Statistical Analysis Plan (SAP)

Prospective Cohort Study of Children with Suspected SARS-CoV-2 Infection Presenting to Pediatric Emergency Departments: A Pediatric Emergency Research Networks (PERN) Study

Clinicaltrials.gov registration #: NCT04330261

SAP Version 1.0

October 31st 2020

Table of Contents

Background **4**

Objectives and hypotheses **5**

Study Methods **6**

Study design **6**

Power considerations **6**

Statistical repetitive analyses **8**

Timing of final analysis **8**

Timing of outcome assessments and timing of loss-to-follow-up **8**

Study Population **9**

Screening, eligibility, recruitment **9**

Potential confounding covariates **10**

Analysis **11**

Outcome definitions **11**

Analysis methods **12**

Statistical analysis methods related to the first primary objective of the study 12

Statistical analysis methods related to the second primary objective of the study 16

Statistical analysis methods related to the first secondary objective of the study 17

Statistical analysis methods related to the second secondary objective of the study 17

Choice of variables to be included in predictive models 18

Treatment of missing data 18

Other analyses and considerations 19

Statistical software 20

Reporting 20

References 20

Background

Relatively limited data are available regarding pediatric COVID-19. Although most children appear to have mild or asymptomatic infections, infants and those with comorbidities may be at increased risk of experiencing more severe illness and requiring hospitalization due to COVID-19. The recent but uncommon association of SARS-CoV-2 infection with development of a multisystem inflammatory syndrome has heightened the importance of understanding pediatric SARS-CoV-2 infection.

Though reports related to pediatric hospitalization and mortality exist, there are few large-scale, multi-national descriptions of the clinical features and disease course of children with COVID-19. Thus, a detailed examination of the risk factors for infection, clinical characteristics of infected children, predictors of severe outcomes, and treatments provided, is urgently needed for this unique and potentially vulnerable population. As the symptoms of SARS-CoV-2 infection are a common reason for pediatric presentation to hospitals, the early identification of high-risk children and clinical phenotypes are vital to optimizing care. Moreover, as viral shedding may occur in minimally symptomatic children, an in-depth study of pediatric infection characteristics, both mild and severe, will contribute to a better understanding of transmission risks. It will also be important to characterize and compare COVID-19 disease with other common pediatric illnesses to contextualize the severity of illness.

Objectives and hypotheses

The primary aim of this study is to fully evaluate the clinical characteristics and outcomes of SARS-CoV-2 positive and SARS-CoV-2 negative children who were tested in the emergency department (ED) due to suspected infection.

Specific primary objectives and hypotheses:

(1) Describe and compare the characteristics of SARS-CoV-2 positive children (i.e. nucleic acid test-positive) and SARS-CoV-2 negative children (i.e. nucleic acid test-negative).

→ We hypothesize that the characteristics of children positive and negative for SARS-CoV-2 will differ.

(2) Identify factors associated with severe outcomes in SARS-CoV-2 positive children (i.e. intensive care unit admission, assisted ventilation, vasoactive medication use, MIS-C, death).

→ We hypothesize that various factors will be associated with severe outcomes of SARS-CoV-2 infection in children, including but not limited to the age of the participant and the presence of underlying medical conditions.

Specific secondary objectives and hypotheses:

(1) Describe and compare health care resource utilization for patient management (e.g. isolation, testing, imaging, supportive care) of both SARS-CoV-2-positive

and SARS-CoV-2 negative children according to changes in national and regional policies.

→ We hypothesize that health care resource utilization for patient management of SARS-CoV-2 positive and negative children will differ based on changes in national and regional policies.

(2) Describe the effects of changing case screening policies for the detection of SARS-CoV-2 in children.

→ We hypothesize that case screening policies that require either a known close-contact with a suspected case or symptoms specific to upper respiratory traction infections, will lead to decreased overall screening numbers but increased positivity rates, by site.

Study Methods

Study design

This is a cohort study of children tested for SARS-CoV-2 at 47 EDs in 13 countries. This cohort study includes a final 90-day follow-up survey.

