# SEAIR Framework Accounting for a Personalized Risk Prediction Score: Application to the Covid-19 Epidemic

Olivier Boulant[1*], Mathilde Fekom[1*], Camille Pouchol[2*], Theodoros Evgeniou[3],
Anton Ovchinnikov[4], Raphaël Porcher[5], Nicolas Vayatis[1]

[*] The first three authors wrote the code. All authors developed the modelling framework.
[1] Université Paris-Saclay, ENS Paris-Saclay, CNRS, Centre Borelli, 94235 Cachan, France
[3] INSEAD, Bd de Constance, 77300 Fontainebleau, France
[4] Smith School of Business, Queen's University, Kingston, ON, K7L3N6, Canada
[5] Université de Paris CRESS, INSERM, INRA, 75004 Paris, France
[2] Université de Paris, FP2M, CNRS FR 2036, MAP5 UMR 8145, F-75006 Paris, France

*Communicated by* Gregory Randall     *Demo edited by* Olivier Boulant

## Abstract

The aim of the present work is to provide an SEAIR framework which takes a personalized risk prediction score as an additional input. Each individual is categorized depending on his actual status with respect to the disease – moderate or severe symptoms –, and the level of risk predicted – low or high. This idea leads to a 4-fold extension of the ODE model in classical SEAIR. This model offers the possibility for policy-makers to explore differentiated containment strategies, by varying sizes for the low risk segment and varying dates for 'progressive release' of the population, while exploring the discriminative capacity of the risk score, for instance through its AUC. Differential contact rates for low-risk/high-risk compartments are also included in the model. The demo allows to select contact rates and time-depending exit strategies. The hard-coded parameters correspond to the data for the Covid-19 epidemic in France, and the risk refers to the probability of being admitted in ICU upon infection. Some examples of simulations are provided.

## Source Code

The reviewed source code and documentation for this algorithm are available from the web page of this article[1]. Usage instructions are included in the `readme.md` file of the archive.

**Keywords:** COVID-19; SIR; isolation measures; risk prediction model

---

[1]https://doi.org/10.5201/ipol.2020.305

# 1 Introduction

In the context of the Covid-19 epidemics, several extensions of the classical SIR compartmental model (Susceptible-Infected-Recovered) have been proposed to cover some of the features of this particular contagion. Among the reference models used in the case of France, one of the most popular is the *SEAIR model* (Susceptible-Exposed-Asymptomatic-Infected-Recovered) further developed by [3] to account for people having moderate or severe symptoms upon infection.

The present work introduces an extension to the SEAIR model in the case where a *personalized risk prediction score* would be available. By high score, we mean here a high probability of being transmitted to an Intensive Care Unit (ICU) upon infection. Through the use of a threshold value, the risk prediction score could support a decision-system that would categorize the population into two segments *adaptively*: a high-risk segment and a low-risk segment. Depending on the context (presence of the virus, modalities of lockdown policy, resources for protection, etc.), the respective size of these two segments can be adjusted. Such risk-prediction models are about to become available as epidemiological studies track the dependence between comorbidity factors and individual outcome with respect to the disease [1]. From there, it will be quite straightforward to derive a personalized risk prediction score based on either logistic regression or more advanced machine learning algorithms for bipartite ranking. We refer to https://www.covidanalytics.io/infection_calculator for an example of such a score.

Our aim is to investigate to which extent the quality of a given prediction model – captured for example by the Area Under a ROC Curve (AUC) –, together with a policy that differentiates high and low risk individuals, can "flatten the curve". Here "flattening the curve" is quantified with respect to the demand in ICU beds that has to remain below a given acceptable level. This approach can be used both for tuning lockdown/exit strategies or for dealing with possible further epidemic waves. We refer to [4] for an investigation of differentiated lockdown strategies in France based on this approach.

# 2 SEAIR with High-risk/Low-risk Compartments

## 2.1 A Compact View of the Model

We propose to start with a variation of the SEAIR model with an extra state denoted by $U$, which means 'admitted in ICU'. As usual $S$, stands for susceptible, $E$ for exposed (asymptomatic and not yet contagious), $A$ for asymptomatic (asymptomatic and contagious), $I$ for infectious (symptomatic and contagious), $R$ for recovered, and finally a last compartment $D$ is considered, counting people who died from the disease.

The standard model is first generalized by dividing

- the susceptible state $S$ into two states depending on each individual's *isolation policy* which is based on their associated personalized risk score, namely those predicted to be low-risk may resume normal activities up to some extent ('r'), while those predicted to be high risk are recommended to remain confined ('c'). These two states are denoted $S^{(r)}$ and $S^{(c)}$. Correspondingly, all other states $Q$ are split up into states $Q^{(r)}$ and $Q^{(c)}$.

- the infected state $I$ into two states depending on the *severity of symptoms* they will develop upon infection ('s' for severe symptoms requiring ICU, 'm' for moderate symptoms). The resulting states are denoted $I_m$ and $I_s$. Correspondingly, all other states $Q$ are split up into states $Q_m$ and $Q_s$.

## 2.2 Key Parameters of the Epidemic Model

Standard epidemiological and additional behavioral parameters come into play in the model. Their interaction with the main compartmental time-dependent variables is schematically shown in Figure 1.

*Standard parameters.* The usual parameters of the model are the rates of passage from a given compartment to any other compartment which here read $\varepsilon$, $\sigma$, $\eta$, $\gamma_m$, $\gamma_s$ and $\alpha$. Recall that the inverse of these variables represents the mean time spent in the associated compartment. The parameter $\beta$, as is usual, represents the transmission rate.

All these parameters are hard-coded and the proposed values can be found in Table 1.

*Additional behavioral parameters.* Given an isolation policy based on the personalized risk prediction score, the behavior of individuals is also determined by different levels of compliance to the policy, which could potentially change according to the risk segment they belong to. We represent this behavioral aspect by the two following variables $\delta_r$ and $\delta_c$, with values in $[0, 1]$. They capture how much released/confined individuals protect themselves from the disease, whether it is by stricter hygiene, avoiding public transports, limiting outdoor excursions, etc. The user should always choose $\delta_r < \delta_c$ to reflect the modelling approach.

The closer the $\delta$-variables are to 1, the stricter the protection. As a result, the usual quadratic $\beta I S$ term in SIR models, which in SEAIR models reads $\beta(A + I)S$, becomes

$$(1 - \delta_i)\beta I^{\text{eff}} S^{(i)}, \quad i = r, c,$$

where, to shorten notations, we have defined

$$I^{\text{eff}} = (1 - \delta_r)(A^{(r)} + I^{(r)}) + (1 - \delta_c)(A^{(c)} + I^{(c)}).$$

These parameters are behavioral and depend on how people respond to policies regarding isolation. One of the aims of the present work is to investigate the effect of their value on the success of differentiated isolation policies.

