

ARTICLE OPEN



International comparisons of laboratory values from the 4CE collaborative to predict COVID-19 mortality

Griffin M. Weber^{1,152}, Chuan Hong^{1,2,152}, Zongqi Xia^{3,152}, Nathan P. Palmer¹, Paul Avillach¹, Sehi L'Yi¹, Mark S. Keller¹, Shawn N. Murphy⁴, Alba Gutiérrez-Sacristán¹, Clara-Lea Bonzel¹, Arnaud Serret-Larmande⁵, Antoine Neuraz⁶, Gilbert S. Omenn⁷, Shyam Visweswaran⁸, Jeffrey G. Klann⁹, Andrew M. South¹⁰, Ne Hooi Will Loh¹¹, Mario Cannataro¹², Brett K. Beaulieu-Jones¹, Riccardo Bellazzi¹³, Giuseppe Agapito¹⁴, Mario Alessiani¹⁵, Bruce J. Aronow¹⁶, Douglas S. Bell¹⁷, Vincent Benoit¹⁸, Florence T. Bourgeois¹⁹, Luca Chiovato²⁰, Kelly Cho²¹, Arianna Dagliati²², Scott L. DuVall²³, Noelia García Barrio²⁴, David A. Hanauer²⁵, Yuk-Lam Ho²¹, John H. Holmes^{26,27}, Richard W. Issitt²⁸, Molei Liu²⁹, Yuan Luo³⁰, Kristine E. Lynch²³, Sarah E. Maidlow³¹, Alberto Malovini³², Kenneth D. Mandl³³, Chengsheng Mao³⁰, Michael E. Matheny³⁴, Jason H. Moore²⁷, Jeffrey S. Morris³⁵, Michele Morris⁸, Danielle L. Mowery²⁶, Kee Yuan Ngiam³⁶, Lav P. Patel³⁷, Miguel Pedrera-Jimenez²⁴, Rachel B. Ramoni³⁸, Emily R. Schriver³⁹, Petra Schubert²¹, Pablo Serrano Balazote²⁴, Anastasia Spiridou²⁸, Amelia L. M. Tan¹, Byorn W. L. Tan⁴⁰, Valentina Tibollo³², Carlo Torti⁴¹, Enrico M. Trearichi⁴¹, Xuan Wang¹, The Consortium for Clinical Characterization of COVID-19 by EHR (4CE)*, Isaac S. Kohane¹, Tianxi Cai¹✉ and Gabriel A. Brat¹✉

Given the growing number of prediction algorithms developed to predict COVID-19 mortality, we evaluated the transportability of a mortality prediction algorithm using a multi-national network of healthcare systems. We predicted COVID-19 mortality using baseline commonly measured laboratory values and standard demographic and clinical covariates across healthcare systems, countries, and continents. Specifically, we trained a Cox regression model with nine measured laboratory test values, standard demographics at admission, and comorbidity burden pre-admission. These models were compared at site, country, and continent level. Of the 39,969 hospitalized patients with COVID-19 (68.6% male), 5717 (14.3%) died. In the Cox model, age, albumin, AST, creatine, CRP, and white blood cell count are most predictive of mortality. The baseline covariates are more predictive of mortality during the early days of COVID-19 hospitalization. Models trained at healthcare systems with larger cohort size largely retain good transportability performance when porting to different sites. The combination of routine laboratory test values at admission along with basic demographic features can predict mortality in patients hospitalized with COVID-19. Importantly, this potentially deployable model differs from prior work by demonstrating not only consistent performance but also reliable transportability across healthcare systems in the US and Europe, highlighting the generalizability of this model and the overall approach.

