Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Journal of the Formosan Medical Association 120 (2021) S69—S76

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.jfma-online.com

Original Article

Probabilistic forecasts of COVID-19 deaths with the progression rate from pneumonia to ARDS: An open-data-based global study

Wei-Jung Chang [a,1], Yee-Chun Chen [b,1], Chen-Yang Hsu [a,c], Chih-Dao Chen [d], Sam Li-Sheng Chen [e], King-Jen Chang [f,*]

[a] *Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan*
[b] *Department of Internal Medicine, National Taiwan University Hospital and College of Medicine, Taipei, Taiwan*
[c] *Dachung Hospital, Miaoli, Taiwan*
[d] *Department of Family Medicine, Far Eastern Memorial Hospital, Taipei, Taiwan*
[e] *School of Oral Hygiene, College of Oral Medicine, Taipei Medical University, Taipei, Taiwan*
[f] *Department of Surgery, National Taiwan University Hospital, Taipei, 100, Taiwan*

Received 30 March 2021; received in revised form 9 May 2021; accepted 10 May 2021

KEYWORDS
Case-fatality rate;
Disease progression;
Forecasting;
COVID-19;
ARDS

Background: Cumulative data of case-fatality rates (CFR) of COVID-19 varied across countries. A forecasting model generated based on detailed information from three countries during the initial phase of pandemic showed that progression rates from pneumonia to ARDS (PRPA) varied by country and were highly associated with CFR. We aim to elucidate the impact of the PRPA on COVID-19 deaths in different periods of pandemic.
Methods: We used the country-based, real-time global COVID-19 data through GitHub repository to estimate PRPA on the first period (January to June), second period (July to September), and third period (October to December) in 2020. PRPA was used for predicting COVID-19 deaths and assessing the reduction in deaths in subsequent two periods.
Results: The estimated PRPA varied widely from 0.38% to 51.36%, with an average of 15.99% in the first period. The PRPA declined to 8.44% and 6.35% in the second and third period. The CFR declined stepwise and was 4.94%, 2.61%, and 1.96%, respectively. Some countries exhibited a decrease in the PRPA from the second to the third period whereas others showed the opposite, particularly where selected viral mutants were prevalent. Overall, the number of observed deaths was lower than that of the predicted deaths in the second and third periods, suggesting an improvement in management of COVID-19 patients. Besides, the degree of improvement depends on the extent of change in PRPA.

* Corresponding author. No.7, Chung Shan S. Rd., Zhongzheng Dist., Taipei, 100, Taiwan.
E-mail address: kingjen@ntu.edu.tw (K.-J. Chang).
[1] These authors have contributed equally.

https://doi.org/10.1016/j.jfma.2021.05.016

Conclusion: PRPA is a useful indicator to facilitate decision making and assess the improvement of clinical management and medical capacity by forecasting deaths.

Introduction

Elucidating the disease spectrum of coronavirus disease 2019 (COVID-19) facilitates policy making for containing the spread of COVID-19 infection and optimal allocation of resource including manpower in the context of COVID-19 pandemic. The spectrum of COVID-19 vary from asymptomatic, mild respiratory disease (MRD), pneumonia, acute respiratory distress syndrome (ARDS) or multi-organ failure resulting in death or recovery.[1,2] As of 30 April 2021, confirmed cases had reached over 150 million globally, with more than 3 million deaths attributed to CIVID-19.[3] In addition to limited availability and accessibility of effective therapy and vaccination, some factors are associated with the high case fatality rate and the progression of pneumonia to ARDS is one of the major concerns.[4]

Wu et al. evaluated the risk factors for clinical outcomes of COVID-19 pneumonia patients and found older age, neutrophilia, organ and coagulation dysfunction were independently associated with the development of ARDS and progression from ARDS to death.[5] High fever and receiving methylprednisolone treatment for COVID-19 ARDS were significantly associated with lower likelihood of death.

