



# Restricted Prevalence Rates of COVID-19's Infectivity, Hospitalization, Recovery, Mortality in the USA and Their Implications

Ramalingam Shanmugam<sup>1</sup> 

Received: 28 May 2020 / Revised: 7 September 2020 / Accepted: 25 September 2020 /

Published online: 9 October 2020

© Springer Nature Switzerland AG 2020

## Abstract

This article constructs and demonstrates an alternate probabilistic approach (using incidence rate restricted model), compared with the deterministic mathematical models such as SIR, to capture the impact of healthcare efforts on the prevalence rate of the COVID-19's infectivity, hospitalization, recovery, and mortality in the eastern, central, mountain, and pacific time zone states in the USA. We add additional new properties for the *incidence rate restricted Poisson* probability distribution. With new properties, our method becomes feasible to comprehend not only the patterns of the *prevalence rate* of the COVID-19's infectivity, hospitalization, recovery, and mortality but also to quantitatively assess the effectiveness of *social distancing*, healthcare management's efforts to hospitalize the patients, the patient's immunity to recover, and lastly the unfortunate mortality itself. To make regional comparisons (as the people's movement is far more frequent within than outside the regional zone on daily basis), we group the COVID-19 data in terms of eastern, central, mountain, and pacific zone states. Several non-intuitive findings in the data results are noticed. They include the existence of imbalance, different vulnerability, and risk reduction in these four regions. For example, the impact of healthcare efforts is high in the recovery category in the pacific states. The impact is less in the hospitalization category in the mountain states. The least impact is seen in the infectivity category in the eastern zone states. A few thoughts on future research work are cited. It requires collecting rich data on COVID-19 and extracting valuable information for better public health policies.

**Keywords** Operational efficiency · Data analytics · Imbalance · Vulnerability · Risk reduction

---

Honorary Professor of International Studies, School of Health Administration, Texas State University.

✉ Ramalingam Shanmugam  
Shanmugam@txstate.edu

<sup>1</sup> International Studies, School of Health Administration, Texas State University, San Marcos, TX 78666, USA

## 1 Introduction

In this frightening time of COVID-19, professionals in the field of public health are drowning with lots of data. Not much practical findings are extracted from the data either in favor of the current public health policies such as lockdown, no travel, masking the face, no public interactions in shopping, and in restaurants or in need new policies to reduce if not total elimination of the COVID-19's incidence rate. This article, using modeling as an approach, addresses whether a significant reduction in the parametric space of the incidence rate has occurred due to various current public health policies cited above with respect to each one of the four categories: infectivity, hospitalization, recovery, mortality. Mathematical models are the generally convenient vehicle to grasp the dynamics of the epidemics and uncertainty that prevent or spread the epidemics in the population (see Okabe and Shudo [8] for details). Once the significant factors that spread the epidemics are identified and quantified, that would assist the policymakers to formulate accurate decisions to reduce first and eliminate later the epidemics (see Yang et al. [15] for details). Though several variations exist, the basic model that describes the transitions of a population in an epidemic to susceptible, infectious, and recovery state based on ordinary differential equations (connecting the contact and cure rates) is called the SIR model. The SIR model does not incorporate the uncertainty principle. Sometimes, an extended version of the SIR is recognized as a compartment model in the literature of infectious diseases (see Tolles, Luong [13] for details). The ideas behind the SIR model originated from Kermack and McKendrick [6] which discussed an epidemic of an infectious disease in a large population. The SIR model perhaps oversimplifies a complex epidemic process like COVID-19. It assumes a homogeneously mixing population, meaning that all individuals in the population are assumed to have an equal probability of meeting one another. This assumption does not match human social structures, in which most of the contact occurs within limited vicinity or networks. In addition, the parameters in a traditional SIR model do not allow quantification of uncertainty as model parameters, although no model is perfect in so far as predicting closely the incidence rate of the epidemics and/or offering some help to formulate viable public health policy. Recently, using the SIR model, Ambrosio and Aziz-Alaoui [1] took into consideration the daily fluxes between New Jersey (NJ) and New York (NY) states to notice a decrease in the COVID-19 incidences due to their lockdown policies. They also noted that a wave in NJ occurred following the NY wave, illustrating the mechanism of spread from one attractive hot spot to its neighbor. Therefore, we offer a viable alternative approach based on probabilistic principles to deal with uncertainty in the spread and reduction of the COVID-19 incidences in US states.

Our approach is new, novel, and interesting from the view of extracting and interpreting pertinent data informatics for the sake of better public health. Our approach helps the administrators and policymakers to formulate and adapt refined healthcare public policies to reduce the incidences of the COVID-19 because of the interpretable parameters and computable expressions of the IRRP model. The US data are portioned into four regions (eastern including District of Columbia, central, mountain, and pacific time zone states, Alaska and Hawaii are added to the pacific states) to compare the regional performances and/or differences. Some factors like population density, reaction time to shut down the state, or even the temperature could affect the prevalence rate

directly. Also, the people's movements are more frequent within than out of the regions on daily basis and it is a reason for the partitioning of data in terms of time zones. Even inside these four time zones, the data are highly imbalanced. There is no significant outlier in any time zone and hence, our results based on the mean and variance are robust. However, the manuscript discovers some fascinating differences and similarities among the four regions including non-intuitive findings about the existence of imbalance, different vulnerability, and risk reduction.

Whether it is a natural or human manufactured, a mystic pathogenic coronavirus (also known as COVID-19) was detected to have caused respiratory infections in humans as early as 24 January 2020, in Wuhan city (capital of the Hubei province), China [5]. The name coronavirus is derived from the Latin word “corona” meaning “crown.” The government of China reported that there were 1287 coronavirus cases, causing 41 death among them (see [16] for details). The first coronavirus case was traced to have happened sometime during December 2019 (see [14]).

When the World Health Organization [7] was informed by China, this deadly coronavirus was given the name *COVID-19* as deadly disease of the year 2019. First, it was falsely mentioned that the COVID-19 was contagious from dead or alive animals and the virus did not spread from a human to another human [9].

How does coronavirus spread? The COVID-19 is a single-stranded RNA. Even after a full recovery from treatment in a hospital, a person could get inflicted by the coronavirus again, according to Japan's Broadcasting Corporation *Nippon Hōsō Kyōkai* (NHK). This dampens the hope of the drug or vaccination discovery process. Even with the availability of effective vaccination, everyone needs to be annually vaccinated. However, as of now, there is no vaccination or preventive medicine to consume to avoid coronavirus. The best people can do is to practice *social distancing, lockdown, no travel, masking face, no public interactions in shopping, in restaurants, etc.* to prevent COVID-19's spread.