Power considerations

At the time of study development, there were many unknown factors related to the epidemiology of COVID-19 that precluded a robust sample size estimate. Our sampling strategy aims to enable the recruitment of more positive than negative cases relative to the prevalence in the general population, creating a cohort whereby 20 – 40% of enrolled

participants are confirmed to be SARS-CoV-2 positive. According to preliminary data from the United States, the proportion of positive children with severe outcomes may be as high as 2%.[1] Among children hospitalized or with a high likelihood of requiring hospitalization who were tested for SARS-CoV-2 in Spain, approximately 10% needed intensive care.[2] In calculating a sample size, we estimated that 2% of SARS-CoV-2 positive children will experience severe outcomes. This is a conservative estimate as the power of our study for assessing discriminative performance depends on the number of the least-frequent outcome level (i.e. severe vs. non-severe outcomes in SARS-CoV-2 positive children). For the severe outcomes, the predictive model will be limited to approximately 10 degrees of freedom. Recruiting 12,500 SARS-CoV-2 screened participants (~250 screened children, on average/site) to identify 50 severe outcomes in SARS-CoV-2 positive children using the most conservative assumptions (12,500 x 20% x 2%) we will have 93.9% power to detect when the predictive model discriminating severe from non-severe outcomes among this sample truly (i.e. in the larger population) has a c-statistic of 0.70. These calculations used a variance inflation factor of 2 to account for model complexity (as measured by degrees of freedom).[3] Power increases with higher assumed values for the c-statistic and with a higher number of assumed cases, therefore attaining near certainty (>99.9%) to detect c=0.70 for the infection risk model of SARS-CoV-2 positive vs SARS-CoV-2 negative children (the other primary objective of the study). We will analyze data and recruitment totals, alongside updated information from other studies on the behavior of COVID-19, weekly, in order to determine the optimal timing and method of data dissemination. Sample size calculations were performed using the SAS ROCPOWER macro,[4] as described by Obuchowski.[5]

Statistical repetitive analyses

This study aims to provide frequent (up to weekly) basic descriptive analyses regarding the outcomes of children with SARS-CoV-2; the primary reason for these analyses will be to monitor data quality and workflow. We aim to perform interim analyses (at an approximate mid-point of the study in terms of participant recruitment) for all primary aims and objectives, under the condition that the sample power at that stage would allow for meaningful conclusions for that specific question – no decisions for continuing or stopping the study will be based on these analysis.

Timing of final analysis

We aim to conduct the final analyses for this study within five months following recruitment of the last enrolled participant. This five-month time period will allow for completion of the 90-day follow-up encounter for the final participant as well as an additional 30 days for final data entry and resolving data queries.

Timing of outcome assessments and timing of loss-to-follow-up

There are two key time points for assessment of outcomes, corresponding with timing to define loss-to-follow-up, among children screened for SARS-CoV-2:

- **Four weeks following the initial ED visit**, there will be an assessment of whether or not the child had a severe outcome following SARS-CoV-2 positivity.

This outcome will be based on data collected between the participant's recruitment and that time point. At that time, participants will be considered lost to follow-up if: five successive attempts to contact the child/caregiver at the time-point two weeks following the ED visit were unsuccessful *and* a medical record review at a time point four weeks following the ED visit was not possible (e.g. unable to link child's healthcare number). In the case that a particular site or region may have a high amount of missing data for severe outcomes when relying on the medical record review (e.g. site with medical records not harmonized across a region, making it unknown if a child had a severe outcome at a different hospital) then loss to follow-up will be defined based only on the five unsuccessful attempts to contact the child/caregiver two-weeks following enrolment.

- **90 days following the ED visit**, there will be an assessment of persistent symptoms or longer-term outcomes of participating children. At this time point, a participant will be considered lost to follow-up if: five successive attempts to contact the child/caregiver between 90 and 120 days since emergency department enrolment are unsuccessful.

Study Population

Screening, eligibility, recruitment

We intend to record the number of children screened per week, overall by site, on a standardized form that collects information on screening and isolation measures in place at that hospital and in the surrounding region. These data reflect the number of children

eligible for study inclusion at the site each week (as eligibility includes screening for SARS-CoV-2 based on suspicion of infection). The number of caregivers of children contacted for enrolment will be recorded on a confidential form that will be kept locally. We will record the number of potentially eligible participants refusing participation or who are unable to be contacted on the form. The number of eligible children contacted for recruitment, as well as the eventual number of those recruited, will be reported in subsequent manuscripts through use of flow-diagrams.

Potential confounding covariates

At present, the below list of covariates are deemed to potentially confound the results of analyses regarding the study's primary objectives to investigate risk factors for infection (comparing SARS-CoV-2 positive and SARS-CoV-2 negative children) and risk factors for severe infection (amongst SARS-CoV-2 positive children). *Although these variables have been identified as possible confounders, they may also be factors predictive of SARS-CoV-2 positivity and of severe outcomes of COVID-19.* This list may be expanded as international understanding of the natural history of SARS-CoV-2 develops.

