



Original article

Epidemiological and clinical characteristics of 671 COVID-19 patients in Henan Province, China

Yifei Nie,^{1†‡} Jitian Li,^{2†} Xueyong Huang,^{1†} Wanshen Guo,^{1†}
Xiaobai Zhang,³ Yan Ma,² Haifeng Wang,¹ Muge Qi,¹ Xiaoyan Tang,¹
Xiaoqing Shen¹ and Xiaofeng Dai^{4*‡}

¹Henan Center for Disease Control and Prevention/Henan Key Laboratory of Pathogenic Microorganisms, Zhengzhou, China, ²Henan Provincial Orthopedic Institute, Zhengzhou, China, ³The First Affiliated Hospital of Henan University of Science and Technology, Luoyang, China and ⁴Wuxi School of Medicine, Jiangnan University, Wuxi, China

[†]These authors contributed equally to this work.

[‡]Joint first authors.

*Corresponding author. Wuxi School of Medicine, Jiangnan University, Lihu Avenue 1800, Wuxi 214122, China.

E-mail: xiaofeng.dai@jiangnan.edu.cn

Editorial decision 16 April 2020; Accepted 20 April 2020

Abstract

Background: Despite many reports on the characteristics of coronavirus disease 2019 (COVID-19) in Wuhan, China, relatively little is known about the transmission features of COVID-19 outside Wuhan, especially at the provincial level.

Methods: We collected epidemiological, demographic, clinical, laboratory, radiological and occupation information, along with contact history, of 671 patients with laboratory-confirmed COVID-19 reported from January 23 to February 5, 2020, in Henan province, China. We described characteristics of these cases, compared the diagnostic accuracy and features of blood testing, computed tomography (CT) scans and X-rays, and analysed SARS-CoV-2 transmission sources and patients' occupations in Henan province.

Results: The mean age of patients in this case series was 43 years, 56.2% were male and 22.4% had coexisting medical disorders. The death rate was 0.3%. Fourteen patients did not show any symptoms. Lymphocyte percentage was associated with disease severity ($\chi^2 = 6.71$, $P = 0.035$) but had a large variation in each sample group. The mean time from illness onset to diagnosis was 5.6 days. A total of 330 patients had ever lived in or visited Wuhan, 150 had contact with confirmed cases, 323 had been to a hospital and 119 had been to a wet market. There were 33 patients who did not have a traceable transmission source, with 21.2% of these being farmers and 15.2% being workmen.

Conclusions: Lymphocyte percentage was a sign of severe COVID-19 in general but was not a good diagnostic index. Longer time from illness onset to diagnosis was associated with higher COVID-19 severity, older age, higher likelihood of having coexisting cardiovascular diseases including hypertension, and being male. Farming was found to be a high-risk occupation in Henan province, China.

Key words: SARS-CoV-2, COVID-19, Henan province

Key Messages

- This study presents the epidemiological and clinical characteristics of a patient cohort of 671 confirmed COVID-19 cases in Henan province, the largest province in terms of population size and ranked as the third most severely affected province in China.
- To the best of our knowledge, our study is the largest report of COVID-19 clinical characteristics in China outside Wuhan at the provincial level that offers direct evidence on SARS-CoV-2 transmission manifestations.
- Insights gained from this study could guide clinical practice on COVID-19 diagnosis and pathogen transmission control at both governmental and individual levels.

Introduction

Ever since the outbreak of novel coronavirus disease (COVID-19) infection in Wuhan, China, this disease has rapidly spread from Wuhan to other areas in China and globally. It has now affected >200 countries, with the epicenter shifted from Asia to Europe and North America. The pathological agent, severe acute respiratory syndrome (SARS) coronavirus 2 (SARS-CoV-2), was identified on 3 January 2020, and confirmed (through full genome sequencing and phylogenetic analysis) as a distinct clade from the beta-coronaviruses that had caused SARS and Middle East respiratory syndrome (MERS) in 2003 and 2012, respectively.¹ Bats were considered to be the original carriers of SARS-CoV-2, and pangolins were recently identified as the intermediate host. Human-to-human transmission is evident.² Many studies have reported the epidemiological and clinical features of COVID-19 in Wuhan,^{1–4} whereas relatively little has been reported on COVID-19 outside Wuhan in China, except for a report on 13 confirmed cases admitted to hospitals in Beijing.⁵

Here, we present a case series that encompasses 671 patients with COVID-19 in Henan province, which has the largest population (>10 million permanent residents) and was ranked as the third most affected region in China. The global picture on the COVID-19 situation in China was recently published, covering 31 provinces and including 1099 clinical cases.⁴ Yet the present study is, to our knowledge, the largest report on COVID-19 at the provincial level outside Wuhan in China to study the manifestations of SARS-CoV-2 in its transmitted region.

