the system simultaneously;
II. **Self-distancing and teleworking** are simulated by taking a random fraction of the population and removing all their interactions in the community+workplace layer;
III. **Self-distancing and teleworking plus School closure.**
IV. **Restaurants, nightlife and cultural closures** are simulated by removing the interactions that take place in any place that falls into that category according to Foursquare's taxonomy of places;
V. **Non-essential workplace closures**, leaving open Hospitals, Salons, Barbershops, Grocery Stores, Dispensaries, Supermarkets, Pet Stores, Pharmacies, Urgent Care Centers, Dry Cleaners, Drugstores, Maternity Clinics, Medical Supplies and Gas Stations;
VI. **Total confinement** with school and non-essential workplace closures within the metropolitan area.

In all cases we test the impact of establishing these strategies at different stages of the epidemic curve and for different temporal duration.

What if scenarios for several social distancing strategies

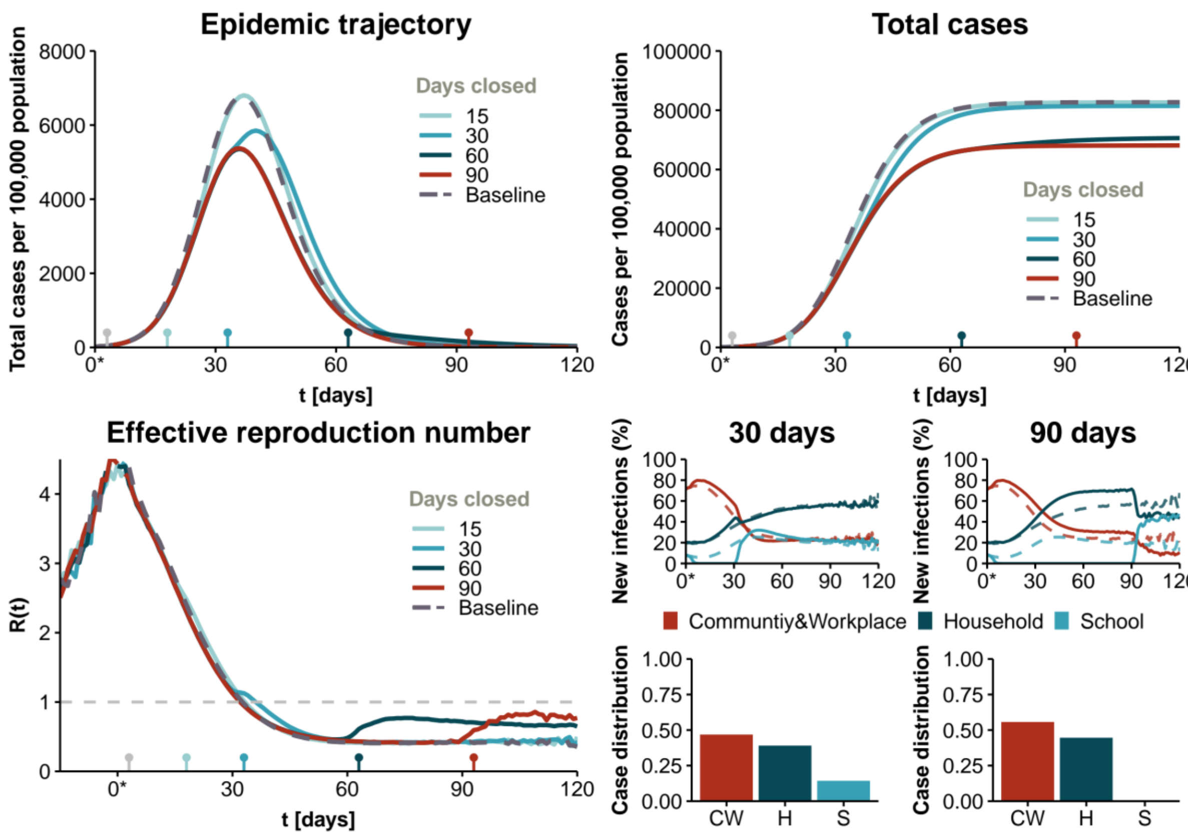
We have tested what would happen if the restrictions are applied after 100 cases have been detected and then removed after 15, 30, 60 and 90 days. The time when the restrictions are implemented will be denoted by a gray vertical line, while the time when they are lifted will be represented by a line matching the color of the legend. Besides, all simulations are aligned to the time when there are 50 cases in the population, this is denoted by an asterisk next to 0 in the plots.

In all cases we compute: the evolution of the number of infected individuals in the population; the evolution of the total number of cases; the effective reproduction number; how new infections are distributed across the different settings as a function of time and in total.

