

Translational Research,
Design and Analysis
Research Article

Cite this article: Bergquist T, Wax M, Bennett TD, Moffitt RA, Gao J, Chen G, Telenti A, Maher MC, Bartha I, Walker L, Orwoll BE, Mishra M, Alamgir J, Cragin BL, Ferguson CH, Wong H-H, Deslattes Mays A, Misquitta L, DeMarco KA, Sciarretta KL, and Patel SA. A framework for future national pediatric pandemic respiratory disease severity triage: The HHS pediatric COVID-19 data challenge. *Journal of Clinical and Translational Science* 7: e175, 1–10. doi: [10.1017/cts.2023.549](https://doi.org/10.1017/cts.2023.549)

Received: 12 January 2023
Revised: 28 April 2023
Accepted: 5 May 2023

Keywords:

Pediatrics; COVID-19; community challenges; machine learning; evaluation

Corresponding author:




M. Wax, PhD; Email: marie.thomas@gmail.com

*These two authors contributed equally to this work.

© The Author(s), 2023. Published by Cambridge University Press on behalf of The Association for Clinical and Translational Science. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided that no alterations are made and the original article is properly cited. The written permission of Cambridge University Press must be obtained prior to any commercial use and/or adaptation of the article.



A framework for future national pediatric pandemic respiratory disease severity triage: The HHS pediatric COVID-19 data challenge

Timothy Bergquist^{1,*} , Marie Wax^{2,*}, Tellen D. Bennett³, Richard A. Moffitt⁴, Jifan Gao⁵, Guanhua Chen⁵ , Amalio Telenti⁶, M. Cyrus Maher⁶, Istvan Bartha⁶, Lorne Walker⁷, Benjamin E. Orwoll⁷, Meenakshi Mishra⁷, Joy Alamgir⁸, Bruce L. Cragin⁹, Christopher H. Ferguson², Hui-Hsing Wong², Anne Deslattes Mays¹⁰ , Leonie Misquitta¹¹, Kerry A. DeMarco², Kimberly L. Sciarretta² and Sandeep A. Patel²

¹Sage Bionetworks, Seattle, WA, USA; ²United States Department of Health and Human Services, Biomedical Advanced Research and Development Authority, Administration for Strategic Preparedness and Response, Washington, DC, USA; ³University of Colorado, School of Medicine, Aurora, CO, USA; ⁴Stony Brook University, Stony Brook, NY, USA; ⁵University of Wisconsin-Madison, Madison, WI, USA; ⁶Vir Biotechnology, San Francisco, CA, USA; ⁷Oregon Health & Science University, Portland, OR, USA; ⁸ARIScience, Wayland, MA, USA; ⁹Wind City Applied Research, Lempster, NH, USA; ¹⁰United States Department of Health and Human Services, National Institutes of Health, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD, USA and ¹¹United States Department of Health and Human Services, National Institutes of Health, National Center for Advancing Translational Sciences, Bethesda, MD, USA

Abstract

Introduction: With persistent incidence, incomplete vaccination rates, confounding respiratory illnesses, and few therapeutic interventions available, COVID-19 continues to be a burden on the pediatric population. During a surge, it is difficult for hospitals to direct limited healthcare resources effectively. While the overwhelming majority of pediatric infections are mild, there have been life-threatening exceptions that illuminated the need to proactively identify pediatric patients at risk of severe COVID-19 and other respiratory infectious diseases. However, a nationwide capability for developing validated computational tools to identify pediatric patients at risk using real-world data does not exist. **Methods:** HHS ASPR BARDA sought, through the power of competition in a challenge, to create computational models to address two clinically important questions using the National COVID Cohort Collaborative: (1) Of pediatric patients who test positive for COVID-19 in an outpatient setting, who are at risk for hospitalization? (2) Of pediatric patients who test positive for COVID-19 and are hospitalized, who are at risk for needing mechanical ventilation or cardiovascular interventions? **Results:** This challenge was the first, multi-agency, coordinated computational challenge carried out by the federal government as a response to a public health emergency. Fifty-five computational models were evaluated across both tasks and two winners and three honorable mentions were selected. **Conclusion:** This challenge serves as a framework for how the government, research communities, and large data repositories can be brought together to source solutions when resources are strapped during a pandemic.

Introduction

Difficulty Identifying Children at Risk of Respiratory Infectious Diseases in Overwhelmed Healthcare Settings

Seroprevalence studies of pediatric patients (age 0–17) reported by the Centers for Disease Control and Prevention suggest that at least 75% of the U.S. pediatric population have been infected with SARS-CoV-2 [1]. As of December 2022, there have been more than 15.2 million reported cases of COVID-19 among U.S. children and adolescents [2]. Despite the high seropositivity of children, hospitalization rates and deaths among children and adolescents remain relatively low compared with the older population. Because of the disproportionate impact of SARS-CoV-2 on the older population, there was significantly more data to analyze and predict the risk factors associated with severe disease in the adult population.

However, as the pandemic progressed through waves of variants, leading up to the Delta variant during the summer of 2021, it became clear that some children were being hospitalized at a greater rate than in previous influenza seasons [3]. Additionally, healthcare providers began to notice previously healthy children presenting with a constellation of symptoms occurring weeks

after COVID-19 infections: fever, myocarditis and shock, abdominal pain and mucocutaneous findings. This postinfection inflammatory response to SARS-CoV-2 was termed multisystem inflammatory syndrome (MIS-C) [4–8]. Given the presentation of MIS-C and the general lack of data in pediatric hospitalizations, healthcare providers caring for children and adolescents did not know whether the risk factors that resulted in severe disease in adults, also translated into children and adolescents.

Hospital providers, particularly hospitals not specialized in pediatrics, needed enhanced tools to help determine children at risk for severe disease outcomes.

Need for a Community-Based Computational Model Challenge to Inform on Infection Severity

Feedback from healthcare workers, heads of large hospital systems, and pediatric government working groups indicates that emergency departments and ICUs faced two major pain points during the COVID-19 surge that still remains today: assessing patients' need for hospitalization and escalated care to provide available interventions. Children's hospitals faced an additional burden in protecting their already vulnerable patients. Due to the rare occurrence and distributed nature of pediatric COVID-19 hospitalizations, approximately 70% of those pediatric hospitalizations occur in non-pediatric-specific hospitals [9]. Many prediction models were based on small sample sizes and resulted in limited diagnostic accuracy.

The Biomedical Advanced Research Development Authority (BARDA), part of the Administration for Strategic Preparedness and Response (ASPR) within the U.S. Department of Health and Human Services (HHS), wanted to catalyze the development of EHR-based algorithms with the ability to predict risk of severe outcomes for earlier care, stratify pediatric patients for available interventions, and help overburdened healthcare workers triage patients. The community-based challenge offered an additional ability to reach a broader community. They developed two key clinical questions for the pediatric COVID-19 data challenge: (1) Of pediatric patients who test positive for COVID-19, who are at risk for hospitalization? and (2) Of pediatric patients who test positive for COVID-19 and are hospitalized, who are at risk for needing mechanical ventilation or cardiovascular interventions?

The National COVID Cohort Collaborative [10]

The National Institutes of Health (NIH) National Center for Advancing Translational Sciences (NCATS) National COVID Cohort Collaborative (N3C) has spearheaded the collection and harmonization of a large EHR repository that, at the time of this writing, represents over 7 million COVID-19-positive patients and over 11 million confirmed COVID-19 negative patients from 76 sites across the U.S. These data-contributing sites include academic medical centers, pediatric hospitals, and healthcare networks. The N3C data repository is updated on a weekly basis with data from contributing sites as well as data from newly onboarded contributing sites. The harmonization process includes mapping site data to the Observational Medical Outcomes Partnership (OMOP) common data model and converting the clinical codes to the OMOP standard vocabularies (Systematized Nomenclature of Medicine, Medical Prescription Normalized, Current Procedural Terminology 4th Edition, etc.). COVID-19-positive patients were defined as having at least one positive COVID-19 antigen test. The repository includes over 900,000 pediatric patients who have tested positive for COVID-19 and

serves as one of the largest, centralized databases of pediatric clinical records in the U.S. At the time of challenge inception, the enclave was being used for research study purposes and many clinical sites were in the process of onboarding but the enclave was an optimal choice for the Pediatric COVID-19 Data Challenge due to the scope of data available, the robust governance oversight, and the sophisticated computing platform. The clinical questions were developed into two challenge tasks with N3C.

