PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Description of symptom course in a telemedicine monitoring clinic for acute symptomatic covid-19: a retrospective cohort study
AUTHORS	O'Keefe, James B; Tong, Elizabeth; O'Keefe, Ghazala; Tong, David

VERSION 1 – REVIEW

REVIEWER	Pieter Cohen Harvard Medical School I receive compensation for a chapter in UptoDate regarding outpatient management of COVID-19.
REVIEW RETURNED	13-Sep-2020

GENERAL COMMENTS	The research addresses an important topic, what are the risk factors for severe disease among patients with COVID-19 in the outpatient setting. However, the authors need to be careful to align their conclusions with the methodology of their study. Methods The authors stated goal is to understand the natural history of COVID-19, but the most severely affected patients – those hospitalized – and those least affect – those that reported symptoms infrequently—were excluded. These clinical outcomes are only known after-the-fact, therefore these exclusion criteria limits the study to only typical symptoms of moderately affected patients who are never hospitalized and do not recover swiftly. If that is the purpose, to understand the persistence of symptoms at 30 days for that subgroup of this patients, then the decision is reasonable, but the authors should focus their entire manuscript on only that specific subgroup of patients and be clear that these results cannot predict disease course or help counsel patients in realtime. For example, the statement in the Discussion that "Patients and providers may be reassured that gradual resolution of symptoms is typical based on our findings" is not supported by the study, since all patients who would go on to be hospitalized were excluded from the study. Nor can the authors' results be used to help patients anticipate recovery time as they excluded patients who recovered rapidly and those that deteriorated. It would also be useful to add additional information about what were the factors that led to patients being tested for COVID-19 during this time in Atlanta? What was the total number of patients who tested positive for SARS-CoV-2 compared to the number that required outpatient telemedicine management? It is described that patients were tested in testing clinics or emergency departments. The number of each is important as well as providing more details

	about the testing sites: Were these sites medical offices? Drive-through testing? Or another system.

REVIEWER	Miguela Caniza St. Jude Children's Research Hospital
REVIEW RETURNED	24-Sep-2020

GENERAL COMMENTS	Summary O'Keefe JB et al, present Predictors of disease duration and symptom course of outpatients with acute COVID-19: a retrospective cohort study a single-center experience, retrospective cohort study of non-hospitalized patients with COVID-19 to evaluate factors predicting duration of symptoms. Other results were published from the same patient cohort (O'Keefe JB, et al. Initial MedRxiv, 2020). In this current report, the cohort consisted of 273 patients with COVID-19, and among the criteria for exclusion were hospitalization (risk factors for hospitalization were already reported in a previous publication, O'Keefe JB, MedRxiv, 2020). The investigators studied each of the symptoms and established the duration of such symptoms; they also correlated initial overall COVID severity to the duration of symptoms. They concluded that the progression of symptoms was from systemic symptoms to lower respiratory symptoms, finding that cough and loss of sense of smell and teste were the symptoms that persisted the longest; and the PASS is the best predictor of symptoms duration. Comments 1. Title: appropriate 2. Abstract: Appropriate Introduction, 3. please clarify here that the cohort that required hospitalization was reported already [#19. O'Keefe JB, et al. Initial Experience in Predicting the Risk of Hospitalization of 496 Outpatients with COVID-19 Using a Telemedicine Risk Assessment Tool. MedRxiv, 2020]. Providing this information, for readers unfamiliar with the work of these investigators, will effectively clarify why hospitalized patients were excluded. Methods 4. (Box 1, line 37), how these risks match the PASS categorization? Please explain if those patients with "mild" PASS also belongs to patients who are in "at low risk" also, please clarify how the patients for low, intermediate and high risks were categorized, for what risk? 5. (page 8/34, line 13), please mention what were the reasons for outpatient COVID testing. 6. (page9/34, line 56), is "VOMC provider intake assessment" similar to "intake telemedicine visit" in Box 1. If yes, please reconcile the name differences. Also, "VOMC follow-up telephone call" with "follow-up phone calls..." in Box 1, if these refer to similar items. Providing similar designation will facilitate reading. 7. Please establish what is day 0 from where counting begins. 8. The analysis was overall very descriptive. I am glad to see multiple testing was adjusted; however, it was a longitudinal study and I do not see any models tailored toward that. In other words, symptoms should be time-varying or evolve. The methods presented currently (ANOVA) do not accommodate the time-dynamic feature readily. Time course should be considered when characterizing risk factors.

	Results 9. Because this was a longitudinal study, how many entries were missing, and how missingness was handled was unclear. 10. (page 11/34, line 3), assure to list all the symptoms, some of them are missing (example fever). 11. Explain why those who were categorized as “severe” were not hospitalized? Discussion 12. To put in context this information, please include a thought of how COVID symptoms are similar/different from other coronavirus and respiratory viral illnesses. 13. References: appropriate Table/Figures 14. (page 29/34) Figure 1, please explain the color used to best appreciate the figure.