To obtain the effective reproduction number we keep track of how many individuals each person infects. Then, we associate this *R_i* to the time where the individual showed symptoms for the first time. That is, if someone shows symptoms at t=1 and infects someone at t=2 and t=3, we would have R(t=1) = 2, R(t=2) = 0, R(t=3) = 0.

1. School closures

We observe that the overall effect of the closure is not too high, especially if it is imposed for only a month or less. The spreading is characterized by one large peak, after which the epidemic is not able to sustain itself in the population. The effective reproduction number clearly increases once the restrictions are lifted, but there are not enough susceptible individuals in the population to go above 1 and cause a new outbreak.



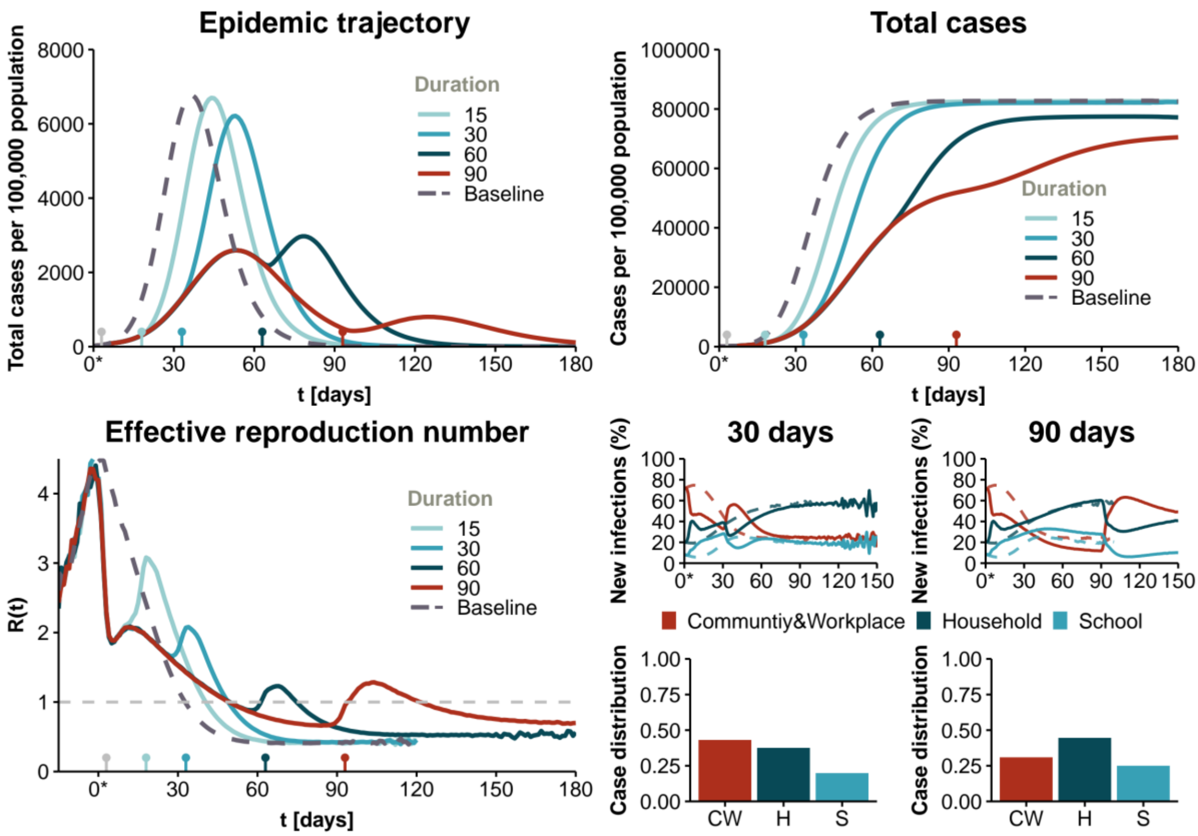
S1. School closure after 100 cases

2. Telework and social distancing

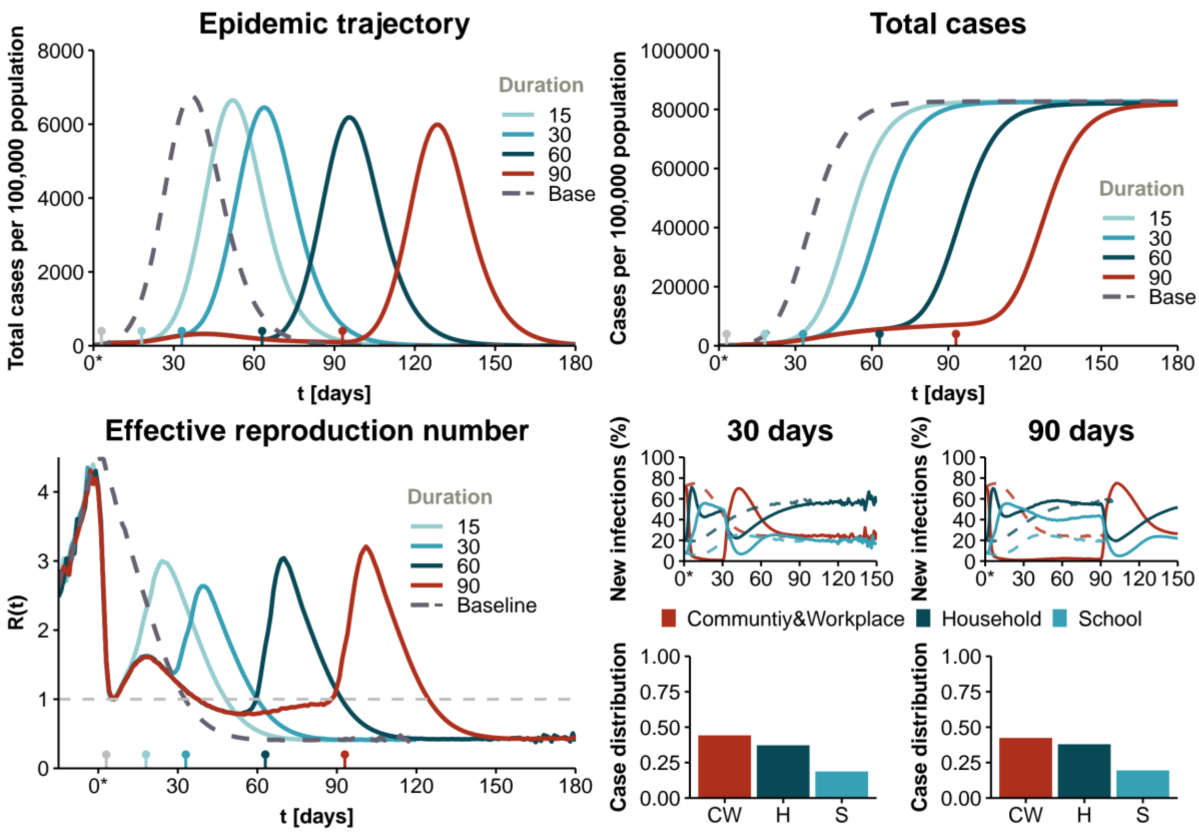
In this strategy, we flag all individuals that have at least one interaction in the community + workplace (CW) layer. Then, in each run, we select a random fraction (adopters) of this subset of individuals and remove their connections from the CW layer (for as long as the strategy is imposed). After that, they recover their connections.

When the number of adopters is relatively low (50% of the population with connections in the community layer), if the restrictions are lifted within a month there is a small delay but the total number of cases is the same. On the other hand, if the restrictions are applied for 2 months or more, there is a first medium-size peak, followed by a slow decay. Once the restrictions are lifted, the epidemic is able to spread again and produce a second peak of noticeable size, so that the total number of cases is quite close to the baseline scenario. The effective reproduction number clearly shows that even if it is below 1, if the restrictions are removed too early it can go again beyond that threshold and produce the second peak.

Conversely, when the fraction of adopters is very high (90%), we observe that the first peak is really small. Then, the disease might be able to spread for long periods of time with an effective reproduction number close to 1 through the school layer. Once the restrictions are lifted, since almost the whole population is susceptible, we recover a behavior equivalent to the baseline scenario.



