Comments from the editors

* Extend length of follow-up to at least 21 days (data that now should be available), to reduce risk of bias from right censoring.

> **Response:** We have now updated our data to include at least 28 days of follow up after positive test result. We include patients tested on or before April 8, with follow up through May 5. In fact, the median duration of follow up for patients from positive test result is 35 days (IQR 29-41), in which follow up extends from time of first positive test to time of death or to May 5 for those still alive. Among hospitalized patients, all but 12 (0.4%) have been discharged or experienced critical illness, essentially eliminating right censoring as a concern for critical illness analyses. Moreover, among hospitalized patients, 93.8% have now been discharged alive or have died, resulting in much less right censoring even for a stand-alone mortality outcome, which we have now added to this analysis. Among the subset of 23.6% of hospitalized patients who required mechanical ventilation, 76.6% have died or been discharged, and an additional 10.0% have been extubated and are awaiting discharge, leaving only 13.3% still ventilated.

* As Reviewer 2 and our staistical consultant (Reviewer 4) emphasize, your study includes 2 groups, those who test positive and those who were hospitalized. There is substantial risk for bias in the first. Can you provide any additional information about those who tested negative? In view of this issue (and other peer review comments), perhaps your paper would best be considered a case series rather than a prospective cohort?

> **Response:** We have now added **Appendix Table S1** to show characteristics of test negative patients. In general, patients testing negative were more likely to be younger, white and female, and had slightly lower rates of most comorbidities (though not all).
>
> We certainly agree that our dataset does not represent an unbiased sample of all patients with COVID-19 disease, as our testing strategy (and that of most of the nation) is focused on sicker patients. We think of this as a cohort study not in the sense that it represents an unbiased sample of all COVID-19 patients, but in the classic sense that it is a group of patients with a common characteristic (positive test) followed prospectively forward. In this our dataset is different from most other published case series, which only include individual hospitalization episodes and do not have as comprehensive a sense of outcomes over time. While we are open to calling it a case series, we welcome your feedback on how we can best characterize this group of patients to help readers understand how our results differ from others. For now we have retained the prospective cohort language.

* Conduct a survival analysis

> **Response:** For the hospitalization outcome, survival analysis is not meaningful since most patients who are hospitalized first test positive on the day of admission. For the critical care outcome, we now have only 12 cases in which there is uncertainty about whether the patient will be discharged or develop critical illness (0.6%). There is, therefore, essentially no risk of bias from right censoring for our critical care outcome, now making logistic regression a valid approach. Nonetheless, we tried to construct a survival analysis for this outcome, but several of the variables (age, race/ethnicity, hypertension) did not meet the proportional

hazards assumption and the model was therefore not valid. Accordingly, we have retained the logistic regression analysis for this outcome.

We suspect from some of the reviewer comments, however, that what they were really seeking is a new survival model analysis for mortality alone, not the broader critical care outcome. Accordingly, we have now added this analysis. We used a competing risk model because discharge was a competing risk for the identification of mortality; unless a patient was readmitted, we have no further information post-discharge on mortality. Similarly, we do not have reliable information about mortality for non-hospitalized patients. Therefore, we elected to do this analysis with only the hospitalized cohort.

New text, page 9: Finally, we fitted a competing risk model for the mortality or hospice outcome with time from first positive test as the start point, including only hospitalized patients.[19] We considered discharge from hospital to be a competing risk, since mortality data are limited after that point unless the patient is readmitted to our system (in which case the newest hospitalization would be included). Patients still hospitalized as of May 5, 2020 were counted as censored. The model was fitted with the R library cmprisk[20] and the proportionality assumption was checked with the gofftest library.[21] We fitted two competing risk models, one adjusting for demographics and comorbidities, and one adding admission vitals and laboratory studies.

New table 4

New Figures 2-4

* Omit the decision tree analysis

Response: We have now omitted these analyses

* Clarify patient flow (e.g., did patients first arrive at ambulatory clinics?)

Response: We have now included a table of site of first positive test by ultimate admission status as a supplementary appendix (**Appendix Table S2**). Indeed, as suspected by the reviewers, testing site differs by ultimate disposition. 96.7% of patients requiring admission were first tested in the emergency department or as inpatients. Conversely, only 19.9% of patients not requiring admission were first tested in the emergency department.

This would cause a bias in our assessment of the association of demographics and comorbidity with hospitalization risk if one of two conditions were met:

1) If patients tested as outpatients systematically went elsewhere for hospitalization. We believe this to be relatively unlikely, as these are all patients with prior visits in the system and electing testing at NYULH suggests these patients might also elect hospitalization there.
2) If patients with Covid-19 who were sicker (and therefore potentially with more comorbidities) preferentially presented to the emergency department for care but were discharged without testing. Such patients would not be captured in our non-hospitalized cohort, potentially making that cohort healthier. This is a real possibility: at our ED such patients are often discharged without testing.

To address this potential bias, we have now added a sensitivity test in which we add to the non-hospitalized cohort 5,914 adult patients treated for fever, cough or dyspnea in the ED during the study period who were neither admitted nor tested. As suspected, this combined cohort now has a higher rate of comorbidity (see **Appendix Table S4**), such that most comorbidities now have a slightly lower OR – though directionally all the same, and nearly all still significant (**Appendix Table S5**). Note, however, that this sensitivity analysis itself has limitations, as not all patients with influenza like illness actually have Covid-19. We are likely now diluting the non-hospitalized group with non-Covid-19 patients. In fact, of these patients, 118 were later tested (after April 8), and 92 (78%) were negative; 22% were positive. Because of this concern, we have not replaced our main analysis with this one, but retain it as a sensitivity analysis only.

* Clarify missing data

Response: Most missing data in our patient cohorts are not missing at random. In the hospitalization analysis, missing data is usually a protective factor, as data is less comprehensively collected for non-hospitalized patients. Among hospitalized patients, missing vitals usually represent increased risk, since patients with missing data are usually those too sick or too rapidly expiring to collect data on, but missing laboratory studies are typically associated with lower severity of illness. For these reasons we consistently include a term for “missing” in every categorical predictor rather than dropping those patients or attempting to impute data.