## 2.3 A Simplified Version of the ODE System

For the reader's convenience, a compact form of the ODE model used may be given. Here, $\dot{Q}$ stands for the time derivative of the variable $Q$. The corresponding equations are given below and Figure 1 provides a schematic view of the resulting ODE system, where the rate from one compartment to another is indicated above the corresponding arrow.

$$\dot{S}^{(r)} = -(1 - \delta_r)\beta I^{\text{eff}} S^{(r)}$$
$$\dot{E}^{(r)} = (1 - \delta_r)\beta I^{\text{eff}} S^{(r)} - \varepsilon E^{(r)}$$
$$\dot{S}^{(c)} = -(1 - \delta_c)\beta I^{\text{eff}} S^{(c)}$$
$$\dot{E}^{(c)} = (1 - \delta_c)\beta I^{\text{eff}} S^{(c)} - \varepsilon E^{(c)}$$
$$\dot{A} = \varepsilon E - \sigma A$$
$$\dot{I}_m = \sigma A_m - \gamma_m I_m$$
$$\dot{I}_s = \sigma A_s - \eta I_s$$
$$\dot{U} = \eta I_s - (\gamma_s + \alpha)U$$
$$\dot{R} = \gamma_m I_m + \gamma_s I_s$$

The number of deaths is computed with the relation:
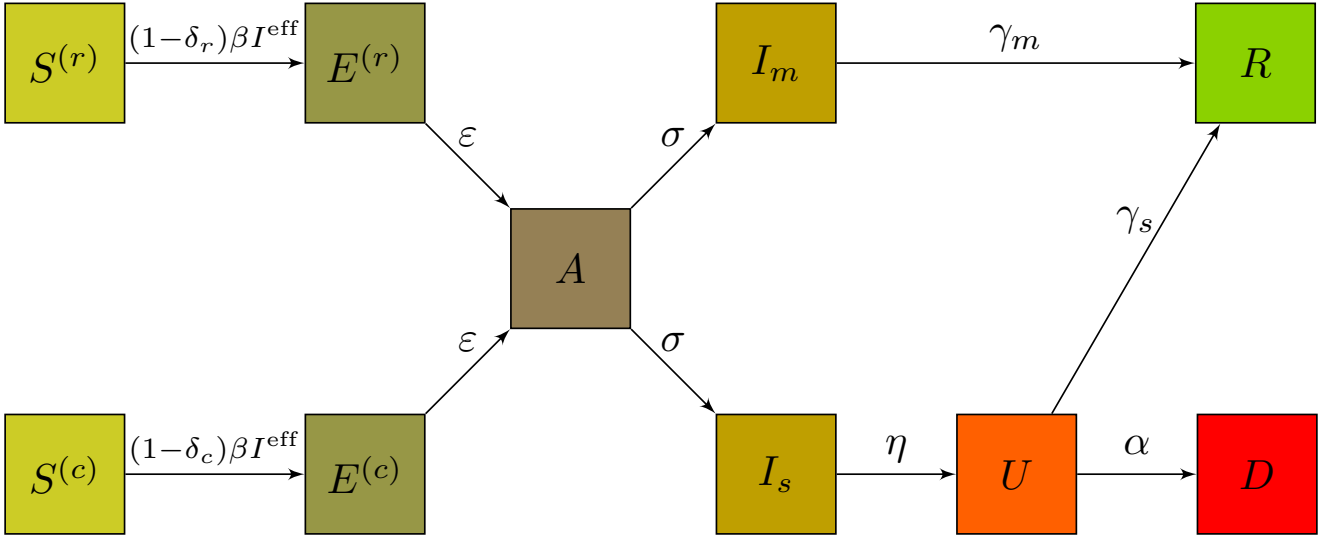
$$\dot{D} = \alpha\, U.$$



Figure 1: A compact view of the dynamics of the SEAI(U)R model. $S$ stands for susceptible, $E$ for exposed, $A$ for asymptomatic, $I$ for infectious, $U$ for people in ICU, $R$ for recovered, $D$ for dead. Superscripts $(r)/(c)$ indicate released/confined people, subscripts $m/s$ indicate moderate/severe symptoms. Values above arrows indicate the rate of passage from the associated compartments (see Table 1 for more details), with $I^{\text{eff}} = (1 - \delta_r)(A^{(r)} + I^{(r)}) + (1 - \delta_c)(A^{(c)} + I^{(c)})$.

## 3 The Complete ODE System

### 3.1 Description of Compartments

In order to take into account the status of each state with respect to the risk prediction in the previous model, each compartment is split according to the outcome of the binary decision for the segment considered based on the personalized predictive risk score. Indeed, we assume that the decision cannot be perfect and that errors will occur. Therefore, each binary decision leads to the segmentation of the population into 4 segments which correspond to the 4 cases of the confusion matrix of a classification rule: true positive (TP), false positive (FP), true negative (TN), false negative (FN). Hence, every compartment $Q \in \{S, E, A, I, R\}$ is divided into 4 compartments, whereas the compartment $U$ is divided only into two because only positive cases will fall into this one. Thus, the total number of compartments equals 22. We recall that we use superscripts to mark the decision made in terms of risk prediction, where "r" indicates the *low-risk* segment ("r" for "released"), while "c" refers to the *high-risk* segment ("c" for "confined"). On the other hand, we use subscripts for the actual severity status of the segment, where "s" corresponds to the segment of individuals who will develop *severe* symptoms requiring ICU care, and "m" for the segment of individuals who will develop *moderate* symptoms upon infection. For a generic compartment $Q \in \{S, E, A, I, R\}$, the following notations are used to account for the four cases of the confusion matrix:

- $Q_s^{(r)}$: number of individuals in $Q$ with severe symptoms requiring ICU care upon infection, categorized low-risk by the predictor, i.e., false negatives,

- $Q_s^{(c)}$: number of individuals in $Q$ with severe symptoms requiring ICU upon infection, categorized high-risk by the predictor, i.e., true positives,

- $Q_m^{(r)}$: number of individuals in $Q$ with moderate symptoms upon infection, categorized low-risk by the predictor, i.e., true negatives,

- $Q_m^{(c)}$: number of individuals in $Q$ with moderate symptoms upon infection, categorized high-risk by the predictor, i.e., false positives.

In addition, we will also use the cumulative count either with respect to risk prediction or severity status for any $Q \in \{S, E, A, I, U, R\}$, and as already introduced with the compact view of the model, this leads to

- $Q^{(r)} = Q_m^{(r)} + Q_s^{(r)}$: number of individuals in $Q$ categorized *low-risk*,

- $Q^{(c)} = Q_s^{(c)} + Q_m^{(c)}$: number of individuals in $Q$ categorized *high-risk*,

- $Q_s = Q_s^{(c)} + Q_s^{(r)}$: number of individuals in $Q$ who will develop *severe* symptoms requiring ICU upon infection,

- $Q_m = Q_m^{(r)} + Q_m^{(c)}$: number of individuals in $Q$ who will develop *moderate* symptoms upon infection,

- $Q = Q_m + Q_s = Q^{(r)} + Q^{(c)}$: total count of the population on which the dynamics of the epidemic is observed.

