medRxiv preprint doi: https://doi.org/10.1101/2020.04.06.20056127; this version posted April 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Prediction of the clinical outcome of COVID-19 patients using T lymphocyte subsets with 340 cases from Wuhan, China: a retrospective cohort study and a web visualization tool

Qibin Liu[1], Xuemin Fang[1], Shinichi Tokuno[3], Ungil Chung[4], Xianxiang Chen[5], Xiyong Dai[6], Xiaoyu Liu[7], Feng Xu[8], Bing Wang[9], Peng Peng*

[1] Wuhan Pulmonary Hospital, Wuhan Institute for Tuberculosis Control, No. 28 Baofeng Road, Qiaokou District, Wuhan City, Hubei Province, China; Liuqibin0221@163.com

[1] Graduate School of Health Innovation, Kanagawa University of Human Services, 2F Bldg.2-A, 3-25-10, Tonomachi Kawasaki-ku, Kawasaki City, Kanagawa 210-0821, Japan; mindy.fang@gmail.com

[3] Graduate School of Health Innovation, Kanagawa University of Human Services, 2F Bldg.2-A, 3-25-10, Tonomachi Kawasaki-ku, Kawasaki City, Kanagawa 210-0821, Japan; tokuno@m.u-tokyo.ac.jp

[4] Graduate School of Health Innovation, Kanagawa University of Human Services, 2F Bldg.2-A, 3-25-10, Tonomachi Kawasaki-ku, Kawasaki City, Kanagawa 210-0821, Japan; tei@tetrapod.t.u-tokyo.ac.jp

[5] Wuhan Pulmonary Hospital, Wuhan Institute for Tuberculosis Control, No. 28 Baofeng Road, Qiaokou District, Wuhan City, Hubei Province, China; 2272534937@qq.com

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.06.20056127; this version posted April 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

[6] Wuhan Pulmonary Hospital, Wuhan Institute for Tuberculosis Control, No. 28 Baofeng Road, Qiaokou District, Wuhan City, Hubei Province, China; daixiyong71@126.com

[7] Wuhan Pulmonary Hospital, Wuhan Institute for Tuberculosis Control, No. 28 Baofeng Road, Qiaokou District, Wuhan City, Hubei Province, China; liuxiaoyu9-102@163.com

[8] Wuhan Pulmonary Hospital, Wuhan Institute for Tuberculosis Control, No. 28 Baofeng Road, Qiaokou District, Wuhan City, Hubei Province, China; victor_fsxq@163.com

[9] Wuhan Pulmonary Hospital, Wuhan Institute for Tuberculosis Control, No. 28 Baofeng Road, Qiaokou District, Wuhan City, Hubei Province, China; 779238021@qq.com

[*] Wuhan Pulmonary Hospital, Wuhan Institute for Tuberculosis Control, No. 28 Baofeng Road, Qiaokou District, Wuhan City, Hubei Province, China; pengpengwg@126.com; Tel:+86-027-83605535; Fax: +86-027-83602499

medRxiv preprint doi: https://doi.org/10.1101/2020.04.06.20056127; this version posted April 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Abstract

Background

Wuhan, China was the epicenter of the 2019 coronavirus outbreak. As a designated hospital, Wuhan Pulmonary Hospital has received over 700 COVID-19 patients. With the COVID-19 becoming a pandemic all over the world, we aim to share our epidemiological and clinical findings with the global community.

Methods

In this retrospective cohort study, we studied 340 confirmed COVID-19 patients from Wuhan

medRxiv preprint doi: https://doi.org/10.1101/2020.04.06.20056127; this version posted April 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Pulmonary Hospital, including 310 discharged cases and 30 death cases. We analyzed their demographic, epidemiological, clinical and laboratory data and implemented our findings into an interactive, free access web application.

Findings

Baseline T lymphocyte Subsets differed significantly between the discharged cases and the death cases in two-sample t-tests: Total T cells ($p < 2{\cdot}2e\text{-}16$), Helper T cells ($p < 2{\cdot}2e\text{-}16$), Suppressor T cells ($p = 1{\cdot}8\text{-}14$), and TH/TS (Helper/Suppressor ratio, $p = 0{\cdot}0066$). Multivariate logistic regression model with death or discharge as the outcome resulted in the following significant predictors: age (OR 1·05, p 0·04), underlying disease status (OR 3·42, p 0·02), Helper T cells on the log scale (OR 0·22, p 0·00), and TH/TS on the log scale (OR 4·80, p 0·00). The McFadden pseudo R-squared for the logistic regression model is 0·35, suggesting the model has a fair predictive power.

Interpretation

While age and underlying diseases are known risk factors for poor prognosis, patients with a less damaged immune system at the time of hospitalization had higher chance of recovery. Close monitoring of the T lymphocyte subsets might provide valuable information of the patient's condition change during the treatment process. Our web visualization application can be used as a supplementary tool for the evaluation.

Funding

The authors report no funding.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.06.20056127; this version posted April 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Prediction of the clinical outcome of COVID-19 patients using T lymphocyte subsets with 340 cases from Wuhan, China: a retrospective cohort study and a web visualization tool

Introduction

The first case of COVID-19 infection was detected in December, 2019 in Wuhan, China. The situation quickly escalated into an epidemic all over China with tens of thousands of infections and several thousand deaths[1-2]. Further on March 11, 2020, the World Health Organization (WHO) declared COVID-19 a pandemic, pointing to the over 118,000 cases of infections in over 110 countries around the world and the sustained risk of further global spread[3-5].