* Regarding the lipid findings: 1) How did you defined hyperlipidemia? 2) Can you separate high LDLc, statin treatment, and high triglyeride-to-HDLc ratio? The hypothesized protective effects involve LDLc (anti-viral or anti-inflammatory), whereas high trigs-to-HDLc ratio is a biomarker of metabolic syndrome, a potential major risk factor. Statin treatment might may be protective independly of LDLc, with anti-inflammatory effects reported for this drug. Ideally, run these 3 baseline covariates in a multivariate model of risk for bad outcome.

Response: Hyperlipidemia was defined based on any diagnosis of hyperlipidemia in the problem list, ambulatory or inpatient encounters, or past medical history in the past 18 months. We were unable to include lipid levels in statistical analyses, as we were missing results for 50% of hospitalized patients with hyperlipidemia (and 75% of hospitalized patients overall). We did obtain pre-admission statin use; however, it is highly collinear with diagnosis of hyperlipidemia, and we were therefore unable to include both statin use and hyperlipidemia simultaneously in the model. We were reluctant to replace hyperlipidemia with statin use, as we do not treat other comorbidities the same way (i.e. glucose lowering agents in place of diabetes). We believe the best way to determine the effect of statin use would be through a separate, dedicated analysis making use of a propensity matched cohort or other methods to account for selection bias. That is beyond the scope of this paper, but such analyses are now underway.

In your response please provide, point by point, your replies to the comments made by the reviewers and the editors, explaining how you have dealt with them in the paper.

Response: Following are point-by point responses to each comment. For ease of review, we summarize all major changes to the manuscript here. In this revision, we:

1) Removed patients less than 19 years old, because their admission criteria and outcomes are systematically different than those of adults
2) Increased denominator population to those tested on or before April 8, from April 5 [now 2,741 hospitalized patients from 2,390]
3) Refreshed outcome data as of May 5, ensuring at least 28 days follow up for all patients
    a. >99% of hospitalized patients have now reached study endpoint (discharge or need for critical care), eliminating right censoring concerns [up from 95% previously]
    b. 94% of hospitalized patients have now been discharged alive or died [up from 83% previously], including 77% of ventilated patients [up from 45% previously]
4) Added comorbidities identified from past medical history and inpatient/outpatient encounters in addition to those on the problem list, yielding approximately 5 percentage point increases in most comorbidity rates
5) Added new survival analysis for new outcome: mortality among hospitalized patients
6) Added sensitivity analysis for the hospitalization outcome by adding patients treated and released from the ED without testing who were at high risk for Covid (presenting with fever, cough or shortness of breath; or given “presumptive Covid-19” or “suspected Covid-19” diagnosis; or placed into isolation) to the non-hospitalized cohort
7) Changed all logistic regression models to hierarchical logistic regression to account for facility at which treated
8) Added week of year as a covariate to account for time-varying admission and treatment protocols
9) Removed the model of critical illness among all patients

** Comments from the external peer reviewers**

Reviewer: 1
The manuscript “Factors associated with hospitalization and critical illness among 4,594 patients with COVID-19 disease in New York City: A prospective cohort study” presents a comparison of clinical observations and risk factors among confirmed COVID-19 cases who are or are not hospitalized and/or admitted to intensive care by time of writing. As the authors note this is the largest case series I have seen described and the first comparison of hospitalized versus non-hospitalized patients, which is valuable. In addition, many risk factors are assessed although there are some important limitations to the design and statistical analysis which may limit the value of these analyses in my view. Some of these can be corrected just by using survival analysis methods, which I believe is important to prevent misinterpretation of the observations regarding clinical outcomes.

Main comments:

1) My first of two key concerns is that observations are reported/analyzed only for patients with observed death/hospice/discharge outcome, resulting in right-censoring (and exclusion) of about 20% of all hospitalized patients; more troubling, we have right-censoring/exclusion of 40% of patients receiving mechanical ventilation. Since this censoring affects a unique class of patients (those requiring longer stay/longer ventilation than those with outcomes observed at time of writing, thus probably sicker patients than those with outcomes observed), interpretation of the clinical outcome data is greatly undermined. The unaddressed censoring issues do not support the authors’ aim to “improve the reliability of future mortality rate estimation”. This issue should be corrected with survival analysis methods rather than those used here as the reported CFR from this study will be extensively quoted by press, etc. and non-statistical audiences.

Response: In fact, the critical illness analyses only excluded patients who had not yet been discharged *or* experienced critical illness (including ICU care or mechanical ventilation, not just death/hospice). In the original submitted paper, therefore, only 110 (5%) patients were excluded from the critical care analyses. In the current paper, only 12 (0.4%) of patients are excluded from the critical care analyses. Because of this very small N of patients without definite outcome, we retained the logistic regression, with modifications as described above, because it requires fewer assumptions and is more robust. However, in response to the reviewer's comment, we have now added a new competing risk survival model using mortality as the outcome. In this analysis, 7% of all hospitalized patients were right censored (including 24% of ventilated patients). These results are all directionally similar as the critical care analysis, and the HR for age, for instance, is still very large – e.g., 10.3 for age>74.

New results, page 12: The competing risk mortality analysis showed similar characteristics to confer increased hazard as the critical illness model, but fewer were significant. The hazard ratio increased proportionally by age, with HR 10.3 (95% CI, 6.4-16.8) for age ≥75. Other significant factors included male sex (HR 1.3, 95% CI 1.1-1.5), heart failure (HR 1.8, 95% CI 1.4-2.2) and cancer (HR 1.3, 95% CI, 1.1-1.6). As with the critical illness model, once vital signs and laboratory studies were added, few patient characteristics remained significant. Only age (e.g., >75 HR 7.7, 9% CI 4.6-12.8) and cancer (HR 1.3, 95% CI 1.0-1.6) remained. Again similar to the critical illness model, vitals and laboratory results on presentation carried significant additional hazard, chiefly hypoxia on presentation (HR 2.0, 95% CI 1.6-2.5 for SpO2<88%), C reactive protein (all abnormal levels had HR >3.5), d-dimer (HR 2.2, 95% CI 1.6-3.0 for first result >2500) and troponin (HR 2.1, 95% CI 1.4-3.2 for first result >1) (Table 4). Representative cumulative incidence functions are shown in Figure 2 (age groupings), Figure 3 (heart failure, cancer, diabetes, male sex) and Figure 4 (admission oxygen saturation, C-reactive protein, d-dimer, lymphocyte count).