However, the quantitative disease progression of COVID-19 on population level has been rarely addressed. We have developed a quantitative model for evaluating the disease progression of COVID-19 based on detailed information reported from China (as of January 29, 2020), Japan (as of March 13, 2020), and Italy (as of March 6, 2020) during the initial period of COVID-19 pandemic.[4] Notably, the progression rate from MRD to pneumonia and progression rate from ARDS to death were very close across three countries. However, the progression rates from pneumonia to ARDS (PRPA) varied substantially. These estimates were highly associated with case fatality rates reported for each country. It can be inferred from these findings that compared with two other progression rates, the PRPA plays a key role in accounting for the difference in the case fatality rates across countries. External validation of the model confirmed that the numbers of expected deaths were consistent with the observed data.

The pandemic exposed the weaknesses of the health system or insufficient preparedness of emerging infection diseases in several countries.[6] It is anticipated that COVID-19 patients have better outcome during the subsequent periods of pandemic due to increase in the awareness and experience of clinicians, availability of testing and therapeutic options, and surge capacity of the healthcare system. Alternatively, patients are less likely to have sufficient and appropriate care if rapid surge in COVID-19 cases outnumbers the capacity of healthcare settings which might further compromised due to hospital outbreak involving healthcare personnel.

Forecasting models provide specific, quantitative, and evaluable predictions that inform short-term decisions such as healthcare staffing needs, school closures, and allocation of medical supplies. A real-time global data through GitHub repository[7] which are shared publicly provide a grand opportunity to quantify the disease burdens. We aim to apply the disease progression model using worldwide data as of December 2020 and compare observed death and predicted death based on estimated PRPA by period and country or region.

Materials and methods

Date source

Information on the reported number of confirmed COVID-19 cases and deaths were collected from a GitHub repository, an online interactive dashboard, created and hosted by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University, Baltimore, MD, USA.[7,8] The dashboard is updated in real time and reports cases at the province level in China; at the city level in the USA, Australia, and Canada; and at the country level otherwise. This center has collected confirmed COVID-19 cases, cased with recovery, and deaths. Data of case numbers have been validated with regional and local health departments, including the respective centers for disease control and prevention (CDC) of Taiwan, China, and Europe, the Hong Kong Department of Health, the Macau Government, and WHO, as well as city-level and state-level health authorities.

Data used in this analysis was those updated through December 31 2020 and were divided into three phases, the first period (January 1 — June 30), the second period (July 1 — September 30), and the third period (October 31 — December 31) in parallel with the evolution of new virus variants.[9–11]

Selected country or region and cases for analysis

For the reason of better comparison of PRPA, those countries or regions with the small number of COVID-19 deaths in a period (under 100 cases) were excluded from our analysis because these countries had less chance to face the problem of insufficient medical care. The country or region with high infection cases in each continent was purposive selecting to present the individual disease spectrum. Due to the increasing capacity of COVID-19 testing, more and more asymptomatic cases were included in reported confirmed cases in the second and third periods and it is less likely to become fatal COVID-19, the pooled proportion of COVID-19 cases for asymptomatic cases was estimated as 17%.[12] Therefore, we excluded 17% of

