PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Prospective Cohort Study of Children with Suspected SARS-CoV-2 Infection Presenting to Pediatric Emergency Departments: A Pediatric Emergency Research Networks (PERN) Study Protocol
AUTHORS	Funk, Anna; Florin, Todd; Dalziel, Stuart; Mintegui, Santiago; Salvadori, Marina; Tancredi, Daniel; Neuman, Mark; Payne, Daniel C.; Plint, AMy; Klassen, Terry; Malley, Richard; Ambroggio, Lilliam; Kim, Kelly; Kuppermann, Nathan; Freedman, Stephen

VERSION 1 – REVIEW

REVIEWER	Alessio Pini Prato Umberto Bosio Center for Digestive Diseases, The Children Hospital, AO SS Antonio e Biagio e Cesare Arrigo, Alessandria, Italy, EU
REVIEW RETURNED	11-Jul-2020

GENERAL COMMENTS	Rules for patients enrollement as well as those for controls are unclear and refer to continuously changing guidelines. Those guidelines are developed by WHO but at present it seems that their statements need to be re-addressed in each center participating into the study. More consistent rules for enrollment are needed to provide reliable and teproducible results.

REVIEWER	Nitin Dhochak All India Institute of Medical Sciences, New Delhi, India
REVIEW RETURNED	10-Aug-2020

GENERAL COMMENTS	I would like to appreciate the author's efforts to plan and conduct such huge study with focus on assessing data of large population. The authors primarily aim to describe the clinical features of COVID-19 in children and identify predictors to differentiate patients with COVID-19 from other acute respiratory diseases and predictors of severe COVID-19. Comments to the authors are summarized below: 1. Authors aim to compare clinical features and health care resource utilization between confirmed and suspected COVID-19 patients. Most of pediatric COVID-19 samples are nasopharygeal and oropharyngeal swabs. Sensitivity of common PCR assays in respiratory samples ranges up to 30 - 70%. Up to 30%, or even more SARS CoV 2 infected patients will be included as suspected cases. This is a large number and will interfere with ability to identify predictors accurately differentiating COVID-19 from non-COVID-19 infections. More intensive investigations to accurately diagnose and allocate patients to COVID-19 disease and other respiratory diseases.

	2. Reason for comparison between confirmed and suspected patients is not clear. The authors need to make more effort to truly identify COVID-19 such as testing specific immunoglobulin level during convalescent period. This will lead to reduction in false negative cases. 3. Verbal consent: details can be provided about verbal consent if it will be recorded? 4. Authors need to define skill assessment/ educational qualification of study staff collecting data as these are partly subjective.

REVIEWER	Dale Steele Alpert Medical School
REVIEW RETURNED	10-Sep-2020

GENERAL COMMENTS	pg. 20, lines 10-38: Regarding calculation of sample size. Please provide citation(s) for the assumptions and methods used to estimate required sample size. If possible, provide the reproducible code. Were methods used consistent with recent recommendations (citations below): Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. BMJ [Internet] 2020;368:m441. Available from: http://dx.doi.org/10.1136/bmj.m441 Riley RD, Snell KI, Ensor J, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. Stat Med [Internet] 2018;Available from: http://dx.doi.org/10.1002/sim.7992 pg. 21 line 3: "Probable Case - Suspected case with inconclusive SARS-CoV-2 result." How will a suspected case be operationally defined? pg 21 lines 24-31: Regarding the "capacity of the independent variables (e.g. clinical characteristics such as age and sex) to discriminate" Discrimination will vary due to site level factors (season, community prevalence at a given time, northern vs. southern hemisphere). Multilevel models should used that reflect the importance of site-level features.

REVIEWER	Janet Peacock Geisel School of Medicine at Dartmouth USA
REVIEW RETURNED	15-Sep-2020