Methods

Data collection

Data on all consecutive patients with confirmed COVID-19 infection in Henan province from 23 January to 5 February 2020 were collected from 279 hospitals. Oral consent was

obtained from patients. All 671 patients with COVID-19 enrolled in this study were diagnosed according to World Health Organization interim guidance.⁶ Clinical outcomes (i.e. discharges, mortality) and disease severity were monitored up to 5 February 2020, the final date of follow-up.

The medical records of patients were analysed by the research team of the Center for Disease Control (CDC) of Henan Province. Epidemiological, clinical, laboratory and radiological characteristics and outcomes data were obtained with data collection forms from electronic medical records. Information recorded included demographic data, medical history, exposure history, underlying comorbidities, laboratory findings, X-rays and chest computed tomography (CT) scans. The date of disease onset was defined as the day when the first symptoms were noticed. Laboratory test values, X-rays and chest CT scans were collected.

Cardiovascular diseases are a group of disorders of the heart and blood vessels, including cerebrovascular disease and coronary heart disease, peripheral arterial disease and rheumatic heart disease. Hypertension was defined according to the American Heart Association expert consensus.⁷ Respiratory diseases were defined as many disorders affecting the lungs, including asthma, chronic bronchitis, bronchiectasis, chronic obstructive pulmonary disease and tuberculosis. Diabetes was defined according to the American Diabetes Association's diagnostic criteria.⁸

Laboratory testing

Throat or nose swab samples were collected for extracting SARS-CoV-2 RNA from patients suspected of having SARS-CoV-2 infection. After sample collection, total RNA was extracted within 2 h using a respiratory sample RNA isolation kit (Shanghai bio-germ Medical Technology Co Ltd, Catlog number: ZC-HX-201–2) and was tested using a SARS-CoV-2 nucleic acid real-time reverse transcription

polymerase chain reaction (RT-PCR) detection kit, following the manufacturer's protocol (Shanghai bio-germ Medical Technology Co Ltd) in biosafety level 2 facilities at the Henan (provincial) CDC. The primers and sequences targeting open reading frame (ORF) 1ab were: forward primer CCCTGTGGGTTTTACACTTAA, reverse primer ACGATTGTGCATCAGCTGA, and probe 5'-FAM-CCGTCTGCGGTATGTGAAAGGTTATGG-BHQ1-3'. The primers and sequences targeting nucleocapsid protein (N) were: forward primer GGGGAACCTTCTCCTGCTAGAAT, reverse primer CAGACATTTTGCTCTCAAGCTG and probe 5'-FAM-TTGCTGCTGCTTGACAGATT-TAMRA-3'. Conditions for the amplifications were 50°C for 10 min, 95°C for 5 min, followed by 40 cycles of 95°C for 10 s and 55°C for 40 s. A cycle threshold value (Ct-value) <37 was defined as a positive test, and a Ct-value ≥ 40 was defined as a negative test. A medium load, defined as a Ct-value of 37 to <40, required confirmation by retesting. If two targets (ORF 1a or 1b, nucleocapsid protein) tested positive by specific real-time RT-PCR in the initial or second test, the case would be considered to be laboratory-confirmed. These diagnostic criteria were based on the recommendation of the National Institute for Viral Disease Control and Prevention (China) (http://ivdc.chinacdc.cn/kyjz/202001/t20200121_211337.html).

Statistical analysis

Continuous variables are described using mean, standard deviation (SD), median and interquartile range (IQR) values, and categorical variables as frequency and percentage. Normality testing was performed using the Shapiro–Wilk test. Means for continuous variables were compared using independent group tests when the data were normally distributed; otherwise, Mann–Whitney *U* test and Kruskal–Wallis *H* test were used (adjusted by Bonferroni correction). Proportions for categorical variables were compared using the χ^2 test or Fisher exact test. The correlation of the two variables was compared using the Spearman rank correlation. The influencing variables for the severity and non-severity of COVID-19 were analysed using a logistic regression model. All statistical analyses were performed using SPSS (Statistical Package for the Social Sciences) version 23.0 software (SPSS Inc).

Ethics approval

This study was approved by the ethics commissions of the Henan Center for Disease Control and Prevention, with a waiver of written consent.

Results

The mean and median ages of the patients were 43 ± 15.09 and 44 (31–53) years, respectively; 56.2% were male and 22.4% (151) had coexisting medical disorders or a previous record. Seventy patients (10.4%) had cardiovascular diseases, of whom 60 (85.7%) had hypertension, 12 (1.8%) had diabetes, 17 (2.5%) had respiratory diseases, 19 (2.8%) had both cardiovascular disease and diabetes, and two (0.3%) had both respiratory disease and diabetes. Thirty-three patients (4.9%) had other complications, such as hepatitis, nephrosis, gastritis and trauma (Table 1). The distribution of disease severity did not follow a normal distribution (Shapiro–Wilk statistic = 0.751, $P = 3.05 \times 10^{-30}$), wherein 14 patients (2.1%) were symptom-free, 163 (24.3%) had light symptoms, 420 (62.6%) were normal, 66 (9.8%) had severe disease and six patients (0.9%) were at the terminal stage (Table 1). Sixteen patients (2.4%) were admitted to the intensive care unit (ICU). A total of 615 patients (91.7%) remained in hospital, 54 patients (8%) were discharged and two patients (0.3%) had died by February 5 (Table 1). The mortality rate was 0.3%; much lower than that reported for Wuhan (4.3%).³

