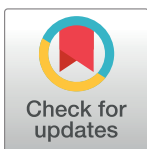# PLOS ONE

RESEARCH ARTICLE

Understanding the implications of under-reporting, vaccine efficiency and social behavior on the post-pandemic spread using physics informed neural networks: A case study of China

Samiran Ghosh[1], **Alonso Ogueda-Oliva**[2], **Aditi Ghosh**[3]*, **Malay Banerjee**[1], **Padmanabhan Seshaiyer**[2]

1 Indian Institute of Technology Kanpur, Kanpur, India, **2** George Mason University, Fairfax, VA, United States of America, **3** Texas A&M University-Commerce, Commerce, TX, United States of America

* amathematics@gmail.com

Abstract

In late 2019, the emergence of COVID-19 in Wuhan, China, led to the implementation of stringent measures forming the zero-COVID policy aimed at eliminating transmission. Zero-COVID policy basically aimed at completely eliminating the transmission of COVID-19. However, the relaxation of this policy in late 2022 reportedly resulted in a rapid surge of COVID-19 cases. The aim of this work is to investigate the factors contributing to this outbreak using a new SEIR-type epidemic model with time-dependent level of immunity. Our model incorporates a time-dependent level of immunity considering vaccine doses administered and time-post-vaccination dependent vaccine efficacy. We find that vaccine efficacy plays a significant role in determining the outbreak size and maximum number of daily infected. Additionally, our model considers under-reporting in daily cases and deaths, revealing their combined effects on the outbreak magnitude. We also introduce a novel Physics Informed Neural Networks (PINNs) approach which is extremely useful in estimating critical parameters and helps in evaluating the predictive capability of our model.

1 Introduction

COVID-19, caused by the novel coronavirus SARS-CoV-2, first emerged in Wuhan, Hubei province, China, in late 2019. As the virus continued to spread, China implemented stringent measures to control its transmission. Since then, China continued with the same methods, making them part of its zero-COVID policy, even after the availability of vaccines. Zero-COVID policy refers to an approach aimed at completely eliminating the transmission of COVID-19 within a specific region or country. This approach typically involves strict measures such as extensive testing, contact tracing, quarantine, travel restrictions and lockdowns to prevent the spread of the virus.

The zero-COVID policy in China has both positive and negative aspects. It aims to control the disease, protect public health, prevent healthcare system overload, and maintain economic stability. However, it also leads to economic disruptions, social and mental health impacts, potential human rights concerns, and faced challenges in feasibility and sustainability [1]. At the end of 2022, when the zero-COVID policy was relaxed gradually in China, it faced a rapid spread of the infection [2]. While we do not have access to official real-time data, there have been reports in the media suggesting a significant increase in daily COVID-19 cases in China after the relaxation of their zero-COVID policy [3–6]. According to these sources, the number of cases has allegedly surged to the magnitude of approximately ten millions per day [3, 4]. This sudden surge in epidemic progression forced researchers to think about the possible reasons behind it.

The implementation of a zero-COVID policy aims to eliminate the transmission of COVID-19, resulting in fewer infections. However, this approach may limit the development of acquired immunity in the population, leaving a significant portion susceptible to the infection. An increase in the total level of immunity within a population results in a reduction of the proportion of individuals who are susceptible to the disease. The total level of immunity in a population comes from two sources:

1. Vaccine induced immunity: Immunity obtained through vaccination of the individuals.

2. Acquired Immunity: 'Acquired immunity' refers to the adaptive immunity developed within a recovered individual after recovery from the infection.

Since, zero-COVID policy resulted in a fewer infections, the contribution of the acquired immunity in the total level of immunity is negligible. Consequently, the proportion of susceptible individuals is comparatively high. The vaccine induced immunity is the major contributor to reduce the proportion of susceptible population. However, the waning of immunity can increase the proportion of susceptibility over the time. This waning of immunity solely depends upon the efficacy of the vaccine. In the context of COVID-19, the level of immunity acquired through prior infection is long lasting compared to the immunity obtained through vaccination.

Vaccine efficacy refers to the effectiveness of a vaccine in preventing a disease. It is typically expressed as a percentage and represent the reduction in disease incidence among vaccinated individuals compared to a similar group of unvaccinated individuals. In the context of COVID-19, multiple vaccines have been developed in different countries by different manufacturers. Different COVID-19 vaccines employ different mechanisms to trigger an immune response. For example, mRNA based vaccines, such as those developed by Pfizer-BioNTech and Moderna, introduce genetic material into cells to produce viral proteins. In contrast, vector-based vaccines like the AstraZenca and Johnson & Johnson vaccines use weakened viruses to deliver viral proteins. These variations in vaccine mechanisms can lead to different types of vaccine efficacies [7, 8].

There were several domestic vaccine candidates in China, among which, most notably, China used Sinopharm-Beijing, Sinopharm-Wuhan, Sinovac, CanSinoBio [1, 9, 10]. The efficacies of these domestic vaccines are less as compared to other vaccines like Pfizer, Moderna, Novavax etc. (see Fig 2 in [11]). Vaccine efficacy can also be a reason behind such a large epidemic peak in the beginning of 2023 in China. Vaccine efficacy, along with its long-lasting protection, plays a crucial role in shaping the course of a new outbreak. A highly effective vaccine not only prevents infection and reduces illness, severity, but also provides durable immunity. Long-lasting vaccine efficacy ensures that individuals remain protected against the disease for an extended period, thereby contributing to the sustained control of the epidemic. By

maintaining high level of immunity in the population, the vaccine helps prevent resurgence of the virus.

Under-reporting in daily cases and death data refers to the situation where the actual number of cases or deaths due to the disease is higher than what is officially reported. Under-reporting in daily cases and deaths in China has been a subject of concern and scrutiny, particularly, during disease outbreaks [12, 13]. This can have a significant impact on the progression of an epidemic by influencing human behavior. When the true magnitude of an epidemic is not accurately reflected in the official statistics, it can create a false perception of the severity of the situation. This can lead to a sense of complacency or decreased vigilance among the population, as individuals may underestimate the risk of infection and adopt less stringent preventive measures. Consequently, under-reporting can result in increased transmission of the virus as people may be less likely to adhere to essential precautions. This can further fuel the spread of the epidemic. Accurate and transparent reporting of daily cases and deaths is crucial to ensure that individuals maintain a realistic understanding of the situation, leading to appropriate behavior that helps curb the transmission of the virus.

In our study, we aim to investigate the factors contributing to a large outbreak, such as the one observed in China, in the beginning of 2023, and to achieve this, we develop a modified SEIR epidemic model [14–16]. One of the important component of our SEIR-type model is that the model incorporates a time-dependent level of immunity, considering both the number of vaccine doses administered and the efficacy of the vaccine. Notably, we consider the vaccine efficacy as a function of the time-post-vaccination because the efficacy is not same all the time after vaccination.

Our computational results demonstrate that the post-vaccination dependent vaccine efficacy plays a significant role in determining the size of the epidemic outbreak. Furthermore, we enhance the precision of our model by integrating *imitation dynamics* into our framework. These dynamics capture individuals' behavior in switching strategies for reporting infections. We also consider that the decision to switch reporting strategies may depend on the perception of disease severity, which is influenced by available death data. Our modelling results reveal that although under-reporting of deaths may not directly impact disease incidence, it can create a false perception of reduced disease severity, leading to implicit under-reporting of daily cases. The combined effects of under-reporting in daily cases and deaths can influence the magnitude of the outbreak.

Along with new ideas on modeling, we also introduce neural network based approaches to quantify parameters in the model learning from the data. Specifically, we use Physics Informed Neural Networks (PINNs) approaches that have been applied to a variety of differential equations [17–19]. Building on this, we also delve into the development of an algorithm based on PINNs in this work applied to the system of differential equations proposed. This algorithm serves the purpose of estimating critical parameters that play a crucial role in understanding the dynamics of the epidemic.

Approaches such as PINNs helps with the realization of the need for artificial intelligence and machine learning (AI/ML) modeling platforms that facilitate discovery of novel biological phenomena, rules, and theories. This work aligns with calls from funding agencies such as the US National Science Foundation that promote the need for mathematical, computational, and biological scientists to work together to develop MODels for Uncovering Rules and Unexpected Phenomena in Biological Systems (MODULUS (https://www.nsf.gov/pubs/2021/nsf21069/nsf21069.jsp)) as well as incorporating human behavior in epidemiological models (IHBEM (https://www.nsf.gov/pubs/2023/nsf23546/nsf23546.htm)).

The goals of our study can be summarized as follows:

- Develop a new SEIR-type epidemic model with time-dependent level of immunity, considering vaccine doses administrated and the time-post-vaccination dependent vaccine efficacy,

- Analyze the significant role of time-post-vaccination dependent vaccine efficacy in determining the size of the outbreak,

- Explore the explicit and implicit effects of the under-reporting in daily cases and deaths on the magnitude of the outbreak and,

- Apply a predictive algorithm to estimate important parameters for understanding the epidemic progression.

The content of the paper is as follows. We discuss the data sources, model formulation and calculate the basic epidemic quantities of the model in section (2). In section (3), we introduce under-reporting in infected and death in the model formulation. Parameter estimation using Physics Informed Neural Networks (PINNs) is discussed in section (4). Computational results are discussed in section (5).

2 Materials and methods

2.1 Data sources

We first describe the data sources utilized in our current study. To examine the vaccination administration in China, we collected the time-series data of vaccination in China, from [10]. The trend in vaccination coverage can be visualized from the Fig 1. Furthermore, to support our modelling results, we gathered data from a secondary source [11], which presents a comparative analysis of vaccine efficacies. The visual representation of these comparisons, displayed in Fig 2, provides valuable insights into the relative effectiveness of different vaccines. Moreover, we used some reliable parameter values associated to our modelling, available in some sources, as referenced in Table 1.

Note that, some of the parameters in Table 1 can be obtained from reliable sources, but, to the best of our knowledge, the rest of the parameters in Table 1 are not available in the existing literature, in the context of COVID-19 in China after the relaxation of the zero-COVID
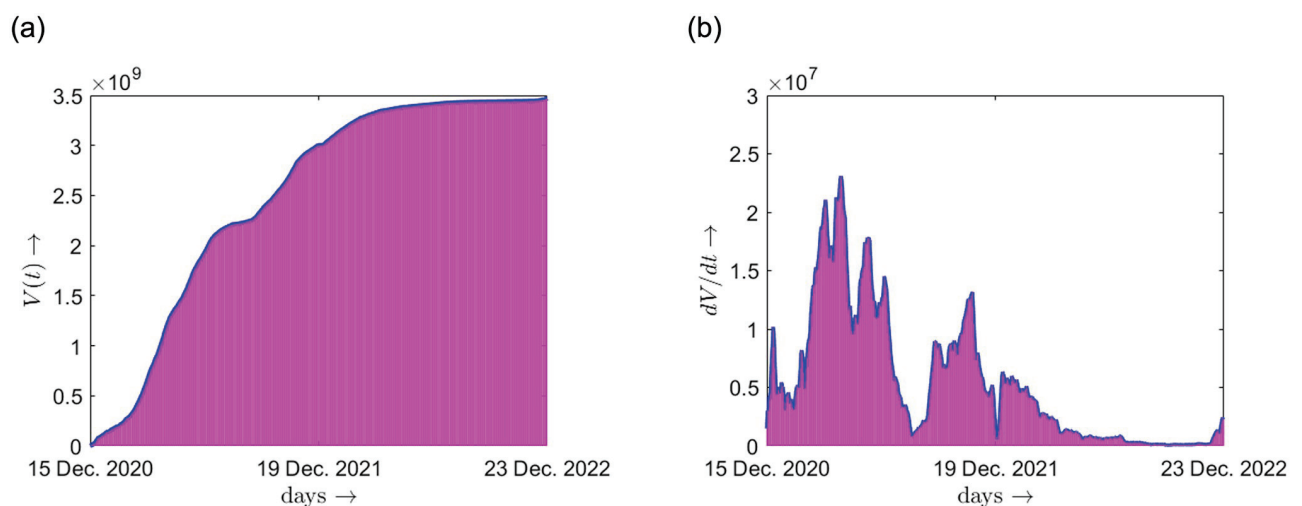
(a) (b)

Fig 1. Real data of vaccination in China [10]. (a) Cumulative vaccination data $V(t)$, (b) Daily vaccination data which approximately gives dV/dt.

