

# Journal of Applied Statistics



ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/cjas20

# A continuous age-specific standardized mortality ratio for estimating the unascertained rates in the early epidemic of COVID-19 in different regions

Peipei Du, Peihua Cao, Xiaodong Yan, Daihai He, Xiaotong Zhang, Weixiang Chen, Jiawei Luo, Ziqian Zeng, Yaolong Chen, Lin Yang, Shu Yang & Xixi Feng

**To cite this article:** Peipei Du, Peihua Cao, Xiaodong Yan, Daihai He, Xiaotong Zhang, Weixiang Chen, Jiawei Luo, Ziqian Zeng, Yaolong Chen, Lin Yang, Shu Yang & Xixi Feng (2023) A continuous age-specific standardized mortality ratio for estimating the unascertained rates in the early epidemic of COVID-19 in different regions, Journal of Applied Statistics, 50:11-12, 2504-2517, DOI: <u>10.1080/02664763.2021.1947995</u>

To link to this article: <u>https://doi.org/10.1080/02664763.2021.1947995</u>

9	© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group	+	View supplementary material 🗷
	Published online: 08 Jul 2021.		Submit your article to this journal $arsigma$
111	Article views: 1829	Q	View related articles 🗹
CrossMark	View Crossmark data 🗹	ආ	Citing articles: 1 View citing articles 🗹

#### APPLICATION NOTE



OPEN ACCESS

Check for updates

# A continuous age-specific standardized mortality ratio for estimating the unascertained rates in the early epidemic of COVID-19 in different regions

Peipei Du <sup>(ba,b\*</sup>, Peihua Cao <sup>(bc\*</sup>, Xiaodong Yan<sup>d\*</sup>, Daihai He <sup>(be</sup>, Xiaotong Zhang<sup>b</sup>, Weixiang Chen<sup>b</sup>, Jiawei Luo<sup>f</sup>, Ziqian Zeng<sup>a</sup>, Yaolong Chen<sup>g</sup>, Lin Yang <sup>(bh</sup>, Shu Yang <sup>(bb</sup>) and Xixi Feng<sup>a</sup>

<sup>a</sup>School of Public Health, Chengdu Medical College, Chengdu, People's Republic of China; <sup>b</sup>College of Medical Information Engineering, Chengdu University of Traditional Chinese Medicine, Chengdu, People's Republic of China; <sup>c</sup>Clinical Research Center, Zhujiang Hospital, Southern Medical University, Guangzhou, People's Republic of China; <sup>d</sup>Zhongtai Securities Institute for Financial Studies, Shandong University, Jinan, People's Republic of China; <sup>e</sup>Department of Applied Mathematics, Hong Kong Polytechnic University, Hong Kong, People's Republic of China; <sup>f</sup>West China Biomedical Big Data Center, West China Clinical Medical College of Sichuan University, Chengdu, People's Republic of China; <sup>g</sup>Evidence-Based Medicine Center, School of Basic Medical Sciences, Lanzhou University, Lanzhou, People's Republic of China; <sup>h</sup>School of Nursing, Hong Kong Polytechnic University, Hong Kong, People's Republic of China

#### ABSTRACT

The difference in age structure and aging population level was an important factor that caused the difference in COVID-19's case fatality rate (CFR) in various regions. To eliminate the age effect on estimating the CFR of COVID-19, our study applied nonlinear logistic model and maximum likelihood method to fit the age-fatality curves of COVID-19 in different countries and regions. We further computed the standardized mortality ratio from the age-fatality curves of COVID-19 in the above regions and found that the risk of COVID-19 death in Wuhan was of a moderate level, while the non-Hubei region was even lower, compared with other regions. Regarding the disparity of CFRs among different regions in the country, we believed that there might be an unascertained phenomenon in highendemic regions. Based on age-fatality rate curves, we estimated unascertained rates in cities with severe epidemics such as Wuhan and New York, and it was found that the total unascertained rates in Wuhan and New York were 81.6% and 81.2%, respectively. Meanwhile, we also found that the unascertained rates varied greatly with age.

#### ARTICLE HISTORY

Received 16 August 2020 Accepted 21 June 2021

#### **KEYWORDS**

Standardized mortality ratio; COVID-19; case fatality rate; nonlinear logistic model; age-specific

**CONTACT** Shu Yang Sishiyu1978@qq.com College of Medical Information Engineering, Chengdu University of Traditional Chinese Medicine, Chengdu, People's Republic of China; Xixi Feng Fengxixi@163.com School of Public Health, Chengdu Medical College, Chengdu, People's Republic of China \*These authors contributed equally to this work.