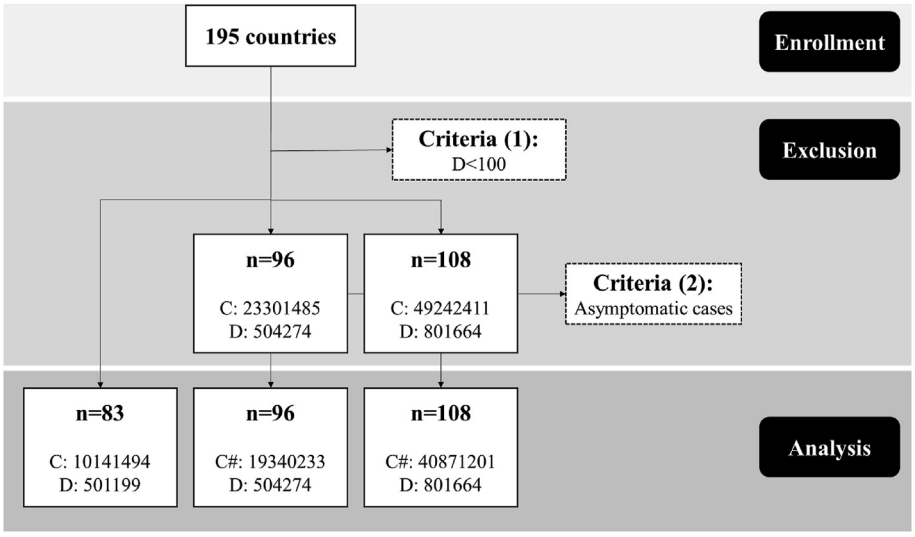

Fig. 1 Flowchart of data enrolled and excluded from a GitHub repository. C: COVID-19 cases; D: COVID-19 deaths. # Exclude estimated asymptomatic SARS CoV-2 infections.

confirmed cases in second and third periods in data analysis. Fig. 1 demonstrated how data were enrolled and excluded and number of countries evaluated by period.

Statistical analysis

Model specification

The disease progression for each patient is assumed to follow the evolution from MRD, pneumonia, ARDS, to death or recovery. The disease progression quantitative model was described previously.[4] We focus on three progression rates: MRD to pneumonia, pneumonia to ARDS, and ARDS to death. As most countries only have information of MRD and death, we assumed that the first (MRD to pneumonia) and third (ARDS to death) progression rates are constant across countries based on previous study.[4] We estimated the progression rate from pneumonia to ARDS (PRPA) by the following formula.

$$\frac{\text{Death}/\text{P(ARDS to death)}}{\text{MRD} \times \text{P(MRD to pneumonia)}}$$

P (MRD to pneumonia) is the pooled estimate of the rate from MRD to pneumonia (73%) and P (ARDS to death) is the pooled estimate of the rate from ARDS to death (43%).[4]

The comparison of observed and expected period-specific case-fatality

A simple linear regression model has been constructed and trained in the first period to establish the relationship between PRPA and case-fatality rate.[4] The PRPA can be treated as an independent variable to project the case-fatality rate.[4] The relationship between these two variables was formed as follows.

Case-fatality rate (%) = 1.4076 + 0.2732 × (progression rate from pneumonia to ARDS, %).

Based on this equation, the numbers of expected death following the progression rate obtained in the first period were estimated to compare with the numbers of observed

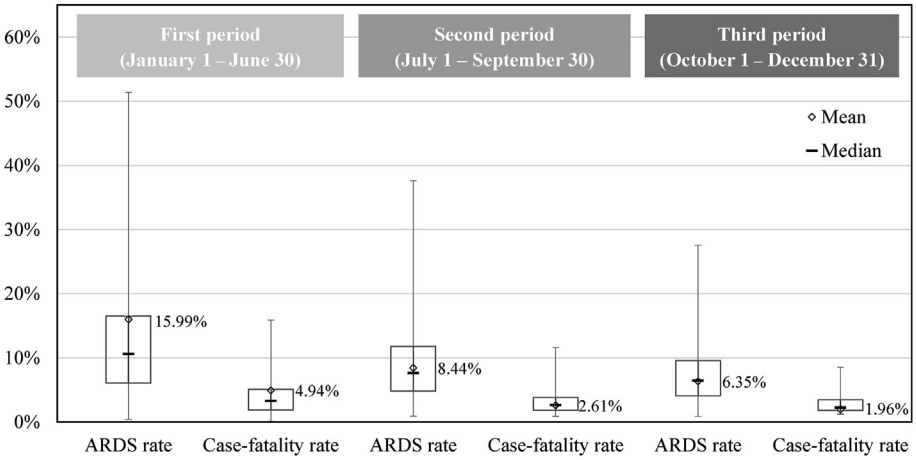

Fig. 2 Pooled progression rates from pneumonia to ARDS and pooled case-fatality rates by periods.