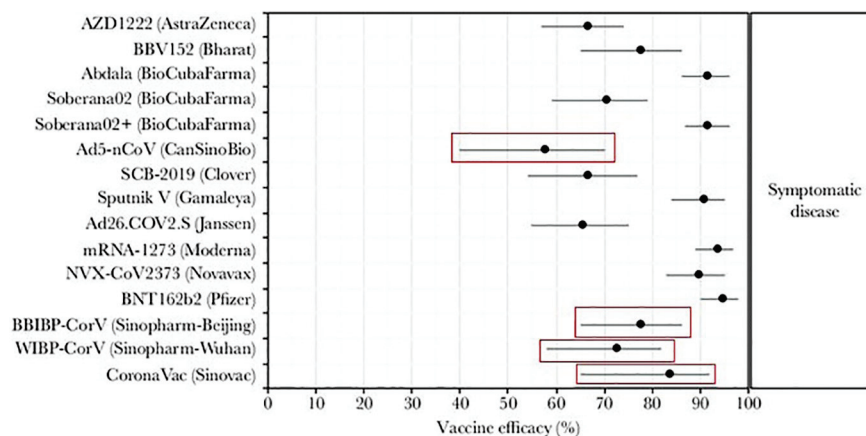https://doi.org/10.1371/journal.pone.0290368.g001

Fig 2. Comparison among the effectiveness of different vaccines ofCOVID-19. In China, mainly Sinopharm-Bejing, Sinopharm-Wuhan, Sinovac, CanSinoBio vaccines were administrated (the red squares) [1, 9, 10]. This figure is taken from [11].

policy. On the other hand, according to some sources [3, 4], after the relaxation of the zero-COVID policy, the number of daily cases in China was in the magnitude of few millions per day. Since, detailed time series data for COVID-19 in China is not available in the literature, with the minimal information available at the sources [3, 4], we initially assumed some of the parameter values which are not available in reliable sources, as indicated in Table 1. Assumption of such parameter values is motivated by the daily number of cases obtained from our model outcome reaches the magnitude of few millions per day (e.g., green curve in Figs 17(c) and 18). However, later in this work, we employed a PINNs algorithm that helped to estimate some of those parameters involved in our model, using the initial assumptions of the parameter values.

Table 1. Parameter values.

Parameters	Description	Value	Units	Source
N	Total population	1.453×10^9	no. of people	[20]
β	Transmission rate	11	day^{-1}	—
σ	Reciprocal of incubation period	1/4.35	day^{-1}	[21]
r	Rate at which exposed individuals become symptomatic	0.227	unit-less	[22]
α	Relative transmission rate of asymptomatic infection	0.43	unit-less	[23]
η_1	Rate of quarantine	0.001	day^{-1}	—
η_2	Rate of hospitalization	0.0689	day^{-1}	[24]
ξ	Rate of hospitalization of quarantined	0.03	day^{-1}	—
δ_I	Recovery rate of infected	0.9975	day^{-1}	[25]
μ_I	Death rate of infected	0.0025	day^{-1}	[25]
δ_A	Recovery rate of asymptomatic	0.9975	day^{-1}	—
δ_H	Recovery rate of hospitalized	0.9975	day^{-1}	—
μ_H	Death rate of hospitalized	0.0015	day^{-1}	—
δ_Q	Recovery rate of quarantined	0.9975	day^{-1}	—
$V(t)$	Daily total vaccination data of China	Fig 1	no. of people	[10]

2.2 Model description

Let us suppose vaccination in China was started at time $t = t_0 = 0$ and the zero-COVID policy was relaxed from $t = t_1$. Since there is negligible number of infection due to zero-COVID policy during the time interval $[t_0, t_1]$, we neglect the acquired immunity due to infection, whereas, during the same time interval, the level of immunity in the population increased due to vaccination. Our main concern is to study the sudden large outbreak in a relatively shorter period of time starting from t_1 (i.e., the time when zero-COVID policy was relaxed), and we assume that those who get recovered after time t_1 do not become susceptible during our time interval of study.

Here we consider following compartments: susceptible (S), exposed (E), asymptomatic (A), infected (I), quarantine (Q), hospitalized (H), recovered (R) and deceased (D) with total population N is supposed to be a fixed constant.

Time dependent growth rates of each compartments, except S and D are given by

$$\frac{dE}{dt} = \frac{\beta S}{N}(I + \alpha A) - \sigma E \tag{1a}$$

$$\frac{dI}{dt} = r\sigma E - (\eta_1 + \eta_2 + \delta_I + \mu_I)I \tag{1b}$$

$$\frac{dA}{dt} = (1 - r)\sigma E - \delta_A A \tag{1c}$$

$$\frac{dQ}{dt} = \eta_1 I - (\xi + \delta_Q)Q \tag{1d}$$

$$\frac{dH}{dt} = \eta_2 I + \xi Q - (\delta_H + \mu_H)H \tag{1e}$$

$$\frac{dR}{dt} = \delta_I I + \delta_A A + \delta_H H + \delta_Q Q. \tag{1f}$$

Let $m(t)$ denote the level of immunity due to vaccination given by

$$m(t) = \frac{1}{N}\int_0^t \phi(t - \eta)V'(\eta)d\eta, \tag{2}$$

where $V'(\eta)$ is nonnegative function which represents number of daily vaccinations and $\phi(\eta)$ is the efficacy of vaccine at the time-post-vaccination η. **Note.** Here, the level of immunity $m(t)$ in Eq (2) is defined only in terms of the vaccine induced immunity. This definition of $m(t)$ does not consider the acquired immunity developed within an individual after recovering from the infection. This is because of the fact that there was negligible number of infection due to zero-COVID policy.

Then the equation for susceptible can be written as:

$$S(t) = N - (E(t) + I(t) + A(t) + Q(t) + H(t) + R(t) + D(t) + Nm(t)). \tag{3}$$

A compartmental flow diagram of the model (1)-(3) is shown in Fig 3. Though the Eq (3) determines $S(t)$ in terms of all other compartments, this relation does not clearly explains how the susceptible becomes exposed due to the interaction with the compartments I and A.
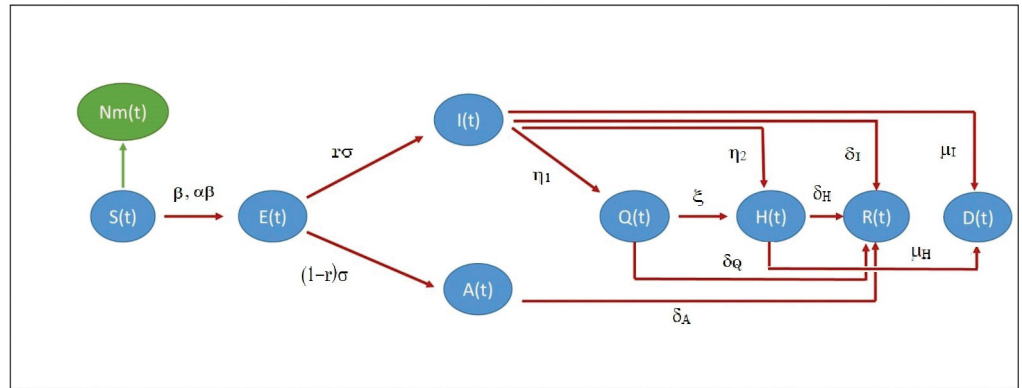
Fig 3. Compartmental flow diagram of the model (1).

Hence, we derive the governing differential equation for the S compartment which is determined by the interaction with the compartments I, A and the level of immunity $m(t)$.

Differentiating both sides of the Eq (2) and using the Leibnitz rule for differentiation under the integral sign, we get

$$\frac{dm(t)}{dt} = \frac{1}{N}\frac{d}{dt}\int_0^t \phi(t-\eta)V'(\eta)d\eta$$

$$= \frac{1}{N}\left[\int_0^t \phi'(t-\eta)V'(\eta)d\eta + \phi(0)V'(t)\right].$$

Assuming $\phi(0) = 0$ we get

$$\frac{dm(t)}{dt} = \frac{1}{N}\int_0^t \phi'(t-\eta)V'(\eta)d\eta. \tag{4}$$

Now, differentiating (3) with respect to t and using the equations in system (1), we get,

$$\frac{dS(t)}{dt} = -\frac{\beta S}{N}(I + \alpha A) - N\frac{dm(t)}{dt}. \tag{5}$$

Substituting the expression of dm/dt from (4) in (5), we get,

$$\frac{dS(t)}{dt} = -\frac{\beta S}{N}(I + \alpha A) - \int_0^t \phi'(t-\eta)V'(\eta)d\eta. \tag{6}$$

The Eq (6) is the governing equation for the rate of change of the S compartment which clearly explains how the rate of change of the S compartment depends on (i) the interaction of the individuals in S compartment with the individuals in the compartments I, A, (ii) the vaccine efficacy function $\phi(t)$ and (iii) the number of vaccinated individuals $V(t)$.

The growth rate of $D(t)$ is given by

$$\frac{dD}{dt} = \mu_I I + \mu_H H. \tag{7}$$

Hence the full system can be written as follows:

$$\frac{dS}{dt} = -\frac{\beta S}{N}(I + \alpha A) - \int_0^t \phi'(t - \eta)V'(\eta)d\eta \tag{8a}$$

$$\frac{dE}{dt} = \frac{\beta S}{N}(I + \alpha A) - \sigma E \tag{8b}$$

$$\frac{dI}{dt} = r\sigma E - (\eta_1 + \eta_2 + \delta_I + \mu_I)I \tag{8c}$$

$$\frac{dA}{dt} = (1 - r)\sigma E - \delta_A A \tag{8d}$$

$$\frac{dQ}{dt} = \eta_1 I - (\xi + \delta_Q)Q \tag{8e}$$

$$\frac{dH}{dt} = \eta_2 I + \xi Q - (\delta_H + \mu_H)H \tag{8f}$$

$$\frac{dR}{dt} = \delta_I I + \delta_A A + \delta_H H + \delta_Q Q \tag{8g}$$

$$\frac{dD}{dt} = \mu_I I + \mu_H H. \tag{8h}$$

2.3 Positivity and boundedness of solution

We assume that the vaccine efficacy function $\phi(t)$ is zero at $t = 0$ and monotonically increases up to $t = \tilde{t}$, and after that it is monotonically decreasing. Under this assumption we have,

$$\phi'(t) \geq 0 \text{ for } 0 < t \leq \tilde{t}, \text{ and } \phi'(t) \leq 0 \text{ for } t > \tilde{t}.$$

Also, we assume that when the number of susceptible is close to 0 then we can stop new vaccination. Without any loss of generality we suppose that $t = t_*$ is the first time when $S(t_*) = 0$. Then from the epidemiological point of view, we assume

$$V'(t) = 0, \quad \text{for } t \in [\max\{0, t_* - \tilde{t}\}, \ t_*]. \tag{9}$$

