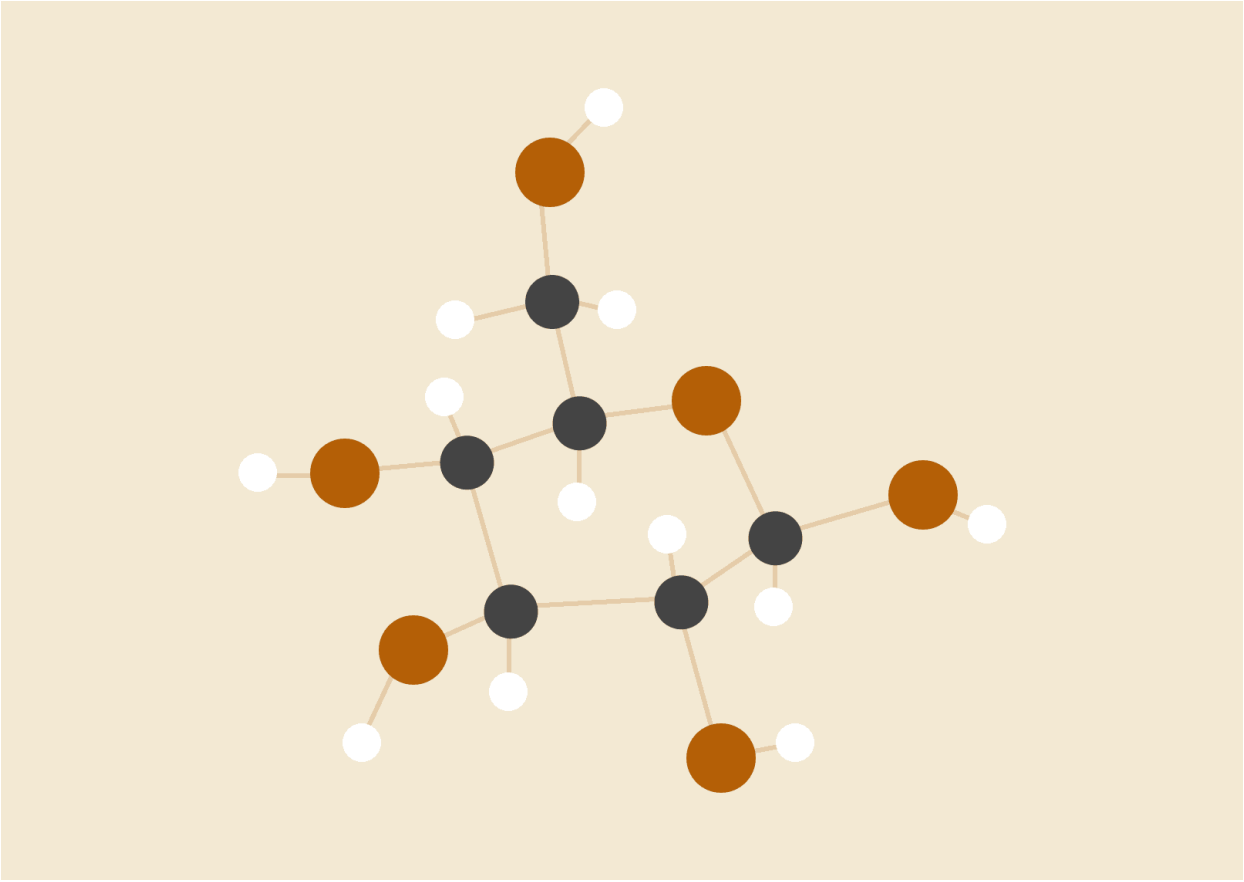
MODELLING THE EFFECTS OF COVID - 19 VACCINES



INTRODUCTION

The World Health Organization (WHO) declared the SARS-CoV-2 (COVID-19) outbreak in Wuhan to be a pandemic on March 11, 2020. Since then, COVID-19 has become a serious global health threat due to its rapid spread, transmission through asymptomatic infected individuals and complex epidemiological dynamics. As of May 2021, already more than 3 million lives have been lost due to the virus. The spread of SARS-CoV-2 has thus far been extremely difficult to contain.

