medRxiv preprint doi: https://doi.org/10.1101/2020.04.22.20076190; this version posted April 27, 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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 .

Characterization of clinical progression of COVID-19 patients in Shenzhen, China

Qifang Bi[1, 2, #], Chengcheng Hong[3,4, #], Juan Meng[1, #], Zhenke Wu[5,6], Pengzheng Zhou[4], Chenfei Ye[4], Binbin Sun[3,4], Lauren M. Kucirka[7], Andrew S. Azman[2], Tong Wang[3], Jiancong Chen[4], Zhaoqin Wang[1], Lei Liu[1], Justin Lessler[2,*], Jessie K. Edwards[8,*], Ting Ma[3,4,*], Guoliang Zhang[1,*]

1. National Clinical Research Center for Infectious Diseases, Shenzhen Third People's Hospital, Southern University of Science and Technology, Shenzhen, Guangdong, China
2. Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA
3. Harbin Institute of Technology at Shenzhen, Shenzhen, Guangdong, China
4. Peng Cheng Laboratory, Shenzhen, Guangdong, China
5. University of Michigan School of Public Health, Ann Arbor, MI, USA
6. Michigan Institute for Data Science, University of Michigan, Ann Arbor, MI, USA
7. Johns Hopkins University School of Medicine, Baltimore, MD, USA
8. University of North Carolina Gillings School of Global Public Health, Chapel Hill, NC, USA

[#]These authors contributed equally.

*Corresponding authors:
Dr. Justin Lessler, justin@jhu.edu;
Dr. Jessie K. Edwards, jessedwards@unc.edu;
Dr. Ting Ma, tma@hit.edu.cn;
Dr. Guoliang Zhang, szdsyy@aliyun.com;

Abstract

The COVID-19 pandemic has stressed healthcare care systems throughout the world. Understanding clinical progression of cases is a key public health priority that informs optimal resource allocation during an emergency. Using data from Shenzhen, China, where all cases were monitored in hospital and symptom profiles and clinical and lab results were available starting from early stages of clinical course, we characterized clinical progression of COVID-19 cases and determined important predictors for faster clinical progression to key clinical events and longer use of medical resources. Epidemiological, demographic, laboratory, clinical, and outcome data were extracted from electronic medical records. We found that those who progressed to the severe stage, developed acute respiratory distress syndrome, and were admitted to the intensive care unit (ICU) progressed on average 9.5 days (95%CI 8.7,10.3), 11.0 days (95%CI 9.7,12.3), and 10.5 days (95%CI 8.2,13.3) after symptom onset, respectively. We estimated that patients who were admitted to ICUs remained there for an average of 34.4 days (95%CI 24.1,43.2) and the average time on a ventilator was 28.5 days (95%CI 20.0,39.1) among those requiring mechanical ventilation. The median length of hospital stay was 21.3 days (95%CI, 20.5, 22.2) for the mild or moderate cases who did not progress to the severe stage, but increased to 52.1 days (95%CI, 43.3, 59.5) for those who required ICU admission. Clear characterization of clinical progression informs planning for healthcare resource allocation during COVID-19 outbreaks and provides a basis that helps assess the effectiveness of new treatment and therapeutics.

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/10.1101/2020.04.22.20076190; this version posted April 27, 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 .

Main

The epidemic of coronavirus SARS-CoV-2 has led to 1.6 million infections and over 100,000 deaths over 4 months after the first case was detected[1], causing severe shortage of essential medical supplies and equipment, medical staff, and hospital beds [2]. Complementary to data from COVID-19 epicenters like Wuhan (China) or Lombardy (Italy), data from places where healthcare capacity was not exceeded and patients were treated early and free of charge has the potential to shed light on the near complete clinical trajectory of cases. Clear characterization of COVID-19 clinical trajectory under the current standard of care informs planning for healthcare resource allocation during COVID-19 outbreaks and provides a basis that helps assess the effectiveness of new treatment and therapeutics.

Here, we use rich data on clinical progression of all COVID-19 cases diagnosed and treated in the only designated hospital in Shenzhen, China. Because all clinically confirmed cases, including a sizable portion detected through contact tracing, were required to be hospitalized for isolation purposes regardless of their clinical presentation and symptom profile, this dataset allows us to examine clinical progression of cases without the considerable selection bias typically seen in hospital-based studies. We estimate time from symptom onset to key clinical events, such as first clinical diagnosis, progression to severe clinical stages, development of acute respiratory distress syndrome (ARDS), admission to the critical care unit (ICU), and discharge. We also estimate duration hospitalized, in the ICU, and on ventilators. We determine the key predictors of faster clinical progression to a series of clinical events and longer use of healthcare resources.

Methods

Study design and study population

This single-centre, observational study was conducted at Shenzhen Third People's Hospital, which is the designated hospital to treat all patients with COVID-19 in Shenzhen. We prospectively collected data of all 420 patients diagnosed and hospitalized with COVID-19 in Shenzhen between January 11th and March 10th 2020, regardless of their clinical severity and symptom profile.

Data collection

Epidemiological, demographic, laboratory, clinical, and outcome data were extracted from electronic medical records using a standardised data collection form. All information was updated as of April 7th 2020. Data were reviewed by multiple reviewers (BS, JC, JZ, PZ), and any disagreement between reviewers was resolved by consultation with an attending physician (JM). Data from patients with severe or critical clinical assessment were extracted by an attending physician (JM) and reviewed with a reviewer (CH) to ensure data quality.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.22.20076190; this version posted April 27, 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 .

We obtained information on demographic characteristics, mode of detection, and date of onset for each symptom. Date of symptom onset before admission was self-reported and date of symptom onset after admission was recorded by an attending physician. Clinical severity was defined based on guidelines issued by the National Health Commission of the People's Republic of China, and the severity definition was generally consistent over time (see supplemental table 3 for guidelines in each version) [7]. We recorded clinical severity at initial diagnosis and date of severity progression. Clinical severity was assessed daily when mild or moderate and was assessed twice a day when severe or critical.