Now, from (8a), we can write

$$\frac{dS}{dt} = -\frac{\beta S}{N}(I + \alpha A) - \int_0^t \phi'(t - \eta)V'(\eta)d\eta$$

$$= -\frac{\beta S}{N}(I + \alpha A) - \int_0^t \phi'(\eta)V'(t - \eta)d\eta.$$

Evaluating $\frac{dS}{dt}$ at $t = t_*$, we find,

$$\left.\frac{dS}{dt}\right|_{t=t_*} = -\frac{\beta S(t_*)}{N}(I(t_*) + \alpha A(t_*)) - \int_0^{t_*} \phi'(\eta)V'(t_* - \eta)d\eta$$

$$= -\int_0^{t_*} \phi'(\eta)V'(t_* - \eta)d\eta.$$

Now, if $t_* < \tilde{t}$ then using (9) we get $V'(t) = 0$ when $t \in [0, t_*]$. Then we get

$$\frac{dS}{dt}\bigg|_{t=t_*} = -\int_0^{t_*} \phi'(\eta) V'(t_* - \eta) d\eta = 0. \tag{3}$$

If $t_* \geq \tilde{t}$ then using (9) we get $V'(t) = 0$ when $t \in [t_* - \tilde{t}, t_*]$. Then we get

$$\frac{dS}{dt}\bigg|_{t=t_*} = -\int_{0\tilde{t}} \phi'(\eta) V'(t_* - \eta) d\eta - \int_{\tilde{t}}^{t_*} \phi'(\eta) V'(t_* - \eta) d\eta$$

$$= -\int_{\tilde{t}}^{t_*} \phi'(\eta) V'(t_* - \eta) d\eta \geq 0,$$

since, $\phi'(\eta) \leq 0$ for $\tilde{t} \leq \eta \leq t_*$, and $V'(\eta) \geq 0$ for all $\eta \geq 0$. Hence in both the cases we find, $\frac{dS}{dt}\big|_{t=t_*} \geq 0$. This proves that if $S(0) > 0$, then $S(t) \geq 0$ for all $t > 0$. Now to prove the positivity of the other compartments we follow the same approach as in [26]. Let

$$\mathbb{R}_+^7 = \{(E, I, A, Q, H, R, D) \in \mathbb{R}^7 : E, I, A, Q, H, R, D \geq 0\}$$

denote the non-negative cone of $\mathbb{R}^7$ and denote

$$X(t) = (E(t), I(t), A(t), Q(t), H(t), R(t), D(t)) \in \mathbb{R}_+^7.$$

Then we have the following inequalities:

$$\frac{dE}{dt}\bigg|_{E=0, X(t) \in \mathbb{R}_+^7} = \frac{\beta S}{N}(I + \alpha A) \geq 0, \quad \frac{dI}{dt}\bigg|_{I=0, X(t) \in \mathbb{R}_+^7} = r\sigma E \geq 0,$$

$$\frac{dA}{dt}\bigg|_{A=0, X(t) \in \mathbb{R}_+^7} = (1-r)\sigma E \geq 0, \quad \frac{dQ}{dt}\bigg|_{Q=0, X(t) \in \mathbb{R}_+^7} = \eta_1 I \geq 0,$$

$$\frac{dH}{dt}\bigg|_{H=0, X(t) \in \mathbb{R}_+^7} = \eta_2 I + \xi Q \geq 0, \quad \frac{dR}{dt}\bigg|_{R=0, X(t) \in \mathbb{R}_+^7} = \delta_I I + \delta_A A + \delta_H H + \delta_Q Q \geq 0,$$

$$\frac{dD}{dt}\bigg|_{D=0, X(t) \in \mathbb{R}_+^7} = \mu_I I + \mu_H H \geq 0.$$

This ensures that if we start with a non-negative initial condition then the system is invariant in $\mathbb{R}_+^7$. Now using the Eq (3) we can write,

$$0 \leq S(t) + E(t) + I(t) + A(t) + Q(t) + H(t) + R(t) + D(t) \leq N,$$

for all time $t \geq 0$, which proves the boundedness of solution of the system (8).

2.4 Controlled reproduction number

In this article we are considering the situation not from the onset of epidemic progression rather we are starting from an intermediate time point. Hence we will be interested on effective reproduction number rather than so called basic reproduction number. Further, at the targeted initial time the vaccination drive is already started and hence the obtained threshold will be somewhat controlled reproduction number.

Suppose vaccination in China was started at time $t = t_0 = 0$ and the zero-COVID policy was relaxed from $t = t_1$, and since, in China there was negligible infection as compared to the total population during $[t_0, t_1]$, we neglect infections, recovery and death in the interval $[t_0, t_1]$, and we suppose the infection enters the population at time $t = t_1$. Hence our interest is to determine the reproduction number at time $t = t_1$. Consider the equilibrium point $(S(t_1), 0, 0, 0, 0, 0, 0, 0)$. Now,

$$S(t_1) = N(1 - m(t_1)) = N - \int_0^{t_1} \phi(t_1 - \eta) V'(\eta) d\eta.$$

We use the next generation matrix approach [27], to calculate the controlled reproduction number the details of which is in Sec.7.

The controlled reproduction number $\mathcal{R}_c$ at time $t = t_1$ is given by:

$$
\begin{aligned}
\mathcal{R}_c &= \frac{\beta S(t_1) r}{N(\eta_1 + \eta_2 + \delta_I + \mu_I)} + \frac{\alpha \beta (1-r) S(t_1)}{\delta_A N} \\
&= (1 - m(t_1)) \left(\frac{\beta r}{\eta_1 + \eta_2 + \delta_I + \mu_I} + \frac{\alpha \beta (1-r)}{\delta_A} \right) \\
&= \left(1 - \frac{1}{N} \int_0^{t_1} \phi(t_1 - \eta) V'(\eta) d\eta \right) \left(\frac{\beta r}{\eta_1 + \eta_2 + \delta_I + \mu_I} + \frac{\alpha \beta (1-r)}{\delta_A} \right).
\end{aligned}
\tag{10}
$$

The above expression of the controlled reproduction number depends upon the vaccine efficacy $\phi(\eta)$ and the daily number of vaccinations $V'(\eta)$. This expression of reproduction number suggests that a lower level of immunity at time t_1 leads to larger value of $\mathcal{R}_c$ and consequently a larger peak of epidemic.

3 Under-reporting in infected and in death: A reason for larger epidemic peak

Under-reporting in infected is an important aspect that can significantly influence the spread of epidemic. Here, by the term *under-reporting in infected*, we mean a subclass of the symptomatic infected individuals where individuals do not move from the symptomatic infected compartment to the quarantine or hospitalization. The symptoms onset of an infected individual is mainly determined by the total viral load in the upper respiratory track of an infected individual. Asymptomatic individuals generally have lower viral loads [28–30] and consequently the disease transmission rate is also smaller for them. Whereas, a symptomatic infected has a higher viral load inside his/her body and consequently the transmission rate is higher for them. Thus under-reporting in the symptomatic infected compartment can play a crucial role.

Since, the asymptomatic individuals do not develop any infection related symptoms, they are not tested for the infection and their numbers remain unaccounted for in the official records. In contrast, the complete reporting of symptomatic individuals can be made compulsory but which is not the case for COVID-19 due to various reasons. One possible explanation for under-reporting lies in the individual-level use of testing kits. In this scenario, individuals who test positive may not be included in the reported data due to the lack of awareness of the individuals. Furthermore, symptomatic individuals, despite being conscious of their infectiousness, may not encourage reporting due to many reasons like, fear of social stigma and discrimination, government policies etc. These factors collectively contribute to the challenge of achieving complete reporting of symptomatic cases in the context of COVID-19. As a matter of fact, from the controlling perspectives, we consider under-reporting as an important factor which is defined as the proportion of symptomatic cases who were not reported to the surveillance authority [31]. To account for under-reporting in symptomatic compartment, we

introduce a new parameter $u \in (0, 1]$ representing a fraction of symptomatic cases that are reported. Accordingly, $I_1(t)$ and $I_2(t)$ represent the reported and under reported infected compartments which are two sub-compartments of the symptomatic compartment I in system (8). To be precise, $I_1 = ur\sigma E$ and $I_2 = (1 - u)r\sigma E$. In the presence of under-reporting, the growth equations for I_1 and I_2 compartments are given by:

$$\frac{dI_1}{dt} = ur\sigma E - (\eta_1 + \eta_2 + \delta_I + \mu_I)I_1 \tag{11a}$$

$$\frac{dI_2}{dt} = (1 - u)r\sigma E - (\delta_I + \mu_I)I_2. \tag{11b}$$

The complete model with I_1 and I_2 compartments is described in (16) at Appendix. (7.2) in S1 Appendix. In the complete model, the new parameter α_u is the relative transmission rate of the under-reported compartment and all other parameters are same as described earlier for the model (8).

3.1 Controlled reproduction number

Now, we find the controlled reproductive number for the under-reported system (see 16 in Sec. 7). Consider the equilibrium point $(S(t_1), 0, 0, 0, 0, 0, 0, 0, 0)$. Now,

$$S(t_1) = N(1 - m(t_1)) = N - \int_0^{t_1} \phi(t_1 - \eta)V'(\eta)d\eta.$$

We use the next generation matrix approach [27], to calculate the controlled reproduction number. Details of the calculations are given in 7.2.

The reproduction number at time $t = t_1$ is the spectral radius of the matrix FV^{-1} and is given by:

$$\begin{aligned}
\mathcal{R}_c^1 &= \frac{\beta S(t_1)ur}{N(\eta_1 + \eta_2 + \delta_I + \mu_I)} + \frac{\alpha_u\beta(1 - u)rS(t_1)}{\delta_I + \mu_I} + \frac{\alpha\beta(1 - r)S(t_1)}{\delta_A N} \\
&= (1 - m(t_1))\left(\frac{\beta ur}{\eta_1 + \eta_2 + \delta_I + \mu_I} + \frac{\alpha_u\beta(1 - u)r}{\delta_I + \mu_I} + \frac{\alpha\beta(1 - r)}{\delta_A}\right) \\
&= \left(1 - \frac{1}{N}\int_0^{t_1} \phi(t_1 - \eta)V'(\eta)d\eta\right)\left(\frac{\beta ur}{\eta_1 + \eta_2 + \delta_I + \mu_I} + \frac{\alpha_u\beta(1 - u)r}{\delta_I + \mu_I} + \frac{\alpha\beta(1 - r)}{\delta_A}\right).
\end{aligned}$$

Note: If $u = 1$, then $\mathcal{R}_c^1 = \mathcal{R}_c$, and for any $u \in (0, 1)$, $\mathcal{R}_c^1 \geq \mathcal{R}_c$. In Fig (4), a comparison between the controlled reproduction numbers with under-reporting and without under-reporting is shown. Note that $u = 1$ corresponds to the case of no under-reporting, and in that case both the reproduction numbers $\mathcal{R}_c^1$ and $\mathcal{R}_c$ are equal. As the value of u decreases, the reproduction number $\mathcal{R}_c^1$ increases (blue curve).

