Protocol

This trial protocol has been provided by the authors to give readers additional information about their work.

Protocol for: Guan W, Ni Z, Hu Y, et al. Clinical characteristics of coronavirus disease 2019 in China. N Engl J Med. DOI: 10.1056/NEJMoa2002032

Study protocol

Background

In early December 2019, the first pneumonia cases of unknown origins were identified in Wuhan city, Hubei province, China. High-throughput sequencing has revealed a novel betacoronavirus that is currently named 2019 novel coronavirus (2019-nCoV), which resembled severe acute respiratory syndrome coronavirus (SARS-CoV). The 2019-nCoV is the seventh member of enveloped RNA coronavirus (subgenus *sarbecovirus, Orthocoronavirinae* subfamily). Evidence pointing to the person-to-person transmission in hospital and family settings has been accumulating.

The World Health Organization has recently declared the 2019-nCoV a public health emergency of international concern. As of February 5th, 2020, 24,554 laboratory-confirmed cases have been documented globally (i.e., the USA, Vietnam, Germany). 28,018 laboratory-confirmed cases and 563 death cases in China as of February 6th, 2020. Despite the rapid spread worldwide, the clinical characteristics of 2019-nCoV acute respiratory disease (ARD) remain largely unclear. In two recent studies documenting the clinical manifestations of 41 and 99 patients respectively with laboratory-confirmed 2019-nCoV ARD who were admitted to Wuhan, the severity of some cases with 2019-nCoV ARD mimicked that of SARS-CoV.

Objective

We sought to identify the defining epidemiological and clinical characteristics with greater precision, but also unravel the risk factors associated with mortality.

Hypothesis

The clinical characteristics of patients with 2019-nCoV in our study cohort might differ from those reported recently given the increased sample size; greater disease severity might be associated with an increased risk of reaching to the composite endpoint.

Study design

Retrospective study on the patients from provinces/autonomous regions/provincial municipalities across China. The time frame started from the ‘first case’ (November 2019) to the latest date when available (late January 2020).

Study participants

Diagnostic criteria

Cases will be diagnosed based on the WHO interim guidance. A confirmed case with 2019-nCoV ARD will be defined as a positive result to high-throughput sequencing or real-time reverse-transcriptase polymerase-chain-reaction (RT-PCR) assay for nasal and pharyngeal swab specimens

Inclusion criteria

Only the laboratory-confirmed cases were included the analysis.

Exclusion criteria

No exclusion criteria apply provided that the clinical profiles of the patients are as complete as possible, which would not prevent from the analysis of the composite endpoint.

Outcome assessment

The primary composite endpoint will be the admission to intensive care unit (ICU), or mechanical ventilation, or death. Secondary endpoints comprised mortality rate, the time from symptom onset to the composite endpoint and each of its component. Because clinical observations were still ongoing, fixed time frame (i.e. within 28 days) will not be applied to these endpoints.

Definitions of exposure and clinical complications

** The exposure to wildlife denote that a person in close contact with wildlife animals (bats, snakes, civet cats, etc.) or visiting either a wildlife retailer or a market selling wildlife within two weeks before the onset of respiratory symptoms. However, the cases with regular visit to the market without recalling the exposure date will not be considered as having definite exposure to the wildlife.

** Pneumonia will be diagnosed as an acute respiratory disorder characterized by the presence of cough and at least one of the new-onset focal chest signs, fever for more than 4 days or dyspnoea/ tachypnoea.

** Shock and acute respiratory distress syndrome (ARDS) will be defined in accordance with the WHO interim guidance.

** Acute kidney injury will be defined based on the highest serum creatinine level and urine output. Specifically, the diagnosis could be made based on any of the following criterion: an increase in serum creatinine levels by 0.3 mg/dl or greater (26.5 μmol/l or greater) within 48 hours; or increase in serum creatinine levels to 1.5 times of the baseline level or greater, which was known or

presumed to have occurred within 7 days; or urine volume of below 0.5 ml/kg/h for 6 consecutive hours.