Supplemental data for this article can be accessed here. https://doi.org/10.1080/02664763.2021.1947995

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

### Introduction

Since the case fatality rate (CFR, number of deaths over number of diagnosed cases) is associated with age and sociological characteristics in the public health and health economics studies, one classical way in comparing different CFRs is to calculate the age-standardized CFR [1,2]. There are two forms of calculating standardized rate: the first was a direct method, in which weighted sums of rates are calculated for different regions based on a standardized age structure, the results of the calculation can be compared and sorted directly. If the standardized age structure is unknown or cannot be used, the indirect method should be present. Assuming that we want to compare the occurrence rates of region A and region B, the actual population of each age group in region A should be multiplied by the occurrence rate of each age group in region B and sum them up, this means the total number of expected occurrences in region A under the circumstance that the occurrence rate of each age group in region A is the same as in region B. After that, we take the total number of actual occurrences in region A over the total number of expected occurrences in region A that we just calculated to generate the standardized mortality ratio (SMR). The indirect standardized rate loses its practical significance because it does not use a standardized population composition, but the SMR makes tremendous sense in public health and is often used as an indicator for evaluating disease burden and disease lethal intensity. If the SMR is greater than 1, it can be considered that the occurrence intensity of region A is higher than that of region B, otherwise, the result is the opposite.

In practical applications, age variables are often described discretely in groups. This study used age as a continuous variable to calculate SMR and applied it to the comparison of COVID-19 fatality rates among different regions. We extended this method to estimate the unascertained rate and extent that possibly existed in high-endemic regions. Meanwhile, we analyzed the regularity of the unascertained phenomenon affected by age, and this study started with establishing a nonlinear logistic regression curve of age-related fatality.

The ongoing pandemic of the COVID-19 had caused over 82 million infections and more than 1.8 million deaths around the world by the end of 2020 [3]. However, the CFRs vary across different countries in mid-June 2020 [4,5], such as China (5.57%), South Korea (2.31%), Italy (13.92%), Spain (8.27%), France (18.90%), Canada (8.11%), Germany (4.72%), and the United States (5.39%), and the differences in CFRs might be heavily affected by the age structure of the population in each country [6,7]. Even in China, the CFR of COVID-19 in Wuhan (7.69%) was still much higher than that in the non-Hubei region (0.81%). Non-Hubei refers to all regions in China except Hubei. Unlike Wuhan, where COVID-19 emerged since early December 2019, non-Hubei region of China had enough medical resources that could provide quick response and timely medical test and treatment to the public [8]. The data from non-Hubei region could therefore be more representative of the real CFR of COVID-19 and could not be affected by the unascertained rate, which allowed us to estimate the unascertained rate in Wuhan.

As there was no uniform standard composition that could be used to calculate the direct standardized rate, the direct comparisons of the CFRs could easily lead to misjudgment due to variation of age or sociological characteristics, we used the indirect method to compare the severity of CFRs in different regions.

# **Data sources**

We retrieved the age information of confirmed cases and deaths of COVID-19 infection from 11 January to 15 June 2020 in eight countries (*China, Italy, Spain, Canada, Germany, France, the United States, and South Korea*), two cities (*Wuhan and New York*), and one region (*non-Hubei region in China*), from the official websites from each country (Supplementary Table 1). Since the officially reported data from different country and region were not divided into age groups with the same cut-off, we restructured the number of deaths and confirmed cases at 10-year age intervals, with the boundaries beginning at ten and ending at nine. As there are 20 deaths of COVID-19 without reporting age information in non-Hubei region, we assigned them into ten age groups by weighting the number of age-specific deaths in this region.

# **Methods**

We define our target dataset at age *i* as:

$$\{(Z_i, Y_i, X_i) : i = 1, \cdots, 100\},\tag{1}$$

where  $X_i$  denotes the *i*th age,  $Z_i$  denotes the number of confirmed cases and  $Y_i$  denotes the number of deaths of COVID-19 at the age of  $X_i$ .

