

# Risk Factors for Death Among the First 80 543 Coronavirus Disease 2019 (COVID-19) Cases in China: Relationships Between Age, Underlying Disease, Case Severity, and Region

Yanping Zhang,<sup>1,a</sup> Wei Luo,<sup>1,a</sup> Qun Li,<sup>1,a</sup> Xijie Wang,<sup>2</sup> Jin Chen,<sup>1</sup> Qinfeng Song,<sup>2</sup> Hong Tu,<sup>1</sup> Ruiqi Ren,<sup>1</sup> Chao Li,<sup>1</sup> Dan Li,<sup>1</sup> Jing Zhao,<sup>1</sup> Jennifer M. McGoogan,<sup>1</sup> Duo Shan,<sup>1</sup> Bing Li,<sup>1</sup> Jingxue Zhang,<sup>2</sup> Yanhui Dong,<sup>2</sup> Yu Jin,<sup>2</sup> Shuai Mao,<sup>2</sup> Menbao Qian,<sup>1</sup> Chao Lv,<sup>1</sup> Huihui Zhu,<sup>1</sup> Limin Wang,<sup>1</sup> Lin Xiao,<sup>1</sup> Juan Xu,<sup>1</sup> Dapeng Yin,<sup>1</sup> Lei Zhou,<sup>1</sup> Zhongjie Li,<sup>1</sup> Guoqing Shi,<sup>1</sup> Xiaoping Dong,<sup>1</sup> Xuhua Guan,<sup>3</sup> George F. Gao,<sup>1</sup> Zunyou Wu,<sup>1,b</sup> and Zijian Feng<sup>1,b</sup>

<sup>1</sup>Chinese Center for Disease Control and Prevention, Beijing, China; <sup>2</sup>Peking University, Beijing, China; and <sup>3</sup>Hubei Center for Disease Control and Prevention, Wuhan, China

**Background.** Knowledge of COVID-19 epidemiology remains incomplete and crucial questions persist. We aimed to examine risk factors for COVID-19 death.

**Methods.** A total of 80 543 COVID-19 cases reported in China, nationwide, through 8 April 2020 were included. Risk factors for death were investigated by Cox proportional hazards regression and stratified analyses.

**Results.** Overall national case-fatality ratio (CFR) was 5.64%. Risk factors for death were older age ( $\geq 80$ : adjusted hazard ratio, 12.58; 95% confidence interval, 6.78–23.33), presence of underlying disease (1.33; 1.19–1.49), worse case severity (severe: 3.86; 3.15–4.73; critical: 11.34; 9.22–13.95), and near-epicenter region (Hubei: 2.64; 2.11–3.30; Wuhan: 6.35; 5.04–8.00). CFR increased from 0.35% (30–39 years) to 18.21% ( $\geq 70$  years) without underlying disease. Regardless of age, CFR increased from 2.50% for no underlying disease to 7.72% for 1, 13.99% for 2, and 21.99% for  $\geq 3$  underlying diseases. CFR increased with worse case severity from 2.80% (mild) to 12.51% (severe) and 48.60% (critical), regardless of region. Compared with other regions, CFR was much higher in Wuhan regardless of case severity (mild: 3.83% vs 0.14% in Hubei and 0.03% elsewhere; moderate: 4.60% vs 0.21% and 0.06%; severe: 15.92% vs 5.84% and 1.86%; and critical: 58.57% vs 49.80% and 18.39%).

**Conclusions.** Older patients regardless of underlying disease and patients with underlying disease regardless of age were at elevated risk of death. Higher death rates near the outbreak epicenter and during the surge of cases reflect the deleterious effects of allowing health systems to become overwhelmed.

**Keywords.** case-fatality ratio; China; COVID-19; risk factors; SARS-CoV-2.

As the first country to encounter coronavirus disease 2019 (COVID-19), experience a large outbreak, and achieve epidemic control [1, 2], a broad range of epidemiologic studies have been conducted in China [3–21]. These studies used data from China's 4 complementary infectious disease information systems [2, 22]. However, most have included few cases from small areas or single centers during short periods in January–February 2020 [3–21]. To date, the largest COVID-19 case series in China included all 72 314 patients with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) as of 11 February 2020 [6, 22]. This report provided the first

epidemiologic curves, a timeline of discovery and response events, and distributions of cases across age, case severity, geography, and time. Importantly, it highlighted higher case-fatality ratios (CFRs) among older adults and adults with underlying disease [6, 22].

However, patient data were abruptly cut off on 11 February and analyzed immediately while many remained hospitalized. Therefore, for many of these patients, the dataset did not include follow-up through the entire course of illness to recovery or death. Moreover, in April 2020, Wuhan officials added 1615 records of confirmed COVID-19 cases (325 survivors and 1290 deaths), most of which had occurred in January and February but had gone unreported due to overwhelmed systems during the case surge. Thus, our understanding of COVID-19 epidemiology in China remains incomplete, and many gaps in the evidence persist. For instance, risk factors for death from COVID-19 in China have been identified but have yet to be quantified and thoroughly investigated in a large cohort [23, 24]. Therefore, our primary aim was to investigate risk factors for death and explore the relationships between them using all confirmed COVID-19 cases nationwide followed for their entire clinical course in the complete first “wave” of the epidemic

Received 16 March 2021; editorial decision 20 May 2021; published online 27 May 2021.

<sup>a</sup>Y. Z., W. L., and Q. L. contributed equally.

<sup>b</sup>Z. W. and Z. F. contributed equally as joint senior authors.

Correspondence: Z. Wu, Chinese Center for Disease Control and Prevention, 155 Changbai Rd, Changping District, Beijing 102206, China (wuzy@263.net).

Clinical Infectious Diseases® 2022;74(4):630–8

© The Author(s) 2021. Published by Oxford University Press for the Infectious Diseases Society of America. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)  
DOI: 10.1093/cid/ciab493

in China. Second, we aimed to more fully characterize COVID-19 epidemiology.

## METHODS

### Design

A nationwide retrospective cohort study design was used to investigate all confirmed COVID-19 cases in China as of 8 April 2020, followed through 16 May 2020. This study was approved by the Chinese Center for Disease Control and Prevention Institutional Review Board. Individual informed consent was not required.