In this case series, 330 patients had ever lived in (40.7%) or visited (8.5%) Wuhan, accounting for up to 49.2% of the cohort. Among patients without a direct association with Wuhan, 150 (22.4%) had contact with confirmed cases, 131 (19.5%) had contact with suspected cases, 253 (37.7%) had contact with individuals who had ever lived in or visited Wuhan, 132 (19.7%) were close to one or several cluster outbreaks, 323 (48.1%) (including 17 medical workers) had visited a hospital and 119 (17.7%) had been to a wet market (Table 1). There were 33 patients (4.9%) with no traceable contact history, according to the design of our questionnaire (Table 1).

The frequency of COVID-19 infection by occupation in patients in Henan province was highest in farmers (89, 13.3%), followed by workmen (81, 12.1%) (Figure 1A). Among the 33 patients without a known transmission source, the same trend was seen (i.e. farmers: 7, 21.2%; followed by workmen: 5, 15.2%) (Figure 1B). There were 17 patients who were medical workers, 7 were children (aged ≤ 8 years) and 3 were pregnant.

The average time from illness onset to diagnosis was 5.6 days, ranging from 0 to 32 days. Among the 671 individuals at admission, 239 received X-rays, with 83 showing clear pneumonia features (34.7%); 562 had chest CT scans, with 539 exhibiting pneumonia characteristics (95.9%); and 606 patients received routine blood tests, with 606, 596, 597 and 593 patients having available information on leucocyte count, lymphocyte count, lymphocyte percentage and neutrophil percentage, respectively, and 598 having available information on all routine blood

Table 1. Baseline characteristics, contact histories, diagnosis and clinical outcomes of 671 patients with COVID-19 in Henan Province (23 January to 5 February 2020)

Clinical characteristics	Statistics ^a
Age (years)	
Mean (SD)	43±15.09
Median (IQR)	44 (31–53)
Range	0.5–88
Sex	
Male	377 (56.2%)
Female	294 (43.8%)
Coexisting disorder	
Cardiovascular diseases	70 (10.4%)
Hypertension	60
Diabetes	12 (1.8%)
Respiratory diseases	17 (2.5%)
Cardiovascular diseases and diabetes	19 (2.8%)
Respiratory diseases and diabetes	2 (0.3%)
Other	33 (4.9%)
None	518 (77.2%)
Disease severity	
Symptom-free	14 (2.1%)
Light	163 (24.3%)
Normal	420 (62.6%)
Severe	66 (9.8%)
Terminal	6 (0.9%)
NA ^b	2 (0.3%)
Admission to ICU	
Yes	16 (2.4%)
No	653 (97.3%)
NA	2 (0.3%)
Clinical outcome	
Remained in hospital	615 (91.7%)
Discharged	54 (8.0%)
Died	2 (0.3%)
Animal–human transmission source contact	
Direct contact	330 (49.2%)
Live	273 (82.7%)
Visit	57 (17.3%)
Indirect contact	306 (45.6%)
Contact with confirmed cases	
Yes	150 (22.4%)
No	516 (76.9%)
NA	5 (0.7%)
Contact with suspected cases	
Yes	131 (19.5%)
No	461 (68.7%)
NA	79 (11.8%)
Contact with individuals who had been to Wuhan	
Yes	253 (37.7%)
No	339 (50.5%)
NA	79 (11.8%)
Close to cluster outbreak	
Yes	132 (19.7%)
No	492 (73.3%)
NA	47 (7.0%)

(Continued)

Table 1. Continued

Clinical characteristics	Statistics ^a
Visited hospital	
Yes	323 (48.1%)
No	269 (40.1%)
NA	79 (11.8%)
Visited wet market	
Yes	119 (17.7%)
No	534 (79.6%)
NA	18 (2.7%)
No contact	33 (4.9%)
NA	2 (0.3%)
Days from illness onset to diagnosis	
Mean (SD)	5.6±3.7
Range	0–32
X-ray with pneumonia features	
Yes	83 (12.4%)
No	156 (23.2%)
Untested	353 (52.6%)
NA	79 (11.8%)
CT with pneumonia features	
Yes	539 (80.3%)
No	23 (3.4%)
Untested	30 (4.5%)
NA	79 (11.8%)
Blood routine test	
Leucocytes ($\times 10^9/L$; normal range 3.5–9.5)	4.98 (3.83–6.42)
Increased	30 (4.5%)
Normal	473 (70.5%)
Decreased	103 (15.4%)
NA	65 (9.7%)
Lymphocytes ($\times 10^9/L$; normal range 1.1–3.2)	1.18 (0.86–1.74)
Increased	42 (6.3%)
Normal	292 (43.5%)
Decreased	262 (39.0%)
NA	75 (11.2%)
Lymphocyte percentage (normal range 20–50%)	24.50 (16.25–31.90)
Increased	23 (3.4%)
Normal	356 (53.1%)
Decreased	218 (32.5%)
NA	74 (11.0%)
Neutrophil percentage (normal range 50–70%)	66.10 (56.50–74.85)
Increased	235 (35.0%)
Normal	278 (41.4%)
Decreased	80 (11.9%)
NA	78 (11.6%)