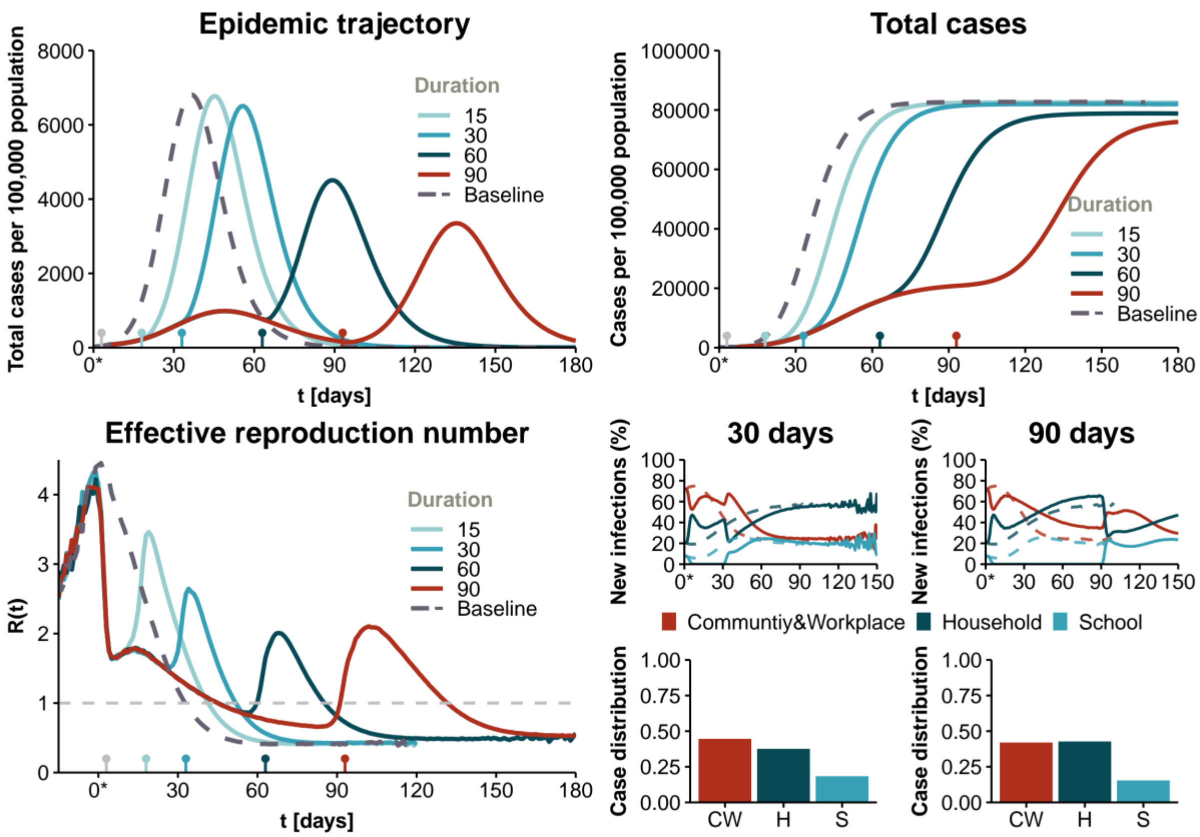
S2. Social distancing (50%) after 100 cases



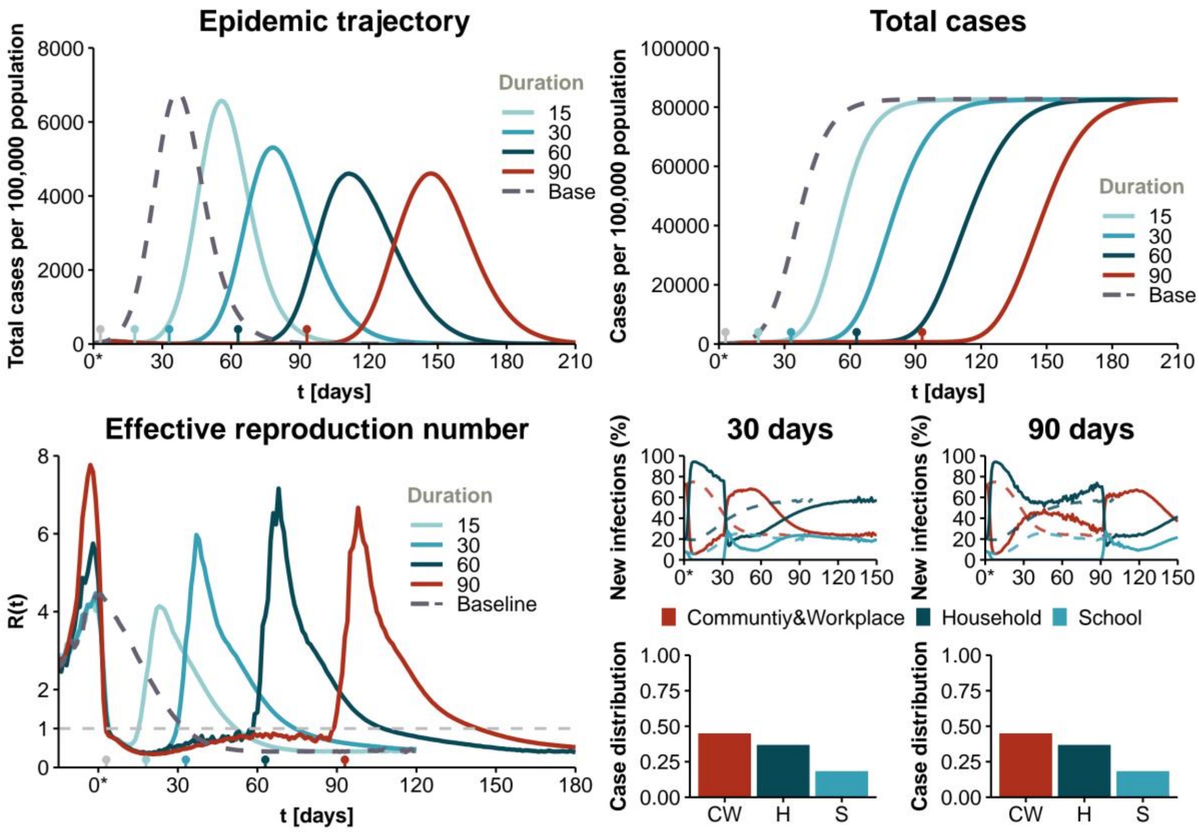
S2. Social distancing (90%) after 100 cases