3.2 Imitation dynamics of under-reporting in infected

In the context of COVID-19 in China, when the zero-COVID policy was relaxed, then there might be sudden significant changes in human behavior. Since, the govt. relaxed the restrictions and the BA.2 sub-variant of Omicron was comparatively less severe, people may be inclined towards under-reporting. In addition to that, the imitation dynamics can also play a crucial role in under-reporting. An individual who initially was inclined towards reporting, may change his/her mind towards under-reporting when he/she finds people around him/her who are inclined towards under-reporting, and vice-versa. Thus in more realistic scenario, the
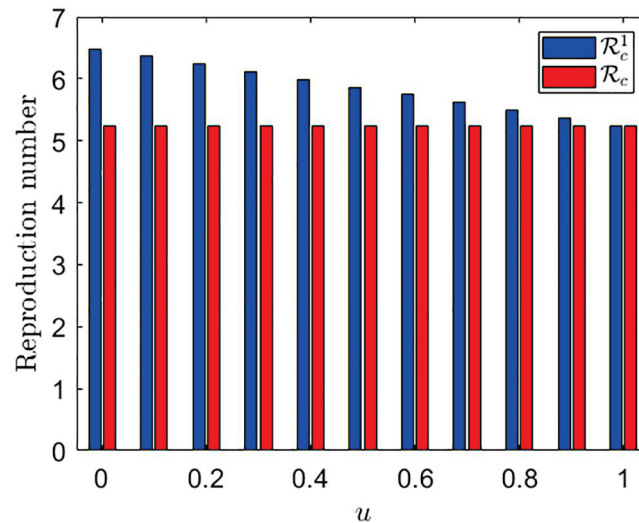
Fig 4. The red and blue color correspond to the controlled reproduction numbers with under-reporting and without under-reporting respectively. All other parameter values are same as in the green curves in Fig (17).

proportion of under-reporting can depend on the behaviors of individuals and one of the important behavior is the imitation.

Under-reporting can be considered as a type of game dynamics, similar to the imitation game. In the imitation game, individuals observe the behavior of their peers and adopt similar strategies in order to achieve a certain outcome. Similarly, in situations where reporting is required, individuals may observe the behavior of others and adopt similar strategies. The payoff regarding the imitation game dynamics, represents the benefit or cost associated with a particular action or strategy, and the payoffs are typically taken as negative for the following reasons:

i. Reporting carries some cost: Reporting often involves some monetary cost, as well as the time and effort required to be quarantined or hospitalized. This cost is typically reflected as a negative payoff.

ii. Some people may perceive quarantine, a consequence of reporting, as an unpleasant experience, which may also be reflected as a negative payoff.

iii. The negative consequences of under-reporting: Under-reporting can result in negative consequences, such as spreading the disease to others, severity of the infection and potentially even death. These negative consequences are typically reflected as negative payoffs.

Suppose, $-p_r$ denotes the payoff related to reporting, where, p_r accumulates some monetary cost, as well as the time and effort required to be quarantined or hospitalized, and $-p_u$ is the payoff related to under-reporting, where, p_u is the perceived risk of suffering serious conditions and sometimes death due to infection. These perceived risk (or cost) factors p_r and p_u typically lie between 0 and 1 because they are usually expressed as the probabilities or percentages, where, 0 represents no risk (or cost) and 1 represents highest possible risk (or cost). The net payoff gain $\Delta\mathcal{P}$ is defined by

$$\Delta\mathcal{P} = -p_r - (-p_u) = -p_r + p_u.$$

Then the human behavior based imitation dynamics of u is described by (following the derivation by Bauch [32]):

$$\frac{du}{dt} = c\Delta\mathcal{P}u(1-u) = c(-p_r + p_u)u(1-u), \tag{12}$$

where, c is the imitation rate. It is important to mention here that the initial condition for u lies between 0 and 1, i.e., $0 \leq u(0) \leq 1$.

Constant rate of under-reporting in infected. Note that, if the net payoff gain is zero, i.e., $\Delta\mathcal{P} = 0$, then $\frac{du}{dt} = 0$, which is the situation of constant rate of under-reporting.

4 Parameter estimation using physics informed neural networks

One of the challenges in using compartmental models with associated differential equation system to describe the dynamics is the *estimation of parameters* for given data. Usually, parameters may be estimated from observed patterns in the data, but transmission rates often have to be computed using heuristic algorithms that are computationally or statistically motivated. In the recent years, there have been new approaches with *machine learning* to discover parameters [17]. One of these approaches includes Artificial Neural Networks (ANN) which are motivated by the human neural system where each neuron is represented with a node, signals are inputs, and synapse is represented as a function evaluation. Each neuron is connected to different neurons in multiple layers, which leads to decrease the error using non-linear approximations. ANN have been used in regression and classification tasks in the last few years. By adding the physics behind the given problem, i.e., the system of equations, a new approach was proposed called Physics-Informed Neural Networks or PINNs [18]. These neural networks encode model equations, like Partial Differential Equations (PDEs), as a component of the neural network itself [33].

This approach aims to solve two main classes of problems: *data-driven solution* and *data-driven discovery* of differential equations. It is a fast and mesh-free method. However, it is still in early development but is starting to be applied to infectious diseases. Recently, Disease Informed Neural Networks (DINNs) was proposed to leverage the hidden physics of infectious diseases and infer the latent quantities of interest by approximating them using PINNs [19]. The neural network architecture for system (8a) and (8b) is showed in Fig 5.

For the function $t \rightarrow (S, E, I, A, Q, H, R, D)$, the system (8a)-(8h) corresponds to the following residuals:

$$
\begin{aligned}
\mathcal{L}_S &= \frac{dS}{dt} + \frac{\beta S}{N}(I + \alpha A) + \int_0^t \phi'(t-\eta)V'(\eta)d\eta, \\
\mathcal{L}_E &= \frac{dE}{dt} - \frac{\beta S}{N}(I + \alpha A) + \sigma E, \\
\mathcal{L}_I &= \frac{dI}{dt} - r\sigma E + (\eta_1 + \eta_2 + \delta_I + \mu_I)I, \\
\mathcal{L}_A &= \frac{dA}{dt} - (1-r)\sigma E + \delta_A A, \\
\mathcal{L}_Q &= \frac{dQ}{dt} - \eta_1 I + (\xi + \delta_Q)Q, \\
\mathcal{L}_H &= \frac{dH}{dt} - \eta_2 I - \xi Q + (\delta_H + \mu_H)H, \\
\mathcal{L}_R &= \frac{dR}{dt} - \delta_I I - \delta_A A - \delta_H H - \delta_Q Q, \\
\mathcal{L}_D &= \frac{dD}{dt} - \mu_I I - \mu_H H.
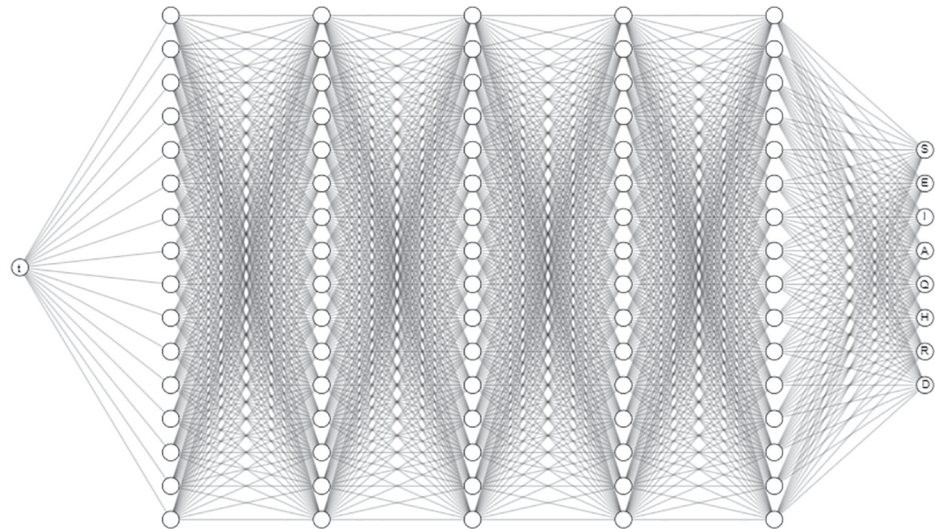\end{aligned}
$$

Fig 5. Neural Network architecture for DINNs applied to system (8a)-(8h).

This neural network is trained using temporal data, which estimate the best values of the model parameters without much prior information. Let the unknown solution be a vector of eight components such that

$$u(t; \lambda) = [S(t; \lambda), E(t; \lambda), I(t; \lambda), A(t; \lambda), Q(t; \lambda), H(t; \lambda), R(t; \lambda), D(t; \lambda)]^\top$$

where λ are the parameters related to the disease dynamics, and a known initial condition $u(0)$.

The training data has been discretized, as time $\{t^j\}$ and solution $\{u^j\}$, where $j = 0, 1, \ldots, N_{\text{data}}$ such that $t^0 = 0$ corresponds to an initial condition. The goal then is to train a neural network with parameters $(\lambda^\top, \theta^\top)^\top$ where θ is a concatenation of weights and biases for each artificial neuron. The optimization process obtains a vector $(\hat{\lambda}^\top, \hat{\theta}^\top)^\top$ and an approximation $\hat{u}^j(\hat{\lambda}, \hat{\theta})$ using the following loss function

$$\mathcal{L}(\hat{\lambda}, \hat{\theta}) = \omega_{\text{ode}} \mathcal{L}_{\text{ode}}(\hat{\lambda}, \hat{\theta}) + \omega_{\text{ic}} \mathcal{L}_{\text{ic}}(\hat{\lambda}, \hat{\theta}) + \omega_{\text{data}} \mathcal{L}_{\text{data}}(\hat{\lambda}, \hat{\theta})$$

where ω_{ode}, ω_{ic} and ω_{data} are the loss weights of the loss functions of the system of differential equations, initial conditions and training data, respectively. While the overall loss function is decomposed in other three parts, the loss function from the differential equation may be expressed as:

$$\begin{aligned} \mathcal{L}_{\text{ode}}(\hat{\lambda}, \hat{\theta}) \quad &= \mathcal{L}_S(\hat{\lambda}, \hat{\theta}) + \mathcal{L}_E(\hat{\lambda}, \hat{\theta}) + \mathcal{L}_I(\hat{\lambda}, \hat{\theta}) + \mathcal{L}_A(\hat{\lambda}, \hat{\theta}) \\ &+ \mathcal{L}_Q(\hat{\lambda}, \hat{\theta}) + \mathcal{L}_H(\hat{\lambda}, \hat{\theta}) + \mathcal{L}_R(\hat{\lambda}, \hat{\theta}) + \mathcal{L}_D(\hat{\lambda}, \hat{\theta}). \end{aligned}$$

The loss function corresponding to the data maybe expressed as:

$$\mathcal{L}_{\text{data}}(\hat{\lambda}, \hat{\theta}) = \sum_{i=1}^{8} \frac{1}{N_{\text{data}}} \sum_{j=1}^{N_{\text{data}}} (u_i^j - \hat{u}_i^j(\hat{\lambda}, \hat{\theta}))^2.$$

The loss function corresponding to the initial condition may be expressed as:

$$\mathcal{L}_{ic}(\hat{\lambda}, \hat{\theta}) = \sum_{i=1}^{8} (u_i^0 - \hat{u}_i^0(\hat{\lambda}, \hat{\theta}))^2$$

where u_i for $i = 1, \ldots, 8$ corresponds to solution to each compartment in system ([8a])–([8h]), for example, u_1 corresponds to S, u_2 corresponds to E and so on until u_8 corresponds to D. Algorithm 1 shows how to estimate $u(t; \lambda)$ and the parameters λ.