Dates of ICU admission and discharge were recorded, as well as dates beginning and ending invasive ventilator use. Patients were eligible for discharge from the hospital or transfer to a non-COVID ward for treatment if they met all of the following: 1) no fever for over 3 days, 2) drastic improvement in respiratory symptoms, 3) pulmonary imaging showing significant reduction in inflammation, and 4) two consecutive negative RT-PCR results from respiratory sampling conducted over one day apart [7]. We counted time in non-COVID wards for treatment of COVID-19 related complications towards the duration of hospitalization.

We recorded patients' self-reported medical history (See table 1 for a list of baseline comorbidities). We reviewed laboratory results and created binary variables indicating presence of any abnormalities and the date when such abnormalities were detected (see Table1 for a complete list of lab indicators and abnormality cutoff). The x-ray computed tomography (CT) results were extracted from radiological examinations. We recorded the lowest cycle threshold values from the available RT-PCR testings and the date when the testing was performed. We also recorded complications developed during hospitalization.

The primary endpoints in this study include patients' time from symptom onset to clinical progression beyond the moderate stage, ICU admission, invasive ventilator use, and discharge. We also examined other endpoints including time to when PaO_2/FiO_2 dropped under 300mmHg and time to developing ARDS.

Statistical analyses

We estimated cumulative incidence of developing key clinical events in the presence of competing risks (i.e., death and hospital discharge) using the Aalen-Johansen estimator [8,9]. Outcomes included progression to severe stage, low PaO2/FiO2 ratio, ARDS, ICU admission, use of invasive ventilator, and hospital discharge.

We estimated patients' duration of hospitalization by calculating the area above the cumulative incidence of hospital discharge or death, which was estimated using the Aalen-Johansen estimator [8,9]. In addition, we estimated the time patients admitted to the ICU ultimately spent in the ICU as the area between the cumulative incidence functions for ICU entry and ICU exit among those who were admitted into ICU. We treated death while in the ICU and ICU discharge as a composite event when estimating cumulative incidence of ICU exit. Similarly, we estimated the duration of invasive ventilator use among patients requiring ventilation, as the area between

medRxiv preprint doi: https://doi.org/10.1101/2020.04.22.20076190; this version posted April 27, 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 .

the cumulative incidence functions for ventilator initiation and ventilator discontinuation. We treated death during ventilator use and withdrawal of ventilator support as a composite event when estimating ventilator discontinuation. We compared times in each state estimated using the nonparametric approach described above with times estimated using a parametric accelerated failure time models to examine improvements in precision seen when invoking a parametric approach (see Text S2 for detailed method). We used bootstrap simulation to construct confidence intervals (2000 bootstrap simulations for time to recovery and time to ARDS, and 200 bootstrap simulations for time to ICU entry/discharge and ventilator use/withdrawal). 95% confidence intervals were the 2.5th and 97.5th percentiles of the distribution of point estimates from the bootstrap samples.

We used competing risk regressions according to the methods of Fine and Gray [10] to estimate subdistribution hazard ratios comparing the rate of clinical progression between subgroups that were defined a priori (See Table 3 for the list of subgroups). We compared the rate of clinical progression to severe stage, ICU admission, ARDS, and hospital discharge between subgroups. Except for the models where time to hospital discharge was the outcome, hospital discharge and death were treated as competing events and end of study as administrative censoring for those still in treatment.

We used a flexible approach to stratify cases into three risk strata for the purpose of visualizing different clinical trajectories. We constructed a random survival forest model (RSF) and divided cases into low, medium, and high risk groups based on tertiles of RSF out-of-bag predictions. The candidate predictors used in the RSF model included 1) demographic information, 2) baseline comorbidities, 3) symptom profile, lab and CT results within 5 days of any symptom onset. Hospital discharge was treated as a competing event. All four deaths occurred after cases progressed beyond the severe stage, thus they were not treated as competing events. We calculated AUC over time since symptom onset (tAUC), providing a measure of model performance across all possible classification thresholds and based on the observed number of cases entering the severe stage by each time point [11]. We fit 1000 trees, considered 5 random splits for each candidate splitting variables, and used the log-rank test as the split function.

Results

Characteristics of cases and description of baseline comorbidities

Four-hundred and twenty cases were admitted and hospitalized to Shenzhen Third People's Hospital between January 11th and March 10th, 2020 (Figure S1). Fifteen percent (63/420) of cases were detected through contact tracing (Table 1). On average, the first clinical diagnosis occurred 1.9 days (95% CI 1.6,2.3) and hospitalization occurred 4.2 days (95%CI 3.8,4.6) after symptom onset.

Of the 420 cases, there were approximately equal numbers of males (47.6%, n=200) and females (52.4%, n=220) (Table 1). A large portion (38.8%, n=163) were under the age of 40 and the majority (83.1%, n=349) were detected through symptom-based surveillance. 21.9%

medRxiv preprint doi: https://doi.org/10.1101/2020.04.22.20076190; this version posted April 27, 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 .

(92/420) of cases had at least one self-reported comorbidity on admission, with hypertension (n=49) and diabetes (n=24) being the most prevalent.

Fever, cough, sputum production were the most common initial symptoms, with 68.6% (288) of patients showing fever, 60.0% (252) with cough, and 31.2% (131) with sputum production within 5 days of initial symptom onset (see supplemental figure for onset time distribution of each symptom). 25.2% (106) of patients never had fever, and 26.9% (113) of patients never had cough.

At the initial clinical assessment, 23 patients (5.5%) were clinically mild, the vast majority (93.8%, n=394) were moderate, and only 3 patients were clinically severe or critical (Figure S2).

Clinical progression of cases

We estimated the proportion of the initially mild or moderate cases in each stage (mild/moderate, severe, ICU, death or discharge) over time following symptom onset, taking into account patients both transitioning into and out of each stage (Figure 1A). The total number of patients in the severe stage reached its peak 12 days after symptom onset. Among the 417 patients who were classified as mild or moderate at the time of initial assessment, 21.6% (90/417) progressed to the severe stage. 9.6% (95%CI, 6.8%, 12.4%) progressed to the severe stage within 7 days after symptom onset, and 20.4% (95%CI, 16.5%, 24.3%) progressed within 14 days (Figure 1 and 2). Those who progressed to the severe stage progressed on average 9.5 days (95%CI 8.7,10.3) after symptom onset. Among the 417 patients who were classified as mild or moderate at the time of initial assessment, 8.6% (36/417) developed ARDS. 2.6% (95%CI 1.1%,4.2%) developed ARDS within 7 days from symptom onset, and 7.7% (95%CI 5.1%, 10.2%) within 14 days. Those who developed ARDS developed ARDS on average 11.0 days (95%CI 9.7,12.3) after symptom onset.