Table 1 Number of new confirmed cases, COVID-19 deaths, pneumonia, ARDS by three study periods.

Country	January 01 to June 30					July 01 to September 30[#]					October 01 to December 31[#]				
	COVID-19 cases	COVID-19 death	Pneumonia*	ARDS*	ADRS rate	COVID-19 cases[#]	COVID-19 death	Pneumonia*	ARDS*	ADRS rate	COVID-19 cases[#]	COVID-19 death	Pneumonia*	ARDS*	ADRS rate
Argentina	64,530	1307	46,913	3075	6.56%	569,771	15,630	414,223	36,776	**8.88%**	725,846	26,308	527,690	61,901	**11.73%**
Brazil	1,402,041	59,594	1,019,284	140,221	**13.76%**	2,829,382	84,358	2,056,961	198,489	**9.65%**	2,377,982	50,997	1,728,793	119,993	**6.94%**
Egypt	68,311	2953	49,662	6948	**13.99%**	28,956	2977	21,051	7005	**33.27%**	28,937	1701	21,037	4002	**19.03%**
France	204,244	29,846	148,485	70,226	**47.29%**	333,369	2132	242,359	5016	2.07%	1,719,572	32,781	1,250,129	77,132	6.17%
Germany	195,418	8990	142,069	21,153	**14.89%**	80,921	505	58,829	1188	2.02%	1,218,114	24,296	885,569	57,167	6.46%
India	585,481	17,400	425,645	40,941	9.62%	4,753,495	81,278	3,455,791	191,242	5.53%	3,281,895	50,060	2,385,937	117,788	4.94%
Iran	227,662	10,817	165,510	25,452	**15.38%**	190,532	15,352	138,517	36,122	**26.08%**	637,376	29,054	463,372	68,362	**14.75%**
Italy	240,578	34,767	175,381	81,613	**46.53%**	61,655	1127	44,946	2646	5.89%	1,487,613	38,265	1,084,470	89,824	**8.28%**
Japan	18,615	972	13,217	2266	**17.14%**	53,930	603	38,290	1406	3.67%	126,343	1717	89,703	4002	4.46%
Korea, South	12,850	282	9342	664	7.10%	9162	133	6661	313	4.70%	31,440	502	22,857	1181	5.17%
Mexico	226,089	27,769	164,367	65,339	**39.75%**	429,215	49,877	312,040	117,358	**37.61%**	566,789	48,161	412,055	113,320	**27.50%**
South Africa	151,209	2657	109,929	6252	5.69%	434,198	14,077	315,662	33,122	**10.49%**	317,742	11,735	230,999	27,612	**11.95%**
Spain	249,271	28,355	181,220	66,718	**36.82%**	431,531	3436	313,723	8085	2.58%	962,034	19,046	699,399	44,814	6.41%
United Kingdom	285,216	40,479	207,352	95,245	**45.93%**	141,623	1754	102,960	4127	4.01%	1,693,520	31,389	1,231,189	73,856	6.00%
United States	2,631,894	127,831	1,913,387	300,779	**15.72%**	3,810,519	79,396	2,770,247	186,814	6.74%	10,578,521	138,510	7,690,585	325,906	4.24%

Abbreviation: ARDS, Acute Respiratory Distress Syndrome. * Estimated numbers. [#] Exclude cases of asymptomatic COVID-19 infections.
The country-specific progression rates from pneumonia to ARDS in bold indicate those higher than the median global progression to ARDS in indicated period.

COVID-19 death in the second and third periods. All analyses were performed with SAS version 9.4 (SAS Institute Inc., Cary, NC, USA).