We consider one additional compartment, denoted by $D$, for the segment of individuals who die from the disease. The number of people in this compartment does not participate in the dynamics of the ODE but may be computed from it.

## 3.2 ODE System: Complete Set of Equations

The ODE system reads

$$\dot{S}_m^{(r)} = -(1 - \delta_r)\beta I^{\text{eff}} S_m^{(r)}$$
$$\dot{E}_m^{(r)} = (1 - \delta_r)\beta I^{\text{eff}} S_m^{(r)} - \varepsilon E_m^{(r)}$$
$$\dot{A}_m^{(r)} = \varepsilon E_m^{(r)} - \sigma A_m^{(r)}$$
$$\dot{I}_m^{(r)} = \sigma A_m^{(r)} - \gamma_m I_m^{(r)}$$
$$\dot{R}_m^{(r)} = \gamma_m I_m^{(r)}$$
$$\dot{S}_m^{(c)} = -(1 - \delta_c)\beta I^{\text{eff}} S_m^{(c)}$$
$$\dot{E}_m^{(c)} = (1 - \delta_c)\beta I^{\text{eff}} S_m^{(c)} - \varepsilon E_m^{(c)}$$
$$\dot{A}_m^{(c)} = \varepsilon E_m^{(c)} - \sigma A_m^{(c)}$$
$$\dot{I}_m^{(c)} = \sigma A_m^{(c)} - \gamma_m I_m^{(c)}$$
$$\dot{R}_m^{(c)} = \gamma_m I_m^{(c)}$$
$$\dot{S}_s^{(c)} = -(1 - \delta_c)\beta I^{\text{eff}} S_s^{(c)}$$
$$\dot{E}_s^{(c)} = (1 - \delta_c)\beta I^{\text{eff}} S_s^{(c)} - \varepsilon E_s^{(c)}$$
$$\dot{A}_s^{(c)} = \varepsilon E_s^{(c)} - \sigma A_s^{(c)}$$
$$\dot{I}_s^{(c)} = \sigma A_s^{(c)} - \eta I_s^{(c)}$$
$$\dot{U}^{(c)} = \eta I_s^{(c)} - (\gamma_s + \alpha)U^{(c)}$$
$$\dot{R}_s^{(c)} = \gamma_s I_s^{(c)}$$
$$\dot{S}_s^{(r)} = -(1 - \delta_r)\beta I^{\text{eff}} S_s^{(r)}$$
$$\dot{E}_s^{(r)} = (1 - \delta_r)\beta I^{\text{eff}} S_s^{(r)} - \varepsilon E_s^{(r)}$$
$$\dot{A}_s^{(r)} = \varepsilon E_s^{(r)} - \sigma A_s^{(r)}$$
$$\dot{I}_s^{(r)} = \sigma A_s^{(r)} - \eta I_s^{(r)}$$
$$\dot{U}^{(r)} = \eta I_s^{(r)} - (\gamma_s + \alpha)U^{(r)}$$
$$\dot{R}_s^{(r)} = \gamma_s U_s^{(r)},$$

and the number of deaths are computed using the equation

$$\dot{D} = \alpha \left( U^{(c)} + U^{(r)} \right),$$

where the dot above each variable stands for a time derivative. A given simulation of a Cauchy problem with the above ODE system requires an initial condition for the 23 compartments. We shall formally denote ODE the function solving the ODE, requiring an initial condition and a time-interval on which to solve the problem.

## 3.3 Generic Discretization Scheme for the ODE

The code fixes the time-discretization at 500 points per day. All plots and variables are to be considered in days. ODEs are integrated using the Python function odeint from the Scipy package.

# 4 ODE Integration with Time-dependent Policies

## 4.1 Computing Initial Conditions

For a given choice of number of individuals in each category, denoted by $S_0$, $E_0$, $A_0$, $I_0$, $U_0$, $R_0$, the resulting initial numbers in each compartment may be computed by taking into account

- the proportion of people who will not require an ICU bed, denoted $p \in [0,1]$, for which estimates are available in the literature [6],

- the proportion of people considered to be low-risk, denoted $\rho \in [0,1]$, which depends on a choice made by the policy-makers,

- the risk-prediction model.

For a given risk-prediction model, $\rho$ will uniquely determine the number of false positives and false negatives, respectively denoted $q^{FP}$, $q^{FN}$. All these variables $p$, $\rho$, $q^{FP}$, $q^{FN}$ are linked by the equation

$$(1 - q^{FP})p + q^{FN}(1 - p) = \rho. \tag{1}$$

This equation (which can be seen as a total probability formula) is readily obtained by counting that $\rho$ will be split up into the true negative and the true positive people. The former is obtained by multiplying the true negative rate, i.e. $q^{TN} = 1 - q^{FP}$, with the number of true negative $p$, while the latter is taken by multiplying the false negative rate $q^{FN}$ with the number of true positive $1 - p$.

Let us denote $F_s$ and $F_m$ the cumulative distribution functions for the probability densities of the people needing ICU and those having moderate symptoms upon infection, respectively. For a choice of $\rho$ and assuming continuous random variables, there is a unique threshold $a$ which makes the proportion of low-risk people exactly equal $\rho$, obtained by solving

$$pF_m(a) + (1 - p)F_s(a) = \rho.$$

The pseudo-code for how $q^{FP}$, $q^{FN}$ are computed from $\rho$ (through the function denoted `compute_qs`), is given in Algorithm 1.

---

**Algorithm 1** Pseudo-code for computing false negative and false positive rates.

$a \leftarrow \text{solve}(pF_m(a) + (1 - p)F_s(a) = \rho)$          $\triangleright$ Solves for the unique threshold
$q^{FP} \leftarrow 1 - F_m(a)$
$q^{FN} \leftarrow F_s(a)$
The outcome is $(q^{FP}, q^{FN})$.

---

We will denote by `compute_init` the function which, given a choice of $S_0$, $E_0$, $A_0$, $I_0$, $U_0$, $R_0$ on the one hand, and of $\rho$ on the other hand, computes the corresponding initial conditions. See the pseudo-code given in Algorithm 2.

## 4.2 Updating the ODE System over Time

As a given policy makes $\rho$ vary over time, the model must specify how to update the ODE whenever the policy changes. More precisely, a policy is defined by the times at which restrictions are gradually relaxed (thus augmenting the proportion of people labeled as low-risk), together with the actual proportions of people who are in the low-risk group.