3. Self-distancing plus School closure

Since we have observed that the epidemic is able to spread through schools even if the social distancing is very high, our next strategy is to also close schools. In this case, if the fraction of adopters is low the first peak is much lower, but then the second peak is much higher than in the previous case, resulting in an epidemic of roughly the same size. If the number of adopters is high, we note that in this case, even for restrictions of 3 months, the disease is able to spread through the few links that are present in the household and community layers.



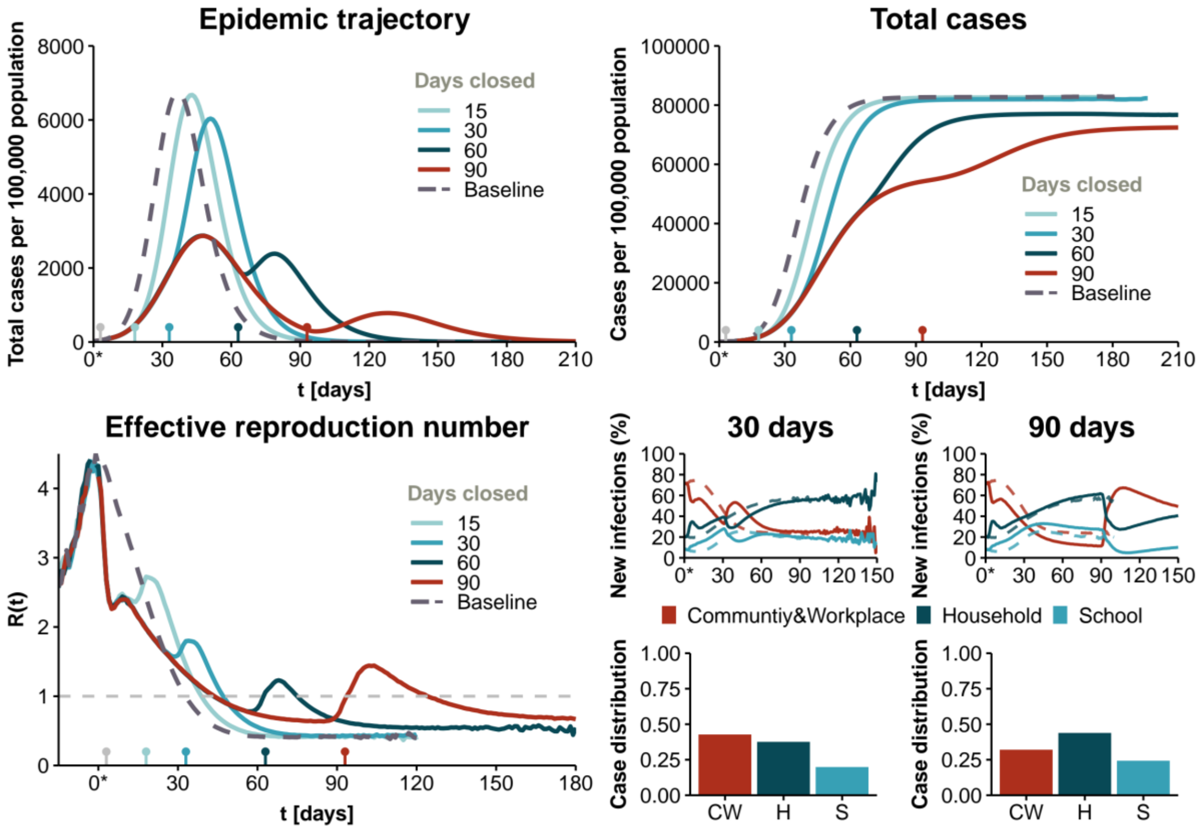
S3. Social distancing (50%) + School closure after 100 cases