There is no known pre-existing immunity to COVID-19 in humans. In the absence of any anti-new corona virus therapy, infection of the virus weakens the host's immune system, leading to Corona Virus Disease 2019 (COVID-19). Current estimates are that 2019-nCoV has an incubation period of 2 to 14 days, with potential asymptomatic transmission[6-7].

Hubei province is located in the central region of China and Wuhan is the capital city of Hubei. With the traffic to and from the rest parts of the country, Wuhan had become the gateway for COVID-19 epidemic in China and is the most affected area in the country. The situation was further worsened by the Chinese Spring Festival travel rush (from mid-January to the end of February), when the world's largest annual human migration occurs. On January 23, a lockdown in Wuhan and other cities in Hubei province started in an effort to quarantine the epicenter. Recent statistics suggest that almost 95% of reported COVID-19 cases in China originated from Wuhan [2].

Wuhan Pulmonary Hospital was a tertiary infectious disease hospital before the outbreak, with 610 beds and ten ICU beds. When the COVID-19 epidemic started in Wuhan, it became one of the first Covid-19 designated hospitals and has received more than 700 COVID-19 patients so far. In this study we aim to make full use of our experience with the COVID-19 patients and provide insight for the global community in evaluating the patient's risk and progress. We have further implemented our findings into an interactive web tool for more practical uses.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.06.20056127; this version posted April 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Methods

Ethics approval

The study protocol was reviewed and approved by the Ethics committee of Wuhan Pulmonary Hospital (WPE 2020-8). Informed consents were obtained from all participants before enrollment. Patient records and information were de-identified prior to the analysis.

Process Flow at the Hospital

Figure 1 illustrates the process flow in Wuhan Pulmonary Hospital. The patients were received from two sources: 1. Outpatient visit. People with symptoms such as fever, cough, or diarrhea made an Outpatient visit and took a blood test to exclude the possibility of bacterial infection. They then took an Influenza test to exclude the possibility of Influenza. Next a lung CT was taken to identified the infected area. Finally, a COVID-19 throat swab was taken to confirm the infection; 2. COVID-19 positive cases from the square cabin hospitals. These patients were transferred to Wuhan Pulmonary Hospital if their condition became severe or critical. Once the patient was hospitalized, regular examinations on T lymphocyte Subsets, COVID-19 PCR test, and lung CT were taken place to closely monitor the patient's progress. Most patients who were hospitalized from January 28, 2020 to March 8, 2020 received treatment according to the fifth edition of the Ministry of Health guidelines[8-9]: 1. Bed rest and supportive treatment; 2. Regular examination on blood routine, CRP, biochemistry, coagulation function, myocardial enzymes, lung CT; 3. Oxygen therapy, including nasal catheter oxygen, transposal high-flow oxygen therapy, ventilator; 4. Antiviral therapy, including Lopinavir/Ritonavir tablets, Arbidol Hydrochloride Tablets, etc. If the patient maintained a normal body temperature for at least three days, and tested negative with COVID-19 (including throat swab, stool, and sputum) at least twice and 24+ hours apart, as well as showing significant improvement in lung CT, the patient would be considered recovered and discharged from the hospital. The discharged patients were transported by the Chinese government using designated vehicles and they were further isolated for at least two weeks.

Blood samples

2 mL venous blood samples were obtained from each patient to measure T cell subsets, and all the analyses were completed within 4 hours of sampling. FACSCalibur flow cytometer (BD Biosciences, San Jose, CA) was used for flow cytometry acquisition and analysis. The absolute count of each lymphocyte subset was determined using CD3/CD4/CD8/CD45 BD Multitest reagents according to the manufacturer's protocol (BD Biosciences).

medRxiv preprint doi: https://doi.org/10.1101/2020.04.06.20056127; this version posted April 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Study population and data collection

From January 31 to March 8, 2020, a total of 721 patients who tested positive of COVID-19 were admitted to Wuhan Pulmonary Hospital. Among these patients, 430 completed the treatment and were discharged, 62 died, and 229 still remain in hospitalization. Excluding four patients whose direct cause of death was not COVID-19 infection, and selecting patients who had at least one T cell Subsets test available, we had a total of 340 patients in the study, including 310 discharged cases and 30 death cases. We reviewed laboratory test results and chest CT examinations of these 340 patients, and collected all the T lymphocyte subsets tests data. If multiple T lymphocyte subsets tests were performed, we chose the earliest one as the baseline. Two researchers independently reviewed the collected data to ensure data accuracy.

Selection of analysis data

There have been a number of descriptive analyses about the epidemiological, clinical, laboratory, and radiological characteristics of the COVID-19 patients[10-14]. Little as yet known about how these characteristics can be used in guiding the practice of the healthcare providers[15]. As the world face the expanding COVID-19 pandemic, with the number of infected cases increase exponentially every day, any quick and easy method of understanding the patient's condition can be valuable. With this in mind we seek to build a statistical model with only a few strong predictive characteristics. As several other research teams pointed out, the potential risk factors of older age, high SOFA score, and d-dimer greater than 1 μg/L could help clinicians to identify patients with poor prognosis at an early stage[12]. Older patients (>65 years) with comorbidities and ARDS are at increased risk of death[14]. Additionally, during our experience of treating the patients, we have noticed that the T lymphocyte subsets are closely correlated to the patient's progress. We found that all the patients showed varying degrees of decline in T lymphocyte subsets at hospital admission. And the patient's condition improved or worsened with the rise or fall of the T lymphocyte subsets. We have also considered some other immune indicators and inflammatory indicators, as well as the blood routine and lung CT, but they all have certain limitations. Immune indicators and inflammatory indicators such as thyrotropin and white blood cells were not as sensitive as the T lymphocyte subsets; the blood routine is not specific enough to differentiate the condition of the patients[11], and although lung CT is an accurate way to assess the patient's condition, it cannot be repeated too often especially for those critical patients who needed oxygen and ventilator. In conclusion, we decided to focus our research on patient's age, underlying disease status, and the T lymphocyte subsets measures: Helper T cells, Suppressor T cells, and TH/TS (Helper T cells and Suppressor T cells ratio).