To assess the age (X) effect on the CFR  $\pi$ (X), a nonlinear logistic regression model is used to define the function:

$$\pi(x) = \frac{e^{\beta + s(x)}}{1 + e^{\beta + s(x)}},$$
(2)

The penalized basis smoother is adopted to estimate the nonlinear effect of the 'age' by penalizing the maximum binomial distribution-based log-likelihood function and to obtain the intercept  $\hat{\beta}$  and nonlinear effect estimators  $\widehat{S(\mathbf{x})}$ , that is,

$$\begin{pmatrix} \stackrel{\wedge}{\beta}, \stackrel{\wedge}{s}(\cdot) \end{pmatrix} = \arg \max_{(\beta, s(\cdot))} \left\{ \sum_{i=1}^{100} \left[ Y_i \log(e^{\beta + s(X_i)}) - Z_i \log(1 + e^{\beta + s(X_i)}) \right] - \lambda \int \left[ s'(t) \right]^2 dt \right\}.$$
(3)

We apply the maximum likelihood method to fit the age-specific CFR curve. We define  $W_k = 1$  when the *k*th confirmed case is dead, else  $W_k = 0$  with  $k = 1 \dots N$ , where *N* is the total number of the confirmed cases.  $Z_i$  represents the number of the confirmed cases at the age *i*, the log-likelihood function at the age  $X_i$  is defined by

$$\log\left\{\sum_{k=1}^{Z_i} \pi \left(X_i\right)^{w_k} \left[1 - \pi(X_i)\right]^{1 - w_k}\right\} = Y_i \log(e^{\beta + S(X_i)}) - Z_i \log(1 + e^{\beta + S(X_i)})$$
(4)

We use Root Mean Square Error (RMSE) to assess the degree of difference between the estimator and the actual quantity, that is,

$$RMSE = \sqrt{\frac{1}{100} \sum_{i=1}^{100} \left\{ \frac{Y_i}{Z_i} - \hat{\pi}(X_i) \right\}^2},$$
(5)

where  $\overset{\wedge}{\pi}(x) = \frac{e^{\overset{\wedge}{\beta} + \overset{\wedge}{s}(x)}}{1 + e^{\overset{\wedge}{\beta} + \overset{\wedge}{s}(x)}}$ .

We then obtain the estimated CFR by then formulation

$$\widehat{CFR} = \int_0^{99} \widehat{\pi}(x) \widehat{f}(x) dx, \qquad (6)$$

where  $\widehat{f(\mathbf{x})}$  was the kernel estimator of  $f(\mathbf{x})$ , where  $f(\mathbf{x})$  represents the density function of the age of diagnosed patients.

In this study, we also provide a novel method to compare the CFRs in different regions by adjusting the age effect of COVID-19 patients. For example, the SMR between other regions and Wuhan could be given by

$$SMR = \frac{\int_0^{99} \hat{f} Wuhan(x) \hat{\pi} nation(x)dx}{\int_0^{99} \hat{f} Wuhan(x) \hat{\pi} Wuhan(x)dx}$$
(7)

where region denoted other nations or regions.

We further define the Phased-age-specified unascertained rate (PUR) for Wuhan as the rate of patients with actual infection who were wrongly diagnosed as non-infected due to the mild and asymptomatic infection at different ages based on the CFR in the non-Hubei region in China, that is,

$$PUR = \frac{H - \int_{a}^{b} \widehat{f} Wuhan(x) dx}{H},$$
(8)

where  $H = \int_{a}^{b} \frac{\hat{\pi} Wuhan(x)}{\hat{\pi} non-Hubei(x)} \hat{f} Wuhan(x) dx.$ 

# Results

### Age-specific CFR function estimation

Figure 1 showed the numbers of deaths and confirmed cases of COVID-19 across different age groups in different regions. By 15 April 2020, the United States had the highest number of confirmed cases of COVID-19, while the CFR in France was the highest. The mean age of confirmed cases of COVID-19 was approximately 45.52, 56.17, 63.05, 61.48, 49.29, 57.40, 44.88, 52.53, 52.17, and 50.91 in the non-Hubei region, Wuhan, Italy, Spain, Germany, France, South Korea, Canada, the United States and New York, respectively. The intercept  $\hat{\beta}$  and nonlinear effect estimators  $\hat{S}(\mathbf{x})$  of the different regions were shown in the Table 1 and Figure 2. The estimated intercept and the nonlinear effect of age were significant for all countries and regions (P < .001), while small RMSE indicated that our model was wellfitted (Table 1). Figure 2(A) showed that France was the most affected country by COVID-19 in terms of age-specific CFR. We also found that the CFRs of COVID-19 remained steady with younger age and increased dramatically over the age of 65 in almost all regions. Only Canada rose sharply after the age of 80 years. Figure 2 also showed that the agespecific CFRs of age between 35 and 50 years remained lower level in non-Hubei, Germany, Canada and South Korea, and moderate in Wuhan, Italy and Spain, while relatively higher in France, the United States and New York. The CFR increases rapidly in line with ageing in France, New York, Wuhan and Italy, where the age-specific CFRs over 76 were much higher than that of the rest of regions.