Results

Progression rate from pneumonia to ARDS by period and country

Fig. 2 shows the PRPA and the case-fatality rate by three periods. The average PRPA was 15.99% in the first period, and decreased to 8.44% and 6.35% in the second and the third periods, respectively. The case-fatality was 4.94%, 2.61%, and 1.96% in the first, second, and third periods, respectively. The similar trends were observed between PRPA and case-fatality rate. The lowest case-fatality with the narrowest confident interval was observed in the third period obtained from 108 countries.

The detailed numbers of the confirmed cases, COVID-19 deaths, estimated pneumonia, estimated PRPA by periods and country are provided in Supplementary Tables 1—3. The progression rate varied widely in the first period among 83 countries/regions, only 0.38% in Qatar and 4.25% in Australia, and up to 47.29% in France and 51.36% in Belgium (Supplementary Table 1). Data from selected countries which were representative of each continent are shown in Table 1. Countries in Europe, such as UK, Italy, Spain, and France had the very high PRPA in the first period but became lower and less than the median of global data in the second and third periods. The declining trends of PRPA were noted in most of countries. Nevertheless, countries including Argentina, Brazil, and Mexico in America, Egypt and South Africa in Africa had worse PRPA in the second period (Fig. 3, upper left panel, first group). For the countries in Africa and South America, the PRPA were higher than the median global progression rates in the second and third periods, especially for Mexico and Egypt. To compare with the change pattern of PRPA in the second and third period, we found that the PRPA were reduced from the second to the third period in India, USA, Brazil, Iran, Egypt, and Mexico, whereas the increasing trend from the second to the third period was noted in countries such as Germany (from 2.02% to 6.46%), France (from 2.07% to 6.17%), UK (from 4.01% to 6.00%), Argentina (from 8.88% to 11.73%), and South Africa (from 10.49% to 11.95%) (Fig. 3, lower panel).

Comparison of expected and observed Covid-19 death by period and country

The parameters of linear regression model trained in the first period were used to calculate the expected numbers of COVID-19 deaths in the second and third period. The country-specific observed and expected deaths and the corresponding percent change are shown in Fig. 4. The numbers of expected deaths were overall higher than those of observed ones in the second and third periods. For countries with a decreasing trend of PRPA, the gap between observed and expected death became larger in the third period than those countries with an increasing trend. For example, the percent changes are 136% and 88% in the third period and second period, respectively in United States (Fig. 4, right panel, second group). For countries such as Italy, Spain, UK, Germany, and France in Europe, the percent change between observed and expected deaths in the second period were higher than third period.

Discussion

The high association between progression from pneumonia to ARDS and case-fatality was demonstrated by using the global COVID-19 pandemic open data during the first full calendar year. The progression from pneumonia to ARDS as a key factor responsible for the variation of case-fatality rate of COVID-19 was identified based on the previous study.[4]

The current study modeled the PRPA using real world big data and showed wide variation among 83 countries evaluated in the first period. Similarly, the PRPA varied across published data. For example, Huang et al. reported that 12

