# E-Companion: Pandemic Lock-down, Isolation, and Exit Policies Based on Machine Learning Predictions

Theodoros Evgeniou
INSEAD, Bd de Constance, 77300 Fontainebleau, France

Mathilde Fekom
Université Paris-Saclay, ENS Paris-Saclay, CNRS, Centre Borelli, 91190 Gif-sur-Yvette, France

Anton Ovchinnikov
Smith School of Business, Queen's University, Kingston, ON, K7L3N6, Canada, and INSEAD, Bd de Constance, 77300 Fontainebleau, France

Raphael Porcher
Université de Paris CRESS, INSERM, INRA, 75004 Paris, France

Camille Pouchol
MAP5 Laboratory, FP2M, CNRS FR 2036, Université de Paris, 75006 Paris, France

Nicolas Vayatis
Université Paris-Saclay, ENS Paris-Saclay, CNRS, Centre Borelli, 91190 Gif-sur-Yvette, France

## Appendix A: Extensions and Variations

In this section, we discuss several extensions and variations to our model. In particular, because multiple approaches to modelling epidemics have been proposed, it is important to understand how our approach of combining an epidemic and a machine learning model would work in those cases and whether the insights would be qualitatively different.

### A.1. Nonlinear infection transition term in SIR models

One key modelling phase necessary for the development of any SIR model is to decide at which rate susceptible individuals are infected – that is, to choose a so-called *force of infection* $\lambda$ in equation $\dot{S} = -\lambda S$. Our model is based on the *density-dependent* approach for transmission, namely $\lambda = \beta I$, as opposed to the *frequency-dependent* approach for transmission, which reads $\lambda = \frac{\beta}{N} I$, where $N$ is the total number of individuals, Begon et al. (2002). The former choice is usually advocated for when the contact rate between individuals increases with density, as is the case for COVID-19.

The situation becomes more intricate when containment comes into play (see Acemoglu et al. (2021) for an in-depth discussion of this scenario). The key point is that in the presence of containment policies, a fine-modelled infection probably should exist between the two approaches. Hence, one way to generalize our model is to interpolate between the two scenarios as Acemoglu et al. (2021) did by introducing an additional parameter $\eta \in [0,1]$ so that $\lambda = \frac{\beta}{N^{\eta}} I$.

Importantly, taking $\eta = 0$, as we do, is a conservative approach because this maximizes the transmission rate among possible $\eta$ choices. In other words, any intermediate value for the parameter $\eta$ will lead to better results in terms of curbing the epidemic, all else equal: faster exits, fewer people in confinement, etc.

### A.2. Heterogeneity in model parameters – a "contact-matrix"-like approach

Another extension could deal with heterogeneity in model parameters. For instance Baqaee et al. (2020) consider an SIR-like model where each compartment is split by age into five groups: 0-19, 20-44, 45-64, 65-74, and 75+ (Duque et al. (2020) consider a similar split). Then, a susceptible adult of age $a_1$ who comes into contact with an adult of age $a_2$ has an instantaneous infection probability of $\beta$ times the probability that the latter adult is infected. The total instantaneous probability of infection is the sum over the expected transmission by contacts of different ages.

Our model has important similarities to and differences from such an approach. We also split each compartment but the split is not static; by choosing $\rho$, we "move" individuals between the groups and the risk model's ROC curve dictates if such moves are correct or are mistakes. Other than that, the instantaneous infection probability behaves identically; most importantly, it also depends on the behavior and sizes of the released and confined groups, even though we use the same level of $\beta$ at every $\rho$.

A model that combines the contact-matrix approach and ours is not difficult to build. Interestingly, in a context like that of COVID-19, the qualitative outcome of such an exercise is evident. The severity risk increases with age, while the degree of social interactions decreases with age. Therefore, interactions with the elderly will be restricted both because they are confined with $\delta_c$ and because they otherwise interact less than young people, i.e., as if $\delta_c$ decreases in $\rho$. This, again, makes our results conservative with respect to incorporating the heterogeneous interactions between age groups, and the qualitative results will hold. That said, in contexts other than COVID-19, such combined models may prove valuable.

