medRxiv preprint doi: https://doi.org/10.1101/2022.12.02.22283013; this version posted December 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Research Article

Abhigayan Adhikary, Manoranjan Pal, Raju Maiti, and Palash Ghosh*

Impact of COVID-19 Vaccinations in India - A Statewise Analysis

Abstract:
Objectives. The aim of this paper is to perform a Statewise Analysis of the Second Covid Wave experienced by India using the Gompertz Curves and to assess the role played by vaccinations in reshaping the trajectory of Covid Infections for India. A total of 21 prominent states are chosen for the analysis encompassing 97% of the Indian population. Since the vaccination program in India was rolled out after the First wave of infections had almost subsided, the current analysis is only relevant for the Second Wave.
Methods. We will try to explore how the different properties of the Gompertz Curves can be used as a convenient tool to study the COVID-19 outbreak in India. The impact of vaccinations has also been studied at the state level to assess the extent to which the roll out of vaccination program has augmented the covid scenario of India. Vaccinations have been incorporated in the analysis by taking the daily cumulative number of individuals having the first and second shots of vaccine in each state as the explanatory variables.
Results. The preliminary question that the paper tries to investigate is whether the vaccines capable of checking infection growth. Out of the chosen 21 states, 16 states show positive outcomes with some observable ambiguity for the states of Telangana, West Bengal, Tamil Nadu, Rajasthan and Kerala.
Conclusions. Our analysis found that most of the states (16 out of 21 states) has positive impact of vaccination in reducing the Covid-19 cases.

Keywords: COVID-19; Disease modelling; Generalized Gompertz Curves; Forecasting; Time Series; Vaccinations.

1 Introduction

1.1 Background

Ever since the emergence of COVID -19 and its consequent spread across continents engulfing both advanced and developing nations, there has been a voluminous amount of literature on various aspects of COVID-19. Although a sizable proportion of these contemporary studies highlight the possible robust estimation techniques for modelling the Covid infections, the question of assessing vaccine efficacy is yet to be studied rigorously at the *"macro level"*. The existing literature on *"covid estimations"* indicates that the Logistic model and the ARIMA models have been the two most popular choices among the researchers. Needless to say that a plethora of other sophisticated techniques have also been implemented. But our focus for this paper shall solely be on the studies on India and the Indian subcontinent in general.

In this context, Gupta and Pal (2020) used exploratory data analysis to report the situation during January to March 2020 and used the ARIMA model to predict the future trends. They inferred that a huge surge in the number of likely COVID-19 positive cases was predicted in April and May. The average that was forecasted was a detection of approximately 7000 patients in a total span of 30 days in April. However,

Abhigayan Adhikary, Economic Research Unit, Indian Statistical Institute, Kolkata
Manoranjan Pal, Economic Research Unit, Indian Statistical Institute, Kolkata
Raju Maiti, Economic Research Unit, Indian Statistical Institute, Kolkata
***Corresponding author: Palash Ghosh,** Department of Mathematics, and, Jyoti and Bhupat Mehta School of Health Sciences and Technology, Indian Institute of Technology, Guwahati; AND, Duke-NUS Medical School, National University of Singapore, Centre for Quantitative Medicine, Singapore, Singapore

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2022.12.02.22283013; this version posted December 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

in reality the figures were much higher. Similar studies have been performed by Tak et al. (2021) and Maan et al. (2022) on India, Yousaf et al. (2020) for Pakistan and Aslam et al. (2021) for a comparative study of India, Bangladesh and Pakistan. Ghosh et al. (2020) take a unique approach to forecasting Covid Infections for India. They consider the State wise data of infections and model them using the logistic and exponential curves. They infer that the predictions from one model might be misleading and hence suggest a linear combination of the exponential and logistic curves for the purpose of realistic predictions.

With regard to vaccinations papers have come up on both the epidemiological and empirical fronts in order to understand the vaccine-induced immunity responses. Acuña-Zegarra et al. (2021) considers an extension of the classical Kermack–McKendrick model incorporating vaccinations to explore the disease dynamics that lasts from six months to one year. They infer that vaccine response and its induced immunity are strongly related on the mitigation prevalence. However, vaccine-induced immunity period remains poorly understood and validated their claim using the data on COVID-19 deaths in Mexico-City and Mexico-State However, unambiguously they observed that natural and vaccine-induced immunity play an essential role in reducing COVID-19 disease mortality. Miłobędzki (2022) goes on the anylyze the EU countries by estimating a nonstationary dynamic panel exhibiting the dynamics of confirmed deaths, infections and vaccinations per million population for January to July 2021. The study infers that vaccinations alone would be hardly enough to curb the current and the next waves of the COVID-19 pandemic in the EU countries. Thus, it becomes evident that the debate on vaccine efficacy is far from being resolved.

For India Kumar et al. (2021) has provided some insights on the vaccination status of India through an exploratory data analysis till April 2021. However,the impact of covid vaccinations in India is yet to be analyzed rigorously. The present paper attempts to fill this gap. We go on to explore how the different properties of the Gompertz Curves can be used as a convenient tool to study the COVID-19 outbreak in India and try to evaluate the role of vaccinations to and offer multifaceted benefits in raising timely alarm for country to prepare itself and take measures to tap the onset of another probable covid wave.

1.2 The Second Covid Wave in India

With India yet to settle down from the obnoxious First Wave, and contemplate on its future strategies, the rather new and even more deadly "delta variant" became the watchword for the nation since its first detection in November 2020. Although its presence remained mostly insignificant till the end of February 2021, the India-level data shows that soon after the vaccination programme started on 16th January, the number of variants with significant presence grew in number. A wide range of variants showed up in good proportions by early March. By the end of March, the Delta variant became *"fitter than others"* and started pushing out the other variants. By the end of May 2021, Delta had a presence of 95%.