** The diagnosis of secondary bacterial or fungal infection will be made in case of the occurrence of hospital-acquired pneumonia or bacteremia, plus a positive result of new pathogen culture from the blood and lower respiratory tract specimen (including sputum, bronchoalveolar lavage fluid or tracheal aspirate) obtained at least 8 hours after admission.

** Acute heart failure will be defined as the clinical syndrome characterized by typical symptoms (e.g. breathlessness, ankle swelling and fatigue) that may be accompanied by signs (e.g. elevated jugular venous pressure, pulmonary crackles and peripheral edema) caused by a structural and/or functional cardiac abnormality.

** Rhabdomyolysis will be diagnosed if the muscle pain or muscle weakness took place on admission and the creatine kinase level was greater than 10 times the upper limit of normal.

Assessment of exposure

In this study, the history of contact with wildlife will be documented through self-report by the patients. Among the patients who live outside Wuhan, the absolute count and percentage of patients who are the residents of Wuhan, patients who have a recent travel to Wuhan and patients who have a contact with people from Wuhan will be reported, respectively. A recent travel to, nor contact with people from, Wuhan will be inquired among the patients living outside of Wuhan.

Sampling and sample size estimation

Because all laboratory-confirmed cases will have to be reported to the National Health Commission which is the main coordinator of our study, our study population would not be derived from random sampling. Since early this January, the medical staff in Hubei (especially in Wuhan) suffered from a major burn-out of the workload given the rapid surge of cases. A team consisting of three experts of the lead site will be dispatched to Wuhan to assist in extracting the clinical data from the electronic hospital information system.

No previous study has documented the percentage of patients who reach to the composite endpoint or either of the outcome variable. Hence, it would not be possible to estimate for the sample size based on literature reports. However, it is believed that the large sample sizes derived from hospitals across China would be sufficient to power the statistical analysis in our study.

Data management

Data will be collected from the electronic medical records (EMR) or paper-based medical records that managed by the National Health Commission. All medical records will be transmitted or copied and sent to the data processing center in Guangzhou under the coordination of the National Health Commission. All needed variables with necessary explanations will be defined firstly by the research team. An experienced respiratory clinicians team will review the copies of the medical records and abstract the data. The abstracted data will be entered into a computerized database, and double-entry is required for all variables. The data-entry team ensures that all data needed are collected. If the core data were missing, requests of clarification will be immediately sent to the coordinators who subsequently contacted the attending clinicians. Data cleaning, including logical check, outlier check, and variables engineering, will be performed by experienced programmers. An experienced clinician will assist the variables engineering, the original variables will be transformed if need, which may include, but is not restricted to, converting continuous variables to categorical variables, and combining multiple variables into single variables for information integration.

Statistical analysis plan

Sampling

The convenience sampling method will be used because random sampling will not be feasible under the epidemic situation of 2019-nCoV. Sampling frame will include all hospitals which admit the patients with 2019-nCoV infection.

Sample size estimation

The sample size will not be estimated based on the statistical power. We will collect as many cases as possible.

Analysis principles

- All tests are two-sided, the nominal level of type I error will be 5% and the confidence level for all confidence intervals will be 95%.
- There will be no imputing of missing values. The number of observations used in an analysis will be reported.
- Subgroup analyses will be carried out irrespective of whether there is a significant treatment effect on the primary outcome.
- Analyses will be conducted using R packages (version 3.6.2). Distribution map will be plotted by using ArcGis (version 10.2.2).

Trial profile

Flow chart of inclusion will be displayed in a diagram. The report will include the number of patients who met the inclusion criteria and the number of non-included patients. The distribution of laboratory-confirmed cases throughout china will also be reported by using a distribution map.

Patients characteristics and baseline comparisons

Description and statistical inference of the patients' characteristics, radiographic and laboratory findings will be presented by the disease severity and the composite endpoint.

Discrete variables will be summarized by frequencies and percentages. Percentages will be calculated according to the number of patients for whom data is available. Continuous variables will be summarized by median and 25%, 75% quartiles [Median (IQR)].