S3. Social distancing (90%) + School closure after 100 cases

4. Restaurants, nightlife and cultural closures

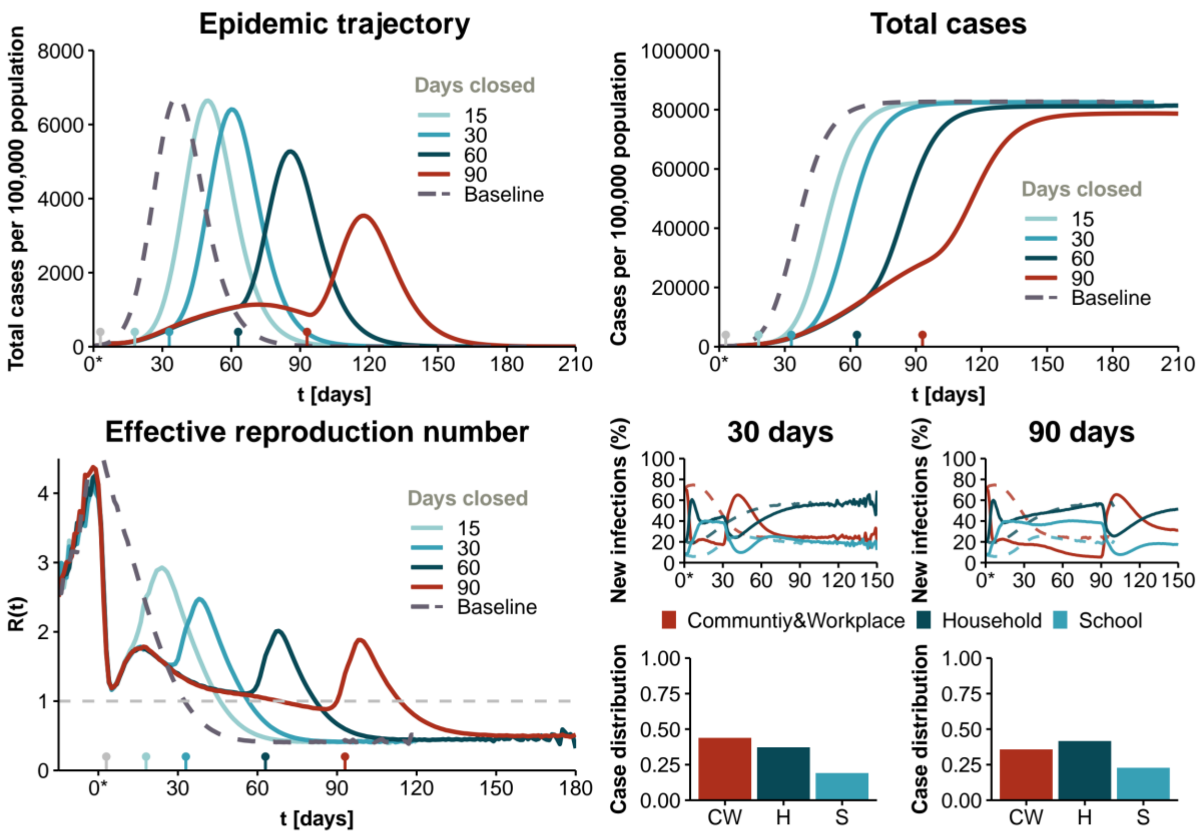
If restrictions are focused on a small set of public places such as restaurants, nightlife and cultural places, we obtain results compatible with strategy 2 with 50% adopters. The difference is that in this case the first peak is slightly larger, resulting in a smaller second peak.