^aStatistics include one of the following: mean \pm SD, range, number, number (percentage), median (IQR), as indicated;

^bNA, not applicable.

test indexes (Table 1). The median leucocyte count was $4.98 \times 10^9/L$ (IQR, 3.83–6.42), lymphocyte count was $1.18 \times 10^9/L$ (IQR, 0.86–1.74), lymphocyte percentage was 24.5% (IQR, 16.25–31.9) and neutrophil percentage was 66.1% (IQR, 56.5–74.85) (Table 1).

By removing patients without COVID-19 severity information or symptoms from the analysis, we collected data on 655 patients to study the sensitivity and accuracy of traditional pneumonia diagnostic approaches. Among X-ray, CT scan and routine blood test (including leucocyte and

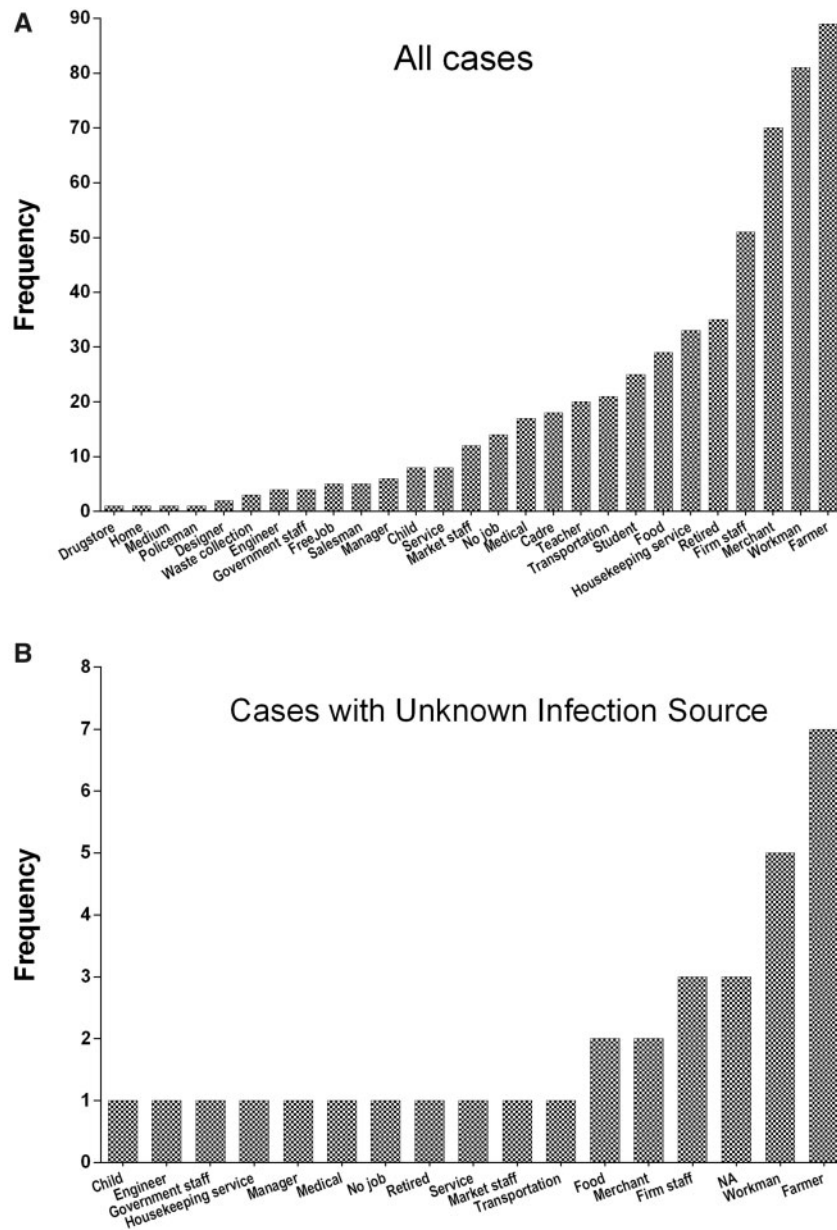


Figure 1 Occupation distribution of confirmed COVID-19 cases. (A) All 669 COVID-19 patients. (B) 33 cases with unknown transmission source. Two cases without available information on occupation were excluded.