### A.3. Alternative epidemic models: stochastic, agent-based, etc.

A completely different modelling approach is to use agent-based models that explicitly model each individual whose status evolves according to some probabilistic rules. Such models have been developed to analyze the spread of COVID-19 and other diseases such as seasonal influenza, e.g., Chao et al. (2010, 2020), Koo et al. (2020).

Rahmandad and Sterman (2008) provide an excellent comparison of the ODE- and agent-based models utilizing controlled experiments. Their overarching conclusion is that the differences between the two model types are small. They further note, *"In the realistic situation where policymakers face time pressure, imperfect data and uncertainty about public behaviour, the wise course may be to use a computationally efficient deterministic compartment model with a broad model boundary, to enable sensitivity and policy tests in time for action, rather than a computationally intensive individual-level model with a constrained boundary and limited ability to carry out parametric and structural sensitivity and policy tests."* This clearly favors ODE-based models for a new pathogen like SARS-CoV-2.

An additional drawback of agent-based models is their computational load. For instance, Chao et al. (2010) note that a single run of their simulation *"takes about 6 hours (192 hours of total CPU time)" ... [on a] "cluster of 32 processors."* This makes a combination of such a model with the optimal control problem like ours hardly feasible. In fact, the other studies that implemented optimal control, such as Acemoglu et al. (2021), Duque et al. (2020), Baqaee et al. (2020), all use SIR-like models. Taken together, these factors suggest that the use of the SIR-like model may be preferred for a study like ours.

### A.4. Alternative approaches to modelling severity risk

An implicit assumption is that our risk model assigns *individual* risk scores (i.e., a chance that a specific individual, given his or her features, would require an ICU if infected), and that these scores are constant over time. However, this narrow interpretation is not the only possible one, as the following example illustrates: Imagine a household in which two young adults live with their elderly parents. The individual risks for the young adults are low and our model would classify them as released, but an individual risk for an elderly parent may be high, and they would need to be confined. How can our model handle this?

First, because confinement in our model is imperfect, $\delta_c < 1$ already implicitly captures multi-generational household situations like the above. Surveys suggest that $\sim 25\%$ of households in the developed world are

multi-generational. Combined with the base rate of needing an ICU in single percent, less than $\sim 1\%$ of high-risk individuals would be impacted by this. Because our simulations use $\delta_c \in \{0.7, 0.9\}$, the multi-generational situations are already captured implicitly.

Second, one can redefine the risk score on a household, rather than individual, level, taking the probability that someone in a household would require an ICU if one of the members gets infected. In practice, this would, of course, require data about the composition of the households and such data would add to the features that the risk classifier uses. With such an approach, one also could implement the dependency of the risk scores on time. Importantly, because the household risk score would be the highest of the individual risk scores, such scores only would decrease over time as high-risk individuals are infected and removed. This, again, makes our results conservative vis-a-vis the efficacy of pandemic management policies in practice.

### A.5. Alternative objective functions and constraints

Lastly, we comment on the formulation of our optimization model. Unlike some other authors, who evaluate the detailed socio-economic consequences of lockdowns, e.g., Acemoglu et al. (2021), Baqaee et al. (2020), our implicit goal is to design policies that impose lockdowns that affect the smallest number of people and for the shortest period of time. Duque et al. (2020) incorporate this goal explicitly, by minimizing the total number of "people-days" during the lockdown. With our risk model, this is unnecessary because the highest-risk individuals are released last and only when their chance of contracting the disease are small enough because most others have been removed. In other words, it is sub-optimal to confine a lower-risk individual at time $t_i$ if releasing him or her is feasible. This is because releasing strictly increases the number of those removed, allowing for the faster release of the remaining high-risk individuals.