S4. Restaurants and cultural closure after 100 cases

5. All non-essential places closed

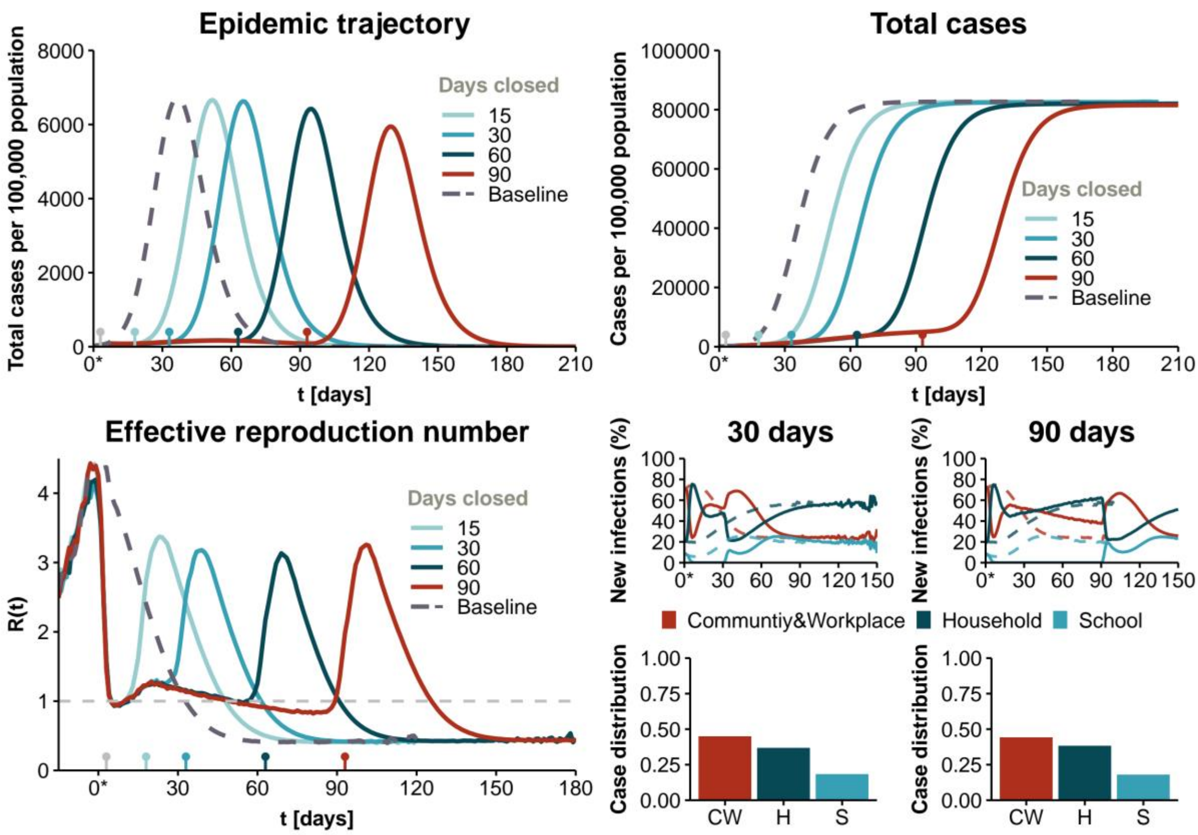
If the previous policy is extended to all non-essential places, we obtain results compatible with strategy 3 with a small fraction of adopters.



S5. Non–essential closure after 100 cases

6. Total confinement

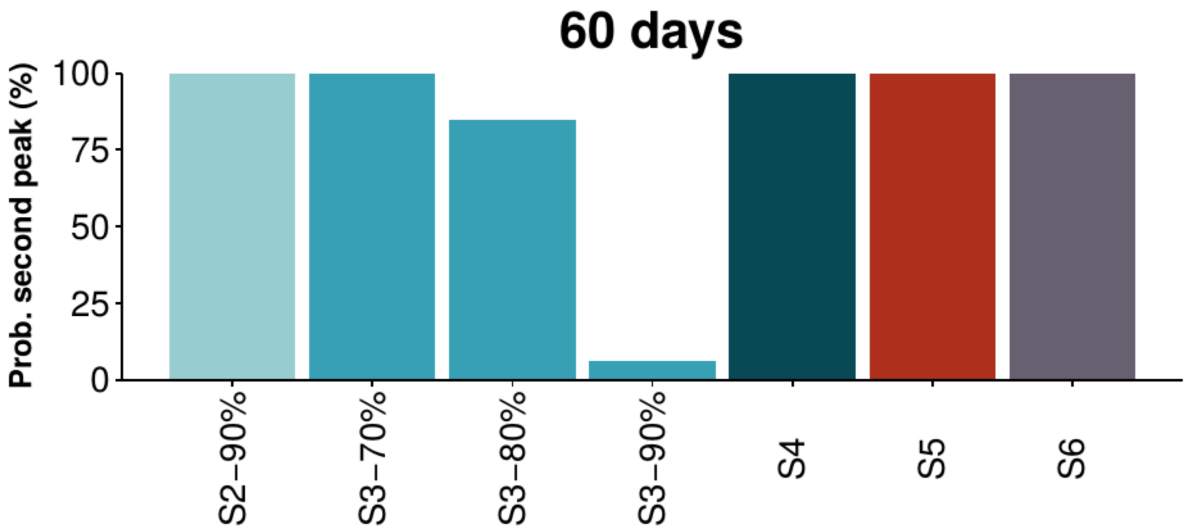
Lastly, we analyze the situation in which all non-essential places and schools are closed. As expected, the results are similar to strategy 3 with a large fraction of adopters.