Materials and Methods

Designing the Challenge Tasks with N3C Data

The challenge utilized available de-identified pediatric patient data from the N3C enclave that included age, gender, height, weight, medical history, lab results, county-level social determinants of health data, and available medications. The challenge cohort included patients who were 18 or under and who had a COVID-19-positive PCR, antigen, or serum antibody test (Supplemental Table 1). The training data available to the challenge participants included accumulated data from August 2020 to July 30, 2021, consisting of 203,508 pediatric COVID-19-positive patients from 55 contributing sites (Table 1). The testing holdout set used prospectively-collected data that was accumulated during initial model development from July 30, 2021 to December 9, 2022 consisting of 201,083 pediatric COVID-19 patients from 64 sites (Table 1). This accumulated data was not available to challenge participants during the model development phase and was used for the evaluation. Patients that appeared in the training data were removed from the prospectively collected testing data.

The data used for this challenge was derived from the safe harbor, de-identified data. The use of de-identified data lowered the barrier for entry into the challenge, since not all participating institutions required Institutional Review Board (IRB) approval for access and expanded the number of eligible participants to include international participants who were not permitted to access the limited dataset at the time of the challenge. Key differentiations between de-identified data and limited data include masked zip codes to the first three numbers, patient-level date shifting of ± 180 days of each clinical record, and abstraction of birth dates to the birth year.

COVID-19-specific severity can be difficult to accurately define when using EHR data. Most of the time, the critical information to determine if a severe health outcome was caused by COVID-19 or simply occurred after a COVID-19-positive test does not exist. When designing the challenge questions, the limitations of using EHR data were an important consideration.

Task 1 Design

For task 1, participants were asked to address the following question: *Of pediatric patients who test positive for COVID-19 in an outpatient setting, who are at risk for hospitalization?* The task was designed to have teams build predictive tools that could be used in an outpatient setting to assess the risk of a given pediatric patient progressing to a level of COVID-19 severity that warranted a trip to the hospital (including an emergency department encounter (Supplemental Table 4) or inpatient hospital visit (Supplemental Table 3)). COVID-19-related outpatient visits were defined as outpatient visits (Supplemental Table 2) that occurred within seven subsequent days of a patient's earliest COVID-19-positive test. The presence of an inpatient visit (Supplemental Table 3) or emergency department visit (Supplemental Table 4) within 35 days

Table 1. Demographic breakdown of the training data and the prospectively collected testing data. The training data was derived from the N3C data enclave from August 2020 to July 30, 2021 and the prospectively collected testing data was derived from the N3C data enclave as of from July 30, 2021 to December 9, 2021. Testing data also included eight new contributing sites.

| Demographic | | Training data (n = 203,508) | | | | | | | | Testing data (n = 201,083) | | | | | | | |
|-------------|------------------|-----------------------------|----------|----------------|----------|----------------|----------|----------------|----------|----------------------------|----------|----------------|----------|----------------|----------|----------------|----------|
| | | Task 1 | | | | Task 2 | | | | Task 1 | | | | Task 2 | | | |
| | | True positives | | True negatives | | True positives | | True negatives | | True positives | | True negatives | | True positives | | True negatives | |
| Total | | 4,062 | | 142,840 | | 830 | | 8,292 | | 4,753 | | 139,737 | | 708 | | 9,270 | |
| Ethnicity | Not Hispanic | 2,852 | (70.21%) | 93,513 | (65.47%) | 504 | (60.72%) | 5,130 | (61.87%) | 3,569 | (75.09%) | 101,801 | (72.85%) | 464 | (65.54%) | 6,487 | (69.98%) |
| | Hispanic | 849 | (20.9%) | 27,885 | (19.52%) | 220 | (26.51%) | 2,199 | (26.52%) | 867 | (18.24%) | 18,832 | (13.48%) | 199 | (28.11%) | 2,220 | (23.95%) |
| | Unknown | 361 | (8.89%) | 21,442 | (15.01%) | 106 | (12.77%) | 963 | (11.61%) | 317 | (6.67%) | 19,104 | (13.67%) | 45 | (6.36%) | 563 | (6.07%) |
| Gender | Female | 2,137 | (52.61%) | 70,792 | (49.56%) | 370 | (44.58%) | 4,226 | (50.96%) | 2,405 | (50.6%) | 69,327 | (49.61%) | 312 | (44.07%) | 4,664 | (50.31%) |
| | Male | 1,920 | (47.27%) | 71,284 | (49.9%) | 460 | (55.42%) | 4,063 | (49%) | 2,348 | (49.4%) | 70,053 | (50.13%) | 396 | (55.93%) | 4,605 | (49.68%) |
| | Unknown | >20 | (>0.49%) | 764 | (0.53%) | >20 | (>2.41%) | >20 | (>0.24%) | >20 | (>0.42%) | 357 | (0.26%) | >20 | (>2.82%) | >20 | (>0.22%) |
| Age | 0-4 | 963 | (23.71%) | 19,643 | (13.75%) | 208 | (25.06%) | 2,646 | (31.91%) | 1,258 | (26.47%) | 20,382 | (14.59%) | 168 | (23.73%) | 3,199 | (34.51%) |
| | 5-11 | 810 | (19.94%) | 40,656 | (28.46%) | 236 | (28.43%) | 1,599 | (19.28%) | 1,189 | (25.02%) | 50,505 | (36.14%) | 181 | (25.56%) | 2,161 | (23.31%) |
| | 12-19 | 2,289 | (56.35%) | 82,541 | (57.79%) | 386 | (46.51%) | 4,047 | (48.81%) | 2,306 | (48.52%) | 68,850 | (49.27%) | 359 | (50.71%) | 3,910 | (42.18%) |
| Race | White | 2,714 | (66.81%) | 96,660 | (67.67%) | 361 | (43.49%) | 4,171 | (50.3%) | 3,254 | (68.46%) | 98,889 | (70.77%) | 288 | (40.68%) | 4,573 | (49.33%) |
| | Black | 624 | (15.36%) | 14,269 | (9.99%) | 221 | (26.63%) | 1,894 | (22.84%) | 698 | (14.69%) | 14,907 | (10.67%) | 181 | (25.56%) | 2,314 | (24.96%) |
| | Asian | 44 | (1.08%) | 1,720 | (1.2%) | 22 | (2.65%) | 145 | (1.75%) | 57 | (1.2%) | 1,821 | (1.3%) | >20 | (>2.82%) | 153 | (1.65%) |
| | Pacific Islander | >20 | (>0.49%) | 330 | (0.23%) | >20 | (>2.41%) | 37 | (0.45%) | >20 | (>0.42%) | 181 | (0.13%) | >20 | (>2.82%) | 25 | (0.27%) |
| | Other | 364 | (8.96%) | 13,676 | (9.57%) | 73 | (8.8%) | 622 | (7.5%) | 400 | (8.42%) | 9,874 | (7.07%) | 58 | (8.19%) | 747 | (8.06%) |
| | Unknown | 306 | (7.53%) | 16,185 | (11.33%) | 153 | (18.43%) | 1,423 | (17.16%) | 340 | (7.15%) | 14,065 | (10.07%) | 162 | (22.88%) | 1,458 | (15.73%) |

of a COVID-19 outpatient visit was determined to be the true positive outcome. True negatives were COVID-19-positive patients who tested positive within seven days of an outpatient visit but had no inpatient or emergency room visit within 35 days of the outpatient visit. A prediction window of 35 days was set in order to capture patients who were being hospitalized for either acute COVID-19 disease or MIS-C. A sensitivity analysis confirmed that patients were still being admitted with MIS-C up to five weeks after initial COVID-19-positive results (Supplementary Analysis 1). Excluded were patients with same-day hospitalizations, in which the hospitalization or emergency room visit occurred on the same day as the outpatient visit making inpatient and outpatient data difficult to discern. Additional information can be found in Supplemental Materials: Task 1 Outcome Definitions.