npj Digital Medicine (2022)5:74 | <https://doi.org/10.1038/s41746-022-00601-0>

¹Department of Biomedical Informatics, Harvard Medical School, Boston, USA. ²Department of Biostatistics and Bioinformatics, Duke University, Durham, USA. ³Department of Neurology, University of Pittsburgh, Pittsburgh, USA. ⁴Department of Neurology, Massachusetts General Hospital, Boston, USA. ⁵Department of biomedical informatics, Hôpital Européen Georges Pompidou, Assistance Publique - Hôpitaux de Paris, Paris, France. ⁶Department of biomedical informatics, Hôpital Necker-Enfants Malade, Assistance Publique Hôpitaux de Paris (APHP), University of Paris, Paris, France. ⁷Department of Computational Medicine & Bioinformatics, Internal Medicine, Human Genetics, and School of Public Health, University of Michigan, Ann Arbor, USA. ⁸Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, USA. ⁹Department of Medicine, Massachusetts General Hospital, Boston, USA. ¹⁰Department of Pediatrics-Section of Nephrology, Brenner Children's Hospital, Wake Forest School of Medicine, Winston Salem, USA. ¹¹Department of Anaesthesia, National University Health System, Singapore, Singapore. ¹²Department of Medical and Surgical Sciences, Data Analytics Research Center, University Magna Graecia of Catanzaro, Italy, Catanzaro, Italy. ¹³Department of Electrical, Computer and Biomedical Engineering, University of Pavia, Italy, Pavia, Italy. ¹⁴Department of Legal, Economic and Social Sciences, University Magna Graecia of Catanzaro, Italy, Catanzaro, Italy. ¹⁵Department of Surgery, ASST Pavia, Lombardia Region Health System, Pavia, Italy. ¹⁶Departments of Biomedical Informatics, Pediatrics, Cincinnati Children's Hospital Medical Center, University of Cincinnati, Cincinnati, USA. ¹⁷Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, USA. ¹⁸IT department, Innovation & Data, APHP Greater Paris University Hospital, Paris, France. ¹⁹Department of Pediatrics, Harvard Medical School, Boston, USA. ²⁰Unit of Internal Medicine and Endocrinology, Istituti Clinici Scientifici Maugeri SpA SB IRCCS, Pavia, Italy. ²¹Massachusetts Veterans Epidemiology Research and Information Center (MAVERIC), VA Boston Healthcare System, Boston, USA. ²²Department of Electrical Computer and Biomedical Engineering, University of Pavia, Italy, Pavia, Italy. ²³VA Informatics and Computing Infrastructure, VA Salt Lake City Health Care System, Salt Lake City, USA. ²⁴Health Informatics, Hospital Universitario 12 de Octubre, Madrid, Spain. ²⁵Department of Learning Health Sciences, University of Michigan, Ann Arbor, USA. ²⁶Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, USA. ²⁷Institute for Biomedical Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, USA. ²⁸Digital Research, Informatics and Virtual Environments (DRIVE), Great Ormond Street Hospital for Children, UK, London, UK. ²⁹Department of Biostatistics, Harvard School of Public Health, Boston, USA. ³⁰Department of Preventive Medicine, Northwestern University, Chicago, USA. ³¹Michigan Institute for Clinical and Health Research, University of Michigan, Ann Arbor, USA. ³²Laboratory of Informatics and Systems Engineering for Clinical Research, Istituti Clinici Scientifici Maugeri SpA SB IRCCS, Pavia, Italy. ³³Computational Health Informatics Program, Boston Children's Hospital, Boston, USA. ³⁴VA Informatics and Computing Infrastructure, Tennessee Valley Healthcare System Veterans Affairs Medical Center, Nashville, USA. ³⁵Department of Biostatistics, Epidemiology, and Biostatistics, University of Pennsylvania Perelman School of Medicine, Philadelphia, USA. ³⁶Department of Biomedical informatics, WISDM, National University Health Systems Singapore, Singapore, Singapore. ³⁷Department of Internal Medicine, Division of Medical Informatics, University of Kansas Medical Center, Kansas City, USA. ³⁸Office of Research and Development, Department of Veterans Affairs, Washington, DC, USA. ³⁹Data Analytics Center, University of Pennsylvania Health System, Philadelphia, USA. ⁴⁰Department of Medicine, National University Hospital, Singapore, Singapore, Singapore. ⁴¹Department of Medical and Surgical Sciences, Infectious and Tropical Disease Unit, University Magna Graecia of Catanzaro, Italy, Catanzaro, Italy. ¹⁵²These authors contributed equally: Griffin M. Weber, Chuan Hong, Zongqi Xia. *A list of authors and their affiliations appears at the end of the paper.