Statistical inference of continuous variables will be performed using t-test or Wilcoxon rank-sum tests as appropriate. The Pearson's Chi-square test or Fisher exact test will be used, as appropriate, for categorical data.

Baseline measures for all patients will be tabulated.

Outcomes

For the primary outcome of time to the composite endpoint and other outcomes, the multivariable analysis will be performed by Fine-Gray competing-risk models in which recovery will be included as competing for risk. The candidate risk factors include exposure history, age, radiologic and laboratory findings, and the development of complications. Candidate risk factors that are of biological interest (age and sex) or clinical important (smoking status) or statistically significant will be included in the multivariable models. The sub-distribution hazard ratios with 95% confidence interval will be reported.

Subgroup analysis

Subgroup analyses will be carried out for the primary outcome.

Planned subgroup analysis:

- disease severity

Sensitivity analysis

The proportional hazard Cox model analysis will be conducted as sensitivity analyses.

Ethics approval

The study has been approved by the National Health Commission and the institutional board of each participating site. Written informed consent has been waived in light of the urgent need to collect clinical data.

Data presentation schemes

Table 1. Clinical characteristics of the patients with 2019-nCoV

Clinical characteristics, symptoms or signs	**All patients N=**	**Mild-moderate N=**	**Severe N=**	**P value**
Age				
Median (range) – yrs				
Age groups – No., %				
0-14 yrs				-
15-49 yrs				-
50-64 yrs				-
≥ 65 yrs				-
Female sex – No.,%				
Current smokers – No., %				
Exposure to live poultry				
Within previous 14 days – No., %				
Median incubation time (IQR) – days				
Hospitalization – No., %				
Recent travel to infected region – No., %				
None				-
Hubei province				-
Elsewhere				-
Incubation period – days				
Median (range)				
On admission No./Total No. (%)				
Respiratory symptoms – No., %	-	-	-	-
Fever				
37.5-38.0℃				-
38.1-39.0℃				-
> 39.0℃				-
Conjunctival congestion				

Nasal congestion				
Rhinorrhea				
Headache				
Cough				
Sore throat				
Sputum production				
Fatigue				
Hemoptysis				
Shortness of breath				
Nausea or vomiting				
Diarrhea				
Mylgia or arthralgia				
Chill				
Signs – No., %	-	-	-	-
Throat congestion				
Tonsil swelling				
Enlargement of lymph nodes				
Rash				
Coexisting disorders – No., %	-	-	-	-
Any				
Chronic obstructive pulmonary disease				
Diabetes				
Hypertension				
Coronary heart disease				
Cerebrovascular diseases				
Hepatitis B infection				
Cancer				
Chronic renal diseases				
Immunodeficiency				
Pregnancy				

P values denoted the comparison between mild-moderate cases and severe cases.

Table 2. Radiographic and laboratory testing findings of 2019-nCoV

Variables	All patients N=	Mild-moderate N=	Severe N=	P value
Radiographic findings	-	-	-	-

Abnormalities on chest X-ray – No./total No. (%)				
Ground-glass opacities				
Local patchy shadowing				
Diffuse patchy shadowing				
Interstitial abnormalities				
Abnormalities on chest CT – No./total No. (%)				
Ground-glass opacities				
Local patchy shadowing				
Diffuse patchy shadowing				
Interstitial abnormalities				
Laboratory findings	-	-	-	-
Median PaO_2: FiO_2 (interquartile range)				
Blood leukocyte count	-	-	-	-
Median, interquartile range – per mm^3				
>10,000 per mm^3 – No., %				
<4,000 per mm^3 – No., %				
Lymphocyte count	-	-	-	-
Median, interquartile range – per mm^3				
Lymphocytopenia – No., %				
CD4+-T lymphocyte count – per mm^3				
CD8+-T lymphocyte count – per mm^3				
CD4:CD8 ratio <1.4 – No./total No. (%)				
Platelet count	-	-	-	-
Median, interquartile range – per mm^3				
Thrombocytopenia – No., %				
Haemoglobin level – g/dl				
C-reactive protein level ≥ 10 mg/liter – No./total No. (%)				
Procalcitonin level ≥ 0.5 ng/ml – No./total No. (%)				
Lactose dehydrogenase ≥ 250 U/liter – No./total No. (%)				
Aspartate aminotransferase > 40 U/liter – No./total No. (%)				
Alanine aminotransferase > 40 U/liter – No./total No. (%)				