In mathematical terms, we are given

---

**Algorithm 2** Pseudo-code for computing an initial condition.

$(q^{FP}, q^{FN}) \leftarrow \texttt{COMPUTE\_QS}(\rho)$
**for** $Q \in \{S, E, A, I, U, R\}$ **do**
$\quad Q_m^{(r)}(0) \leftarrow p\,(1 - q^{FP})Q_0$
$\quad Q_m^{(c)}(0) \leftarrow p\,q^{FP}Q_0$
$\quad Q_s^{(c)}(0) \leftarrow (1 - p)\,(1 - q^{FN})Q_0$
$\quad Q_s^{(r)}(0) \leftarrow (1 - p)\,q^{FN}Q_0$
**end for**

The outcome is

$$(S_m^{(r)}, E_m^{(r)}, A_m^{(r)}, I_m^{(r)}, R_m^{(r)}, S_m^{(c)}, E_m^{(c)}, A_m^{(c)}, I_m^{(c)}, R_m^{(c)},$$
$$S_s^{(c)}, E_s^{(c)}, A_s^{(c)}, I_s^{(c)}, U^{(c)}, R_s^{(c)}, S_s^{(r)}, E_s^{(r)}, A_s^{(r)}, I_s^{(r)}, U^{(r)}, R_s^{(r)})(0)$$

---

- a vector of proportions of releases

$$(\rho_0, \rho_1, \ldots, \rho_{N-1}),$$

  satisfying $\rho_{i-1} \leq \rho_i$, $i = 1, \ldots N - 1$,

- the corresponding times at which the policy changes

$$(T_0, T_1, \ldots T_{N-1}, T_N)$$

  with $T_0 = 0$ and $T = T_N$ a final horizon of interest.

Each $\rho_i$ that says how many people will be predicted to be low risk, determines the number of false positives and false negatives in agreement with (1) through

$$(1 - q_i^{FP})p + q_i^{FN}(1 - p) = \rho_i.$$

Assume that the solution at time $T_i$ has been computed for a given $i$. Since the number of individuals in the low-risk group changes at time $T_i$, the ODE must be integrated for a corrected initial condition on the interval $[T_i, T_{i+1}]$.

This initial condition is obtained from reallocating people depending on their new labeling, which will both make the number of false negatives increase and the number of false positives decrease.

Let us explain how the changes are made at a given time $T_i$, by letting $Q \in \{S, E, A, I, R\}$.

**Case of $Q_m$.** At $T_i$, the number of people in $Q_m^{(r)}$ increases, while the number of people in $Q_m^{(c)}$ decreases. More precisely, a portion $\frac{q_{i-1}^{FP} - q_i^{FP}}{q_{i-1}^{FP}}$ of people previously in $Q_m^{(c)}$ moves to compartment $Q_m^{(r)}$. This ratio corresponds to the proportion of people considered low-risk with the policy defined by $\rho_i$, conditioned on the fact that they were considered high-risk with the policy defined by $\rho_{i-1}$, among people not requiring ICU upon infection.

**Case of $Q_s$.** At $T_i$, the number of people in $Q_s^{(r)}$ increases, while the number of people in $Q_s^{(c)}$ decreases. More precisely, a portion $\frac{q_i^{FN} - q_{i-1}^{FN}}{1 - q_{i-1}^{FN}}$ of people previously in $Q_s^{(c)}$ moves to compartment $Q_s^{(r)}$. This ratio corresponds to the proportion of people considered low-risk with the policy defined by $\rho_i$, conditioned on the fact that they were considered high-risk with the policy defined by $\rho_{i-1}$, among people requiring ICU upon infection.

More compactly, denoting $y^{\text{old}}$ and $y^{\text{new}}$ the value of $y(T_i)$ before and after relabeling, the new value is given as a function of the previous one by

$$y_j^{\text{new}} = y_j^{\text{old}} + \frac{q_{i-1}^{FP} - q_i^{FP}}{q_{i-1}^{FP}} y_{j+5}^{\text{old}}, \qquad\qquad j = 1, \ldots, 5$$

$$y_{j+5}^{\text{new}} = y_{j+5}^{\text{old}} - \frac{q_{i-1}^{FP} - q_i^{FP}}{q_{i-1}^{FP}} y_{j+5}^{\text{old}}, \qquad\qquad j = 1, \ldots, 5$$

$$y_j^{\text{new}} = y_j^{\text{old}} - \frac{q_i^{FN} - q_{i-1}^{FN}}{1 - q_{i-1}^{FN}} y_j^{\text{old}}, \qquad\qquad j = 11, \ldots, 16$$

$$y_{j+6}^{\text{new}} = y_{j+6}^{\text{old}} + \frac{q_i^{FN} - q_{i-1}^{FN}}{1 - q_{i-1}^{FN}} y_j^{\text{old}}, \qquad\qquad j = 11, \ldots, 16.$$

We shall denote `change_policy` the function which given $y^{old}$ and $\rho_i, \rho_{i-1}$, computes $y^{new}$.

With the above rule, the corresponding solution

$$y = (S_m^{(r)}, E_m^{(r)}, A_m^{(r)}, I_m^{(r)}, R_m^{(r)}, S_m^{(c)}, E_m^{(c)}, A_m^{(c)}, I_m^{(c)}, R_m^{(c)},$$
$$S_s^{(c)}, E_s^{(c)}, A_s^{(c)}, I_s^{(c)}, U^{(c)}, R_s^{(c)}, S_s^{(r)}, E_s^{(r)}, A_s^{(r)}, I_s^{(r)}, U^{(r)}, R_s^{(r)})$$

is computed up until the final time $T$.

**Effect of a change of policy.** Note that a change in policy induces discontinuities for the different variables, but continuity is restored at the level of the total number of people in a given category $Q_m$ or $Q_s$.

## 4.3 Pseudo-code for Time-dependent Policies

Algorithm 3 provides a pseudo-code outlining how, given a policy, the ODE is integrated. In other words, all other parameters being fixed and for a choice of $S_0, E_0, A_0, I_0, U_0, R_0, (\rho_0, \rho_1, \ldots, \rho_{N-1})$, $(T_0, T_1, \ldots T_{N-1}, T_N)$, the algorithm writes as follows.

---

**Algorithm 3** Pseudo-code for integrating the ODE along time-dependent policies.

$y_0 \leftarrow \texttt{COMPUTE\_INIT}(S_0, E_0, A_0, I_0, U_0, R_0, \rho_0)$
**for** $i \leftarrow 1, \ldots, N$ **do**
    $\text{sol}_i \leftarrow \texttt{ODE}(y_{i-1}, [T_{i-1}, T_i])$         $\triangleright$ Solves the ODE over $[T_{i-1}, T_i]$
    $y^{old} \leftarrow \text{sol}_i(T_i)$            $\triangleright$ Gets value of solution at time $T_i$
    $y_i \leftarrow \texttt{CHANGE\_POLICY}(y^{old}, \rho_{i-1}, \rho_i)$    $\triangleright$ Updates numbers with the new policy
**end for**
The outcome is the concatenation of $\text{sol}_i$, $i = 1, \ldots, N$.