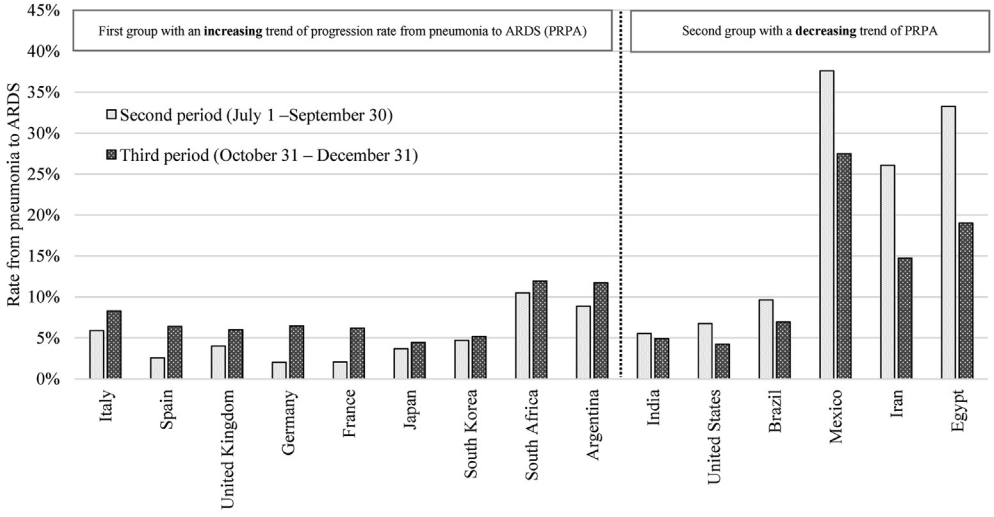

Fig. 3 Group classification by the trend of estimated progression rates from pneumonia to ARDS.