As of April 7th, 19 patients had been admitted to ICU, among which 18 patients required invasive mechanical ventilation support, 4 patients died (1 patient died after initial hospital discharge with viral clearance), and 4 patients remained hospitalized in critical condition. We estimated that among the 417 patients who were classified as mild or moderate at the time of initial assessment, 3.3% (95%CI, 1.6%, 5.1%) of patients required ICU admission within 14 days from symptom onset (same for patients who required mechanical ventilators support). Those who required ICU admission were admitted into ICU on average 10.5 days (95%CI 8.2,13.3) after symptom onset. Using data from the 19 patients who were admitted into ICU, we estimated the average time in ICU was 34.4 days (95%CI 24.1,43.2) (Table 2). Using data from the 18 patients who required mechanical ventilator support, we estimated the average time on a ventilator was 28.5 days (95%CI 20.0,39.1).

The median length of hospital stay was 21.3 days (95%CI, 20.5, 22.2) for the mild or moderate cases who did not progress to the severe stage, and increased to 30.3 days (95%CI, 26.7, 31.4) for cases who reached the severe stage but did not enter ICU and 52.1 days (95%CI, 43.3, 59.5) for the cases who required ICU admission. Of note, patients in Shenzhen were required to

medRxiv preprint doi: https://doi.org/10.1101/2020.04.22.20076190; this version posted April 27, 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 .

be hospitalized for about 2 weeks for isolation; 21% (86/417) of the cases that were initially mild or moderate were discharged within 14 days from hospitalization, 34% (140/417) within 16 days. The duration of hospitalization for the mild or moderate cases was likely inflated as a result. All patients who were clinically mild at the time of initial assessment stayed mild until discharge (Figure S1).

Characteristics associated with difference in clinical progression

We then identified a priori-defined patients' characteristics that were associated with faster clinical progression. We found that having hypertension and diabetes at baseline was strongly associated with faster clinical progression to various clinical events, including progression to the severe stage (sHR=3.2, 95%CI 2.1,5.1 for hypertension and sHR=3.0, 95%CI 1.6,5.4 for diabetes), to developing ARDS (sHR=3.6, 95%CI 1.8,6.9 for hypertension and sHR=5.5, 95%CI 2.7,11.3 for diabetes), and to ICU admission (sHR=4.6, 95%CI 1.8,11.6 for hypertension and sHR=4.7, 95%CI 1.6,14.1 for diabetes) (Table 3). Having more baseline comorbidities was also associated with a higher rate of clinical progression to these events (Table 3). Although many lab abnormalities measured within 5 days of symptom onset were strongly predictive of faster clinical progression, including low lymphocyte count, low platelet count, high concentration of C-reactive protein, and high concentration of D-dimer, notably, a low PaO2/FiO2 ratio close to symptom onset was very strongly associated with faster clinical progression. We observed a 22.3 times (95%CI 8.4, 58.8) increase in the subdistribution hazard of ICU admission among those with early measures of low PaO2/FiO2 ratio (Table 3).

Older age was one of the most important predictors of faster clinical progression (Table 3). All four patients who died were male over the age of 60. About half (46.3%, 50/108) of cases aged 60 or above progressed to the severe or critical stage (Table 1). Although the vast majority of those under the age of 40 did not progress beyond the moderate stage, 6% (9/163) of cases in this younger age group became clinically severe or critical and none of them had any known underlying comorbidities.

Sex was strongly associated with clinical progression of cases, though the difference by sex was mostly driven by the difference among older patients. We did not observe a significant difference in time to ICU admission between males and females under the age of 60 (sHR = 0.7, 95%CI 0.1,4.3 comparing males with reference to females; Table S1, Figure 2b). However, we observed a noteable difference in progression to require ICU admission between males and females aged 60 or above; males in this older age group had a 10-fold increase in the subdistribution hazard of ICU admission compared to females in the same age group (sHR=10.5, 95%CI 1.0, 108.6), despite the similar sex-specific age distribution in this age group (males: mean 67, IQR 63,69 vs. females: mean 65, IQR 62,66) (Table S1). Differences in baseline comorbidities between older males and females did not explain the disparity; after adjusting for having any underlying comorbidity, the subdistribution hazard ratio remained unchanged (Table S1). Similarly, we found that males aged 60 or above had a lower rate of hospital discharge compared to females in the same age group, and no significant difference by

medRxiv preprint doi: https://doi.org/10.1101/2020.04.22.20076190; this version posted April 27, 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 .

sex in the younger age group (Figure 2c). However, we did not find significant disparities in clinical progression to severe stage or to developing ARDS by sex in any age group (Table S1).

Using linear regression, we found that the minimum RT-PCR cycle threshold values for the severe cases were significantly lower than the mild cases after adjusting for time of sample collection with reference to symptom onset. We observed a general trend of lower minimum cycle threshold values in patients with more severe clinical presentation and in patients in the older age group though the association with age was not statistically significant (Figure S5).

Risk strata for clinical progression

Based on the random survival forest results, the most important predictors of faster progression to the severe stage were low PaO2/FiO2 ratio, low platelet count, and high C-reactive protein concentration (Figure S8a). We observed very different clinical trajectories of patients in each risk group (Figure 1b), highlighting the effectiveness of risk stratification produced by the RSF model. In the low-risk group, no case required ICU admission and only 3% (95%CI, 0.1%,6%) of cases became severe within 14 days from symptom onset. Whereas in the high-risk group, we estimated that 43% (95%CI, 35, 52) cases became severe and 9% (95%CI, 4%,14%) required ICU admission within 14 days from symptom onset. All but one case that required ICU admission were classified into the high-risk group, and we estimated that the duration in ICU for those in the high-risk group was 35.1 days (95%CI, 26.1,45.1) (Table 2, Figure S8). Risk stratification produced by a RSF model that excluded lab and CT results from the candidate predictors also differentiated clinical trajectory of patients well (Figure S7, Figure S8 c-d).