- Site (hospital) of enrolment
- Season and/or time period of enrolment as it relates to stages of the pandemic
- Participant age
- Participant biological sex
- Presence of underlying medical conditions
- Participant ethnicity / race

Analysis

Outcome definitions

The following definitions will be used during data analysis:

- **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.
- **Severe outcomes** (within four weeks of enrolment)- Positive pressure ventilation (invasive or noninvasive) OR intensive care unit admission with ventilatory or inotropic support or > 48 hour hospitalization OR renal replacement therapy OR multi-system inflammatory syndrome OR death; other outcomes may be added as the understanding of the epidemic evolves.
- **90 day outcomes / persistent symptoms** – a child will be considered as having a persistent symptom if the parents have indicated, at 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. Both respiratory and psychosocial persistent symptoms are collected as a Yes/No outcome, along with free text to further describe the symptoms experienced. The presence of 'other' symptoms is collected only as free text; which of these symptoms (e.g. chronic fatigue) are considered as among those possibly linked to COVID-19 will be

determined based on expert consensus and through scientific literature published at the time of analysis.

Analysis methods

In all cases, effect size estimation with 95% confidence intervals will be prioritized. Statistical significance will be considered at the traditional level of $p<0.05$ for comparative analyses. In cases where multiple testing is a concern, we will use methods to control the false discovery rate at 10% within the family of related inferences.[6]

Statistical analysis methods related to the first primary objective of the study

The first primary objective of this study is to describe and compare the characteristics of SARS-CoV-2 positive children (i.e. nucleic acid test-positive) and SARS-CoV-2 negative children (i.e. nucleic acid test-negative). This will be accomplished through three types of analyses:

1) **Descriptive analysis characterizing SARS-CoV-2 positive children:** We will describe in detail the follow-up of the cohort of SARS-CoV-2 positive children, from the time of their ED visits until 90 days later. Descriptive statistics will be used to summarize participant characteristics and the risk of various outcomes at each follow-up time-point. The descriptive statistics used to summarize participant characteristics (demographic, epidemiological, laboratory, clinical) and outcomes will include proportions (with 95% confidence intervals), means (with standard deviations), or medians (with interquartile ranges), as appropriate depending on the variable type and

value distribution. Loss-to-follow-up occurring between each follow-up time point will be reported.

2) Unadjusted (bivariable) analysis of the entire cohort of children tested for SARS-CoV-2

Two types of unadjusted bivariable analyses will be done:

- Among SARS-CoV-2 positive children, stratified outcome risks will be estimated based on variables including the age of participants, the country of enrolment, and the specific site of enrolment (when possible based on the number of children enrolled at the site). Stratified risk estimates of severe outcomes will be presented along with 95% confidence intervals, and will be compared to each other using the Chi-square or Fisher's exact test (when the number of events is less than 10 for any stratum). Patient characteristics, outside of the stratified variable, will be summarized and compared using one-way analysis of variance, Kruskal-Wallis, or Pearson Chi-square tests, as appropriate.
- The stratified outcome risk estimates in SARS-CoV-2 positive children derived through the above analyses will be compared to these risks in selected subsets of the SARS-CoV-2 negative controls. The subsets of participants selected for comparison may include but will not be limited to: 1) SARS-CoV-2 negative children with specific symptom complexes (e.g. symptoms of an upper respiratory tract infection such as rhinorrhea and cough) 2) SARS-CoV-2 negative children, regardless of presenting symptoms or diagnosis. Stratified risk estimates for SARS-CoV-2 positive and SARS-CoV-2 negative participants will

be presented along with 95% confidence intervals, and will be compared using the Pearson Chi-square, Fisher's exact (when the number of events is less than 10 for either group), one-way analysis of variance, or Kruskal-Wallis tests, as appropriate. Patient characteristics, outside of the stratified variable, will be summarized and compared using one-way analysis of variance, Kruskal-Wallis, or Pearson Chi-square tests, as appropriate.