Task 2 Design

Following the expected clinical progression from the task 1 scenario for severe COVID-19 patients, the second task focused on predicting the risk of a patient needing additional medical intervention once they were hospitalized. Task 2 asked participants to address the question, *Of pediatric patients who tested positive for COVID-19 and were hospitalized, who are at risk for needing mechanical ventilation or cardiovascular interventions?* COVID-19-positive hospitalized patients were defined as any patients who tested positive for COVID-19 within seven days of being admitted to an inpatient visit. The true positives included all patients who, during their hospitalization, needed mechanical ventilation (Supplemental Table 5), Extracorporeal Membrane Oxygenation (Supplemental Table), cardiovascular support (Supplemental Tables 7-14), or expired during their hospital stay. Additional information can be found in Supplemental Materials: Task 2 Outcome Definitions.

Developing Evaluation Criteria

Quantitative Metrics

The desired outcome for this challenge was the development of models that could be used to identify at-risk pediatric COVID-19 patients to prioritize medical interventions or preventative measures. Sensitivity, precision, and specificity were considered in the quantitative evaluation, but additional weight was put on sensitivity to prioritize the identification of more at-risk patients. All submitted models were evaluated using Area Under the Precision-Recall Curve (AUPR), the $F_{\beta_{\max}}$ statistic ($\beta=2$), and the Area Under the Receiver Operator Curve (AUROC). The F_{\max} was calculated by finding the threshold at which the models had the highest score. This was done to give teams flexibility in deciding their model thresholds. F_{β} and AUPR were used due to the imbalanced nature of the data, and F_2 was used to put additional weight on sensitivity. AUROC was used to calculate cross-site generalizability of the models. Because N3C is a central repository from a wide range of healthcare centers, the performance of each model's AUROC could be calculated separately by site.

Qualitative Metrics

Historically, most data challenges have focused on the quantitative aspect of model assessment, asking participants to maximize their model's performance. However, clinical models trained on EHR data can produce accurate models, but the model itself has methodological issues and when further scrutinized, the performance drastically decreases when given new data due to concept drift, changes in clinical practice, changes in

informatics practices, or changes in the presentation of disease. Successful clinical models typically require a multidisciplinary team of informaticists, clinicians, and statisticians to develop and implement accurate and impactful models. A panel of subject matter experts in medical diagnostics, clinical, statistical, and informatics domains evaluated the submitted models for their potential clinical utility and for their reproducibility. Clinical utility was further evaluated by interpretability, timeliness, and utility. Reproducibility was evaluated by prediction, technical, and documentation reproducibility. Additional information can be found in Supplemental Materials: Model Evaluation.

Challenge Administration and Governance

The Pediatric COVID-19 Data Challenge was held from August 19, 2021 to December 17, 2021 in three phases: Onboarding (August 19 to September 15, 2021), Model Development (September 15 to December 17, 2021), and Evaluation (December 17, 2021 to March 9, 2022) (Supplemental Figure 1). In order to receive access to the data, participants needed to follow all the governance procedures established in challenge.gov and N3C. This includes having a signed Data Use Agreement between their host institution and N3C, receiving IRB approval if required by their home institution, agreeing to not download any data from the enclave without permission, and agreeing to the terms of use of the N3C enclave. Usually, when researchers access the N3C enclave, they receive access to all available N3C data as either de-identified or limited data. For this challenge, researchers were given access to a subset of the de-identified N3C dataset and were not able to access the full N3C data while they were working in the challenge-specific project space. Additional information on the challenge administration is available in the supplemental materials.

Many of the participants of the challenge were new to the N3C enclave. To familiarize them with the enclave, a series of webinars were provided to participants with relevant enclave tools and methods for participating in the challenge, a "Getting Started" document was created to highlight key tutorials available in the training portal, and informal office hours were conducted on a weekly basis to answer questions and troubleshoot issues. The enclave also had a robust support infrastructure that enabled this challenge, including responsive technical support for when technical issues arose, an extensive library of tutorials developed by the N3C community, and a library of code repositories and concept sets that were developed by the N3C community for the N3C community [11,12].

Model Evaluation

The evaluation team consisted of government, industry, and academic experts representing clinical, data science, medical countermeasure product development, and informatics domains from BARDA, Sage Bionetworks, University of Colorado, Stony Brook University, Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), NCATS, and Health Resources and Services Maternal and Child Health Bureau (HRSA MCHB). Models were scored based on the quantitative and qualitative metrics defined above. Additionally, the top-scoring models were explored for further government utility and interest. The winners and honorable mentions were then explored post-challenge for performance for evolving variants. Additional information can be found in Supplemental Materials: Model Evaluation.

Table 2. Highest quantitatively scoring models for both Task 1 and Task 2. The highest scoring model in Task 1 achieved an AUROC of 0.756, while the highest scoring model in Task 2 achieved an AUROC of 0.853. Cross-site AUROC is the AUROC when the model's performance is calculated on each individual data site and the mean of those AUROCs is calculated

| Team | Task 1 scores | | | | Team | Task 2 scores | | | |
|--------|---------------|--------|------------------|---------------------------|--------|---------------|--------|------------------|---------------------------|
| | F2 | AUPR | Cross site AUROC | Cross site AUROC variance | | F2 | AUPR | Cross site AUROC | Cross site AUROC variance |
| Team M | 0.2865 | 0.1438 | 0.7561 | 0.0039 | Team Q | 0.594 | 0.5907 | 0.8533 | 0.0028 |
| Team K | 0.2767 | 0.1202 | 0.752 | 0.0041 | Team L | 0.5771 | 0.4928 | 0.8541 | 0.0049 |
| Team L | 0.2439 | 0.1162 | 0.7412 | 0.0057 | Team X | 0.5895 | 0.5064 | 0.8356 | 0.0055 |
| Team H | 0.2539 | 0.1149 | 0.7241 | 0.005 | Team J | 0.572 | 0.4962 | 0.8323 | 0.0053 |
| Team F | 0.2514 | 0.1006 | 0.7241 | 0.0038 | Team E | 0.5763 | 0.4795 | 0.836 | 0.0061 |
| Team W | 0.2144 | 0.0892 | 0.7058 | 0.0064 | Team K | 0.5414 | 0.4432 | 0.8094 | 0.0069 |
| Team U | 0.2366 | 0.0671 | 0.696 | 0.0044 | Team H | 0.5411 | 0.39 | 0.8045 | 0.0045 |
| Team J | 0.2394 | 0.0976 | 0.6871 | 0.0079 | Team A | 0.5193 | 0.3774 | 0.7824 | 0.0055 |
| Team D | 0.2052 | 0.0777 | 0.6768 | 0.0051 | Team O | 0.4992 | 0.385 | 0.777 | 0.0036 |
| Team S | 0.2153 | 0.081 | 0.6635 | 0.0051 | Team C | 0.4879 | 0.3363 | 0.7115 | 0.0089 |

Evaluation on COVID-19 Variants Post-Challenge

After the close of the challenge, new data was continually collected in the N3C enclave. Toward the end of the challenge, a new variant of COVID-19, Omicron, became the dominant strain in the U.S. Patients infected with Omicron have shown different clinical phenotypes than previous strains. To test the generalizability of the winning and honorable mention models with different variants, the performance of the models was evaluated on data collected from December 9, 2021 to April 7, 2022 to include Omicron data. This data included new patients who appeared in the N3C enclave during this time frame and who were part of the same sites that were contributing data partners during the challenge.