There is no perfect diagnostic either to determine the existence of coronavirus in a person. The high body temperature, dry cough, lack of smelling odor or tasting food, shortness of breath, and/or respiratory distress are primitive symptoms. Any diagnostic of the above symptoms exhibited much more false negatives and false positives to frustrate the healthcare professionals to recommend quarantine the potential coronavirus cases.

Despite this pandemonium, China discovered that on 25 March 2020, an estimated 81,285 persons contracted the coronavirus and 3287 of them died. Alarmed by this rapid increase, the Chinese government quarantined the people in Wuhan city in the beginning and in Hubei province later based on an unproven knowledge that the incubation time between the exposure and disease onset was estimated to be 2 to 14 days [2].

There was (in fact, even now) no specific effective anti-viral medicine to treat COVID-19. Seeing the coronavirus infectivity gets out of the manageable situation, the people were advised to wash their hands using soap, not to touch their eyes, nose or face, wear a facial mask, and stay away from each other by 6 ft to maintain a *social distancing* to avert the spread of this contagious virus.

On 30 January 2020, the Center for Disease Control (CDC) in the USA came to know the first COVID-19 case in Chicago as he returned 17 days earlier from Wuhan city in China. Subsequently, five more COVID-19 cases occurred in the USA. On 31

January 2020, the President of the USA signed an executive order to deny entry to anyone originating or traveling through China.

After noticing 12,307 COVID-19 cases with death of 259 persons worldwide as of 1 February 2020, the WHO declared COVID-19 a *pandemic* as it is a global health emergency. The pandemic spread to almost all continents now on earth. Some nations resorted quickly to quarantine the COVID-19 suspected cases and others imposed a severe curfew. Many nations, including the USA, adapted social distancing of people from each other first and totally shutting down the nation later. Consequently, the commercial sectors, travels, parks, schools, and universities were locked down in the USA and in almost 198 nations around the world.

As of today (17 May 2020), the prevalence of COVID-19 cases is more than 4,425,485 and at least 302,059 deaths have occurred already worldwide. The numbers are staggering and changing daily. The gravity of the health hazards and the fear of death cannot be adequately described. The day to day activities were crippled. The productivity reached near zero everywhere on earth.

The research theme of this article is therefore to construct an approach to detect how effective the social distancing, locking down the nation, etc. have been to reduce the prevalence of COVID-19, to detect how efficiently the infected cases have been hospitalized by the local/state/federal agencies, to detect how optimally the COVID-19 patients recover in the treatment, and lastly to detect how best the COVID-19 mortality is disposed of. For this purpose, we construct an approach based on the incidence of restricted Poisson chance principles and derive much needed algebraic expressions to detect the abovementioned goals. To check whether our novel approach works, we involve the prevalence of COVID-19 cases, the number of them hospitalized, the number of recovered cases from coronavirus, and how many died in the 50 states and the District of Columbia (DC). The data are drawn from a public domain webpage <http://kff.org>. For optimal data analyses, we grouped the data in terms of eastern, central, mountain, and pacific time zone states because we need enough samples to estimate the sampling errors to make a statistical assessment. Of course, the Alaska and Hawaii states were included in the pacific zone, while the DC is included in the eastern zone.

Utilizing the derived core analytic expressions, we analyze and interpret the data evidence to compare and contrast the operational efficiencies in all the four zones with respect to social distancing, adapting hospitalizations, offering optimal treatment for recovery, and the pattern of inevitable mortality due to either the ugly virality of the coronavirus and/or individual's lack of strong immunity to resist. We believe that the approach and research findings (not thorough or complete in any sense) for the US regions will form a basis to come forward with a viable, consistent, robust future healthcare delivery of services, drugs, and facilities.

## 2 Realistic Underlying Probability Model for COVID-19-Driven Healthcare System

Of course, a pandemic like COVID-19 does not follow a mathematical regulation or its precision. We, the healthcare management analysts, postulate a chance-oriented model

only for the sake of easy comprehension of its complex outcomes. Several critics of apriority model selection argue that all models are wrong. Their point of view also makes sense. Sometimes, some model works. If a model works, then why not benefit from it?

One such model for the COVID-19-driven health system could be the *incidence rate restricted Poisson* (IRRP). The IRRP distribution was originally presented by Shanmugam [10]. It fits our scenario here because the rates in all categories (they are infectivity, hospitalization, recovery, and mortality) receive an impact due to practical restrictions with respect to COVID-19. If the IRRP works, it helps to learn and interpret the patterns and predictability of the COVID-19 as they are illustrated in the article. Also, we will be knowledgeable enough to present a feasible reform for the optimal healthcare delivery to deal with any future contingency due to pandemics like this COVID-19.

The COVID-19 data for fifty-one US states (including the DC) in <http://kff.org> are categorized in terms of the *confirmed corona cases*,  $y_1$ , the *number hospitalized*,  $y_2$ , the *number recovered*,  $y_3$ , and the *number of mortality*,  $y_4$  due to corona. These are nonnegative Poisson random variables (RV) with positive prevalence rates  $\theta_1 > 0$ ,  $\theta_2 > 0$ ,  $\theta_3 > 0$ , and  $\theta_4 > 0$  respectively. The core idea in our approach is based on recognizing the existence of impacts on the prevalence rate due to the imposed restrictions. The impact is here quantified as an unknown parameter,  $\delta > 0$ . In other words, without the restrictions, the domain for the prevalence rate in every category would have been  $(0, \infty)$  with the existence of restrictions and their impact, the prevalence rate is upper bounded to be in the domain  $(0, \delta)$  in the sense  $0 < \theta < \theta + \delta$  in every category. The social distancing, closure of transportation, closing of schools, shops, industries, curfew, etc. are restrictions on the COVID-19's infectivity. Making extra beds in a hospital, quickly constructing new hospitals, swift transferring COVID-19 infected cases to nearby hospitals, etc. are a few examples of impacting the hospitalization rate. Importing additional physicians, nurses, medicines, drugs, and ventilators to some hospitals in which there is an overcrowding of COVID-19 patients are impacts on the recovery. The final but not least important is the COVID-19's mortality rate which is impacted by the vitality and individual's strong immunity to resist, to fight the ugly virality of the coronavirus. Smaller an estimated  $\hat{\delta}$  amount would refer to a stronger impact due to the imposed restrictions and vice versa. It is in this framework; we bring in the IRRP model (1) below to educate us.