**Figure 1.** Histogram and density plots of the age-specific deaths and confirmed cases in different regions. We present the age distribution of confirmed and dead COVID-19 cases in different regions as of June 15. The distribution of confirmed cases and deaths by age groups are shown by the light grey and dark grey histograms, respectively. The solid black line is a density curve that reflects the overall distribution of the sample. The dotted lines indicate the mean age of confirmed cases.

			Intercept $\hat{eta}$				Smooth terms ŝ (age)			
Region	RMSE	Est.	SE	t-value	<i>p</i> -value	edf	Ref.df	F-value	<i>p</i> -value	
Wuhan	0.1378	-3.64	0.06	-60.48	< .0001	5.10	6.07	1059.01	< .0001	
Non-Hubei	0.1122	-6.02	0.26	-23.38	< .0001	3.52	4.39	237.11	< .0001	
Italy	0.0100	-3.42	0.07	-46.42	< .0001	5.05	5.77	1778.62	< .0001	
Spain	0.0100	-4.02	0.06	-63.90	< .0001	5.00	5.82	2371.47	< .0001	
United States	0.0245	-5.15	0.33	-15.39	< .0001	1.29	1.52	40.56	< .0001	
France	0.0100	-5.80	0.53	-10.95	< .0001	2.82	3.50	144.93	< .0001	
Korea	0.0173	-7.03	0.59	-11.91	< .0001	2.60	3.26	251.84	< .0001	

**Table 1.** The behavior of estimates of intercept  $\beta$  and nonlinear effect of ages, and the RMSE among the regions.

Note: edf and Ref.df denote the estimated degrees of freedom for each model parameter.

#### The whole CFR estimation and the ratio of CFR estimation

In all countries and regions, our estimates of overall CFRs, ranged from 0.0094 for the non-Hubei region to 0.0828 for New York, were very much close to the actual CFRs (Table 2). Our estimates of SMR suggested that the age-specific CFR in New York was 1.77 times of Wuhan, which was the highest among all countries and regions. However, when compared to the non-Hubei region, France was of the highest age-specific CFR, with an SMR of 7.23 (Table 3).

Figure 3 showed the ratio of CFR function between other regions and Wuhan. The CFRs of all age groups in non-Hubei and South Korea were lower than that in Wuhan. France surpassed Wuhan in CFR in middle and advanced age groups, while Italy, Germany, Canada, Spain and the United States mainly exceeded Wuhan in older age groups. We also found that the CFRs of infant and under 2-years old children in the United States were higher than those in Wuhan.

The ratio of CFR function between other regions and non-Hubei region were shown in Figure 4. For almost all age groups in Italy, Spain, France, New York, and the United States and the middle and older age groups in Canada and Germany, the CFRs were higher than that in non-Hubei. The CFRs in South Korea in older age groups were higher than those in the non-Hubei region.

#### Phased-age-specific unascertained rate estimation

As mentioned above, the CFRs in the 10 countries or regions were quite different due to different age and sociological characteristics, various detection levels and confirmation definitions. However, the difference of CFR between Wuhan and non-Hubei, which were both from China, could not be easily explained by the above reasons. We speculated that the insufficient knowledge of the novel virus and limited testing capacity in Wuhan at the early stage of the outbreak resulted in a large number of infected individuals going unascertained. We assume that the age-specific CFR in Wuhan was consistent with that of non-Hubei and calculate age-specified unascertained rate.

As shown in Figure 3, there were two turning points of age at 46 and 84, in the comparison of age-specific CFR between Wuhan and non-Hubei regions in China, from which the unascertained rates in Wuhan were estimated to be 88.9% for ages 0–45, 76.0% for ages 46–84, and 48.1% for ages 85–99, respectively (Table 4). The total unascertained rate was 2510 (<del>`</del> P. DU ET AL.





Figure 2. CFR function estimates across different ages in different regions. (A) The age-specific CFR estimates for each country or region combined. (B). The age-specific CFR estimates for each country or region. The solid lines in black represent CFR of COVID-19 at different ages. The grey shadow is the confidence interval of the function estimate.