New table 4, figures 2-4

Related to the point above is that we also have a fundamental censoring problem of who, among all those diagnosed, will ultimately become a hospitalized case: since time elapses from initial presentation at an outpatient visit to hospitalization (at least for cases whose first interaction is not an inpatient admission) the concerns above apply to this comparison. Same goes for critical care/no critical care among the hospitalized.

Response: After refreshing the data, the median follow up in our sample now from first positive test is 35 days, which is well beyond the period (7-12 days) at which most patients become sick enough to require hospitalization, even assuming that all outpatient tests were obtained on day 1 of symptoms. Among hospitalized patients, there are now only 12 patients who are still hospitalized without critical care or discharge. Therefore, we believe censoring of hospitalization among those tested as outpatients and of critical illness among those still hospitalized is now of little concern. We do have a separate risk of missing hospitalizations in those tested as outpatients who are then hospitalized outside our system. We hope that the fact that they were tested in our facility makes it more likely that they would be hospitalized here if sicker, but we have added this point to the limitation section.

New text, page 17: Finally, our outcome assignments may be imperfect: some patients in the non-hospitalized group may have been hospitalized at other institutions

2) My second concern is the changing definition of the non-severe case group over time due to changes in testing practices (assays, conducted by NYC DOHMH vs. Langone, recommendation to restrict testing of patients with mild/moderate illness after March 26). However, not much can be done about this and there is still a value of comparing hospitalized/non-hospitalized patients.

Response: We agree this is a bias and have added to the limitations section.

New text, page 16: They are also a heterogenous group made even more heterogeneous by changing testing thresholds over time.

In light of these considerations it is probably optimistic to call the study a prospective cohort rather than a case series. In my view a proper prospective cohort of COVID-19 cases would catch all patients meeting a given clinical threshold and assess how many go on to hospitalization, critical illness, ventilation, death, etc. Here patients are being ascertained at differing thresholds over the course of the study.

Response: We agree that this is one definition of a cohort study; however, a typical definition of a prospective cohort is a group of subjects with a common characteristic (here, positive test) that are identified at a given time point and followed forward. Our cohort meets this latter, common, definition. Nonetheless, as noted above, we would be open to calling this a case series if the editors prefer.

3) The patient EHR data used for analyses of risk factors like tobacco use, BMI, and comorbidities come from previous interactions with the health system. However it is not clear if all the patients routinely receive care from this system. For instance, 140 patients have only age and sex data and are excluded from the analysis—presumably they do not. Is further information available on this issue? The concern is that the absence of an indication of a comorbidity/risk factor could arise if the comorbidity/risk factor is truly absent, or patient is not routinely receiving care from the Langone system such that this is simply not recorded in the medical record. This becomes a problem of confounding when we consider that those who interact with the health system often (either because they are very sick or the worried well) probably differ from those who do not. This concern is well illustrated in the smoking analyses, for instance. The fact that smoking appears protective could suggest that simply having EHR risk factor data available suggests individuals are better linked to care/in better health.

Response: We fully agree and already list this issue first in the limitations section. In this revision, we have expanded our exclusions to exclude those with no prior interaction with the health system and only a visit for testing, such that we exclude 287 patients instead of 140. This helps to ensure that the remaining patients had opportunities to have comorbidity data documented. Moreover, we have now expanded our search for prior comorbidities. Previously, we only included comorbidities listed on the problem list, which our institution considers most trustworthy. We have now added comorbidities described in the past medical history section or on encounter diagnoses (billing data). Most comorbidity rates are now about 5 percentage points higher because of these additions. Of note, we do not believe this limitation explains the smoking results, as most comorbidities once documented show an association with *increased* likelihood of hospitalization, not less.

Existing text: Most important, data on non-hospitalized patients was more limited because many did not have vital signs or blood samples collected, and may not have had as detailed

a medical history taken. We may therefore be overestimating the importance of chronic disease in hospitalization risk.

4) Are too many covariates being included if hyperlipidemia is associated with lower hospitalization risk/critical illness when all the other findings are opposing this? Is there a scientific reason to assess the association with hyperlipidemia after controlling for BMI, overweight, etc.? The findings are hard to interpret given multiple forms of censoring and a time-varying spectrum of patients who would be ascertained as COVID-19 cases without meeting the hospitalization threshold. Another explanation is offered in the discussion but I do not know the merits/significance.

Response: We also don't understand the reason for this finding and can only speculate. We have described several hypotheses in the discussion section. We have now formally tested all the variables for multiple simultaneous interactions using the determinate test and find no significant interactions.

5) What is the interpretation/value of the decision tree? It doesn't inform who benefits most from hospitalization (deaths averted etc.) and therefore should be hospitalized; moreover, the information presented in the tables already addresses risk factors—the "use case" of this analysis is not clear to me. The issues around censoring are masked even further in this analysis and ultimately I don't think it adds much to the paper, unless I am missing something.

Response: Given the large number of covariates assessed in this analysis, we thought it would help to conduct an analysis that shows relative prioritization of risks in terms of best ability to separate patient cohorts; moreover in observational analyses we think it is a strength to show similar results with very different approaches. However, at the recommendation of the editor, we have removed these analyses.

Minor/specific comments

6) The endpoint of the paragraph beginning "After adding admission vitals…" is not clear. Is this still critical illness? Presumably the finding that children ages 0-18 had high risk is hard to interpret given only very susceptible children would experience such illness to begin with. Age probably isn't the risk factor so much as a basis for selection.

Response: Yes, this endpoint is still critical illness. We agree about pediatric interpretability challenges and have now removed patients < 19 years from this analysis.

7) I remain concerned that the discussion puts undue weight on the interim results among patients with observed definitive outcomes, i.e. more than 20% are likely to die here once outcomes are observed for those with very long hospitalizations and/or readmissions.

