**The Urgent Need for Research Coordination to Advance Knowledge on COVID-19 in Children**

Florence T. Bourgeois, MD, MPH1,2, Paul Avillach, MD, PhD1,2,3, Mark A. Turner, FRCPCH, PhD4

**Short title:** Research Coordination in Pediatric COVID-19

**Affiliations:** 1 Department of Pediatrics, Harvard Medical School, Boston, USA; 2 Computational Health Informatics Program, Boston Children's Hospital, Boston, USA; 3 Department of Biomedical Informatics, Harvard Medical School, Boston, USA; and 4 Department of Women's and Children's Health, Institute of Life Course and Medical Sciences, University of Liverpool, Liverpool Health Partners, Liverpool, UK

**Address Correspondence to:** Florence T. Bourgeois, MD, MPH, Boston Children’s Hospital, 300 Longwood Ave, Boston, MA 02115. Phone: 617-355-6624; E-mail: florence.bourgeois@childrens.harvard.edu

**Conflict of Interest Disclosures:** The authors report no financial relationships relevant to this article to disclose.

**Funding/Support:** No funding was secured for this work.

**Abbreviations:** Coronavirus Disease 2019, COVID-19; Repository of Aggregated Pediatric International Data on COVID-19, RAPID-19; Observational Medical Outcomes Partnership, OMOP

**Word count:** 1197

**References:** 8

**Table:** 1

The emergence of Coronavirus Disease 2019 (COVID-19) prompted a surge in research activity. Funding bodies swiftly allocated resources to establish research infrastructures and partnerships to study the novel virus. The scientific community realigned existing research and launched new studies to define the clinical course of COVID-19 and identify therapeutic candidates. Overall, comparatively fewer studies have been initiated in children relative to adults,1 in part due to the lower prevalence and disease morbidity recorded in pediatric populations. However, characterizing the disease in pediatric patients is critical to elucidate transmission dynamics, inform public health measures, and generate evidence on best practices for clinical care and therapeutic interventions. The life-threatening multisystem inflammatory syndrome further underscores the need for natural history studies and drug development in pediatric populations.2

Numerous observational studies have been established to collect information on pediatric patients with laboratory-confirmed COVID-19, including children in inpatient and outpatient settings, and those with specific underlying conditions (Table). These efforts have been driven largely by single institutions and research organizations. As would be expected during a pandemic response, there was initially limited opportunity to establish extensive regional or international collaborations and networks, and there has been little coordination across these initiatives. However, as we approach a steady state in morbidity rates and gain a better view of the longterm trajectory of the pandemic, it is important that we re-assess our approaches and take a deliberate and strategic approach to further resource investments. If the current fragmented approach is maintained, many studies will not yield the sample sizes needed to support robust analyses—especially as infection rates decline in certain areas—and resources will be wasted on inefficient enrollment and duplicative efforts. For example, we identified at least 3 unique non-interventional studies focused on pediatric patients requiring critical care in hospitals in North America.

Building on what has been accomplished to date, pediatric clinical data collected in studies could be aggregated to develop a larger, representative dataset, from multiple countries as well as global patient populations. This resource would not replace any individual studies, but provide a collaborative entity to complement local efforts. As an illustration of this approach, we have adopted this method with the launch of the Repository of Aggregated Pediatric International Data on COVID-19 (RAPID-19).3 This infrastructure leverages contemporary informatics methodologies that efficiently pool data, as has been previously accomplished with routine electronic health record and patient cohort data.4 RAPID-19 aggregates de-identified clinical data on pediatric patients with PCR confirmed SARS-CoV-2 infection curated across existing research activities. All data contributors are able to access the integrated dataset, which is maintained on a secure HIPAA and HITRUST compliant, cloud-based, Apache 2.0 open source platform (<https://github.com/hms-dbmi/pic-sure-documentation>)~~.~~ The infrastructure includes embedded analytic tools and export functionality which are managed by an API-enabled user interface with access controls for individual investigations. A steering group with representation from contributing sites oversees access and export of patient-level data. To ensure efficient and equitable use of the dataset, a number of mechanisms are in place to provide full transparency on site participation, research activities, data use, and opportunities for collaboration and authorship.