Total bilirubin 17.1 μmol/liter – No./total No. (%)				
Creatinine kinase ≥ 200 U/liter – No./total No. (%)				
Creatinine ≥ 133 μmol/liter – No./total No. (%)				
D-dimer ≥ 0.5 mg/liter – No./total No. (%)				
Sodium – mmol/liter				
Potassium – mmol/liter				
Chloride – mmol/liter				

P values denoted the comparison between mild-moderate cases and severe cases.

Table 3. Complications, treatment and clinical outcomes of patients with 2019-nCoV

Characteristics	All patients N=	Mild-moderate N=	Severe N=	P value

Complications – No., %				
Septic shock				
Pneumonia				
Acute respiratory distress syndrome				
Acute kidney injury				
Disseminated intravascular coagulation				
Rhabdomyolysis				
Treatment – No., %				
Administration of intravenous antibiotics				
Timing from onset of illness to antibiotic administration				
0-2 days				
3 days or longer				
Administration of oseltamivir				
Timing from onset of illness to antiviral administration				
0-2 days				
3 days or longer				
Administration of antifungal medications				
Administration of systemic corticosteroids				
Maximal corticosteroid dose > 0.1g per day				
Oxygen therapy				
Mechanical ventilation				
Non-invasive				
Invasive				
Intensive care unit admission				
Use of extracorporeal membrane oxygenation				
Use of continuous renal replacement therapy				
Use of intravenous immunoglobin				
Clinical outcomes	-	-	-	-
Discharge from hospital				
Death				
Cause of death	-	-	-	-

Respiratory failure				
Secondary bacterial or fungal infections				
Acute heart failure				
Acute kidney failure				
Shock				
Coagulopathy				

Table 4. Multivariate risk factors associated with the development of acute respiratory distress syndrome and mortality

	Acute respiratory distress syndrome				Mortality		
Risk factor	**OR**	**95%CI**	**P value**		**OR**	**95%CI**	**P value**
Age							
Any coexisting medical conditions							
Duration of disease > XX days							
Time from onset of symptoms to initiation of antiviral medications > X days							
Alanine or aspartate aminotransferase level > 40 U/liter							
Blood leukocyte count < 1*10^9/L							
Blood lymphocyte count < 1*10^9/L							
Serum creatinine level > XX U/liter							

Table 5. Summary characteristics of 2019-nCoV, SARS-CoV, MERS-CoV and highly pathogenic influenza

Clinical characteristics	2019-nCoV	SARS-CoV	MERS-CoV	Highly pathogenic influenza
Median incubation period (days)				
Median age (yrs)				
% of the elderly people				
Median period of viral shedding (days)				
High-grade fever (%)				
Gastrointestinal symptoms (%)				
Leukopenia (%)				
Lymphocytopenia (%)				
Thrombocytopenia (%)				
Abnormal liver function (%)				
Coagulatory disorders (%)				
Needing invasive ventilation (%)				
Needing EMCO (%)				
Needing NRRT (%)				
Mortality (%)				

Table E1. Clinical characteristics of patients with 2019-nCoV stratified by clinical outcomes

Clinical characteristics, symptoms or signs	**Alive** **N=**	**Died** **N=**	**P value**
Age			
Median (range) – yrs			
Age groups – No., %			
0-14 yrs			-
15-49 yrs			-
50-64 yrs			-
≥ 65 yrs			-
Female sex – No.,%			
Current smokers – No., %			
Exposure to live poultry			
Within previous 14 days – No., %			
Median incubation time (IQR) – days			
Hospitalization – No., %			
Recent travel to infected region – No., %			
None			-
Hubei province			-
Elsewhere			-
Incubation period – days			
Median (range)			
On admission **No./Total No. (%)**			
Respiratory symptoms – No., %			
Fever	-	-	-
37.5-38.0℃			-
38.1-39.0℃			-
> 39.0℃			-
Conjunctival congestion			