Conclusions

The analysis of clinical data from COVID-19 patients in Shenzhen Third People's hospital provides insights into clinical progression of cases starting early in the course of infection. We estimate the proportion of cases in each severity stage over 80 days following symptom onset. We present patient characteristics associated with faster clinical progression and longer use of medical resources.

Because treatment of COVID-19 was free of charge and hospitalization for all cases was mandatory in Shenzhen, we expect that our results are not strongly affected by the selection bias that plagues many hospital-based studies. Therefore, our results provide a clear picture of the composition of cases in a city in terms of disease severity and the clinical trajectories of these cases. Although healthcare resources in Shenzhen were rarely overwhelmed by influx of COVID-19 patients, we show that the trajectory of clinical progression of cases in Shenzhen were similar to the trajectory in Wuhan, China, with mean time to ICU admission and mean time to developing acute respiratory distress syndrome to be around 10 days [3,4]. Patient characteristics previously reported to be associated with ARDS and death [5] including hypertension, diabetes, and various lab results were also highly predictive of faster clinical

medRxiv preprint doi: https://doi.org/10.1101/2020.04.22.20076190; this version posted April 27, 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 .

progression to various key clinical events including progression to severe stage, ARDS, and ICU admission. Previous studies reported that most cases were males and median age was over 45 [3,6] and a higher percentage of men required ICU care [3]. We further show differences in clinical progression between males and females that were primarily driven by the stark difference in the older age group independent of differences in baseline comorbidities.

This study has a number of limitations. Dates of symptom onset were extracted from physicians' notes that were not recorded to explicitly ascertain information on symptom onset. When the date of symptom onset could not be determined, we assumed onset date was the date of initial diagnosis. We performed sensitivity analyses to assess patient characteristics for clinical progression to severe stage and ICU admission and the results remained qualitatively the same (Table S4). We also performed the RSF analyses using symptom profile and lab results within 5 days of initial diagnosis as candidate predictors, and the integrated AUC remained relatively unchanged (Figure S9). Our estimates of duration of ICU stay and ventilator use were somewhat imprecise, likely due to the small sample size for those reaching the critical stage. Even though estimating duration using our non-parametric approach showed improvement in precision compared to times estimated using parametric accelerated failure time models, the estimated time within subgroups needs to be interpreted with caution because of the small sample size (Table S3). Finally, the association presented between clinical progression and patient characteristics should not be interpreted as causal given the variation in treatment and numerous confounders that were not accounted for in this study.

We demonstrate that patient characteristics near symptom onset have tremendous potential to inform COVID-19 triage, grouping patients into risk sets with different outlook of clinical progression. While our RSF model performs well based on out-of-bag predictions (Figure S8; Text S4), we would be highly cautious of triaging patients in other settings using the important variables identified here due to our limited sample size. However, this is an important first step towards an applicable triage risk screening tool once well recorded data on clinical courses for more patients become available. Strategic response and allocation of medical resources for ongoing outbreaks may also benefit from a dynamic risk scoring system that incorporates new patient-level lab and symptom information as it is updated over time.

In conclusion, we provided quantitative characterization of the clinical progression of COVID-19 patients beginning from early clinical stages. Our estimates form the basis for assessing effectiveness of new treatments and inform planning for healthcare resource allocation during COVID-19 outbreaks.

Acknowledgements

We thank all patients and their families involved in the study; as well as the front line medical staff and public health workers who collected this critical data. GZ was funded by the National Science and Technology Major Project for Control and Prevention of Major Infectious Diseases

medRxiv preprint doi: https://doi.org/10.1101/2020.04.22.20076190; this version posted April 27, 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 .

of China (No. 2017ZX10103004). CH, PZ, CY, BS, TW, JC, TM were funded by the Emergency Response Program of Harbin Institute of Technology (HITERP010) and Emergency Response Program of Peng Cheng Laboratory (PCLERP001). JL and QB were funded by a grant from the US Centers for Disease Control and Prevention (NU2GGH002000).

Author Contributions

QB, Z Wu, JE conducted statistical analyses and drafted the figures
QB, Z Wu, JE, AA, JL, LK, CH, JM, GZ drafted and provided critical feedback on the manuscript
BS, JC, JZ, PZ, CH, JM, Z Wang, CY extracted the data
QB, TM, JL, LL, GZ, TW conceived the study

Competing Interests statement

Authors declare no conflict of interests

Ethics Statement

This work was conducted in support of an ongoing public health response, hence was determined not to be human subjects research after consultation with the Johns Hopkins Bloomberg School of Public Health IRB. The study was approved by the ethics committees of Shenzhen Third People's Hospital.

References

1. Novel Coronavirus (2019-nCoV) situation reports. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports.
2. Ranney, M. L., Griffeth, V. & Jha, A. K. Critical Supply Shortages — The Need for Ventilators and Personal Protective Equipment during the Covid-19 Pandemic. *New England Journal of Medicine* (2020) doi:10.1056/nejmp2006141.
3. Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497–506 (2020).
4. Guan, W.-J. *et al.* Clinical Characteristics of Coronavirus Disease 2019 in China. *N. Engl. J. Med.* (2020) doi:10.1056/NEJMoa2002032.
5. Wu, C. *et al.* Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan, China. *JAMA Intern. Med.* (2020) doi:10.1001/jamainternmed.2020.0994.
6. Grasselli, G. *et al.* Baseline Characteristics and Outcomes of 1591 Patients Infected With SARS-CoV-2 Admitted to ICUs of the Lombardy Region, Italy. *JAMA* (2020) doi:10.1001/jama.2020.5394.
7. Website. http://www.nhc.gov.cn/yzygj/s7653p/202002/3b09b894ac9b4204a79db5b8912d4440.shtml.
8. Aalen, O. & Johansen, S. *AN EMPIRICAL TRANSITION MATRIX FOR NON-HOMOGENEOUS MARKOV CHAINS BASED ON CENSORED OBSERVATIONS*. (1977).
9. Aalen, O. Nonparametric Estimation of Partial Transition Probabilities in Multiple Decrement Models. *The Annals of Statistics* vol. 6 534–545 (1978).
10. Fine, J. P. & Gray, R. J. A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association* vol. 94 496–509 (1999).
11. Heagerty, P. J., Lumley, T. & Pepe, M. S. Time-Dependent ROC Curves for Censored Survival Data and a Diagnostic Marker. *Biometrics* vol. 56 337–344 (2000).
12. Kalbfleisch, J. D. & Prentice, R. L. The Statistical Analysis of Failure Time Data. *Wiley Series in Probability and Statistics* (2002) doi:10.1002/9781118032985.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.22.20076190; this version posted April 27, 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 1. Demographic characteristics, surveillance method, baseline comorbidity, symptom profile, and lab and CT results by the highest clinical severity assessment. Comorbidities were self-reported on admission.