medRxiv preprint doi: https://doi.org/10.1101/2020.04.06.20056127; this version posted April 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Statistical analysis results and the interactive web tool

Among the 310 discharged cases, 155 were male and 155 were female. The average age was 56·4 years (SD 14·0). 107 (34·5%) of them had underlying diseases (the most common underlying disease is Hypertension, 23·9% for male and 20·0% for female). The average duration of the hospitalization was 11·1 days (SD 6·94), and it took 15·1 days (SD 10·3) on average for them to reach our hospital since their first symptoms (Table 1). Elderly patients tended to take longer to recover and they also had a higher probability to have underlying diseases (Figure 2a). The summary statistics for their baseline T lymphocyte subsets tests are the following: Total T cells percentage (mean 60·4%, SD 12·7%), Total T cells counts (mean 773, SD 549), Helper T cells percentage (37·2%, SD 10·6%), Helper T cells counts (mean 457, SD 342), Suppressor T cells percentage (mean 25·0%, SD 8·72%), Suppressor T cells counts (mean 297, SD 220), TH/TS (mean 1·71, SD 0·89), and Total Lymphocyte counts (mean 1160, SD 744) (Table 3).

Among the 30 death cases, 17 were male and 13 were female. The average age was 69·0 years (SD 7·87). 23 (76·7%) of them had underlying diseases (the most common underlying disease is also Hypertension, 64·7% for male and 38·5% for female). The average duration of the hospitalization was 15·1 days (SD 8·78), and it took 11·7 days (SD 8·20) on average for them to reach our hospital since their first symptoms (Table 2). The death event occurred more quickly among elderly patients (Figure 2b). The summary statistics for their baseline T lymphocyte subsets tests are the following: Total T cells percentage (mean 52·7%, SD 14·2%), Total T cells counts (mean 228, SD 168), Helper T cells percentage (33·8%, SD 11·8%), Helper T cells counts (mean 139, SD 98), Suppressor T cells percentage (mean 17·3%, SD 8·72%), Suppressor T cells counts (mean 80·9, SD 97·7), TH/TS (mean 2·41, SD 1·28), and Total Lymphocyte counts (mean 425, SD 254) (Table 4).

Two sample T-test between the discharged cases and the death cases in terms of their baseline T lymphocyte subsets measures yielded significant p values: Total T cells ($p < 2{\cdot}2e\text{-}16$), Helper T cells ($p < 2{\cdot}2e\text{-}16$), Suppressor T cells ($p = 1{\cdot}8e\text{-}14$), and TH/TS ($p = 0{\cdot}0066$) suggesting that patients with a poor prognosis had a more damaged immune system at the time of hospitalization (Figure 3).

Figure 4 are the scatterplots showing the pair-wise display of the three absolute T cell counts at the baseline: Helper T cells, Suppressor T cells, and the Total T cells (sum of the first two). Red dots indicate the death cases and green dots indicate the discharged cases. In all of the three scatterplots in Figure 4, there was a clear tendency that the death cases had lower cell counts than the discharged cases.

We also performed a multivariate logistic regression model using age, underlying disease status, and the baseline T lymphocyte subsets test as the predictors to predict the patient outcome (death or hospital discharge). The significant predictors are age (OR 1·05, p 0·04), underlying disease status (OR 3·42, p 0·02), Helper T cells on the log scale (OR 0·22, p 0·00), and TH/TS on the log scale (OR 4·80, p 0·00). The McFadden pseudo R-squared[16] for the logistic regression model is 0·35, suggesting the

medRxiv preprint doi: https://doi.org/10.1101/2020.04.06.20056127; this version posted April 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

model has a fair predictive power (The McFadden pseudo R-squared measure ranges from 0 to below 1, with values closer to 0 indicating lack of predictive power). Total T cells or Suppressor T cells, and Total Lymphocyte did not turn out to be significant predictors in our logistic regression model (Table 5).

We believe by looking at some of the patient's basic characteristics such as age and underlying diseases, together with the T lymphocyte subsets measures, could be a quick way to shed light on the patient's prognosis, during the time of pressure and emergency. In order for our findings to be more applicable for public health workers fighting at the frontline, we have developed an interactive web data visualization tool to implement the algorithm and made it accessible for the world at the following web address: https://rpubs.com/mindyfang/covid19.

Simply by accessing to the web address and inputting a new patient's information, the UI will show where the new patient is positioned (by a large blue dot) comparing to the reference panel of the 310 patients included in this research (Figure 5).

This data visualization tool applied a k-means clustering method to detect the differentiation in their baseline profile patterns among all patients. K-means clustering is a type of unsupervised learning method, enabling us to find and analyze the intrinsic underlying patient subgroups without any pre-defined subgroup labels. In real-world scenarios, clustering is a widely used technique for customer segmentation[17].