REVIEWER	James Bentham University of Kent, UK
REVIEW RETURNED	08-Oct-2020

GENERAL COMMENTS	The authors present an interesting and detailed description of Covid patients' symptoms over a period of time, and I think the heat maps in the paper are particularly effective. I have some major comments, particularly on the statistical analysis: 1. It seems sensible to transform the data, but it's not clear what the model is here. What are the predictors in the models that are being compared using the AIC and BIC? How does this relate to the ANOVA? It would seem more natural to use a single GLM for duration of symptoms with a log link, and then include factors such as symptom severity or sex as covariates. The authors should justify the rationale behind their choice of model, or adjust this choice. 2. On page 12, what is the justification for dividing the p-value by 10 for multiple comparisons? 3. Overall, it isn't clear what corrections have been made for multiple testing, for example in the ANOVA and chi-squared tests. This should be stated explicitly in the methods. 4. A difference in AIC or BIC of 50 is highly significant, so I think the better fitting model should be chosen, particularly given that the estimates are being transformed back to the original scale. 5. In Table 1, it isn't clear what the various p-values are. The ANOVA should give a single p-value with a null hypothesis of the group means being the same. It is possible to calculate p-values for comparisons between pairs of groups, but this isn't stated in the methods. 6. In Table 1, the counts for mild, moderate and severe symptom severity add to 254, while the totla sample size is 273. It should be made clearer in the text why this is. I also have some minor comments: 1. Consistent numbers of decimal places should be used, e.g., in the abstract, there is a confidence interval 1.15 to 1.4. I think 1 dp is appropriate accuracy. 2. In the abstract, the confidence interval for cough should be 11.0 to 13.6, I think. Similarly, loss of smell or taste should be 10.0 to 12.2. 3. In Table 1, the lines for symptom onset to first visit and phone call should have units.

	4. There are a few typos, e.g, for joint pain in Table 1, which should be checked. 5. On p15, "symptoms at initial visit", the value for joint pain is 21%. Later in the sentence, 11% is for shortness of breath at rest. 6. p16, first paragraph, loss of smell/taste for the moderate group has a maximum 63% (at day 8) 7. p16, first paragraph. The authors mention that the heat map for severe PASS has 75% for various symptoms, but this doesn't match the figure, which has 100% in various categories, including 100% on day 30 for shortness of breath with exertion 8. p17 "Symptom duration by Provider-Assessed Symptom Severity". The authors state that PASS was signficantly correlated with more symptoms on ANOVA. Is this a correlation or association? I think it might be the latter. 9. p17. Given that multi-way ANOVA was not performed, the methods should be updated to reflect this. 10. p17. The description of signficanct differences in durations isn't clear, as it doesn't appear to match the p-values above the bars in the figures. This should be edited for clarity. 11. The heat maps should be reordered to have all patients, mild, moderate and severe in order.

REVIEWER	Zhiying You CU Anschutz Medical Campus Denver, CO USA
REVIEW RETURNED	12-Oct-2020

GENERAL COMMENTS	It is a very interesting and important study in the midst of the pandemic. I have a minor statistical concern as follows. There were 26 patients who did not receive follow up call were excluded. These patients might have very short duration of symptoms and were more likely to NOT be missing at random. This might imply potential bias in favor of longer duration and should be considered a potential limitation. I have one minor comment: Survival analysis can be an alternative to ANOVA. For example, the Cox model does assume distribution of the time/duration if applicable

VERSION 1 – AUTHOR RESPONSE

Reviewer(s)' Comments to Author:		
Reviewer: 1, Pieter Cohen		
The research addresses an important topic, what are the risk factors for severe disease among patients with COVID-19 in the outpatient setting. However, the authors need to be careful to align their conclusions with the methodology of their study.	We appreciate the expert clinical attention of this reviewer to the study population selection. We believe the reviewer mis-states our study aims (stating "what are the risk factors for severe disease"). Our aim is to describe the symptom course in outpatients (telemedicine cohort). We have studied outpatient risk assessment (the topic suggested by the reviewer) in a separate manuscript (reference 19).	Line 126-129

	We have clarified our study aims in the introduction and cited our other paper for risk factors.	
Methods		
The authors stated goal is to understand the natural history of COVID-19, but the most severely affected patients – those hospitalized – and those least affect – those that reported symptoms infrequently—were excluded. These clinical outcomes are only known after-the-fact, therefore these exclusion criteria limits the study to only typical symptoms of moderately affected patients who are never hospitalized and do not recover swiftly. … Nor can the authors' results be used to help patients anticipate recovery time as they excluded patients who recovered rapidly and those that deteriorated.	We agree that the exclusion criteria (our exclusion of hospitalized patients and minimally symptomatic patients) were too broad. We have now included these patients in the study cohort so that the paper is more applicable to the general outpatient population seen in a telemedicine clinic after COVID-19 diagnosis.	Line 154-155, also updated result section with newly included patients
It would also be useful to add additional information about what were the factors that led to patients being tested for COVID-19 during this time in Atlanta? What was the total number of patients who tested positive for SARS-CoV-2 compared to the number that required outpatient telemedicine management? It is described that patients were tested in testing clinics or emergency departments. The number of each is important as well as providing more details about the testing sites: Were these sites medical offices? Drive-through testing? Or another system.	Thank you for bringing this to our attending. These data have been added to the paper. Test criteria are in Box 1 and overall testing numbers in the results section.	Line 166-179, Line 328-332
Reviewer: 2, Miguela Caniza		
Comments		
Title: appropriate		
Abstract: Appropriate		
Introduction,		
please clarify here that the cohort that required hospitalization was reported already [#19. O'Keefe JB, et al. Initial Experience in Predicting the Risk of Hospitalization of 496 Outpatients with COVID-19 Using a Telemedicine Risk Assessment Tool. MedRxiv, 2020]. Providing this information, for readers unfamiliar with the work of these investigators, will effectively clarify why hospitalized patients were excluded.	Per other reviewer, hospitalized patients are now included. We do clarify this modification in the paper and we do address the concern about clearly reporting our other study (risk factors for hospitalization) by citing our other study prominently in the introduction.	Line 182