	mild (n=23)	moderate (n=304)	severe (n=74)	critical (n=19)	total (N=420)	p-value
sex						
female	16 (69.6%)	170 (55.9%)	29 (39.2%)	5 (26.3%)	220 (52.4%)	0.002
male	7 (30.4%)	134 (44.1%)	45 (60.8%)	14 (73.7%)	200 (47.6%)	
age						
0-9	6 (26.1%)	14 (4.6%)	0 (0.0%)	0 (0.0%)	20 (4.8%)	<0.001
10-19	3 (13.0%)	11 (3.6%)	0 (0.0%)	0 (0.0%)	14 (3.3%)	
20-29	4 (17.4%)	34 (11.2%)	0 (0.0%)	0 (0.0%)	38 (9.0%)	
30-39	8 (34.8%)	74 (24.3%)	8 (10.8%)	1 (5.3%)	91 (21.7%)	
40-49	0 (0.0%)	54 (17.8%)	10 (13.5%)	1 (5.3%)	65 (15.5%)	
50-59	0 (0.0%)	61 (20.1%)	20 (27.0%)	3 (15.8%)	84 (20.0%)	
60-69	2 (8.7%)	49 (16.1%)	26 (35.1%)	12 (63.2%)	89 (21.2%)	
70+	0 (0.0%)	7 (2.3%)	10 (13.5%)	2 (10.5%)	19 (4.5%)	
surveillance method						
contact-based	7 (30.4%)	50 (16.4%)	6 (8.1%)	0 (0.0%)	63 (15.0%)	0.055
symptom-based	16 (69.6%)	248 (81.6%)	67 (90.5%)	18 (94.7%)	349 (83.1%)	
minimum cycle threshold value	30.4 (27.7,33.9)	28.6 (25.0,32.2)	28.0 (25.0,30.5)	25.6 (22.5,25.6)	28.4 (24.6, 32.1)	0.033
baseline comorbidity						
hypertension	1 (4.3%)	23 (7.6%)	18 (24.3%)	7 (36.8%)	49 (11.7%)	<0.001
diabetes	0 (0.0%)	11 (3.6%)	9 (12.2%)	4 (21.1%)	24 (5.7%)	0.001
coronary heart disease	0 (0.0%)	4 (1.3%)	7 (9.5%)	0 (0.0%)	11 (2.6%)	0.001
chronic lung disease	0 (0.0%)	8 (2.6%)	2 (2.7%)	3 (15.8%)	13 (3.1%)	0.011
cerebrovascular disease	0 (0.0%)	0 (0.0%)	1 (1.4%)	1 (5.3%)	2 (0.5%)	0.008
chronic liver disease	0 (0.0%)	10 (3.3%)	3 (4.1%)	0 (0.0%)	13 (3.1%)	0.66
chronic kidney disease	0 (0.0%)	0 (0.0%)	1 (1.4%)	0 (0.0%)	1 (0.2%)	0.20
chronic intestinal inflammatory disease	0 (0.0%)	1 (0.3%)	0 (0.0%)	0 (0.0%)	1 (0.2%)	0.94
cancer	0 (0.0%)	2 (0.7%)	2 (2.7%)	1 (5.3%)	5 (1.2%)	0.16
history of tuberculosis	0 (0.0%)	8 (2.6%)	1 (1.4%)	0 (0.0%)	9 (2.1%)	0.69
symptom						
fever	11 (47.8%)	216 (71.1%)	70 (94.6%)	17 (89.5%)	314 (74.8%)	<0.001
chill	0 (0.0%)	9 (3.0%)	4 (5.4%)	1 (5.3%)	14 (3.3%)	0.55
shortness of breath	0 (0.0%)	19 (6.2%)	30 (40.5%)	17 (89.5%)	66 (15.7%)	<0.001
difficulty breathing	0 (0.0%)	7 (2.3%)	7 (9.5%)	10 (52.6%)	24 (5.7%)	<0.001
chest tightness	1 (4.3%)	35 (11.5%)	26 (35.1%)	5 (26.3%)	67 (16.0%)	<0.001
chest pain	1 (4.3%)	10 (3.3%)	4 (5.4%)	2 (10.5%)	17 (4.0%)	0.41
nausea	1 (4.3%)	33 (10.9%)	10 (13.5%)	2 (10.5%)	46 (11.0%)	0.67
vomit	0 (0.0%)	15 (4.9%)	6 (8.1%)	1 (5.3%)	22 (5.2%)	0.47
diarrhea	2 (8.7%)	53 (17.4%)	17 (23.0%)	2 (10.5%)	74 (17.6%)	0.34
abdominal pain	1 (4.3%)	10 (3.3%)	6 (8.1%)	2 (10.5%)	19 (4.5%)	0.18
cough	13 (56.5%)	214 (70.4%)	65 (87.8%)	15 (78.9%)	307 (73.1%)	0.01
sputum production	3 (13.0%)	133 (43.8%)	51 (68.9%)	9 (47.4%)	196 (46.7%)	0.00
nasal congestion	1 (4.3%)	23 (7.6%)	6 (8.1%)	1 (5.3%)	31 (7.4%)	0.92
rhinorrhea	2 (8.7%)	29 (9.5%)	8 (10.8%)	2 (10.5%)	41 (9.8%)	0.99
sore throat	3 (13.0%)	39 (12.8%)	13 (17.6%)	4 (21.1%)	59 (14.0%)	0.59
headache	1 (4.3%)	32 (10.5%)	11 (14.9%)	1 (5.3%)	45 (10.7%)	0.41
fatigue	2 (8.7%)	71 (23.4%)	36 (48.6%)	11 (57.9%)	120 (28.6%)	<0.001
muscle soreness	2 (8.7%)	27 (8.9%)	21 (28.4%)	5 (26.3%)	55 (13.1%)	<0.001
joint soreness	0 (0.0%)	2 (0.7%)	5 (6.8%)	1 (5.3%)	8 (1.9%)	0.004

medRxiv preprint doi: https://doi.org/10.1101/2020.04.22.20076190; this version posted April 27, 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 .