GENERAL COMMENTS	Thank you for asking me to review this protocol paper. I write this review as a biostatistician/epidemiologist who has worked largely in paediatric studies. This is such an important area and a timely study. I enjoyed reading it and have a few comments and suggestions which I hope will be helpful: The authors seek to collect information on children attending the ED with COVID symptoms and in particular to be able to predict the minority of children who will develop a severe outcome. Severe

	is defined as positive pressure ventilation, or ICU admission with ventilatory or inotropic support or multisystem hyper-inflammatory syndrome or death. They estimate that among the 12500 children to be recruited about 50 will have a severe episode. They give sample size calculations for AUC (c-statistic) of 0.7 and state this has high power. I was able to replicate these calculations. My questions are: 1. Why choose AUC=0.70? This is quite low and lower than the usual recommended minimum for a predictive model, 0.80. (the value 0.50 indicates the predictive model is no better than tossing a coin). The difficulty is that with a low area under the curve, there will be incorrect predictions. Please can the authors discuss this. 2. The sample size of around 50 severe cases is understandably small and the team has said they will only fit a model with up to 10 degrees of freedom, which is appropriate for this number of events. They haven't said what these factors will be or how they will be chosen and/or what statistical strategies will be used. They haven't discussed validating the model which may mean splitting the sample into subsamples, making the sample even smaller. 3. The sampling strategy includes oversampling positive cases to ensure sufficient positives are included. Have the team considered oversampling or better still fully sampling the hospitalizations as these are the real focus and are few in number. This would provide a dataset with greater numbers of severe cases and therefore greater ability to develop an accurate predictive model. 4. I didn't see any mention of potentially missing data – what would be anticipated and how these might impact the analyses and results? 5. A full statistical analysis plan will be needed to ensure this study can deliver on its aims. A short version could be included as an appendix to this protocol. 6. Playing devil's advocate, perhaps a case-control design might serve this question better and be more efficient. I'm guessing this was considered and rejected?

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1

COMMENT 1.1: Rules for patients enrolment as well as those for controls are unclear and refer to continuously changing guidelines. Those guidelines are developed by WHO but at present it seems that their statements need to be re-addressed in each center participating into the study.
More consistent rules for enrolment are needed to provide reliable and reproducible results.

RESPONSE 1.1: Thank you for your review and comment. All participants (both those who are SARS-CoV-2 positive and SARS-CoV-2 negative) are enrolled using the same eligibility criteria, namely that the child needs to have been tested for SARS-CoV-2 due to a suspected infection (e.g. presence of symptoms or other risk factors). However, as you point out, the indications for testing children identified as potentially being infected vary by country, by region, and over time. These

elements are beyond the control of our study and cannot be standardized and thus we decided to take a pragmatic approach for this global study. We have added the following statement to the 'Participants and Recruitment' description of our methods section in order to highlight this:

> "This recruitment strategy will identify children who are both SARS-CoV-2 positive and SARS-CoV-2 negative. The indications for testing children identified as potentially infected will differ by country, region, and over time; however standardization of these protocols is beyond the control of this study."

We agree that our lack of standardized screening criteria across all hospitals could introduce selection bias by site and season. To address this, since study initiation, each site has been recording a weekly log of their hospital-specific screening criteria using a standardized data collection form. This is outlined at the end of the 'data collection' section of the manuscript. We will use this information to understand how screening criteria differ between hospitals, and to generate site classifiers that may be included in statistical models. Furthermore, for our analyses, we will use multivariable conditional regression models, where matched groups are based on hospital of enrolment. In our statistical models, in addition to site, we will include terms for specific symptom complexes, whether or not there was a case-contact, season of recruitment, and other factors that may be associated with screening criteria.

We have highlighted the conditional modelling approaches in the statistical analysis section of our manuscript as follows:

> "Conditional multiple logistic regression models, with matched groups based on hospital of enrolment, will be used to identify a set of independent variables that are able to: 1) discriminate among the two main case-statuses (i.e. SARS-CoV-2 positive and SARS-CoV-2 negative), and 2) predict severe outcomes of COVID-19."

Please note that we have now removed mention of the ordinal logistic regression across three case-statuses (suspected, probable, and confirmed) from the statistical analysis section of our manuscript as we believe, based on our data accrued to date, that the number of probable cases will be very minimal. Thus, we have restricted our primary analyses to include only SARS-CoV-2 negative (formerly called 'suspected') and SARS-CoV-2 positive (formerly called 'confirmed').

We have also included a detailed description of these models in our detailed statistical analysis plan, which has been added as an appendix to this manuscript.