✉email: tcgai@hsph.harvard.edu; gabriel_brat@hms.harvard.edu

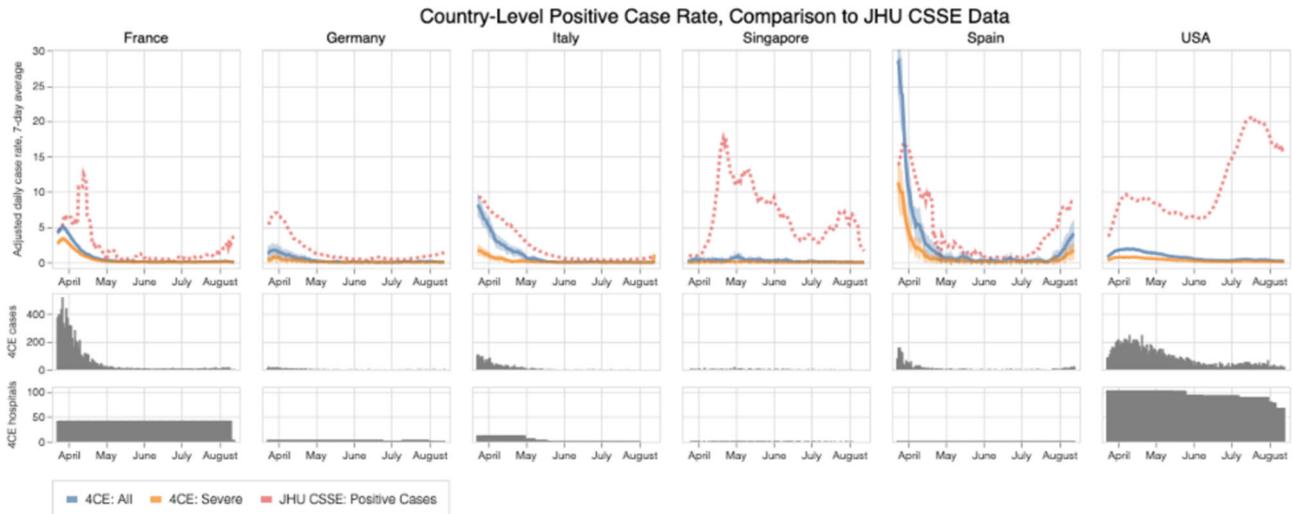


Fig. 1 Comparison of National Hospitalization Rates by Data Source. Adjusted 7-day average new hospitalization rate and rate of ever-severe disease per 100,000 people by country based on 4CE contributors along with 95% confidence intervals compared with 7-day average new case rates collected by Johns Hopkins Center for Systems Science and Engineering (JHU CSSE).

INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has caused millions of cases of coronavirus disease 2019 (COVID-19) in nearly every country. While most patients with COVID-19 have a mild form of viral pneumonia, an appreciable subgroup develops rapid onset of severe disease. Several large national studies have demonstrated that a variable and potentially significant proportion (ranging from 5% to 70%)^{1–3} of hospitalized patients with COVID-19 develop cardiorespiratory failure, require mechanical ventilation and hemodynamic support, and may ultimately die. The early identification of patients at high risk for death can improve triage and resource allocation, particularly when numbers of COVID-19 cases overwhelm health systems⁴.

Numerous studies have reported models using clinical data, including laboratory values, to predict patients at high risk of death for COVID-19². However, most models have not been tested across hospital systems and countries to determine generalizability. Few studies have included patients from multi-national cohorts. The international nature of this disease begs the question of whether models derived using data from one site or one country can be used in another. Is transportability possible if the experience of one site or country could help another make better decisions?

We formed the 4CE Consortium⁵ as an international research collaborative of nearly 300 hospitals from four countries in order to collect standardized patient-level electronic health record (EHR) data to examine the epidemiology, pathophysiology, management, and healthcare system dynamics of COVID-19. Using the 4CE data, we examined the relationship between pre-selected laboratory values⁶ and mortality across institutions and countries. We compared prediction models using single laboratory values at admission to a prediction model containing multiple laboratory values. Across all models, we evaluated geographical differences (national and continental) among the outcome prediction models to better understand if models trained on data from one country and institution can be used elsewhere.

RESULTS

Characteristics of the study population

In this study population of 39,969 patients, the incidence of hospitalization for COVID-19 largely tracked with population dynamics of COVID-19 cases⁷ across different countries during

the initial pandemic period (Fig. 1). Both the COVID-19 case rate and the COVID-19 hospitalization rate dropped significantly from the first peak in April 2020. While hospitalization rates remained relatively low for all countries, case rates increased in France, Germany, Spain and United States after June 2020.

Consistent with prior studies^{4,8}, the study population of patients hospitalized with COVID-19 showed a higher prevalence of men and older populations. See Supplementary Fig. 1 for demographic characteristics and percentages among age group, race/ethnicity, and sex. International comparisons were consistent and showed across three countries that most patients (79.6%) were 50 years of age or older and male (68.6%).

International comparisons of individual laboratory tests at admission for mortality risk prediction

The prediction performances of individual laboratory test across all sites, at country level and continent level were summarized using random-effects meta-analysis. On average, albumin, creatinine, neutrophil count, CRP and white blood cell were stronger predictors of mortality than the other labs (Supplementary Fig. 2). The predictiveness of the laboratory tests for mortality within the next few days after admission tends to be slightly higher than for 1 or 2-week mortality although the decrease in predictiveness over time was moderate. The predictiveness of the labs varies substantially across sites. Albumin has low predictiveness in European sites but higher in the US, CRP appears to be slightly more predictive in Europe than in US, while other labs performed similarly in the US and in Europe on average.

International comparisons of mortality risk prediction model

The estimated log hazard ratios for demographic, nine laboratory tests and Charlson comorbidity index from a comprehensive Cox model are largely consistent across different healthcare systems with respect to their directions and magnitudes (Supplementary Fig. 3). The estimated log hazard ratios across all sites and at country level were summarized using random-effects meta-analysis. The risk models indicate that age, albumin, AST, creatine, CRP, and white blood cell are most predictive of mortality. For example, the risk model predicts a protective effect against mortality from those who are <50 years old, report higher albumin values and lymphocyte count values, and report lower AST, creatinine and CRP values. The average AUC of the full risk model is about 0.80, 0.79 and 0.77 for predicting both 3-day, 1-week, or