Algorithm 1: DINN algorithm.

```
Input: Training Data {tʲ}, {uʲ} where j = 0, 1, ..., N_data
Output: û and λ̂
1 Initialize λ̂₀ and θ̂₀
2 Define time interval where the solution will be found.
3 Define loss function L(λ̂,θ̂), related to residual errors, initial
  conditions and training data.
4 Create a fully connected neural network with 1 neuron in the input
  layer and 8 neuron in the output layer (one per compartment) and such
  that it normalize the input data.
5 Choose optimization hyper-parameters (e.g. Adam optimizer, learning
  rate and loss weights).
6 for iter = 1, ..., max_iter do
7   Compute total loss L(λ̂_iter-1,θ̂_iter-1), in particular is necessary to use
    auto-differentiation for ODE residuals.
8   Train neural network with optimizer algorithm and update θ̂_iter-1 to
    θ̂_iter.
9   Get approximation û_iter and λ̂_iter.
10 Return û_max_iter and λ̂_max_iter.
```

For our computations, we first estimated,

$$\frac{dm}{dt} = \frac{1}{N} \int_0^t \phi'(t - \eta) V'(\eta) d\eta. \tag{13}$$

Specifically, we first estimated the best-fit sigmoid function for $V(\eta)$ using the real-data for vaccinations from Jan 2021 to Jan 2023 [10] given by:

$$\tilde{V}(t) = v_1 \left(1 - \frac{v_2}{v_2 + t^4}\right)$$

where $v_1 = 0.904733394$ and $v_2 = 1.16081503 \times 10^9$. Fig 6 shows the fit for this function in comparison to the vaccination data and we use the derivative of the function in Eq (13).

We then calculated $\phi(t)$ using specific values of $(b, c) = (0.231, 0.0084)$, for a particular choice of the vaccine efficacy function $\phi(t)$ by the acquisition-fading kernel, as given in Eq (14). From the explicit form that we obtain for $\phi(t)$, we compute $\phi'(t - \eta)$ to be used in Eq (13).

Once the integral in Eq (13) is numerically computed, this is explicitly used in solving the system ([8a])-([8h]) using the DINNs approach explained earlier. The neural network was able to learn the dynamics of the system ([8a])-([8h]) (see Fig 7).

As a first experiment, we fixed every parameter but we applied the framework only for estimating β, the rest of the parameters were fixed. Fig 8 shows the learning history of the DINNs framework. For this experiment in particular we fixed $\beta = 11$ and the approximation was $\hat{\beta} = 10.6782$, which means a relative error equals to 0.02924.

In order to study the robustness of the framework we have run the model 40 times. Fig 9 shows there is no a large variation of the error. Similarly, Fig 10 shows the average learning
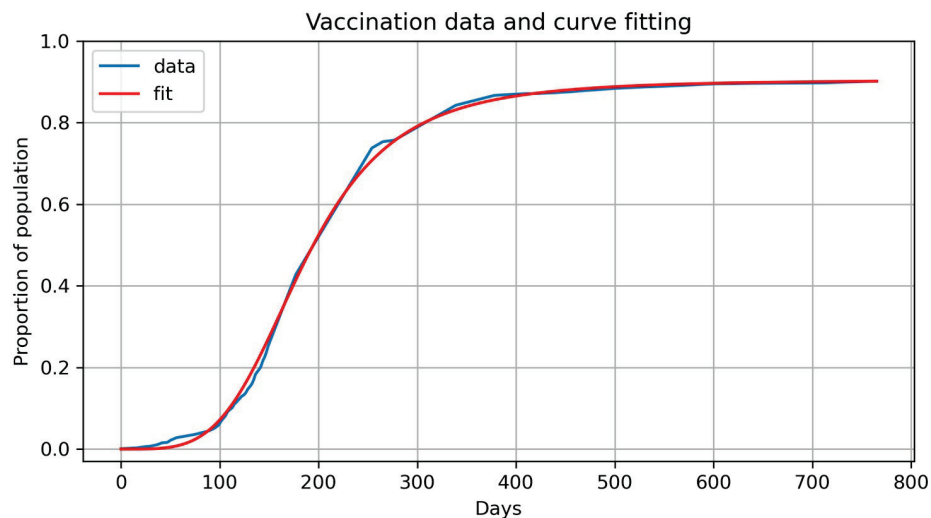
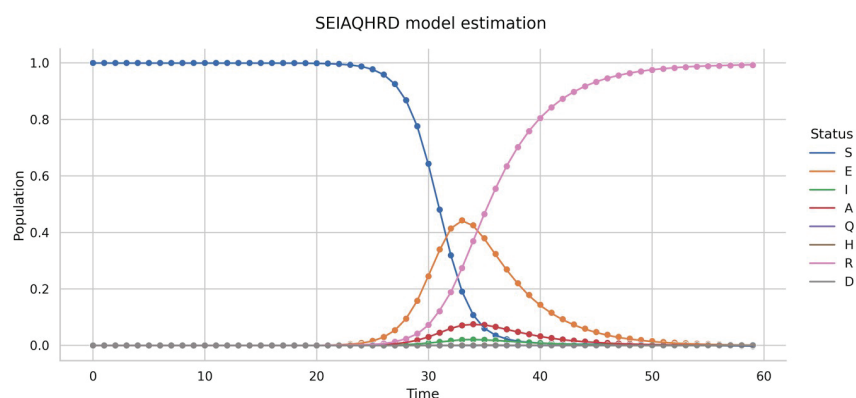**Fig 6. Proportion of population vaccinated and curve fitting.**

Fig 7. System ([8a](#))-([8h](#)) estimation using synthetic data, dots correspond to training data and lines correspond to DINNs prediction.
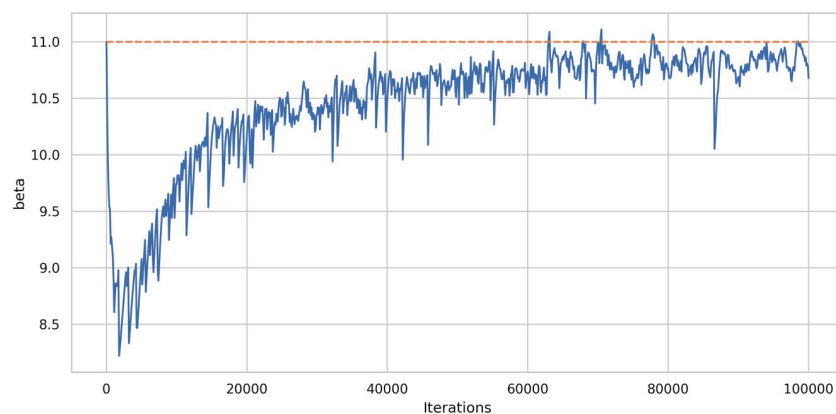
Fig 8. Learning history of parameter estimation (only β) of system ([8](#)) using DINNs.
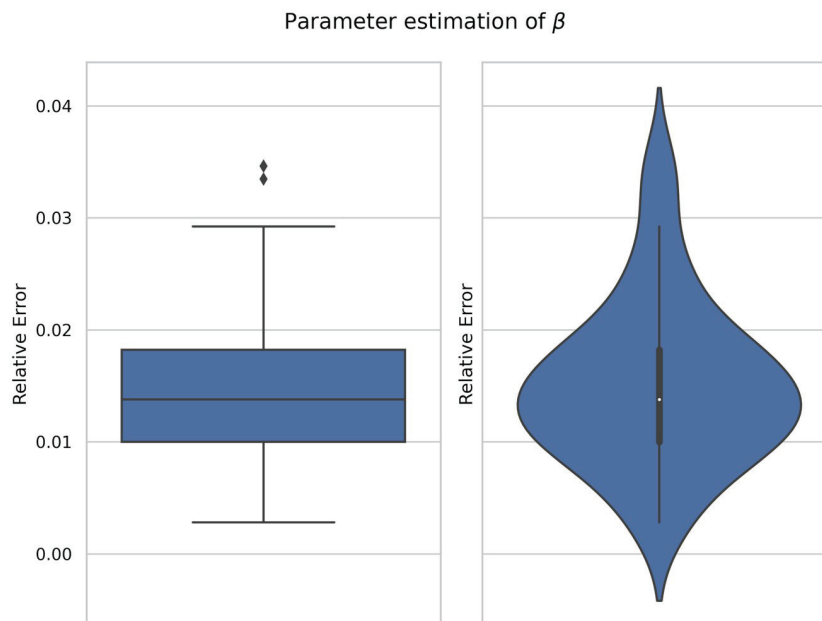
Fig 9. Box-plot (left) and violin-plot (right) of the relative error of β estimation corresponding to 40 simulations.
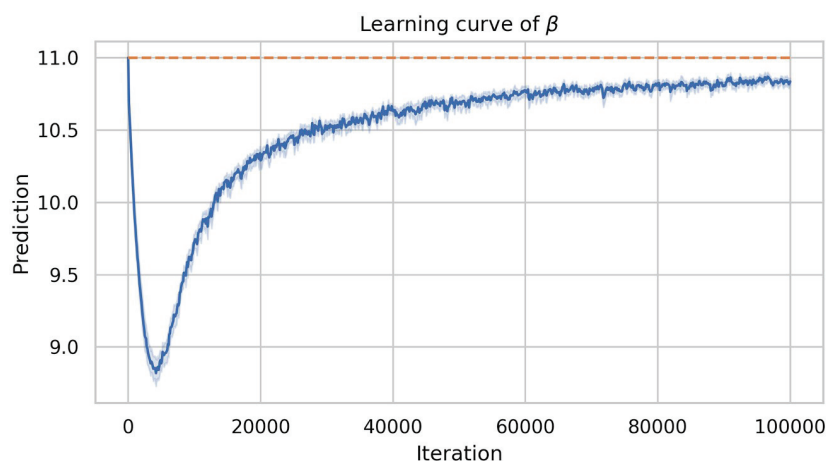
Fig 10. Mean learning history and its 95% confident interval of β estimation corresponding to 40 simulations.

history of β with a 95% confidence interval corresponding to the 40 simulations. Since this confidence interval is small we can support the robustness of this model as well.

4.1 Multi-parameter estimation

The predictive reliability of any inverse algorithm depends on its ability to apply the technique to estimate multiple parameters. In traditional non-linear least-squares approaches, it is well known that the approximations start to become worse when such optimization techniques try to estimate more parameters. Here we apply the DINNs approach described in this article to estimate more parameters next.
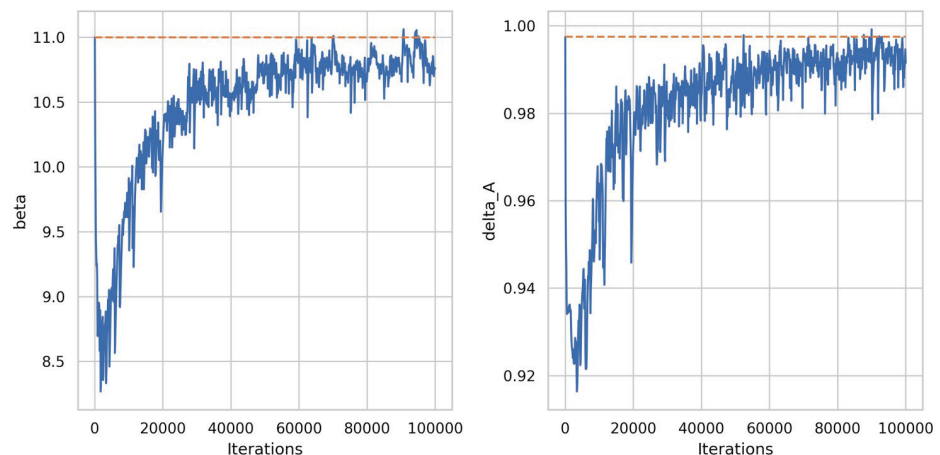
Fig 11. Learning history of parameter estimation (β and δ_A) of system (8) using DINNs.