However, from the economic viewpoint, the heterogeneity in the way the coronavirus has diffused across the states might render an iota of doubt on the correctness of the results if the analysis is conducted based on the India level data. Most of the Indian states are quite large in geographic area and population. Analyzing coronavirus infection data, considering the entirety of India to be on the same page may not provide us the right picture as suggested by Ghosh et al. (2020) . This is because the new infection rate, preventive measures taken by state governments and the pace at which vaccinations were carried out are different for each of the states. Hence, there arises a need to analyze the states separately. Thus, we select 21 Indian states for this analysis based on population which account for 97% of the total Indian population as per the 2020 estimates. Figure 1 provides the list of the chosen states.

medRxiv preprint doi: https://doi.org/10.1101/2022.12.02.22283013; this version posted December 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

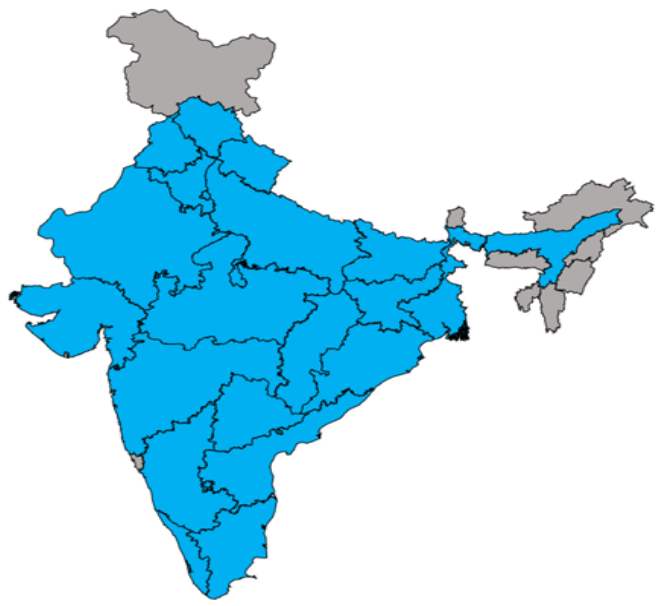

Fig. 1: The highlighted states are included in the study

In light of this above background, the paper is divided into the following sections. Section 2 elaborates on the Methods used in this paper followed by the Data Analysis covered in Section 3. Section 4 highlights our key findings and Section 5 renders the Conclusion of the paper.

2 Methods

2.1 Gompertz Curves and it's properties

Gompertz Curves have nowadays gained ample popularity for modelling economic and biological phenomena and particularly as epidemic curves. However only a handful papers have adapted this for their analysis. Mazurek and Nenickova (2020) used the Gompertz Curve to model the COVID 19 cases of USA and also applied it to the data on COVID-19 deaths. They inferred that in both cases, the Gompertz Curve was able to provide a reasonably good approximation to the data. Medina-Mendieta et al. (2022) also provided a similar conclusion from their analysis on Italy, Spain and Cuba. They considered the Logistic curve and the Gompertz Curve and showed that for both the countries, the Gompertz model had better estimates for the peak in confirmed cases and deaths. Rypdal and Rypdal (2020) made a detailed comparison of epidemic curves for Sweden and Norway using the Gompertz curves and they too observed that the epidemic curves for COVID-19 related deaths for most countries with a reliable reporting system are surprisingly well described by the Gompertz growth model.

The present section reiterates some of the important mathematical properties of this curve. For practical purposes, we consider the Gompertz Curve as:
For practical purposes, we consider the Gompertz Curve(GC) as:

$$y = e^{a-be^{-ct}} \tag{1}$$

where $a, b, c > 0$.Throughout the paper we denote y as the cumulative frequency of daily infections, N as the maximum cumulative frequency that can be attained (e^a), i.e. the asymptote, μ_m be the maximum growth (maximum noncumulative frequency) ($e^{a-1}c$), t^* as the time point at which the maximum growth is attained ($(\ln b)/c$).

medRxiv preprint doi: https://doi.org/10.1101/2022.12.02.22283013; this version posted December 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Note that b is the displacement on the x-axis and c as the growth rate. Essentially, b and c are the shape parameters.
The differential equation representing the Gompertz Curve can be written as:

$$\frac{dy}{dt} = \gamma(y)\, y$$

With the Gompertz function, the daily number of new cases dy/dt, and the daily increase of infected people who tested positive, shows an asymmetric time profile rather than the symmetric one found in the prediction of some of the standard epidemiological models. The following are the desirable properties of the Gompertz Curve for such an analysis:

Property 1:
(1) can also be extended to construct a generalized version of the Gompertz Curve as:

$$y = e^{a - b e^{f(t)}}$$

where $a, b > 0$.

Where $f(t) = c_1 t + c_2 t^2 + \ldots$,is a polynomial in time.
It is to be noted that if we wish to use only a finite number of terms in the power series, we must keep an odd power of t for our highest term, if our curve is to run from y = 0 to an upper limit y = k.
Hence, we consider our Generalized Gompertz Curve(GGC) as:

$$y = e^{a - b e^{c_1 t + c_2 t^2 + c_3 t^3}} \tag{2}$$

where $a, b > 0$.

For the Generalized Gompertz Curve also $N = e^a$ provided $c_3 < 0$.
Subsequently, we will also go on to a further extension and construct the vaccination augmented Gompertz Curves.

Property 2:
Statistically, we can say that in general the sum, or the average, of several Gompertz curves will not be a Gompertz curve, just as several logistics do not, in theory, give a logistic when added or when averaged. But it has been found in practice that the sum of a number of logistics does in fact often approximate closely a logistic as shown by Reed and Pearl (1927). Winsor (1932) comments that it will be true that Gompertz curves will often add to give something very close to a Gompertz curve.

Property 3:
Ohnishi et al. (2020) points out that the Gompertz function appears when the infection probability is an exponentially decreasing function of time. One of the characteristic features of the Gompertz function is the asymmetry of its derivative has a fast rise and slow decay This feature is found in the daily new covid cases $\frac{dy}{dt}$.Hence, with the asymmetric time profile exhibited by the data, the Gompertz curve is expected to be a reasonably good fit to the data.