By the end of 2020, the successful development of effective vaccines and the onset of their widespread distribution in most of the world's countries, was hailed as the decisive mean to contain the pandemic. However, important questions linger on whether the vaccination effort will succeed in effectively eradicating the disease. The appearance and wide spread of more contagious SARS-Cov-2 strains, the onset and scale of the vaccine deployment and high levels of vaccine hesitancy/denial in the society, are among the key factors hindering the vaccination effort and the achievement of herd immunity. Modeling the impact of these key factors on the evolution of the pandemic is of critical importance for assessing the vaccination effectiveness against it.

In studying past epidemics, scientists have systematically applied "random mixing" compartmental models which assume that an infectious individual can spread the disease to any susceptible member of the population before becoming recovered or removed, as originally considered by Kermack and

McKendrick. These models constrain the total population in compartments by considering stages of the infection and flows among them.

In the present study we propose a new model named SAIVR, which incorporates two important characteristics of the COVID-19 epidemic, namely the considerable transmission of the disease by asymptomatic infected individuals and the vaccination campaign with World Health Organization (WHO) approved vaccines. More recent modeling approaches involve agent-based simulations, heterogeneous social networks, and Bayesian inference models. Although a large number of research studies are currently investigating the COVID-19 epidemiological characteristics, we believe that a simple but efficient model, which can capture the basics of the complex behavior of the pandemic including the vaccine roll-out, can offer useful guidance for the pandemic's near-term and longer-term evolution. By using a recently developed semi-supervised machine learning approach we systematically reproduced the pandemic dynamics during the 2021 spring in several different countries. We then used the model to assess the importance of a rapid vaccination campaign to prevent future outbreaks driven by more infectious variants.

The SAIVR model

One of the first attempts to mathematically describe the spread of an infectious disease is due to Kermack and McKendrick [1]. In 1927 they introduced the so-called Susceptible-Infectious-Removed (SIR) model. The SIR model describes the dynamics of a (fixed) population of N.

Individuals split into three compartments:

S(t) is the Susceptible compartment that counts the number of individuals susceptible but still not infected by the disease;

I(t) is the Infectious compartment that counts the number of infectious individuals;

R(t) is the Removed compartment. It represents the number of those who can no longer be infected either because they recovered and gained long-term immunity or because they passed away.

The model involves two positive parameters, β and γ which govern the flow from one compartment to the other:

β is the transmission rate or effective contact rate of the disease: an infected individual comes into contact with β other individuals per unit time (the fraction that are susceptible to contracting the disease is S/N);

γ is the removal rate. γ^{-1} is the mean number of days who is infected spends in the Infectious compartment.

The SIR model obeys the following system of ordinary differential equations (ODE):

$$\frac{dI}{dt} = \beta I \frac{S}{N} - \gamma I \tag{1a}$$

$$\frac{dS}{dt} = -\beta I \frac{S}{N} \tag{1b}$$

$$\frac{dR}{dt} = \gamma I \tag{1c}$$

Since they often avoid contact tracing due to the absence of symptoms, they can spread the disease while remaining undetected. Furthermore, in December 2020 a global vaccination campaign has started. Vaccinating is a safe way to transfer people from the Susceptible to the Removed compartment bypassing the Infectious one thus reducing the likelihood of an outbreak.

We emphasize that in the SAIVR model the fully vaccinated population is already included in the model as part of the 'Removed' compartment, while the 'Vaccinated' population is an additional compartment that adds further predictive power to the model by taking into consideration the fact that it takes a few weeks (often due to the necessity of a second vaccine shot) to reach full immunization; full immunization, in the context of the model, is equivalent to moving a person to the 'Removed' compartment.

The SAIVR model ODEs read:

$$\frac{dI}{dt} = \beta_1 I\frac{S}{N} + \alpha_2 A\frac{S}{N} + \zeta I\frac{V}{N} - \gamma I, \tag{2a}$$

$$\frac{dA}{dt} = \alpha_1 A\frac{S}{N} + \beta_2 I\frac{S}{N} + \eta A\frac{V}{N} - \gamma A, \tag{2b}$$

$$\frac{dS}{dt} = -\beta I\frac{S}{N} - \alpha A\frac{S}{N} - \delta\frac{S}{N} + (1-\lambda)\,\epsilon V, \tag{2c}$$

$$\frac{dV}{dt} = \delta\frac{S}{N} - \eta A\frac{V}{N} - \zeta I\frac{V}{N} - \epsilon V, \tag{2d}$$

$$\frac{dR}{dt} = \gamma I + \gamma A + \lambda\epsilon V. \tag{2e}$$

The compartment inter-dependencies and flow are presented in Fig. 1 .