Results

The Pediatric COVID-19 Data Challenge began on August 19, 2021 and was closed to submissions on December 17, 2021. Over the course of the challenge, 200 participants and 88 teams were fully on-boarded into the N3C enclave. Fifty-five total models were submitted for evaluation across both tasks.

The highest quantitatively scoring model for Task 1 to predict need for hospitalization achieved an F2 of 0.286, an AUPRC of 0.144, and a cross-site AUROC of 0.756. The highest quantitatively scoring model for Task 2 to predict the need for respiratory and cardiovascular intervention achieved an F2 of 0.594, an AUPRC of 0.591, and a cross-site AUROC of 0.853. (Table 2)

The highest qualitatively scoring models in Task 1 could predict hospitalization up to 5 weeks out (Fig. 1) and Task 2 could predict the need for cardiac and respiratory interventions the most accurately up to 4 days before the outcome (Fig. 2). These models performed well across different races, gender, age, and BMI percentile with minimal variability within groups. Of note, the highest qualitative scoring model for Task 2 was not the best at capturing patients that would develop MIS-C but was in the top four. Additionally, the top models performed well across different clinical site types – from small rural clinics to large pediatric hospitals. Incidentally, the two highest quantitatively scoring models also had some of the highest qualitative scores. Additional information can be found in Supplemental Materials: Model Evaluation.

Information Leakage

During evaluation, the subject matter experts discovered variables that were included in the dataset that were either highly correlated with the true positive outcomes or were proxies for true positive outcomes. In Task 2, all the interventions of interest that defined true positive outcomes were removed from the patient's record during the covid-related hospitalization; however, many models used same-day data from the patient's record that was indicative of severe patient status. In a follow-up evaluation, all same-day records were removed except for measurement data. The top model remained the top model, however, many models drastically reduced their overall standing, highlighting the need to better identify outcome information leakage in challenge data.

The winner of Task 1 was from the Department of Biostatistics & Medical Informatics (BMI) at the University of Wisconsin-Madison (UWisc-Madison-BMI). They used a high-performing gradient boosting method and handcrafted features extracted from multisite EHR data to build their model. In particular, their feature extraction procedure summarized patients' medical conditions and drug exposures using medical meaning concepts such as International Classification of Diseases and Anatomical Therapeutic Chemical codes, which reduced the dimensionality of EHR data, and enhanced model interpretability. They also used a subset of COVID-19-related lab measurements and recent values prior to the patient's COVID-19 diagnosis and customized the model training/tuning procedure, to make it resistant to sample size bias and more generalizable across multiple sites. Additional information can be found in Supplemental Materials: Winners and Honorable Mentions Methods.

The winner of Task 2 was from Vir Biotechnology (Vir). This team used a gradient-boosting tree classifier, capable of extracting patterns from the complex set of EHRs. The team focused on extracting data from laboratory measurements, disease conditions, and past medical interventions to employ manual data cleaning, creation of new aggregate variables, and further harmonization of the data model. Not only did this group have the highest quantitative score but they also employed a missingness-aware classifier, capable of learning from the patterns of data availability and which avoids the imputation of missing data and overfitting by evaluating their trained classifier. In an additional analysis where

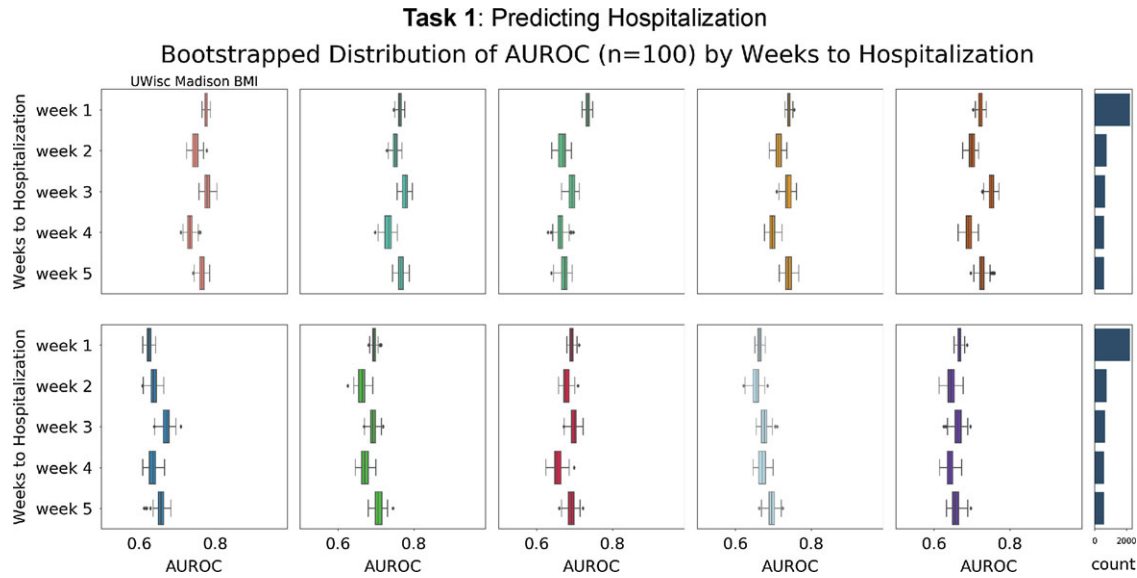


Figure 1. Time to outcome assessments for the top 10 quantitative scoring models in Task 1 predicting hospitalization by weeks 1–4 from the time of the outpatient visit. All team names have been removed except for the winning team. Outcome time bins are defined as noncumulative counts of patients who have the outcome in question within the given time window. Patients included in week 1 are not included in week 4 and are excluded from AUROC calculation. The boxplot represents the bootstrapped distribution of the AUROC ($n = 100$) for the given time window. The bar chart on the right shows the number of patients who have the outcome of interest during the given time window.

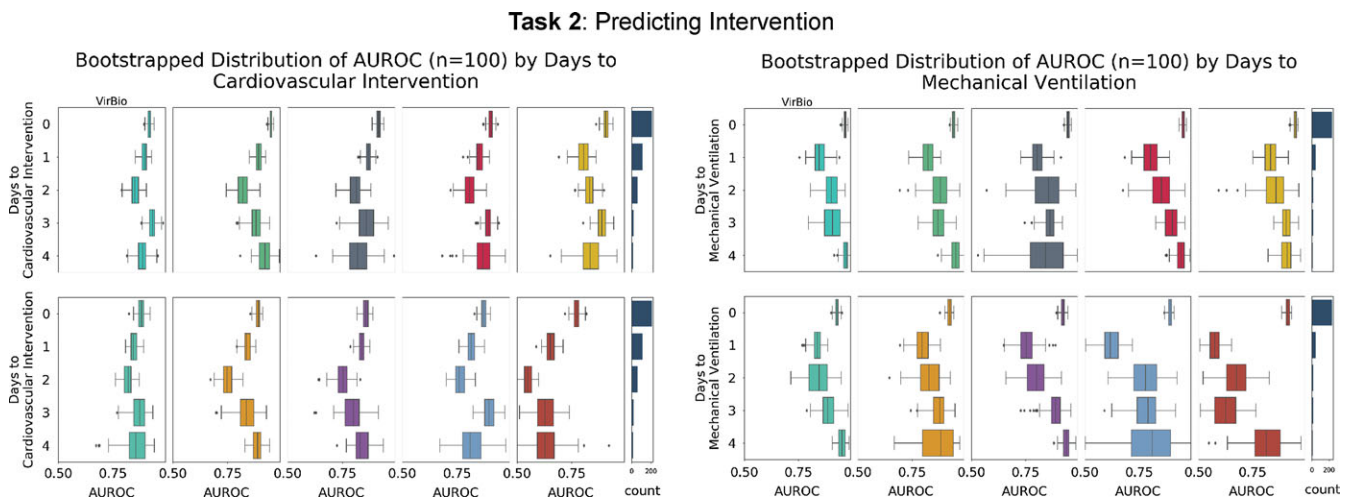


Figure 2. Time to outcome assessments for the top 10 quantitative scoring models in Task 2 predicting cardiovascular interventions and mechanical ventilation separately by days 0–4 from the time of hospitalization. All team names have been removed except for the winning team. Outcome time bins are defined as noncumulative counts of patients who have the outcome in question within the given time window. Patients included in day 0 are not included in day 4 and are excluded from AUROC calculation. The boxplot represents the bootstrapped distribution of the AUROC ($n = 100$) for the given time window. The bar chart on the right shows the number of patients who have the outcome of interest during the given time window.

all clinical information except for measurements were removed from the hospitalization period, their model maintained its high performance, scoring the highest among all submitted models (Supplemental Analysis 2). Additional information can be found in Supplemental Materials: Winners and Honorable Mentions Methods.