The variables used in the k-means clustering included age, underlying disease status, Helper T cells (log scale), Suppressor T cells (log scale), and the Helper T cells and Suppressor T cells ratio (log scale) for the following reasons: 1. Age, underlying disease status, Helper T cells, and TH/TS were significant predictors in our multivariate logistic regression model. Although Suppressor T cells was not statistically significant, it was used in the TH/TS ratio. In fact a principal component analysis[18] showed that all of the selected five indexes carry certain proportions of independent information (Figure 6a); 2. We did not use other lab tests such as the regular blood test items in our analysis because they are less differentiative than the T cell subset measures; and 3. We hope our interactive web tool could be utilized for quick use, therefore keeping as few input items as needed seems to be a more practical choice.

Figure 6b shows the k-means clustering result using the Wuhan Pulmonary Hospital data. After multi-dimensional data transformation, the algorithm separates the death group and the discharged group as shown in the graph. A proportion of the discharged cases had similar profiles with the death cases, which had made it difficult for the algorithm to differentiate them apart. However, by using this algorithm, it is possible to identify a large number of patients with relatively good prognosis.

We have also uploaded a dummy date set with de-identified and randomly modified patient data , as well as all the source code used for the current analysis as well as the interactive web application to:

medRxiv preprint doi: https://doi.org/10.1101/2020.04.06.20056127; this version posted April 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

https://github.com/mindy-fang/COVID-19. All of our source code were written with the R programming language and the interactive web application was developed with the shiny package[19-21] (Rstudio Version 1.2.1335, R version 3.6.0). Other fellow researchers can substitute the dummy data with their own data, or modify the source code to make their own applications. Meanwhile we will keep updating our reference panel as we include more patients, so that the algorithm would gain more and more statistical power over time.

Discussion

To our best knowledge, the current research is the largest retrospective study of COVID-19 patients with known clinical outcomes so far. Significant reductions in T cells are very common in severe COVID-19 patients. Age-dependent deficits in T cell and B cell functions and overproduction of type 2 cytokines may cause inadequate viral replication control and longer pro-inflammatory responses, which may lead to poor prognoses[22]. Lymphopenia is a prominent part of SARS-CoV infection and lymphocyte counts may be useful in predicting the severity and clinical outcomes[23-25]. We know that 2019-nCoV was once called Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-20)[26-27]. The level and the speed of T cell recovery are important factors of assessing disease prognoses and guiding early intervention in critically ill patients. In this study we have identified that older age, underlying diseases, and low T cell counts may be risk factors for poor clinical outcomes in COVID-10 positive patients. We have also implemented an interactive web tool to visualize a new patient's risk using these risk factors. When a new patient's data is entered through the web UI, it will be shown where he is as compared to the 310 patients included in this study.

It should be noted that our intension is not for the web tool to be considered as a 100% gold standard for determining the final clinical outcome of the patients. The visualization result simply suggests whether the new patient is more likely to recover or to have adverse outcomes. Another important significance of our tool is the monitoring of patient's progress in real time during the treatment. For example, the patient's longitudinal profile gradually shifting towards the red centroid (poor outcome) or the green centroid (good outcome) may provide insight on the patient's progress (Figure 5). However, the actual final prognosis of the patient depends on many other factors, including the starting time of the treatment, treatment compliance, degree of treatment, and so on.

Our algorithm has the following limitations: It may not predict accurately for younger patients, or patients with no symptoms, since our training data contains relatively old and severe patients; Secondly, the model was built based on a sample size of 340, which is not a large number. But we will keep updating our web application as we collect more data.

In addition, the reference panel used in our analysis were infected population in Wuhan, China. Although there was no clear evidence that the underlying mechanism of the T cell depletion under the

medRxiv preprint doi: https://doi.org/10.1101/2020.04.06.20056127; this version posted April 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

COVID-19 infection is similar or different across ethnicity groups, it is only natural to assume that the baseline values and the degree of T cell depletion would be slightly different. We found that the T-lymphocytes, B-lymphocytes, and NK cells did differ among people in different regions[28-30]. We encourage researchers around the world to download the source code and customize it with their own data, as they accumulate more experience and knowledge with patients from their own hospitals or regions.

To end our writing with the most recent update (March 13, 2020), the epidemic in Wuhan has gradually passed its peak. All the cabin hospitals were closed, and other hospitals in Wuhan started to return to their normal track. Medical volunteer teams have returned to their home cities. We give thanks to all the support that we have received during our most difficult time and hope the situation would improve quickly for other parts of the world.

Acknowledgments

Our research was supported by medical teams from Shanxi, Inner Mongolia and other parts of China. The authors would like to thank Prof Li Ming, Prof Zhu Qi, Prof Yang Chengqing, Prof Guo Guangyun, Prof Du Juan, Prof Du Ronghui for their technical support and Prof Chen Xianxiang for guidance in interpretation of the results.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.06.20056127; this version posted April 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

medRxiv preprint doi: https://doi.org/10.1101/2020.04.06.20056127; this version posted April 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Reference

1. WHO. Novel coronavirus – China. Jan 12, 2020. http://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/

2. WHO. Clinical management of severe acute respiratory infection when novel coronavirus (nCoV) infection is suspected. Jan 11, 2020. https://www.who.int/internal-publications-detail/clinical management-of-severe-acute-respiratory-infection-when-novelcoronavirus-(ncov)-infection-is-suspected (accessed Jan 19, 2020)

3. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). 2019, 16–24 (2020).