Methods		
(Box 1, line 37), how these risks match the PASS categorization? Please explain if those patients with “mild” PASS also belongs to patients who are in “at low risk” also, please clarify how the patients for low, intermediate and high risks were categorized, for what risk?	(1) We have changed our analysis to use "symptom severity" at the initial VOMC visit rather than "Provider-Assessed Symptom Severity" i.e. PASS, which is now removed). We use provider-assessed symptom severity if available and patient-reported symptom severity if not. Both fields were in our initial note template and we find results are significant using either provider-assessed and/or patient reported and that the focus on the "provider-assessed symptom severity (PASS)" in our initial version detracted from the overall message. We do leave the definitions of provider-assessed severity in Box 1 to help explain the provider perspective on symptom assessment. (2) The “risk” groups are not included in this study except in Box 1 as a reference and they are detailed fully in reference 19. To clarify the difference between symptom severity and “risk”: Symptom severity was one factor in overall risk of hospitalization (i.e. a patient with moderate symptoms is higher risk than mild symptoms), but age and comorbidity also factor in for overall risk score (i.e. an older patient can be “high risk” for hospitalization even with “mild” symptom severity).	Line 260-263
(page 8/34, line 13), please mention what were the reasons for outpatient COVID testing.	Thank you for this request - we have clarified the outpatient testing criteria, settings, and volumes in Box 1 and in the results section.	Line 166-179, Line 329-333
(page9/34, line 56), is “VOMC provider intake assessment” similar to "intake telemedicine visit" in Box 1. If yes, please reconcile the name differences. Also, “VOMC follow-up telephone call” with "follow-up phone calls..." in Box 1, if these refer to similar items. Providing similar designation will facilitate reading.	Thank you for the clarification. The initial telemedicine visit for all patients in our cohort is now referred to as the VOMC Intake Visit throughout and this is defined in the introduction.	Line 122

Please establish what is day 0 from where counting begins.	Thank you for clarifying. Day 1 was the first day a patient had any symptoms to allow for a clear denotation of patients with only a single day of symptoms as "day 1." We have added this definition to the paper.	Line 245
The analysis was overall very descriptive. I am glad to see multiple testing was adjusted; however, it was a longitudinal study and I do not see any models tailored toward that. In other words, symptoms should be time-varying or evolve. The methods presented currently (ANOVA) do not accommodate the time-dynamic feature readily. Time course should be considered when characterizing risk factors.	Thank you for this important insight. We have changed methods to survival analysis to assess for time course. We did not want to look at covariates over time because the goal was for a provider to be able to give some guidance as to how long clinically relevant symptoms may last at the first intake visit.	Line 281-312
Results		
Because this was a longitudinal study, how many entries were missing, and how missingness was handled was unclear.	There was 15% missing data between first VOMC visit and last phone call. Nothing was done to impute these values so the data for the heat maps may be a slight underestimation. The missing data does not affect the survival analysis because duration is time from each symptom's onset to resolution (or censoring) of the symptom so missing data in the middle does not affect the analysis.	N/A
(page 11/34, line 3), assure to list all the symptoms, some of them are missing (example fever).]	Thank you for clarifying. We did add Fever to "Systemic" and Wheezing to "Lower Respiratory." We did not include rarely reported symptoms that are not commonly grouped in other papers in the literature: Rash and Confusion.	Line 257-260
Explain why those who were categorized as "severe" were not hospitalized?	Thank you for this important insight (that severe symptoms usually merit hospitalization based on our definitions and standard practice). This is a retrospective study so we can only gleam what we can from chart review. The four patients with severe symptoms who were not hospitalized had significant systemic symptoms (e.g. severe myalgias, headache) and lower respiratory symptoms (e.g. breathless on minimal exertion) and were all monitored by the high-risk team (twice daily APP calls) using pulse	Line 342-347