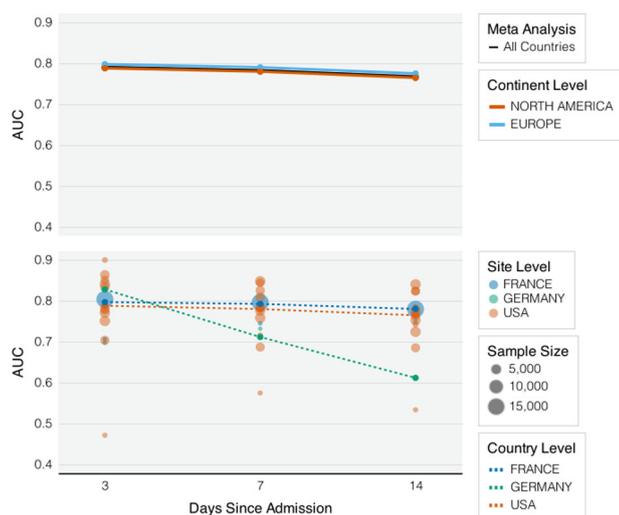


Fig. 2 Risk Model Performance Across Countries and Continents. AUCs of cox regression models with nine common laboratory tests (missing rate <30%) in predicting death adjusting for demographic variables and Charlson comorbidity index.

2-week mortality (Fig. 2). While the performance of the locally trained site-level models varies across healthcare systems, the average performance of the full model is similar in the US versus Europe.

Portability of mortality algorithms across sites, countries, and continents

The AUCs of the locally trained mortality risk models for 1-week mortality when porting to external sites were summarized in Fig. 3 (refer to Supplementary Table 4 for numerical results). The averaged AUCs across all sites and at country level were summarized using random-effects meta-analysis. The algorithms trained from sites with large cohort size tend to have better performance both locally and when transported to other sites. For example, the AUCs of the model trained at SITE1 (France) are always close to or higher than those of the local trained model. We additionally compared the portability performance across continents. In general, when porting to North America sites, the algorithms trained at both continents perform equally well. For example, when porting to SITE5 (US), the maximum AUC was 0.842 and 0.847 for algorithms trained at North America sites and at European sites, respectively, which are very close to the maximum AUC of the local SITE5 algorithm. On the other hand, when porting to Europe sites, the algorithms trained at North America sites perform slightly better than those trained at Europe sites, due to the relatively smaller sample size of the Europe sites. For example, when porting to SITE1 (France), the maximum AUC was 0.813 and 0.791 for algorithms trained at North America sites and at European sites, respectively.

DISCUSSION

In this large-scale multi-national study, we reported a mortality prediction model for patients hospitalized with COVID-19 that retained accuracy across healthcare systems and countries. Building on the growing literature of COVID-19 mortality prediction, our study is unique in leveraging international cohorts to validate the generalizability of the prediction model, which has the following specific features. First, a predictive model containing nine commonly measured laboratory test values performed better than the model containing 17 laboratory test values: CRP, creatinine, white blood cell count, lymphocyte count, AST, ALT,

total bilirubin, neutrophil count, and albumin. From a list of 17 laboratory tests associated with worse outcomes in patients with COVID-19 based on prior reports⁶, we selected the subset of nine tests based on their low rate of missing data in our data set. Second, we identified albumin, CRP, creatinine, neutrophil count, and white blood cell count as better individual predictors than other individual laboratory tests. Third, a comprehensive model containing the nine commonly measured laboratory tests as well as baseline demographic features and comorbidity burden indicates that age, albumin, AST, CRP, creatine, and white blood cell count are most predictive of mortality. Interestingly, the baseline covariates are more predictive of mortality in the early days after admission for COVID-19, likely because other features gain importance as hospital course prolongs. Finally, when comparing prediction models between North American and European sites, the final model showed crucial consistency across international sites, highlighting its potential generalizable application.

The study has several strengths. Chief among them is the international consortium with a federated data sharing approach that facilitated the pooling of laboratory values across 283 hospitals with diverse healthcare practices and populations, enabling the examination of model transportability. Second, while the accuracy (AUC) of individual laboratory test in predicting mortality after hospital admission for COVID-19 varies substantially cross countries, the accuracy of the mortality risk prediction model is remarkably consistent between US and Europe. Further, the estimated log hazard ratios from the best-performing Cox model are largely consistent across different healthcare systems with respect to their directions and magnitudes. Third, the mortality prediction model using commonly measured laboratory tests and baseline demographic and comorbidity burden trained at healthcare systems performs well both locally and externally when transported to other sites. Interestingly, the transportability does not appear to depend on the continent or country. Taken together, the key innovation of our study that differs from prior studies is the transportability and the potential generalizability of the COVID-19 mortality prediction model that seems independent of the specific healthcare system.

The study also has several limitations that we took measures to mitigate. First, EHR data have variable degree of intrinsic noise, missing data, and available documentation due to differences in clinical practice that contribute to differences among healthcare systems. Indeed, we found healthcare system-level (within-healthcare system and between-healthcare system) differences were greater than country-level differences. By leveraging our federated system of common EHR data elements and capturing healthcare system-level heterogeneity, the 4CE consortium is uniquely positioned to identify international differences in patient characteristics and outcomes as well as to test model transportability. To mitigate the quality issue of EHR data, we performed extensive and iterative quality controls at *each* participating healthcare system with local collaborators and centrally to address potential imprecision due to healthcare system-specific variations in data extraction and incompleteness of datasets (e.g., incomplete mapping of local EHR codes to desired data elements). These critical quality control steps, which are often underappreciated in multi-center EHR data research, further differentiate the 4CE research efforts from other COVID-19 research efforts. Second, we observed a significant level of heterogeneity in the predictiveness of individual laboratory tests and the locally trained mortality risk models across the participating healthcare systems. The heterogeneity could result from differences in patient population, clinical practice and EHR system. To address this concern, we performed random-effects meta-analyses to account for the heterogeneity across sites. Importantly, the best-performing model showed evidence of good transportability despite of the heterogeneity.