CT results						
ground glass opacity	1 (4.3%)	293 (96.4%)	73 (98.6%)	19 (100.0%)	386 (91.9%)	<0.001
local patchy shadowing	1 (4.3%)	270 (88.8%)	72 (97.3%)	19 (100.0%)	362 (86.2%)	<0.001
bilaterial patchy shadowing	0 (0.0%)	211 (69.4%)	70 (94.6%)	19 (100.0%)	300 (71.4%)	<0.001
interstitial abnormality	0 (0.0%)	1 (0.3%)	5 (6.8%)	3 (15.8%)	9 (2.1%)	<0.001
lab abnormality						<0.001
PaO2/FiO2 <=300mmHg	0 (0.0%)	0 (0.0%)	66 (89.2%)	19 (100.0%)	85 (20.2%)	<0.001
WBC count >10K per mm3	4 (17.4%)	37 (12.2%)	28 (37.8%)	18 (94.7%)	87 (20.7%)	<0.001
WBC count <4K per mm3	4 (17.4%)	119 (39.1%)	47 (63.5%)	10 (52.6%)	180 (42.9%)	<0.001
lymphocyte count <1500 per mm3	8 (34.8%)	211 (69.4%)	72 (97.3%)	19 (100.0%)	310 (73.8%)	<0.001
lymphocyte count <1000 per mm3	3 (13.0%)	107 (35.2%)	62 (83.8%)	19 (100.0%)	191 (45.5%)	<0.001
platelet count <150k per mm3	3 (13.0%)	71 (23.4%)	45 (60.8%)	15 (78.9%)	134 (31.9%)	<0.001
platelet count <100K per mm3	0 (0.0%)	18 (5.9%)	8 (10.8%)	9 (47.4%)	35 (8.3%)	<0.001
C-reactive protein >=10mg/liter	3 (13.0%)	171 (56.2%)	71 (95.9%)	19 (100.0%)	264 (62.9%)	<0.001
Procalcitonin >=0.5 ng/ml	1 (4.3%)	2 (0.7%)	4 (5.4%)	11 (57.9%)	18 (4.3%)	<0.001
Alanine aminotransferase >40U/liter	2 (8.7%)	109 (35.9%)	49 (66.2%)	18 (94.7%)	178 (42.4%)	<0.001
Aspartate aminotransferase >40 U/liter	4 (17.4%)	72 (23.7%)	41 (55.4%)	19 (100.0%)	136 (32.4%)	<0.001
Total bilirubin >17.1 umol/liter	7 (30.4%)	125 (41.1%)	47 (63.5%)	18 (94.7%)	197 (46.9%)	<0.001
Creatinine >=133 umol/liter	0 (0.0%)	0 (0.0%)	4 (5.4%)	6 (31.6%)	10 (2.4%)	<0.001
D-dimer >=0.5 mg/liter	1 (4.3%)	100 (32.9%)	63 (85.1%)	19 (100.0%)	183 (43.6%)	<0.001
clinical events and complications						
acute respiratory distress syndrome	0 (0.0%)	0 (0.0%)	20 (27.0%)	19 (100.0%)	39 (9.3%)	<0.001
acute heart injury	0 (0.0%)	0 (0.0%)	0 (0.0%)	4 (21.1%)	4 (1.0%)	<0.001
acute kidney injury	0 (0.0%)	2 (0.7%)	1 (1.4%)	8 (42.1%)	11 (2.6%)	<0.001
liver dysfunction	11 (47.8%)	187 (61.5%)	64 (86.5%)	19 (100.0%)	281 (66.9%)	<0.001
peneumothorax	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	-
sepesis	0 (0.0%)	1 (0.3%)	2 (2.7%)	8 (42.1%)	11 (2.6%)	<0.001
deep vein thrombosis	0 (0.0%)	0 (0.0%)	2 (2.7%)	4 (21.1%)	6 (1.4%)	<0.001
bacterial pneumonia	0 (0.0%)	5 (1.6%)	5 (6.8%)	11 (57.9%)	21 (5.0%)	<0.001

*sample size for the cycle threshold value was 11, 222, 56, 19 for the mild, moderate, severe, critical group; mean (IQR).

chronic lung disease included bronchiectasis, COPD, emphysema, and chronic bronchitis, but excludes asthma. Chronic liver disease included hepatitis B and cirrhosis.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.22.20076190; this version posted April 27, 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: Estimated length of hospital stay, duration in ICU, and length of invasive ventilator use. We estimated duration in ICU among those who were admitted into ICU and estimated duration on ventilator among those who required ventilation.