		Wuhan	Non-Hubei	Italy	Spain	Germany	France	Korea	Canada	United States	New York
Act.	CFR	0.0516	0.0092	0.1392	0.0827	0.0472	0.1886	0.0231	0.0811	0.0539	0.0836
Est.	CFR	0.0517	0.0094	0.1382	0.0821	0.0464	0.1876	0.0223	0.0812	0.0536	0.0828

Table 2. The actual and estimated age-specific CFRs in different countries and regions.

Table 3.	The SMR	calculation	of CFRs function	between Wuhan,	, non-Hubei and	l other rec	gions
----------	---------	-------------	------------------	----------------	-----------------	-------------	-------

	Italy	Spain	Germany	France	Korea	Canada	United States	New York	Wuhan	Non-Hubei
Wuhan	1.659	0.950	0.871	2.628	0.612	1.121	1.086	1.777	1.000	0.430
Non-Hubei	4.412	2.502	2.115	7.230	1.416	2.933	2.799	5.389	2.923	1.000

estimated to be 81.6%. We also estimated the total unascertained rate for New York City as 81.2%. The results were shown in Table 5.

#### Discussion

Compared with the previous discrete calculation method of SMR, the continuous method was more accurate and more stable. For the age-specific CFR curve newly designed in this study, it not only visually presented the effect of age on the fatality rate but also had a more practical value in calculating SMR and estimating the unascertained rate at different ages.

Our research delineated the age-specific CFR of COVID-19 in three Asian regions, four European regions, and three regions of the United States, which were mainly the worst-hit regions in the early stages of the global pandemic. The results proved that there was a precise relationship between age and the CFR of COVID-19, especially a rapid growth over the age of 65 in almost all regions in this study. Although the age-specific CFRs under 60 were similar in various regions, they become significant different in elderly people, which means something that happened in these regions have caused a difference in the CFR among the elderly people, particularly the age-specific CFRs over 76 were much higher in France, New York, Wuhan, and Italy. This phenomenon meant something happened in these regions that had caused a difference in the CFR among the elderly people there. The regions researched in this study were facing varying degrees of population aging, for instance, the population aged 65 years or older in Italy, France, and Wuhan is approximately 23%, 20%, and 15%, respectively. More than half of the SARS-CoV-2 infected cases in Italy were the people aged 60 and above [9]. Although people were generally susceptible to SARS-CoV-2, young people seemed to have mild symptoms or even be asymptomatic, while older people might have significantly worse outcomes considering their pre-existing health conditions [10].

Many scientific publications have evaluated the impact of comorbidities on the clinical outcomes in patients with COVID-19 [11–15]. In Guan *et al.*'s study, they found that patients with any comorbidity yielded poorer clinical outcomes than those without among laboratory-confirmed cases of COVID-19 [16]. Among patients with COVID-19, circulatory diseases (including hypertension and coronary heart diseases) and endocrine diseases such as diabetes remained the most common categories of comorbidity [16–18]. Our previous studies also echoed with recently published studies in terms of identifying the risk factors of hypertension and diabetes in patients with COVID-19 [19,20]. We have to admit that the individual data on COVID-19 patient comorbidities status were not available in



**Figure 3.** The ratio of CFR function between other areas and Wuhan. We used CFRs at all ages in Wuhan as a baseline to compare with other regions, and the horizontal solid line is the base line equal to 1. The CFR comparison between different regions and Wuhan at all ages can be reflected by the curved solid line. Ages that lie above the baseline have a higher CFR than Wuhan and vice versa. We also showed the turning points of age.



**Figure 4.** The ratio of CFR function between other regions and non-Hubei region. Same as Figure 3 but for non-Hubei region.

#### 2514 😉 P. DU ET AL.

Table 4. The estimation of the PUR in Wuha
--

Wuhan	a–b (years)	0–45	46-84	85–99	0–99
	PUR	0.888	0.760	0.481	0.816

#### Table 5. The estimation of the PUR in New York.

New York	a-b (years)	0–24	25–36	37–45	46–68	69–99	0–99
	PUR	0.857	0.806	0.839	0.844	0.334	0.812

the current study. Since the comorbidity are highly prevalent among elderly people and the older age is also been regarded as one of the risk factors for severe outcomes among patients with COVID-19, it is reasonable that higher CFR was found among elderly.