Property 4:
The Gompertz Curves provides us with an add-on benefit of not only enabling us to make short run predictions (say for a period of 10 days, 20 days etc) but long-term predictions as well. An estimate of the maximum possible cumulative infections can be obtained as $y_{max} = \lim_{t \to \infty} y \; = e^a$.

medRxiv preprint doi: https://doi.org/10.1101/2022.12.02.22283013; this version posted December 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

2.2 Estimation of the models

The Gompertz Curve and the Generalized Gompertz Curves can now be fitted to the data using the method of non-linear least squares. However, to begin with non-linear least sqaures, we first need to obtain initial starting values of our parameters. We get the initial estimates by using the following linearized model. To illustrate the steps, we use the Gompertz Curve.

$$\begin{aligned} y &= e^{a-be^{-ct}} \\ \ln(y) &= a - be^{-ct} \\ \ln(\overline{a} - ln(y)) &= \ln(b) - ct \\ \ln(\overline{a} - \ln(y)) &= b^* + c^* t \end{aligned} \tag{3}$$

Hence with known $\overline{a}$ we can obtain $b = e^{b^*}$ and $c = -c^*$

Thus, the expression for the Modified Gompertz Curve becomes:

$$\ln(\overline{a} - ln(y)) = b^* + c_1 t + c_2 t^2 + c_3 t^3 \tag{4}$$

Hence with known $\overline{a}$ we can obtain $b = e^{b^*}, c_1, c_2, c_3$.

For obtaining the starting values of a we proceed in the following way:
For a, we know $y_{max} = e^a$. From the dataset, this can be taken as the cumulative number of infections on the last day of the wave. We increment this by 20% and initialize a as $\overline{a} = \ln(1.2\ y_{max})$.

2.3 The Vaccination augmented Gompertz Curves(VGC)

Vaccination has been understood as the most prominent external intervention in curbing covid. However as newer and newer variants of the virus are emerging and leading to new waves of infections, the question of how effective the vaccines are for the Indians is yet to be answered rigorously. Thus, we now go on construct a framework to assess how vaccinations have reshaped the covid trajectory of India.

For modelling vaccinations, we consider the following notations and definitions and then suggest a viable extension of the Generalized Gompertz Curve.

Let N_{1t} be the cumulative number of individuals having the first dose and N_{2t} be the cumulative number of individuals having the second dose. Further, define X_t as the cumulative proportion of individuals with the first dose, and Z_t as the cumulative proportion of the individuals with the second dose. Then $X_t = \frac{N_{1t}}{P}$ and $Z_t = \frac{N_{2t}}{P}$, where P denotes the population of the state. Note here that Z_t can also be interpreted as kind of capturing the interaction between the first and the second doses respectively. (i.e to assess the joint contribution as to how much have the first and the second doses done together). However, for our purpose it is sufficient to consider only X_t (because here we want to analyze how an external intervention has modulated one's ability to resist covid).

We start off by assuming that vaccinations have indeed been a meaningful venture. Some initial medical reports suggest that the efficacy of vaccines is observable approximately after a month (i.e after a lag of 30 days). Hence, to introduce vaccines into the Gompertz model, we proceed as follows:

We will show in the subsequent sections that the Generalized Gompertz Curves(GGC) turns out to be more suitable for modelling the Second Wave. Hence, the following modifications are under consideration:

$$\begin{aligned} VM_1 :\ y_t &= e^{a+\delta_1 X_{t-30}+\delta_2 t X_{t-30} - be^{c_1 t + c_2 t^2 + c_3 t^3}} \\ VM_2 :\ y_t &= e^{a - be^{c_1 t + c_2 t^2 + c_3 t^3} + \delta_1 X_{t-30} + \delta_2 t X_{t-30}} \end{aligned}$$

where $\phi_t = \delta_1 X_{t-30} + \delta_2 t X_{t-30}$ captures the impact of vaccinations. We shall be referring to this as the vaccination function. Note that vaccination is expected to lower both the positivity rate and intensity

medRxiv preprint doi: https://doi.org/10.1101/2022.12.02.22283013; this version posted December 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

of infections and thereby exerting a downward force on the number of cumulative infections. Thus, for vaccinations to be deemed meaningful, we expect ϕ_t to be a decreasing function of t.

Thus, to introduce the impact of an external intervention in the form of vaccination, we have considered the above two models. Our motivation behind suggesting this modification has been indicated earlier and can be summarized as follows: One possibility is introduction of vaccination is likely to reduce the maximum number of cumulative infections and thereby lowering the asymptote of the Gompertz curve which is the modification considered in VM_1. Also, another possibility is vaccination has a significant contribution in reducing the pace (i.e the growth rate) at which the infections occur which we have captured in VM_2.

Since the relationship need not necessarily be linear, the cumulative number of cases and vaccinations have been related in the above way. In particular, it is suspected that there exists a critical level of vaccination (say X_{t_0}) beyond which vaccination will be observed to be significantly effective in each of the states. In other words, we are willing to begin with the assumption that there exists a critical level of cumulative vaccination beyond which the effect of vaccinations will be remarkably observable in a state and thereby hinting towards a possibility that 100% vaccinations need not be a necessary target to be fulfilled to guarantee the effectiveness of vaccines.

3 Data analysis

3.1 Data

Data on vaccinated individuals is available from 16th January, 2021 to 9th August,2021 . Data on cumulative number of cases is available till 31st October, 2021. Both datasets were accessed on 29th January, 2022.

The updated data on infections can be downloaded from www.kaggle.com/datasets/vinitshah0110/covid19-india?resource=download

And the statewise data on vaccinations is available at www.kaggle.com/datasets/harveenchadha/india-covid19-vaccination-data?select=vaccination.csv

The entire dataset is segregated into two parts, i.e for the First and the Second Waves and then modelled individually by the Gompertz Curves with some modifications.

medRxiv preprint doi: https://doi.org/10.1101/2022.12.02.22283013; this version posted December 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