	length of hospital stay (95%CI)	duration in ICU (95%CI)	duration on ventilator (95%CI)
all cases	24.0 (23.0,25.1)	34.4 (24.1,43.2)	28.5 (20.0,39.1)
sex			
female	23.4 (22.1,24.6)	27.2 (20.4,33.0)	15.0 (13.2,52.7)
male	24.8 (23.2,26.5)	37.0 (25.9, 50.8)	33.7 (16.9,39.1)
age			
0-39	20.3 (19.2, 21.5)	-	-
40-59	24.3 (22.8,25.9)	23.0 (12.5,38.0)	13.0 (5.5,24.0)
60+	29.3 (26.4,32.1)	37.9 (25.9,48.7)	34.1 (23.0,46.3)
surveillance method			
contact-based	22.7 (20.7, 24.7)	-	-
symptom-based	24.2 (22.9, 25.4)	34.8 (26.2,45.1)	29.0 (18.4,40.4)
RSF-defined risk group^			
low	20.6 (19.4,21.8)	-	-
medium	22.4 (21.0,24.0)		
high	28.3 (26.2,30.6)	35.1 (26.1,45.2)	30.0 (17.9,41.4)
baseline comorbidity			
any comorbidity	28.1 (25.1,31.3)	46.5 (30.1,59.4)	41.3 (15.3,44.7)
hypertension	30.3 (25.9,34.9)	48.5 (33.5,66.0)	42.3 (13.3,44.5)
diabetes	32.7 (26.2,40.1)	-	-
coronary heart disease	25.4 (18.9,32.5)	-	-
chronic liver disease	22.5 (20.4,24.5)	-	-
chronic lung disease	37.5 (23.7,52.8)	-	-
early symptom			
fever	24.9 (23.6,26.3)	33.5 (24.6,42.8)	28.0 (18.1,40.4)
cough	24.1 (22.8,25.5)	31.6 (22.2, 40.7)	26.2 (17.3,45.3)
shortness of breath	31.3 (25.1,38.6)	26.3 (13.0,41.8)	24.7 (11.1, 39.5)
fatigue	27.4 (24.5,30.6)	-	-
muscle soreness	27.7 (24.2,31.4)	-	-

^Three cases whose initial severity was either severe or critical were excluded from estimating risk-group specific duration of hospital stay, duration in ICU, and duration on ventilator.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.22.20076190; this version posted April 27, 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 3: The association of demographic characteristics, baseline comorbidity, initial symptoms, and initial lab results with rate of clinical progression to severe stage, acute respiratory distress syndrome, and ICU admission

	progression to severe stage			progression to acute respiratory distress syndrome			progression to ICU admission		
	sHR	2.5%	97.5%	sHR	2.5%	97.5%	sHR	2.5%	97.5%
sex									
female	ref			ref			ref		
male	2.0	1.3	3.0	1.4	0.8	2.7	3.1	1.1	8.7
age									
0-39	ref			ref			ref		
40-59	4.5	2.1	9.3	6.2	1.4	28.1	4.4	0.5	39.6
60+	10.1	5.0	20.5	21.9	5.2	91.7	22.4	3.0	168.2
surveillance method									
contact-based	ref			ref			-	-	-
symptom-based	2.7	1.2	6.2	3.4	0.8	13.9	-	-	-
cormorbidity									
hypertension	3.2	2.1	5.1	3.6	1.8	6.9	4.6	1.8	11.6
diabetes	3.0	1.6	5.4	5.5	2.7	11.3	4.7	1.6	14.1
coronary heart disease	3.7	1.9	7.0	2.0	0.5	7.4	0.0	0.0	0.0
chronic lung disease	2.0	0.8	4.7	4.2	1.5	11.8	6.2	1.9	20.1
cerebrovascular disease	5.2	4.2	6.6	5.6	1.2	24.9	21.7	1.6	300.5
total comorbidity									
0	ref			ref			ref		
1	2.6	1.6	4.2	3.0	1.4	6.3	2.1	0.7	6.7
2,3	3.6	2.0	6.5	4.8	2.2	10.6	6.0	2.1	17.3
early symptom									
fever	3.4	1.9	6.1	2.2	1.0	4.9	4.0	0.9	17.2
cough	2.0	1.2	3.1	1.5	0.8	3.0	1.5	0.6	3.8
shortness of breath	4.7	2.4	9.3	7.2	3.2	16.0	15.5	6.1	39.2
muslcle soreness	3.2	2.0	5.2	3.0	1.5	6.1	2.2	0.8	6.7
fatigue	2.8	1.8	4.3	3.1	1.7	5.9	3.9	1.6	9.5
lab results near symptom onset									
PaO2/FiO2 <= 300mmHg	38.3	14.6	100.8	16.7	7.4	37.8	22.3	8.4	58.8
lymphocyte count < 1500 per mm3	2.1	1.4	3.1	2.1	1.1	4.0	3.6	1.3	10.0
lymphocyte count < 1000 per mm3	2.1	1.4	3.3	2.9	1.5	5.5	4.5	1.8	11.0
platelet count <150,000 per mm3	3.1	2.0	4.7	2.1	1.1	4.1	3.0	1.2	7.5
C-reactive protein >= 10mg/L	3.6	2.4	5.4	3.1	1.6	5.7	10.6	3.5	31.6
D-dimer >= 0.5mg/L	3.0	1.9	4.8	3.2	1.6	6.3	5.0	2.0	12.3

medRxiv preprint doi: https://doi.org/10.1101/2020.04.22.20076190; this version posted April 27, 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 .