### Data

In China, all COVID-19 cases must be entered into the Notifiable Infectious Disease Reporting System (NIDRS) within 2 hours of discovery [2, 25]. Since all NIDRS case reports contain individuals' unique national identification (ID) numbers, the system contains no duplicate reports [2, 25]. Each NIDRS case record must be investigated by local public health specialists within 24 hours of reporting [2]. Investigation results are collected, stored, and managed in the Epidemiological Investigation Information System (EIIS) [2]. Each record in EIIS contains both the individual's national ID number and NIDRS case report number to prevent duplicate records and facilitate records matching [2].

### Cases

The case definition in China—positive SARS-CoV-2 polymerase chain reaction (PCR) test results and documented symptoms—has been consistent throughout the pandemic and our entire study with 1 exception: in Hubei Province (including in Wuhan City), early in the outbreak when testing had not yet scaled up sufficiently, some patients were clinically diagnosed based on symptoms and lung imaging [6]. Notably, this case definition excludes asymptomatic infection. Thus, all data from all records in NIDRS and EIIS with positive SARS-CoV-2 PCR test results and documented symptoms through 8 April 2020 were included, but extracted approximately 6 weeks later on 16 May 2020. This delay was required to ensure records of all previously unreported cases and deaths had been completed, quality checked, and released and to allow all patients to be followed through their entire clinical course to recovery or death.

### Deaths

Since all persons with positive SARS-CoV-2 PCR test results were hospitalized in isolation in China [2], all deaths occurred in hospitals and thus were ascertained by physicians. All deaths among COVID-19 cases were recorded as COVID-19 deaths regardless of any other circumstances surrounding their deaths (eg, coinfection, myocardial infarction, stroke). No deaths were ascertained postmortem as COVID-19.

### Variables

For occupation, the health worker category was defined as any type of active employment in any health facility. For underlying disease, having no major pre-existing clinical diagnosis was categorized as “no” while having any was categorized as “yes.” Cases with a single underlying disease were further categorized by disease (eg, hypertension) and by number of diseases (eg,  $\geq 3$  diseases). For case severity, categories were mild, moderate, severe, and critical. Mild cases had mild clinical symptoms and no sign of pneumonia on imaging. Moderate cases had fever and respiratory symptoms with radiological findings of pneumonia. Severe cases were characterized by dyspnea, respiratory rate (RR) 30 or more breaths per minute, oxygen saturation of 93% or less, PaO<sub>2</sub> to FiO<sub>2</sub> ratio (ie, ratio of arterial oxygen partial pressure to fractional inspired oxygen) less than 300 for adults, and any one of the following for children: (1) tachypnea (RR,  $\geq 60$  breaths/minute for infants aged <2 months,  $\geq 50$  breaths/minute for infants aged 2–12 months,  $\geq 40$  breaths/minute for children aged 1–5 years, and  $\geq 30$  breaths/minute for children >5 years) independent of fever and crying; (2) oxygen saturation of 92% or less; (3) labored breathing (ie, moaning, nasal fluttering, and infrasternal, supraclavicular, and intercostal retraction), cyanosis, and intermittent apnea; (4) lethargy and convulsion; or (5) difficulty feeding and signs of dehydration. Critical cases had respiratory failure, septic shock, and/or organ failure that required intensive care. For region, 3 were defined as China except for Hubei Province, Hubei except Wuhan City, and Wuhan only. For epidemic stage, cases were categorized into 5 periods based on the date of symptom onset.

### Analysis

Cases and deaths are presented as numbers and cases at the national level only as percentages calculated as the number of cases in a category (numerator) divided by the total number of cases (denominator). Incidence and mortality were calculated as the number of cases and deaths, respectively (numerator), divided by the total population (denominator), presented as /100 000. CFR was calculated as the number of deaths (numerator) divided by the total number of confirmed cases (denominator), presented as a percentage. Univariate and multivariate Cox regression models, based on survival analysis, were used to examine risk factors associated with death from COVID-19, producing unadjusted and adjusted hazard ratios (aHRs) and 95% confidence intervals (CIs). Age, sex, residence, underlying disease, case severity, and region variables were adjusted in the multivariate model (STATA version 14.0; StataCorp, College Station, TX, USA). Stratified analysis with chi-square test was used to examine the relationships between variables. Regional differences were evaluated using Pearson chi-square and Fisher's exact test. National and regional epidemiological curves were plotted as daily number of confirmed cases versus date

of symptom onset. Confirmed cases by province and date of symptom onset were presented as a heatmap time series to illustrate geographic spread of COVID-19 (RStudio v3.6.3; RStudio, PBC, rstudio.com).

## RESULTS

### Participants

A total of 80 543 confirmed COVID-19 cases were used in the analysis (asymptomatic infections excluded but summarized in [Supplementary Table 1](#)). In [Table 1](#), most were aged 40–69 years (60.68%) and were urban residents (75.81%). A total of 34 320 cases (42.61%) had information on underlying disease in their records. Among them, 25 743 had no underlying disease (75.01%), 4923 had 1 underlying disease (14.34%), 1458 had 2 underlying diseases (17.00%), and 391 had 3 or more underlying diseases (4.56%). Among those with 1 underlying disease, hypertension was most common (53.14%). Most cases were mild (41.96%) or moderate (39.17%) at diagnosis; severe cases were less common (14.89%) and critical cases were rare (3.46%).

### Risk Factors for Death

In [Table 2](#), compared to being younger than 30 years, older age was associated with greater risk of death: 3.5 times greater for 50–59 (aHR, 3.52; 95% CI, 1.91–6.51), approximately 7 times greater for 60–69 (aHR, 6.71; 95% CI, 3.66–12.30), 9 times greater for 70–79 (aHR, 8.95; 95% CI, 4.86–16.48), and 12.5 times greater for 80 years or older (aHR, 12.58; 95% CI, 6.78–23.33). Having any underlying disease (compared with none) was associated with a 33% greater risk of death (aHR, 1.33; 95% CI, 1.19–1.49). Having only vascular disease (aHR, 1.43; 95% CI, 1.07–1.91) or only kidney disease (aHR, 2.19; 95% CI, 1.13–4.24) and having 2 diseases (aHR, 1.43; 95% CI, 1.21–1.69) or 3 or more diseases (aHR, 2.16; 95% CI, 1.68–2.77) were also associated with greater risk of death. Compared with mild cases, severe (aHR, 3.86; 95% CI, 3.15–4.73) and critical (aHR, 11.34; 95% CI, 9.22–13.95) cases had a significantly higher risk of death. Finally, increased risk of death was associated with diagnosis in Hubei (aHR, 2.64; 95% CI, 2.11–3.30) and Wuhan (aHR, 6.35; 95% CI, 5.04–8.00), compared with elsewhere.