lymphocyte counts, and lymphocyte and neutrophil percentages), lymphocyte percentage played a significant role in differentiating patients' COVID-19 infection severity ($\chi^2 = 6.71$, $P = 0.035$, Table 2), followed by neutrophil percentage ($\chi^2 = 4.943$, $P = 0.084$, Table 2). In general, lymphocyte percentage and neutrophil percentage decreased and increased, respectively, with more severe disease. The IQR of lymphocyte percentage largely overlapped with the lower 50% percentile of the range typically considered healthy (i.e. COVID-19, 13.75–32.91% vs healthy, 20–50%). The IQR of neutrophil percentage among infected individuals spanned and encompassed the normal range (i.e. COVID-19, 55.48–81.8% vs healthy,

50–70%). The IQR of leucocyte count fell in the normal range. Lymphocyte count was, in general, decreased in COVID-19 infected individuals, but its IQR overlapped with the normal range (i.e. COVID-19, $0.84\text{--}1.83 \times 10^9/\text{L}$ vs healthy, $1.1\text{--}3.2 \times 10^9/\text{L}$). CT scans showed a higher detection rate than did X-ray (96 vs 35%) (Table 2). Thirteen cases showing scattered opacities on X-ray were not detected from CT scans, and 147 cases showing bilateral ground glass opacities on CT scans did not show typical pneumonia features on X-ray. There were 5 cases undetected from either X-ray or CT scans. These 5 patients were all female, except for a 16-year-old boy, and all had light or normal severity of infection.

Table 2. Comparisons of canonical pneumonia detection methods in the accuracy of COVID-19 severity diagnosis

Diagnostic approach	Parameter	Number	Light	Normal	Severe	χ^2	<i>P</i>
X-ray			163	420	72		
Yes	Number (percentage)	82	22 (31.4%)	47 (33.8%)	13 (46.4%)	2.08	0.353
No	Number (percentage)	155	48 (68.6%)	92 (66.2%)	15 (53.6%)		
Untested/NA ^a	Removed	418	93	281	44		
CT							
Yes	Number (percentage)	535	150 (97.4%)	333 (94.9%)	52 (100%)	4.166	0.125
No	Number (percentage)	22	4 (2.6%)	18 (5.1%)	0		
Untested/NA	Removed	98	9	69	20		
Routine blood test							
Leucocyte count ($\times 10^9/L$, normal range 3.5–9.5)	Median (IQR)	606	5.19 (3.61–6.66)	4.79 (3.86–6.13)	5.52 (4.03–6.79)	3.334	0.189
	Range		1.56–16.5	1.11–16.07	2.3–15.48		
Lymphocyte count ($\times 10^9/L$, normal range 1.1–3.2)	Median (IQR)	596	1.22 (0.86–1.83)	1.18 (0.86–1.74)	1.16 (0.84–1.63)	1.051	0.591
	Range		0.02–70.3	0.07–77	0.36–78		
Lymphocyte percentage (normal range 20–50%)	Median (IQR)	597	26.25 (16.83–32.91)	24.75 (16.70–31.70)	18.70 (13.75–28.21)	6.71	0.035
	Range		0.4–85.1	0.2–83.5	0.2–80.28		
Neutrophil percentage (normal range 50–70%)	Median (IQR)	593	64.60 (55.48–74.50)	66.01 (56.70–74.40)	71.70 (57.40–81.80)	4.943	0.084
	Range		1.45–91.7	1.1–90.9	0.74–93.04		

^aNA, Not applicable.

Table 3. Spearman correlation of days from illness onset to diagnosis with comorbidities, age and COVID-19 severity

	Comorbidities ^a	Age	Sex	COVID-19 severity
Correlation	0.089*	0.086*	−0.075	0.186**
95% CI	0.014 to 0.171	0.010 to 0.163	−0.149 to −0.001	0.110 to 0.259
<i>P</i>	0.022	0.027	0.055	2×10^{-6}

^aComorbidities are classified as cardiovascular diseases including hypertension (CCH), non-CCH and no disorder.

**P* < 0.05 (two-tails);

***P* < 0.01 (two-tails).

The number of days from illness onset to diagnosis was associated with COVID-19 infection severity [correlation: 0.186, 95% confidence interval (CI): 0.110–0.259, $P = 2 \times 10^{-6}$, Table 3], wherein patients with severe disease, in general, took a longer time from illness onset to diagnosis (Table 3). The number of days from illness onset to diagnosis was also positively correlated with comorbidities (correlation: 0.089, 95% CI: 0.014–0.171, $P = 0.022$) and age (correlation: 0.086, 95% CI: 0.010–0.163, $P = 0.027$) and was marginally associated with sex (Table 3). Female patients experienced slightly fewer days from illness onset to diagnosis than male patients, with significance (correlation: −0.075, 95% CI: −0.149 to −0.001, $P = 0.055$, Table 3).