4. Coronavirus disease 2019 (COVID-19) Situation Report – 53

5. 5.Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H et al. First Case of 2019 Novel Coronavirus in the United States. N. Engl. J. Med. 929–936 (2020) doi:10.1056/nejmoa2001191

6. Phelan AL, Katz R, Gostin LO. The novel coronavirus originating in Wuhan, China: challenges for global health governance. JAMA 2020

7. Ling Y, Xu SB, Lin YX, Tian D, Zhu ZQ, Dai FH et al. Persistence and clearance of viral RNA in 2019 novel coronavirus disease rehabilitation patients. Chin. Med. J. (Engl). 133, E007–E007 (2020).

8. WHO. Clinical management of severe acute respiratory infection when Novel coronavirus (nCoV) infection is suspected: interim guidance. Jan 11, 2020.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.06.20056127; this version posted April 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

9. National Health Commission of the People's Republic of China. Chinese management guideline for COVID-19 (version 6.0). Feb 19, 2020.

10. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. JAMA 2020

11. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 497–506 (2020) doi:10.1016/s0140-6736(20)30183-5.

12. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. 6736, 1–9 (2020).Lancet. 2020 Mar 11. pii: S0140-6736(20)30566-3. doi: 10.1016/S0140-6736(20)30566-3

13. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 395, 507–513 (2020).

14. Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. Lancet Respir Med 2020.

15. Xu Z, Shi L, Wang Y, Zhang J, Huang L, Zhang C et al. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. Lancet Respir. 2600, 19–21 (2020).

16. LANGER, Wolfgang. The assessment of fit in the class of logistic regression models: A pathway out of the Jungle of Pseudo-R^2s. Halle: Institut für Soziologie, 2000.

17. Pascal, C., Ozuomba, S. & kalu, C. Application of K-Means Algorithm for Efficient Customer Segmentation: A Strategy for Targeted Customer Services. Int. J. Adv. Res. Artif. Intell. 4, 40–44 (2015).

18. Lever, J., Krzywinski, M. & Altman, N. Points of Significance: Principal component analysis. Nat. Methods 14, 641–642 (2017).

19. Beely, C. Web Application Development with R Using Shiny 1st edn. (PACKT Publishing, Birmingham, United Kingdom, 2013)

20. R Core Team. R: A language and Environment for Statistical Computing. (R Foundation for Statistical Computing, Vienna, Austria, 2014).

21. RStudio Inc. shiny: Web Application Framework for R. R package version 0.10.1. (2014).

22. Wang, D., Shao, Y., Zhang, X., Lu, G. & Liu, B. IL-23 and PSMA-targeted duo-CAR T cells in Prostate Cancer Eradication in a preclinical model. J. Transl. Med. 18, 23 (2020).

medRxiv preprint doi: https://doi.org/10.1101/2020.04.06.20056127; this version posted April 16, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

23. He Z, Zhao C, Dong Q, Zhuang H, Song S, Peng G et al. Effects of severe acute respiratory syndrome (SARS) coronavirus infection on peripheral blood lymphocytes and their subsets. Int. J. Infect. Dis. 9, 323–330 (2005).

24. Chen J, Lau YF, Lamirande EW, Paddock CD, Bartlett JH, Zaki SR et al. Cellular Immune Responses to Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) Infection in Senescent BALB/c Mice: CD4+ T Cells Are Important in Control of SARS-CoV Infection. J. Virol. 84, 1289–1301 (2010).

25. Peng H, Yang LT, Wang LY, Li J, Huang J, Lu ZQ et al. Long-lived memory T lymphocyte responses against SARS coronavirus nucleocapsid protein in SARS-recovered patients. Virology, 2006, 351.2: 466-475.

26. Aziz N, Jamieson BD, Quint JJ, Martinez-Maza O, Chow M, Detels R.Longitudinal Intra- and Inter-individual variation in T-cell subsets of HIV-infected and uninfected men participating in the LA Multi-Center AIDS Cohort Study.Medicine (Baltimore). 2019 Oct; 98(41): e17525. doi: 10.1097/MD.0000000000017525.

27. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus ADME, Fouchier RAM. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N Engl J Med 2012; 367: 1814–20

28. Kokuina E, Breff-Fonseca MC, Villegas-Valverde CA, Mora-Díaz I. Normal Values of T, B and NK Lymphocyte Subpopulations in Peripheral Blood of Healthy Cuban Adults. MEDICC Rev. 2019 Apr-Jul;21(2–3):16–21.

29. Garrido-Rodríguez D, Ávila-Ríos S, García-Morales C, Valenzuela-Ponce H, Ormsby C, Reyes-Gopar H et al.Killer cell immunoglobulin-like receptor and human leukocyte antigen gene profiles in a cohort of HIV-infected Mexican Mestizos. Immunogenetics. 2016 Oct;68(9):703-17. doi: 10.1007/s00251-016-0920-3. Epub 2016 May 30.

30. Haverkos BM, Pan Z, Gru AA, Freud AG, Rabinovitch R, Xu-Welliver M et al. Extranodal NK/T Cell Lymphoma, Nasal Type (ENKTL-NT): An Update on Epidemiology, Clinical Presentation, and Natural History in North American and European Cases. Curr Hematol Malig Rep. 2016 Dec;11(6):514–527. doi: 10.1007/s11899-016-0355-9.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.06.20056127; this version posted April 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Figure 1: Process Flow at the Wuhan Pulmonary Hospital