---

# 5 Experimental Design and Value Recommendations for Key Parameters

In order to run an experiment, the user can select the following parameters:

- disease-related parameters,

- a risk-prediction model,

- behavioral parameters $\delta_r$ and $\delta_c$,

- a set of initial conditions,

- a policy regarding high-risk and low-risk people over time.

## 5.1 Disease-related Parameters

Disease-related parameters are hard-coded but may be changed by the user. They are summarized in Table 1.

| Symbol | Description | Value(s) | Reference |
|---|---|---|---|
| $p$ | proportion with moderate symptoms | 0.993 | [6] |
| $\beta$ | transmission rate | computed | [3] |
| $\mathcal{R}_0$ | basic reproduction number | 2.9 | [2] |
| $\varepsilon$ | waiting rate to viral shedding | $1/3.7$ day$^{-1}$ | [2] |
| $\sigma$ | waiting rate to symptom onset | $1/1.5$ day$^{-1}$ | [2] |
| $\eta$ | waiting rate from symptom onset to ICU | $1/7$ day$^{-1}$ | [6] |
| $\gamma_m$ | recovery rate from moderate symptoms | $1/2.3$ day$^{-1}$ | [2] |
| $\gamma_s$ | recovery rate for people in ICU | $1/17$ day$^{-1}$ | [6] |
| $\alpha$ | mortality rate for people in ICU | $1/11.7$ day$^{-1}$ | [6] |

Table 1: Disease-related parameters used.

The parameter $\beta$ may be computed as in [3] by stability analysis. Denoting $\mathcal{R}_0$ the basic reproduction number, as estimated without lockdown (*i.e.*, for $\delta_r = \delta_c = 0$), the formula writes

$$\beta = \frac{\mathcal{R}_0}{S_{\text{init}}} \frac{1}{p(\sigma^{-1} + \gamma_m^{-1}) + (1-p)(\sigma^{-1} + \eta^{-1})},$$

where $S_{\text{init}}$ is the initial number of susceptible people for the period during which $\mathcal{R}_0$ has been estimated.

Since all estimates for $\mathcal{R}_0$ have been obtained for the period prior to lockdown, we choose $S_{\text{init}} = N_0$ where $N_0$ is the total number of people in the population of interest.

## 5.2 Choosing a Risk-prediction Model

For the sake of illustration, the simulation may run with a synthetic risk-prediction model which uses Beta distributions to represent class-conditional distributions (with respect to severity of symptoms). We denote $\alpha_p, \beta_p$, the parameters of the Beta-distribution for the score of positive (severe symptoms) and $\alpha_n, \beta_n$, for the score of negative (moderate symptoms).

The quality of a test can be summarized by the so-called AUC (Area under the Curve) from the corresponding ROC (Receiver Operating Characteristic) curve. The AUC has value 0.5 for a dummy risk-prediction model (the class-conditional distributions are equal), and 1 for a perfect risk-prediction model (the supports of class-conditional distributions are disjoint). Some preliminary work developing risk-prediction models on Covid-19 using age, BMI (body mass index), lab values, and some comorbidity factors reach an AUC ranging between 0.83 and 0.92 (see https://www.covidanalytics.io/infection_calculator).

Let us denote $\text{AUC}(\alpha_p, \beta_p, \alpha_n, \beta_n)$ the AUC associated to a given choice of $(\alpha_p, \beta_p, \alpha_n, \beta_n)$.

Such an AUC is obtained by computing false positive and true positive rates on a grid for thresholds, and then using the built-in function `metrics.auc` of Scikit-learn [5], which takes false positive and true positive rates as an input and solves for the corresponding AUC. The default parameter for the grid size $n_{grid}$ is 500, see the pseudo-code Algorithm 4.

---

**Algorithm 4** Pseudo-code for computing AUC.

$a \leftarrow$ `UNIFORM_GRID`$([0, 1], n_{grid})$
$q^{FP} \leftarrow 1 - F_m(a)$
$q^{TP} \leftarrow 1 - F_s(a)$                                     ▷ Computes false positive and true positive rates.
The outcome is `metrics.auc`$(q^{FP}, q^{TP})$.

---

We let the user choose an AUC between 0.5 and 1 by choosing symmetric distributions, i.e., $\beta_p = \alpha_n$ and $\beta_n = \alpha_p$. Choosing $\beta_p = \alpha_n = 2$ as a default value and for a given choice $z \in [0.5, 1]$ of AUC given by the user, we solve for $x$ the equation

$$\text{AUC}(x, 2, 2, x) = z.$$

This equation is uniquely solvable when $z \in [0.5, 1[$. For $z = 1$, the solution is formally given by $x = +\infty$, in which case we set it to $x = 1000$.

## 5.3 Behavioral Parameters

The user can also select values for $\delta_c$, $\delta_r$. Note that these variables may be linked to the so-called *reduction of contact rates*. Denoting reduction of contact rates with the letter $c$, the equation for the susceptible individuals write $\dot{S} = -(1 - c)\beta IS$ in the standard SIR model.

With this in mind, and given that in our case the number of individuals that fall in the low-risk or in the high-risk group depends on the parameter $\rho$, reduction of contact rates for populations in these two groups may correspondingly be defined by the relation

$$1 - c_i = (1 - \delta_i)\left((1 - \delta_r)\rho + (1 - \delta_c)(1 - \rho)\right), \quad i \in \{r, c\}.$$

For a given choice of $\delta_r$, $\delta_c$, a plot displaying the evolution of $c_r$ and $c_c$ according to $\rho$ is provided to the user. As an example, the reduction in contact rates has been estimated to range from 70% to 80% during lockdown in France (that is, $c = 0.7$ to $c = 0.8$).

## 5.4 Initial Conditions

Initial conditions should be chosen depending on the country/region. More precisely, we let the user choose

- $N_0$: total number of individuals in the population,

- $E_0 + A_0 + I_0$: number of individuals who were carrying the disease at time 0,

- $U_0$: number of individuals in ICU at time 0,

- $R_0$: number of individuals who recovered from the disease at time 0,

The initial number of susceptible individuals is then computed by

$$S_0 = N_0 - (E_0 + A_0 + I_0 + U_0 + R_0).$$

We do not explicitly require the user to choose $E_0$, $A_0$ and $I_0$, as it is already difficult to have an estimate for these. Instead, for a given choice of $E_0 + A_0 + I_0$, we choose to compute $E_0$, $A_0$, $I_0$ relative to the mean time spent in each of these categories for the majority type, namely individuals not needing ICU upon infection. Denoting

$$x := \frac{1}{1 + \varepsilon \left( \frac{1}{\sigma} + \frac{1}{\gamma_m} \right)},$$

this leads to the formulae

$$E_0 = x(E_0 + A_0 + I_0),$$
$$A_0 = \frac{\varepsilon}{\sigma} x(E_0 + A_0 + I_0),$$
$$I_0 = \frac{\varepsilon}{\gamma_m} x(E_0 + A_0 + I_0).$$

Finally, the user may choose a value (which is not required for computations but is then displayed when plotting ICU occupation over time), representing a hospital capacity in ICU beds, denoted $I_{\max}$.