### Age Versus Underlying Disease

In [Table 3](#), among those with no underlying disease, CFR increased significantly with age from 0.7% for 40–49 years to 1.9% for 50–59 years, 6.7% to 60–69 years, and 18.2% for 70 years or older ( $P < .001$ ). Significant increases in CFR with increasing age were observed for those with 1 underlying disease ( $P < .001$ ), 2 diseases ( $P < .001$ ), and 3 or more diseases ( $P = .001$ ). Conversely, CFR also increased significantly with the number of underlying diseases for single age groups. For example, for those aged 50–59 years, CFR significantly increased from 1.9% for no underlying disease to 2.8% for 1 underlying disease, 6.2% for 2 diseases, and 10.2% for 3 or more diseases ( $P < .001$ ). Significant increases in CFR with increasing age

were also observed for single, specific underlying conditions ([Supplementary Table 2](#)). Among those with a single underlying disease, CFR remained highest in the oldest age groups. For example, in the 70-years and older age group, CFR was 33.3% for liver disease, 26.7% for kidney disease, 26.3% for cancer, 21.0% for diabetes, 20.9% for hypertension, and 20.4% for vascular disease.

### Case Severity and Region

[Figure 1](#) shows progression from initial case severity at diagnosis through worst status reached during hospitalization to final outcome of recovery or death. Among those diagnosed as mild (56.2%) or moderate (30.0%), most did not progress further (74.2% and 93.3%, respectively) and recovered (97.2% and 97.8%, respectively). However, among those already severe (12.1%) or critical (1.6%) at diagnosis and among those who were ever severe (15.5%) or critical (3.7%) during hospitalization, CFRs were high ( $\geq 12\%$  and  $>47\%$ , respectively; [Supplementary Table 3](#)). As shown in [Table 4](#), those diagnosed at the critical case status had very high CFRs in all regions (58.87% in Wuhan, 49.80% in Hubei, and 18.39% in the rest of China), regardless of proximity to the outbreak epicenter. However, the highest regional CFR was observed in Wuhan (7.59%) and increasing CFR with increasing proximity to the outbreak was observed across all age groups and regardless of underlying disease or case severity.

### COVID-19 Epidemiology

Nationally, incidence was 5.79 per 100 000 population. Incidence was highest for 60- to 69-year-olds (13.38/100 000) and urban residents (7.38/100 000). Altogether, 4545 deaths were recorded for national mortality of 0.33 per 100 000 and a CFR of 5.64%. Elevated mortality and CFR were observed for individuals aged 80 years and older (3.38/100 000; 32.08%), retirees (12.90%), severe and critical cases (12.51% and 48.60%), cases with underlying disease (8.69%), and cases with symptom onset early in the epidemic (15.84%) ([Table 1](#)). Notably, CFR increased with age, numbers of underlying diseases, and case severity, and decreased with later epidemic stage in all regions ([Table 4](#)). Additional data on cases, incidence, deaths, mortality, and CFRs by region are presented in [Supplementary Table 4](#).

As shown in [Supplementary Figure 1](#), the epidemiologic curves for the national level and the 3 regions have similar shapes but notably different amplitudes, with Wuhan clearly having the largest numbers of daily confirmed cases. Daily cases remained at a very low level through 31 December 2019 and then began to increase on 1 January 2020 at both national and regional levels. At the national level, daily new cases from early to mid-January increased to more than 2000 per day. For most of the period from 23 January to 5 February, the number of daily new cases exceeded 3000, with a peak of 3734 cases on 1 February. Then, the number of cases

**Table 1. Characteristics of COVID-19 Cases and Incidence, Deaths, Mortality, and Case-Fatality Ratios in China**

Characteristics	Cases, <sup>a</sup> n (%)	Incidence <sup>b</sup>	Deaths, n	Mortality <sup>b</sup>	CFR, %
Overall	80 543 (100)	5.79	4545	0.33	5.64
Age, y					
<10	936 (1.16)	0.56	2	<0.01	0.21
10-19	1035 (1.29)	0.69	3	<0.01	0.29
20-29	6101 (7.57)	2.54	19	0.01	0.31
30-39	12 928 (16.05)	6.65	70	0.04	0.54
40-49	14 796 (18.37)	5.92	166	0.07	1.12
50-59	17 925 (22.26)	10.78	556	0.33	3.10
60-69	16 156 (20.06)	13.38	1325	1.10	8.20
70-79	7499 (9.31)	10.16	1388	1.88	18.51
≥80	3167 (3.93)	10.54	1016	3.38	32.08
Sex					
Male	39 843 (49.47)	5.60	2912	0.41	7.31
Female	40 700 (50.53)	5.98	1633	0.24	4.01
Residence					
Urban	61 060 (75.81)	7.38	3960	0.48	6.49
Rural	16 322 (20.26)	2.91	457	0.08	2.80
Missing	3161 (3.92)	...	128	...	4.05
Occupation					
Service industry	6487 (8.05)	...	83	...	1.28
Farmer/laborer	14 148 (17.57)	...	492	...	3.48
Health worker	3282 (4.07)	...	23	...	0.70
Retiree	18 373 (22.81)	...	2371	...	12.90
Other/none	38 253 (47.49)	...	1576	...	4.12
Underlying disease					
Missing	46 223 (57.39)	...	3157	...	6.83
No	25 743 (31.96)	...	643	...	2.50
Yes <sup>a</sup>	8577 (10.65)	...	745	...	8.69
One disease	4923 (57.40)	...	380	...	7.72
Hypertension	2616 (30.50)	...	210	...	8.03
Diabetes	732 (8.53)	...	47	...	6.42
Vascular disease	502 (5.85)	...	54	...	10.76
Lung disease	692 (8.07)	...	37	...	5.35
Kidney disease	105 (1.22)	...	10	...	9.52
Liver disease	161 (1.88)	...	5	...	3.11
Cancer	115 (1.34)	...	17	...	14.78
Two diseases	1458 (17.00)	...	204	...	13.99
Three or more diseases	391 (4.56)	...	86	...	21.99
Missing	1805 (21.04)	...	75	...	4.16
Case severity					
Mild	33 799 (41.96)	...	945	...	2.80
Moderate	31 547 (39.17)	...	721	...	2.29
Severe	11 994 (14.89)	...	1500	...	12.51
Critical	2784 (3.46)	...	1353	...	48.60
Missing	419 (0.52)	...	26	...	6.21
Region					
China except Hubei	12 958 (16.09)	0.97	121	0.01	0.93
Hubei except Wuhan	17 795 (22.09)	36.98	643	1.34	3.61
Wuhan only	49 790 (61.82)	457.09	3781	34.71	7.59
Epidemic stage					
Early (8–31 December)	202 (0.25)	...	32	...	15.84
Rise (1–22 January)	14 105 (17.51)	...	1203	...	8.53
Plateau (23 January–5 February)	47 183 (58.58)	...	2572	...	5.45
Decline (6–29 February)	18 803 (23.35)	...	731	...	3.89
Late (1 March–8 April)	250 (0.31)	...	7	...	2.80