Fig. 3 Transportability of the Mortality Prediction Model Across Sites and Countries. Heatmap of transportability of the Cox regression model across different sites and countries. Each part of the figure represents performance when the model is trained at one site and evaluated at another.

As the pandemic persists and new SARS-CoV-2 variants emerge, two clinically relevant questions remain unanswered: (1) does the mortality prediction model continue to perform well across healthcare systems and countries? (2) can the prediction model predict long-term mortality after COVID-19 hospitalization? To address these questions, we are planning future analyses using patient-level data at each participating healthcare system to assess the temporal trends of the model performance throughout the pandemic waves and at individual patient-level over longer period. We will revise and adapt to temporal changes in clinical scenarios. In this study, we observed that AUCs are generally consistent across genders. Since age is a significant risk factor for mortality, conditioning on the age group, the model performance for distinguishing high-risk vs. low-risk patients within the age group is expected to be lower than the overall accuracy. Further developing age-specific risk prediction models warrants further research. Beyond mortality prediction, the 4CE consortium has established a platform of harmonized data capture through its federated system with iterative and methodical expansion of data elements to enable the clinical investigation of a wide range of domains pertaining to COVID-19 such as coagulopathy and

thrombotic events, acute renal failure, pediatric manifestation, neurological complications as well as the post-acute sequelae syndrome (i.e., long-hauler). We will apply the approach from this study to assess other prediction model transportability within our international network of participating healthcare systems.

We make several noteworthy observations of clinical relevance. First, the laboratory tests predictive of mortality in patients hospitalized for COVID-19 represent the combination of acute inflammatory response (as indicated by CRP, white blood cell, lymphocyte, and neutrophil count) and underlying physiological function as well as the acute response of critical organ systems (general nutritional status as indicated by albumin, renal function as indicated by creatinine, and hepatic function as indicated by AST, ALT, and bilirubin). These routinely collected laboratory indicators of systemic response to the SARS-CoV-2 viral infection in conjunction with easily ascertainable baseline demographic and comorbidity burden formulate a clinically deployable prediction tool of mortality risk following hospital admission for COVID-19. Second, the relatively modest accuracy of individual laboratory values in predicting mortality is likely due to its large variation within each participating healthcare system. This combination of

commonly measured clinical laboratory tests dramatically improved the prediction performance over individual laboratory tests, and performed better than a larger panel of clinical laboratory tests. A key clinical insight is that clinical laboratory tests beyond the commonly measured routine tests may not inform mortality, which is the most important clinical outcome. Third, the performance of the final model was relatively stable over the hospital course and did not improve beyond the initial hospital days. This finding suggests that additional factors contribute to mortality as the hospital course for COVID-19 patients prolongs. Of particular clinical relevance, it supports the utility of commonly measured routine clinical laboratory test values (and other routine clinical and demographic features) at admission to identify patients at high risk for mortality who would warrant early and aggressive intervention as well as close monitoring, particularly in the setting of limited healthcare resources.

METHODS

Cohort identification

We included all patients hospitalized at participating 4CE sites with an admission date from 7 days before to 14 days after the date of their first reverse transcription polymerase chain reaction (PCR)-confirmed SARS-CoV-2 positive test result. The first admission date within this 21-day time window was considered the index admission date. Throughout this work, “days since admission” refers to this index date.

Participating sites

Data were available from 39,969 patients from 284 hospitals (affiliated with 16 sites) across four countries: France, Germany, Spain, and the United States. See Supplementary Table 2 for details about participating sites. Several sites collected data from multiple hospitals. In the United States, 170 medical centers of the US Department of Veterans Affairs were grouped into five regional divisions called Veterans Integrated Service Networks.