$$P(y_i) = \left(1 + \frac{y_i}{\theta_i + \delta_i}\right)^{y_i-1} \left(\theta_i e^{-\left\{\frac{\theta_i}{\theta_i + \delta_i}\right\}}\right)^{y_i} / y_i! e^{\theta_i}; y_i = 0, 1, 2, \dots, \infty; i = 1, 2, 3(1)$$

When the restrictions are weaker or non-existent, it would imply that  $\delta \rightarrow \infty$ , and the model (1) becomes a disastrous Poisson scenario in which the prevalence rate.  $\theta$  could be unbounded in its vicinity  $(0, \infty)$ . The characteristic property of the (disastrous) Poisson situation would echo in the relationship between its expected value and variance and that relationship is  $E(y) = Var(y)$ . Otherwise (that is, with a finite amount of impact due to the restrictions), there would arise an *imbalance* between the expected value and the variance as shown below in (2). That is,

$$E(y_i) = \left(1 + \frac{\theta_i}{\delta_i}\right)^{-2} \text{Var}(y_i) \tag{2}$$

where the *imbalance factor*,  $\left(1 + \frac{\theta_i}{\delta_i}\right)^{-2}$  approaches value one (see Fig. 1 for the imbalance equation) as  $\delta_i$  moves to infinite for  $i = 1, 2, 3$ . The parameters  $\delta_1, \delta_2, \delta_3$ , and  $\delta_4$ , therefore, portray the impact level due to the restrictions on COVID-19’s infectivity, hospitalization, recovery, and mortality rates. When the imbalance factor attains one, the configuration in Fig. 1 will unbend to look like a 3-dimensional plate. The maximum likelihood estimate (MLE) of the prevalence rate is  $\hat{\theta} = \sqrt{\frac{y^3}{s_y^2}}$  and the impact parameter is  $\hat{\delta} = \left\{ \frac{\sqrt{y}}{\sqrt{s_y^2 - y}} \right\} \hat{\theta}$ .

Alternately, the proportion,  $\left(\frac{\delta_i}{\theta_i + \delta_i}\right)^2$  is interpreted as a measure of *effective diagnostic* to confirm the *coronavirus*, a measure of *sufficiency to hospitalize*, a measure of *appropriate recovery treatment*, and a measure of inevitable *mortality* with  $i = 1, 2, 3, 4$  respectively.

In a data analysis, one wonders whether the MLE  $\hat{\delta}$  of the impact parameter is statistically significant. To answer the above question, we must deal with the bivariate stochasticity of the correlated sample mean and variance. Only when the sample is drawn from a Gaussian population (see [12]), the sample mean and variance are stochastically independent. Otherwise, they (sample mean,  $\bar{y}$  and variance  $s_y^2$ ) are correlated. When the sample mean is less (greater) than the variance, it is recognized as overdispersion (underdispersion) in data informatics. Realize our COVID-19 data are drawn from an IRRP, not a Gaussian population. After algebraic simplifications (see in Appendix at the end of this article), we have an expression to find the  $p$  value for the MLE  $\hat{\delta}$  to be significantly low enough to cap the incidence rate and it is

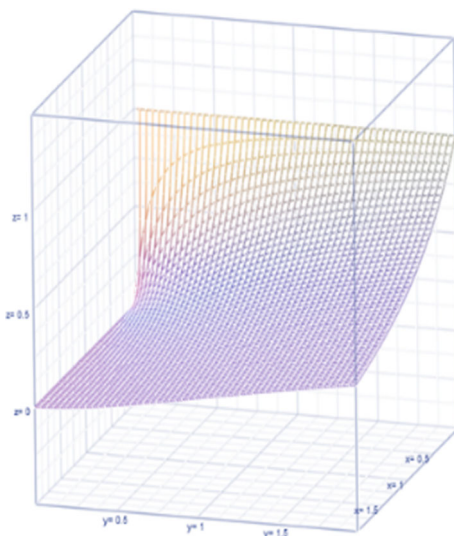


Fig. 1 Imbalance factor

$$p \text{ value} \approx \Phi \left\{ \left| 1 - \frac{\bar{y}}{s_y^2} \right| \sqrt{n\hat{\theta} / (1 + \hat{\theta})} \right\} \tag{3}$$

The asymptotic (as  $n \rightarrow \infty$ ) statistical power is the probability of accepting a true alternative hypothesis  $H: \delta = \delta^*$ , where  $\delta^*$  is a chosen value to assess the impact and it is

$$\text{power} \approx 1 - \Phi \left( \frac{\left\{ z_\epsilon \sqrt{\left( 1 + \frac{1}{\hat{\theta}} \right)} \right\}}{\sqrt{1 + \frac{1}{\hat{\theta}} + \frac{1}{\delta^*}}} \right) \tag{4}$$

after algebraic simplifications.

Ideally, every healthcare management desires to have zero COVID-19 in each category. Its complement is then to have one or more cases. The probability of noticing one or more cases in any category among the corona confirmation, hospitalization,

recovery, and mortality category (see the Appendix for details) is  $S_y[1, \theta, \delta] \approx$

$e^{-\theta \left[ e^{-\left\{ \frac{\theta}{\theta + \delta} \right\}} - 1 \right]} \Pr \left( \chi_{2df}^2 < 2\theta e^{-\left\{ \frac{\theta}{\theta + \delta} \right\}} \right)$  after algebraic simplifications, where  $\chi_m^2$  is the

chi-squared random variable with  $m$  degrees of freedom. Now, we define the *odds* of attaining the ideal situation of no incidence amid the best possible healthcare impact

due to the restrictions is  $\text{odds}_{\delta \neq \infty} = \frac{1 - S_y[1, \theta, \delta]}{S_y[1, \theta, \delta]} \approx \frac{e^{-\theta \left[ e^{-\left\{ \frac{\theta}{\theta + \delta} \right\}} - 1 \right]}}{\Pr \left( \chi_{2df}^2 < 2\theta e^{-\left\{ \frac{\theta}{\theta + \delta} \right\}} \right)} - 1$ . The counterpart

odds of attaining the ideal situation of no incidence amid no restriction is  $\text{odds}_{\delta = \infty} \approx$

$\frac{1}{\Pr \left( \chi_{2df}^2 < 2\theta \right)} - 1$ . Hence, the constant one adjusted odds ratio (OR) to attain the ideal

situation under restrictions compared with no restriction is

$$\text{OR} = \frac{1 + \text{odds}_{\delta \neq \infty}}{1 + \text{odds}_{\delta = \infty}} \approx \theta \left\{ 1 - e^{-\left\{ \frac{\theta}{\theta + \delta} \right\}} \right\} \tag{5}$$