3.2 The Indian Scenario

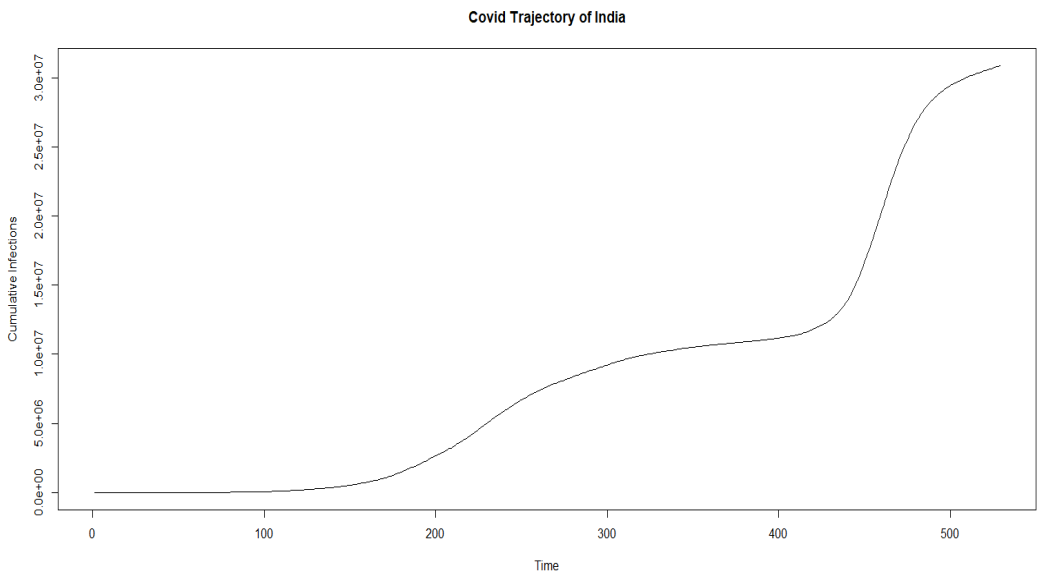

Fig. 2: Daily Cumulative Infections overtime at the All-India level

Figure 2 graphs the country level data over time. As expected, the cumulative number of daily infections follows a sigmoid shape. However, since two major variants of Covid-19 drove the number of affected individuals, we observe two covid waves each having the expected sigmoid shape. Similar plots have been obtained for each of the states chosen for the analysis. However, analysis of the two waves cannot be done together and hence there arises a need to split our datasets into 2 parts for the first and the second waves respectively. However, scientifically it has not been possible to demarcate a well-defined cutoff point for marking the end of the first wave and the beginning of the second wave. Hence, for the purpose of proceeding with the analysis we need to subjectively decide on the cutoff points based on the above plot. Another crucial observation that's worth mentioning is that it has been empirically observed that subjective determination of the cutoff points is not a point of grave concern because the predictions obtained by varying the cutoff points only differ marginally.

An empirically relevant question that needs to be addressed at this juncture is why not use the Logistic Model intead? Well, it is to be noted that the Gompertz Curves is asymmetric with respect to the point of inflexion whereas the Logistic is symmetric. When we desire to fit growth curve which show a point of inflection in the early part of the growth cycle, when approximately 35% to 40% of the total growth has been realized, we may use the Gompertz curve with the expectation that the approximation to the data will be good.

For the coronavirus, there is enough scientific evidence to suggest that the distribution of the daily covid infected cases have a long tail, i.e in the initial phases growth rate of covid infection was much higher as compared to the later phases. The inflection point of the cumulative infections is expected to occur before the half-life of the wave. Hence, Gompertz appears to be a more relevant choice.

3.3 Exploratory data analysis of the Vaccination Program

Note that India started its vaccination program from 16th January, 2021 with Covishield and Covaxin initiating the vaccination drive in the country. Note that this was the time when the first wave had nearly

medRxiv preprint doi: https://doi.org/10.1101/2022.12.02.22283013; this version posted December 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

subsided in most of the states and the second wave was yet to begin. Thus, the analysis on vaccinations is only relevant for the Second Wave(often referred as the *delta wave*).

The best possible statewise data that is available is on the cumulative number of individuals having the first and the second dose respectively for each of the states (daily data is available till 9th August, 2021). Hence, the standard techniques used for assessing the efficacy of vaccines in clinical trials cannot be applied to this aggregate level data simply because an appropriate control group is not available. Hence, evaluating the impact of the vaccination program at the macro level needs further considerations.

The roll out of the vaccination program in India was done in a staggered manner with only the senior citizens(60 years and above) being eligible for getting vaccinated 16th January,2021 onwards. All individuals 18 years and above became eligible for a shot from 1st May,2021. Hence, from Figure 3 it is evident that the pace of the vaccination program escalated sharply only after all adults were incorporated in the program.

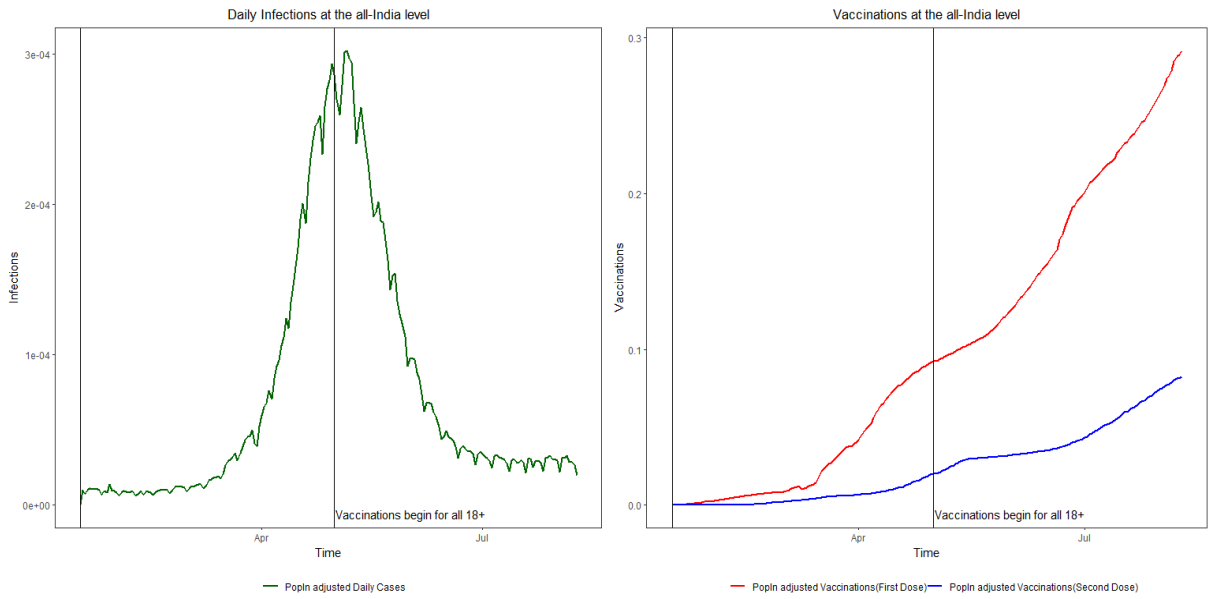

Fig. 3: Daily new infections and Vaccinations at the all-India level

Given the plethora of heterogeneities amongst the states with respect to their implementation and response to the vaccination program, Figure 3 does hint at some association between number of vaccinated individuals and cumulative number of infections. Similar plots are obtained for the states as well. However, a finer assessment needs to be done to evaluate the efficacy of vaccines.