Response: We agree; for this reason we avoided constructing a mortality model initially and focused on critical illness, for which we had near complete data. Given more follow up time now, we have added a formal survival analysis for mortality. Recall that because we follow patients forward, we include rehospitalizations (if occurring in our health system and within the follow up period, which is to May 5). We have added an upper bound in the discussion section.

New text, page 14: If all remaining hospitalized patients died, the overall mortality would be a maximum of 30.5%.

8) The authors claim that the findings can help epidemiologists improve projections around hospital beds, staffing, etc. although I think these claims may be unsupported as the paper does not address key aspects like incidence and trends in incidence/transmission intensity (nor can it since there is no population-at-risk denominator to work from), and does not address duration of hospital stay.

Response: We have removed these comments from the abstract and introduction section.

Reviewer: 2
This is an important paper representing the largest COVID-19 clinical case series from the U.S. with 4,734 cases identified in an NYC private health system identified between March 1 through April 5th , including 2,390 who were hospitalized. For reference, this is more total cases than over 25 U.S. states. The goal of this study was to describe the characteristics of patients that tested positive for COVID-19 within a health system and the association between those characteristics and adverse outcomes. The rapid reporting and analysis of these data is impressive and commendable, providing useful clinical and public health evidence as areas around the world are currently grappling with how to manage this disease. However, there are number of concerns and limitations. The majority of these concerns are addressable. If these concerns are addressed, this paper will make a high impact contribution to the literature particularly with the understanding of obesity and inflammatory parkers as being associated with adverse outcomes among COVID19 patients requiring hospitalization, as well as understanding the overall rates of critical illness in a large US cohort of hospitalized COVID19 patients.

Major concerns

1. Design of the study cohort(s)

The study has two study cohorts, one nested within the other:

Cohort #1 All patients who tested positive for COVID-19 via ambulatory or ED testing in the NYU Lagone system. Outcome = hospitalization in NYU Lagone system.

Cohort #2: All patients who were hospitalized in the NYU Lagone system and were not still hospitalized on an inpatient ward at the time of followup. Outcome = critical illness (ICU, mechanical ventilation, discharge to hospice or death)

Problems with Cohort #1: There are multiple potential sources of bias in the construction of this cohort for answering the question which demographic and clinical factors are most associated with need for hospitalization.
- Heterogeneity in site of test/hospitalization decision resides in the emergency department (ED). The cohort includes patients tested in ambulatory settings, concerned employees, and those presenting to the emergency department. However, no data are provided on proportion of the sample by location of testing and whether there are demographic and clinical differences according to location of testing. Therefore, imbalances in demographics and severity of illness by location of testing could account for associations between these factors and hospitalization. For example, if Hispanic patients are overrepresented in the population of patients tested in the ED, the higher hospitalization rates could simply represent higher

unobserved severity of illness in this population. Furthermore, the decision to hospitalize is made exclusively in the ED Therefore, those who were tested as outpatients but never presented to or were referred to the emergency department could never experience the outcome of hospitalization. To make these data useful and generalizable for clinicians practicing in the ED, a cohort could just be limited to those who presented to the ED. Likewise, to assist outpatient clinician determine which patients should be referred to the ED for evaluation and consideration of hospitalization, a separate cohort could be created addressing that decision.

Response: We have now provided data on proportion of the sample by location of testing (**Appendix Table S2**), the demographic characteristics by location of first positive test (**Appendix Table S3**), and the demographic characteristics of patients treated in the ED for suspected Covid-19 without testing. As can be seen in these tables, indeed Hispanic patients are overrepresented in ED treat and release visits. When these (untested, suspected) patients are included in analysis, the increased risk of Hispanic patients is attenuated (**Appendix Table S5**). We now comment on this in the paper. More generally, however, we believe the heterogeneity of the comparison cohort is actually a strength. If we restricted for instance only to ED patients, our comparison group would have a higher proportion of comorbidities than the general public and we might falsely conclude that comorbidity is not associated with severity of illness. Finally, we disagree that those who were tested as outpatients could never experience the outcome of hospitalization – presumably, if they became sick enough, they would present to an emergency department and have the opportunity to be admitted. If at our institution, we would then capture that event. We have added to the limitations section that we may not capture all of these events.

New text, page 12: Being self-reported Hispanic was associated with increased risk of admission but not critical illness; this increased risk was eliminated once ED treat-and-release patients were included (**Appendix Table S5).**

New text, page 17: Finally, our outcome assignments may be imperfect: some patients in the non-hospitalized group may have been hospitalized at other institutions, and some discharged patients may have been readmitted elsewhere with critical illness, or may have died post-discharge.

New table: Appendix Tables S2-5

- Potential biases in who gets tested. We do not know from the data what the total denominator of patients who were eligible to be tested based on influenza or COVID-19 like illness - https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/04172020/covid-like-illness.html. (Obviously, it takes a lot more work to define that cohort). Therefore, it cannot be ruled out that there were biases in who was tested all else equal even though there are number of other reports indicating that blacks and Hispanics have been tested at lower rates. So even if just limiting to the patients presenting to the emergency department, there could be unobserved differences in severity of illness according to demographic groups. Finally, for patients presenting the ED, the decision to admit is based on assessment of clinical presentation/severity. If the decision is made to admit, a COVID-19 test is sent. If the decision is made that the patient is well and can be discharged, these patients often do not get tested for COVID-19 at all or do not receive any other tests (see assessment of well patients: https://emupdates.com/cv3w/). The authors could make the case this is not a major deal if a high proportion of those presenting to the ED with COVID-19 like illness and were discharge did receive testing.

Response: The reviewer is quite correct that patients treated and released from the ED were typically not tested. In our dataset, 20% of non-hospitalized Covid-19 patients were tested in the ED; the rest as outpatients. To address this concern, we have added a sensitivity analysis including 5,914 patients treated and released from the ED during the study period with possible Covid-19 by the following: presenting with fever, cough or shortness of breath; or given "presumptive Covid-19" or "suspected Covid-19" diagnosis; or placed into isolation. (**Appendix Table S5**). As noted above, this attenuates most ORs but most remain significant.