Specifically, we try to estimate two parameters, namely, β and δ_A, where data prediction is similar to the first experiment. Fig 11 shows the learning history of these two parameters and Table 2 summarizes predictions and relative errors.

The next experiment considered the under-reported scenario corresponding to system (16) where we estimated β, δ_A and u. Fig 12 shows the learning history of these two parameters and Table 3 summarizes predictions and relative errors.

Next, we used the same system (16) but now considering u as a function, as defined in Eq (15) and $p_u(\theta) = \rho(1 - \theta)$. We ran a similar experiment to learn the parameters β, δ_A and θ. See Fig (13) for parameter prediction history and Table 4 summarize parameters estimation with its relative errors.

4.2 Influence of noise

To measure the robustness of our approach for the estimation approach to system (8), we perturb the data with different levels of noise (1%, 5% and 10%) on each compartment and then estimate β. Similar to the first estimation (single parameter) for studying robustness of the framework, we ran 30 simulations of each noise level. Figs 14 and 15 are box-plots and violin-plots of each noise level, showing what would be expected, this is, to increasing the noise results also increase the relative error and its spread as well. The history learning (see Fig 16) shows the same behavior, more noise increase the error and the confidence interval, but it is proportional to the noise level added to the training data.

4.3 Influence of missing data

Another measurement of robustness of an optimization technique is how the method predicts solutions when data is limited. In order to simulate it, we removed randomly 10%, 25% and 50% data points for the estimation of $\beta = 11$ in the system of Eq (8). Table 5 shows that even

Table 2. Parameter predictions and relative errors for system (8) using DINNs.

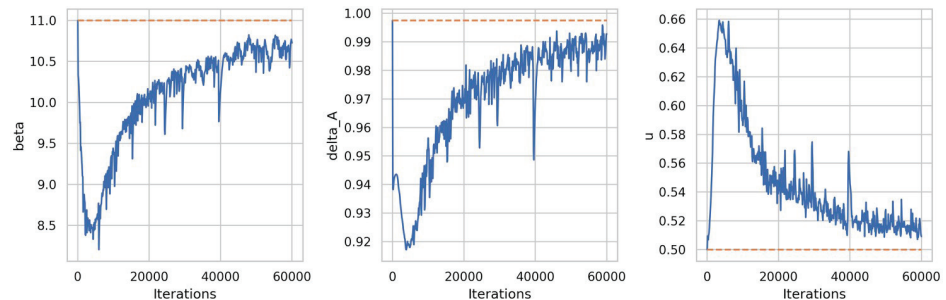|	Real	Predicted	Rel. Error
β	11	10.7587	0.02193
δ_A	0.9975	0.9915	0.00596

Fig 12. Learning history of parameter estimation (β, δ_A and u) of system (16) using DINNs.

with increasing amount of missing data, the neural network method proposed herein, performs very well.

5 Results and findings

5.1 Effect of vaccine efficacy on epidemic progression

In this subsection, we study the effect of the vaccine efficacy on the disease progression with the help of numerical simulation. Here, we set $t = 0$ at the time when the vaccination started i.e., 15 December, 2020 for China, and t_1 is 1st November, 2022 (because the outbreak started around the first week of November 2022, as per Worldometer data). The plot of $V(t)$ for real data of cumulative vaccination in China starting from 15 December, 2020 to 23 December, 2022 is shown in Fig 1.

Effect of $\phi(\eta)$ on $\mathcal{R}_c$. Here, we study the effect of vaccine efficacy on the controlled reproduction number $\mathcal{R}_c$. Vaccine efficacy begins to rise shortly after vaccination, reaching its peak after a few weeks (or months), and then it gradually declines over time-post-vaccination, due to waning of immunity [34]. To represent this pattern of vaccine efficacy, we consider a particular choice of the vaccine efficacy function by the acquisition-fading kernel, given by

$$\phi(t) = A(e^{-bt} - e^{-ct}), \tag{14}$$

where, b is the fading rate, c is the acquisition rate, $b \neq c$, and $A = 1/((c/b)^{(b/(b-c))} - (c/b)^{(c/(b-c))})$. For this particular choice of ϕ, in Fig 17a, we plot the controlled reproduction number $\mathcal{R}_c$ as a function of b and c. From Fig 17, we observe that $\mathcal{R}_c$ significantly depends upon the fading and acquisition parameters b and c respectively.

Note. The value of $\mathcal{R}_c$ is significantly small in case of smaller values of both the fading and acquisition parameters b and c respectively. Thus, a vaccine efficacy function with slow acquisition and slow decaying nature, can effectively control the epidemic progression. If the vaccine efficacy rapidly decays (as in the green curve in Fig 17(b)), then the $\mathcal{R}_c$ is very high (around 5.25 for the green curve), whereas, if the vaccine efficacy decays slowly (as in the red curve in Fig 17(b)), then the $\mathcal{R}_c$ is significantly smaller (around 1.42 for the red curve).

Table 3. Parameter predictions and relative errors for system (16) using DINNs.

	Real	Predicted	Rel. Error
β	11	10.7269	0.02481
δ_A	0.9975	0.9927	0.00472
u	0.5	0.5091	0.01825
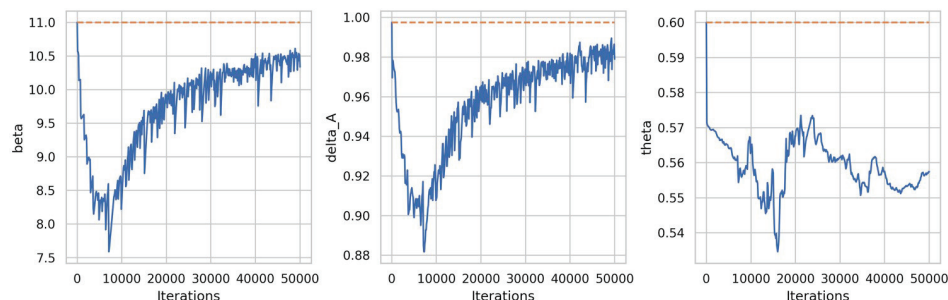
Fig 13. Learning history of parameter estimation (β, δ_A and θ) of system (16) and Eq (15) using DINNs.

Effect of $\phi(\eta)$ on daily cases. To numerically investigate the effect of vaccine efficacy and its duration on the daily number of cases of the infection, we numerically simulate the model (8) with parameter values given in Table 1, along with the efficacy function defined in (14). The vaccination data $V(t)$ is shown in Fig 1. The initial conditions are: $S(t_1) = N - 1$, $E(t_1) = 0$, $I(t_1) = 1$, $A(t_1) = 0$, $Q(t_1) = 0$, $H(t_1) = 0$, $R(t_1) = 0$ and $D(t_1) = 0$, where t_1 is 1st November, 2022 (as the outbreak started around the first week of November 2022, as per Worldometer data). We assumed the value of $\beta = 11$ to get the daily cases in the order of 10^7. We simulate the model (8) with the vaccine efficacy function $\phi(\eta)$ (as in (14)) for the set of values of the acquisition and rate parameters as follows: $(b, c) = (0.231, 0.0084)$, $(0.061, 0.0033)$, $(0.271, 0.0043)$, and $(0.1, 0.004)$, corresponding to the green, red, black and blue curves, respectively, in Fig 17a and 17b. The simulation results for the daily cases are shown in Fig 17c by respective colors in Fig 17a and 17b. We observe that there is no outbreak corresponding to the red curves, whereas, the green curves produce a very large peak of the outbreak. Thus, although the number of vaccinated individuals are same, the less efficacy of the vaccine (i.e., the efficacy is decaying quickly) can lead to significantly different type of epidemic progression. In other words, the less effectiveness of the vaccine may be one of the reasonable factor to have faster epidemic growth and a large number of daily infected cases.

5.2 Under-reporting in death

Based upon the information available in various news media, there were under-reporting in death in China [12, 13]. Although, the under-reporting in death has no direct impact on the disease transmission, but it can indirectly influence the human behavior. If people receive suppressed data of death, then based upon that suppressed data people may decide that the severity and mortality of the disease is less, and this decision has an impact on $-p_u$ (the payoff related to under-reporting of infected). To capture this implicit influence of under-reporting in death on the epidemic progression, we introduce another parameter $\theta \in [0, 1]$, which represents the proportion of under-reporting in death. Since the perceived payoff related to under-reporting of infected depends upon the information about the morbidity due to infection, we

Table 4. Parameter predictions and relative errors for system (15) and (16) using DINNs.

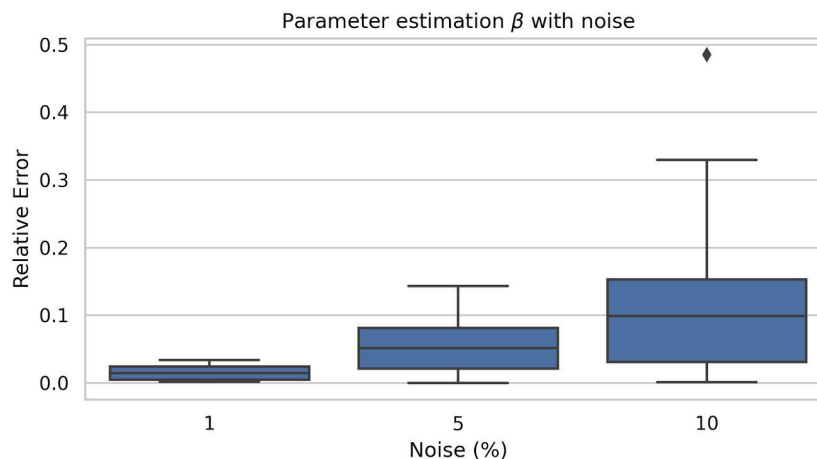|	Real	Predicted	Rel. Error
β	11	10.33896916	0.060093713
δ_A	0.9975	0.979012692	0.018533643
θ	0.6	0.557463108	0.070894819

Fig 14. Box-plot of the relative error of β estimation corresponding to 30 simulations of each noise level (1%, 5% and 10%).
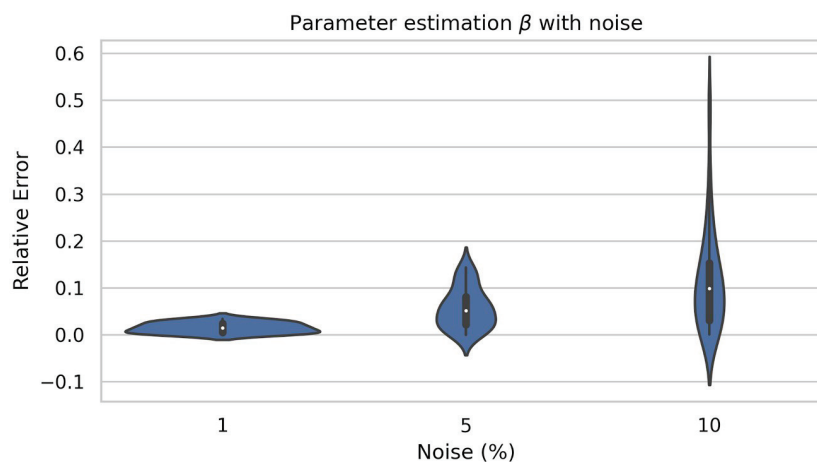
Fig 15. Violin-plot of the relative error of β estimation corresponding to 30 simulations of each noise level (1%, 5% and 10%).

can assume that the payoff related to under-reporting of infected is a function of θ, i.e., $-p_u(\theta)$. Then the Eq (12) becomes:

$$\frac{du}{dt} = c(-p_r + p_u(\theta))u(1-u).$$ (15)