One fundamental challenge inherent to collaborative networks, is the need for data standardization across sites. This is a limitation we have encountered, as studies we have identified for participation in the Repository collect different variables and use distinct definitions for data elements. This limits options for combining datasets or variables across studies. While not all variables will be amenable to harmonization, a core set of data elements can be standardized to support queries and analyses across healthcare sites and study networks. We aim to also promote standardization of new data collection and have made study documents and tools available at [www.RAPID-19.org](http://www.RAPID-19.org) to support COVID-19 research by other investigator teams and facilitate addition of new cohorts.

Our experiences building the RAPID-19 Repository highlight key principles the scientific community should consider in the study of any pediatric condition or population.

Specifically, three key principles should be adopted. First, as a research community, we must build consensus on the importance of sharing pediatric data and adopt multi-institutional approaches as the default design. Research infrastructures and networks must implement contemporary informatics methodologies that support systematic computational approaches across research entities. At a minimum, all data should be FAIR: Findable, Accessible, Interoperable and Reusable.5 Whenever possible, new studies should deploy standard definitions that have been established for many pediatric specialties, such as the Pediatric Terminology sets developed by the National Institute of Child Health and Human Development.6 Further, harmonized approaches to structuring datasets are also essential to data sharing and reuse, including common formats and representations (e.g. terminologies and coding schemes). Pediatric observational studies need to use existing data models, such as the Observational Medical Outcomes Partnership (OMOP), the i2b2 framework, or the PCORNet Common Data Model to support transformation to a shared format amenable to reproducible analyses.

Second, pediatric studies must be developed with data sharing and combined analyses in mind, even if a prospective network is not possible or all relevant collaborations are not yet established at the outset of a study. Specifically, collaborative work must be considered in the study organization, design, and documentation to enable future partnerships. This includes using informed consent documents that specify sharing of de-identified patient information with external collaborators and allow secondary data use beyond primary research questions.7 Data use agreements for registries and other multi-institutional studies should be written to allow for additional sharing and research analyses outside of the original study, building in measures for embargo periods, country-specific laws for protection of personal and medical information, and appropriate authorship attributions. Study governance must also allow for flexibility to enable onboarding of new co-investigators and expansion of study infrastructure and oversight.

Third, participation in collaborative studies should be incentivized. This can be achieved, for example, by providing access to integrated datasets for approved investigators and by offering authorship opportunities to data contributors, such as group authorship by default for all data contributors. Study committees and working groups should include representation across sites and networks, including steering committees governing policies and procedures for research activities. More broadly, academic institutions need to recognize data sharing and input to patient-level datasets as an important contribution to the research enterprise and endorse these, for example, in promotion and tenure criteria.8 Finally, resources for data sharing need to be built into grants and institutional research infrastructure to make them feasible and sustainable.

The COVID-19 pandemic has highlighted the fragmentation of pediatric research. Ideally, research involves prospective formation of multi-institutional networks, but absent such infrastructure, it is essential that pediatric data be collected with the potential for collaboration and reuse. Sharing pediatric COVID-19 data and resources across existing studies will strengthen our research capacity during the current pandemic, including long-term study of natural history and identification of rare complications. The principles suggested in this Commentary are being applied in other scientific domains so that pediatric research is at risk of being left behind. We also note that insights gained from this pandemic should also serve as a stimulus to advance research programs involving other conditions and pave the way for routine use of coordinated approaches and application of sophisticated methodologies for efficient and scalable data integration.

**References**

1. 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

2. Feldstein LR, Rose EB, Horwitz SM, et al. Multisystem Inflammatory Syndrome in U.S. Children and Adolescents. *N Engl J Med*. 2020. doi:10.1056/nejmoa2021680

3. Repository of Aggregated Pediatric International Data on COVID-19 (RAPID-19). https://www.rapid-19.org. Accessed July 13, 2020.

4. 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

5. GO FAIR: FAIR Principles. https://www.go-fair.org/fair-principles/. Accessed July 13, 2020.

6. National Cancer Institute. Pediatric Terminology. https://www.cancer.gov/research/resources/terminology/pediatric. Accessed July 13, 2020.

7. 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

8. Rice DB, Raffoul H, Ioannidis JPA, Moher D. Academic criteria for promotion and tenure in biomedical sciences faculties: cross sectional analysis of international sample of universities. *BMJ*. 2020. doi:10.1136/bmj.m2081