The vulnerability change (see the Appendix for its derivation and the Fig. 2 for its dynamics) under the restrictions is

$$\text{VulnerabilityChange} = \frac{\left\{ \delta^2 - (\delta + \theta)^2 \right\}}{2 \left\{ \delta^2 + (\delta + \theta)^2 \right\}} \tag{6}$$

When the impact parameter,  $\delta \rightarrow \infty$ , the vulnerability (6) approaches 50% which portrays equally likely among the mutually complementary possibilities by a random chance, synonymous to flipping a coin. The needed reduction in the tail value at risk is

$$\hat{\mathcal{N}} \approx \left| 1 - \hat{\theta} \left( \sqrt{\frac{\bar{y}}{s_y^2} - 1} \right) \right| \quad (7)$$

### 3 Comparison of COVID-19 in US Four Time Zones

It is worth matching the IRRP model (1) and the data analytics of COVID-19 data as reported in <http://kff.org> for all US states. We grouped the data in terms of the time zones (see Tables 1, 2, 3, and 4 for the eastern, central, mountain, and pacific zones) on the assumption that the restrictions in each category vary in the zones.

The COVID-19 data exhibit overdispersion consistently in all four regional incidences and it is the data evidence that the prevalence rate of COVID-19 is not of a regular Poisson type but rather IRRP. Notice that the mean  $\bar{y}$  is less than the variance,  $s_y^2$  in all four categories of the four zones, which is indicative of the validity of the IRRP (1). The MLE of the prevalence rate,  $\theta$  and the impact,  $\delta$  are displayed in Tables 1, 2, 3, and 4. The prevalence rate is highest in the eastern zone and the lowest in the mountain zone for the COVID-19 infectivity. The prevalence rate is diverse and highest in the eastern states rather than in other zones (see Fig. 3 for details).

The impact is largest in the recovery category of the pacific states and smallest in the mortality category of the eastern states. The estimated impact,  $\hat{\delta}$  shows the widest

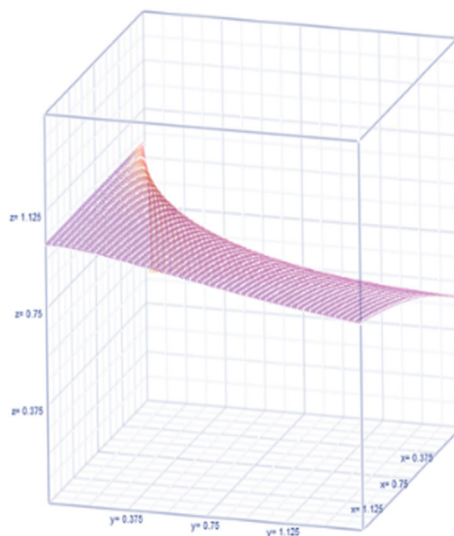


Fig. 2 Vulnerability change



**Table 1** COVID-19 report for eastern time zone

State	Cases	Died	Recovered	Hospitalized
Connecticut	27,700	2257		
Delaware	4734	152	1275	
District of Columbia	4323	224	660	
Florida	33,690	1268		5589
Georgia	26,237	1131		5186
Indiana	17,835	1007		
Kentucky	4708	240		
Maine	1095	53	631	170
Maryland	21,742	1047	1432	4559
Massachusetts	62,205	3562		5942
Michigan	41,379	3789		
New Hampshire	2054	66	980	259
New Jersey	118,652	7228		
New York	304,372	18,321		
North Carolina	10,509	378		
Ohio	18,027	975		3533
Pennsylvania	45,763	2292		
Rhode Island	8621	266		
South Carolina	6095	244		
Tennessee	10,735	199	5338	1045
Vermont	866	49		
Virginia	15,846	552		2322
Average	35,781	2059	1719	3178
Variance	4,314,004,644	16,129,699	3,245,602	5,296,115
$\hat{\theta}$	103.04	23.26	39.57	77.86
$\hat{\delta}$	0.29	0.26	0.93	1.95
Imbalance	0.000008	0.0001	0.0005	0.0006
<i>p</i> value	0.000001	0.000002	0.000001	0.000001
power	0.93	0.92	0.93	0.93
Odds ratio	65.02	14.60	24.67	48.50
Vulnerability Change	0.000008	0.00013	0.00055	0.00063
Reduction in tail value at risk	0.29	0.26	0.91	1.91

variance in the recovery category compared with other categories (see Fig. 4). For the regular Poisson distribution (that is,  $\delta = \infty$  in the IRRP (1), the impact factor is one. The impact factor in all zones of infectivity and of mortality is less than one. The eastern and central zones have less than one impact in the recovery category. The impact factor is less than one in the hospitalization category in the mountain zone and the pacific zones only. However, the least impact has occurred in the infectivity category of the eastern zone. These findings confirm what we heard in the media about COVID-19 in the USA.

**Table 2** COVID-19 report for central time zone

State	Cases	Died	Recovered	Hospitalized
Alabama	7019	269		978
Arkansas	3281	61	1339	414
Illinois	52,918	2355		
Iowa	7145	162	2697	
Kansas	4238	129		523
Louisiana	28,001	1862		
Minnesota	5136	343	2172	1044
Mississippi	6815	261		
Missouri	7582	329		
Nebraska	4281	70		
North Dakota	1067	19	458	85
Oklahoma	3618	222	2401	
South Dakota	2449	17	1573	173
Texas	28,087	782	13,353	
Wisconsin	6854	316		1512
Average	11,233	480	3428	676
Variance	201,822,753.5	481,367.2	19,716,349	269,977.6
$\hat{\theta}$	83.79	15.14	45.19	33.79
$\hat{\delta}$	0.62	0.49	0.60	1.77
Imbalance	0.00005	0.0009	0.0001	0.002
$p$ value	0.00005	0.00008	0.00006	0.00007
Power	0.65	0.63	0.78	0.86
Odds ratio	52.74	9.39	28.34	20.72
Vulnerability change	0.00005	0.00106	0.00017	0.00276
Reduction in tail value at risk	0.62	0.47	0.59	1.69

The  $p$  value (3) is the probability of the null hypothesis  $H_o : \delta = \infty$  to be true. The  $p$  values are smaller in all categories (infectivity, mortality, recovery, and hospitalization) in every zone (see Fig. 5 for details). Note that not only smaller  $p$  values identify the best match of the IRRP (1) but also that there is a finite and significant impact due to the imposed restrictions. The  $p$  values are smaller in all categories across the zones consistently. The statistical power (4) of accepting the true value of  $\delta^* = \hat{\delta}$  is displayed in all four categories for all zones. The power is at least 0.63 and it increases to as much as 0.93. See Fig. 6 for details on the statistical power to accept the true alternative hypothesis  $H_a : \delta^* = \hat{\delta}$  across the categories. The high power suggests that the methodology based on IRRP (1) works excellent to recognize the significant impact due to the restrictions when it occurs.