	oximetry, where appropriate. Two were referred to the in-person respiratory clinic and determined to be stable for outpatient monitoring and the other two were managed by telemedicine only.	
Discussion		
To put in context this information, please include a thought of how COVID symptoms are similar/different from other coronavirus and respiratory viral illnesses.	For initial symptom data, the most pertinent available comparison is to SARS and MERS and have added this to the paper. Comparison to influenza (Powers et al, Value Health. 2018 Feb) and rhinovirus (JAMA Nov 6 1967, Vol 202, No 6) may be possible but we do not think merit the expansion of the discussion due to target audience familiarity with these entities.	Line 507-519
References: appropriate		
Table/Figures		
(page 29/34) Figure 1, please explain the color used to best appreciate the figure.	Legend added.	Figure 1
Reviewer: 3, James Bentham		
I have some major comments, particularly on the statistical analysis:		
1. It seems sensible to transform the data, but it's not clear what the model is here. What are the predictors in the models that are being compared using the AIC and BIC? How does this relate to the ANOVA? It would seem more natural to use a single GLM for duration of symptoms with a log link, and then include factors such as symptom severity or sex as covariates. The authors should justify the rationale behind their choice of model, or adjust this choice.	Thank you for this important insight; we have changed our analysis to survival analysis and re-written the paper to reflect this change.	Line 281-312
2. On page 12, what is the justification for dividing the p-value by 10 for multiple comparisons?	As above, we have changed to survival analysis.	N/A
3. Overall, it isn't clear what corrections have been made for multiple testing, for example in the ANOVA and chi-squared tests. This should be stated explicitly in the methods.	As above, we have changed to survival analysis.	N/A
4. A difference in AIC or BIC of 50 is highly significant, so I think the better fitting model should be chosen, particularly given that the estimates are being transformed back to the original scale.	As above, we have changed to survival analysis.	N/A

5. In Table 1, it isn't clear what the various p-values are. The ANOVA should give a single p-value with a null hypothesis of the group means being the same. It is possible to calculate p-values for comparisons between pairs of groups, but this isn't stated in the methods.	We have changed this to a single p value	Table 1
6. In Table 1, the counts for mild, moderate and severe symptom severity add to 254, while the totla sample size is 273. It should be made clearer in the text why this is.	We have made changes to make the numbers match: These numbers did not match because not all patients had a provider-assessed symptom severity recorded on intake visit template. We realize that excluding those patients (with missing provider-assessed severity) may have resulted in worse biases and so therefore we now include them using patient-reported initial symptom severity if provider severity is missing. Details of this approach are added to methods section.	Line 260
I also have some minor comments:		
1. Consistent numbers of decimal places should be used, e.g., in the abstract, there is a confidence interval 1.15 to 1.4. I think 1 dp is appropriate accuracy.	This is corrected for ages and dates to 1 decimal place	N/A
2. In the abstract, the confidence interval for cough should be 11.0 to 13.6, I think. Similarly, loss of smell or taste should be 10.0 to 12.2.	N/A - new analysis has been done and these numbers removed	N/A
3. In Table 1, the lines for symptom onset to first visit and phone call should have units.	This is added.	Table 1
4. There are a few typos, e.g, for joint pain in Table 1, which should be checked.	This is corrected.	Table 1
5. On p15, "symptoms at initial visit", the value for joint pain is 21%. Later in the sentence, 11% is for shortness of breath at rest.	This sentence has been removed (we decided not to list results for the entire column that can be easily seen in the table).	Table 1
6. p16, first paragraph, loss of smell/taste for the moderate group has a maximum 63% (at day 8) 7. p16, first paragraph. The authors mention that the heat map for severe PASS has 75% for various symptoms, but this doesn't match the figure, which has 100% in various categories, including 100% on day 30 for shortness of breath with exertion 8. p17 "Symptom duration by Provider-Assessed Symptom Severity". The authors state that PASS was signficantly correlated with more	We have re-written the heatmap description in the results to note the correct symptom frequency on each illness day. We have re-done the analysis of the paper to show the association of symptom severity with duration.	Lines 362-395

symptoms on ANOVA. Is this a correlation or association? I think it might be the latter.		
9. p17. Given that multi-way ANOVA was not performed, the methods should be updated to reflect this.	As above, we have changed to survival analysis.	N/A
10. p17. The description of signficanct differences in durations isn't clear, as it doesn't appear to match the p-values above the bars in the figures. This should be edited for clarity.	As above, we have changed to survival analysis.	N/A
11. The heat maps should be reordered to have all patients, mild, moderate and severe in order.	Thank you for this comment, we have arranged heat maps in recommended order.	Figure 1
Reviewer: 4, Zhiying You		
It is a very interesting and important study in the midst of the pandemic. I have a minor statistical concern as follows. There were 26 patients who did not receive follow up call were excluded. These patients might have very short duration of symptoms and were more likely to NOT be missing at random. This might imply potential bias in favor of longer duration and should be considered a potential limitation.	Thank you for this comment. We have re-done the analysis, now to include hospitalized and patients with short symptoms durations.	Line 182
I have one minor comment: Survival analysis can be an alternative to ANOVA. For example, the Cox model does assume distribution of the time/duration if applicable.	As above, we have changed to survival analysis.	Line 281-312

VERSION 2 – REVIEW

REVIEWER	Pieter Cohen Harvard Medical School, USA I receive payments from UptoDate for the chapter on Outpatient Management of COVID-19.
REVIEW RETURNED	08-Dec-2020