The parameters of the SAIVR model are the following:

-$\beta_1$describes the rate at which individuals are exposed to symptomatic infection. An infected symptomatic individual comes into contact and infects β_1 susceptible individuals per unit time;

-α_1 is the asymptomatic infection rate. An infected asymptomatic individual comes into contact with α_1 susceptible individuals per unit time;

-β_2 describes the rate at which susceptible individuals become asymptomatic infected after entering in contact with a symptomatic individual;

-α_2 describes the rate at which who's susceptible becomes symptomatic after entering in contact with an asymptomatic individual;

-γ retains the same meaning as in the SIR model, representing the mean removal rate. γ^{-1} is the mean amount of time individuals spend either in the Infectious or Asymptomatic compartments;

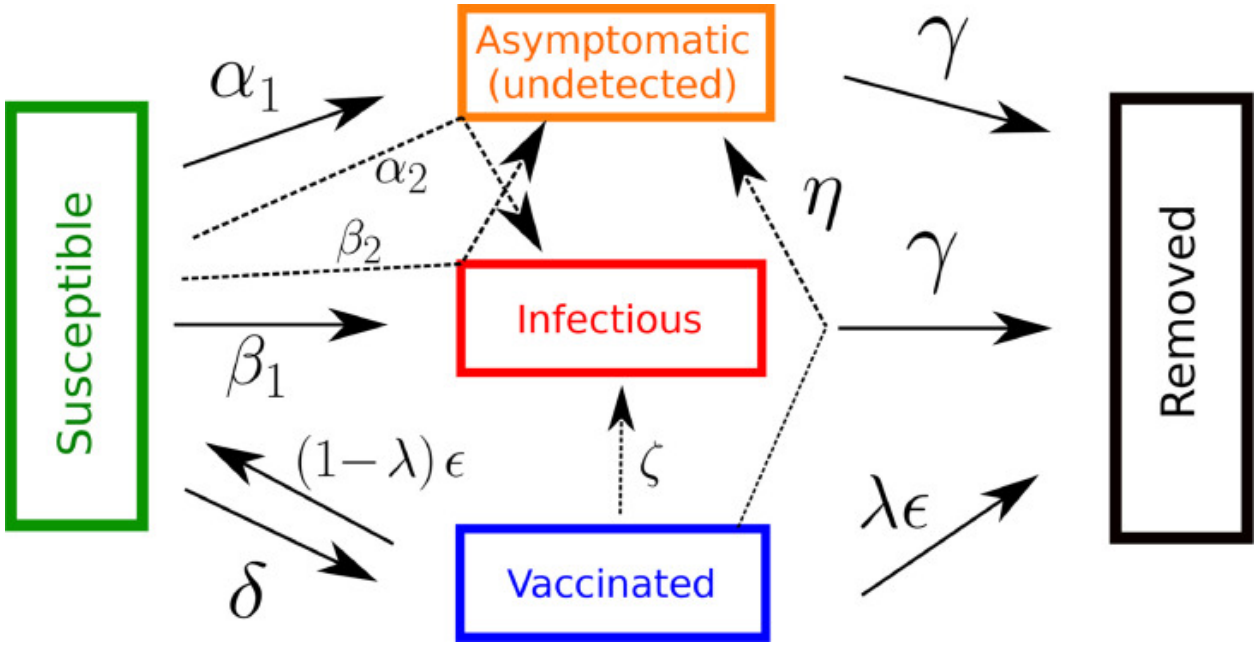
-ζ is the rate at which a vaccinated (but still not immune) individual enters in contact with a symptomatic infectious;

-η describes the transmission rate at which who's asymptomatic comes into contact and infects vaccinated (but still not immune) individuals;

-δ is the first shot vaccination rate;

-λ is the vaccine efficacy;

-ϵ^{-1} is the mean amount of time an individual spends in the Vaccinated compartment before reaching immunity and moving to the Removed compartment.