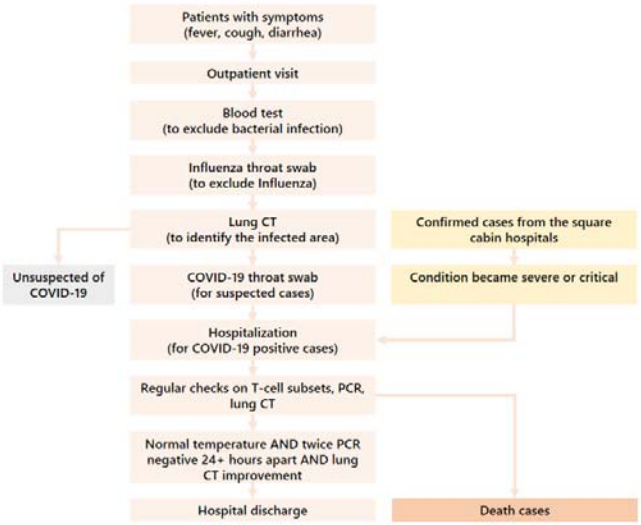


Table 1: Summary table of the discharged patient characteristics and the duration of hospitalization

	Female (n=155)	Male (n=155)	Overall (n=310)
Age			
Mean (SD)	57.3 (13.8)	55.5 (14.2)	56.4 (14.0)
Median [IQR]	58.0 [48.0, 68.0]	56.0 [44.5, 66.0]	57.0 [46.0, 66.0]
Underlying Diseases*			
All diseases	48 (31.0%)	59 (38.1%)	107 (34.5%)
Hypertension	31 (20.0%)	37 (23.9%)	68 (21.9%)
Cardiovascular Disease	9 (5.8%)	10 (6.5%)	19 (6.1%)
COPD	3 (1.9%)	7 (4.5%)	10 (3.2%)
Diabetes	3 (1.9%)	3 (1.9%)	6 (1.9%)
Chronic Liver Disease	3 (1.9%)	3 (1.9%)	6 (1.9%)
Hyperlipidemia	2 (1.3%)	2 (1.3%)	4 (1.3%)
Duration of Hospitalization (days)			
Mean (SD)	10.4 (5.89)	11.9 (7.78)	11.1 (6.94)
Median [IQR]	9.00 [5.00, 14.0]	10.0 [6.00, 15.5]	9.00 [5.00, 14.0]
From the First Symptom to Hospitalization (days)			
Mean (SD)	15.6 (10.5)	14.6 (10.1)	15.1 (10.3)
Median [IQR]	12.0 [8.00, 20.0]	11.0 [7.00, 18.5]	12.0 [7.00, 20.0]
From the First Symptom to Hospital Discharge (days)			
Mean (SD)	26.0 (9.78)	26.6 (10.6)	26.3 (10.2)
Median [IQR]	25.0 [18.0, 33.5]	24.0 [19.0, 34.0]	24.5 [19.0, 34.0]

* Some patients had a combination of underlying diseases

medRxiv preprint doi: https://doi.org/10.1101/2020.04.06.20056127; this version posted April 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Table 2: Summary table of the death patient characteristics and the duration of hospitalization

	Female (n=13)	Male (n=17)	Overall (n=30)
Age			
Mean (SD)	68.9 (6.42)	69.1 (9.01)	69.0 (7.87)
Median [IQR]	69.0 [65.0, 71.0]	69.0 [62.0, 77.0]	69.0 [62.5, 73.5]
Underlying Diseases*			
All diseases	9 (69.2%)	14 (82.4%)	23 (76.7%)
Hypertension	5 (38.5%)	11 (64.7%)	16 (53.3%)
Cardiovascular Disease	2 (15.4%)	3 (17.6%)	5 (16.7%)
COPD	1 (7.7%)	0 (0%)	1 (3.3%)
Diabetes	0 (0%)	2 (11.8%)	2 (6.7%)
Chronic Liver Disease	1 (7.7%)	1 (5.9%)	2 (6.7%)
Hyperlipidemia	0 (0%)	0 (0%)	0 (0%)
Duration of Hospitalization (days)			
Mean (SD)	15.2 (9.93)	15.0 (8.12)	15.1 (8.78)
Median [IQR]	12.0 [7.00, 21.0]	13.0 [10.0, 17.0]	12.5 [9.25, 20.3]
From the First Symptom to Hospitalization (days)			
Mean (SD)	13.5 (11.2)	10.2 (4.72)	11.7 (8.20)
Median [IQR]	10.0 [7.00, 14.0]	9.00 [7.00, 13.0]	10.0 [7.00, 13.0]
From the First Symptom to Death (days)			
Mean (SD)	28.7 (12.2)	25.2 (8.00)	26.7 (10.0)
Median [IQR]	27.0 [23.0, 38.0]	23.0 [20.0, 28.0]	25.5 [20.3, 30.5]