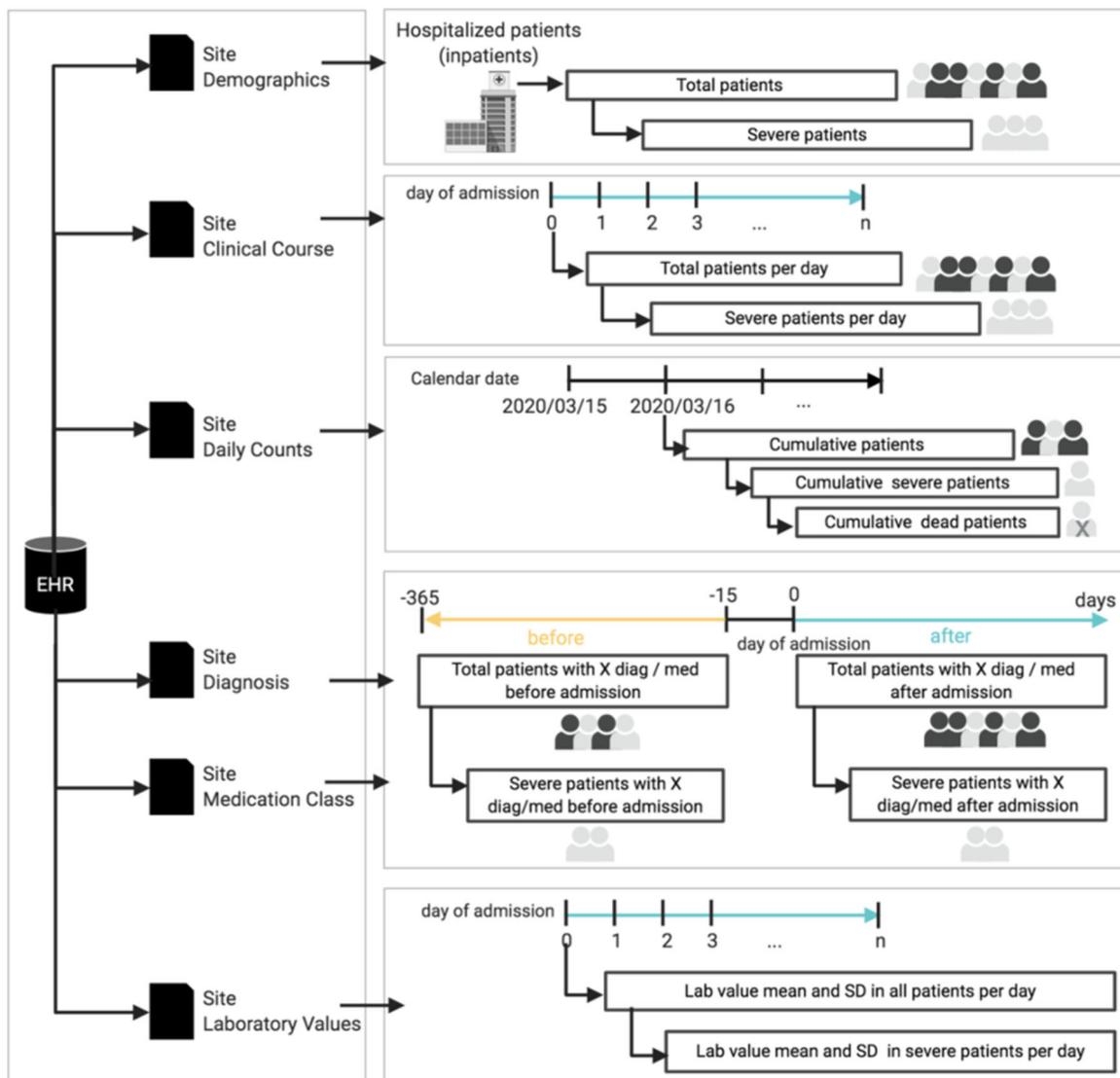


Fig. 4 Schematic of the federated EHR-based study involving healthcare systems from three countries. Each site generated three data tables (comma-separated files) containing patient level data: 1) local patient clinical course indicates which days the patient was in the hospital and when the patient died; 2) local patient observation includes first three-character ICD9/10 diagnosis code and laboratory tests, where laboratory test has a numerical value; 3) local patient summary contains demographic variables including age, sex and race. Sites then conduct analysis using these individual level data within their firewall (see Methods).

Patient and public involvement

Patients and the public were not involved in the design, conduct, or reporting, or dissemination plans of the research.

Outcome

We consider death as the main COVID-19 outcome. Death was identified via standard coding and discharge data aggregation from each site. Each partner institution used local criteria to identify in-hospital mortality.

Local data collection

Patient-level data. Sixteen sites representing 284 Hospitals assembled patient-level data for detailed analyses, including twelve US sites, and four international sites. Individual healthcare systems then ran separate analyses using the patient-level data within their local firewall and only reported the final analytic results to the central institution for meta-analysis. A schematic of our workflow is presented in Fig. 4, and further details of collected data are reported in Supplementary Table 3.

Software platform. Most sites used the open source i2b2 (Informatics for Integrating Biology and the Bedside) software platform to obtain the data. More than 200 organizations worldwide use i2b2 for purposes that include identifying participants for clinical trials, drug safety monitoring, and clinical and epidemiological research. Those 4CE sites with i2b2 used database scripts to directly query their i2b2 repository, calculate the counts and statistics, and export the data files. The 4CE sites without i2b2 used the Observational Medical Outcomes Partnership (OMOP) Common Data Model or their own clinical data warehouse solutions (e.g., Epic Caboodle) and querying tools to create the required files.

Selection of laboratory tests

We focused on nine laboratory tests that are commonly measured (missing rate <30% at most sites) and associated with mortality in patients with COVID-19 based on prior reports⁶. We provided each site with a single standard Logical Objects, Identifiers, Names and Codes (LOINC) identifier for each test, but sites often needed to map tests to additional LOINC or custom codes within their EHR. We addressed barriers that arose during initial efforts to extract these laboratory values by stratifying region-specific laboratory test types to reduce extraction errors and enable standardization.