The CFR was still changing along with the ongoing pandemic, yet what we have learned from available data might help a lot in dealing with this complicated situation. In this study, we provide a novel method to compare the CFRs in different regions after continuous standardized using the age of COVID-19 patients. According to the SMR calculation, we found that the CFR in Italy, France, Canada, the United States, and New York was higher than that in Wuhan, China, while other regions showed lower CFRs. But compared with the non-Hubei in China, none of the nine regions was found lower CFRs. The age-specific CFRs of France was the highest among the nine regions, most likely due to the lowest test rate of less than 8000 per 1 million populations. The test rate was over 10,000 in South Korea, the United States, and Spain, and 20,000 in German and Italy, 30,000 in New York [21]. Compared with Wuhan, the top two countries with the SMR were France (2.628) and Italy (1.659). The age-specific CFR was relatively higher at the age groups of 56-88 years with a peak age of 69 in Italy, and at the age groups of 57-77 with a peak age of 70 in France. Therefore, these two countries should attach great importance to the diagnosis and treatment of people over age of 55 who were infected with the COVID-19 and took timely measures to avoid the deterioration of the disease and adverse outcomes. In addition, it is worth noting that there was a peak of CFR in the age group of 80-85 of non-Hubei, South Korea, France, and Germany, suggesting that these four regions should pay special attention to the prognosis of infected persons in the older age groups.

Studies had shown that there were 30–60% asymptomatic cases in SARS-CoV-2 infected [22]. As the first epicenter of COVID-19, Wuhan encountered a shortage of both test kit and manpower, which means only the cases with typical manifestation might have the priority of laboratory examination. Although Wuhan had reported more than half of the confirmed case in China, there might be still a significant portion of mild and asymptomatic cases under detected. Given that the epidemic in China was nearing its end at the time of data collection, the CFR was estimated to be relatively stable, we estimated the unascertained rates of COVID-19 infected cases in Wuhan according to the CFR of non-Hubei, China. The unascertained rate was the highest at the age group of 45 and below and gradually decreases with age. This was in line with the epidemiological characteristics of previous studies finding that most of the COVID-19 infections presented mild and asymptomatic symptoms among younger people [23–25]. Because the prevalence of comorbidity was higher in the older age group, the symptoms after infection were more evident and severe; the unascertained rate of people over 84 years old was lower. Our estimate of the

unascertained rate in Wuhan was slightly lower than Li *et al.*'s estimate of 86% [26]. Compared to their method of estimation using the Susceptible-Exposed-Infected-Recovered (SEIR) model, this study did not need to set parameters and could easily demonstrate the relationship between unascertained phenomenon and age.

There are several limitations in our study. First, our cross-sectional study design could not make an accurate estimation of the CFR with the time effect. Second, the prevalence of comorbidity might be a confounding factor that affected the relationship between age and the CFR. Lack of the individual-level information of age-specific comorbidity for COVID-19 patients might remain uncertain in our estimates. Our estimate of the unascertained rate still needs to be confirmed by further research such as field investigations and seroepidemiological studies. Third, the data we collected from mid-June had a varying degree of impact on CFR in different regions as they were in different time spans during the ongoing epidemic. Non-Hubei region of China, for instance, had a relatively stable CFR, because it was in the middle and late stages of the epidemic. However, the United States was in the early stages of the epidemic, which made the CFR uncertain. We also found that age was a risk factor for death in patients with COVID-19, although there was no difference in the risk of death across different epidemic stages.

In conclusion, our study provides a novel method for estimating the SMR and unascertained rate. Moreover, it found an additional evidence that the risk of COVID-19 deaths increases with age, and most of the deaths observed in different regions occurred in people aged 65 and above. Younger people who infected SARS-CoV-2 might be undiagnosed, due to mild and asymptomatic infections. However, they were essential factors of transmission and lack of treatment that might contribute to the deaths of older people. Therefore, we should pay more attention to the mild and asymptomatic infections among young and middle-aged people to avoid the spread to the older people who are more vulnerable.

#### **Disclosure statement**

No potential conflict of interest was reported by the author(s).

#### Funding

The study was funded by the National Natural Science Foundation of China [grant number 81903406], the National Natural Science Foundation of China [grant number 11901352), Natural Science Foundation of Shandong Province [grant number ZR2019BA017], Social Science Foundation of Shandong Province [grant number 35019DTJJ03], the Young Scholars Program of Shandong University [YSPSDU: 11020088964008] and 'Coronavirus Disease Special Project' of Chengdu University of Traditional Chinese Medicine [XGZX2013].