By including all epidemiological, clinical and laboratory factors, as well as contact histories, in a univariate logistic regression model explaining patient severity, we identified

five factors that potentially influenced COVID-19 severity. These factors are age, contact with confirmed cases, days from illness onset to diagnosis, lymphocyte percentage and neutrophil percentage. By fitting these into the multivariate logistic regression model, age [$P = 0.003$, odds ratio (OR) = 1.026, 95% CI: 1.009–1.043], contact with confirmed cases ($P = 0.039$, OR = 0.456, 95% CI: 0.213–0.976), days from illness onset to diagnosis ($P = 0.004$, OR = 1.065, 95% CI: 1.003–1.132), lymphocyte percentage ($P = 0.015$, OR = 0.98, 95% CI: 0.957–1.003) and neutrophil percentage ($P = 0.033$, OR = 1.004, 95% CI: 0.989–1.019) were closely associated with disease severity (Table 4). By analysing variables identified from the multivariate logistic regression model and using univariate models separately, age ($P = 0.002$, OR = 1.026, 95% CI: 1.009–1.043), contact with confirmed cases ($P = 0.043$, OR = 0.456, 95% CI: 0.213–0.976), days from illness

Table 4. Statistics of the multivariate logistic regression model for severe versus non-severe COVID-19 as stratified by epidemiological, clinical and laboratory factors, as well as contact histories of 655 patients with COVID-19 in Henan Province (23 January to 5 February 2020)

Clinical characteristics	Non-severe (<i>n</i> = 583)	Severe (<i>n</i> = 72)	<i>z</i> / χ^2 ^a	<i>P</i>	OR (95% CI)
Age (years)					
Median (IQR)	43 (31–52)	48.5 (34–58)	–3.020	0.003	1.026 (1.009–1.043)
Sex					
Male	324 (55.6%)	43 (59.7%)	0.447	0.504	0.844 (0.512–1.389)
Female	259 (44.4%)	29 (40.3%)			
Coexisting disorder					
Cardiovascular diseases	62	8	0.015	0.902	0.809 (0.306–2.142)
Hypertension	55	5	0.477	0.490	0.494 (0.168–1.451)
Diabetes	11	0	0.475	0.491	—
Respiratory diseases	15	2	<0.001	1.000	0.910 (0.191–4.333)
Cardiovascular diseases and diabetes	15	4	1.104	0.293	2.062 (0.605–7.035)
Respiratory diseases and diabetes	2	0	<0.001	1.000	—
Other	26	5	0.413	0.520	1.442 (0.476–4.366)
Animal/human transmission source contact					
Direct	284	39	0.414	0.520	1.062 (0.695–1.622)
Live	238 (83.8%)	29 (74.4%)	2.134	0.144	1.738 (1.152–2.622)
Visit	46 (16.2%)	10 (25.6%)			
Indirect	266	31			
Contact with confirmed cases					
Yes	128 (22.1%)	8 (11.4%)	4.274	0.039	0.456 (0.213–0.976)
No	452 (77.9%)	62 (88.6%)			
NA	3	2			
Contact with suspected cases					
Yes	112 (21.5%)	9 (16.1%)	0.886	0.347	0.701 (0.333–1.474)
No	410 (78.5%)	47 (83.9%)			
NA	61	16			
Contact with individuals who had been to Wuhan					
Yes	221 (42.3%)	21 (37.5%)	0.486	0.486	0.817 (0.463–1.442)
No	301 (57.7%)	35 (62.5%)			
NA	61	16			
Close to cluster outbreak					
Yes	104 (19.2%)	14 (20.9%)	0.111	0.739	1.112 (0.595–2.082)
No	438 (80.8%)	53 (79.1%)			
NA	41	5			
Visited hospital					
Yes	285 (54.6%)	38 (67.9%)	3.607	0.058	1.756 (0.976–3.157)
No	237 (45.4%)	18 (32.1%)			
NA	61	16			
Visited wet market					
Yes	102 (17.9%)	15 (22.1%)	0.692	0.406	1.296 (0.703–2.389)
No	467 (82.1%)	53 (77.9%)			
NA	14	4			
No contact	32	1			
NA	1	1			

(Continued)

Table 4. Continued

Clinical characteristics	Non-severe (<i>n</i> = 583)	Severe (<i>n</i> = 72)	<i>z</i> / χ^2 ^a	<i>P</i>	OR (95% CI)
Days from illness onset to diagnosis					
Median (IQR)	5 (3–8)	6 (4–9)	–2.896	0.004	1.065 (1.003–1.132)
X-ray with pneumonia features					
Yes	69 (33.0%)	13 (46.4%)	1.964	0.161	1.758 (0.793–3.901)
No	140 (67.0%)	15 (53.6%)			
CT with pneumonia features					
Yes	483 (96.7%)	52 (92.9%)	0.367	0.544	—
No	22 (4.2%)	4 (7.1%)			
Blood routine test					
Leucocyte count ($\times 10^9/L$; normal range 3.5–9.5)	4.83 (3.82–6.35)	5.52 (4.03–6.79)	–1.656	0.098	1.103 (0.996–1.222)
Increased	24 (4.5%)	6 (9.4%)	2.883	0.237	1.370 (0.760–2.469)
Normal	420 (78.5%)	48 (75.0%)			
Decreased	91 (17.0%)	10 (15.6%)			
Lymphocyte count ($\times 10^9/L$; normal range 1.1–3.2)	1.20 (0.86–1.75)	1.16 (0.84–1.63)	–0.941	0.347	1.022 (0.992–1.053)
Increased	36 (6.8%)	6 (9.8%)	2.027	0.363	0.926 (0.598–1.433)
Normal	264 (49.9%)	25 (41.0%)			
Decreased	229 (43.3%)	30 (49.2%)			
Lymphocyte percentage (normal range 20–50%)	25.20 (16.70–31.98)	18.70 (13.75–28.21)	–2.439	0.015	0.980 (0.957–1.003)
Increased	21 (4.0%)	2 (3.2%)	8.265	0.016	0.517 (0.315–0.848)
Normal	324 (61.4%)	27 (43.5%)			
Decreased	183 (34.7%)	33 (53.2%)			
Neutrophil percentage (normal range 50–70%)	65.90 (56.40–74.50)	71.70 (57.40–81.80)	–2.135	0.033	1.004 (0.989–1.019)
Increased	199 (38.0%)	33 (52.4%)	5.658	0.059	1.355 (0.909–2.021)
Normal	253 (48.4%)	21 (33.3%)			
Decreased	71 (13.6%)	9 (13.7%)			