One possibility is that the introduction of vaccination is likely to reduce the maximum number of cumulative infections. This however cannot be directly observed due to the absence of a proper control group.

Also, another possibility is vaccination has a significant contribution in reducing the pace (i.e the growth rate) at which the infections occur. This can be analyzed provided certain adjustments are made to account for the different time points at which the Second wave started in the different states.

To have some further insights, we will now look at the population adjusted daily new infections(=New cases/ Population) and the population adjusted vaccinations(=Cumulative first dose/ Population) post the beginning of vaccinations for different states. To refrain from further heterogeneities, we will now look at the states with comparable populations like Kerala and Assam, each accounting for 2.604% and 2.598% of the Indian population respectively.

Figure 4 presents a comparative analysis of Kerala and Assam from 21st April,2021(the subjectively determined starting point of the Second Wave in Kerala). Note that till this date, the cumulative proportion

medRxiv preprint doi: https://doi.org/10.1101/2022.12.02.22283013; this version posted December 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

of the first dose stood at 15.44% and 4.08% respectively. Thus, vaccinations have been quite high in Kerala as compared to that of Assam. Hence, it is natural to expect that the growth of infections will be lower for Kerala or at least will gradually become lower overtime relative to Assam. However, as evident from Figure 4 this is not the case. The new reported infections show a drop but seems to be bouncing back swiftly and one cannot attribute this observation to a slowdown of the vaccination program in Kerala because the cumulative proportion of vaccinations has been consistently higher in Kerala at all time points for the period under consideration as observed from the second panel Figure 4.

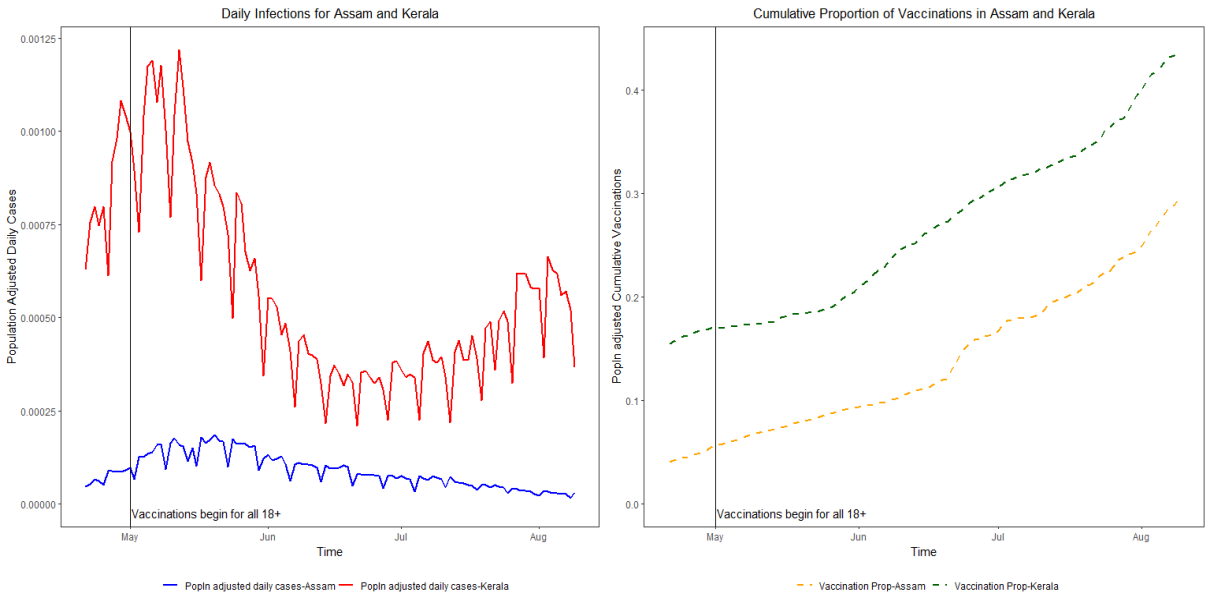

Fig. 4: Daily new infections and Vaccinations for Assam and Kerala

An iota of doubt might linger around the observations from Figure 4 owing to the different geographical attributes of Kerala and Assam. To account for this, another comparative study has been done between Telangana and Kerala, each accounting for 2.8% and 2.604% of the Indian population with the cumulative proportion of First dose standing at 7.7% and 15.44% respectively as on 21st April, 2021. Figure 5 further confirms the fact that higher vaccination rates has not necessarily translated into lower infection rates for India. This means that we cannot unambiguously claim a cent percent efficiency of the vaccines in controlling the spread of covid.

medRxiv preprint doi: https://doi.org/10.1101/2022.12.02.22283013; this version posted December 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