GENERAL COMMENTS	The authors provide important information on the natural history of outpatient COVID-19 managed by a telemedicine team. The revision is greatly improved, and the authors have done significant work to address this reviewer and other reviewers' initial comments. The manuscript will make a valuable contribution to the literature on COVID-19. Addressing several additional issues will further strengthen the manuscript. Consider carefully if the patients included are sufficiently alike to be a useful cohort. The authors emphasize as their principal

	findings that this study describes “the longitudinal symptom course…for a telemedicine cohort”. That is important, but the current combination of patients testing positive in the outpatient setting, ED and inpatient does not entirely support the focus of the study. From prior studies, we already have a lot of information about the presentation of patients found to have COVID-19 in the ED and then hospitalized. The strength of the current research is that it focuses on patients who, at least at the beginning, are managed by telemedicine. Given that strength, I am confused by the inclusion of two patients who were tested and found to have COVID-19 as inpatients. These two patients, in my opinion, should be excluded given the purpose (and title) of this paper is understanding the outpatient presentation/natural history. Likewise, I am not sure how many of the 34 patients tested in the ED were hospitalized at that initial ED visit. It would seem that only patients tested in the ED but then discharged to outpatient care should be included in the current study. The authors mention an “in person covid-19 clinic” only in passing on page 13. The overarching structure of care, including this clinic, should be carefully described earlier in the Methods “study setting” section. It would also be important to explain if only 2 patients of the entire cohort required evaluation in this clinic or if others were evaluated. If patients were evaluated in this clinic, were the notes from those visits reviewed and incorporated into the symptom reported on the heatmap? Title should be reconsidered. Does this study truly capture outpatients? Or is this a study of a subset of symptomatic outpatients who could be managed effectively by telemedicine. Does “predictors of” refer only to “disease duration” or also to “symptom course” if so, what is the difference? Does this study capture the natural history of mild to moderate COVID-19? Should telemedicine be incorporated into the title? Introduction. Since the authors submitted this revision, a similar study has now been published in Annals of Internal Medicine, https://doi.org/10.7326/M20-5926, this recently published study should be referenced and the similarity and differences should be now included in the Discussion. Methods. It’s unclear if the outpatient testing was performed in a drive-through setting or if this involved clinician evaluation or a combination of the two. This should be clarified and specifics provided. Methods. What was the delay between the testing day and the results being reported? How might have this affected the results? Bias section (p10). The authors suggest that the voluntary nature of the enrollment biased the study toward the more symptomatic patients. But without a comparison of those enrolled and those not enrolled, this is not supported. Might it also be possible, that more anxious patients, maybe with only mild symptoms, would be more likely to enroll? Bias section (p10). The authors describe “screening in emergency room” – it should be clarified if the ED was screening all patients or using the test diagnostically, not as a screening test, as suggested in the Box.

	Results. This section should begin with the overall number of patients testing positive for SARS-CoV-2 in the outpatient setting (appears to be 730) and explain how that 730 positive tests led to 340 patients included in the current study. As it's hard to determine how representative the findings are without additional details: for example, how many patients were eligible for VOMC enrollment but declined participation. Furthermore, it would be ideal to compare basic demographic information from the patients not enrolled to those enrolled to help determine how this might have affected the results. Limitations: An additional limitation is that the study captures the natural history of COVID-19 patients managed by telemedicine. As there was not screening for COVID-19 at the time, it does not capture milder cases not requiring contacting physicians nor many cases requiring in person outpatient management (or maybe it does, this is not clear from the current manuscript). Comparison with other studies. The sentence (page 22) "In our experience…convalescent illness" is not clear. Figure 1. Recommend removing the heatmap of 4 patients with severe initial symptoms. This small number of patients is too small to be representative and, visually, gives the misperception that there is as much data about this group as the other two groups. Figures 2 & 3: it is not entirely clear why this information needs figures.

REVIEWER	James Bentham University of Kent, UK
REVIEW RETURNED	06-Jan-2021

GENERAL COMMENTS	The paper is now ready for publication. I only have two very minor comments: 1. I think the 'Bias' section should be renamed 'Potential biases' 2. On p18, it looks like the confidence interval for 178% should be 57-392.

REVIEWER	Zhiying You CU Anschutz Medical Campus, USA
REVIEW RETURNED	14-Dec-2020

GENERAL COMMENTS	There are some statistical concerns found as follows. 1. Page 11, including some more details about how the data has been summarized will help readers understand the tables and figures. 2. Page 17-22, it does not sounds enough to not use the Cox model in analysis because of that the assumption of proportional hazard is not met. Given its advantages, a Cox model with nonproportional hazards should be tried if appropriate. 3. Page 14, Table 1, because of censored data, ANOVA is not applicable in the comparison of days to symptom onset and duration of symptom. If there is not censored data, then what the distributions of the duration look like? ANOVA is most applicable for data from population of normal distribution.