Abbreviations: CFR, case-fatality ratio; COVID-19, coronavirus disease 2019.

<sup>a</sup>Proportions for cases were calculated using all cases (N = 80 543) as the denominator except for subcategories under “Yes” for the Underlying disease variable, which use total cases with underlying disease (n = 8577) as the denominator for each different number of underlying diseases and uses total with 1 underlying disease (n = 4923) as the denominator for each different single specific disease.

<sup>b</sup>Incidence and mortality are expressed as /100 000 population. For additional context, historical values for mortality in China, nationally, and in Wuhan City specifically are provided in Supplementary Table 5.

**Table 2. Risk Factors for Death Among Confirmed COVID-19 Cases in China**

Characteristics	Unadjusted		Adjusted	
	HR (95% CI)	P <sup>a</sup>	HR (95% CI)	P <sup>a</sup>
<b>Age, y</b>				
<30	1		1	
30–39	1.81 (1.14–2.87)	.012	1.24 (.61–2.52)	.547
40–49	3.65 (2.38–5.60)	<.001	2.28 (1.20–4.34)	.012
50–59	9.21 (6.12–13.85)	<.001	3.52 (1.91–6.51)	<.001
60–69	20.71 (13.83–31.02)	<.001	6.71 (3.66–12.30)	<.001
70–79	39.64 (26.46–59.40)	<.001	8.95 (4.86–16.48)	<.001
≥80	50.55 (33.66–75.92)	<.001	12.58 (6.78–23.33)	<.001
<b>Sex</b>				
Male	1		1	
Female	.68 (.64–.72)	<.001	.84 (.75–.95)	.005
<b>Residence</b>				
Urban	1		1	
Rural	.57 (.52–.63)	<.001	1.03 (.89–1.19)	.702
<b>Underlying disease</b>				
No	1		1	
Yes	2.68 (2.41–2.99)	<.001	1.33 (1.19–1.49)	<.001
Hypertension	1.92 (1.65–2.24)	<.001	1.18 (1.00–1.38)	.050
Diabetes	1.86 (1.39–2.49)	<.001	1.36 (.99–1.86)	.057
Vascular disease	2.38 (1.82–3.11)	<.001	1.43 (1.07–1.91)	.015
Lung disease	1.07 (.78–1.49)	.663	.96 (.69–1.35)	.825
Kidney disease	3.12 (1.68–5.81)	<.001	2.19 (1.13–4.24)	.020
Liver disease	.84 (.35–2.02)	.699	1.64 (.68–3.97)	.270
Cancer	2.32 (1.44–3.75)	.001	1.53 (.94–2.51)	.088
Two diseases	2.38 (2.06–2.75)	<.001	1.43 (1.21–1.69)	<.001
Three or more diseases	4.20 (3.39–5.20)	<.001	2.16 (1.68–2.77)	<.001
<b>Case severity</b>				
Mild	1		1	
Moderate	.81 (.74–0.90)	<.001	.73 (.57–.95)	.018
Severe	3.50 (3.22–3.80)	<.001	3.86 (3.15–4.73)	<.001
Critical	6.41 (5.83–7.05)	<.001	11.34 (9.22–13.95)	<.001
<b>Region</b>				
China except Hubei	1		1	
Hubei except Wuhan	3.43 (2.83–4.17)	<.001	2.64 (2.11–3.30)	<.001
Wuhan only	5.61 (4.67–6.73)	<.001	6.35 (5.04–8.00)	<.001

Abbreviations: CI, confidence interval; COVID-19, coronavirus disease 2019; HR, hazard ratio.  
<sup>a</sup>P values < .05 were considered statistically significant.