#### ORCID

*Peipei Du* http://orcid.org/0000-0001-6582-8384 *Peihua Cao* http://orcid.org/0000-0003-4269-3899 *Daihai He* http://orcid.org/0000-0003-3253-654X *Lin Yang* http://orcid.org/0000-0002-5964-3233 *Shu Yang* http://orcid.org/0000-0002-1859-6362

# References

- [1] Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, Wu Y, Zhang L, Yu Z, Fang M, Yu T, Wang Y, Pan S, Zou X, Yuan S, Shang Y. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. Lancet Respir Med. 2020 May;8(5):475-481.
- [2] Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, Xiang J, Wang Y, Song B, Gu X, Guan L, Wei Y, Li H, Wu X, Xu J, Tu S, Zhang Y, Chen H, Cao B. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet. 2020 Mar 28;395(10229):1054-1062.
- [3] WHO-Coronavirus disease (COVID-19) Weekly Epidemiological Update and Weekly Operational Update. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situationreports
- [4] Dudel C, Riffe T, Acosta E, van Raalte A, Strozza C, Myrskylä M. Monitoring trends and differences in COVID-19 case-fatality rates using decomposition methods: Contributions of age structure and age-specific fatality. PLoS One. 2020 Sep 10;15(9):e0238904.
- [5] Green MS, Peer V, Schwartz N, Nitzan D. The confounded crude case-fatality rates (CFR) for COVID-19 hide more than they reveal-a comparison of age-specific and age-adjusted CFRs between seven countries. PLoS One. 2020 Oct 21;15(10):e0241031.
- [6] Iosa M, Paolucci S, Morone G. Covid-19: A Dynamic Analysis of Fatality Risk in Italy. Front Med (Lausanne). 2020 Apr 30;7:185.
- [7] Rubino S, Kelvin N, Bermejo-Martin JF, Kelvin D. As COVID-19 cases, deaths and fatality rates surge in Italy, underlying causes require investigation. J Infect Dev Ctries. 2020 Mar 31;14(3):265-267.
- [8] Liang WH, Guan WJ, Li CC, Li YM, Liang HR, Zhao Y, Liu XQ, Sang L, Chen RC, Tang CL, Wang T, Wang W, He QH, Chen ZS, Wong SS, Zanin M, Liu J, Xu X, Huang J, Li JF, Ou LM, Cheng B, Xiong S, Xie ZH, Ni ZY, Hu Y, Liu L, Shan H, Lei CL, Peng YX, Wei L, Liu Y, Hu YH, Peng P, Wang JM, Liu JY, Chen Z, Li G, Zheng ZJ, Qiu SQ, Luo J, Ye CJ, Zhu SY, Cheng LL, Ye F, Li SY, Zheng JP, Zhang NF, Zhong NS, He JX. Clinical characteristics and outcomes of hospitalised patients with COVID-19 treated in Hubei (epicentre) and outside Hubei (non-epicentre): a nationwide analysis of China. Eur Respir J. 2020 Jun 4;55(6):2000562.
- [9] Epidemia COVID-19 Aggiornamento nazionale 9 aprile 2020-ore 16:00. Published April 10, 2020. Accessed April 20, 2020.
- [10] WHO-COVID-19 STRATEGY UPDATE 14 April 2020. doi; file:///Users/daocaorenjie/ Downloads/covid-strategy-update-14april2020.pdf
- [11] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J, Cao B. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020 Feb 15;395(10223):497-506.
- [12] Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, Qiu Y, Wang J, Liu Y, Wei Y, Xia J, Yu T, Zhang X, Zhang L. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet. 2020 Feb 15;395(10223):507-513.
- [13] Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, Wang B, Xiang H, Cheng Z, Xiong Y, Zhao Y, Li Y, Wang X, Peng Z. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. JAMA. 2020 Mar 17;323(11):1061-1069.
- [14] Liu K, Fang YY, Deng Y, Liu W, Wang MF, Ma JP, Xiao W, Wang YN, Zhong MH, Li CH, Li GC, Liu HG. Clinical characteristics of novel coronavirus cases in tertiary hospitals in Hubei Province. Chin Med J (Engl). 2020 May 5;133(9):1025-1031.
- [15] Xu XW, Wu XX, Jiang XG, Xu KJ, Ying LJ, Ma CL, Li SB, Wang HY, Zhang S, Gao HN, Sheng JF, Cai HL, Qiu YQ, Li LJ. Clinical findings in a group of patients infected with the 2019 novel coronavirus (SARS-Cov-2) outside of Wuhan, China: retrospective case series. BMJ. 2020 Feb 19;368:m606.
- [16] Guan WJ, Liang WH, Zhao Y, Liang HR, Chen ZS, Li YM, Liu XQ, Chen RC, Tang CL, Wang T, Ou CQ, Li L, Chen PY, Sang L, Wang W, Li JF, Li CC, Ou LM, Cheng B, Xiong S, Ni ZY,