Finally, based on the feedback provided, we have added the following limitation in the 'Ethics and Dissemination' section of the manuscript:

"As case-screening criteria will vary both by site and by time period, a limitation includes the potential for selection bias; our statistical modelling approaches are designed to account for some of this bias. "

Reviewer: 2

I would like to appreciate the author's efforts to plan and conduct such huge study with focus on assessing data of large population. The authors primarily aim to describe the clinical features of COVID-19 in children and identify predictors to differentiate patients with COVID-19 from other acute respiratory diseases and predictors of severe COVID-19. Comments to the authors are summarized below:

COMMENT 2.1: Authors aim to compare clinical features and health care resource utilization between confirmed and suspected COVID-19 patients. Most of paediatric COVID-19 samples are nasopharyngeal and oropharyngeal swabs. Sensitivity of common PCR assays in respiratory samples ranges up to 30 - 70%. Up to 30%, or even more SARS-CoV-2 infected patients will be included as suspected cases. This is a large number and will interfere with ability to identify predictors accurately differentiating COVID-19 from non-COVID-19 infections. More intensive investigations to accurately diagnose and allocate patients to COVID-19 disease and other respiratory diseases.

RESPONSE 2.1:Thank you for this feedback, which we considered carefully. The sensitivity of real-time polymerase chain reaction (RT-PCR) performed on nasopharyngeal and oropharyngeal swabs to detect SARS-CoV-2 varies considerably depending on several factors. This includes the time since exposure (linked to time of symptom onset), as evidenced by a pooled analysis that examined the performance of RT-PCR to detect the virus on 1330 upper respiratory tract samples.(1) This study found that at the time of symptom onset, which was an average of five days following exposure, the false negative rate among those tested for SARS-CoV-2 was 38% (62% sensitivity), further decreasing to 20% (80% sensitivity) three days after symptom onset.(1) In order to understand how test sensitivity might affect our control population of SARS-CoV-2 negative children, we need to relate test sensitivity to the pre-test probability of being infected.(1-3) Although the precise pre-test probability in the general paediatric population at each participating site is unknown, it is reasonable to believe that this is fairly low. For example, a recent study from the United States estimated a seroprevalence of 9.3% in the adult population as of mid-June 2020, representing an *accumulation* of all infections during the first three months of the pandemic.(4) Among children with suspected infections presenting to emergency departments, the pre-test probability would likely be higher than that of the general paediatric population, and would vary based on presenting symptoms (e.g. cough + fever + difficulty breathing versus headache only) and other risk factors (e.g. history of case-contact). With a test sensitivity of 70% and a pre-test probability of 15% (a conservatively high estimate) in children with any symptom potentially suggestive of SARS-CoV-2 presenting to participating emergency departments, the percentage of misclassification of the SARS-CoV-2 negative control group would be only 5%.(2)

We have now added the following statements, summarizing these points, as discussion of limitations in the second paragraph of the 'Ethics & Dissemination' section of the manuscript as follows:

> "There is also the potential for misclassification (e.g false negatives) within the SARS-CoV-2 positive and SARS-CoV-2 negative groups. In order to understand how the sensitivity of real-time polymerase chain reaction (RT-PCR) performed on nasopharyngeal and oropharyngeal swabs might lead to misclassification in our control group of SARS-CoV-2 negative children, we need to also consider the pre-test probability of being infected.[34-36] It is reasonable to believe that the pre-test probability among children tested for SARS-CoV-2 due to suspected infection in EDs would still be relatively low at any one time, due to the highly sensitive screening criteria that are commonly used (e.g. any child with fever). With a test sensitivity of 70% and a pre-test probability of 15% (a conservatively high estimate) in children with any symptom potentially suggestive of SARS-CoV-2 presenting to participating EDs, the percentage of misclassification of the SARS-CoV-2 negative control group would be 5%.[35] These limitations are addressed in our statistical analysis plan, which outlines the use of measurement error models to account for varying levels of misclassification in our study groups"

To mitigate the impact of misclassification in the control group, we have devised various strategies. One strategy employs the use of measurement error models to account for inclusion of both false-positives and false-negatives in our exposed and control groups. This has been outlined in the manuscript in the Data analysis section under 'Statistical analysis':

> "As there is a potential for misclassification with respect to infection status (e.g. false negatives) due to the sensitivity and specificity of the SARS-CoV-2 tests used, sensitivity analyses using measurement error models and other strategies will be conducted.[31,32]"

The following further strategies are outlined in our statistical analysis plan (now attached as an appendix):

- A statistical exploration of the proportion of children with positive SARS-CoV-2 test result in relation to their varying times since first symptom onset. This will include sensitivity analyses that consider the types of symptoms experienced.
- We will conduct sensitivity analyses for predictive models where SARS-CoV-2 positive and negative children have been compared. These sensitivity analyses will exclude from the SARS-CoV-2 negative control group, participants with factors related to increased pre-test probability, including symptoms experienced, length of time since symptom onset, and history of close contact with a known case.