Quality control

We conducted site-specific quality control. Each site ran an R script for the following additional quality control checks: consistency of the total counts of total cases across all datasets within each site, consistency between the 3-digit diagnosis codes and the ICD dictionary, and consistency of the range of laboratory data from each site with the normal range observed from all sites. Sites checked and fixed the data if their laboratory values were consistently lower or higher than the other sites or otherwise implausible.

Statistical analysis

We estimated the country-level daily incidence of new patients hospitalized with COVID-19 during the study period from March 1, 2020 to September 30, 2020. Specifically, for each country, we summed the daily incidence of new patients hospitalized with COVID-19 at each site within that country per 100,000 people of the country and multiplied this by an adjustment factor, defined as the ratio between the country's overall inpatient discharge rate and the overall inpatient discharge rate of all 4CE sites in that country irrespective of COVID-19 status. We then reported the adjusted 7-day average incidence of new COVID-19 hospitalizations per 100,000 of the country population.

We divided our analysis into two parts: (1) prediction of mortality using individual laboratory values and a comprehensive algorithm derived from multiple laboratory values, comorbid conditions, and demographics available at each site and (2) comparison of these models across sites, countries, and continents.

We evaluated the ability of a biomarker and demographics-based algorithm to predict mortality using admission data. We removed patients who died at admission. We developed mortality risk prediction models using a set of nine common laboratory tests with missing rates <30% at most sites, adjusting for demographic variables and the Charlson

comorbidity index. We derived the risk models by fitting penalized Cox proportional hazards model. We evaluated the accuracy of the risk models for predicting mortality by t -days since admission based on the time-specific AUC⁹. We used the 10-fold cross-validation to estimate the AUC when evaluating the model performance within each local site. The mortality risk prediction model was not trained at Spain because the data were not available at the time when we collected the model training results. To assess the transportability of the mortality risk prediction models across different sites, we validated the algorithm trained at local individual healthcare centers using independent dataset from remaining external sites including the healthcare center from Spain. We used random effects meta-analysis on the prediction performance measures across sites to summarize country level, continent level, and overall average performances.

IRB Approval was obtained at Assistance Publique—Hôpitaux de Paris, Beth Israel Deaconess Medical Center, Bordeaux University Hospital, Hospital Universitario 12 de Octubre, Massachusetts General Brigham, Northwestern University, Medical Center, University of Freiburg, University of Pittsburgh, VA North Atlantic, VA Southwest, VA Midwest, VA Continental, and VA Pacific. An exempt determination was made by the IRB at University of California Los Angeles, University of Michigan, and University of Pennsylvania.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

Only aggregate data was shared by sites for this study. All aggregate data in a de-identified fashion can be found and downloaded at www.covidclinical.net.

CODE AVAILABILITY

The SQL and R scripts used in this work can be found and downloaded at <https://github.com/covidclinical>.

Received: 4 February 2021; Accepted: 11 March 2022;

Published online: 13 June 2022

REFERENCES

1. Wu, Z. & McGoogan, J. M. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. *JAMA* **323**, 1239–1242 (2020).
2. Wynants, L. et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ* **369**, m1328 (2020).
3. Goyal, P. et al. Clinical characteristics of Covid-19 in New York City. *N. Engl. J. Med.* **382**, 2372–2374 (2020).
4. Fried, M. W. et al. Patient characteristics and outcomes of 11,721 patients with COVID19 hospitalized across the United States. *Clin. Infect. Dis.* <https://doi.org/10.1093/cid/cia1268> (2020).
5. Brat, G. A. et al. International electronic health record-derived COVID-19 clinical course profiles: the 4CE consortium. *NPJ Digit. Med.* **3**, 109 (2020).
6. Lippi, G. & Plebani, M. Laboratory abnormalities in patients with COVID-2019 infection. *Clin. Chem. Lab. Med.* <https://doi.org/10.1515/cclm-2020-0198> (2020).
7. *COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)* (Johns Hopkins University (JHU), accessed 7 October 2020); <https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6>.
8. Guan, W. -J. et al. Clinical characteristics of coronavirus disease 2019 in China. *N. Engl. J. Med.* <https://doi.org/10.1056/NEJMoa2002032> (2020).
9. Uno, H., Cai, T., Tian, L. & Wei, L. J. Evaluating prediction rules for t -year survivors with censored regression models. *J. Am. Stat. Assoc.* **102**, 527–537 (2007).