New text, page 11: A sensitivity analysis adding patients seen in the ED for suspected COVID-19 but not tested produced similar results (**Appendix Table S5**).

New text, page 16: This limitation may be further exacerbated by the fact that patients treated and released from the ED were not commonly tested and thus omitted from our analysis unless later hospitalized, yet might be more likely to have comorbidities than those tested in outpatient settings.

New table: Appendix Table S5.

- Substantial missing physiologic data between hospitalized and non-hospitalized patients. Table 1 indicates that Temperature was missing for 74% of non-hospitalized patients and that there substantially higher proportions of oxygen saturation measurements were made while supplemental oxygen was being administered among hospitalized patients (99% vs. 23%). These major observable imbalances are problematic and there are likely major unobservable imbalances as well for the reasons above.

- Patients testing positive in outpatient settings may have been hospitalized in non NYU EDs

Response: We present the physiologic data in case readers are curious about presentation, but do not include it in models for precisely the reason the reviewer raises.

Cohort #2: This is a much cleaner and more generalizable population. The methodological problems with Cohort #1, detract from the important information that can be gleaned from Cohort #2.

2. Statistical Analysis and Modeling
- Given the goal of this paper is to "describe the clinical and laboratory characteristics associated with severity of illness" and not the distribution of clinical and laboratory characteristics among those with severe illness, it seems that Tables 1 and 2 would provide more clarity to clinicians if reporting row percentages instead of column percentages. (eg https://www.annalsthoracicsurgery.org/article/S0003-4975(15)01520-9/pdf).

Response: We respectfully disagree with this suggestion. Column percents are in fact much more commonly displayed in papers comparing populations (e.g. https://www.nejm.org/doi/full/10.1056/NEJMoa2002032 https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)30566-3/fulltext) because they provide readers with an easy means of assessing disproportionate share of events between cohorts, independent of group size. For instance, 53% of the cancer comorbidities in our study are in non-critical patients and 47% are in critical patients (row percent). That makes it appear at first glance that cancer is less common in critical patients.

However, since critical patients only account for 36.1% of the total patients, in fact cancer is overrepresented in this group. This is hard for a reader to grasp at first look. By contrast, reporting column percents, we show that 8.9% of non-critical patients and 13.9% of critical patients have cancer, making the overrepresentation immediately obvious to the casual reader. This formulation also aligns better with the manuscript text. Thus, we have retained column percents in virtually all tables.

-The model presented in Table 1 includes a common outcome (2,390/4,594 = 52% experienced event of hospitalization). The "risk of hospitalization" is presented as an odds ratio, but this overinflates relative "risk" given this is a common outcome (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3348192/). Hence an odds ratio of 48.53 for patients 75 and older. As a clinician, a really high odds ratio loses meaning. Could consider presenting as predicted probabilities, relative risks, or average marginal effects to make the data more meaningful and useful for clinicians. As AMEs are not familiar to most medical readers, we have retained ORs in the abstract but can change to AMEs if the editors prefer.

Response: We appreciate this suggestion. In this revision, we have added average marginal effects to the hospitalization model to make the results more interpretable.

New text, page 9: We also calculated average marginal effects for each predictor by using the margins library in R, which uses a discrete first-difference in predicted outcomes to obtain the AME.

New text, page 11: In multivariable analysis of the full COVID-19 positive cohort, the factors most strongly associated with hospitalization were age, including 75 years or older (OR 37.9, 95% CI, 26.1-56.0; average marginal effect [AME] 58%), age 65-74 (OR 8.7, 95% CI, 76.8-11.2, AME 40%), heart failure (OR 4.4, 95% CI 2.6-8.0, AME 22%), male sex (OR 2.8, 95% CI 2.4-3.2, AME 16%), chronic kidney disease (OR 2.6, 95% CI, 1.9-3.6, AME 14%), and any elevation in BMI (e.g., for BMI>40 OR 2.5, 95% CI, 1.8-3.4, AME 14%).

New column, Table 1

- Given the data on critical illness come from 4 hospitals, which are in different neighborhoods, may have different surge/ICU capacity and practice patterns, seems like the multivariable models should consider adjusting for hospital effects. Seems especially important given that one of the hospitals is a children's hospital.

Response: We have now included hospital as a random effect to account for hospital-level variation. (Of note, as this revision excludes pediatric patients, the children's hospital is not a separate site.)

- Table 3 – because of all the potential sources of bias with Cohort # 1 (all patients), the model looking at risk of critical illness among all patients tested in the health system is going to have all the same limitations and biases. I think this detracts from the more useful and more robust results of critical illness among hospitalized patients.

Response: We agree and have now dropped this model.

- It appears that the laboratory data in the multivariate regression model in Table 3 are not missing at random.

Response: We agree; for this reason we include “missing” as a level in all analyses.

- Give the dynamics of the surge, and changes in practice over time and adjustment to available hospital capacity, consideration should be given to adjusting for study week in the multivariate models

Response: We have now added study week as a predictor. In fact the results of this variable were rather interesting, and we have commented on them in the discussion.

New text, page 16: Last, we were interested to note that, while risk of hospitalization was constant across the study period, risk of critical illness (and directionally, but not significantly, mortality) decreased over time. Our institution was stretched but not overwhelmed by the epidemic and did not experience significant equipment or treatment shortages. The improvement in outcomes over time (in the setting of a functioning health system) raises the possibility that familiarity with the disease, ongoing iteration of protocols and practices in response to observed outcomes, and initiation of new treatments may improve outcomes even in the absence of vaccination or regimens known to be effective.

- Given that critical illness could occur upon admission or much later during hospitalization, may consider using a survival model approach instead

Response: Our data did not meet the proportional hazards assumptions required for survival modeling, and we have only 0.4% of patients right censored, so we retained logistic regression for this analysis. However, we added a competing risks mortality model.

Other issues to address:
1. Introduction: Page 4, last paragraph and Page 5, 1st paragraph. This important paper suffers a little from trying to accomplish too much for too many audiences (epidemiologists/policy makers, outpatient docs, ED docs, inpatient docs). The focus on critical illness among those who are hospitalized is the strongest contribution.