The odds ratio for having corona free safe situation attests to the fact that the imposed restrictions in each category have yielded desirable and significant impacts across all zones compared with what would have been without those restrictions. These findings offer great comfort and relief to those in governing agencies and the healthcare

**Table 3** COVID-19 report for mountain time zone

State	Cases	Died	Recovered	Hospitalized
Arizona	7962	330		1169
Colorado	15,284	777		2697
Idaho	1984	60	1121	175
Montana	453	16	392	61
New Mexico	3411	123		
Utah	4672	46	1939	390
Wyoming	559	7	373	56
Average	4904	194	956	758
Variance	27,765,910.3	78,323.14	550,499.6	1,076,514
$\hat{\theta}$	65.16	9.66	39.85	20.11
$\hat{\delta}$	0.87	0.50	1.73	0.54
Imbalance	0.00017	0.002	0.0017	0.0007
<i>p</i> value	0.004	0.005	0.004	0.004
Power	0.78	0.71	0.85	0.72
Odds ratio	40.87	5.92	24.56	12.51
Vulnerability change	0.00018	0.00273	0.00188	0.00074
Reduction in tail value at risk	0.86	0.48	1.66	0.53

management professionals who came up with the restrictions to begin with for the sake of good public health.

**Table 4** COVID-19 report for pacific time zone

States	Corona cases	Died	Recovered	Hospitalized
Alaska	355	9	252	36
California	48,917	1982		
Hawaii	618	16	526	70
Nevada	4998	243		
Oregon	2510	103		578
Washington	14,070	801		
Average	11,911.33	525.66	389	228
Variance	354,269,334.3	596,393.5	37,538	92,164
$\hat{\theta}$	69.06	15.60	39.59	11.34
$\hat{\delta}$	0.40	0.47	4.48	0.59
Imbalance	0.00003	0.0008	0.01	0.002
<i>p</i> value	0.0075	0.008	0.008	0.009
Power	0.68	0.70	0.90	0.73
Odds ratio	43.51	9.69	23.47	6.95
Vulnerability change	6.96672 E-09	3.41 E-06	5.33 E-06	1.74 E-05
Reduction in tail value at risk	912.31	93.34	138.18	53.50

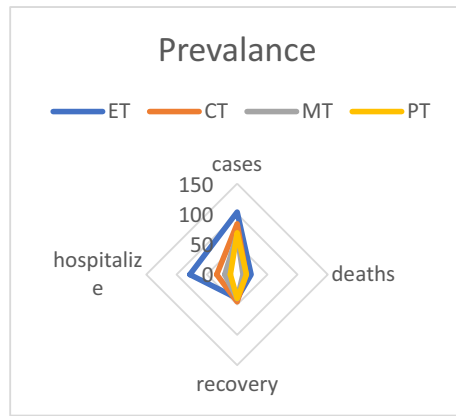


Fig. 3 Prevalance,  $\hat{\theta}$

The estimated imbalance factor across the categories (see Figs. 7, 8, and 9 for details) in every zone confirms that the IRRP (1) fits better than a simple Poisson in the COVID-19 data. The expected value is clearly tilted from the variance. Only when the estimated imbalance factor is unity, the regular Poisson distribution (in which  $\delta = \infty$ ) is appropriate. The regular Poisson is synonymous with no restriction.

The vulnerability changes to contract coronavirus, to be hospitalized, or to die are significantly lesser under the restrictions than what it would have been (that is, 50% in every category across the zones) without any imposed restriction. Only in the eastern zone, the vulnerability change to corona infectivity is remarkably low due to the imposed restrictions.

A comparative discussion of the results in these four time zones is worthwhile for COVID-19-related public health policymakers and the administrators of such policies. One would anticipate that the USA is one nation, ought to have the same health policies, at least with respect to COVID-19. Has it happened that way?

First, let us discuss the incidence rate ( $\hat{\theta}$ ) of COVID-19 and it is highest in the eastern zone, next high in the central zone, next high in the pacific zone, and the least in the mountain zone. Smaller the estimated value of the restriction rate ( $\hat{\delta}$ ) portrays better

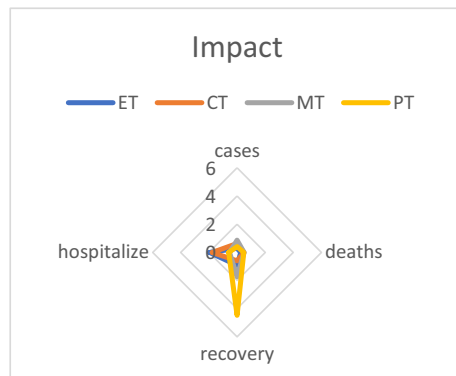


Fig. 4 Impact,  $\hat{\delta}$

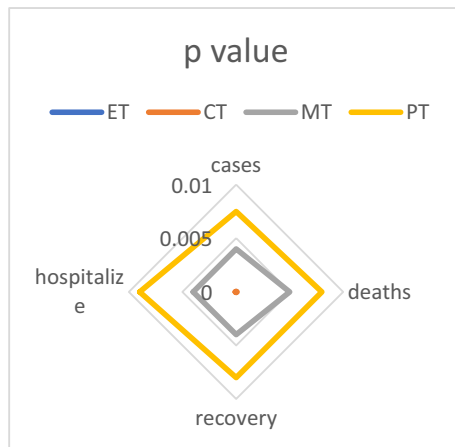


Fig. 5  $p$  value for  $H_0 : \delta = \infty$

effectiveness in the implementation of social distancing and/or face making. The most effective to the least effective is the eastern zone, pacific zone, central zone, and mountain zone in the incidence of COVID-19.