We provide an example of these parameters for France at the end of lockdown (May 11) in Table 2, using some currently available estimates for the total number who were infected at that date, namely $E_0 + A_0 + I_0 + R_0$:

| Symbol | Description | Value(s) | Reference |
|---|---|---|---|
| $N_0$ | total initial number of people in the population | $6.7 \, 10^7$ | |
| $E_0 + A_0 + I_0 + R_0$ | total initial number of people who had been infected | $5.7 \, 10^6$ | [6] |
| $I_{max}$ | hospital capacity for COVID-19 ICU beds | $10\,000$ | assumed |

Table 2: Example in France for May 11.

## 5.5  Gradual Relaxation Policy

Finally, the user has to choose a policy regarding who is considered to be low-risk and subsequently in a less strict isolation. This is done by choosing an array of proportions of people in low-risk

$$(\rho_0, \rho_1, \ldots, \rho_{N-1}),$$

together with the times at which the policy changes

$$(T_1, \ldots T_{N-1}, T_N)$$

with $T_N$ a final horizon of interest.

For instance, if the user chooses:

- For proportions of people in low-risk : $(\rho_0 = 0.3, \rho_1 = 0.7, \rho_2 = 1)$,

- For policy change times : $(T_1 = 50, T_2 = 120, T_3 = 200)$

Then, this means that :

1. 30% of people are in low isolation from the beginning of the simulation and for 50 days,

2. Then 70% people are in low isolation (that is, 40% people more) during the next $120 - 50 = 70$ days,

3. Finally after 120 days there are no more people recommended to be in strict isolation.

The ODE is integrated up until day 200.

## 5.6 Example

In this section, we present the graphical outputs produced by a simulation run. In order to do so, we have to set the parameters involved in the model and presented above.

### 5.6.1 Population Classification and Accuracy of Risk Prediction Score

For the discriminative capacity of the risk score classifier to differentiate people, we choose for instance $AUC = 0.96$. As we assume that the underlying population probability densities are symmetric Beta distributions with $\alpha_n = \beta_p = 2$, it then translates into the following Beta distribution parameters:

- Probability density of people with severe symptoms upon infection, as a function of the risk score

  - $\alpha_p = 4.5$
  - $\beta_p = 2$

- Probability density of people with severe symptoms upon infection, as a function of the risk score

  - $\alpha_n = 2$
  - $\beta_n = 4.5$

### 5.6.2 Population Contact behavior

Let us consider the following behavioral parameters:

- $\delta_r = 0.1$

- $\delta_c = 0.6$

### 5.6.3 Gradual Relaxation Policy

Let us keep the gradual relaxation policy introduced above. In other words, we choose

- times at which the policy changes: $(T_1 = 50, T_2 = 120, T_3 = 200)$

- proportion ratios of people in the low-risk segment over time: $(\rho_0 = 0.3, \rho_1 = 0.7, \rho_2 = 1)$,

### 5.6.4 Population Characteristics

We choose the following parameters for the

- Initial population size $= 67\,000\,000$,

- Population with moderate symptoms $p = 99.5\%$,

- ICU capacity $= 10\,000$ beds

### 5.6.5 Initial Conditions for the Virus Spread

As for the initial conditions of the virus spread, we choose the following parameters:

- Initially immune population $= 5\,690\,000$

- Initially infected population $= 10\,000$

- Initial number of people in ICU $= 4\,000$

### 5.6.6    Simulation Graphical Outputs

All the model parameters have now been chosen. A simulation run then outputs the following figures[2].
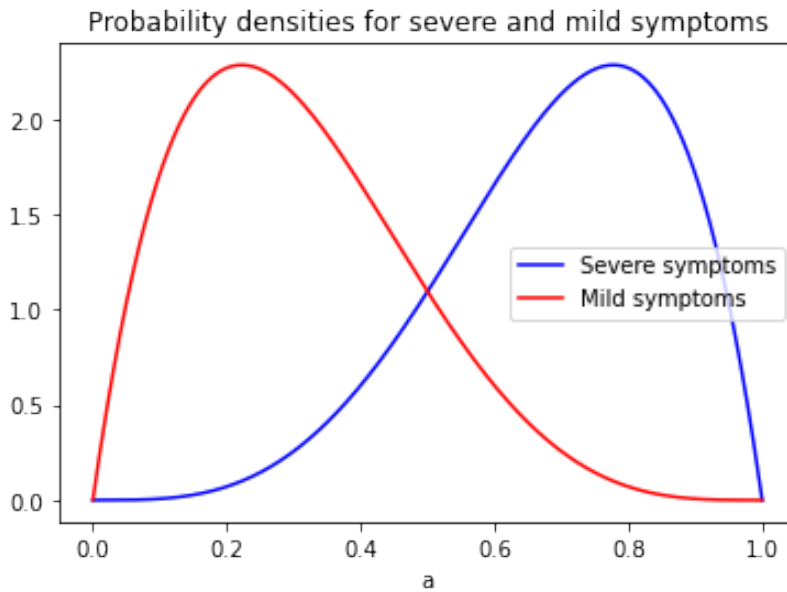


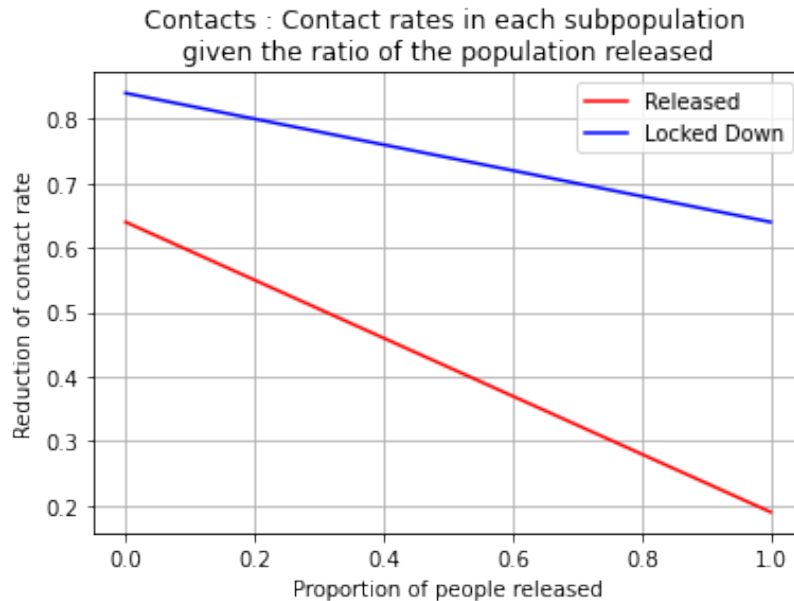Figure 2: Example of population distributions as a function of the personalized risk prediction score.



