

---

**PROTOCOL TITLE:** Repository of Aggregated Pediatric International Data on COVID-19 (RAPID-19)

**RESEARCH SITE:** Boston Children's Hospital

**PRINCIPAL INVESTIGATOR:** Florence Bourgeois, MD, MPH

## **A. Summary and Specific Aims**

We are proposing to build and host an international repository of data on pediatric Coronavirus Disease 2019 (COVID-19) cases in order to support research studies aiming to characterize the full clinical course and outcomes of children with COVID-19. De-identified patient-level data from patients treated at BCH and contributed from other organizations will be hosted on BCH's HIPAA compliant Amazon Web Service infrastructure. 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 the BCH network.

The pooled data source will be made available to 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 practices 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.<sup>1</sup> This will necessitate collaborative efforts across sites to build a robust resource to fully define the clinical course and outcomes of disease in children. This is particularly urgent as clinical trials have been for the most part excluding participation from children—as of mid- April, among 275 interventional trials studying COVID-19 and registered on the trial registry ClinicalTrials.gov, only 30 (11%) were open to any patients less than 18 years.<sup>2</sup>

The resource 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 disease
- 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

In order to provide a rapid solution to the ongoing fragmented collection of pediatric data, we are proposing a phased approach in which we begin now with a feasibility pilot and, if successful, scale our approach to enable participation by a range of registries across countries and networks, covering both general and specific pediatric patient populations. The feasibility pilot will serve to identify obstacles and opportunities, better define use cases and value of aggregate data, and inform prioritization of registries for future inclusion. We have formed a committee to begin to identify data

contributors and have participation from leaders of several large trial networks in Canada and the UK, as well the European Medicines Agency.

### C. Preliminary Studies

The approach is modeled on the principles underlying the BCH Biobank for Health Discovery and The Consortium for Clinical Characterization of COVID-19 by EHR (4CE). Florence Bourgeois (PI on this protocol) is the scientific lead of the BCH Biobank, which aims to establish standardized, compatible protocols for the enrollment of broadly consented participants.<sup>3</sup> A common Biobank platform provides search capability across annotated biospecimens, electronic medical records, and omics data to support feasibility queries and study design. To date greater than 16K patients have been enrolled at BCH and the project was recently awarded an NCATS grant to extend participation to additional sites ([www.grinnetwork.org](http://www.grinnetwork.org)).<sup>4</sup>

The 4CE consortium, built under the leadership of Paul Avillach (co-investigator on this protocol), leverages electronic medical record data from 96 hospitals across 5 countries ([www.covidclinical.net](http://www.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. Data covered 27,584 COVID-19 cases with 187,802 laboratory tests. Case counts and laboratory trajectories were concordant with existing literature and laboratory tests at the time of diagnosis showed hospital-level differences equivalent to country-level variation across the consortium partners.<sup>5</sup>

The proposed Pediatric COVID-19 Consortium will build off of these models and technical infrastructures, 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 aggregation of de-identified, patient-level data on pediatric patients diagnosed with COVID-19. Data will be hosted on BCH's HIPAA compliant Amazon Web Service infrastructure, which allows for embedded analytic tools and applications and permits controlled access and functionality. The pooled data source will be made available to approved collaborators who can query, analyze, and export de-identified patient-level data.
2. **Patient Selection and Inclusion/Exclusion Criteria**  
Clinical data on pediatric patients with lab-confirmed COVID-19 will be included in the repository. Data will be included from patients at BCH 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 clinical course of the disease and the outcomes in this patient population.

## **5. Data Collection Methods, Assessments, Interventions and Schedule (what assessments performed, how often)**

Data on eligible patients at BCH will be collected through extraction from the electronic medical record using a RedCAP database. Data will be collected once initially and the chart re-accessed for additional data collection until the final disease outcome is known. 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 this initial visit. However, if a patient is admitted with COVID-19, we will collect data on the disease progression through final convalescence, which may consist of hospital discharge or an outpatient follow up visit after discharge.

Data on patients diagnosed and treated at other institutions will be uploaded into the data repository as the data are provided to us.

## **6. Study Timeline**

We will maintain the repository 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 this analysis. 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 will be extracted and stored on secure password-protected servers. We will store data from the abstracted records using BCH approved systems, such as REDCap, and it will be stored at BCH on local servers behind the firewall, accessible only at BCH or through VPN. Participant identifying data will be maintained in secure, HIPAA compliant databases, and will only be accessible to the PI and her associates listed on the protocol. 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 analyses.

The repository containing pooled data across institutions will be hosted on BCH's HIPAA compliant Amazon Web Service 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) is signed 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 essentially 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**

A data repository will be built to support different types of future analyses to define COVDP-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**

Florence Bourgeois is the Principal Investigator and Paul Avillach a Co-Investigator. Additional study staff will assist with data extraction and management.

#### **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.
3. Bourgeois FT, Avillach P, Kong SW, et al. Development of the precision link biobank at Boston Children's Hospital: Challenges and opportunities. *J Pers Med*. 2017;7(4). doi:10.3390/jpm7040021
4. Mandl KD, Glauser T, Krantz ID, et al. The Genomics Research and Innovation Network: creating an interoperable, federated, genomics learning system. *Genet Med*. 2020. doi:10.1038/s41436-019-0646-3
5. 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.