	4. For all tables and figures: A footnote should be included if appropriate to make the table/figure self-explanatory, guiding the readers to understand it. For example, in page 3, “Loss of smell or taste had the second longest duration with 14 days (12 to 17), with 38% not resolved at final visit” implies that there are censored data. Then in Table 1 in page 14, how readers should understand the results for “Symptom onset to intake VOMC visit” and “Symptom onset to last phone call”? 5. Page 19, line 3, what is baseline group and how it has been defined? Is it the control group? 6. Page 19, Table 3, the symptom column, that each symptom show up once will make the table more clear and easier to read. 7. Page 3, line 40, “Initial symptom severity is the best predictor of disease duration” sounds better to change to “Initial symptom severity is a significant predictor of disease duration for most considered symptoms” or so.

VERSION 2 – AUTHOR RESPONSE

Response to reviewers

Comment	Author Response	Location of change
Reviewer #1		
Consider carefully if the patients included are sufficiently alike to be a useful cohort. … The strength of the current research is that it focuses on patients who, at least at the beginning, are managed by telemedicine. Given that strength, I am confused by the inclusion of two patients who were tested and found to have COVID-19 as inpatients. These two patients, in my opinion, should be excluded given the purpose (and title) of this paper is understanding the outpatient presentation/natural history.	We agree with the reviewer. We have removed: 1. The two patients diagnosed with covid-19 as inpatients 2. One patient who was tested in ED but admitted to hospital prior to VOMC visit All other patients (including patients tested at outpatient sites and ED) were not admitted prior to VOMC. The cohort is now 337 patients (from 340) and the findings of our study are unchanged.	Methods: Page 5, line 25

Likewise, I am not sure how many of the 34 patients tested in the ED were hospitalized at that initial ED visit. It would seem that only patients tested in the ED but then discharged to outpatient care should be included in the current study.	We again agree. Only 1 other person tested in ED was hospitalized at that ED visit. That person has now been excluded as mentioned above.	Methods: Page 5, line 25
The authors mention an "in person covid-19 clinic" only in passing on page 13. The overarching structure of care, including this clinic, should be carefully described earlier in the Methods "study setting" section. It would also be important to explain if only 2 patients of the entire cohort required evaluation in this clinic or if others were evaluated. If patients were evaluated in this clinic, were the notes from those visits reviewed and incorporated into the symptom reported on the heatmap?	We agree with the reviewer that it is important context to understand the sites of outpatient covid-19 care. This question is about our "Acute Respiratory Clinic" (ARC) site. Similar to an emergency department, the ARC is a site of care available for episodic respiratory care (covid-19 confirmed or not). This site had only modest overlap of patients with our research site (the telemedicine program), with 12% of VOMC patients seen at ARC at any time. We have included background details about the ARC to the Methods section and to Box 1. We did not audit in-person notes (ARC or emergency department) because the VOMC follow-up teams (RN + APP) maintained more complete daily symptom data in template form for patients, even those seen in ARC or the ED. 1. VOMC had a more robust method of collecting onset/initial symptom data than ARC with a previsit nurse call (average 40 min) and physician telemedicine visit (40 min), documenting the symptom onset/offset dates. ARC visit providers would document symptoms present/absent/resolved without taking the time to record specific dates. 2. ARC visits were intentionally shorter (average 7-10 minutes face-to-face) because of infection prevention considerations and providers would not document in the room. Therefore, they were	Page 6 lines 10-14 And Box 1 (page 7 line 11-18)

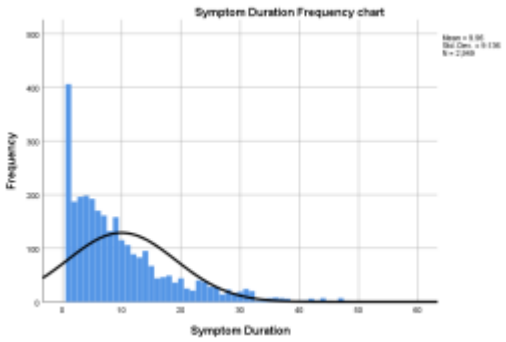
	less detailed in symptom reporting if not relevant to the visit complaint. 3. If a VOMC patient was seen in ARC, it was by specific referral from the VOMC provider. The VOMC would be in close touch (twice daily calls for escalated patients) and would include a summary of the ARC visit and updated symptom list in the VOMC template for that day.	
Title should be reconsidered. Does this study truly capture outpatients? Or is this a study of a subset of symptomatic outpatients who could be managed effectively by telemedicine. Does "predictors of" refer only to "disease duration" or also to "symptom course" if so, what is the difference? Does this study capture the natural history of mild to moderate COVID-19? Should telemedicine be incorporated into the title?	Thank you for this targeted feedback. We agree and have changed the title to removed "outpatients" and add "telemedicine": "Description of symptom course in a telemedicine monitoring clinic for acute symptomatic covid-19: a retrospective cohort study"	Page 1 Title
Since the authors submitted this revision, a similar study has now been published in Annals of Internal Medicine, this recently published study should be referenced and the similarity and differences should be now included in the Discussion.	Thank you for bringing this Annals of Internal Medicine study to our attention and we have included it in our discussion. It is a brief study and our paper adds to this work because many symptoms are still at high prevalence after day 10 (they do not show data during days 11-30) and we analyse predictors of duration.	Page 24 3-8
Methods. It's unclear if the outpatient testing was performed in a drive-through setting or if this involved clinician evaluation or a combination of the two. This should be clarified and specifics provided.	Thank you for pointing out this area for clarification. We have updated our manuscript with details about the testing sites.	Page 6 line 9-10