Next, let us discuss the death rate due to COVID-19. Among many other things, it reflects how vulnerable the people ought to have been in a zone with respect to dying due to COVID-19 once they contract it. The death rate ( $\hat{\theta}$ ) is highest in the eastern zone, next high in the pacific zone on par with the central zone, but least in the mountain zone. The effectiveness ( $\hat{\delta}$ ) of guarding from death is best in the eastern zone, next in the central zone on par with the pacific zone, but least in the mountain zone.

If we wish to discuss comparatively the recovery from COVID-19 among the zones, notice that there are missing data in recovery information across the zones in the USA. The results, based on the available data information, point out that the recovery rate ( $\hat{\theta}$ ) is equally best among the eastern, mountain, and pacific zones but worst in the central zone. The effectiveness ( $\hat{\delta}$ ) to recover from the COVID-19 is best in the eastern zone,

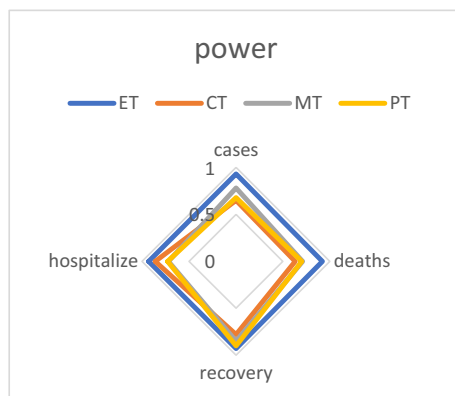
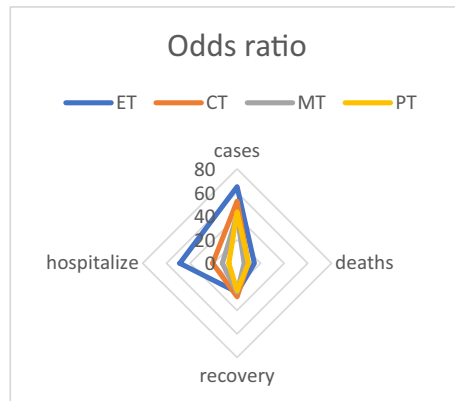


Fig. 6 Statistical power with  $H_a : \delta^* = \hat{\delta}$

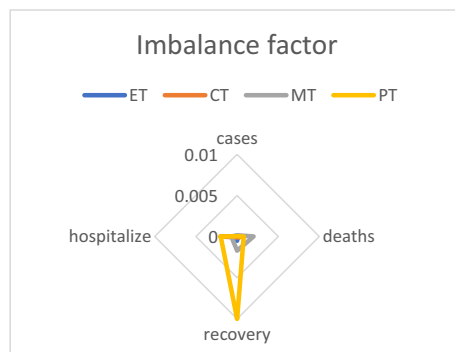


**Fig. 7** Odds ratio

the next best is in the central zone, bad in the mountain zone, and worst in the pacific zone.

Lastly, we will take up the reality of hospitalization to treat COVID-19 infections in the USA. In hospitalization, there is missing information for some states in all zones. Realize that any variation among the zones is reflective of disparities among hospital facilities and/or available medical professionals. The hospitalization rate ( $\hat{\theta}$ ) to treat the COVID-19 case is highest to lowest occur in the eastern zone, central zone, mountain zone, and pacific zone respectively. However, the effectiveness ( $\hat{\delta}$ ) of the administrative policies to hospitalize a COVID-19 case is best to worst occurred respectively in the mountain zone on par with the pacific zone, the eastern zone almost on par with the central zone.

The abovementioned similarities and differences could not have been noticed without our IRRP model to analyze and interpret the COVID-19 data across the zones in the USA. One would have thought that the USA as a nation comprising of homogeneous people and practices, at least with respect to the COVID-19 epidemics but it is not so. Our IRRP model and the results based on it indicate otherwise and that is the take-home lesson for the healthcare policymakers and the administrators of such policies. We all can learn from each other no matter in which zone we live and practice in the USA.



**Fig. 8** Imbalance factor

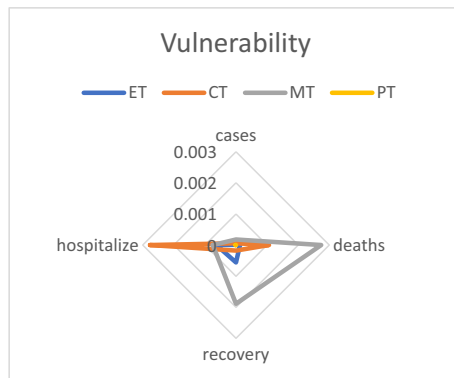


Fig. 9 Vulnerability change

## 4 Conclusion and Thoughts for Future Work

Many of the non-trivialities we noticed in the interpretation of the COVID-19 data would have remained unknown without charting out our approach. The IRRP (1) and the derived analytic expressions based on IRRP (1) are worthy enough to capture and comprehend the importance of the restrictions like social distancing to reduce (if not a total elimination) the COVID-19 infectivity. The lessons we learned from analyzing and interpreting the data evidence to deal with pandemics like COVID-19 will equip us to be ready for any future pandemic. The predictability in any category for any zone is feasible via building appropriate regression or structural equation models if the pertinent data on the covariates such as gender, age, race, socioeconomic status, and health insurance status are collected and made available.

The findings based on our approach impact healthcare efforts as they are noticed in the recovery category in the pacific states, in the hospitalization category in the mountain states, and in the infectivity category in the eastern states. If data have been collected on related variables, one could elaborate on the underlying reasons for such variations between the time zones and also help us to demonstrate the predictive power of the IRRP model.

The potential health policy implications of the work done in this article include the implications of the excellent and consistent data evidence on the significant reduction of the COVID-19's prevalence rate by an upper bound across all regions and it is a relief to the healthcare professionals such as physicians, nurses, lab assistants, and paramedical in ambulance services.

It is interesting to notice that the number of COVID-19 cases increased in Florida, Texas, and California while the number decreased substantially in New York, New Jersey, and Pennsylvania.

Does our model in harmony with this turning changes? Florida, Texas, and California were a member of the eastern, central, and pacific zone states. Our IRRP model pointed out that the estimated incidence rate of COVID-19 was highest in all these three zones. New York, New Jersey, and Pennsylvania were uniquely within the eastern zone which had the highest incidence rate. Recall that the effectiveness of controlling the spread of COVID-19 was the best only in the eastern zone. What must have happened in New York, New Jersey, and Pennsylvania was the spillover lag effect of most

effectiveness. Being a part of the least effective central and pacific zone (as we pointed out in early discussion), Texas and California continued to experience increased COVID-19 incidences. California and Florida relaxed the social distancing too fast and too liberal. Our data analysis did not involve any direct information pertinent to these changes and hence, our IRRP model could not address this issue. The quick reopening up the state from the lockdown and/or social distancing ought to have created super-spreaders (who have the potential to connect with a great number of people) via some social events (such as sport, music, religious gathering, recreation parks). Such information has not entered the current IRRP modeling process and hence, the model is limited in that sense to explain the resurgence or the second wave.