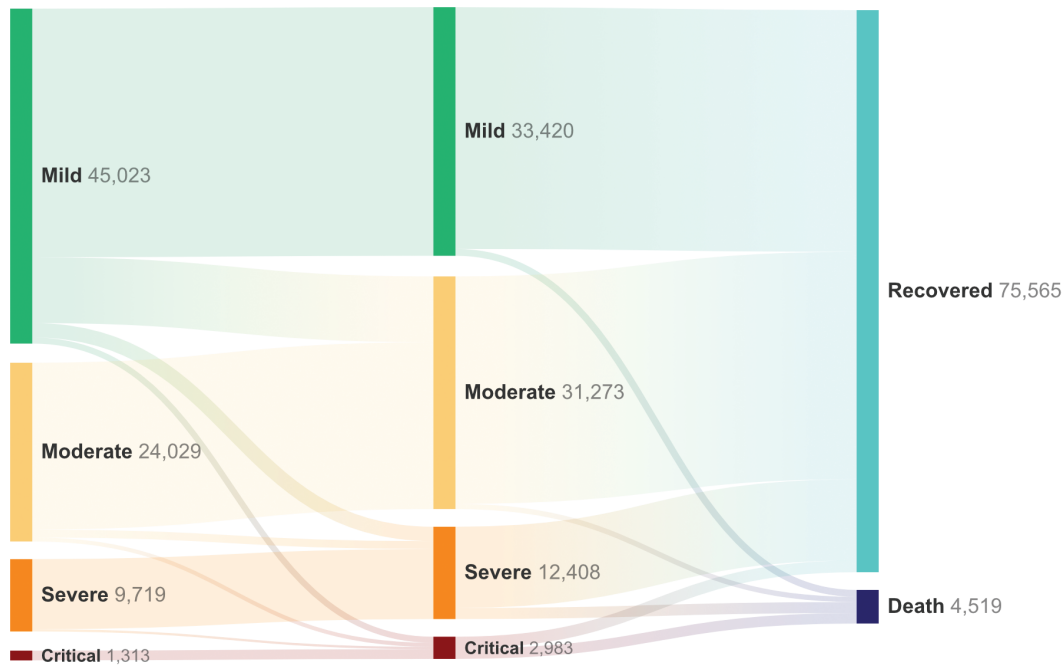
decreased gradually and averaged 550 from 6 to 29 February. Finally, from 1 March and through 8 April, daily new cases were less than 50.

At the provincial level, [Supplementary Figure 2](#) shows the spread of the epidemic using a time series of heatmaps. While on 31 December only 3 provinces had cases who had begun to

**Table 3. Stratification of COVID-19 Cases, Deaths, and Case-Fatality Ratios by Age and Number of Underlying Diseases in China**

Age, y	No. of Underlying Diseases											
	0			1			2			≥3		
	Cases, n	Deaths, n	CFR, %	Cases, n	Deaths, n	CFR, %	Cases, n	Deaths, n	CFR, %	Cases, n	Deaths, n	CFR, %
<30	4182	10	0.2	87	1	1.1	4	0	0	0	0	...
30–39	5680	20	0.4	289	6	2.1	26	2	7.7	5	0	0
40–49	5822	43	0.7	764	12	1.6	112	4	3.6	15	3	20.0
50–59	5580	104	1.9	1445	41	2.8	322	20	6.2	59	6	10.2
60–69	3029	202	6.7	1360	121	8.9	488	55	11.3	135	21	15.6
≥70	1450	264	18.2	978	199	20.3	506	123	24.3	177	56	31.6

Abbreviations: CFR, case-fatality ratio; COVID-19, coronavirus disease 2019.



**Figure 1.** Changes in case severity status (disease progression) over time from initial diagnosis to final outcome in China. Status at initial diagnosis is shown at the left, worst status reached during hospitalization is shown in the center, and final outcome is shown at the right. Case status is color-coded as mild in green, moderate in yellow, severe in orange, and critical in red. Likewise, final outcome is color-coded as recovery in blue and death in purple. Flow between states is indicated and the thickness of bands is proportional to numbers of cases. It is important to note that, in China, all confirmed COVID-19 cases were hospitalized in isolation regardless of the severity of their condition and that all cases remained until either recovery or death, with recovery defined as completely free of all COVID-19–associated symptoms and consistently negative on SARS-CoV-2 PCR testing. Finally, among a total of 80 543 participants, 459 had missing data for case severity status (299 at diagnosis only and 160 at diagnosis and during hospitalization), and therefore, 80 084 participants (99.4%) were included in this analysis. Abbreviations: COVID-19, coronavirus disease 2019; PCR, polymerase chain reaction; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

have symptoms, by 31 January, all 31 provinces had cases and Wuhan and neighboring areas had already been heavily affected. By contrast, relatively little change was observable between 1 March and 8 April. Province-level incidence and mortality are presented in [Supplementary Table 5](#) and, for context, historical overall mortality values for China and for Wuhan are provided in [Supplementary Table 6](#).

## DISCUSSION

This study of all 80 543 confirmed COVID-19 cases in China through 8 April 2020 is the first to provide an epidemiological description of a complete cycle from initial outbreak through rapid epidemic expansion to achievement of long-term control at a national level. The main result was the independent nature of 4 important risk factors for death: older age, greater numbers of underlying diseases, worse case severity at diagnosis, and close proximity to the outbreak epicenter. This important finding indicates that even otherwise healthy older adults and even young adults but with pre-existing disease are at elevated risk of death from COVID-19, which has implications for prevention and vaccine-prioritization strategies. Moreover, it highlights the deleterious consequences of late diagnosis and overwhelmed health systems.

The CFRs were higher in our cohort compared with the prior report of cases up to 11 February 2020, [6] both overall (5.6% vs 2.3%, respectively) and when disaggregated, for 2 reasons. First, we included follow-up time sufficient to ensure that all cases achieved either recovery or death, whereas the prior paper did not. The prior paper's data cutoff occurred very early in the epidemic and many patients were still hospitalized at the time the data were analyzed. Second, we included approximately 1300 deaths that were not previously reported, most of which occurred during the prior report's study period but were not included.

Up to now, the best available evidence on risk factors for COVID-19 death has come from meta-analyses of mostly small studies from China [23, 24, 26–28]; however, they were limited by heterogeneity, particularly with respect to how age groups were categorized and comorbidities were defined (ie, disease that pre-existed SARS-CoV-2 infection was often not differentiated from disease that emerged during COVID-19 disease progression). This presents challenges in isolating risk factors for death. More recently, a very large study in Mexico reported odds of death to be 7-fold greater for those aged 61–80 years; 12-fold greater odds for those aged more than 80 years; 24–31% greater for those with hypertension, obesity, diabetes, chronic

**Table 4. Regional Distribution of COVID-19 Cases, Deaths, and Case-Fatality Ratios by Age, Case Severity, Underlying Disease, and Epidemic Stage in China**

