**PROTOCOL TITLE:** Participation in the Repository of Aggregated Pediatric International Data on COVID-19 (RAPID-19)
**RESEARCH SITE:** [.....]
**PRINCIPAL INVESTIGATOR:** [.....]

## A. Summary and Specific Aims

We are proposing to participate in an international clinical data repository on pediatric Coronavirus Disease 2019 (COVID-19) in order to support research studies aiming to characterize the epidemiology, clinical presentation, and natural history of children with COVID-19. De-identified patient-level data from patients treated at our institution and contributed from other organizations will be hosted on a HIPAA-compliant Amazon Web Service infrastructure managed by Boston Children's Hospital (BCH). This environment allows for embedded analytic tools and applications and permits controlled access and functionality. Access will be limited and data security ensured by establishing a Virtual Private Cloud (VPC), which makes the services hosted in Amazon function as if the environment were located within BCH's own network, thus establishing the same protections that apply to BCH clinical data.

The integrated dataset will be made available to our research team and other approved collaborators who can query, analyze, and export de-identified patient-level data. Data use agreements will be in place to ensure investigators take appropriate measures to maintain data security and patient confidentiality, and that they engage in collaborative and scientifically sound studies to advance our knowledge on COVID-19 in children.

## B. Background and Significance

Multiple pediatric registries, observational studies, and clinical trials have been launched to study COVID-19 in pediatric patients. In order to maximize efforts and augment collective resources, workflows are needed to facilitate aggregation and standardization of de-identified clinical data on pediatric patients with lab-confirmed COVID-19 across healthcare sites and study networks.

The prevalence of COVID-19 in children appears to be substantially lower than in adults and the disease course less severe.[1] This will necessitate collaborative efforts across sites to build a robust resource to fully define the natural history and identify risk factors for severe outcomes in children. This is particularly urgent as many clinical trials have been excluding participation by children[2]—as of mid-July, among 1480 interventional trials studying COVID-19 and registered on the trial registry ClinicalTrials.gov, only 141 (10%) were open to any patients less than 18 years.

RAPID-19 is intended to support a broad range of research topics and questions, including but not limited to:

- clinical characterization of pediatric patients with COVID-19
- descriptive analysis of clinical care and treatments administered
- evaluation of laboratory tests for diagnosis and disease monitoring
- identification of patient sub-groups at higher risk for severe disease manifestations
- ascertainment of country-specific variations

## C. Preliminary Studies

The approach is modeled on the principles underlying the *Phelan-McDermid Syndrome Data Network (PMS_DN)* and *The Consortium for Clinical Characterization of COVID-19 by EHR (4CE)*. PMS_DN is a

platform that aggregates patient clinical data, including electronic health record data, curated test reports, and patient-reported outcomes.[3] To ensure data security and scalability, the repository is hosted on a HIPAA-compliant cloud-based environment hosted by Amazon Web Services. Access to the data is granted to authorized users at two distinct levels: aggregate results across the integrated datasets and de-identified patient-level data. The platform exemplifies the ability to advance disease-specific knowledge through data science and sharing among academic researchers and outside collaborators.

The 4CE consortium leverages electronic medical record data from 96 hospitals across 5 countries (covidclinical.net) to address critical clinical and epidemiological questions about COVID-19. In the first demonstration project, harmonized data were analyzed locally and converted to a shared aggregate form for rapid analysis and visualization of regional differences and global commonalities. By the end of April, data covered 27,584 COVID-19 cases with 187,802 laboratory tests. Case counts and laboratory trajectories were concordant with existing surveillance data, demonstrating the ability of the platform to efficiently extract and aggregate valuable clinical data to address clinical and epidemiological questions about COVID-19.[4]

The proposed RAPID-19 repository will build off of these principles and technical advances, focusing on aggregation of existing data, shared data access and analysis, and stringent data confidentiality and security.

## D. Design and Methods

1. **Study Design:** This study is an observational study consisting of collection of de-identified, patient-level data on pediatric patients diagnosed with COVID-19.

2. **Patient Selection and Inclusion/Exclusion Criteria**
   Clinical data on pediatric patients (0 to 21 years of age) with lab-confirmed COVID-19 will be included in the repository. Data will be included from patients at our institution as well as from other participating organizations.