* Some patients had a combination of underlying diseases

Table 3: Summary table of the initial T cell test of the discharged patients

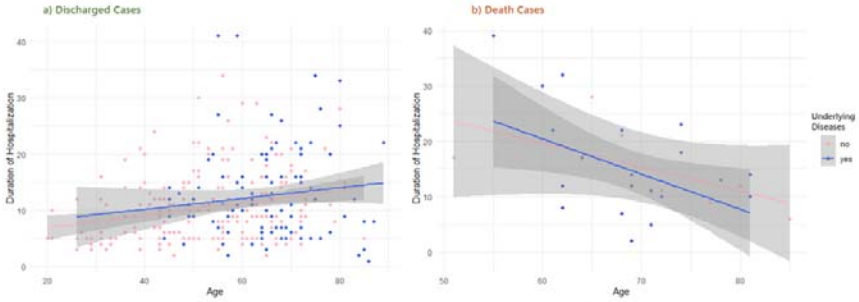
	Female (n=155)	Male (n=155)	Overall (n=310)
Total T Cells Percentage (%)			
Mean (SD)	63.1 (12.3)	64.8 (13.0)	64.0 (12.7)
Median [IQR]	63.9 [55.0, 73.0]	67.0 [58.8, 73.9]	66.2 [56.7, 73.8]
Total T Cells Counts (cells/ul)			
Mean (SD)	737 (478)	809 (611)	773 (549)
Median [IQR]	666 [318.8, 1108.0]	695 [317.9, 1101.9]	678 [317.4, 1105.1]
Helper T Cells Percentage (%)			
Mean (SD)	38.1 (10.6)	36.4 (10.5)	37.2 (10.6)
Median [IQR]	38.7 [31.1, 45.6]	36.8 [30.2, 43.6]	37.5 [31.0, 44.5]
Helper T Cells Counts (cells/ul)			
Mean (SD)	448 (300)	466 (380)	457 (342)
Median [IQR]	403 [188.7, 657.1]	366 [176.1, 645.9]	370 [182.8, 651.0]
Suppressor T Cells Percentage (%)			
Mean (SD)	23.4 (7.06)	26.6 (9.88)	25.0 (8.72)
Median [IQR]	23.1 [19.1, 28.3]	26.4 [20.1, 31.8]	24.2 [19.6, 29.9]
Suppressor T Cells Counts (cells/ul)			
Mean (SD)	269 (188)	324 (244)	297 (220)
Median [IQR]	229 [115.2, 376.4]	267 [130.0, 477.6]	249 [118.8, 425.1]
Th/Ts(TH/TS)			
Mean (SD)	1.82 (0.893)	1.61 (0.886)	1.71 (0.894)
Median [IQR]	1.67 [1.14, 2.17]	1.49 [1.02, 2.00]	1.59 [1.06, 2.04]
Total Lymphocyte Counts (cells/ul)			
Mean (SD)	1120 (643)	1200 (833)	1160 (744)
Median [IQR]	975 [561.3, 1582.1]	1030 [532.7, 1628.3]	1020 [536.3, 1603.7]

medRxiv preprint doi: https://doi.org/10.1101/2020.04.06.20056127; this version posted April 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Table 4: Summary table of the initial T cell test of the death patients

	Female (n=13)	Male (n=17)	Overall (n=30)
Total T Cells Percentage (%)			
Mean (SD)	52.4 (14.5)	52.9 (14.3)	52.7 (14.2)
Median [IQR]	52.0 [45.2, 63.7]	54.3 [43.2, 65.2]	53.6 [43.7, 64.4]
Total T Cells Counts (cells/ul)			
Mean (SD)	250 (137)	211 (190)	228 (168)
Median [IQR]	213 [140.0, 362.1]	111 [90.8, 280.1]	170 [94.0, 339.1]
Helper T Cells Percentage (%)			
Mean (SD)	33.7 (13.0)	33.9 (11.2)	33.8 (11.8)
Median [IQR]	37.2 [19.0, 43.5]	35.8 [24.1, 42.6]	36.2 [24.0, 43.2]
Helper T Cells Counts (cells/ul)			
Mean (SD)	162 (111)	121 (86.3)	139 (98.0)
Median [IQR]	126 [72.5, 231.8]	76.3 [62.8, 136.6]	115 [62.8, 195.1]
Suppressor T Cells Percentage (%)			
Mean (SD)	16.8 (6.70)	17.7 (10.2)	17.3 (8.72)
Median [IQR]	19.3 [12.0, 20.9]	15.5 [12.1, 19.9]	16.3 [12.0, 20.8]
Suppressor T Cells Counts (cells/ul)			
Mean (SD)	76.8 (48.9)	84.1 (124)	80.9 (97.7)
Median [IQR]	72.2 [47.1, 81.1]	45.5 [21.7, 65.4]	56.8 [26.8, 80.8]
Th/Ts(TH/TS)			
Mean (SD)	2.41 (1.36)	2.41 (1.26)	2.41 (1.28)
Median [IQR]	2.17 [1.65, 2.97]	2.32 [1.83, 2.90]	2.25 [1.70, 2.93]
Total Lymphocyte counts (cells/ul)			
Mean (SD)	475 (232)	386 (270)	425 (254)
Median [IQR]	382 [344.9, 653.3]	305 [173.8, 542.1]	375 [211.0, 564.2]

Figure 2: Duration of Hospitalization vs Age

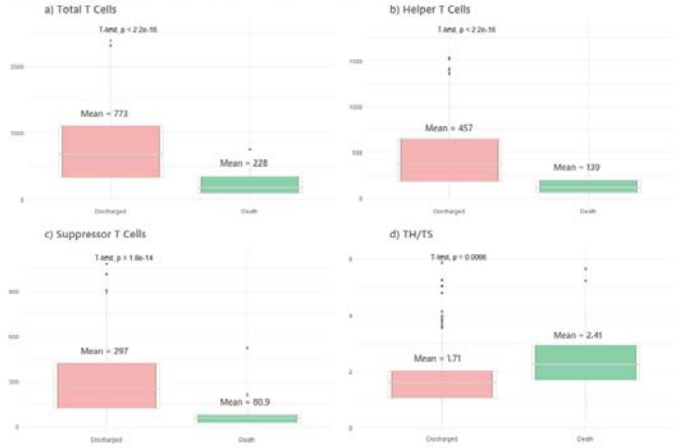


a) Duration of the hospitalization plotted vs Age, for the discharged group. Pink dots indicate the patients with underlying medical conditions and the blue dots indicate those without. There are two observations from this chart: 1. Elderly patients tend to have underlying diseases than the younger ones; 2. The duration of the hospitalization tends to be longer among the elderly patients. The straight lines and the shaded areas show the regression line and the 95% confidence interval.

b) Duration of the hospitalization plotted vs Age, for the death group. Pink dots indicate the patients with underlying medical conditions and the blue dots indicate those without. Elderly patients die more quickly than the younger ones. The straight lines and the shaded areas show the regression line and the 95% confidence interval.