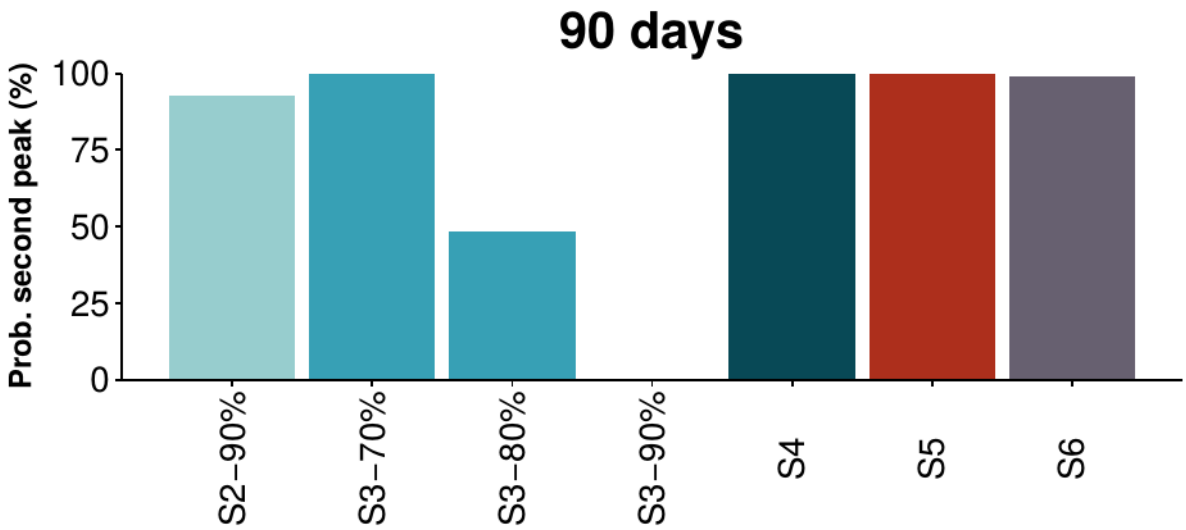


S6. Total non–essential closure after 100 cases

Second peak

In many cases we have seen that there is a second peak after the containment measures are lifted. In this plot we show the probability of having that peak. The best strategy will be the one that has a second peak, but with low probability so that with effective contact tracing and similar measures there is a low probability of having a second peak.





References

[1] Measurability of the epidemic reproduction number in data-driven contact networks, https://www.pnas.org/content/115/50/12680.abstract
[2] Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand
https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf
[3] Evolving epidemiology of novel coronavirus diseases 2019 and possible interruption of local transmission outside Hubei Province in China: a descriptive and modeling study
https://www.medrxiv.org/content/10.1101/2020.02.21.20026328v1
[4] Estimating the risk of sustained community transmission of COVID-19 outside Mainland China
https://www.mobs-lab.org/uploads/6/7/8/7/6787877/estimating_the_risk_of_sustained_community_transmission_of_covid-19_outside_china.pdf

[5] Expected impact of school closure and telework to mitigate COVID-19 epidemic in France

https://www.epicx-lab.com/uploads/9/6/9/4/9694133/inserm_covid-19-school-closure-french-regions_20200313.pdf

Limitations of the study

- Social mixing as given by age is not included. This might be important to evaluate CFT and other age-related quantities.
- Children are not present in the community layer. The impact of this limitation should not be important in most scenarios as they consider confinement at households.
- The mobility data is biased towards neighborhoods with a larger amount of family households than the average.
- Currently we do not distinguish between mild and critical cases.
- Uncertainties on the precise values of several epidemic parameters. This implies that the arrival of the peak or the duration of the restrictions cannot be taken as an exact quantity but an indication of the expected overall behavior. More simulations will be carried out to perform a sensitivity analysis.
- The population is closed. This limitation is not supposed to be severe as the underlying assumption is that inter-subpopulation mobility is reduced to minimal levels.
- We build schools by mixing children from the same US Census tract.
- When we implement restrictions over some places, we consider that the relationships in the others are not affected. Similarly, once restrictions are removed, we consider that the population goes back to normal, but it is likely that a significant fraction of individuals will still adopt self-aware measures, such as wearing masks, reducing some unnecessary contacts, etc. However, these can be considered as examples of active measures rather than passive and currently we have not included them in the model.