Three honorable mentions were named based on unique features or capabilities of their models: a team from the Oregon Health & Science University for feature interpretability & design, a citizen scientist from Wind City Applied Research for clinical utility, and ARIScience for Computational Methodology. All three teams also had very high quantitative scores and unique contributions to the qualitative scores. Oregon Health & Science University used a common set of predictors including

demographics, laboratory values, and associated diagnosis codes to employ an ensemble classifier that combined individual predictions from logistic regression, random forest, gradient-boosted tree, and artificial neural network models. They used Shapley Additive Values to provide individual-level and population-level explanations for model predictions. This high-performing approach provided clinicians with an outcome prediction and an individualized explanation with predictors for intervention. Wind City Applied Research leveraged the extensive clinical experience already utilized in the N3C community to create model features derived from existing electronic health record code sets from N3C to create an XGBoost code and feature importance matrix. Finally, ARIScience used clinical and laboratory indicators from pre-visit and during-visit data that was normalized by age,

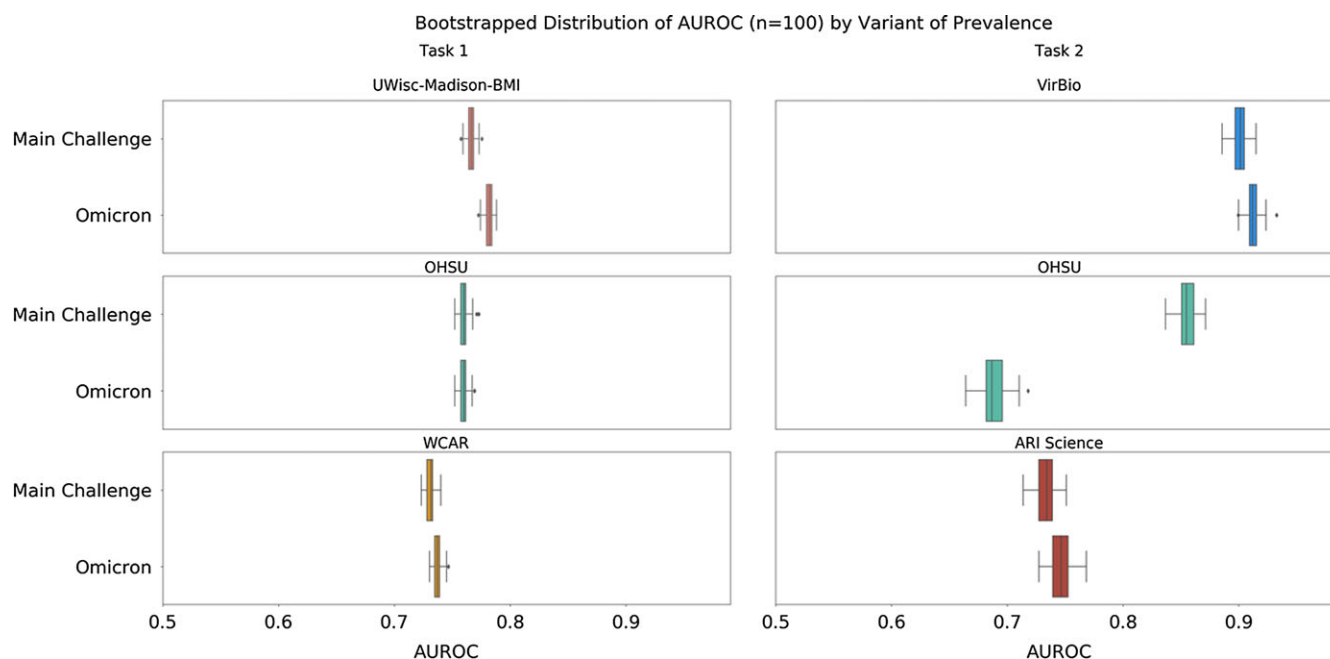


Figure 3. Results of evaluating the winners and honorable mentions on N3C data from the time period when the omicron variant was the most prevalent strain of COVID-19. The main challenge data was used to evaluate models during the challenge. Each boxplot compares the bootstrapped distributions ($n = 100$) of models performance using AUROC between the main challenge data and the omicron data. The results from Task 1 are in the left column and the results from Task 2 are in the right column.

gender, and other demographic attributes and fed into random forest, neural network, regression-based, Naïve Bayes, and neighborhood-based artificial intelligence (AI) models to create ensembles of predictions. Additional information can be found in Supplemental Materials: Winners and Honorable Mentions Methods.

New COVID-19 Variants- Post-Challenge Analysis

The models from the two winners and the honorable mentions were run on a prospectively collected dataset from the time when the Omicron variant was the most prevalent strain of COVID-19. Models were trained on the original challenge training dataset and applied to the newly collected Omicron-era dataset. The dataset was evaluated for UWisc-Madison-BMI, Oregon Health & Science University, and the Wind City team for Task 1 and Vir, Oregon Health & Science (OHSU), and ARIScience team submissions for Task 2 (Fig. 3). For Task 1, none of the models saw a significant decrease in their performance, and in the case of UWisc-Madison-BMI and Wind City, their models saw a nonsignificant increase in their performance (0.016 and 0.006, respectively). For Task 2, the Oregon Health & Science University saw a significant drop in performance (-0.167), while Vir and ARIScience saw a small increase in their performance (0.012 and 0.014, respectively).

Discussion

For unknown emerging infectious diseases, particularly during hospital surges, government decision-makers need to be able to support the development and deployment of accurate and useful clinical decision-support tools to educate healthcare workers and to reduce hospital burden. Nationwide, this approach allows healthcare workers to focus on patients who need escalated care. These clinical decision support tools can help healthcare workers guide available interventions and allow the government to triage resources to hospitals that need them the most.

Organization and Administration

Initial reports of MIS-C were sporadic but what was known was that the Kawasaki-like symptoms seen in COVID-19 children were occurring several weeks after initial diagnosis and in children that were seemingly previously healthy. The challenge needed to include nearly real-time data from pediatric COVID-19 patients nationwide to reflect a sufficient size to ensure diversity and a relevant case number to evaluate for severe COVID-19 and MIS-C. This allowed for the development of computational models with the potential for broad applicability. N3C had the appropriate regulatory, policy, privacy, and security protections in addition to a robust computing platform but also was building broad geographical representation, including at pediatric hospitals. N3C had limited, deidentified, and synthetic data available. The use of the limited data set required each participant to obtain their own IRB determination letter, which the organizers felt would greatly limit the number of organizations that would participate so the use of deidentified data was utilized for the challenge. However, the use of deidentified data limited each team's ability to integrate more granular location data including for social determinants of health, incorporate pandemic time course information and use granular age information of the newborn population. The utilization of limited data set, potentially through a centralized IRB that could be implemented for challenges, from the onset could help address some of these limitations in the future.