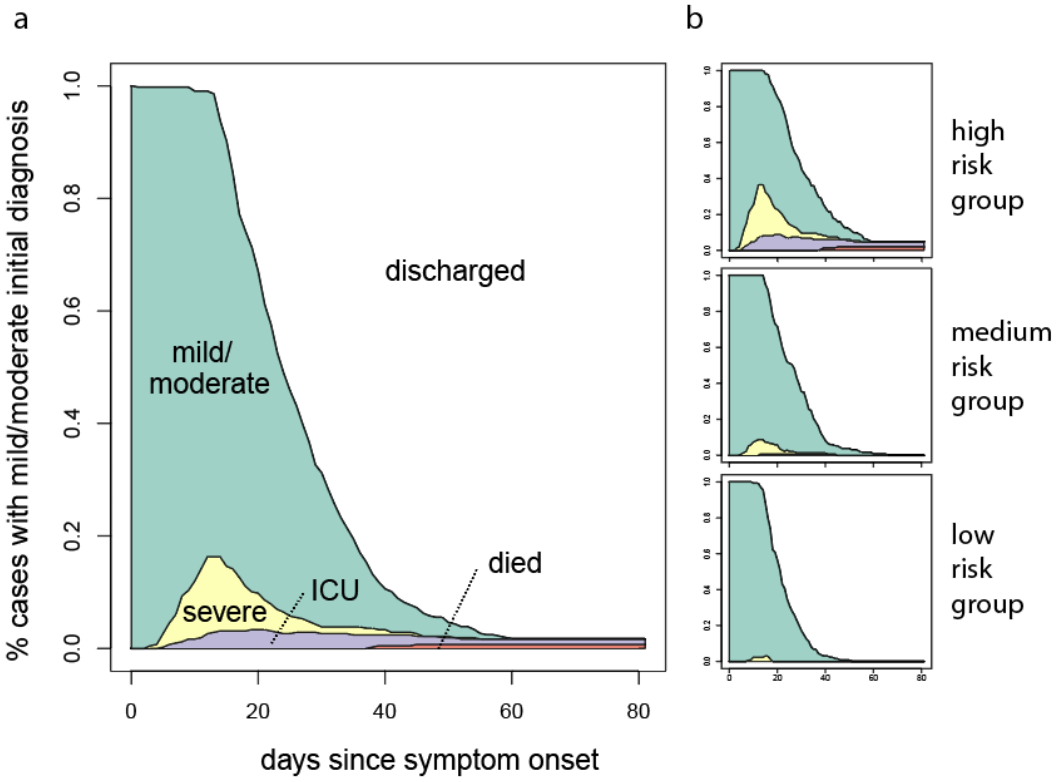

Figure 1: Clinical progression within 80 days following symptom onset for a) all cases with mild or moderate initial assessment, and b) cases in each of three risk subgroups obtained from random survival forest (see *Results* for the list of predictors).
From the top to the bottom, the four curves show the time-varying proportions of all admitted cases who 1) have not been discharged, i.e., still hospitalized or have died in hospital, 2) were severe, in the ICU, or have died, 3) were in the ICU or have died, and 4) have died. Successive differences between the four curves over 80 days from symptom onset were highlighted in distinct colors and show daily composition of cases in each of four stages (mild/moderate, severe, ICU, died).

medRxiv preprint doi: https://doi.org/10.1101/2020.04.22.20076190; this version posted April 27, 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 .

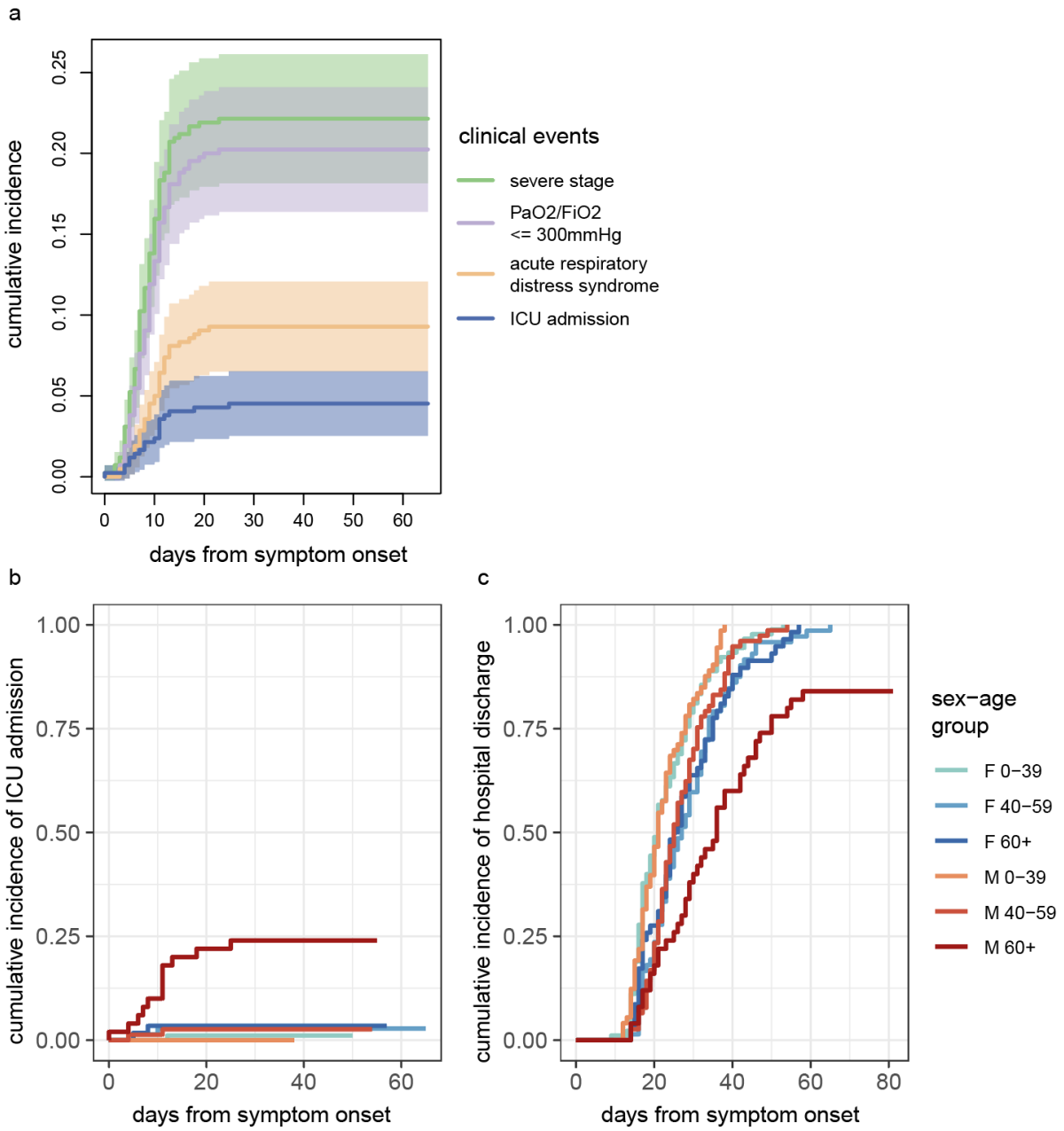

Figure 2: Cumulative incidence to clinical events and difference by age and sex. A) Cumulative incidence of advancing to severe stage, PaO2/FiO2 dropping below 300mmHg, requiring ICU admission, and developing acute respiratory distress syndrome. B) cumulative incidence of ICU admission by age group and sex. C) cumulative incidence of hospital discharge by age group and sex.