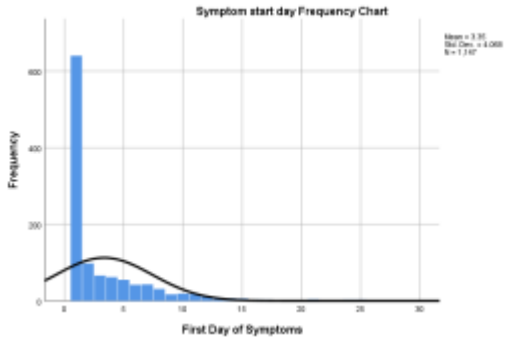
Methods. What was the delay between the testing day and the results being reported? How might have this affected the results?	The usual timing during the study period was 24-48 hours from time of test to patient notification of result. We do not have the specific data (time of test to notification) for this study cohort. Because there are multiple factors that could delay VOMC intake, we selected an exclusion criteria of 10 days from symptom onset to account for these multiple factors and we report the average number of days from symptom onset to VOMC intake visit in Table 1 (5-6 days). We agree with the reviewer in questioning how the timing of testing/results affects the interpretation of the study. We have updated the discussion with more detailed consideration of the possible effects: “Another limitation to the VOMC cohort data is the time delay to the VOMC intake visit. Our usual care requires a positive SARS-CoV-2 test prior to VOMC enrollment, and delays in the testing process or results notification could attenuate patient recall of initial symptoms. It is also possible that delays would reduce the intake of patients with severe symptoms (as they escalate to admission) as well as mild symptoms (as they resolve). To reduce the effects of testing delays on our study, we limited the study to patients within 10 days of symptom onset and used chart review to verify symptoms reported in the testing process.”	Page 26, line 7-14

Bias section (p10). The authors suggest that the voluntary nature of the enrollment biased the study toward the more symptomatic patients. But without a comparison of those enrolled and those not enrolled, this is not supported. Might it also be possible, that more anxious patients, maybe with only mild symptoms, would be more likely to enroll?	Thank you for the feedback. We have revised this sentence to remove the statement "for more symptomatic patients." Also, we are renaming the "Bias" section "Potential Bias" per reviewer #3 below.	Page 10, line 10
Bias section (p10). The authors describe "screening in emergency room" – it should be clarified if the ED was screening all patients or using the test diagnostically, not as a screening test, as suggested in the Box.	Thank you for clarifying. During the study period, the ED was selectively testing more symptomatic patients and the sentence has been revised. We have revised the paper to reduce use of the term "screening" for symptomatic patients and instead use the term "testing."	Page 10, line 8.
Results. This section should begin with the overall number of patients testing positive for SARS-CoV-2 in the outpatient setting (appears to be 730) and explain how that 730 positive tests led to 340 patients included in the current study. As it's hard to determine how representative the findings are without additional details: for example, how many patients were eligible for VOMC enrollment but declined participation. Furthermore, it would be ideal to compare basic demographic information from the patients not enrolled to those enrolled to help determine how this might have affected the results.	We agree with the reviewer that the representativeness of our sample is affected by the VOMC enrollment process. We do not have data on the reason(s) for patients to decline enrollment in VOMC of the patients with positive results. We added to our manuscript that we do not have these data. Of the positive tests (730 outpatient and 170 ED) we report 551 intake visits completed of whom 337 meet inclusion criteria. We have added to Box 1 that telemedicine and telephone were both permitted for intake visit. Possible reasons for patients not to enroll include: (1) patient preference, (2) unable to reach patients for any reason including telephone issues or hospitalization at any location, (3) result notification team error (not providing VOMC referral), (4) VOMC team error (not	Page 26, line 3

	scheduling patient or providing correct telemedicine and/or telephone enrollment options). Because the screening/testing location did not include clinical evaluation and patients were not required to be established prior to testing, it would be difficult to make comparisons between the VOMC enrolled and not-enrolled groups.	
Limitations: An additional limitation is that the study captures the natural history of COVID-19 patients managed by telemedicine. As there was not screening for COVID-19 at the time, it does not capture milder cases not requiring contacting physicians nor many cases requiring in person outpatient management (or maybe it does, this is not clear from the current manuscript).	Thank you for this observation. All patients receiving results of a positive test for SARS-CoV2 were offered VOMC intake visits, including asymptomatic or mildly symptomatic individuals. We acknowledge that, because of the screening criteria (symptoms), few such individuals were included in our cohort. In-person outpatient care was available and did not alter the VOMC care unless a patient was hospitalized (i.e. a patient could enroll in VOMC monitoring and have in-person visit on the same date if needed or at separate times based on clinical need). The details of in-person care are outside of the primary analysis of this manuscript and we have tried to provide this as context in methods as requested.	No change
Comparison with other studies. The sentence (page 22) “In our experience…convalescent illness” is not clear.	We have rephrased this: “In our experience, many patients present for evaluation of non-resolving symptoms in the weeks that follow acute covid-19” and added a clarifying reference.	Page 24, line 9
Figure 1. Recommend removing the heatmap of 4 patients with severe initial symptoms. This small number of patients is too small to be representative and, visually, gives the misperception that there is as much data about this group as the other two groups.	We agree and have removed this part of the figure (removed figure 1d).	Figure