References:

1. Kucirka LM, Lauer SA, Laeyendecker O, Boon D, Lessler J. Variation in False-Negative Rate of Reverse Transcriptase Polymerase Chain Reaction–Based SARS-CoV-2 Tests by Time Since Exposure. *Ann Intern Med.* 2020 May 13: M20-1495. Doi: 10.7326/M20-1495
2. Interpreting a covid-19 test result. *BMJ* 2020;369:m1808. Doi: 10.1136/bmj/m1808
3. Woloshin S, Patel N, Kesselheim AS. False Negative Tests for SARS-CoV-2 Infection — Challenges and Implications. *N Engl J Med* 2020; 383:e38. Doi:10.1056/NEJMp2015897
4. Anand S, Montez-Rath M, Han J, et al., Prevalence of SARS-CoV-2 antibodies in a large nationwide sample of patients on dialysis in the USA: a cross-sectional study. *Lancet* 2020. Published online September 25th, 2020. Doi: 10.1016/ S0140-6736(20)32009-2

COMMENT 2.2: Reason for comparison between confirmed and suspected patients is not clear. The authors need to make more effort to truly identify COVID-19 such as testing specific immunoglobulin level during convalescent period. This will lead to reduction in false negative cases.

RESPONSE 2.2: Our “confirmed” patients are those who test positive for SARS-CoV-2, and our comparator group will be those initially suspected (and tested) but eventually found to be negative for SARS-CoV-2. We believe that our original terminology for this was not clear and we have changed the nomenclature used to describe these groups as either, ‘SARS-CoV-2 positive’ (instead of ‘confirmed’) and ‘SARS-CoV-2 negative’ (instead of ‘suspected’), throughout the manuscript. In adjusted analysis, comparisons of SARS-CoV-2 positive and SARS-CoV-2 negative participants will allow us to determine factors predictive of infection. Through bivariable analysis comparing SARS-CoV-2 positive and SARS-CoV-2 negative children, we will be able to contextualize the risk of various outcomes associated with SARS-CoV-2 with risks of the same outcomes among other children presenting to the ED with similar symptoms but who are determined to be SARS-CoV-2 negative. For example, to understand the risk of hospitalization if a child is infected with SARS-CoV-2 compared to the risk of hospitalization if a child is infected with other respiratory viruses. This approach will allow us to understand whether specific outcomes can be attributed to SARS-CoV-2 infection itself, or rather to other factors affecting all children being screened for SARS-CoV-2 in that time period (e.g. region of study, pandemic-related isolation measures, etc).

Our explanation regarding the potential for false negatives in our study control group, and our statistical approach to address the possibility of some false negatives is described in Response 2.1 above.

COMMENT 2.3:. Verbal consent: details can be provided about verbal consent if it will be recorded?

RESPONSE 2.3: In some cases written consent is obtained immediately upon first contact with participants/caregivers and in other cases, verbal consent is obtained over the phone during first contact, this is recorded in the patient file, and then a detailed consent form is emailed to the caregiver. The use of these strategies depends on the human subjects/ ethics requirements at each site.

We have clarified this in the third paragraph in the ‘Participants and Recruitment’ section of the Methods, as follows:

> “In the instance of verbal consent attained via telephone, a detailed consent form is then emailed to the caregiver. All consent is recorded in the participant’s case report forms.”

COMMENT 2.4: Authors need to define skill assessment/ educational qualification of study staff collecting data as these are partly subjective.

RESPONSE 2.4: The research staff collecting data may include research assistants, research nurses, medical doctors, or other qualified, trained research staff. These individuals all have at least an undergraduate university degree, and have been trained on research procedures. To standardize data collection across all study sites, a detailed data collection manual of operations, which includes explanations and examples for any potentially subjective questions, has been composed and distributed to all sites. Research staff have been trained using this guidance, and are continually encouraged to reach out to the study leadership team with questions about data collection.