Response: For now we have elected to retain the analysis of hospitalization risk; however, if the editor prefers, we can remove it.

2. Methods: Study setting. It would be helpful to list the hospital names (so people whether NYU Bellevue or the Children’s Hospital are included).

Response: We have added the hospital names; in this revision we have restricted to adult patients.

Revised text, page 5: The study was conducted at NYU Langone Health, which includes over 260 outpatient office sites and four acute care hospitals (Tisch Hospital and NYU Langone Orthopedic Hospital in Manhattan, NYU Langone Hospital – Brooklyn in Brooklyn, and NYU Winthrop on Long Island), ranging from a quaternary care hospital to a safety net institution.

3. Methods, p 6, Lines 24-33. Were emergency department testing results available on the same day (e.g under 3 hours?). In many EDs, same day tests are only sent if the patient is going admitted to the hospital. If patients are going to be discharged, patients may or may not be tested, and is often the case in many EDs, if tests are going to be sent, they will be send out tests that don't result for several days. This should be clarified. It may be helpful to include an author from the emergency medicine as the decision to hospitalize a patient is made by clinicians working in the ED practices may have change dramatically in the surge.

Response: Once testing moved in-house March 16 (before which only 115 patients had tested positive), all results whether outpatient, ED or inpatient were available within 24 hours. It is difficult for us to assess the frequency with which testing was NOT performed. However, the reviewer is correct that testing was relatively rarely done for patients treated and released from the ED. For this reason we have added a sensitivity analysis of ED treat and release not tested patients, as described above. We have also added this as a limitation. As per instructions from the editors, we can not add authors at this time.

New text, page 16: This limitation may be further exacerbated by the fact that patients treated and released from the ED were not commonly tested and thus omitted from our analysis unless later hospitalized, yet might be more likely to have comorbidities than those tested in outpatient settings. However, a sensitivity analysis including these patients showed similar results.

4. Discussion. Page 14, Lines 3-5. "we found comorbidities to be less strongly associated with critical illness once patients were hospitalized." Would consider rewording since the outcome of critical illness includes those who were critically ill upon admission and those who became critically ill after a few days of being hospitalized.

Response: We have reworded for clarity

Revised text: We found comorbidities to be less strongly associated with critical illness in hospitalized patients.

Reviewer: 3
My comments here are in line with my other review of manuscript BMJ-2020-057242

I would publish this article rather than the above mentioned manuscript as it involves more patients and also include etnicity.

I have read the article with great interest but as a non-clinician it is difficult to estimate how useful this article would be for the medical world. I would recommend the BMJ to publish from the most interesting articles on COVID-19 only the abstract with a link to the whole article. And ask one or two very experiences COVID-19 scientists to comment on the particular 'added value' of this article for their colleagues. More or less like the way Mike Makris is doing this on Twitter. As a patient expert, I find that most helpful to get a feeling of the most important developments, lack in research etc.

Overall, this type of overview of patients can be very helpful and certainly from my own experience is very labour-intensive.

Response: We very much appreciate having a patient reviewer comment on the utility of this manuscript for non-clinicians.

This article is helpful in understanding underlying comorbidities for patients and carers. But as etnicity is not mentioned, it doesn't say much about possible lack in access for certain groups as The Lancet recently described in their World Report of April 18, 2020 page 1243-4

Response: We suspect this is inadvertently copied from another review; our paper does include ethnicity as commonly defined in the US (Hispanic vs. non-Hispanic).

The large use of hydroxychloroquine is remarkable and maybe lower as elsewhere in the world. But as I mentioned above something I cann't judge exactly.

Response: Similarly, perhaps this is meant for another paper. We do not comment on use of hydroxychloroquine in this manuscript.

Personally, I believe that public or patient involvement in these type of research could not be very helpful. However, there are certainly lessons to be learnt for the after COVID-19 period in the sense that patient groups should focus much more on the importance of comorbidities and the need for lifestyle coaching for especially those underlying conditions where lifestyle advice can be of assistance to decrease the number with diseases like diabetes, kidney failure, large BMI etc.

And we have to take in account that COVID-19 outcomes in the US may greatly differ from the outcomes in other parts of the world as the US lacks a good health care infrastructure and Insurance as well as another social security system or none at all.

Reviewer: 4
This cohort study of COVID-19 subjects aims to assess the relationship between patient characteristics and hospitalization, and between patient characteristics and 'critical illness'.

There are a number of issues (methodological, analysis, interpretation and presentation of results) which need to be addressed:

1. The 'tested' cohort is a very heterogenous group of subjects: those presenting to ED with symptoms or clinician concern, ambulatory testing with clinician's referral, outpatient testing of symptomatic/concerned employees, and repeat testing of negative specimens at clinician discretion. This results in a very biased group of COVID-19 positive subjects, the structure of which also varies with time (since testing was restricted further during the study period). It would be useful to see information on patient characteristics by reason for testing. The reason for hospitalisation may also vary according to the reason for presenting -see below.

Response: We respectfully disagree that heterogeneity results in a biased group of subjects; rather, we feel it decreases bias by ensuring a wide range of patients are included, from asymptomatic to critically ill. In many other case series, only critically ill or hospitalized patients are included, generating a biased view of patient characteristics. We fully acknowledge, however, that the tested cohort does not represent a comprehensive picture of all patients with COVID-19, given that most patients who are less severely ill do not present to a healthcare setting at all, and conversely that many patients sick enough to

present to the ED were not tested. We do not, unfortunately, have “reason for testing” available. We have now added a sensitivity analysis with more ED patients, as noted above.

2. How subjective is the decision to admit? Does this vary by reason for presenting? Does it vary over the study period? Could this not bias the comparisons between those who were and who were not hospitalised? If so, then it is difficult to interpret the ‘hospitalised’ vs ‘non-hospitalised’ comparisons. Should ‘reason’ and ‘time’ not be included in the statistical model? Or, perhaps the formal statistical analysis here (Table 1) should be omitted? In addition, what was the extent of missing data on comorbidities/health characteristics for those subjects not hospitalised?