Xiang J, Hu Y, Liu L, Shan H, Lei CL, Peng YX, Wei L, Liu Y, Hu YH, Peng P, Wang JM, Liu JY, Chen Z, Li G, Zheng ZJ, Qiu SQ, Luo J, Ye CJ, Zhu SY, Cheng LL, Ye F, Li SY, Zheng JP, Zhang NF, Zhong NS, He JX; China Medical Treatment Expert Group for COVID-19. Comorbidity and its impact on 1590 patients with COVID-19 in China: a nationwide analysis. Eur Respir J. 2020 May 14;55(5):2000547.

- [17] Li X, Xu S, Yu M, Wang K, Tao Y, Zhou Y, Shi J, Zhou M, Wu B, Yang Z, Zhang C, Yue J, Zhang Z, Renz H, Liu X, Xie J, Xie M, Zhao J. Risk factors for severity and mortality in adult COVID-19 inpatients in Wuhan. J Allergy Clin Immunol. 2020 Jul;146(1):110-118.
- [18] Rodriguez-Morales AJ, Cardona-Ospina JA, Gutiérrez-Ocampo E, Villamizar-Peña R, Holguin-Rivera Y, Escalera-Antezana JP, Alvarado-Arnez LE, Bonilla-Aldana DK, Franco-Paredes C, Henao-Martinez AF, Paniz-Mondolfi A, Lagos-Grisales GJ, Ramírez-Vallejo E, Suárez JA, Zambrano LI, Villamil-Gómez WE, Balbin-Ramon GJ, Rabaan AA, Harapan H, Dhama K, Nishiura H, Kataoka H, Ahmad T, Sah R; Latin American Network of Coronavirus Disease 2019-COVID-19 Research (LANCOVID-19). Electronic address: https://www.lancovid.org. Clinical, laboratory and imaging features of COVID-19: A systematic review and meta-analysis. Travel Med Infect Dis. 2020 Mar-Apr;34:101623.
- [19] Cao P, Song Y, Zhuang Z, Ran J, Xu L, Geng Y, Han L, Zhao S, Qin J, He D, Wu F, Yang L. Obesity and COVID-19 in Adult Patients With Diabetes. Diabetes. 2021 May;70(5):1061-1069.
- [20] Ran J, Song Y, Zhuang Z, Han L, Zhao S, Cao P, Geng Y, Xu L, Qin J, He D, Wu F, Yang L. Blood pressure control and adverse outcomes of COVID-19 infection in patients with concomitant hypertension in Wuhan, China. Hypertens Res. 2020 Nov;43(11):1267-1276.
- [21] Reported Cases and Deaths by Country, Territory, or Conveyance. Doi: https://www.worldo meters.info/coronavirus/#countries
- [22] Qiu J. Covert coronavirus infections could be seeding new outbreaks [published online ahead of print, 2020 Mar 20]. Nature. 2020;10.1038/d41586-020-00822-x. doi:10.1038/d41586-020-00822-x
- [23] Mei X, Zhang Y, Zhu H, Ling Y, Zou Y, Zhang Z, Guo H, Liu Y, Cheng X, Liu M, Huang W, Wang J, Yi Z, Qian Z, Lu H. Observations about symptomatic and asymptomatic infections of 494 patients with COVID-19 in Shanghai, China. Am J Infect Control. 2020 Sep;48(9):1045-1050.
- [24] Choi WS, Kim HS, Kim B, Nam S, Sohn JW. Community Treatment Centers for Isolation of Asymptomatic and Mildly Symptomatic Patients with Coronavirus Disease, South Korea. Emerg Infect Dis. 2020 Oct;26(10):2338-2345.
- [25] Ye Y, Fan W, Wang WH, Wang HF, Pan JJ, Nie YF, Yong AG, Huang XY. Difference in epidemic characteristics between asymptomatic infected persons and confirmed cases in COVID-19 clustered epidemics [J]. Chinese Journal of Infection Control,2020,19(06):492-497.
- [26] Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, Shaman J. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). Science. 2020 May 1;368(6490):489-493.