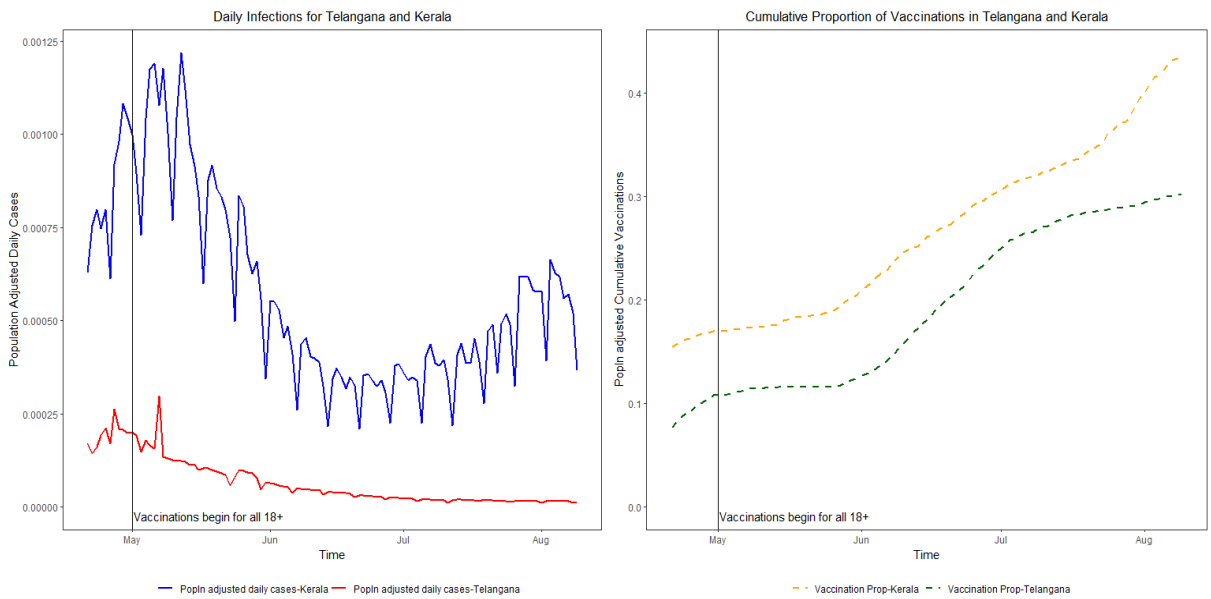

Fig. 5: Daily new infections and Vaccinations for Telangana and Kerala

4 Results

4.1 Modelling the Second Wave through Gompertz Curves

With the subjectively determined cutoff points for the Second Wave for each of the states, we will now use the method of nonlinear least squares for fitting the curves by plugging in the initial estimates. The two models under consideration are:
The Gompertz Curve $M_1 : \ y_t = e^{a-be^{-ct}} + \epsilon_t$
The Generalized Gompertz Curve $M_2 : \ y_t = e^{a-be^{c_1 t + c_2 t^2 + c_3 t^3}} + \epsilon_t$
Which model will fit better is decided on the basis of $\widehat{\sigma}$ and the AIC values as presented in the following tables:

medRxiv preprint doi: https://doi.org/10.1101/2022.12.02.22283013; this version posted December 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Tab. 1: Fitted Models for the Second Wave

State	$\widehat{\sigma}_{M_1}$	$\widehat{\sigma}_{M_2}$
Telangana	8433	2150
Assam	3515	3085
Jharkhand	3155	1030
Bihar	4614	1532
Madhya Pradesh	12440	2277
Himachal Pradesh	3899	2025
Gujarat	14850	2667
Chhattisgarh	9552	3981
West Bengal	12670	5505
Odisha	5436	4779
Uttarakhand	4386	1383
Andhra Pradesh	10410	9589
Karnataka	25580	15840
Maharashtra	77840	50410
Punjab	16350	11240
Tamil Nadu	35330	12340
Haryana	12590	1465
Uttar Pradesh	77840	50410
Rajasthan	12370	1956
Delhi	12680	3670
Kerala	83030	13830

Hence, we can observe that unanimously M_2 is a better fit in all the states for both the waves (as indicated by lower $\widehat{\sigma}$). Note that only the $\widehat{\sigma}$ values are reported as our conclusions are unchanged under AIC comparision.

4.2 Choosing the appropriate Vaccination augmented Gompertz Curves

Hence, now the two models under consideration are VM_1 and VM_2 which were formulated as:

$$VM_1: \; y_t = e^{a+\delta_1 X_{t-30}+\delta_2 t X_{t-30}-be^{c_1 t+c_2 t^2+c_3 t^3}}$$
$$VM_2: \; y_t = e^{a-be^{c_1 t+c_2 t^2+c_3 t^3}+\delta_1 X_{t-30}+\delta_2 t X_{t-30}}$$

Both these models are fitted for all the states and the one with the lower AIC is considered to be a better fit for each of states. Our observations can be summarized as follows:

medRxiv preprint doi: https://doi.org/10.1101/2022.12.02.22283013; this version posted December 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Tab. 2: Comparision of AIC for VM_1 and VM_2

States	$AIC(VM_2)$	$AIC(VM_1)$
Telangana	3019.073	3083.036
Assam	3583.145	3591.811
Jharkhand	3277.295	3349.816
Bihar	3319.826	4095.825
Madhya Pradesh	3533.882	3713.693
Himachal Pradesh	3608.961	3797.106
Gujarat	3723.879	3766.842
Chhattisgarh	3875.273	4232.968
West Bengal	3409.332	3738.131
Odisha	3531.198	5310.312
Uttarakhand	2966.636	4350.991
Andhra Pradesh	4481.98	5733.074
Karnataka	4421.959	4548.075
Maharashtra	4588.743	4966.915
Punjab	3698.787	3858.579
Tamil Nadu	3884.966	4410.484
Haryana	3029.44	3962.11
Uttar Pradesh	3793.34	3882.07
Rajasthan	3969.16	4582.63
Delhi	3716.84	4276.73
Kerala	3051.42	3127.14

Thus, it is evident that VM_2 is better fit for all the states. Hence, we shall now proceed with VM_2 for the subsequent analysis.

4.3 Determination of the Optimal Lag(h)

Having chosen the appropriate model, we now turn to a more important question i.e what is the optimal lag after which the effect of vaccinations is actually observable? Hence, now we will modify the lag for VM_2 and fit the model for each state to obtain the optimal lag. The lag corresponding to which we have the highest number of states exhibiting X_{t_0} will be the optimal. Given the available data, the following are the lags that can be analyzed.

Define: h=lag length, h $\in \{10, 20, \ldots, 50\}$ We can now rewrite VM_2 as:

$$VM_2 : y_t = e^{a - be^{c_1 t + c_2 t^2 + c_3 t^3 + \delta_1 X_{t-h} + \delta_2 t X_{t-h}}} + \epsilon_t \tag{5}$$

The number of states exhibiting this criterion for different lag values can be seen in the Figure 6 below.

medRxiv preprint doi: https://doi.org/10.1101/2022.12.02.22283013; this version posted December 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

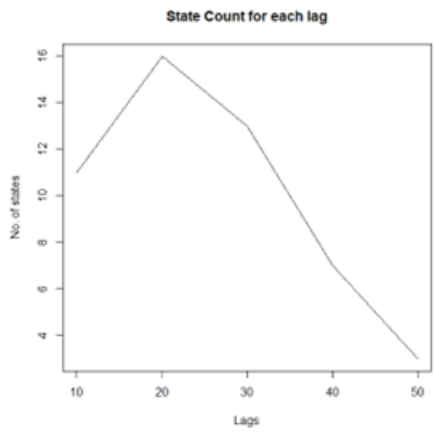

Fig. 6: Plot of No. of States vs Lag length

It is clear that the optimal lag value turns out to be $h^{opt} = 20$ from which we obtain X_{t_0} for 16 out of the selected 21 states. Although X_{t_0} is varying sharply across the states, the existence of such a value for the majority of the chosen states suggests that our assertion was indeed correct. Results from our analysis can be summarized as follows:

Tab. 3: Estimated X_{t_0} *for all states for lag=20.*

States	Cumulative vaccinations (X_T)	$X_{t_0}^{opt}$ for lag=20
Telengana	30.25	_
Assam	29.48	2.66
Jharkhand	21.72	6.56
Bihar	18.71	5.02
Madhya Pradesh	34.82	8.5
Himachal Pradesh	57.03	13.95
Gujarat	44.00	10.81
Chhattisgarh	31.19	15.5
West Bengal	23.35	_
Odisha	30.10	11.34
Uttarakhand	45.07	11.79
Andhra Pradesh	32.70	22.48
Karnataka	38.26	9.22
Maharashtra	28.46	0.02
Punjab	26.56	7.12
Tamil Nadu	26.77	_
Haryana	35.76	7.18
Uttar Pradesh	19.31	4.93
Rajasthan	33.33	_
Delhi	41.88	15.77
Kerala	43.90	_

In third column empty cells indicates that no such $X_{t_0}^{opt}$ is obtained for the particular states.

Having obtained the optimal model and the optimal lag, the impact of vaccinations can be easily observed by looking at the vaccination function: $\phi_t = \delta_1 X_{t-20} + \delta_2 t X_{t-20}$ for each of the states. The following two types of observations are made:

medRxiv preprint doi: https://doi.org/10.1101/2022.12.02.22283013; this version posted December 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Case I: A critical level of vaccination proportion (X_{t_0}) is observable.
For example: Madhya Pradesh

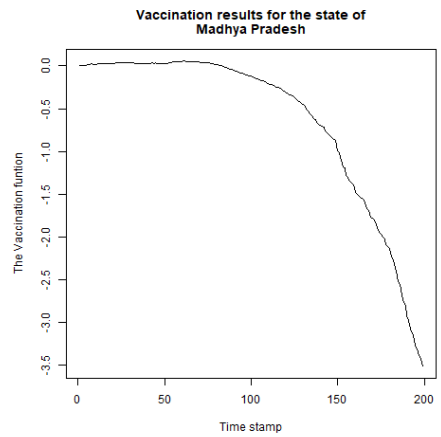

Fig. 7: Plot of ϕ_t vs t for Madhya Pradesh

Case II: A critical level of vaccination proportion (X_{t_0}) is not observable.
For example: West Bengal

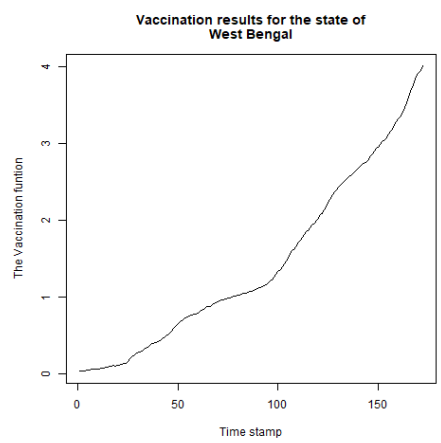

Fig. 8: Plot of ϕ_t vs t for West Bengal

From the table we see that there are 5 states for which there doesn't exist any X_{t_0} namely: Telangana, West Bengal, Tamil Nadu, Rajasthan, Kerala. Now, here it is important to note that with the data at hand, X_{t_0} was not observable for these states. However, if more data is available, these states can be re-examined for this possibility.

4.4 Comparing the GGC and VGC Models

Till now our focus has been on trying to examine the impact of the vaccination program for the Indian states. However, one of the common objectives of any time series analysis is to perform robust predictions. Thus, in order to assert that vaccinations have been a meaningful inclusion in modelling the covid infections,

medRxiv preprint doi: https://doi.org/10.1101/2022.12.02.22283013; this version posted December 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

the VGC has to be compared with the GGC in terms of which of the two models gives us better predictions of the cumulative covid infections.

Hence, we divide the data available for the states into the training and test datasets and provide a short term 10-days prediction and compare the mean square errors(MSE) based on the two models. The model giving the lower test-MSE will be better in terms of prediction. The results are summarized as follows:

Tab. 4: 10 days prediction comparision for the two models

States	Pred(start)	Pred(end)	MSE(Gomp)	MSE(Vac)	Choice
Telangana	31-07-2021	09-08-2021	1807.63	239.77	1
Assam	31-07-2021	09-08-2021	736.63	1940.35	0
Jharkhand	31-07-2021	09-08-2021	1369.24	1222.63	1
Bihar	31-07-2021	09-08-2021	2327.26	1626.53	1
Madhya Pradesh	31-07-2021	09-08-2021	1767.74	1679.33	1
Himachal Pradesh	31-07-2021	09-08-2021	8074.99	5747.32	1
Gujarat	31-07-2021	09-08-2021	2249.47	1459.29	1
Chhattisgarh	31-07-2021	09-08-2021	5662.96	3596.28	1
West Bengal	31-07-2021	09-08-2021	16573.27	16246.99	1
Odisha	31-07-2021	09-08-2021	8749.36	7771.73	1
Uttarakhand	31-07-2021	09-08-2021	1968.76	1316.84	1
Andhra Pradesh	31-07-2021	09-08-2021	31516.36	30401.4	1
Karnataka	31-07-2021	09-08-2021	38141.35	39246.84	0
Maharashtra	31-07-2021	09-08-2021	164319.1	105972.1	1
Punjab	31-07-2021	09-08-2021	22228.6	1722.45	1
Tamil Nadu	31-07-2021	09-08-2021	48744.2	22016.11	1
Haryana	31-07-2021	09-08-2021	2002.17	1760.22	1
Uttar Pradesh	31-07-2021	09-08-2021	2516.02	2580.59	0
Rajasthan	31-07-2021	09-08-2021	1343.03	1530.53	0
Delhi	31-07-2021	09-08-2021	3462.38	3817.16	0
Kerala	31-07-2021	09-08-2021	49609.69	42491.37	1

(Choice =1 indicates that the vaccination augmented model is better and Choice=0 otherwise)

From the above table, we can observe that we get similar conclusions as earlier with the vaccination augmented Gompertz Fit giving better predictions for 16 of the 21 states with the exception of Assam, Karnataka, Uttar Pradesh, Rajasthan and Delhi. Hence, we can observe that the vaccination augmented Gompertz Curve has given us significant insight into the impact of vaccinations in India as it gives a higher prediction accuracy of the majority of the states under consideration.