3. **Description of Study Treatments or Exposures/Predictors**
   All included participants will have COVID-19.

4. **Description of Primary and Secondary Outcomes/Endpoints**
   This study does not include any prespecified endpoints. Observational data will be aggregated on pediatric patients with COVID-19 in order to support analyses characterizing the epidemiology clinical course, and outcomes in this patient population.

5. **Data Collection Methods, Assessments, Interventions and Schedule (what assessments performed, how often)**
   Data on eligible patients will be collected from the electronic medical record and other clinical data sources. The clinical data sources may be accessed once or multiple times, as needed, in order to ascertain final disease outcome. For example, if a patient has a positive test and is sent home without additional follow up, data will be collected only on disease presentation at the initial visit. However, if a patient is admitted to the hospital, data will be collected on disease progression through final outcome.

6. **Study Timeline**

The initial repository will be maintained for 5 years. This timeline may be modified as additional knowledge is gained on the time course of the pandemic.

## E. Adverse Event Criteria and Reporting Procedures

Since this study does not include interventions or any patient interactions, the study involves minimal risks for adverse events and no adverse events are anticipated from participation in the study. In the event that study staff learn of an unanticipated problem, reporting procedures outlined by IRB policy will be promptly followed.

As with all studies involving use of personal health information (PHI), there are certain risks associated with privacy and confidentiality. This study cannot be practically conducted without access to and use of PHI. We will only view and extract the minimal amount of data necessary to achieve our research aims. Personnel conducting the medical record review and extraction will be appropriately trained on best practices for confidentiality and have experience in biomedical research. We will not share any PHI with individuals outside of our study team.

## F. Data Management Methods

Electronic medical record information and other clinical data on patients at our institution will be extracted and stored on secure password-protected computers. We will store data from the abstracted records using approved systems, such as REDCap, and store the data on local servers behind a firewall. PHI will be maintained in secure, HIPAA-compliant databases, and will only be accessible to the study team. Once the patient cohort has been confirmed and all relevant clinical data extracted, identifiers included in study documents will be removed and a de-identified dataset created for contribution to the RAPID-19 repository.

The repository, containing pooled data across institutions, will be hosted on BCH's HIPAA-compliant Amazon Web Service (AWS) infrastructure. AWS is an offering by Amazon that allows easy and secure setup of environments for hosting applications and databases. Amazon provides tools for limiting access to the environments at different levels of granularity as well as backup and recovery mechanisms. A business associate agreement (BAA) has been established between BCH and Amazon. This compliance ensures the proper measures are taken to protect the data held within the secure environment. Access will be limited to the environment by establishing a Virtual Private Cloud (VPC) that makes the services hosted in Amazon function as if the environment was located within BCH’s own network. The same protections that apply to the BCH network apply to the VPC. Only individuals who have been specifically granted access will be able to view and analyze the data.

## G. Quality Control Methods

Study personnel will monitor data collection by periodically reviewing the data and data collection procedures for quality and completeness.

## H. Data Analysis Plan

The RAPID-19 repository will be built to support different types of analyses to define COVPD-19 in children. There are no pre-specified analysis plans for these studies.

## I. Statistical Power and Sample Considerations

Not applicable, as we do not aim to test a specific hypothesis.

## J. Study Organization

[list of investigators]

## K. References

1. Hong H, Wang Y, Chung HT, Chen CJ. Clinical characteristics of novel coronavirus disease 2019 (COVID-19) in newborns, infants and children. *Pediatr Neonatol*. 2020. doi:10.1016/j.pedneo.2020.03.001
2. Hwang TJ, Randolph AG, Bourgeois FT. Inclusion of Children in Clinical Trials of Treatments for Coronavirus Disease 2019 (COVID-19). *JAMA Pediatr*. 2020;Online Fir. doi:10.1001/jamapediatrics.2020.1888
3. Kothari C, Wack M, Hassen-Khodja C, et al. Phelan-McDermid syndrome data network: Integrating patient reported outcomes with clinical notes and curated genetic reports. *Am J Med Genet Part B Neuropsychiatr Genet*. 2018. doi:10.1002/ajmg.b.32579
4. The Consortium for Clinical Characterization of COVID-19 by EHR (4CE). *International Electronic Health Record-Derived COVID-19 Clinical Course Profiles: The 4CE Consortium*.; 2020.