**Compliance with Ethical Standards**

**Conflict of Interest** The author declares that he has no conflict of interest.

**Ethical Approval** This article does not contain any study with human participants or animals performed by the author.

**Appendix**

The expected value of the IRRP distribution (1) is

$$E(y_i) = \sum_{y_i=0}^{\infty} y_i P(y_i) = \sum_{y_i=0}^{\infty} y_i \left(1 + \frac{y_i}{\theta_i + \delta_i}\right)^{y_i-1} \left(\theta_i e^{-\left\{\frac{\theta_i}{\theta_i + \delta_i}\right\}}\right)^{y_i} / y_i! e^{\theta_i}$$

$$= \left\{1 + \frac{\theta_i}{\delta_i}\right\} \theta_i \rightarrow \theta_i, \text{ when } \delta_i \rightarrow \infty.$$

The variance of the IRRP distribution (1) is  $E(y_i^2) - \{E(y_i)\}^2$  and it, after simplifications, becomes  $Var(y_i) = \left(1 + \frac{\theta_i}{\delta_i}\right)^3 \theta_i \rightarrow \theta_i, \text{ when } \delta_i \rightarrow \infty.$

Of interest in scrutinizing the data evidence, whether the MLE of the impact parameter statistically small involves dealing with the correlated sample mean and variance. Only in the case of normally distributed data, the sample mean and variance are independent [4]. Let  $\rho_{\bar{y}, s_y^2}$  be the correlation between the sample mean and sample variance. Utilizing the formulas in Blumenfeld [3], we obtain that

$$E\left(\frac{\bar{y}}{s_y^2} \mid \theta, \delta\right) = \frac{E(\bar{y} \mid \theta, \delta)}{E(s_y^2 \mid \theta, \delta)} \left\{1 + \frac{Var(s_y^2 \mid \theta, \delta)}{E(s_y^2 \mid \theta, \delta)}\right\} - \frac{\rho_{\bar{y}, s_y^2} \sqrt{Var(\bar{y} \mid \theta, \delta) Var(s_y^2 \mid \theta, \delta)}}{\{E(s_y^2 \mid \theta, \delta)\}^2}$$

and

$$Var\left(\frac{\bar{y}}{s_y^2} \mid \theta, \delta\right) = \left\{ \frac{E(\bar{y} \mid \theta, \delta)}{E(s_y^2 \mid \theta, \delta)} \right\}^2 \left\{ \frac{Var(\bar{y} \mid \theta, \delta)}{[E(\bar{y} \mid \theta, \delta)]^2} + \frac{Var(s_y^2 \mid \theta, \delta)}{[E(s_y^2 \mid \theta, \delta)]^2} - \frac{2\rho_{\bar{y}, s_y^2} \sqrt{Var(\bar{y} \mid \theta, \delta) Var(s_y^2 \mid \theta, \delta)}}{E(\bar{y} \mid \theta, \delta) E(s_y^2 \mid \theta, \delta)} \right\},$$



where  $E(\bar{y}|\theta, \delta) = E(y|\theta, \delta)$ ,  $E\left(s_y^2|\theta, \delta\right) = \left(1 + \frac{\theta}{\delta}\right)^3 \theta$  and  $Var(\bar{y}|\theta, \delta) = Var(y|\theta, \delta)/n$ ,  $Var\left(s_y^2|\theta, \delta\right) \approx \left(1 + \frac{\theta}{\delta}\right)^6 \theta/n$ , and

$\rho_{\bar{y}, s_y^2} \approx \sqrt{\frac{\{E(\bar{y}|\theta, \delta)\}^2}{\{E(\bar{y}|\theta, \delta)\}^2 + \{E(s_y^2|\theta, \delta)\}^2}} \approx \sqrt{\frac{\theta(1+\frac{\theta}{\delta})}{1+\theta(1+\frac{\theta}{\delta})}}$ . The survival function,  $S_{y_i}[m, \theta, \delta] = Pr[y \geq m|\theta, \delta]$  for the IRRP distribution (1) is

$$S_{y_i}[m, \theta, \delta] = e^\theta \sum_{y=m}^\infty \left(1 + \frac{y}{\theta + \delta}\right)^{y-1} \left(\theta e^{-\frac{\theta}{\theta+\delta}}\right)^y / y! \approx e^\theta \sum_{y=m}^\infty \left(1 + \frac{y(y-1)}{\theta + \delta}\right)^{y-1} \left(\theta e^{-\frac{\theta}{\theta+\delta}}\right)^y / y! \\ \approx e^\theta \left[ e^{-\left\{\frac{\theta}{\theta+\delta}\right\}-1} \right] \left\{ Pr\left(\chi_{2m}^2 < 2\theta e^{-\left\{\frac{\theta}{\theta+\delta}\right\}}\right) + \frac{\left\{\theta e^{-\left\{\frac{\theta}{\theta+\delta}\right\}}\right\}^2}{(\theta + \delta)} Pr\left(\chi_{2[m-2]}^2 < 2\theta e^{-\left\{\frac{\theta}{\theta+\delta}\right\}}\right) \right\}$$

due to the relationship (see [12]) between the chi-squared distribution function and the cumulative Poisson probabilities indicated. With  $q = 1$ , an interesting particular result

$$S_{y_i}[1, \theta, \delta] \approx e^\theta \left[ e^{-\left\{\frac{\theta}{\theta+\delta}\right\}-1} \right] Pr\left(\chi_{2df}^2 < 2\theta e^{-\left\{\frac{\theta}{\theta+\delta}\right\}}\right)$$

describes the probability of noticing one or more counts.

The constant adjusted odds of having no count are the ratio

$$1 + odds_{\delta \neq \infty} = Pr(y = 0|\theta, \delta) / Pr(y \geq 1|\theta, \delta) = \frac{e^{-\theta e^{-\left\{\frac{\theta}{\theta+\delta}\right\}}}}{Pr\left(\chi_{2df}^2 < 2\theta e^{-\left\{\frac{\theta}{\theta+\delta}\right\}}\right)}$$

In no restriction situation (that is,  $\delta = \infty$ ), the constant adjusted odds are  $1 + odds_{\delta = \infty} = \frac{e^{-\theta}}{Pr\left(\chi_{2df}^2 < 2\theta\right)}$ .