5 Discussion and Conclusion

The current paper takes a sequential approach to facilitate the inclusion of vaccinations in modelling the cumulative number of Covid infections. As observed for the state-level data, the Generalized Gompertz gives a better fit for all the states. Vaccinations seem to be a meaningful inclusion in modelling the daily cumulative infections and the optimal lag has been obtained to be 20 days at the *"macro level"*, i.e the effectiveness of the vaccines can be strongly seen after 20 days of the first dose for 16 of the chosen 21 states. This has been backed by our observations that in these states we have been able to observe a $X_{t_0} < 1$ where the vaccination function becomes a decreasing function of time. Hence, 100% vaccination is not the gold standard that needs to be necessarily achieved to prove that vaccinations have indeed been effective.

An attempt has been made to answer the burning issue: Are the vaccines capable of checking infection growth? Our analysis suggests that the claim cannot be fully upheld with certainty at the aggregate level.

medRxiv preprint doi: https://doi.org/10.1101/2022.12.02.22283013; this version posted December 6, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Although there are a myriad of factors influencing the surge of covid cases, the fact that 5 states which are home to 24.28% of Indians not showing visible effects of vaccinations is indeed an observation that cannot be relegated to the background.

There is no gainsaying that one might attribute the counter intuitive conclusions found in the 5 states coming out as a result of some other factor(s) that might dominate the vaccination effect.However, inclusion of these nonmedical interventions(like lockdowns, quarantine etc) has not been possible in this analysis due to lack of appropriate and reliable data.Further,the study by Mir et al. (2021) on India finds that risk perceptions and social media exposure has nearly insignificant influence on people's attitudes towards Covid-19 vaccinations. Social norms, trust, and people's attitudes towards the Covid-19 vaccinations are are the key factors driving their intentions to take up Covid-19 vaccinations. Hence the question of vaccine hesitancy cannot be ruled out. Further research needs to be done to address the question of acquired immunity vs natural immunity which continues to remain the fulcrum of assessing vaccine efficacy. The future prospects of this analysis can be focussed on developing improved methodologies to include the other seemingly relevant variables subject to the data availability and thereby robustifying our process of filtering out the vaccination effect on the infections spread.

Data Availability Statement:

The updated data on infections can be downloaded from www.kaggle.com/datasets/vinitshah0110/covid19-india?resource=download, and the statewise data on vaccinations is available at www.kaggle.com/datasets/harveenchadha/india-covid19-vaccination-data?select=vaccination.csv

References

Acuña-Zegarra, M. A., Dıaz-Infante, S., Baca-Carrasco, D., & Olmos-Liceaga, D. (2021). Covid-19 optimal vaccination policies: A modeling study on efficacy, natural and vaccine-induced immunity responses. *Mathematical biosciences*, *337*, 108614.

Aslam, F., Awan, T. M., Khan, R., Aslam, M., & Mohmand, Y. T. (2021). Prediction of covid-19 confirmed cases in indo-pak sub-continent. *The Journal of Infection in Developing Countries*, *15*(03), 382–388.

Ghosh, P., Ghosh, R., Chakraborty, B., et al. (2020). Covid-19 in india: Statewise analysis and prediction. *JMIR public health and surveillance*, *6*(3), e20341.

Gupta, R., & Pal, S. K. (2020). Trend analysis and forecasting of covid-19 outbreak in india. *MedRxiv*.

Kumar, S., Kumar, S., Singh, A., & Raj, A. (2021). Covid-19 data analysis and prediction using (machine learning) and vaccination update of india. *Available at SSRN 3847564*.

Maan, S., Devi, G., & Rizvi, S. (2022). Prediction of third covid wave in india using arima model. *J. Sci. Res*, *66*(2).

Mazurek, J., & Nenickova, Z. (2020). Predicting the number of total covid-19 cases and deaths in the usa by the gompertz curve. *Accessed: Jun*, *23*.

Medina-Mendieta, J. F., Cortés-Cortés, M., & Cortés-Iglesias, M. (2022). Covid-19 forecasts for cuba using logistic regression and gompertz curves. *MEDICC review*, *22*, 32–39.

Miłobędzki, P. (2022). Are vaccinations alone enough to curb the dynamics of the covid-19 pandemic in the european union? *Econometrics*, *10*(2), 25.

Mir, H. H., Parveen, S., Mullick, N. H., & Nabi, S. (2021). Using structural equation modeling to predict indian people's attitudes and intentions towards covid-19 vaccination. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, *15*(3), 1017–1022.

Ohnishi, A., Namekawa, Y., & Fukui, T. (2020). Universality in covid-19 spread in view of the gompertz function. *Progress of Theoretical and Experimental Physics*, *2020*(12), 123J01.

Reed, L. J., & Pearl, R. (1927). On the summation of logistic curves. *Journal of the Royal Statistical Society*, *90*(4), 729–746.

Rypdal, K., & Rypdal, M. (2020). A parsimonious description and cross-country analysis of covid-19 epidemic curves. *International Journal of Environmental Research and Public Health*, *17*(18), 6487.

Tak, A., Dia, S., Dia, M., & Wehner, T. C. (2021). Indian covid-19 dynamics: Prediction using autoregressive integrated moving average modelling. *Scripta Medica*, *52*(1), 6–14.

medRxiv preprint doi: https://doi.org/10.1101/2022.12.02.22283013; this version posted December 6, 2022. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-NC-ND 4.0 International license .

Winsor, C. P. (1932). The gompertz curve as a growth curve. *Proceedings of the national academy of sciences*, *18*(1), 1–8.

Yousaf, M., Zahir, S., Riaz, M., Hussain, S. M., & Shah, K. (2020). Statistical analysis of forecasting covid-19 for upcoming month in pakistan. *Chaos, Solitons & Fractals*, *138*, 109926.