3) **Adjusted analyses:** Multiple logistic regression models will be used to identify a set of independent variables (e.g. age, biological sex, presence of an underlying condition, ethnicity/race, etc.) able to discriminate between the two main case-statuses (SARS-CoV-2 positive vs SARS-CoV-2 negative). From site-to-site, there will be variability in the sampling fractions (SARS-CoV-2 positive, SARS-CoV-2 negative), hence it will be important to control for these effects when fitting the multiple logistic regression models. We anticipate that our primary modeling strategy will be conditional logistic regression where the matched groups are based on the ED site. As a backup strategy in case there are computational or theoretical difficulties with use of conditional logistic regression models with ED sites considered as the matched sets, we will either use alternative matched sets (e.g. grouping of sites by region and sampling fraction), or unconditional logistic regression models that include terms for site effects. These models may need to use survey weights to address concerns with sampling fractions. The overall discriminative capacity of the model will be estimated along with the 95% confidence interval by first estimating Somers' D, and the D estimate will be transformed to a c-statistic using the formula

c=0.5*(D+1).[7] We will also estimate out-of-sample performance using 10-fold cross validation. Treatment of missing data and selection of variables to include in models are discussed in later sections of this statistical analysis plan.

Subgroup and sensitivity analyses:

- There is interest in models for specific subsets of the pediatric population, such as: children younger than 90 days of age, school-aged children, and children with chronic illnesses. Two approaches may be used to develop these models. One would be development of a model anew as described above for the larger population. Alternatively, in the case that a model is already available for a larger population, one tailored to the subpopulation (e.g. removing unnecessary variables and refitting coefficients).
- As the sensitivity and specificity of SARS-CoV-2 testing are not 100%, misclassification of children with respect to infection status (e.g. false negatives and positives) could potentially result in bias. To explore the potential impacts of misclassification on our inferences, we will conduct and report various sensitivity analyses under a variety of assumptions regarding the extent of misclassification. We will compare the resulting inferences to our primary inferences in order to allow a judgment of the extent of this misclassification. In some sensitivity analyses, we will use measurement error models for misclassified response variables and assume relevant values for the misclassification parameters.[8 9] Another type of sensitivity analysis will exclude participants with factors related to increased pre-test probability from the SARS-CoV-2 negative control group.

These, include symptoms experienced, lengthy time since symptom onset, and history of close contact with a known case.

Statistical analysis methods related to the second primary objective

The second primary objective is to identify factors associated with severe outcomes in children positive with SARS-CoV-2. This will be done through use of similar multiple logistic regression models, as described above. The salient concerns regarding the variation from site-to-site in terms of sampling fractions is only relevant to the extent that there may be differences among the SARS-CoV-2 positive participants with respect to severity. However, outside of a few sites in very high transmission regions (e.g. New York City at the beginning of the pandemic), we believe most sites will enroll a high fraction of all SARS-CoV-2 positive children that they screen. Therefore, we will not be restricted to conditional logistic regression and other methods reserved for highly stratified data. Subgroup analyses will be considered for outcomes with sufficient sample sizes. These may be performed for specific subsets of the study population that are of particular interest or vulnerability (e.g. children of particular ethnicity/race, children in specific age categories, children with chronic conditions). However, as we expect very few severe outcomes of SARS-CoV-2 infection overall, the modeling strategies for these subgroup analyses will need to be simplified. Treatment of missing data and selection of variables to include in models are discussed in later sections of this statistical analysis plan.

Statistical analysis methods related to the first secondary objective

The first secondary objective is to describe and compare health care resource use for patient management (e.g. isolation, testing, imaging, supportive care) of both SARS-CoV-2 positive and SARS-CoV-2 negative children according to changes in national and regional policies. The analysis will use methods as described for the first primary objective for binary outcomes. For count and continuous outcomes, analogous generalized linear models will be used, as appropriate, with choices for link and distribution functions determined by outcome variable distributions. Treatment of missing data and selection of variables to include in models are discussed in later sections of this statistical analysis plan.

Statistical analysis methods related to the second secondary objective of the study

The second secondary objective is to describe the effects of changing case screening policies for the detection of confirmed SARS-CoV-2 positive children. Sites complete a standardized form that collects information on the number of children screened at that ED along with the number of children found positive for SARS-CoV-2 (regardless of whether or not these children were enrolled in the study) on a weekly basis. Furthermore, sites indicate the case-screening criteria used each week at that ED. The proportion of children that test positive for SARS-CoV-2 each week at each site will be estimated along with 95% confidence intervals, and this will be summarized and compared using descriptive statistics (Pearson Chi-square) across categories of case-screening criteria. The case-screening criteria used and the proportion of children positive for SARS-CoV-2 at each site will be summarized graphically by region and time-period. We will analyze

these data using time-series methods where the unit of analysis is the weekly information pertaining to each site. Heterogeneity in test-positivity rate will be assessed using a Poisson regression model for over-dispersed autocorrelated data. Time-varying independent variables will be used to operationalize specific components of the screening policy, e.g. fever present in child, close-contact known, etc.