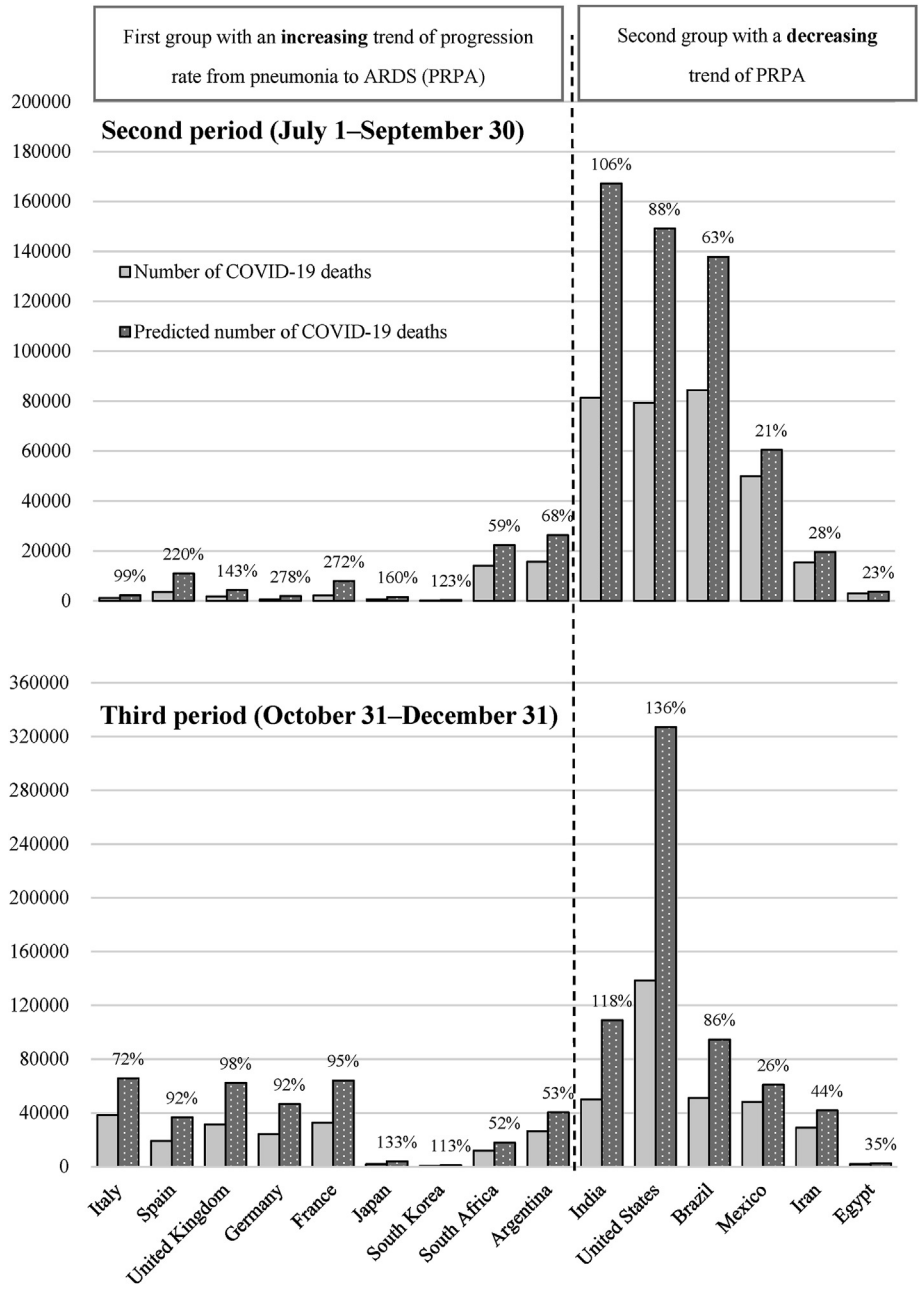

Fig. 4 Predicted and observed COVID-19 deaths and the percent change by period and group. Note: The percentage on the bar chart indicates the percent change by taking the predicted deaths and subtracting the observed deaths then divided by the observed deaths. ARDS, acute respiratory distress syndrome.

of the first 41 patients (29.3%) with COVID-19 pneumonia developed ARDS.[13] Wu et al. reported that 84 of 201 patients (41.8%) with pneumonia developed ARDS in Wuhan.[5] These two were single-center study. Based on nationwide data as of January 29, 2020 in China, ARDS cases were present in 37 of 800 patients (4.6%) with non-severe pneumonia on admission.[14] Our previous study showed that the PRPA estimated according to published data was 5% in 1099 patients from 552 hospitals in China.[4] Single-center studies[5,13] are vulnerable to patient selection bias as the development of ARDS might be affected by the underlying characteristics of COVID-19 patients such as older age,[5] immune response,[14,15] pharmaceutical interventions, and medical capacity.[10]

The global PRPA decreased over time although this model has adjusted for the potential confounding generated due to the increasing capacity of COVID-19 testing and more asymptomatic SARS CoV-2 infections were included in reported cases in the second and third periods. Nevertheless, the escalating PRPA overtime was observed in some countries such as Argentina and South Africa. The higher PRPA in third period than the second period was found in Italy, Spain, UK, Germany, France, Japan, South Korea, Argentina and South Africa. These trends were

probably in part due to the emerging of selected viral variants as more frequent infected cases with emerging variants have been detected in these countries in the third pandemic period.[9—11] A variant of SARS CoV-2 with a D614G substitution in the gene encoding the spike protein became the dominant form of the virus circulating globally by June 2020 which does not cause more severe illness or alter the effectiveness of existing diagnostics, pharmaceutical or non-pharmaceutical interventions.[11] On the other hand, B.1.1.7 variant (also called VOC-202012/1) first detected in Britain in September 2020 was associated with higher case fatality rate.[9,10] B.1.1.7 variant afterwards rapidly spread out over the world. Until December 28 2020, the proportion of B.1.1.7 among clades were 20% in Europe, 17% in Asia, 12% in South America, 10% in Africa, and 9% in North America (data obtained through GISAID).[16] Furthermore, the peak of second wave of COVID-19 in South Africa was January 2021 and was driven by the recently discovered 501Y.V2 variant (also called B.1.351).[11] Note that B.1.351 has accounted for 34% of confirmed cases with genome sequencing in Africa until the end of 2020.[16]

With the parameters estimated by a built linear regression model with the trained data in the first period, we predicted the number of COVID-19 deaths in the second and the third period and compared these predicted numbers with the observed number to assess the extent of the improvement in intervention, such as resource allocation and medical care. As anticipated, the overall number of observed deaths was lower than that of the predicted deaths in the second and third periods. Nevertheless, the observed death by period and gap between expected and observed death varied from country to country. As mentioned above, the gap changes between expected and observed cases among countries could be distinguished as two groups (Fig. 4). The smaller gap between expected and observed deaths in the first group indicated that countries with more emerging viral variants are more likely to have the higher transmission and severe cases, offset the impact of interventions, resulting in the better fitting of prediction model developed in the first period.