The N3C enclave is a unique resource for building scalable data analyses on large datasets. However, the platform uses nonstandard methods to create code repositories and to organize datasets. New users of the platform, therefore, had a steep learning curve, and many participants spent their first few months in the challenge exploring how to use the enclave before they were even able to begin building their models. Once this issue was identified, multiple challenge-specific tutorial seminars were organized to help familiarize new users with the challenge data, available tools,

and common issues. Future challenges using unconventional platforms should plan tutorial sessions within the first few weeks of the challenge to help jumpstart new users in the platform. Additionally, regular office hours should be available starting early in the challenge to field problems and questions that may arise as users learn the platform.

Community Challenge

The Pediatric COVID-19 Data Challenge community participants varied from a single citizen scientist to an 18-person team with representation from multiple countries, pediatric specialties, and data science domains. The international community challenge brought perspectives from all over the world – small rural areas, large metropolitan cities, multiple children’s hospitals, small businesses, and large pharmaceutical and diagnostic companies but also allowed for the utilization of different expertise. Unlike many data challenges, this challenge did not seek to make teams public, allowing competitors to work to leverage strengths, rural communities to participate to gain pediatric insights to bring back to their communities, and small businesses to compete without fear of how poor scores may be viewed. Winners and honorable mentions were given the option to make their computational model public for others to refine post-challenge through open collaboration or to keep their developed computational model private and potentially seek follow-on investments from government agencies and external sources as well as continue analytical and clinical validation toward regulatory pathways.

Evaluation

While the quantitative metrics helped evaluators focus on the models that would be best to assess pediatric patients at risk for severe COVID-19, the qualitative metrics allowed evaluators to explore the best models based on subject matter expert interest and priority. For example, agency representation reflected interest in predictive capabilities of models for community and emergency department triage, features of importance and performance in special cohorts and changing variants for educational material and communication, and overall model performance for medical countermeasure development and preparation for further validation. Subject matter expertise in health data science reflected interest in mapping to EHRs and well-annotated code while pediatric clinicians expressed the most interest in identifying at-risk groups to target available interventions. The deep evaluation of submissions by experts of varied backgrounds allowed government agencies to glean a holistic understanding of the clinical models, and to have deeper insight into the important predictive factors, the methodological robustness, and the potential for clinical impact that a pure quantitative optimization would not have achieved.

The evaluation criteria used for the challenge were based on quantitative metrics to address precision, recall, sensitivity, and specificity as well as qualitative metrics to address utility and reproducibility. While the evaluation criteria resulted in high-performing winners, the resulting winners had too many features to be directly mapped to an EHR system. In future challenges, incorporating a metric such as “translational feasibility” that would score models on their potential for direct validation and incorporation into clinical decision workflows would be beneficial. While large models can be implemented into an EHR, translating models that were built on a harmonized central repository into an EHR brings unique informatics challenges. Smaller models, while

often sacrificing accuracy, are often more easily translated to an EHR due to the manageable task of mapping all the features to their equivalent clinical concepts in use in the EHR. While the metrics for this challenge included interpretability and timeliness, which are key considerations when implementing a clinical prognosis tool, having an additional metric that considers the size and feature space of a given model would be helpful in identifying even more clinically impactful solutions.

Winners and Honorable Mentions

Both UWisc-Madison-BMI’s and Vir’s models used forms of ontology roll-up or concept binning to reduce their feature space. Both team’s also incorporated time into their features to account for the longitudinal properties of EHR data. UWisc-Madison-BMI binned each feature into two-time windows, counting all concept occurrences within three days of the outpatient visit and prior to the three-day window. Vir binned each feature into three-time windows, concept counts with four days of hospitalization, within four to eight days of hospitalization, and prior to eight days before hospitalization. These methods helped the top models maintain robust performance to changing data and improved interpretability.

Both teams showed an increased awareness of the limitations and caveats of using EHR data for predictive modeling. UWisc-Madison-BMI’s ordering of the contribution sites by COVID-19 positive prevalence was an important data curation decision that helped in their improved generalizability. Vir’s model included a wide range of measurement information that gave their model an advantage when other types of clinical data were removed from the test set. When all clinical records, except for measurements, were removed from the first day of hospitalization, Vir’s model decreased in performance (0.90 to 0.83 AUROC) but maintained its status as the highest-scoring model while most other models drastically decreased in performance.

The models that performed well in the qualitative metrics also did well in the quantitative metrics. In fact, the highest-scoring qualitative models both were the highest-scoring quantitative models in their respective tasks. This makes intuitive sense, since models that are well thought out, well documented, and well designed, are most often going to have good quantitative scores.

Impactful features from top-performing models suggested that children at risk for severe outcomes include patients with extreme BMI (both underweight and overweight), cancer, diabetes, and preexisting heart conditions. While many models did not have specific features for cancer or diabetes, many of the features suggested that models were prioritizing patients with tangential clinical elements that indicated diabetic or oncology patients (e.g. glucose measurements, related catheter procedures). During the evaluation, the feedback from the clinician subject matter experts was important to highlight these clinical nuances and to point out where features made sense or were clearly proxy features for other foundational health problems.

By the time the challenge ended, the Delta variant was no longer the dominant strain and N3C had onboarded a much wider geographical spread of contributing sites bringing concern to the continued utility of the high-performing models. The models were evaluated post-challenge on new data that had accumulated in the enclave when Omicron was the dominant variant. When the models were evaluated with new data to include Omicron-dominant data, models in Task 1 did not decrease in performance and in some cases, had a slight increase in performance. This may

indicate that major risk factors for hospitalization may not have changed from Delta to Omicron or that some of the hospitalizations may not have been directly COVID-19 related. In contrast, the significant drop in performance in Task 2 for one of the teams on the omicron-included data suggests that the clinical risk factors for pediatric severity shifted from delta to omicron. For example, while many of the important features in OHSU's model remained the same, numerous features either dropped out of the most important features or new measurements became more important such as mean platelet count and age. Also, practice patterns have possibly changed in terms of either (1) administration of therapies that define severe outcomes, or (2) unobserved effects of vaccination or prior infections.

Limitations

In addition to the limitations outlined by using deidentified data, the N3C enclave currently has a lack of granular vital sign data, specific information about medication dosage, timing, and route data, as well as generalized procedure codes that often resulted in a loss of granular information about the procedures performed. Of note, respiratory rates and dosage information such as inotrope doses were not available, both of which are important data points for assessing the risk of severe COVID-19. Additionally, vaccination data was not sufficiently complete during the challenge but since then has been improved within N3C. The utilization of real-world data comes with the understanding that models being developed will evolve as additional data sets become available.

Pediatric Respiratory Infectious Diseases

The Pediatric COVID-19 Data Challenge was the first step to develop a clinical decision support tool for a population in which healthcare providers did not have clarity on the subset being affected, lacked vaccination and therapeutic options, and were anticipated to face new surges with confounding respiratory illnesses. This challenge serves as a framework for how researchers with varied backgrounds and large data repositories can be brought together, under the governmental oversight of a panel of technical reviewers, to develop healthcare solutions to emerging health crises. The lessons learned and barriers overcome can serve to accelerate future responses to deliver impactful pediatric clinical models for future infectious diseases.

HHS ASPR BARDA, in collaboration with Sage Bionetworks, NIH's NCATS and NICHD as well as HRSA MCHB, sponsored and facilitated a community challenge to develop prediction models to address the pediatric pandemic surge and MIS-C. The tasks centered around the severity of COVID-19 disease but can be envisioned to any respiratory infectious disease. Additionally, one of the difficulties that predictive computational models face when developing robust models for the pediatric population is the lack of available and quality data. N3C represents one of the largest centralized repositories of pediatric COVID-19 cases in the U.S. and it continues to add additional data and sites making it ideal to assess future infectious diseases as well.