We have highlighted this standardized training manual at the end of the first paragraph of the data collection section of the manuscript, which now reads as follows:

> "All research staff will be trained using a standardized and detailed manual of operations that includes explanations and examples for any potentially subjective questions in the study's data collection forms."

Reviewer: 3

COMMENT 3.1: pg. 20, lines 10-38: Regarding calculation of sample size.
Please provide citation(s) for the assumptions and methods used to estimate required sample size. If possible, provide the reproducible code.

RESPONSE 3.1: Our power calculation for determining sample size was performed using the ROCPOWER SAS macro.(1) Power calculations for ROC statistics are developed for a 1-df test. To account for model complexity and our large number of candidate variables, we doubled the sample size. This doubling was based on comparing the ratio of chi-square noncentrality parameters for 10 vs. 1 df (20.53 vs. 10.51, respectively), when power=90% and alpha=5%. This method is based on the same principles that motivate the variance inflation factor for multiple regression and ANCOVA as described by Hsieh and colleagues.(2)

These references are now included in the description of our sample size calculation in the manuscript. We have also added the following detail to the end of the paragraph describing the sample size calculation:

> "Sample size calculations were performed using the SAS ROCPOWER macro.[30]"

References

1. Zepp RC (1995). A SAS® Macro for Estimating Power for ROC Curves One-Sample and Two-Sample Cases. Proceedings of the 20th SAS Users Group International Conference (SUGI), Paper 223.

2. Hsieh FY, Lavori PW, Cohen HJ, Feussner JR. An Overview of Variance Inflation Factors for Sample-Size Calculation. *Evaluation & the Health Professions.* 2003;26(3):239-257. doi:10.1177/0163278703255230

COMMENT 3.2: Were methods used consistent with recent recommendations (citations below):

Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. BMJ [Internet] 2020;368:m441. Available from: http://dx.doi.org/10.1136/bmj.m441

Riley RD, Snell KI, Ensor J, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. Stat Med [Internet] 2018;Available from: http://dx.doi.org/10.1002/sim.7992

RESPONSE 3.2: Our sample size was calculated using a different method than that outlined by the two publications mentioned. The resulting calculation, however, aligns closely with that we would have obtained if using such recommendations. This is highlighted by our use of a variance inflation factor (design effect) of two, which recommends enrolling 2500 SARS-CoV-2 positive children in order to achieve a study power that could have otherwise been achieved after enrolling 1250 SARS-CoV-2 positive children.

COMMENT 3.3: pg. 21 line 3: "Probable Case - Suspected case with inconclusive SARS-CoV-2 result." How will a suspected case be operationally defined?

RESPONSE 3.3: All participants for this study are enrolled using the same inclusion criteria (please see operational definition discussion in response to Reviewer 2, in Response 2.2). All children enrolled in this study need to have been tested for SARS-CoV-2 at a participating emergency department due to a suspicion of acute SARS-CoV-2 infection (e.g. presence of symptoms or other risk factors). However, a 'suspected infection' may be defined differently (with different criteria) at each study site and over time during the pandemic. The pragmatic approach we have taken is representative of clinical care delivery during the pandemic. We discussed this in detail in response to Reviewer 1 (Response 1.1), and we have included details of our use of a conditional logistic regression model in order to account for differences within sites, as follows:

> "Conditional multiple logistic regression models, with matched groups based on hospital of enrolment, will be used to identify a set of independent variables that are able to: 1) discriminate among the two main case-statuses (i.e. SARS-CoV-2 positive and SARS-CoV-2 negative), and 2) predict severe outcomes of COVID-19."

We have included further strategies that will account for potential differences by site, within our statistical analysis plan that is included as an appendix.

Furthermore, based on your comment and that of Reviewer 2 (Comment 2.2) we have realized that our initial terminology for our varying case-status groups was unclear. We have also now removed mention of the ordinal logistic regression across three case-statuses (suspected, probable, and

confirmed) from the statistical analysis section of our manuscript as we believe, based on our data accrued to date, that the number of probable cases will be very minimal. As can be seen in the description of definitions at the beginning of the 'Data Analysis' section of the manuscript; we have changed the mention of the two remaining case-status groups to be as follows:

- SARS-CoV-2 negative - Patient screened (i.e. tested) but with a negative test result for SARS-CoV-2.
- SARS-CoV-2 positive - Patient screened (i.e. tested) with laboratory confirmed SARS-CoV-2 infection.