The study has several limitations. First, individual—level associations could not be incorporated based on public reported country-level data as demonstrated in a recent study.[17] Second, our analyses did not include potential unmeasurable confounding factors such as COVID-19 variants and the non-pharmaceutical interventions as data are available later on in selected countries or regions.[18]

In conclusion, assessing the progression rates from pneumonia to ARDS could be suggested as a surrogate indicator to inform short-term decision making about surging capacity of intensive care and others. Besides, this model could help to quantify the impact of intervention by forecasting the expected deaths and discrepancy between observed deaths.

Declaration of competing interest

The authors have no conflicts of interest relevant to this article.

Acknowledgements

This study was supported by Ministry of Science and Technology, Taiwan (MOST 109-2327-B-002-009). YC Chen was supported by intramural fund of National Taiwan University Hospital, Taiwan (109-P14).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jfma.2021.05.016.

References

1. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early transmission dynamics in wuhan, China, of novel coronavirus—infected pneumonia. *N Engl J Med* 2020;**382**:1199—207.
2. Centers for Disease Control and Prevention. *Symptoms of coronavirus (2019-nCoV)*. 2019. Available at: https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html. [Accessed 22 February 2021].
3. World Health Organization. *COVID-19 situation reports*. 2019. Available at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports.
4. Hsu CY, Lai CC, Yeh YP, Chang CC, Chen HH. Progression from pneumonia to ARDS as a predictor for fatal COVID-19. *J Infect Public Health* 2020;**4**:504—7.
5. Wu C, Chen X, Cai Y, Xia J, Zhou X, Xu S, et al. Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in wuhan, China. *JAMA Intern Med* 2020;**180**:934—43.
6. Armocida B, Formenti B, Ussai S, Palestra F, Missoni E. The Italian health system and the COVID-19 challenge. *Lancet Public Health* 2020;**5**:e253.
7. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* 2020;**20**:533—4.
8. *COVID-19 dashboard by the center for systems science and engineering*. (CSSE) at Johns Hopkins University; 2020. https://arcg.is/0fHmTX.
9. Davies NG, Abbott S, Barnard RC, Jarvis CI, Kucharski AJ, Munday JD, et al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* 2021;**372**:eabg3055.
10. Challen R, Brooks-Pollock E, Read JM, Dyson L, Tsaneva-Atanasova K, Danon L. Risk of mortality in patients infected with SARS-CoV-2 variant of concern 202012/1: matched cohort study. *BMJ* 2021;**372**:n579.
11. World Health Organization. *SARS-CoV-2 variants*. 2020. https://www.who.int/csr/don/31-december-2020-sars-cov2-variants/en/. [Accessed 2 April 2021].
12. Byambasuren O, Cardona M, Bell K, Clark J, McLaws ML, Glasziou P. Estimating the extent of asymptomatic COVID-19 and its potential for community transmission: systematic review and meta-analysis. *J Assoc Med Microbiol Infect Dis Canada* 2020;**5**:223—34.
13. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020;**395**:497—506.
14. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med* 2020;**382**:1708—20.

15. Moore JB, June CH. Cytokine release syndrome in severe COVID-19. *Science* 2020;**368**:473—4.
16. *Global initiative on sharing all influenza data (GISAID)*. 2020. Available at: https://www.gisaid.org/phylodynamics/global/nextstrain/. [Accessed 30 April 2021].
17. Schwab P, Mehrjou A, Parbhoo S, Celi LA, Hetzel J, Hofer M, et al. Real-time prediction of COVID-19 related mortality using electronic health records. *Nat Commun* 2021;**12**:1058.
18. Centers for Disease Control and Prevention. *COVID-19 Forecasts: deaths*. Available at. 2019. Updated April 28, 2021, https://www.cdc.gov/coronavirus/2019-ncov/covid-data/forecasting-us.html. [Accessed 4 May 2021].