Characteristics	Region									P <sup>c</sup>
	China Except Hubei			Hubei Except Wuhan			Wuhan Only			
	Cases, <sup>a</sup> n (%)	Deaths, <sup>b</sup> n (%)	CFR, %	Cases, <sup>a</sup> n (%)	Deaths, <sup>b</sup> n (%)	CFR, %	Cases, <sup>a</sup> n (%)	Deaths, <sup>b</sup> n (%)	CFR, %	
Overall	12 958 (16.09)	121 (2.7)	0.93	17 795 (22.09)	643 (14.1)	3.61	49 790 (61.82)	3781 (83.2)	7.59	<.001 <sup>d</sup>
Age, y										
<10	318 (2.45)	0 (0)	0	152 (0.85)	0 (0)	0	466 (0.94)	2 (0.1)	0.43	.66
10–19	375 (2.89)	0 (0)	0	237 (1.33)	0 (0)	0	423 (0.85)	3 (0.1)	0.71	.25
20–29	1591 (12.28)	1 (0.8)	0.06	1733 (9.74)	8 (1.2)	0.46	2777 (5.58)	10 (0.3)	0.36	.095
30–39	2796 (21.58)	3 (2.5)	0.11	3196 (17.96)	18 (2.8)	0.56	6936 (13.93)	49 (1.3)	0.71	.001 <sup>e</sup>
40–49	2814 (21.72)	1 (0.8)	0.04	3825 (21.49)	38 (5.9)	0.99	8157 (16.38)	127 (3.4)	1.56	<.001 <sup>d</sup>
50–59	2547 (19.66)	12 (9.9)	0.47	4369 (24.55)	106 (16.5)	2.43	11 009 (22.11)	438 (11.6)	3.98	<.001 <sup>d</sup>
60–69	1634 (12.61)	31 (25.6)	1.90	2643 (14.85)	194 (30.2)	7.34	11 879 (23.86)	1100 (29.1)	9.26	<.001 <sup>d</sup>
70–79	641 (0.49)	39 (32.2)	6.08	1276 (7.17)	193 (30.0)	15.13	5582 (11.21)	1156 (30.6)	20.71	<.001 <sup>d</sup>
≥80	242 (1.87)	34 (28.1)	14.05	364 (2.05)	86 (13.4)	23.63	2561 (5.14)	896 (23.7)	34.99	<.001 <sup>d</sup>
Underlying disease										
Missing	1754 (13.54)	16 (13.2)	0.91	576 (3.24)	17 (2.6)	2.95	43 893 (88.16)	3124 (82.6)	7.12	...
No	8427 (65.03)	35 (28.9)	0.42	13 087 (73.54)	287 (44.6)	2.19	4229 (8.49)	321 (8.5)	7.59	<.001 <sup>d</sup>
Yes	2777 (21.43)	70 (57.9)	2.52	4132 (23.22)	339 (52.7)	8.20	1668 (3.35)	336 (8.9)	20.14	<.001 <sup>d</sup>
One disease	1500 (54.0)	35 (50.0)	2.33	2536 (61.4)	186 (54.9)	7.33	887 (53.2)	159 (47.3)	17.93	<.001 <sup>d</sup>
Two diseases	441 (15.9)	19 (27.1)	4.31	706 (17.1)	93 (27.4)	13.17	311 (18.6)	92 (27.4)	29.58	<.001 <sup>d</sup>
Three or more diseases	117 (4.2)	11 (15.7)	9.40	170 (4.1)	39 (11.5)	22.94	104 (6.2)	36 (10.7)	34.62	<.001 <sup>e</sup>
Missing	719 (25.9)	5 (7.1)	0.70	720 (17.4)	21 (6.2)	2.92	366 (21.9)	49 (14.6)	13.39	...
Case severity										
Mild	2908 (22.44)	1 (0.8)	0.03	6475 (36.39)	9 (1.4)	0.14	24 416 (49.04)	935 (24.7)	3.83	<.001 <sup>f</sup>
Moderate	8155 (62.93)	5 (4.1)	0.06	8190 (46.02)	17 (2.6)	0.21	15 202 (30.53)	699 (18.5)	4.60	<.001 <sup>d</sup>
Severe	1401 (10.81)	26 (21.5)	1.86	2105 (11.83)	123 (19.1)	5.84	8488 (17.05)	1351 (35.7)	15.92	<.001 <sup>d</sup>
Critical	485 (3.74)	89 (73.6)	18.39	992 (5.57)	494 (76.8)	49.80	1308 (2.63)	770 (20.4)	58.87	<.001 <sup>d</sup>
Missing	10 (0.08)	0 (0)	0	33 (0.19)	0 (0)	0	376 (0.76)	26 (0.7)	6.91	...
Epidemic stage										
Early (8–31 December)	2 (0.02)	0 (0)	0	3 (0.02)	1 (0.2)	33.33	197 (0.40)	31 (0.8)	15.74	.58
Rise (1–22 January)	2452 (18.92)	35 (28.9)	1.43	3041 (17.09)	160 (24.9)	5.26	8612 (17.30)	1008 (26.7)	11.70	<.001 <sup>d</sup>
Plateau (23 January–5 February)	8414 (64.93)	70 (57.9)	0.83	11 635 (65.38)	403 (62.7)	3.46	27 134 (54.50)	2099 (55.5)	7.74	<.001 <sup>d</sup>
Decline (6–29 February)	2064 (15.93)	16 (13.2)	0.78	3112 (17.49)	79 (12.3)	2.54	13 627 (27.37)	636 (16.8)	4.67	<.001 <sup>d</sup>
Late (1 March–8 April)	26 (0.20)	0 (0)	0	4 (0.02)	0 (0)	0	220 (0.44)	7 (0.2)	3.18	1.0

Abbreviations: CFR, case-fatality ratio; COVID-19, coronavirus disease 2019.

<sup>a</sup>Proportions for cases were calculated using all cases (N = 80 543) as the denominator for overall and regional total cases (n = 12 958, n = 17 795, n = 49 790) as the denominator for all other variable categories in each region, respectively, except for subcategories under "Yes" for the Underlying disease variable, which use total cases with underlying disease (n = 2777, n = 4132, n = 1668) in each of the regions, respectively, as the denominator for each different number of underlying diseases.