Throughout the manuscript we now use this terminology (i.e. SARS-CoV-2 negative, SARS-CoV-2 positive).

COMMENT 3.4: pg 21 lines 24-31: Regarding the "capacity of the independent variables (e.g. clinical characteristics such as age and sex) to discriminate". Discrimination will vary due to site level factors (season, community prevalence at a given time, northern vs. southern hemisphere). Multilevel models should used that reflect the importance of site-level features.

RESPONSE 3.4: Thank you for this important comment. We plan on using conditional multivariable logistic regression models, thereby accounting for differences between sites. These models will include terms that account for the period of sampling (season), and known regional differences in community prevalence.

This is now outlined in the statistical analysis section of the manuscript as follows:

> "Conditional multiple logistic regression models, with matched groups based on hospital of enrolment, will be used to identify a set of independent variables that are able to: 1) discriminate among the two main case-statuses (i.e. SARS-CoV-2 positive and SARS-CoV-2 negative), and 2) predict severe outcomes of COVID-19."

Please note that we have removed mention of the ordinal logistic regression across three case-statuses (suspected, probable, and confirmed) from the statistical analysis section of our manuscript as we believe that the number of probable cases will be very minimal. We have restricted our primary analyses to include only SARS-CoV-2 negative (formerly called 'suspected' here) and SARS-CoV-2 positive (formerly called 'confirmed' here).

We have also added a more detailed statistical analysis plan, which outlines our analytic strategy as, an appendix to this manuscript.

Reviewer: 4

Thank you for asking me to review this protocol paper. I write this review as a biostatistician/epidemiologist who has worked largely in paediatric studies. This is such an important area and a timely study. I enjoyed reading it and have a few comments and suggestions which I hope will be helpful:

The authors seek to collect information on children attending the ED with COVID symptoms and in particular to be able to predict the minority of children who will develop a severe outcome. Severe is defined as positive pressure ventilation, or ICU admission with ventilatory or inotropic support or multisystem hyper-inflammatory syndrome or death. They estimate that among the 12500 children to be recruited about 50 will have a severe episode. They give sample size calculations for AUC (c-statistic) of 0.7 and state this has high power. I was able to replicate these calculations.

My questions are:

COMMENT 4.1: Why choose AUC=0.70? This is quite low and lower than the usual recommended minimum for a predictive model, 0.80. (the value 0.50 indicates the predictive model is no better than tossing a coin). The difficulty is that with a low area under the curve, there will be incorrect predictions. Please can the authors discuss this.

RESPONSE 4.1: Thank you for the insightful question. We chose an AUC of 0.70 because, in the context of beginning this study early in the pandemic, we hoped to allow for detection of even minimally significant risk factors for severe infections in children. We agree that an AUC of 0.70 is relatively low and note that, as a result, our sample size requirement is higher and will result in very high power to detect if the true AUC is even larger. For example, we will have a power of >99.0% if we assume the true c-statistic is 0.75 or higher.

COMMENT 4.1: The sample size of around 50 severe cases is understandably small and the team has said they will only fit a model with up to 10 degrees of freedom, which is appropriate for this number of events. They haven't said what these factors will be or how they will be chosen and/or what statistical strategies will be used. They haven't discussed validating the model which may mean splitting the sample into subsamples, making the sample even smaller.

RESPONSE 4.2: We performed our sample size calculation based on the most rare outcome that we expected (i.e. severe outcomes in children), and this calculation conservatively included a variance inflation factor of 2. If our study does indeed observe only 50 or fewer severe cases, then we will indeed be limited in terms of splitting this sample into subgroups for sensitivity analyses. If there is eventual interest in a model predicting severe outcomes in a subset of children (e.g. only children with chronic conditions) then our model will be very restricted. However, for our other objectives, such as developing a predictive model for infection risk factors, we do expect to have a high power even after splitting our study population of 2500 SARS-CoV-2 children into subgroups.

We have added the following details to the statistical analysis section of the manuscript:

> “For each of these two types of models, subgroup analyses using only a subset of the population of interest (e.g. children younger than 90 days of age, children with chronic illnesses, etc.) may be conducted. ... The selection of variables for inclusion in the models will rely heavily on expert judgment among study investigators and will be supplemented by literature review. Strategies such as elastic-net regression will be used in order to improve the external generalizability of the model.”