Reason: Presenting signs and symptoms are heterogeneous and difficult to classify in a limited enough fashion to permit analysis. However, as the reviewer recommends, we have now added week of study period to all analyses to account for time-varying propensity to admit. With respect to comorbidities, we can not distinguish between a patient that genuinely does not have a comorbidity and a patient in whom the comorbidity was never assessed; both will appear as no comorbidity. We have, however, tried to exclude patients for whom it is most likely that absence of data represents lack of assessment rather than lack of disease (N=287), and we have now added more comorbidities based on past history and encounter diagnoses. In general with COVID-19, we suspect there is relatively less heterogeneity of admission decisions, since patients are typically hospitalized on the basis of clinically important hypoxia, which is a relatively objective finding. Moreover, variation over time or by clinician would be a conservative bias, reducing differences between groups.

3. What proportion of those COVID-19 subjects who were not hospitalised might have been subsequently hospitalised/died (after the study period)?

Response: The prospective nature of our study allows us to capture future hospitalizations to our health system; we therefore know who has been hospitalized up to the present moment if that was in our health system. Moreover, as described above, our minimum follow up period of 28 days is now well beyond the point at which we would expect hospitalization to occur. In general for Medicare patients, about 80% of readmitted patients return to our institution. In this instance, we can not estimate what proportion we might be missing and have added this as a limitation.

New text, page 17: some discharged patients may have been readmitted elsewhere with critical illness, or may have died post-discharge.

4. For the cohort who were hospitalised, an endpoint (critical illness/discharged well) was unavailable for 5% (n=115). These cases (although a small number) are omitted from the logistic regression analyses which assess factors predictive for critical illness, and hence may introduce bias. It would be preferable to use this censored information (and time to critical illness) in a ‘survival-type’ analysis, rather than logistic regression. Also, ‘hospital’ and ‘time’ would seem to be appropriate additional factors to include in the model.

Response: In the updated dataset, we exclude only 0.4% of patients who have not experienced discharge or critical illness, so we have elected to retain the logistic regression. We have now accounted for clustering by hospital and for study week.

5. Furthermore, the large proportion of missing endpoints (ie those still hospitalised/ventilated/extubated but still hospitalised) for those with 'critical illness' (see Figure 1), means that 'final outcomes' for this subgroup are difficult to judge. At the very least a simple sensitivity-type analysis could be carried out with 'good' and 'bad' outcomes substituted for the 'missing' endpoints, to show the possible range in rates.

Response: In this revision, 24% (as opposed to 45%) of patients in this subgroup do not have "final" outcomes. We appreciate this suggestion and have now added an upper mortality bound in the discussion if everyone remaining died. We have also added a new competing risk survival model for mortality that accounts for right censoring.

New text, page 14: If all remaining hospitalized patients died, the overall mortality would be a maximum of 30.5%.

6. Predictors of hospitalisation/critical illness. Some of the 'significant' multivariable-adjusted factors need to be interpreted carefully. For example, in Table 1 the multivariable logistic regression odds ratio for hyperlipidemia of 0.63 (0.50-0.80), is interpreted by the study authors as showing a 'protective' effect of hyperlipidemia for hospitalization. However, from Table 1, the % of those with hyperlipidemia for the hospitalised cohort is higher than that for the non-hospitalised cohort. In fact, the observed (unadjusted) hospitalization rates (Table 1) are 625/866 = 72% for those with hyperlipidemia and (approximately, assuming no missing values) 1765/3728 = 47% for those without hyperlipidemia, ie a higher risk (not a lower one) for those with hyperlipidemia. The 'change' from a higher to a lower risk with statistical adjustment indicates a high degree of intercorrelation between patient characteristics, and can lead to misleading results. This introduces problems in the interpretation of any multivariable result. Tobacco use data seems also to be affected by intercorrelation (and the observed data show 'former' smokers, with perhaps underlying health conditions, have a greater risk of hospitalisation/critical illness)

Response: We agree this is always a concern in multivariable analysis. We have now formally tested interactions across all covariates simulataneously using a determinant of correlation matrix and found no significant results.

New text, page 8: We also tested for overall multicolllinearity among all variables simultaneously using the determinant of correlation matrix implemented in R's mctest library and found no significant results.

7. The maximum information gain decision tree classification analysis does not really add to the results, hence could be omitted from the paper.

Response: We have now omitted these results.

8. Table 1. Temperature results should be omitted here as they were missing for most of the non-hospitalised cohort. Should oxygen saturation also be omitted as these values were obtained from differing proportions of patients on supplemental oxygen?

Response: We included these purely as a point of information for readers who are interested in the clinical presentation of these patients; they are not in the hospitalization model. These are present in most other published case series. For now we have elected to retain them but would be glad to remove if the editors prefer.

9. Tables 1 and 2. The hospitalization/critical illness data might be easier to understand if hospitalization/critical illness rates were shown by characteristic category (ie row percentages rather than column percentages were shown).

Response: As noted above, we feel that differences between cohorts are more readily evident for casual readers using column percents. However, if the editors feel that row percentages are essential, we will change.

10. Table 3. The critical illness risk analysis for 'all patients' (ie those hospitalised plus those not hospitalised) is inappropriate surely as risk is dominated by decision to hospitalise? [None of the subjects not hospitalised had critical illness?]

Response: We have removed this analysis.

Items to include with your revision (see http://www.bmj.com/about-bmj/resources-authors/article-types/research):

1. What this paper adds/what is already known box (as described at http://resources.bmj.com/bmj/authors/types-of-article/research)

2. Name of the ethics committee or IRB, ID# of the approval, and a statement that participants gave informed consent before taking part. If ethics committee approval was not required, please state so clearly and explain the reasons why (see http://resources.bmj.com/bmj/authors/editorial-policies/guidelines.)

Response: We have added the IRB number; the remaining information was present.

3. Patient confidentiality forms when appropriate (see http://resources.bmj.com/bmj/authors/editorial-policies/copy_of_patient-confidentiality).