In a particular we can choose

$$p_u(\theta) = \rho(1-\theta),$$

where, θ is the proportion of death that is suppressed, and consequently, the proportion of reported information of death is $(1-\theta)$. Just to explain, if $\theta = 1$, which means all the deaths are under reported, then people may receive the information that the disease related death rate is
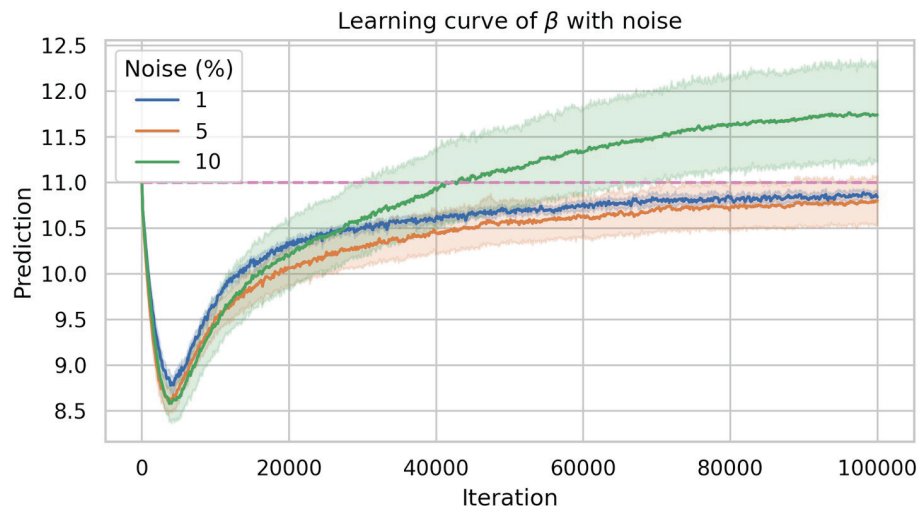
Fig 16. Mean learning history and its 95% confident interval of β estimation corresponding to 30 simulations of each noise level (1%, 5% and 10%).

https://doi.org/10.1371/journal.pone.0290368.g016

zero, then,

$$p_u(\theta) = \rho(1-\theta) = 0,$$

which implies from (15),

$$\frac{du}{dt} = c(-p_r + p_u)u(1-u)$$
$$= -cp_r u(1-u).$$

This shows u is decreasing, which signifies the fact that people are becoming more inclined due to under-reporting. Thus under-reporting in death as well as in infected can influence the epidemic progression.

We explain the implicit effect of under-reporting in death in the decision making of individuals which has an impact on determining the size of the epidemic peak. We simulate system (15) and (16) for three scenarios as follows:

i. Scenario-I: $\theta = 0.05$, which means that only 5% data of death are under-reported or suppressed.

ii. Scenario-II: $\theta = 0.6$, which means that 60% data of death are under-reported or suppressed.

iii. Scenario-III: $\theta = 0.95$, which means that 95% data of death are under-reported or suppressed.

Table 5. $\beta = 11$ estimation and relative errors for system (8) using DINNs with different percentage of missing data.

Noise	Predicted	Rel. Error
10%	10.416714	0.053026
25%	10.738632	0.023761
50%	11.138561	0.012596

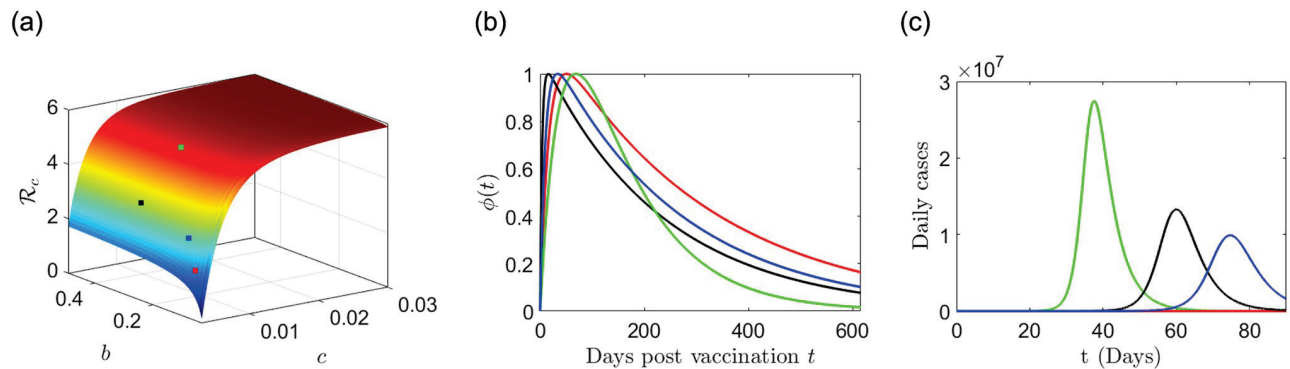https://doi.org/10.1371/journal.pone.0290368.t005

Fig 17. (a) Plot of the controlled reproduction number $\mathcal{R}_c$ as a function of the fading rate b and the acquisition rate c as described in the formula (14); (b) Plot of the acquisition-fading kernel type vaccine efficacy function $\phi(t) = A(e^{-bt} - e^{-ct})$ (see, formula (14)) corresponding to different choice of b and c as follows: $(b, c) = (0.231, 0.0084)$ (green), $(b, c) = (0.061, 0.0033)$ (red), $(b, c) = (0.271, 0.0043)$ (black), $(b, c) = (0.1, 0.004)$ (blue) and the corresponding values of $\mathcal{R}_c$ are indicated by the dots of corresponding color in the panel (a); (c) The number of daily cases for different choice of (b, c) are shown by corresponding colors.

For these three possible scenarios, the simulated results for the daily cases are shown in Fig 18, where, Scenario-I, Scenario-II, and Scenario-III correspond to the blue, green and red curves respectively, in Fig 18. Although, the death data has no direct influence on disease progression, still, Fig 18 shows that, the suppression of death data can influence the epidemic progression implicitly through human behavior. In Fig 18, the red curve which correspond to the suppression of 95% of the death data, gives a higher peak, whereas, the blue curve which correspond to the suppression of only 5% of the death data can reduce the height of the epidemic peak. In Fig 19, we plot the maximum of daily cases as a function of p_r and θ, which explains the effect of both p_r and θ in determining the maximum height of the epidemic.
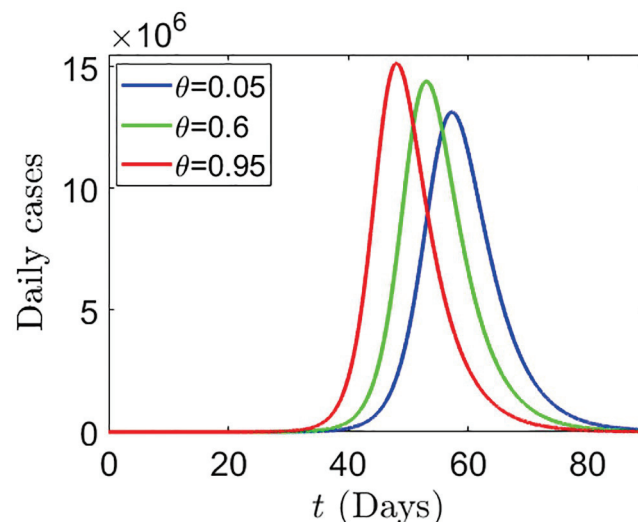

Fig 18. Daily cases for $\theta = 0.05$ (blue), $\theta = 0.6$ (green), $\theta = 0.95$ (red). We assume that $p_u(\theta) = \rho(1 - \theta)$. The parameter values are: $\alpha_u = 1.5$, $p_r = 0.4$, $\rho = 0.9$, $c = 0.8$, $u(0) = 0.8$, and all other parameters are same as before.
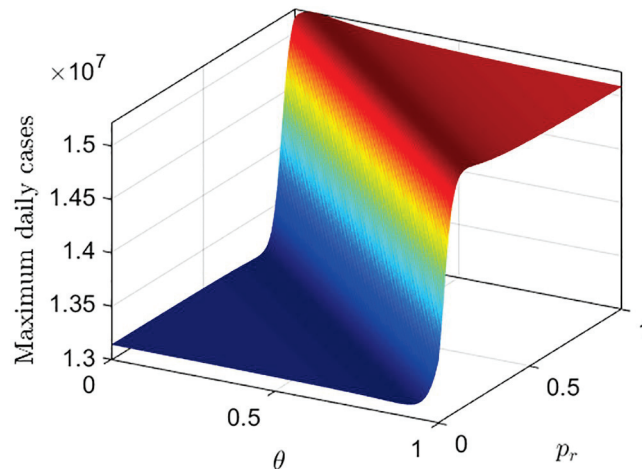
Fig 19. Maximum of daily cases is plotted as a function of p_r and θ. We assume that $p_u(\theta) = \rho(1 - \theta)$. The parameter values are: $\alpha_u = 1.5$, $\rho = 0.9$, $c = 0.8$, $u(0) = 0.8$, and all other parameters are same as before.

https://doi.org/10.1371/journal.pone.0290368.g019

6 Discussion

In the late 2019, the novel coronavirus SARS-CoV-2 emerged in China and rapidly spread to other parts of the world. To control the transmission of the disease, China implemented strict measures through the zero-COVID policy. With positive intentions, zero-COVID policy aimed to completely eliminate the transmission of COVID-19 in China. Apart from the social and economic issues, another drawback of China's zero-COVID policy was the low level of immunity achieved due to negligible infection, which prevented the country from attaining acquired immunity in its population [35]. This low immunity left a significant portion susceptible to the virus, contributing to the subsequent surge in cases after policy relaxation. Towards the end of 2022, China began gradually relaxing its zero-COVID policy and the reports in the media suggested a subsequent rapid spread of epidemic [3–6].

Our study aims to investigate the factors contributing to a large outbreak observed in China in the beginning of 2023. An SEIR-type epidemic model is developed that incorporates a time-dependent level of immunity, taking into account the number of vaccine doses administered and the efficacy of the vaccine.

The sudden outbreak following the relaxation of the zero-COVID policy in early November 2022 can be explained by the controlled reproduction number ($\mathcal{R}_c$) at that time $t = t_1$ (beginning of the outbreak in early November 2022 in China). The expression of $\mathcal{R}_c$ in (10) shows that $\mathcal{R}_c$ is inversely proportional to the level of immunity $m(t)$. Without a zero-COVID policy and a significant prior infection in the population, the level of immunity at time t_1 would be as follows:

$$\tilde{m}(t_1) = m(t_1) + \frac{1}{N}\int_0^{t_1} \psi(t_1 - \eta)R'(\eta)d\eta \; > \; m(t_1),$$

where, $\psi(\xi)$ is the efficacy of the acquired immunity which depends on the time-post-recovery ξ and $R'(\eta)$ is the daily new recovery at time η. Consequently, the immunity level without the zero-COVID policy $\tilde{m}(t_1)$, would exceed the immunity level with the policy $m(t_1)$. This would lead to a comparatively lower value of $\mathcal{R}_c$ in case of no zero-COVID policy, which indicates a lower peak of the outbreak. However, the use of acquired immunity of recovered to achieve herd immunity in the population is advisable only when the infection is less fatal (e.g., Omicron).

Immunity waning is an important factor which leads to a variation in the level of immunity in the population and consequently it has a significant influence in the emergence of new outbreaks. This immunity waning is considered in our definition of $m(t)$ (see formula (2)) by the time post vaccination dependent vaccine efficacy. A vaccinated individuals do not immediately reach maximum immunity after vaccination, but rather experience a gradual increase, peak, and subsequent decrease in immunity. This variation is captured through acquisition fading kernels in our model (see formula (14)).