Similarly, we implicitly assume that the release policies are optimized subject to constraints on ICU availability. In practice, however, overall hospital capacity might be reached before that of the ICU. Incorporating this is straightforward: We add two additional compartments before the ICU (recall Figure 1 in the main body of the paper), together with the corresponding parameters, which can be easily obtained, e.g., from Salje et al. (2020).

Finally, $\delta_c$ in practice may depend on $\rho$. For instance, it may be possible to impose a tougher restriction, $\delta_c^{'} > \delta_c$, on only a small fraction of the population, but restricting a larger group may only be feasible with $\delta_c^{''} < \delta_c$. More generally, one could consider $\delta_c(\rho)$ and optimize as we did in the main body of the paper for

time-dependent $\delta_c$ due to fatigue. This is a plausible extension for our work, but our current results again could be viewed as conservative if one views $\delta_c$ as a lower-bound on such a function. Indeed, if the remaining high-risk individuals would be better confined, our model would release more low-risk individuals and sooner.

## Appendix B: Further sensitivity analyses for a single release

We here gather additional sensitivity analyses in the case of a single release, presented in Figure 1, with respect to key parameters of the model: the percentage $p$ of the population with severe symptoms upon infection plays a crucial role, which we decrease from the posterior average $\sim 0.993$ to 0.98, and the ICU capacity, which we increase from 7250 to 15000. Other parameters are same as in Figure 3 in th main body of the paper.

First, as expected, the lower $p$, the less the impact of a risk prediction model keeping the risk-model AUC constant. Given the limited ICU – and possibly other – resources, a smaller $p$ allows for a smaller range of percentages of the population being released, making all differences between policies smaller in absolute terms. Second, when we compare the main-body Figure 3 (a) and (c) with Figures 1 (a) and (c) here, where the only difference is in the total ICU capacity, we see that the more ICUs available the larger the impact of using personalized policies, keeping everything else constant. More available resources allow for a larger range of percentage of released people making the differences between policies – risk based vs. not – larger in absolute terms. Note that in all cases a risk prediction model approach allows for confining fewer people; this is consistent with the value of information-related arguments, as any test provides information that can be beneficial assuming everything else (including behavioral aspects) kept constant.