<sup>b</sup>Proportions for deaths were calculated using all deaths (N = 4545) as the denominator for overall and regional total deaths (n = 121, n = 643, n = 3781) as the denominator for all other variable categories in each region, respectively, except for subcategories under "Yes" for the Underlying disease variable, which use total deaths with underlying disease (n = 70, n = 339, n = 336) in each of the regions, respectively, as the denominator for each different number of underlying diseases.

<sup>c</sup>P values were calculated for all groups using Pearson chi-square except for age group <10 years and 10–19 years and epidemic stage Early and Late, which were calculated using Fisher's exact test.

<sup>d</sup>Statistically significant differences were found between any 2 out of the 3 regions (China except Hubei, Hubei except Wuhan, Wuhan only).

<sup>e</sup>Statistically significant differences were found in 2 groups: China except Hubei vs Hubei except Wuhan, China except Hubei vs Wuhan only.

<sup>f</sup>Statistically significant differences were found in 2 groups: China except Hubei vs Wuhan only, Hubei except Wuhan vs Wuhan only.

obstructive pulmonary disease, or immunosuppression; and 85% greater for those with chronic kidney disease. However, the relationship between age and underlying disease was not explored [29]. Our results provide a clearer picture of groups

more vulnerable to death, facilitating targeted prevention intervention and vaccine distribution.

We also found that, although overall incidence was 5.79 per 100 000 and mortality was 0.33 per 100 000, these national-level

measures hid dramatic geographic and demographic differences, such as very high incidence and mortality in Hubei and in Wuhan among older and urban adults. Overall, for the 81% of cases that were mild or moderate, CFR was under 3%, but the 15% who were severe faced a CFR of 12.5% and the 3.5% of cases who were critical had a CFR of nearly 50%. Yet, in Wuhan, where the health system quickly became overwhelmed, 17% of cases were severe and had a CFR of 16% and nearly 3% were critical with a CFR of almost 60%.

Even for countries that, like China, took aggressive action to try to contain the infection and break the chain of human transmission [2], enormous numbers of people fell ill and died. For instance, in Italy, the first Western nation to be affected by COVID-19, the number of cumulative cases increased 100 000-fold and deaths increased 429-fold in March 2020 alone [30]. Still lacking therapeutics and vaccines, the only defenses were traditional public health methods [2]. China provides not only an example of how decisive action to thoroughly implement these countermeasures can bring COVID-19 under control [2] but also a cautionary tale of how quickly a new emerging infectious disease can overwhelm a health system. Evidence of China's struggles during the January and February 2020 "surge" of cases is observable in the data. Higher incidence values were to be expected in Wuhan and outlying areas in Hubei, but these areas also suffered higher mortality and CFRs compared with elsewhere in China, which reflects the enormous strain on the local health system during the surge. These areas also suffered reporting delays (ie, 325 survivors and 1290 deaths never before included in other studies) and incompleteness of records due to stresses on the public health system. These issues have also been observed, for example, in Italy where the CFR in Milan was much greater than in some parts of southern Italy, and in the United States the CFR in New York was greater than in other states like Minnesota. Many national, state, and city public health systems have also experienced disruption to timely case and death reporting due to the overwhelming stress of case surges.

However, CFR was not only elevated because of health system strain. Delayed testing because of insufficient capacity and inexperience with a new disease meant late presentation and suboptimal clinical management early in the pandemic. Case-fatality ratios were elevated early in the outbreak in all regions of China compared with later in the outbreak because diagnosis and clinical management improved over time. This has been observed in many nations as they began dealing with COVID-19 for the first time. As physicians and nurses learned from their own experience (and that of others) how to care for patients with COVID-19 and as best practices were developed and more broadly shared, CFRs naturally declined.

### Limitations

Major strengths of this study were its nationwide scope, very large size, and complete inclusion of all cases in China's

COVID-19 epidemic wave followed for their entire clinical course. Nevertheless, our study had some important limitations. First, because of China's official case definition, characteristics of asymptomatic infections were not examined. Second, the data were compiled under the crushing pressure of a rapidly spreading epidemic. This limited our ability to explore other variables (eg, smoking, obesity) and outcomes (eg, rehospitalizations, post-acute phase complications), which are now of interest. Third, this also meant that some records were incomplete or may have contained errors. For example, 57% of records contained no information on underlying conditions, requiring our analyses on this variable to be limited to a subset of cases. Fourth, clinically diagnosed cases during the early outbreak when testing was scarce may have included some misdiagnoses, thereby exaggerating the total case count. Finally, although all confirmed cases were immediately hospitalized, and therefore all deaths among them occurred in the hospital and were recorded as COVID-19 deaths regardless of other contributing factors, there remained a small possibility of under-ascertainment of COVID-19 deaths.

### Conclusions

In conclusion, this study provides new, high-quality evidence of the epidemiology of COVID-19 from a very large nationwide cohort that includes complete follow-up of all cases in the entire first wave of COVID-19 from the first case to epidemic control in China. It further demonstrates how quickly COVID-19 can affect large numbers of people over expansive geographical areas, even with early and aggressive action taken to thoroughly implement countermeasures. Most important, it supports and extends prior observations of poorer outcomes for individuals who are older, have underlying conditions, and are diagnosed at later disease stages or where healthcare services are overwhelmed. These results provide important evidence for prevention intervention, vaccine prioritization, and health system planning that can be applied globally.

### Supplementary Data

Supplementary materials are available at *Clinical Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

### Notes

**Author Contributions.** Z. W. and Z. F. had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Y. Z., W. L., Z. W., and Z. F. Acquisition, analysis, or interpretation of data: Y. Z., W. L., Q. L., X. W., R. R., D. L., J. M. M., and Z. W. Drafting of the manuscript: Y. Z., W. L., and Z. W. Critical revision of the manuscript for important intellectual content: All authors. Statistical analysis: W. L. and X. W. Obtained funding: Z. W. Administrative, technical, or material support: Y. Z., G. F. G., Z. W., and Z. F. Supervision: Z. W. and Z. F.



**Disclaimer.** The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication. The opinions expressed herein reflect the collective views of the coauthors and do not necessarily represent the official position of the authors' affiliated institutions.

**Financial support.** This work was supported by the Ministry of Science and Technology of the People's Republic of China (grant number 2020YFC0846300).