Hence, the *odds ratio* (OR) between under restrictions versus no restriction situation is  $OR = \frac{odds_{\delta \neq \infty}}{odds_{\delta = \infty}} \approx \theta \left\{ 1 - e^{-\left\{\frac{\theta}{\theta+\delta}\right\}} \right\}$ , after algebraic simplifications. See Shanmugam and Chattamvelli [11] for details about the application of the odds and odds ratio in data

analytics. The discrete *hazard rate*,  $h(1|\theta, \delta) = 1 - \frac{S_{y_i}[2, \theta, \delta]}{S_{y_i}[1, \theta, \delta]} = 1 - \frac{Pr\left(\chi_{4df}^2 < 2\theta e^{-\left\{\frac{\theta}{\theta+\delta}\right\}}\right)}{Pr\left(\chi_{2df}^2 < 2\theta e^{-\left\{\frac{\theta}{\theta+\delta}\right\}}\right)}$  in

under restrictions compared with such quantity,  $h(1|\theta, \delta = \infty) = 1 - \frac{Pr\left(\chi_{4df}^2 < 2\theta\right)}{Pr\left(\chi_{2df}^2 < 2\theta\right)}$  in no

restriction situation. The desirable reduction,  $R_i$  in the *hazard rate* due to restrictions is

$$\frac{Pr\left(\chi_{4df}^2 < 2\theta\right)}{Pr\left(\chi_{2df}^2 < 2\theta\right)} - \frac{Pr\left(\chi_{4df}^2 < 2\theta e^{-\left\{\frac{\theta}{\theta+\delta}\right\}}\right)}{Pr\left(\chi_{2df}^2 < 2\theta e^{-\left\{\frac{\theta}{\theta+\delta}\right\}}\right)}$$

A measure of *vulnerability change* under restrictions in a category is defined as  $Vulnerability_i = \frac{E[y|\theta, \delta]}{Var[y|\theta, \delta] + E[y|\theta, \delta]} - 0.5 = \frac{\delta^2 - (\delta + \theta)^2}{2\{\delta^2 + (\delta + \theta)\}}$ . The *tail value at risk* (TVaR<sub>i</sub>) is commonly used in the business world and it can be fitting here

to describe the risk at an outbreak of pandemic like COVID-19. It is

$$TVaR = E[y|y \geq 1, \theta, \delta] = 1 + \frac{\sum_{y=1}^\infty (y-1)Pr[y|\theta, \delta]}{Pr[y \geq 1|\theta, \delta]} = \frac{e^{-\theta} \left[ e^{-\left\{\frac{\theta}{\theta+\delta}\right\}-1} \right]}{Pr\left(\chi_{2df}^2 < 2\theta e^{-\left\{\frac{\theta}{\theta+\delta}\right\}}\right)}$$

whose ideal target amount is  $\left\{ \Pr\left(\chi_{2df}^2 < 2\theta e^{-\frac{\theta}{\theta+\delta}}\right) \right\}^{-1}$  in a perfect healthcare system. The value at risk,  $TVaR_{\beta_i,1}$  under restrictions is reduced by a percent  $e^{-\lambda_i} \left[ e^{-\left\{ \frac{\lambda_i}{\beta_i} \right\}} - 1 \right]$  from that of the value at risk,  $TVaR_{\beta_i=\infty,1}$  at the outbreak of COVID-19. The estimable reduction in the reduction of tail value at risk in a category is therefore  $\hat{\mathcal{M}} \approx \left| 1 - \hat{\theta} \left( \sqrt{\frac{y}{s_y^2}} - 1 \right) \right|$ .

## References

1. Ambrosio B, Aziz-Alaoui MA (2020) On a coupled time-dependent SIR models fitting with New York and New-Jersey states COVID-19 data. arXiv preprint arXiv 2006:05665
2. Backer JA, Klinkenberg D, Wallinga J (2020) Incubation period of 2019 novel COVID-19 (2019-nCoV) infections among travelers from Wuhan, China, 20–28 January 2020. *Euro Surveill* 25(5):1–10
3. Blumenfeld D (2010) Operations research calculations handbook. CRC Press, Boca Raton
4. Chattamvelli R, Shanmugam R (2019) Generating functions in engineering and the applied sciences, Synthesis Lectures on Engineering, Morgan & Claypool Publishers, 82 Winter sport Lane, Williston, VT 05495, USA
5. Cohen J (2020) Wuhan seafood market may not be source of novel virus spreading globally. *Science Mag American Association for the Advancement of Science. (AAAS)*. Archived from the original on 2020-01-27. Retrieved 2020-01-29
6. Kermack W, McKendrick A (1927) A contribution to the mathematical theory of epidemics. *Proc R Soc London A* 115:700–721
7. Naming the COVID-19 disease (COVID-19) and the virus that causes it, World Health Organization (WHO). 2020. Online, Available in the link <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>. Accessed 17 May 2020
8. Okabe Y, Shudo A (2020) A mathematical model of epidemics-a tutorial for students. *Mathematics* 8: 1174 1–16
9. Phan LT, Nguyen TV, Luong QC, Nguyen TV, Nguyen HT, Le HQ, Nguyen TT, Cao TM, Pham QD (2020) Importation and human-to-human transmission of a novel COVID-19 in Vietnam. *N Engl J Med* 382:872–874
10. Shanmugam R (1991) Incidence rate restricted Poissonness. *Sankhya (B)* 53(2):191–201
11. Shanmugam R, Chattamvelli R (2015) Statistics for scientists and engineers. John Wiley Inter-Science Publication, Hoboken
12. Stuart A, Ord K (2015) Kendall's advanced theory of statistics, vol 1. Oxford University Press, London
13. Tolles J, Luong TB (2020) Modeling epidemics with compartmental models. *Am Med Assoc* 323(24): 2515–2516
14. Wang C, Horby PW, Hayden FG, Gao GF (2020) A novel COVID-19 outbreak of global health concern. *Lancet*. 395(10223):470–473
15. Yang W, Zhang D, Peng L, Zhuge C, Liu L. (2020) Rational evaluation of various epidemic models based on the COVID-19 data of China, arXiv:2003.05666v1
16. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J et al (2020) A novel COVID-19 from patients with pneumonia in China, 2019. *N Engl J Med* 382(8):727–733

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.