We provide further details on how we will select for variables included in the models, as details about planned subgroup and sensitivity analyses in the statistical analysis plan that has been added as an appendix to this manuscript.

COMMENT 4.3: The sampling strategy includes oversampling positive cases to ensure sufficient positives are included. Have the team considered oversampling or better still fully sampling the hospitalizations as these are the real focus and are few in number. This would provide a dataset with greater numbers of severe cases and therefore greater ability to develop an accurate predictive model.

RESPONSE 4.3: While one of our primary objectives is to understand risk factors for severe infection, and although we have based our sample size calculation on this rare outcome, we have many other important objectives that do not focus on severe infection. Importantly, because of our recruitment strategy and longitudinal cohort follow-up study design, we will be able to determine the risk of children developing severe infection and the risk of persistent symptoms following infection. As many children present to the emergency department with mild symptoms, our findings on the risk of progressing to these severe outcomes and development of long-term symptoms may then be more generalizable to the paediatric population. This recruitment strategy also allows us to develop models for risk factors of infection with SARS-CoV-2, while taking into account that this is not a community cohort but instead a cohort of children presenting to emergency departments with symptoms or other risk factors of infection.

We have added the following description of 90-day outcomes to the manuscript at the beginning of the ‘Data Analysis’ section:

“

- 90-day outcomes / persistent symptoms – a child will be considered as having a persistent symptom if the parents have indicated, at the 90-day follow-up, that respiratory, psychosocial, or ‘other’ symptoms that began in the immediate time period surrounding the ED visit are persisting until the present day.”

COMMENT 4.4: I didn’t see any mention of potentially missing data – what would be anticipated and how these might impact the analyses and results?

RESPONSE 4.4: We apologise that these important details were missing from our initial version of the manuscript. We have now included the following details in the statistical analysis section of the manuscript:

> "Our primary analyses will be complete case analyses, as we anticipate that key outcome and predictive variables will be available for a large fraction of the study population. However, sensitivity analyses evaluating the impacts of non-ignorable missingness on the soundness of our complete case inferences will be conducted and reported according to principled approaches to missing data.[33]"

This is further expanded upon in the section on missing data included in the statistical analysis plan appendix in the 'Treatment of Missing Data' section, with the following details:

> "However, the validity of the complete case analysis depends on the validity of the missing-at-random assumption. To evaluate the impact of non-ignorable missingness on the soundness of our complete case inferences, we will conduct sensitivity analyses using multiple imputation but alter imputation models to reflect varying degrees of non-missingness (e.g. assuming a higher likelihood of a severe outcome than was predicted in the complete case model)."

COMMENT 4.5: A full statistical analysis plan will be needed to ensure this study can deliver on its aims. A short version could be included as an appendix to this protocol.

RESPONSE 4.5: Thank you for this suggestion. We have now included a statistical analysis plan as an appendix to this manuscript.

COMMENT 4.6: Playing devil's advocate, perhaps a case-control design might serve this question better and be more efficient. I'm guessing this was considered and rejected?

RESPONSE 4.6: As mentioned in more detail to your comment above (COMMENT 4.3), only one or our objectives is to examine risk factors for severe infection. Many of our other important objectives rely on the longitudinal follow-up of children with SARS-CoV-2 starting from a relatively early stage of infection. Therefore, we did not consider a case-control design for this study. We agree that this would have been much more efficient had the study been focused only on characterizing, and determining risk factors for, severe infection.

VERSION 2 – REVIEW

REVIEWER	Nitin Dhochak All India Institute of Medical Sciences, New Delhi, India
REVIEW RETURNED	18-Nov-2020

GENERAL COMMENTS	The revision has clarified the initial concerns.

REVIEWER	Dale Steele Alpert Medical School of Brown University USA

REVIEW RETURNED	25-Nov-2020

GENERAL COMMENTS	Previous reviewer comments have been adequately addressed.

REVIEWER	Janet L Peacock Dartmouth College Geisel School of Medicine, USA
REVIEW RETURNED	20-Nov-2020

GENERAL COMMENTS	Thank you for your clear replies and actions in response to my comments and to the other comments from reviewers. I am satisfied that all my comments are addressed and/or clarified. I wish you well with this important study!