The level of immunity in the population is also influenced by the dynamics of vaccination, for which we directly utilize available data sources. Our approach avoids the introduction of an additional compartment in the model to represent vaccination dynamics, simplifying the model while still considering the impact of vaccination on immunity levels.

Additionally, our model incorporates the imitation dynamics of individuals switching strategies based on net payoffs. Notably, our study explores the implicit effect of under-reporting in death data on individuals' decision-making processes. To the best of our knowledge, there is no existing mathematical study that explicitly addresses this aspect, highlighting the novelty of our research.

The modeling results of the study demonstrate that time-post-vaccination dependent vaccine efficacy plays a significant role in determining the size of the epidemic outbreak. Also, this work describes the fact that accurate and transparent reporting of daily cases and deaths is crucial to ensure individuals maintain a realistic understanding of the situation. In conclusion, the study emphasizes the importance of considering time-dependent vaccine efficacy and accurate reporting of daily cases and deaths in understanding and controlling the magnitude of an outbreak. The findings highlight the need for highly effective vaccines with durable immunity and transparent reporting systems to effectively combat the spread of COVID-19 and minimize the impact on public health and the economy. In addition, we introduce a novel neural network-based algorithm, specifically a Physics Informed Neural Networks (PINNs), to quantify parameters in the model by learning from the data. This algorithm aids in estimating critical parameters that contribute to understanding the dynamics of the epidemic. We strongly believe that this algorithm will help researchers to estimate many crucial parameters effectively, by learning the data.

The model developed in this work is generic and can be used for other epidemic diseases, but this model is based on several simplifying assumptions due to the lack of specific data. We assumed equal effectiveness for all vaccine doses and did not differentiate between individuals who received the first dose, second dose, or booster dose of the vaccine. Future research can incorporate more accurate data on vaccine efficacy and distinguish between different vaccine doses to calculate immunity levels more precisely. Furthermore, we assumed a homogeneous population with similar immune status for each individual. However, factors such as age can influence an individual's immune status, and future investigations can extend the model to introduce population heterogeneity through age-structured modeling. Regarding under-reporting, we assumed that the payoff related to under-reporting in infections is a function of the proportion of under-reporting in deaths, represented by $p_u(\theta) = \rho(1 - \theta)$, where θ represents the proportion of under-reporting in deaths. Future studies can explore different choices of the function $p_u(\theta)$ to further analyze the impact of under-reporting on individuals' decision-making.

Supporting information

S1 Appendix.
(PDF)

Author Contributions

Conceptualization: Aditi Ghosh.

Data curation: Alonso Ogueda-Oliva, Padmanabhan Seshaiyer.

Formal analysis: Samiran Ghosh, Alonso Ogueda-Oliva, Malay Banerjee.

Funding acquisition: Padmanabhan Seshaiyer.

Investigation: Samiran Ghosh, Alonso Ogueda-Oliva, Aditi Ghosh.

Methodology: Alonso Ogueda-Oliva, Malay Banerjee.

Project administration: Aditi Ghosh, Malay Banerjee.

Resources: Padmanabhan Seshaiyer.

Software: Samiran Ghosh, Alonso Ogueda-Oliva, Padmanabhan Seshaiyer.

Supervision: Aditi Ghosh, Padmanabhan Seshaiyer.

Validation: Samiran Ghosh, Alonso Ogueda-Oliva, Aditi Ghosh.

Visualization: Aditi Ghosh.

Writing – original draft: Samiran Ghosh, Alonso Ogueda-Oliva.

Writing – review & editing: Aditi Ghosh, Malay Banerjee, Padmanabhan Seshaiyer.

References

1. Yuan Shawn. Zero COVID in China: what next? *The Lancet*, 399(10338): 1856–1857, 2022. https://doi.org/10.1016/S0140-6736(22)00873-X

2. https://timesofindia.indiatimes.com/world/china/abrupt-relaxation-of-covid-curbs-puts-pressure-on-hospitals-authorities-as-cases-surge-in-china/articleshow/96171117.cms

3. https://www.ptcnews.tv/world-news/china-records-37-crore-fresh-covid-cases-in-a-day-highest-single-day-rise-globally-715283

4. https://www.livemint.com/news/india/coronavirus-live-updates-latest-coronavirus-news-today-in-india-china-11671754262102.html

5. Mallapaty SmritiChina COVID wave could kill one million people, models predict. *Nature*, 2022. https://doi.org/10.1038/d41586-022-04502-w PMID: 36536129

6. Dyer OwenCovid-19: Winter surge feared as China lurches away from zero covid, *British Medical Journal Publishing Group*, 2022. https://doi.org/10.1136/bmj.o2983 PMID: 36593542

7. Sadoff Jerald and Le Gars Mathieu and Shukarev Georgi and Heerwegh Dirk and Truyers Carla and de Groot, et al. COV2. S Covid-19 vaccine. *New England Journal of Medicine* 384(19):1824–1835,2021.

8. Polack Fernando P and Thomas Stephen J and Kitchin Nicholas and Absalon Judith and Gurtman Alejandra and Lockhart Stephen et al. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *New England journal of medicine* 383(27):2603–2615,2020. https://doi.org/10.1056/NEJMoa2034577 PMID: 33301246

9. Chris Baraniuk. What do we know about China's covid-19 vaccines? *bmj*, 373, 2021 British Medical Journal Publishing Group

10. https://github.com/owid/covid-19-data/blob/master/public/data/vaccinations/country_data/China.csv

11. Higdon Melissa M, Wahl Brian, Jones Carli B, Rosen Joseph G, Truelove Shaun A, Baidya Anurima, et al. A systematic review of covid-19 vaccine efficacy and effectiveness against sars-cov-2 infection and disease. In *Open Forum Infectious Diseases*, 2021.

12. Dyer Owen. Covid-19: China stops counting cases as models predict a million or more deaths. *BMJ: British Medical Journal (Online)*, 380(p2),2023. PMID: 36596574

13. Talha BurkiDynamic zero COVID policy in the fight against COVID. *The Lancet Respiratory Medicine*, 10(6):e58–e59,2022. https://doi.org/10.1016/S2213-2600(22)00142-4

14. Brauer Fred and Castillo-Chavez Carlos and Feng Zhilan and others. Mathematical models in epidemiology. *Springer*, 32, 2019.

15. Martcheva Maia. An introduction to mathematical epidemiology. *Springer*, 61, 2015.

16. Samiran Ghosh, Malay Banerjee Multi-strain A Model for COVID-19. Applied Analysis, Optimization and Soft Computing. 2023, Springer Nature Singapore.

17. Maziar Raissi, Ramezani Niloofar and Seshaiyer Padmanabhan. On parameter estimation approaches for predicting disease transmission through optimization, deep learning and statistical inference methods. In *Letters in biomathematics*, pages 1–26, 2019.

18. Maziar Raissi, Perdikaris Paris and George E Karniadakis Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. In *Journal of Computational physics*, pages 686–707, 2019.

19. Shaier Sagi, Raissi Maziar and Seshaiyer Padmanabhan Data-driven approaches for predicting spread of infectious diseases through DINNs: Disease Informed Neural Networks In *Letters in Biomathematics*, pages 71–105, 2019.

20. https://www.worldometers.info/world-population/china-population/#:∼:text=The%20current%20population%20of%20China,of%20the%20total%20world%20population.

21. Li Kangguo, Ruan Feng, Zhao Zeyu, Guo Zhinan, Yang Zimei, Yu Shanshan, et al. Comparative analysis of transmission and vaccine effectiveness in omicron and delta variant outbreaks in china. *Journal of Infection*, 2022. https://doi.org/10.1016/j.jinf.2022.08.018 PMID: 36007658

22. Kathy Leung, Lau Eric HY, Wong Carlos KH, Leung Gabriel M, and Joseph T Wu. Estimating the transmission dynamics of sars-cov-2 omicron bf. 7 in beijing after the adjustment of zero-covid policy in november-december 2022. *Nature Medicine*, pages 1–1, 2023.

23. Lou Lishu, Zhang Longyao, Guan Jinxing, Xiao Ning, Nie Mengli, Wei Yongyue, et al. Retrospective modeling of the omicron epidemic in shanghai, china: Exploring the timing and performance of control measures. *Tropical Medicine and Infectious Disease*, 8(1):39, 2023. https://doi.org/10.3390/tropicalmed8010039 PMID: 36668946

24. Zhao-bing Gao, Yi Wang, Hongying Ji, Xiaoli Zuo, and Bingqing Xia. Inspiration of sars-cov-2 envelope protein mutations on pathogenicity of omicron xbb. *bioRxiv*, pages 2023–01, 2023.

25. Samiran Ghosh, Malay Banerjee, and Vitaly Volpert. Immuno-epidemiological model-based prediction of further covid-19 epidemic outbreaks due to immunity waning. *Mathematical Modelling of Natural Phenomena*, 17:9, 2022. https://doi.org/10.1051/mmnp/2022017

26. Thieme Horst R. *Mathematics in population biology*, volume 1. Princeton University Press, 2018.

27. Pauline Van den Driessche and Watmough James. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Mathematical biosciences*, 180(1-2):29–48, 2002. https://doi.org/10.1016/S0025-5564(02)00108-6 PMID: 12387915

28. Hall Samantha M and Landaverde Lena and Gill Christopher J and Yee Grace M and Sullivan Madison and Doucette-Stamm, et al. Comparison of anterior nares CT values in asymptomatic and symptomatic individuals diagnosed with SARS-CoV-2 in a university screening program. *PLoS One*, 17:7, 2022. https://doi.org/10.1371/journal.pone.0270694

29. Elliott Paul and Haw David and Wang Haowei and Eales Oliver and Walters Caroline E and Ainslie Kylie EC et al. REACT-1 round 13 final report: exponential growth, high prevalence of SARS-CoV-2 and vaccine effectiveness associated with delta variant in England during May to July 2021. *MedRxiv*, Cold Spring Harbor Laboratory Press, 2021.

30. Servellita Venice and Morris Mary Kate and Sotomayor-Gonzalez Alicia and Gliwa Amelia S and Torres Erika and Brazer Noah et al. Predominance of antibody-resistant SARS-CoV-2 variants in vaccine breakthrough cases from the San Francisco Bay Area, California. *Nature microbiology*, 7:2, 2022. https://doi.org/10.1038/s41564-021-01041-4 PMID: 35013591

31. Kazerooni PA and Fararouei M and Nejat M and Akbarpoor M and Sedaghat Z Under-ascertainment, under-reporting and timeliness of Iranian communicable disease surveillance system for zoonotic diseases. In *Public health*, pages 130–135, 2018. https://doi.org/10.1016/j.puhe.2017.10.029 PMID: 29241098

32. Bauch Chris T. Imitation dynamics predict vaccinating behaviour. *Proceedings of the Royal Society B: Biological Sciences*, 272(1573):1669–1675, 2005. https://doi.org/10.1098/rspb.2005.3153

33. S. Cuomo, V. di Cola, F. Giampaolo, G. Rozza, M. Raissi and F. Piccialli Scientific Machine Learning through Physics-Informed Neural Networks: Where we are and What's next In *arXiv*, 2022.

34. Ebrahim Fawzi and Tabal Salah and Lamami Yosra and Alhudiri Inas M and El Meshri Salah Edin and Al Dwigen Samira M et al. Immunity post-COVID-19 recovery boosts the antibody immune response to SARS-CoV-2 vaccination. *MedRxiv*, Cold Spring Harbor Laboratory Press, 2022.

35. https://carnegieendowment.org/2022/12/01/china-s-zero-covid-policy-is-double-edged-sword-pub-88535