^aCalculations were computed by removing NA (not applicable) from each analytical cohort. *z*/ χ^2 means either *z* score or χ^2 statistics.

onset to diagnosis (*P* = 0.04, OR = 1.065, 95% CI: 1.003–1.132) and lymphocyte percentage (*P* = 0.009, OR = 0.517, 95% CI: 0.315–0.848) were identified as prognostic of disease severity (Table 5).

Discussion

The lower mean (43 years) and median (44 years) ages of this cohort, compared with those reported for other regions,^{3,9,10} were driven by the inclusion of a 6-month old infant; on the other hand, this suggested family transmission. Male patients dominated this case series (male: female ratio = 1.28), which is consistent with other reports.^{3,4} Hypertension was the most prevalent medical disorder among patients with comorbidities, followed by respiratory diseases and diabetes.

A total of 14 patients were symptom-free, all of whom were <60 years old and diagnosed within 2 days due to their close contact with infected cases. These patients did not show any symptoms, which might be attributable to the long latency of SARS-CoV-2 and/or the strong immune system of these individuals. Detailed information on the duration from their initial infection source exposure to the diagnosis is required to know whether ‘true’ symptom-free SARS-CoV-2 carriers exist. Immediate laboratory testing of individuals having close contact with confirmed patients is recommended.

Two of three pregnant patients in this study had severe or terminal disease, which might be associated with their reduced immunity during pregnancy. All eight infected children (aged ≤ 8 years) were symptom-free, with disease of light or normal severity, suggesting the existence of extended maternal immunity.¹¹

Table 5. Statistics of the univariate logistic regression models for COVID-19 severity (non-severe, severe) as stratified by significant factors identified from the multivariate logistic regression model

Characteristics ^a	beta	SE	Wald	P	Odds ratio (95% CI)
Age (years)	0.026	0.008	9.534	0.002	1.026 (1.009–1.043)
Contact with confirmed cases	−0.786	0.389	4.088	0.043	0.456 (0.213–0.976)
Days from illness onset to diagnosis	0.063	0.031	4.206	0.040	1.065 (1.003–1.132)
Neutrophil percentage	0.004	0.007	0.268	0.604	1.004 (0.989–1.019)
Lymphocyte percentage	−0.660	0.253	6.818	0.009	0.517 (0.315–0.848)

^aExcept for 'Contact with confirmed cases', which is categorized as 'Yes' or 'No', all other variables are continuous.

Of the 671 infected cases, the transmission source of 42 (including 3 medical workers, 1 of whom had hypertension) could be traced back to a hospital and 5 to a wet market, suggesting that these two places are high-risk areas that warrant reduced exposure for the benefits of self-protection and SARS-CoV-2 transmission control.

In this dataset, 33 patients (4.9%) did not visit Wuhan nor have any close contact with confirmed or suspected cases, suggesting the existence of SARS-CoV-2 transmission via fomite, such as aerosol. Among these cases with an unknown transmission source, 21.2% (7) were farmers. This might be due to the high exposure of farmers to the outdoor environment and, possibly, less education and awareness about self-protection against SARS-CoV-2 infection in China. These 33 individuals were abnormally distributed (Shapiro–Wilk statistic = 0.944, $P = 5.32 \times 10^{-15}$) by stratifying age into 10-year groups, suggesting that fomite-mediated infection is a random event. The majority of these cases were male (19/33). Except for one male patient with comorbidities in the lung, all patients had light (11) or normal (21) disease severity, suggesting that SARS-CoV-2 transmission via fomite might be less infectious than transmission via living organisms, possibly due to reduced virulence or viability. However, we did not find any statistical significance on this hypothesis.