Response: N/A

4. Competing interests statement (see http://resources.bmj.com/bmj/authors/editorial-policies/competing-interests)

Response: added; there are no competing interests

5. Contributorship statement+ guarantor (see http://resources.bmj.com/bmj/authors/article-submission/authorship-contributorship)

Response: Already included

6. Transparency statement: (see http://www.bmj.com/about-bmj/resources-authors/forms-policies-and-checklists/transparency-policy)

Response: added

7. Copyright statement/licence for publication (see http://www.bmj.com/about-bmj/resources-authors/forms-policies-and-checklists/copyright-open-access-and-permission-reuse)

Response: Added

8. Data sharing statement (see http://www.bmj.com/about-bmj/resources-authors/article-types/research)

Response: Already included

9. Funding statement and statement of the independence of researchers from funders (see http://resources.bmj.com/bmj/authors/article-submission/article-requirements).

Response: Already included

10. Patient and public involvement statement

https://docs.google.com/document/d/1djgVLEUFtPQzLpf5HyuiFcMgrNJ_o7Yb8z73zgXxXYM/e

Response: Already included

11. Dissemination plans: At the end of the paper please state how the results of your study have been (or will be) sent to patients and the public under the heading “Dissemination plans”. If you have prepared a lay summary eg for your funders, please include it in a supplementary file.

If you have not disseminated and have no plans to do so, please state why.

Response: Added

12. Patient confidentiality forms when appropriate

Response: N/A

13. Please ensure the paper complies with The BMJ’s style, as detailed below:

a. Title: this should include the study design eg "systematic review and meta-analysis.”

b. Abstract: Please include a structured abstract with key summary statistics, as explained below (also see http://resources.bmj.com/bmj/authors/types-of-article/research). For every clinical trial - and for any other registered study- the last line of the abstract must list the study registration number and the name of the register.

Please report all outcomes that were listed in the trial registry, or explain that you will publish them elsewhere. Please clearly identify each outcome as primary, secondary, or post-hoc in the text, abstract, and any tables or figures. We expect authors to report prespecified outcomes. If outcomes in the trial registry have later been changed, please explain the reasons for the change and the dates of the change in the paper. You may report the changed outcomes, but we will expect you to also report on the originally specified outcomes unless otherwise agreed

with the handling editor for your paper.

Occasionally the outcomes that are prespecified in a trial registry do not match up with those included in the trial protocol. When there are discrepancies between protocol and registry specified outcomes, we expect the paper to report and interpret the registry specified outcomes. You may also report any protocol specified outcomes, but if you do please be sure to include the date of the protocol and the point at which each outcome was added to the protocol, and explain why the registry entry differed from the protocol and why the registry was not updated to reflect any protocol changes.

c. Introduction: This should cover no more than three paragraphs, focusing on the research question and your reasons for asking it now.

d. Methods: For an intervention study the manuscript should include enough information about the intervention(s) and comparator(s) (even if this was usual care) for reviewers and readers to understand fully what happened in the study. To enable readers to replicate your work or implement the interventions in their own practice please also provide (uploaded as one or more supplemental files, including video and audio files where appropriate) any relevant detailed descriptions and materials. Alternatively, please provide in the manuscript urls to openly accessible websites where these materials can be found.

e. Results: Please report statistical aspects of the study in line with the Statistical Analyses and Methods in the Published Literature (SAMPL) guidelines http://www.equator-network.org/reporting-guidelines/sampl/. Please include in the results section of your structured abstract (and, of course, in the article's results section) the following terms, as appropriate:

i. For a clinical trial: Absolute event rates among experimental and control groups; RRR (relative risk reduction); NNT or NNH (number needed to treat or harm) and its 95% confidence interval (or, if the trial is of a public health intervention, number helped per 1000 or 100,000.)
ii. For a cohort study: Absolute event rates over time (eg 10 years) among exposed and non-exposed groups; RRR (relative risk reduction.)
iii. For a case control study:OR (odds ratio) for strength of association between exposure and outcome.
iv. For a study of a diagnostic test: Sensitivity and specificity; PPV and NPV (positive and negative predictive values.)
v. For a systematic review and/or meta-analysis: Point estimates and confidence intervals for the main results; one or more references for the statistical package(s) used to analyse the data, eg RevMan for a systematic review. There is no need to provide a formal reference for a very widely used package that will be very familiar to general readers eg STATA, but please say in the text which version you used. For articles that include explicit statements of the quality of evidence and strength of recommendations, we prefer reporting using the GRADE system.

Please report all outcomes that were listed in the trial registry, or explain that you will publish them elsewhere. Please clearly identify each outcome as primary, secondary, or post-hoc in the text, abstract, and any tables or figures. We expect authors to report prespecified outcomes. If outcomes in the trial registry have later been changed, please explain the reasons for the change and the dates of the change in the paper. You may report the changed outcomes, but we will expect you to also report on the originally specified outcomes unless otherwise agreed with the handling editor for your paper.

Occasionally the outcomes that are prespecified in a trial registry do not match up with those

included in the trial protocol. When there are discrepancies between protocol and registry specified outcomes, we expect the paper to report and interpret the registry specified outcomes. You may also report any protocol specified outcomes, but if you do please be sure to include the date of the protocol and the point at which each outcome was added to the protocol, and explain why the registry entry differed from the protocol and why the registry was not updated to reflect any protocol changes.

f. Discussion: To minimise the risk of careful explanation giving way to polemic, please write the discussion section of your paper in a structured way. Please follow this structure: i) statement of principal findings of the study; ii) strengths and weaknesses of the study; iii) strengths and weaknesses in relation to other studies, discussing important differences in results; iv) what your study adds (whenever possible please discuss your study in the light of relevant systematic reviews and meta-analyses); v) meaning of the study, including possible explanations and implications for clinicians and policymakers and other researchers; vi) how your study could promote better decisions; vi) unanswered questions and future research

g. Footnotes and statements

Online and print publication: All original research in The BMJ is published with open access. Our open access policy is detailed here: http://www.bmj.com/about-bmj/resources-authors/forms-policies-and-checklists/copyright-open-access-and-permission-reuse. The full text online version of your article, if accepted after revision, will be the indexed citable version (full details are at http://resources.bmj.com/bmj/about-bmj/the-bmjs-publishing-model).