**Potential conflicts of interest.** The authors: No reported conflicts of interest. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest.

## References

1. World Health Organization. Report of the WHO-China Joint Mission on coronavirus disease 2019 (COVID-19). Available at: [https://www.who.int/publications-detail/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-\(covid-19\)](https://www.who.int/publications-detail/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-(covid-19)). Accessed 22 April 2021.
2. Zhou L, Wu Z, Li Z, et al. One hundred days of COVID-19 prevention and control in China. *Clin Infect Dis* **2021**; 72:332-9.
3. The 2019-nCoV Outbreak Joint Field Epidemiology Investigation Team; Li Q. Notes from the field: an outbreak of NCIP (2019-nCoV) infection in China—Wuhan, Hubei Province, 2019-2020. *China CDC Wkly* **2020**; 2:79-80.
4. Wang D, Hu B, Hu C, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA* **2020**; 323:1061-9.
5. Xu XW, Wu XX, Jiang XG, et al. 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**; 368:m606.
6. Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19)—China, 2020. *China CDC Wkly* **2020**; 2:113-22.
7. Guan WJ, Ni ZY, Hu Y, et al. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med* **2020**; 382:1708-20.
8. Wu J, Liu J, Zhao X, et al. Clinical characteristics of imported cases of COVID-19 in Jiangsu Province: a multicenter descriptive study. *Clin Infect Dis* **2020**; 71:706-12.
9. Wang R, Pan M, Zhang X, et al. Epidemiological and clinical features of 125 hospitalized patients with COVID-19 in Fuyang, Anhui, China. *Int J Infect Dis* **2020**; 95:421-8.
10. Zhao XY, Xu XX, Yin HS, et al. Clinical characteristics of patients with 2019 coronavirus disease in a non-Wuhan area of Hubei Province, China: a retrospective study. *BMC Infect Dis* **2020**; 20:311.
11. Lei Z, Cao H, Jie Y, et al. A cross-sectional comparison of epidemiological and clinical features of patients with coronavirus disease (COVID-19) in Wuhan and outside Wuhan, China. *Travel Med Infect Dis* **2020**; 35:101664.
12. Zhang G, Hu C, Luo L, et al. Clinical features and short-term outcomes of 221 patients with COVID-19 in Wuhan, China. *J Clin Virol* **2020**; 127:104364.
13. Huang Q, Deng X, Li Y, et al. Clinical characteristics and drug therapies in patients with the common-type coronavirus disease 2019 in Hunan, China. *Int J Clin Pharm* **2020**; 42:837-45.
14. Nie X, Fan L, Mu G, et al. Epidemiological characteristics and incubation period of 7015 confirmed cases with coronavirus disease 2019 outside Hubei Province in China. *J Infect Dis* **2020**; 222:26-33.
15. Duan X, Guo X, Qiang J. A retrospective study of the initial 25 COVID-19 patients in Luoyang, China. *Jpn J Radiol* **2020**; 38:683-90.
16. Lv Z, Cheng S, Le J, et al. Clinical characteristics and co-infections of 354 hospitalized patients with COVID-19 in Wuhan, China: a retrospective cohort study. *Microbes Infect* **2020**; 22:195-9.
17. Liang WH, Guan WJ, Li CC, et al. 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**; 55:2000562.
18. Cao Z, Li T, Liang L, et al. Clinical characteristics of coronavirus disease 2019 patients in Beijing, China. *PLoS One* **2020**; 15:e0234764.
19. Qian GQ, Yang NB, Ding F, et al. Epidemiologic and clinical characteristics of 91 hospitalized patients with COVID-19 in Zhejiang, China: a retrospective, multi-centre case series. *QJM* **2020**; 113:474-81.
20. Sun Y, Tian L, Du X, Wang H, Li Y, Wu R. Epidemiological and clinical characteristics of a familial cluster of COVID-19. *Epidemiol Infect* **2020**; 148: e145.
21. Zhang SY, Lian JS, Hu JH, et al. Clinical characteristics of different subtypes and risk factors for the severity of illness in patients with COVID-19 in Zhejiang, China. *Infect Dis Poverty* **2020**; 9:85.
22. Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. *JAMA* **2020**; 323:1239-42.
23. Romero Starke K, Peterit-Haack G, Schubert M, et al. The age-related risk of severe outcomes due to COVID-19 infection: a rapid review, meta-analysis, and meta-regression. *Int J Environ Res Public Health* **2020**; 17:E5974.
24. Moula AI, Micali LR, Matteucci F, et al. Quantification of death risk in relation to sex, pre-existing cardiovascular diseases and risk factors in COVID-19 patients: let's take stock and see where we are. *J Clin Med* **2020**; 9:E2685.
25. Yang W, Li Z, Lan Y, et al. A nationwide web-based automated system for outbreak early detection and rapid response in China. *Western Pac Surveill Response J* **2011**; 2:10-5.
26. Zheng Z, Peng F, Xu B, et al. Risk factors of critical and mortal COVID-19 cases: a systematic literature review and meta-analysis. *J Infect* **2020**; 81:e16-25.
27. Parohan M, Yaghoubi S, Seraji A, Javanbakht MH, Sarraf P, Djalali M. Risk factors for mortality in patients with coronavirus disease 2019 (COVID-19) infection: a systematic review and meta-analysis of observational studies. *Aging Male* **2020**; 23:1416-24.
28. Espinosa OA, Zanetti ADS, Antunes EF, Longhi FG, de Matos TA, Battaglini PF. Prevalence of comorbidities in patients and mortality cases affected by SARS-CoV2: a systematic review and meta-analysis. *Rev Inst Med Trop Sao Paulo* **2020**; 62: e43.
29. Parra-Bracamonte GM, Lopez-Villalobos N, Parra-Bracamonte FE. Clinical characteristics and risk factors for mortality of patients with COVID-19 in a large data set from Mexico. *Ann Epidemiol* **2020**; 52:93-8.
30. La Maestra S, Abbondandolo A, De Flora S. Epidemiological trends of COVID-19 epidemic in Italy over March 2020: From 1000 to 100000 cases. *J Med Virol* **2020**; 92:1956-61.