Future Direction

The Pediatric COVID-19 Data Challenge successfully executed a framework to launch, develop and evaluate robust models that have the potential to be used on a nationwide scale and in an evolving landscape. The challenge not only outlined a framework to produce robust computational models that assess children for

risk of COVID-19 severe outcomes but the tasks in N3C can be adapted for future pediatric respiratory infectious diseases as well. Significant work is still needed to adapt robust computational models at scale, in complex healthcare environments, and for everyday decision-making. In order to develop a more end-to-end solution for national pediatric respiratory infection severity triage (1) further validation of winning models using the limited data set is needed to look at accuracy over time, to assess indicators of severity in newborn and infant patients, and to determine the impact of social determinants of health on risk of COVID-19 severe outcomes; (2) additional development of a workflow to disseminate key clinical features, insights, and models back to N3C sites is needed for further education, refinement, and clinical validation; (3) additional validation of external data sets linked to N3C need to be explored to bring additional clinical insights, and (4) additional work with government pediatric centers of excellence and networks are needed to evaluate computational models outside of N3C. As government agencies support the development of medical countermeasures (MCMs), such as diagnostics, therapeutics, and vaccines, during a pandemic response, computational models that inform on utility of critical interventions or can be utilized to help manage patients during the disease course can also be developed to help inform healthcare providers and U.S. Government on opportunities to improve patient care.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/cts.2023.549>.

Acknowledgments. We would like to thank the following individuals for their contributions to challenge development and evaluation:

Timothy Buchman, Robert Tamburro, Richard Gorman, and Cheryl Pikora for their clinical expertise.

Kenneth Gersing, Leonie Misquitta, Carlos Arguello, Carlisle Runge, George Plopper, Jason Gilder, Michael Mutty, Allison Harrill, Rui Li, and Allison Kolbe for their scientific, government partnership, and data expertise.

Amin Manna, Kate Bradwell, Saad Ljazouli, and Emily Niehaus for their technical support and enclave expertise.

The analyses described in this publication were conducted with data or tools accessed through the NCATS N3C Data Enclave covid.cd2h.org/enclave and supported by CD2H – The National COVID Cohort Collaborative IDEA CTR Collaboration 3U24TR002306-04S2 NCATS U24 TR002306. This research was possible because of the patients whose information is included within the data from participating organizations (covid.cd2h.org/dtas) and the organizations and scientists (covid.cd2h.org/duas) who have contributed to the ongoing development of this community resource (cite this <https://doi.org/10.1093/jamia/ocaa196>).

We gratefully acknowledge the following core contributors to N3C:

Adam B. Wilcox, Adam M. Lee, Alexis Graves, Alfred (Jerrold) Anzalone, Amin Manna, Amit Saha, Amy Olex, Andrea Zhou, Andrew E. Williams, Andrew Southerland, Andrew T. Girvin, Anita Walden, Anjali A. Sharathkumar, Benjamin Amor, Benjamin Bates, Brian Hendricks, Brijesh Patel, Caleb Alexander, Carolyn Bramante, Cavin Ward-Caviness, Charisse Madlock-Brown, Christine Suver, Christopher Chute, Christopher Dillon, Chunlei Wu, Clare Schmitt, Cliff Takemoto, Dan Housman, Davera Gabriel, David A. Eichmann, Diego Mazzotti, Don Brown, Eilis Boudreau, Elaine Hill, Elizabeth Zampino, Emily Carlson Marti, Emily R. Pfaff, Evan French, Farrukh M Koraishy, Federico Mariona, Fred Prior, George Sokos, Greg Martin, Harold Lehmann, Heidi Spratt, Hemalkumar Mehta, Hongfang Liu, Hythem Sidky, J.W. Awori Hayanga, Jami Pincavitch, Jaylyn Clark, Jeremy Richard Harper, Jessica Islam, Jin Ge, Joel Gagnier, Joel H. Saltz, Joel Saltz, Johanna Loomba, John Buse, Jomol Mathew, Joni L. Rutter, Julie A. McMurry, Justin Guinney, Justin Starren, Karen Crowley, Katie Rebecca Bradwell, Kellie M. Walters, Ken Wilkins, Kenneth R. Gersing, Kenrick Dwain Cato, Kimberly Murray, Kristin Kostka, Lavance Northington, Lee Allan Pyles, Leonie Misquitta, Lesley Cottrell, Lili Portilla, Mariam Deacy, Mark M. Bissell, Marshall Clark,

Mary Emmett, Mary Morrison Saltz, Matvey B. Palchuk, Melissa A. Haendel, Meredith Adams, Meredith Temple-O'Connor, Michael G. Kurilla, Michele Morris, Nabeel Qureshi, Nasia Safdar, Nicole Garbarini, Noha Sharafeldin, Ofer Sadan, Patricia A. Francis, Penny Wung Burgoon, Peter Robinson, Philip R.O. Payne, Rafael Fuentes, Randeep Jawa, Rebecca Erwin-Cohen, Rena Patel, Richard A. Moffitt, Richard L. Zhu, Rishi Kamaleswaran, Robert Hurley, Robert T. Miller, Saiju Pyarajan, Sam G. Michael, Samuel Bozzette, Sandeep Mallipattu, Satyanarayana Vedula, Scott Chapman, Shawn T. O'Neil, Soko Setoguchi, Stephanie S. Hong, Steve Johnson, Tellen D. Bennett, Tiffany Callahan, Umith Topaloglu, Usman Sheikh, Valery Gordon, Vignesh Subbian, Warren A. Kibbe, Wendy Hernandez, Will Beasley, Will Cooper, William Hilleagass, Xiaohan Tanner Zhang. Details of contributions available at covid.cd2h.org/core-contributors

We would like to thank the Pediatric COVID-19 Data Challenge Consortium members for their participation in the challenge

Adam Cross¹, Ahmed Said², Alessandro Petrini³, Aliaksandr Hubin⁴, Alyssa Columbus⁵, Anamika Kumari⁶, Andrea Bratsberg⁴, Anjali Sharathkumar⁷, Ankit Shukla⁸, Anna Muldoon⁸, Annette Möller⁹, Arnoldo Frigessi⁴, Asbjorn Westvik⁴, Ashwin Dhakal¹⁰, Benjamin Orwoll¹¹, Bhagyashree Aras¹², Bing Xue², Blessy Antony¹³, Bowen Li⁵, Brajesh Karna⁸, Brandon Jernigan¹⁴, Brennan Gallamoza¹⁵, Brian Erly¹⁶, Casey Grage⁵, Celine Cunen⁴, Chaoqi Yang¹⁷, Chengyin Li¹⁸, Chris Kurtz⁸, Christiane Fuchs⁹, Christophe Lambert¹⁹, Christos Argyropoulos¹⁹, Corina Rueegg⁴, Corneliu Antonescu¹⁴, Daniel Liu²⁰, Daoyi Zhu², Dario Malchiodi³, David Swanson⁴, David Little⁸, Deepika Thacker²¹, Dimitri Perrin²², Dongxiao Zhu¹⁸, Elena Casiraghi³, Elham Soltani Kazemi¹⁰, Elizabeth Jimenez¹⁹, Ellen Kerns²³, Ethan Goan²², Even Moa Myklebust⁴, Feifan Liu⁶, Frimpong Boadu¹⁰, Guantao Zhao²⁴, Hailey Stevens-Macfarlane⁸, Hailey Stevens⁸, Hamed Fayyaz¹⁵, Hamidreza Moradi²⁵, Hamza Mir¹⁹, Hanqing Yang², Harry Snow¹⁹, Heather Ross⁸, Henri Pesonen⁴, Holger Roth²⁶, Hong Yi²⁷, Hong Seo Lim²⁴, Ibtihal Ferwana¹⁷, James Howard⁵, Jerrod Anzalone²³, Jessica Gliozzo³, Jianlin Cheng¹⁰, Jimeng Sun¹⁷, Joerg Heintz¹⁷, Jonas Bauer⁹, Jonathan Bona²⁰, Jordy Oswaldo Rodriguez Rincon⁸, Julian W. Sche⁹, Junyi Gao¹⁷, Kenneth Mckinley²⁶, Kerri Rittschof⁸, Khanh Luong²², Kimberly Robasky²⁷, Lauren Chan²⁸, Lav Varshtey¹⁷, Manuela Zucknick⁴, Marco Notaro³, Marco Mesiti³, Marissa Leblanc⁴, Marius Lingurar²⁶, Martine De Cock¹², Matthew Davis²⁹, Matthew Kearney⁸, Md Mozahar Mottalib¹⁵, Meghana Kamineni⁴, Melody Greer²⁰, Michael Lau⁸, Michael Simeone⁸, Mina Jafarpoor⁸, Mohamed Ghalwash³⁰, Mohammad Arif Ul Alam³¹, Morten Stakkeland⁴, Neda Jalali³², Neel Shah², Nicholas Rand³³, Nicholet Deschine Parkhurst⁸, Nicolas Cutrona¹⁵, Nirup Menon²⁴, Nishanth Prathap⁸, Pablo Meyer³⁰, Patrick Finley³⁴, Pavan Turaga⁸, Peng Qiu²⁴, Pooneh Roshanabriz²⁶, Pragati Chidanand Patil¹², Prithwish Chakraborty³⁰, Priyal Makwana³⁵, Qingchuan Sun⁹, Rahmatollah Beheshti¹⁵, Raju Rimal⁴, Rakesh Subramani Kaleshwaran²⁴, Raphael Poulain¹⁵, Riccardo De Bin⁴, Richi Nayak²², Rui Maia⁹, Sai Vinnakota²⁴, Sajid Mahmud¹⁰, Sanjoy Dey³⁰, Shawn Walker⁸, Sheetal Shetty⁸, Sophie Thiesbrummel⁹, Steven Erly¹⁶, Supraja Amrutha¹², Suzanne Mccahan²¹, Talia Hernandez⁸, Tamara Orth²², Thomas Woolf⁵, Tiffany Callahan³⁶, Tim Bunnell²¹, Timothy Lant⁸, Timothy Chappell²², Valeria Vitelli⁴, Vignesh Subbian¹⁴, Waldir Leoncio⁴, Wei Hai Deng⁴, Xinling Li²⁴, Xinyang Liu²⁶, Yao Qiang¹⁸, Yu Shao³⁷, Yuqi Lei², Zhifan Jiang²⁶, Ziqi Xu²