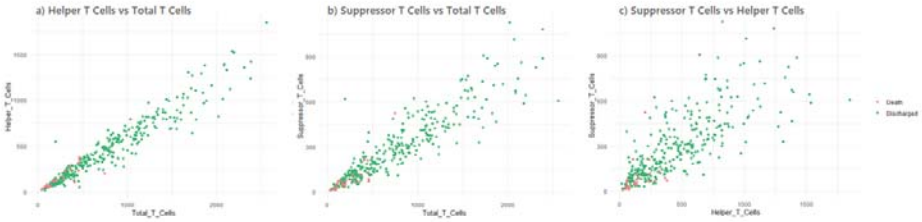
medRxiv preprint doi: https://doi.org/10.1101/2020.04.06.20056127; this version posted April 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Figure 3: Comparison the T-cell Subsets between the Discharged Cases and the Death Cases



a) Boxplot of the difference in the total T cells between the last and the first T-cell Subsets tests during the hospitalization. T-test comparing the means of the two groups returns a p value less than 2.2e-16.
b) Boxplot of the difference in the helper T cells between the last and the first T-cell Subsets tests during the hospitalization. T-test comparing the means of the two groups returns a p value less than 2.2e-16.
c) Boxplot of the difference in the suppressor T cells between the last and the first T-cell Subsets tests during

Figure 4: First Test Result after Hospitalization (Baseline)



a) Helper T cells plotted vs total T cells, red dots indicate the death group and the green dots indicate the discharged group. The death cases have lower helper T cells and total T cells than the discharged group
b) Suppressor T cells plotted vs total T cells, red dots indicate the death group and the green dots indicate the discharged group. The death cases have lower suppressor T cells and total T cells than the discharged group
c) Suppressor T cells plotted vs helper T cells, red dots indicate the death group and the green dots indicate the discharged group. This is a different view from a) and b) showing the same tendency.

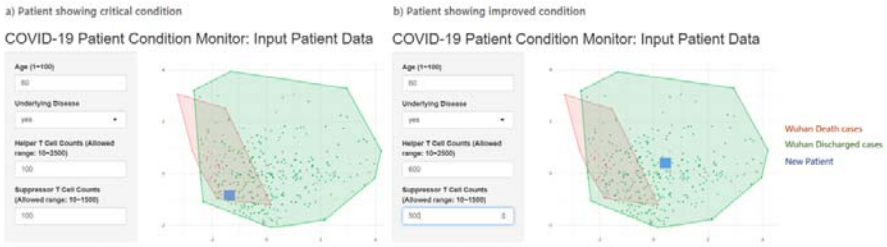
medRxiv preprint doi: https://doi.org/10.1101/2020.04.06.20056127; this version posted April 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Table 5: Multivariate Logistic Regression Model Result using Death as the Outcome

	Estimate	Std. Error	z value	Odds Ratio	P value
Age	0.05	0.02	2.02	1.05	0.04
Underlying Diseases	1.23	0.51	2.40	3.42	0.02
log(Helper T cells)	-1.47	0.30	-4.95	0.22	0.00
log(TH/TS)	1.57	0.44	3.58	4.80	0.00

McFadden pseudo R-squared: 0.35

Figure 5: Screen Shots of the Interactive Web Tool

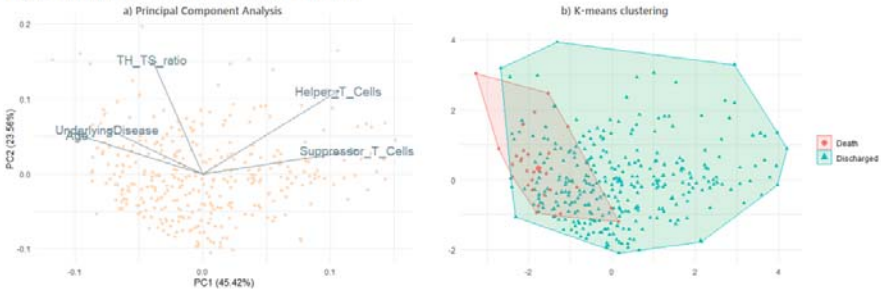


a) A 60 year old patient with underlying diseases, displayed by the large blue square, showing a critical condition with Helper T cell counts of 100 and Suppressor T cell counts of 100

b) The same patient showing improved condition, with Helper T cell counts increased to 600 and Suppressor T cell counts increased to 500

medRxiv preprint doi: https://doi.org/10.1101/2020.04.06.20056127; this version posted April 16, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Figure 6: Principal Component Analysis of the T-cell subsets



a) The three selected T-cell subsets indexes and age, underlying disease indicator were plotted against their first two principal components, which carries 45.42% and 23.56% of the total information respectively. The dots in the background represent each patients. There was very little overlap among the six variables suggesting them carrying exclusive information.

b) Multi-dimentional transformation of the above five variables for the 340 patients in our study, so that the underlying patterns can be recognized to the maximum extent. A proportion of the discharged cases had similar T-cell subsets profiles with the death cases which made it difficult to differentiate between them.