Figures 2 & 3: it is not entirely clear why this information needs figures.	We agree and have made the requested change to tables to replace figures 2 and 3.	New tables 2 and 3
Reviewer 3		
1. I think the 'Bias' section should be renamed 'Potential biases'	We agree and have made this change.	Page 10 line 5
2. On p18, it looks like the confidence interval for 178% should be 57-392.	Thank you for pointing out this error	All numbers for table 5 corrected in text with revised table. Page 19 line 10-Page 20 line 16
Reviewer 4 – “statistical concerns found as follows”		
1. Page 11, including some more details about how the data has been summarized will help readers understand the tables and figures.	We appreciate the feedback in this and the subsequent comment about the difficulty understanding tables and figures. We have attempted to clarify this with improved labeling of the figures and clarifications within the results section.	Results section

2. Page 17-22, it does not sounds enough to not use the Cox model in analysis because of that the assumption of proportional hazard is not met. Given its advantages, a Cox model with nonproportional hazards should be tried if appropriate.	The primary advantage of survival analysis to include censored data applies to both Cox and accelerated failure time models. I did try including time*covariate interaction terms in the cox model to address non proportional hazards. There is another technique, which is what I think you are referring to, that divides time into smaller segments where the proportional hazards assumption is met within each time segment but we ultimately chose the AFT models to present in the paper because the AIC showed significantly better fit than the Cox models even for the covariates that did not violate the proportional hazards assumption. Supplement table 1 includes the final fitted Cox models with time varying covariates for those who would prefer to see the data presented in a Cox model. For the symptoms without time varying covariates the results are similar to the AFT model results in terms of effect size.	
3. Page 14, Table 1, because of censored data, ANOVA is not applicable in the comparison of days to symptom onset and duration of symptom. If there is not censored data, then what the distributions of the duration look like? ANOVA is most applicable for data from population of normal distribution.	Both are very positively skewed. Duration is censored so ANOVA would not be appropriate but assuming the data were not censored here is a graph if the distribution with normal distribution overlay (done for all symptoms since breaking it down by each symptoms would take too much space but none were normal distribution.)  There is no censored data for start day for symptoms organized by system. Once again this is for all systems with normal distribution overlay. KM curves to determine medial day made the most sense due to the 1) non normal distribution and 2) consistency of analysis technique between figure 2 and 3.	

		
4. For all tables and figures: A footnote should be included if appropriate to make the table/figure self-explanatory, guiding the readers to understand it. For example, in page 3, “Loss of smell or taste had the second longest duration with 14 days (12 to 17), with 38% not resolved at final visit” implies that there are censored data. Then in Table 1 in page 14, how readers should understand the results for “Symptom onset to intake VOMC visit” and “Symptom onset to last phone call”?	For table 3, we are clarifying the nature of censored data in a footnote For table 1: “symptom onset to VOMC visit” and “symptom onset to last phone call” - we have clarified: *Number of days from initial symptom(s) of covid-19 to completion of telemedicine intake visit for the Virtual Outpatient Management Clinic (VOMC), inclusive of time required for testing, result notification, and scheduling with VOMC. †Number of days from initial symptom(s) of covid-19 to the final telephone call with VOMC. The calls would end with patient-reported symptom improvement (not necessary resolution) or hospital admission (32 admitted, of whom nine were followed with additional calls post-discharge).	Page 19, line 2 Page 15, lines 3-8
5. Page 19, line 3, what is baseline group and how it has been defined? Is it the control group?	Thank you for this input. Baseline was the wrong choice of words. We have changed it to reference group and clarified it better in the example and the table.	Page 20 line 7
6. Page 19, Table 3, the symptom column, that each symptom show up once will make the table more clear and easier to read.	This is excellent input and we have made the change.	Page 21

7. Page 3, line 40, “Initial symptom severity is the best predictor of disease duration” sounds better to change to “Initial symptom severity is a significant predictor of disease duration for most considered symptoms” or so.	We agree with the reviewer and have made this change and appreciate their suggested terms.	Abstract

VERSION 3 – REVIEW

REVIEWER	Pieter Cohen Cambridge Health Alliance, US Receive compensation from UptoDate to co-author chapter on outpatient management of COVID-19.
REVIEW RETURNED	08-Feb-2021

GENERAL COMMENTS	The authors research will be very helpful in clarifying the natural history of mild covid in the outpatient setting. The authors have addressed all my prior concerns. My final comment is that the manuscript should be reviewed carefully for accurate and up-to-date terminology regarding race. The authors should confer with the journal editor to determine most appropriate terminology but we are now using Black and White (capitalized), and I would recommend carefully reviewing every place in the manuscript, tables and figures where race is mentioned as there are some outdated terms used.

REVIEWER	Zhiying You CU Anschutz Medical Campus
REVIEW RETURNED	15-Feb-2021

GENERAL COMMENTS	All previous concerns have been addressed