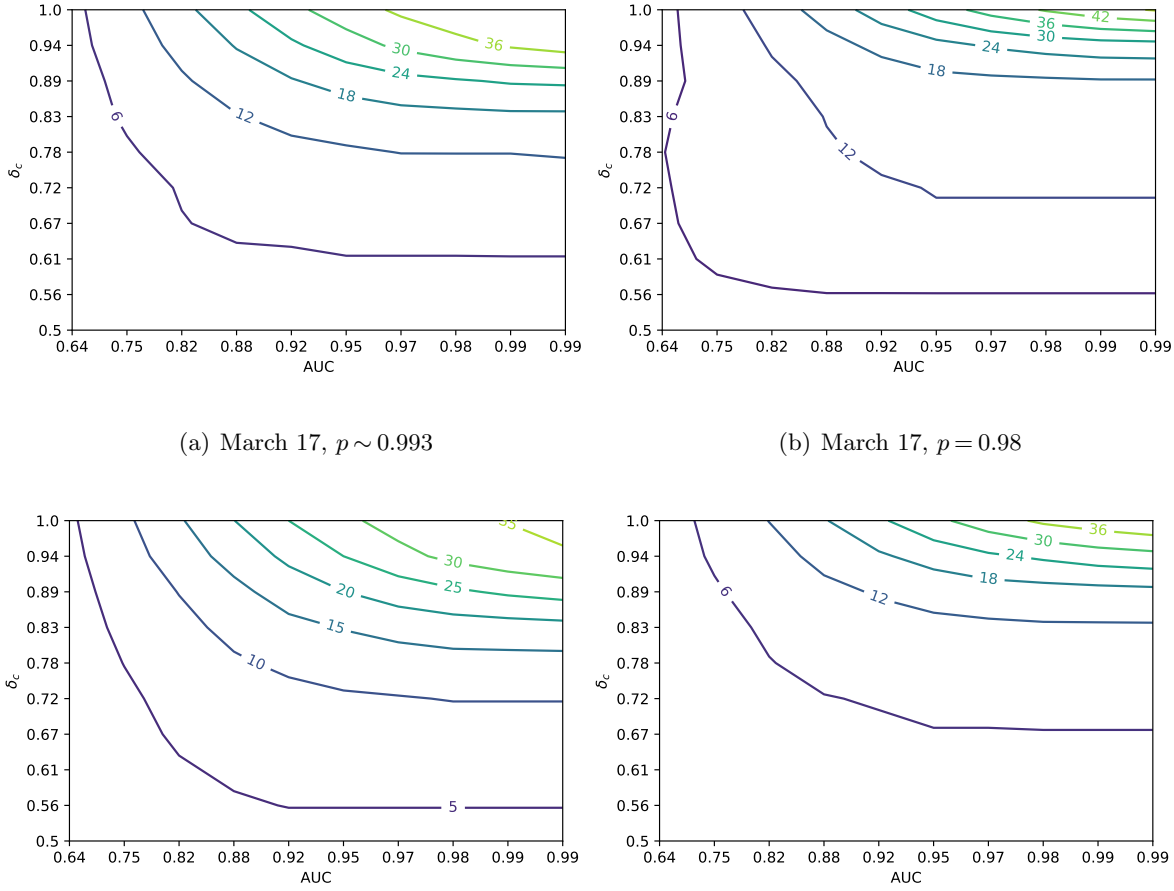
## References

Acemoglu D, Chernozhukov V, Werning I, Whinston MD (2021) Optimal targeted lockdowns in a multigroup sir model. *American Economic Review: Insights* 3(4):487–502.

Baqaee D, Farhi E, Mina M, Stock JH (2020) Policies for a second wave. *Brookings Papers on Economic Activity* 385–431.

Begon M, Bennett M, Bowers RG, French NP, Hazel S, Turner J (2002) A clarification of transmission terms in host-microparasite models: numbers, densities and areas. *Epidemiology & Infection* 129(1):147–153.

Chao DL, Halloran ME, Obenchain VJ, Longini Jr IM (2010) Flute, a publicly available stochastic influenza epidemic simulation model. *PLoS Comput Biol* 6(1).




(a) March 17, $p \sim 0.993$

(b) March 17, $p = 0.98$

(c) May 11, $p \sim 0.993$

(d) May 11, $p = 0.98$

**Figure 1** **Difference in maximal possible percentage of released people without exceeding the ICU capacity, compared to the case of not using a risk prediction model, plotted as a function of the AUC of a risk prediction model and the protection level $\delta_c$ for confined peoplections. $\delta_r = 0.1$ for all figures.**

Chao DL, Oron AP, Srikrishna D, Famulare M (2020) Modeling layered non-pharmaceutical interventions against sars-cov-2 in the united states with corvid. *medRxiv 2020.04.08.20058487* .

Duque D, Morton DP, Singh B, Du Z, Pasco R, Meyers LA (2020) Timing social distancing to avert unmanageable covid-19 hospital surges. *PNAS* 117(33):19873 – 19878.

Koo JR, Cook AR, Park M, Sun Y, Sun H, Lim JT, Tam C, Dickens BL (2020) Interventions to mitigate early spread of sars-cov-2 in singapore: a modelling study. *The Lancet Infectious Diseases* 20:678–688.

Rahmandad H, Sterman J (2008) Heterogeneity and network structure in the dynamics of diffusion: Comparing agent-based and differential equation models. *Management Science* 54(5):iv–1036.

Salje H, Tran Kiem C, Lefrancq N, Courtejoie, et al. (2020) Estimating the burden of sars-cov-2 in france. *Science* 369(6500):208–211.