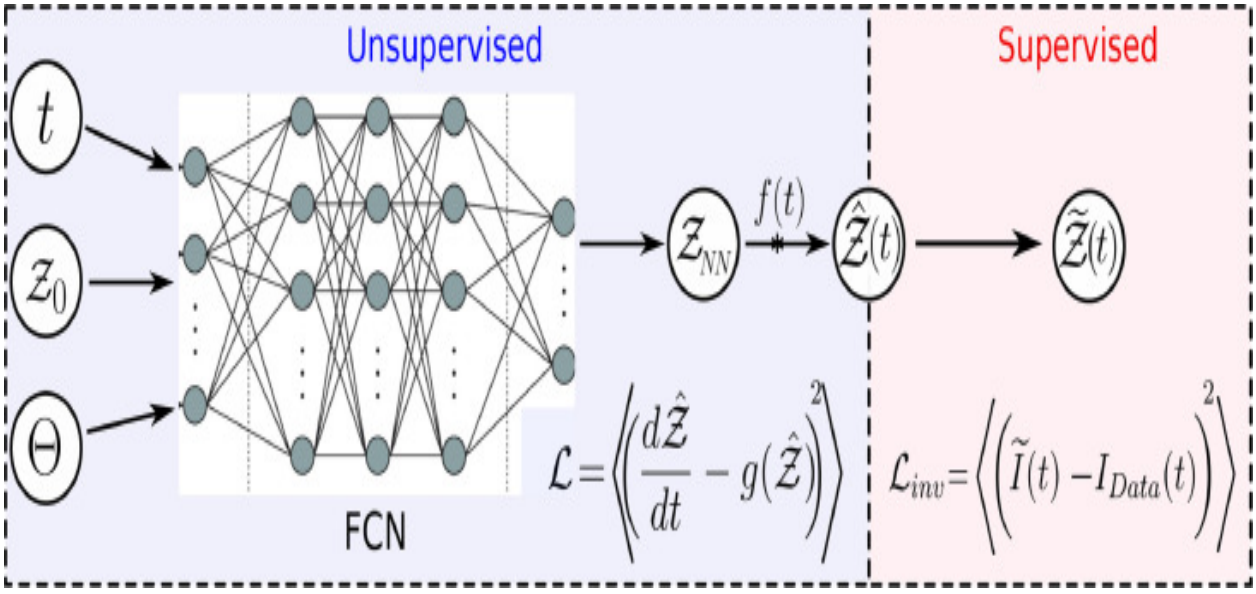


Countries and states do not respond to the disease as static entities passively facing the pandemic. They react by actively imposing (and relaxing) restrictive measures, learning how to effectively treat the infected, adjusting social interactions and by launching vaccination campaigns. Finally, the virus itself evolves in more infectious variants.

Country-specific parameters can be obtained by fitting the SAIVR model to a selected infectious wave occurred in a given country. SAIVR has 14 adjustable parameters or initial conditions that needs to be estimated; given the scarcity of data (only the infectious and vaccinated populations are known) optimizing them presents a challenging problem. To address this, we either fixed some of them or employed a novel fitting method based on semi-supervised neural networks, which we present in the following section.

Solving the SAIVR model with neural networks

In order to apply the SAIVR model we need a realistic estimate of the parameters and initial conditions for the system of Eq. (2). To obtain them we employed machine learning, a powerful method that has been extensively used for disease modeling and dynamical system forecasting. Our approach employs a semi-supervised procedure that determines the optimal set of initial conditions and parameters of the SAIVR model, yielding solutions that best fit a given data set. A sketch of this procedure is shown in Fig. 2 .



Unsupervised learning

The unsupervised part (blue box) consists of a data-free Neural Network (NN) that is trained to discover solutions for an ODE system of the form:

$$\frac{d\mathscr{Z}}{dt} = g(\mathscr{Z}), \qquad \mathscr{Z}(t=0) = \mathscr{Z}_0. \tag{3}$$

where $\mathscr{Z} = (S(t), A(t), I(t), V(t), R(t))$ and $g(\mathscr{Z})$ is given in Eq. (2). The NN takes as an input a time sequence t, a set of initial conditions $\mathscr{Z}_0$, and modeling parameters Θ.