Choice of variables to be included in predictive models

The selection of variables for inclusion in models will rely on expert judgment supplemented by a literature review. Candidates for inclusion in the model will be evaluated based on the following criteria: presumed major confounders, variables considered potentially predictive of the dependent variable (SARS-CoV-2 positivity or severe outcomes) based on PERN-COVID-19 study investigator consensus, factors already found significantly predictive of the dependent variable (SARS-CoV-2 positivity or severe outcomes) in any age group through already published studies. We will consider using elastic-net regression in order to improve the external generalizability of the model. We anticipate that this will result in a parsimonious model.

Treatment of missing data

The proportion of missing data for each variable of interest (characteristics and outcomes) will be reported, as will be the number of participants lost to follow-up at each follow-up time point. Outcome data that would be obtained through follow-up phone calls at day 14 and day 90, and through the medical record review performed at day 30, are likely not to be missing at random, although this is an untestable assumption.

Similarly, some data may be more prone to non-ignorable missingness (e.g. person-reported infection prevention measures, provider-ordered laboratory data) whereas other data relating to participant demographic characteristics (e.g. age), and some clinical characteristics (such as symptoms experienced) may be deemed missing at random (MAR). Our primary analysis will be a complete case analysis, as we anticipate that key outcome and predictive variables will be available for a large fraction of the study population. 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). Sensitivity analyses will be conducted and reported according to principled approaches to missing data.[10]

Other analyses and considerations

To mitigate the impact of potential misclassification (i.e. false negatives) within the SARS-CoV-2 negative control group, we will further perform a statistical exploration of the proportion of children with positive SARS-CoV-2 test results in relation to time since first symptom onset. This will include sensitivity analyses that consider the types of symptoms experienced.

Statistical software

The statistical software that will be used to carry out the analyses will include the following: SAS 9.4 (SAS Institute, Cary NC), SPSS 25.0 (IBM Corp), Stata 16 (StataCorp LP 2019), R statistical software (R Foundation for Statistical Computing; https://www.R-project.org/).

Reporting

Reporting for models will adhere to the TRIPOD checklist.[11]

References

1. Team CC-R. Coronavirus Disease 2019 in Children - United States, February 12-April 2, 2020. *MMWR Morb Mortal Wkly Rep* 2020;69(14):422-26. doi: 10.15585/mmwr.mm6914e4

2. Tagarro A, Epalza C, Santos M, et al. Screening and Severity of Coronavirus Disease 2019 (COVID-19) in Children in Madrid, Spain. *JAMA Pediatr* 2020 doi: 10.1001/jamapediatrics.2020.1346 [published Online First: Apr 8]

3. Hsieh FY LP, Cohen HJ, Feussner JR. . An Overview of Variance Inflation Factors for Sample-Size Calculation. . *Evaluation & the Health Professions* 2003;26(3):239-57. doi: doi:10.1177/0163278703255230

4. RC Z. 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),* 1995:Paper 223.

5. Obuchowski N. Computing Sample Size for Receiver Operating Characteristic Studies. . *Investigative Radiology* 1994;29:238-43. doi: doi:10.1177/096228029800700405

6. B E. Size, power and false discovery rates. *Ann Statist* 2007;35(4):1351-77.

7. R N. Confidence Intervals for Rank Statistics: Somers' D and Extensions. . *The Stata Journal* 2006;6(3):309-34.

8. Küchenhoff H MS, Lesaffre E. A general method for dealing with misclassification in regression: the misclassification SIMEX. *Biometrics* 2006;62(1):85-96.

9. JP B. Measurement Error: Models, Methods, and Applications (Chapman & Hall/CRC Interdisciplinary Statistics). 1 ed: CRC Press 2010.

10. Perkins NJ CS, Harel O, Tchetgen Tchetgen EJ, Sun B, Mitchell EM, Schisterman EF. . Principled approaches to missing data in epidemiological studies. *American Journal of Epidemiology* 2018;187(3):568-75.

11. Moons KG AD, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): Explanation and Elaboration. *Ann Intern Med* 2015;162(1):W1-W73.