Among the three conventional pneumonia diagnostic approaches, lymphocyte percentage from routine blood tests on patient admission conveyed prognostic value. It is worth emphasizing that the range of the statistics considered as healthy for each blood test index does not apply for the diagnosis of COVID-19. High lymphocyte percentage was a sign of poor prognosis but was not a good diagnostic measure for COVID-19 due to its large variation.

Typical pneumonia features on CT scans are bilateral ground glass opacities, and CT had a remarkably higher detection rate than X-ray. The false negative rate of CT scans was 2.3%, and that of X-ray was as high as 62%. False negative cases from CT scans all had light or normal disease severity or were symptom-free, and without comorbidities. The distribution of the disease severities of infected individuals not detected on X-ray did not vary

from that of the full cohort ($\chi^2=1.964$, $P=0.161$, OR = 1.758, 95% CI: 0.793–3.901) (Table 4). This suggests that X-ray was not a feasible approach for COVID-19 diagnosis, and patients without bilateral ground glass opacities on CT scans might not be severely infected.

A longer time from illness onset to diagnosis was associated with higher COVID-19 severity, older age, higher probability of having comorbidities and being male.

Conclusion

This data series is, to our knowledge, by far the largest dataset reported outside Wuhan and at a provincial level in China, which provides information to aid the understanding of COVID-19 clinical manifestation. The main findings include: (i) 'age', 'contact with confirmed cases', 'days from illness onset to diagnosis' and 'lymphocyte percentage' are prognostic of disease severity; (ii) low lymphocyte percentage is associated with severe COVID-19 in general but is not a good index for diagnosis; (iii) X-ray is not a feasible diagnostic approach for COVID-19; (iv) farmers are an occupational group with a high infection rate in Henan province, China; and (v) SARS-CoV-2 transmission via fomite is evident. This study is limited by the missing information in different analytical indexes and lack of follow-up data on the collected cases.

Funding

This work was supported by the Special Emergency Project for New Coronavirus Prevention and Control in Henan Province [grant number 201100310800], the National Natural Science Foundation of China [grant number 81972789], the National Science and Technology Major Project [grant number 2018ZX10302205-004-002], the Six Talent Peaks Project in Jiangsu Province [grant number SWYY-128], Fundamental Research Funds for the Central Universities [grant number JUSRP22011], and Technology Development Funding of Wuxi [grant number WX18IVJN017]. The funders of the study had no role in study design, data collection, data analysis, data interpretation or manuscript preparation. The corresponding author had full access to all the data in the study and approved the decision to submit for publication.

Author Contributions

X.D. conceived, designed, and planned the study. Y.N., X.H., W.G., X.Z., H.W., M.Q., X.T. and X.S. acquired the data. J.L. and Y.M. did the statistical analysis. X.D. and J.L. prepared the first draft of the manuscript. All authors interpreted the results, revised the article and approved the decision to submit for publication.

Conflict of Interest

None declared.

References

1. Zhu N, Zhang D, Wang W *et al.* A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 2020; **382**:727–33.
2. Huang C, Wang Y, Li X *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020; **395**:497–506.
3. 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.
4. Guan WJ, Ni ZY, Hu Y *et al.* Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med* 2020; **382**:1708–20. doi:10.1056/NEJMoa2002032.
5. Chang D, Lin M, Wei L *et al.* Epidemiologic and clinical characteristics of novel coronavirus infections involving 13 patients outside Wuhan, China. *JAMA* 2020; **323**:1092–3.
6. Patel A, 2019-nCoV CDC Response Team, Jernigan DB, Abdirizak F, Abedi G, Aggarwal S, Albina D, Allen E. Initial Public Health Response and Interim Clinical Guidance for the 2019 Novel Coronavirus Outbreak-United States, December 31, 2019-February 4, 2020. *MMWR Morb Mortal Wkly Rep* 2020; **69**:140–6.
7. Aronow WS, Fleg JL, Pepine CJ *et al.* ACCF/AHA 2011 expert consensus document on hypertension in the elderly: a report of the American College of Cardiology Foundation Task Force on Clinical Expert Consensus Documents developed in collaboration with the American Academy of Neurology, American Geriatrics Society, American Society for Preventive Cardiology, American Society of Hypertension, American Society of Nephrology, Association of Black Cardiologists, and European Society of Hypertension. *J Am Soc Hypertens: JASH* 2011; **5**: 259–352.
8. Genuth S, Alberti KG, Bennett P *et al.* Follow-up report on the diagnosis of diabetes mellitus. *Diabetes Care* 2003; **26**:3160–7.
9. Li Q, Guan X, Wu P *et al.* Early transmission dynamics in wuhan, china, of novel coronavirus-infected pneumonia. *N Engl J Med* 2020; **382**:1199–207.
10. Chen N, Zhou M, Dong X *et al.* Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* 2020; **395**: 507–13.
11. Bhargav H, Huilgol V, Metri K *et al.* Evidence for extended age dependent maternal immunity in infected children: mother to child transmission of HIV infection and potential interventions including sulfatides of the human fetal adnexa and complementary or alternative medicines. *J Stem Cells* 2012; **7**:127–53.