As we'll see in the following, t is the set of days involved in a given epidemic wave going from t_0 to $t_0 + \Delta t$, Z_0 are the initial compartment populations and Θ some parameters of the SAIVR model. The initial conditions and parameters are randomly sampled at each iteration n over predefined intervals called bundles [19], [21], so that the network learns an entire family of solutions. The inputs propagate through the network until an output vector $\mathscr{Z}_{NN}$ of the same dimensions as the target solutions $\mathscr{Z}$ is produced. The learned solutions $\widehat{\mathscr{Z}}$ satisfy the initial conditions identically by considering parametric solutions of the form:

$$\widehat{\mathscr{Z}} = \mathscr{Z}_0 + f(t)(\mathscr{Z}_{NN} - \mathscr{Z}_0) \tag{4}$$

where $f(t)=1-e^{-t}$. The loss function:

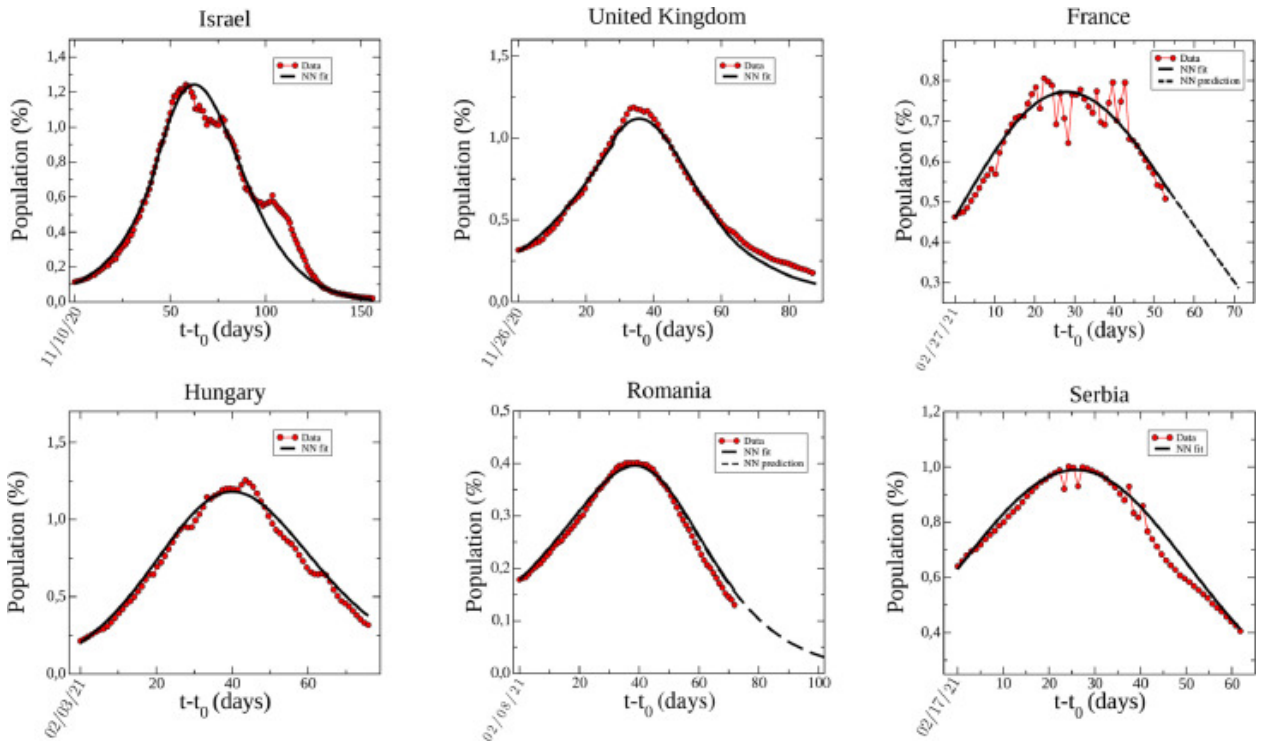
$$L=\langle(dZ/dt-g(Z))^2\rangle \tag{5}$$

solely depends on the network predictions averaged ($\langle\ldots\rangle$) over all the iterations n, providing an unsupervised learning framework. Time derivatives are computed using the automatic-differentiation and back-propagation techniques.

Fitting a dataset

where $I_{Data}(t)$ is the infectious population of a given country/state and $\tilde{I}(t)$ is its NN fit.