Figure 3: Contact rates as a function of the proportion of the population released $\rho$, with the choice $\delta_r = 0.1$, $\delta_c = 0.6$.

- Figure 2 shows probability densities for both populations with severe and moderate symptoms according to the personalized risk prediction score,

- Figure 3 shows, given the choice for $\delta_r$, $\delta_c$, how contact rates for released and confined populations evolve as a function of $\rho$,

---

[2]Please find the notebook used to generate the figures for the current example here: https://reine.cmla.ens-cachan.fr/boulant/seair/blob/paper_ipol/SIR_Score_Final_parameters.ipynb

OLIVIER BOULANT, MATHILDE FEKOM, CAMILLE POUCHOL, THEODOROS EVGENIOU, ANTON OVCHINNIKOV, RAPHAËL PORCHER, NICOLAS VAYATIS
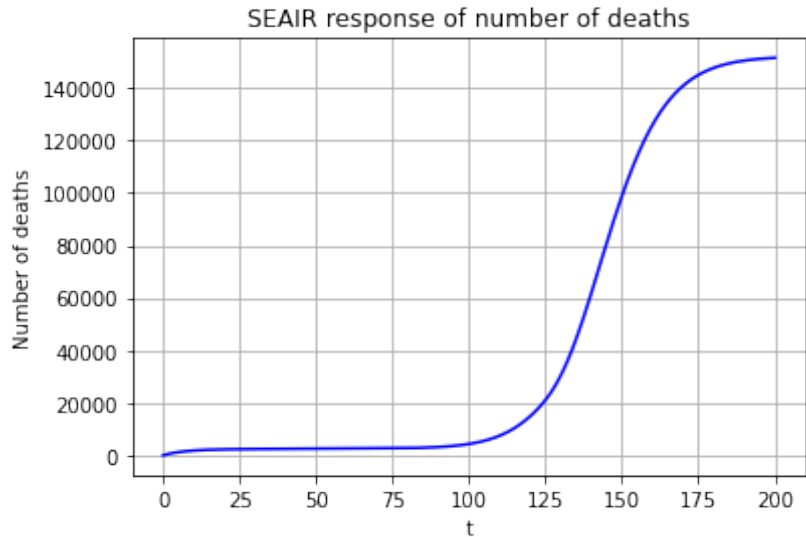
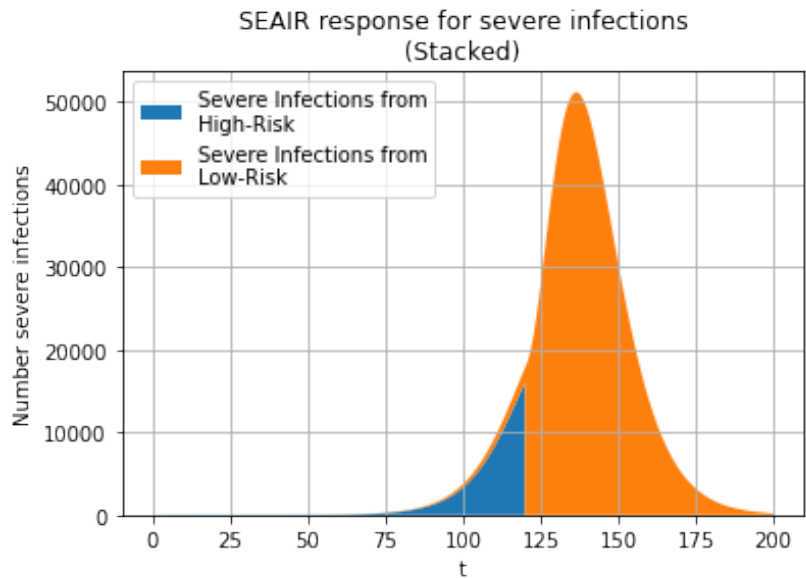

Figure 4: Example response for number of deaths.



Figure 5: Example response for number of severe infections.

- Figure 4 shows the number of deaths over time,

- Figure 5 shows the number of severe infections output by the model over time. On this graph, one can see severe infections from people identified as high risk (blue) as well as from people identified as low risk (orange). Y-axis is stacked, meaning for instance that at time 100 days from simulation start (when 70% of the population is in low isolation), the vast majority of severe infections come from people identified as high risk,

- Figure 6 shows the number of people in ICU over time. The horizontal line of the ICU capacity shows the chosen ICU capacity $I_{\max}$. Then are displayed the number of patients in ICU from people identified as high risk (blue), as well as from people identified as low risk (orange). Y-axis is stacked,

- Figure 7 shows the ratio between the number of recovered people with moderate infections to the total population size. Recovered people with moderate infection from high risk are displayed in
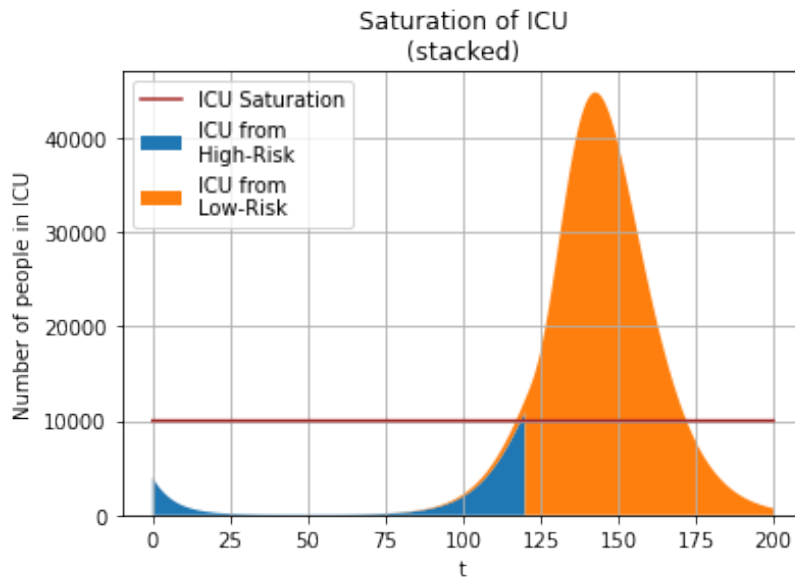
164