Nasal congestion			
Rhinorrhea			
Headache			
Cough			
Sore throat			
Sputum production			
Fatigue			
Hemoptysis			
Shortness of breath			
Nausea or vomiting			
Diarrhea			
Mylgia or arthralgia			
Chill			
Signs – No., %	-	-	-
Throat congestion			
Tonsil swelling			
Enlargement of lymph nodes			
Rash			
Coexisting disorders – No., %	-	-	-
Any			
Chronic obstructive pulmonary disease			
Diabetes			
Hypertension			
Coronary heart disease			
Cerebrovascular diseases			
Hepatitis B infection			
Cancer			
Chronic renal diseases			
Immunodeficiency			
Pregnancy			

Table E2. Radiographic and laboratory testing findings of patients with 2019-nCoV stratified by clinical outcomes

Variables	Alive N=	Died N=	P value
Radiographic findings			
Abnormalities on chest X-ray – No./total No. (%)			
Ground-glass opacities			
Local patchy shadowing			
Diffuse patchy shadowing			
Interstitial abnormalities			
Abnormalities on chest CT – No./total No. (%)			
Ground-glass opacities			
Local patchy shadowing			
Diffuse patchy shadowing			
Interstitial abnormalities			
Laboratory findings			
Median PaO_2: FiO_2 (interquartile range)			
Blood leukocyte count	-	-	-
Median, interquartile range – per mm^3			
>10,000 per mm^3 – No., %			
<4,000 per mm^3 – No., %			
Lymphocyte count	-	-	-
Median, interquartile range – per mm^3			
Lymphocytopenia – No., %			
CD4+-T lymphocyte count – per mm^3			
CD8+-T lymphocyte count – per mm^3			
CD4:CD8 ratio <1.4 – No./total No. (%)			
Platelet count	-	-	-
Median, interquartile range – per mm^3			
Thrombocytopenia – No., %			

Haemoglobin level – g/dl			
C-reactive protein level ≥ 10 mg/liter – No./total No. (%)			
Procalcitonin level ≥ 0.5 ng/ml – No./total No. (%)			
Lactose dehydrogenase ≥ 250 U/liter – No./total No. (%)			
Aspartate aminotransferase > 40 U/liter – No./total No. (%)			
Alanine aminotransferase > 40 U/liter – No./total No. (%)			
Total bilirubin 17.1 µmol/liter – No./total No. (%)			
Creatinine kinase ≥ 200 U/liter – No./total No. (%)			
Creatinine ≥ 133 µmol/liter – No./total No. (%)			
D-dimer ≥ 0.5 mg/liter – No./total No. (%)			
Sodium – mmol/liter			
Potassium – mmol/liter			
Chloride – mmol/liter			

Table E3. Complications, treatment and clinical outcomes of patients with 2019-nCoV stratified by clinical outcomes

Characteristics	Alive N=	Died N=	P value
Complications – No., %			
Septic shock			
Pneumonia			
Acute respiratory distress syndrome			
Acute kidney injury			
Disseminated intravascular coagulation			
Rhabdomyolysis			
Treatment – No., %			
Administration of intravenous antibiotics			
Timing from onset of illness to antibiotic administration			
0-2 days			
3 days or longer			
Administration of oseltamivir			
Timing from onset of illness to antiviral administration			
0-2 days			
3 days or longer			
Administration of antifungal medications			
Administration of systemic corticosteroids			
Maximal corticosteroid dose > 0.1g per day			
Oxygen therapy			
Mechanical ventilation			
Non-invasive			
Invasive			
Intensive care unit admission			
Use of extracorporeal membrane oxygenation			

Use of continuous renal replacement therapy			
Use of intravenous immunoglobin			
Clinical outcomes			
Discharge from hospital			
Death			
Cause of death	-	-	-
Respiratory failure			
Secondary bacterial or fungal infections			
Acute heart failure			
Acute kidney failure			
Shock			
Coagulopathy			