¹University of Illinois at Chicago, ²Washington University in St Louis, ³Università degli Studi di Milano, ⁴Oslo University Hospital, ⁵Johns Hopkins University, ⁶University of Massachusetts Medical School, ⁷University of Iowa, ⁸Arizona State University, ⁹Bielefeld University, ¹⁰University of Missouri, ¹¹Oregon Health & Science University, ¹²University of Washington, ¹³Virginia Tech, ¹⁴University of Arizona, ¹⁵University of Delaware, ¹⁶Salish Research Group, ¹⁷University of Illinois at Urbana Champaign, ¹⁸Wayne State University, ¹⁹University of New Mexico, ²⁰University of Arkansas for Medical Sciences, ²¹Nemours, ²²Queensland University of Technology, ²³University of Nebraska Medical Center, ²⁴Georgia Institute of Technology, ²⁵University of Mississippi Medical Center, ²⁶Children's National Health System, ²⁷University of North Carolina at Chapel Hill, ²⁸Oregon State University, ²⁹Medical University of South Carolina, ³⁰IBM, ³¹University of Maryland, Baltimore County, ³²University of Florida, ³³University of Cincinnati, ³⁴Sandia National Laboratories, ³⁵West Virginia University, ³⁶University of Colorado Anschutz Medical Campus, ³⁷Harvard University

Funding statement. Some authors (Marie Wax, Hui-Hsing Wong) are full-time or contractual employees of the United States Federal Government,

Department of Health and Human Services, the funding agency. Timothy Bergquist, Tellen Bennett, and Richard Moffitt disclosed that the work performed by Sage Bionetworks and its subcontractors under a grant with the National Institute of Health (U24TR002306). Additional funding for Timothy Bergquist was provided through the Bill and Melinda Gates Foundation (INV-018455).

Competing interests. The views expressed are solely those of the authors and do not necessarily represent those of the U.S. Department of Health and Human Services. Timothy Bergquist, Tellen Bennett, and Richard Moffitt disclosed that this work was performed by Sage Bionetworks and its subcontractors under a grant with the National Institute of Health (U24TR002306). Additional funding for Timothy Bergquist was provided through the Bill and Melinda Gates Foundation (INV-018455). Marie Wax and Hui-Hsing Wong disclose that they are government support contractors employed by Aveshka Inc. and Tunnell Government Services Inc. respectively, which receives funds from the U.S. government under contract to provide technical and programmatic support for HHS-BARDA. Joy Almagr discloses that he is a founder and a shareholder of ARIScience. Tellen Bennett has received funding from the National Institutes of Health – NCATS, Eunice Kennedy Shriver NICHD, and National Heart, Lung and Blood Institute (NHLBI).

References

1. Kristie E.N. Seroprevalence of Infection-Induced SARS-CoV-2 Antibodies — United States, September 2021–February 2022. <https://www.cdc.gov/mmwr/volumes/71/wr/mm7117e3.htm>. Accessed April 29, 2022.
2. Children and COVID-19: State-Level Data Report. American Academy of Pediatrics. <https://www.aap.org/en/pages/2019-novel-coronavirus-covid-19-infections/children-and-covid-19-state-level-data-report/>. Accessed December 29, 2022.
3. Delahoy MJ, Ujamaa D, Taylor CA, et al. Comparison of influenza and COVID-19-associated hospitalizations among children < 18 years old in the United States-FluSurv-NET (October-April 2017-2021) and COVID-NET (October 2020-September 2021). *Clin Infect Dis: An Offic Pub Infect Dis Soc Am* 2023; **76**(3): e450–e459.
4. Feldstein LR, Rose EB, Horwitz SM, et al. Multisystem inflammatory syndrome in U.S. children and adolescents. *N Engl J Med*. 2020;**383**(4): 334–346.
5. Feldstein LR, Tenforde MW, Friedman KG, et al. Characteristics and outcomes of US children and adolescents with multisystem inflammatory syndrome in children (MIS-C) compared with severe acute COVID-19. *JAMA*. 2021;**325**(11):1074–1087.
6. Forrest CB, Burrows EK, Mejias A, et al. Severity of acute COVID-19 in children <18 years old March 2020 to December 2021. *Pediatrics*. 2022;**149**(4): e2021055765.
7. Martin B, DeWitt PE, Russell S, et al. Characteristics, outcomes, and severity risk factors associated with SARS-CoV-2 infection among children in the US national COVID cohort collaborative. *JAMA Netw Open*. 2022;**5**(2):e2143151.
8. Martin B, DeWitt PE, Russell S, et al. Acute upper airway disease in children with the omicron (B.1.1.529) variant of SARS-CoV-2-a report from the US national COVID cohort collaborative. *JAMA Pediatr*. 2022;**15**(8):819. doi: [10.1001/jamapediatrics.2022.1110](https://doi.org/10.1001/jamapediatrics.2022.1110).
9. Leyenaar JK, Ralston SL, Shieh MS, Pekow PS, Mangione-Smith R, Lindenaer PK. Epidemiology of pediatric hospitalizations at general hospitals and freestanding children's hospitals in the United States. *J Hosp Med*. 2016;**11**(11):743–749. doi: [10.1002/jhm.2624](https://doi.org/10.1002/jhm.2624).
10. Haendel MA, Chute CG, Bennett TD, et al. The N3C consortium, the national COVID cohort collaborative (N3C): rationale, design, infrastructure, and deployment. *J Am Med Inform Assoc*. 2021;**28**(3):427–443. doi: [10.1093/jamia/ocaa196](https://doi.org/10.1093/jamia/ocaa196).
11. N3C Data Enclave Tools. National COVID Cohort Collaborative. <https://covid.cd2h.org/tools>. Accessed 5 January 2023.
12. National COVID Cohort Collaborative (N3C). National COVID Cohort Collaborative. <https://zenodo.org/communities/cd2h-covid/?page=1&size=20>. Accessed 5 January 2023.