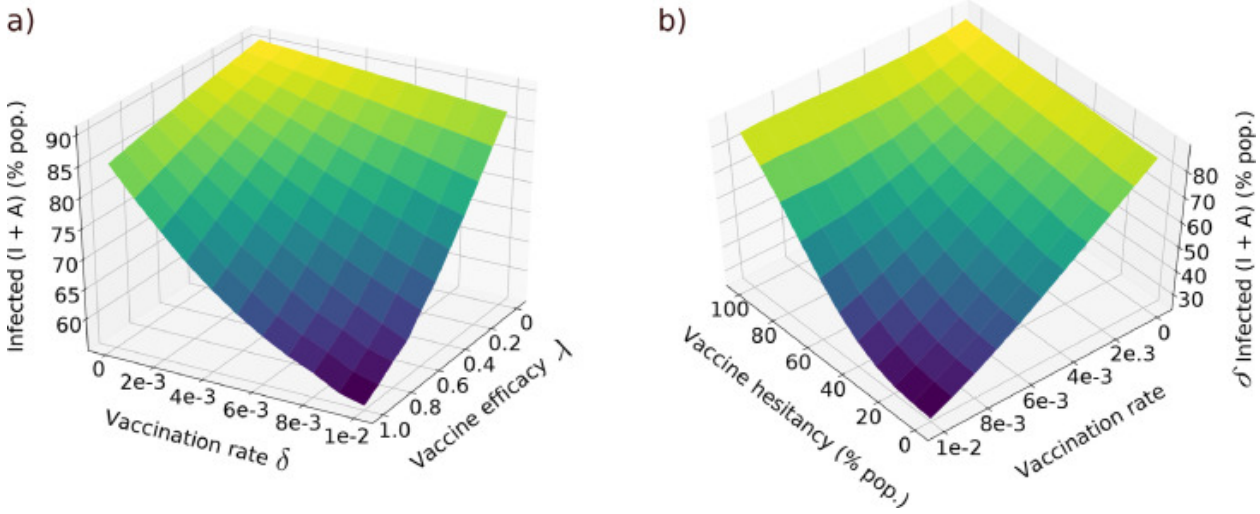
The machine learning approach presented in this work provides numerical solutions to a nonlinear system of ODEs without statistical error (no data is used in the first part of the process). The supervised part is only learning what are the best parameters/conditions of the SAIVR model that fits given data, so the statistical error of the noisy data does not affect the final outcome of the process.



Vaccination efficacy and hesitancy

Fig. 5 presents the total infected (I+A) population under increasing values of vaccination onset times (T_0), vaccination daily rates (δ), vaccine efficacy (λ), and vaccine hesitancy/denial population percentage. In the top panel, the total infected population is shown as a function of the vaccination rate δ and vaccine efficacy λ. As can be seen, even vaccines with a relatively low efficacy can rapidly reduce the infected population. we show how the number of those infected evolves as a function of δ and the percentage of the

population that avoids getting vaccinated. These findings suggest that vaccine hesitancy, which accounts for a significant proportion of the population might seriously threaten the reach of herd immunity, especially if the situation is worsened by the appearance of more infectious COVID-19 strains.



Conclusions

Compartmental models are efficient tools to deal with the time evolution of disease outbreaks. They provide us with useful intuition on the impact of non-pharmaceutical intervention in decreasing the number of infectious incidence rates.

In this work, we have augmented the classic SIR model with the ability to accommodate asymptomatic transmission and vaccinated individuals. The SAIVR model is a straightforward deterministic model, which does not take into consideration age, gender or geographic clustering. Despite this, its simplicity and the insights it offers on how key epidemiological variables affect individuals are among its main strengths. Its power also lies in the fact that, as factors such new variants are added to the model, it is easy to adjust its parameters and provide with best fit curves between the data and the

model predictions.

Since the inclusion of the Asymptomatic and Vaccinated compartments enlarged the number of parameters and initial conditions of the model, we employed a novel semi-supervised framework to estimate most of them. An unsupervised neural network solves the model's differential equations over a range of parameters and initial conditions. A supervised approach then incorporates data and determines the optimal initial conditions and modeling parameters that best fit the 27 epidemic curves considered. As expected due to the heterogeneity of the countries sample, the resulting parameters fit are dissimilar although they follow similar trends.

We used these results to shed light on the impact of the vaccination campaign on the future of the pandemic. We pointed out how vaccine hesitancy is one of the most important hurdles of the campaign and further efforts should be done to support people and give them correct information about vaccines. Because of this, vaccinating the critical number of people that have to be immune in order to prevent future outbreaks (i.e. herd immunity), is likely to be out of reach. Widely circulating coronavirus variants are also a threat as they move the herd immunity threshold to higher values. This points out the importance of rapidly reducing the infection rate by any means, such as by imposing restrictive measures in case highly infective new variants appears before the herd immunity threshold is reached. These results manifest the need for continuing the vaccination effort and the drive for achieving high vaccination coverage in order to contain outbreaks generated by new and possibly more infectious variants.