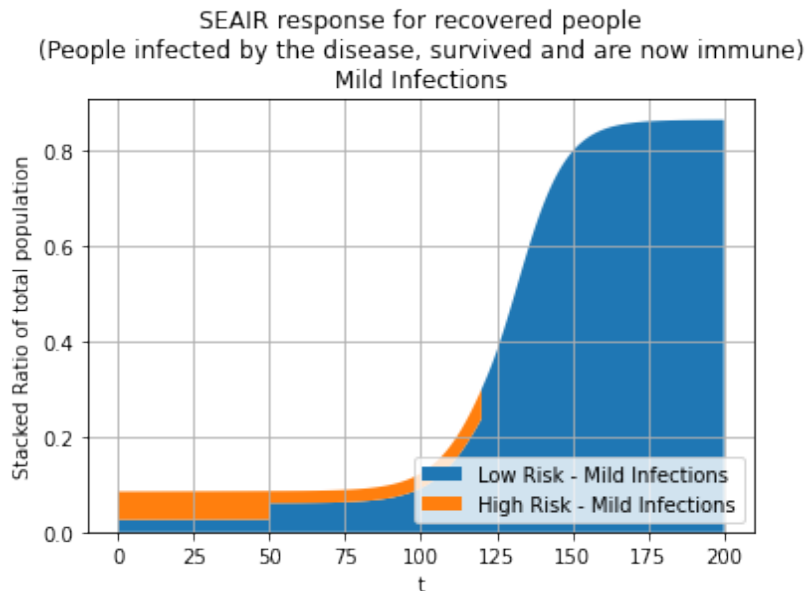Figure 6: Example response for number of ICU admissions.



Figure 7: Example response for number of recovered people with moderate infections.

blue, whereas recovered people with moderate infection from low risk are displayed in orange. For instance, at time 75 days from simulation start (when 70% of the population is in low isolation), ∼10% of the total population has recovered from moderate infection (infected by the disease with moderate symptoms, survived and are now immune),

- Figure 8 shows the ratio between the number of recovered people with severe infections to the total population size. Recovered people with severe infection from high risk are displayed in blue, whereas recovered people with severe infection from low risk are displayed in orange. For instance, at time 75 days from simulation start (when 70% of the population is in low isolation), ∼0.005% of the total population has recovered from severe infection (infected by the disease with moderate symptoms, survived and are now immune).
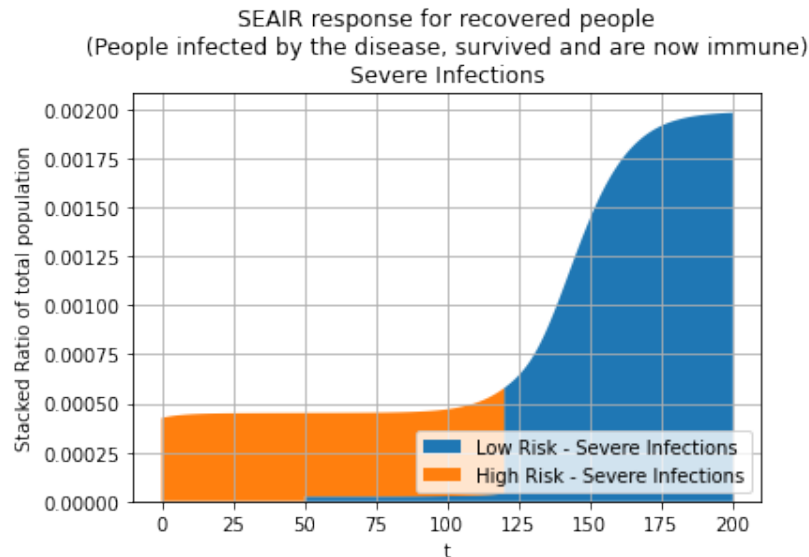
165

Figure 8: Example response for number of recovered people with severe infections.

# References

[1] THE OPENSAFELY COLLABORATIVE, E. WILLIAMSON, A.J. WALKER, K.J. BHASKARAN, S. BACON, C. BATES, C.E. MORTON, H.J. CURTIS, A. MEHRKAR, D. EVANS, P. INGLESBY, J. COCKBURN, H.I. MCDONALD, B. MACKENNA, L. TOMLINSON, I.J. DOUGLAS, C.T. RENTSCH, R. MATHUR, A. WONG, R. GRIEVE, D. HARRISON, H. FORBES, A. SCHULTZE, R.T. CROKER, J. PARRY, F. HESTER, S. HARPER, R. PERERA, S. EVANS, L. SMEETH, AND B. GOLDACRE, *OpenSAFELY: factors associated with COVID-19-related hospital death in the linked electronic health records of 17 million adult NHS patients*, medRxiv, (2020). https://doi.org/10.1101/2020.05.06.20092999.

[2] L. DI DOMENICO, G. PULLANO, C.E. SABBATINI, P-Y. BOËLLE, AND V. COLIZZA, *Impact of lockdown in Île-de-France and possible exit strategies*, medRxiv, (2020). https://doi.org/10.1101/2020.04.13.20063933.

[3] R. DJIDJOU-DEMASSE, Y. MICHALAKIS, M. CHOISY, M.T. SOFONEA, AND S. ALIZON, *Optimal COVID-19 epidemic control until vaccine deployment*, medRxiv, (2020). https://doi.org/10.1101/2020.04.02.20049189.

[4] T. EVGENIOU, M. FEKOM, A. OVCHINNIKOV, R. PORCHER, C. POUCHOL, AND N. VAYATIS, *Epidemic Models for Personalised COVID-19 Isolation and Exit Policies Using Clinical Risk Predictions*, medRxiv, (2020). https://doi.org/10.1101/2020.04.29.20084707.

[5] F. PEDREGOSA, G. VAROQUAUX, A. GRAMFORT, V. MICHEL, B. THIRION, O. GRISEL, M. BLONDEL, P. PRETTENHOFER, R. WEISS, V. DUBOURG, J. VANDERPLAS, A. PASSOS, D. COURNAPEAU, M. BRUCHER, M. PERROT, AND E. DUCHESNAY, *Scikit-learn: Machine learning in Python*, Journal of Machine Learning Research, 12 (2011), pp. 2825–2830.

[6] H. SALJE, C. TRAN KIEM, N. LEFRANCQ, N. COURTEJOIE, P. BOSETTI, J. PAIREAU, A. ANDRONICO, N. HOZÉ, J. RICHET, C-L. DUBOST, Y. LE STRAT, J. LESSLER, D. LEVY-BRUHL, A. FONTANET, L. OPATOWSKI, P-Y. BOELLE, AND S. CAUCHEMEZ, *Estimating the burden of SARS-CoV-2 in France*, Science, (2020). https://doi.org/10.1126/science.abc3517.