ACKNOWLEDGEMENTS

G.W. reports funding from NCATS UL1TR002541, NCATS UL1TR000005, and NLM R01LM013345. S.M. and J.K. report funding from NCATS 5UL1TR001857-05 and NHGRI 5R01HG009174-04. Z.X. reports funding from NINDS R01NS098023. G.O. reports

funding from NIH grants NIEHS P30ES017885 and NCI U24CA210967. S.V. reports funding from NLM R01LM012095 and NCATS UL1TR001857. A.S. reports funding from NHLBI K23HL148394 and L40HL148910, and NCATS UL1TR001420. B.A. reports funding from NHLBI U24 HL148865. D.B. and R.F. report funding from NCATS UL1TR001881. T.G. and T.G. report funding from 01ZZ1801E German Federal Ministry of Education and Research. D.H. reports funding from NCATS UL1TR002240. M.K. reports funding from NHGRI 5T32HG002295-18. D.K. reports funding from MIRACUM Consortium grant 01ZZ1801A. Y.L. reports funding from NLM R01LM01333. J.M. reports funding from NCATS UL1TR001878. D.M. reports funding from NCATS UL1TR001878 Institutional Clinical and Translational Science Award (University of Pennsylvania). L.P. reports funding from NCATS CTSA Award #UL1TR002366.

AUTHOR CONTRIBUTIONS

G.M.W., C.H., N.P.P., P.A., S.N.M., A.G.S., G.S.O., J.G.K., R.B., M.A., B.J.A., D.S.B., F.T.B., K.C., A.D., J.H.M., I.S.K., T.C., and G.A.B. contributed to design and conceptualization of the study. G.M.W., Z.X., N.P.P., P.A., S.N.M., A.S.L., A.N., S.V., J.G.K., A.M.S., N.H.W.L., M.C., B.K. B.J., R.B., G.A., M.A., D.S.B., V.B., L.C., K.C., A.D., S.L.D., N.G.B., D.A.H., Y.L.H., J.H.H., R.W.I., Y.L., K.E.L., S.E.M., A.M., K.D.M., C.M., M.E.M., J.H.M., J.S.M., M.M., D.L.M., K.Y.N., L.P.P., M. P.J., R.B.R., E.R.S., P.S., P.S.B., A.S., A.L.M.T., B.W.L.T., V.T., C.T., and E.M.T. contributed to data collection. G.M.W., C.H., N.P.P., P.A., S.L., M.S.K., S.N.M., A.G.S., C.L.B., G.S.O., S.V., J. G.K., A.M.S., M.C., B.K.B.J., G.A., M.A., B.J.A., D.S.B., F.T.B., A.D., S.L.D., D.A.H., J.H.H., M.L., Y.L., S.E.M., K.D.M., C.M., M.E.M., J.S.M., M.M., L.P.P., A.L.M.T., C.T., E.M.T., X.W., I.S.K., T.C., and G.A.B. contributed to data analysis and interpretation. All authors contributed to drafting and revision of the manuscript and approved the final manuscript. All authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

COMPETING INTERESTS

The authors declare no competing interests.

THE CONSORTIUM FOR CLINICAL CHARACTERIZATION OF COVID-19 BY EHR (4CE)

James R. Aaron⁴², Giuseppe Agapito¹⁴, Adem Albayrak⁴³, Giuseppe Albi¹³, Mario Alessiani¹⁵, Anna Alloni⁴⁴, Danilo F. Amendola⁴⁵, François Angoulvant⁴⁶, Li L. L. J. Anthony⁴⁷, Bruce J. Aronow¹⁶, Fatima Ashraf⁴⁸, Andrew Atz⁴⁹, Paul Avillach¹, Paula S. Azevedo⁵⁰, James Balshi⁵¹, Brett K. Beaulieu-Jones¹, Douglas S. Bell¹⁷, Antonio Bellasi⁵², Riccardo Bellazzi¹³, Vincent Benoit⁵³, Michele Beraghi⁵⁴, José Luis Bernal-Sobrino²⁴, Mélodie Bernaux⁵⁵, Romain Bey⁵³, Surbhi Bhatnagar¹⁶, Alvar Blanco-Martínez²⁴, Martin Boeker⁵⁶, Clara-Lea Bonzel¹, John Booth²⁸, Silvano Bosari⁵⁷, Florence T. Bourgeois¹⁹, Robert L. Bradford⁵⁸, Gabriel A. Brat¹, Stéphane Bréant¹⁸, Nicholas W. Brown¹, Raffaele Bruno⁵⁹, William A. Bryant²⁸, Mauro Bucalo⁴⁴, Emily Bucholz⁶⁰, Anita Burgun⁶¹, Tianxi Cai¹, Mario Cannataro¹², Aldo Carmona⁶², Charlotte Caucheteux⁶³, Julien Champ⁶⁴, Krista Y. Chen⁶⁵, Jin Chen⁶⁶, Luca Chiovato²⁰, Lorenzo Chiudinelli⁶⁷, Kelly Cho⁶⁸, James J. Cimino⁶⁹, Tiago K. Colicchio⁶⁹, Sylvie Cormont¹⁸, Sébastien Cossin⁷⁰, Jean B. Craig⁷¹, Juan Luis Cruz-Bermúdez²⁴, Jaime Cruz-Rojo²⁴, Arianna Dugliati¹⁸, Mohamad Daniar⁷², Christel Daniel⁷³, Priyam Das¹, Batsal Devkota⁷⁴, Audrey Dionne⁶⁰, Rui Duan⁷⁵, Julien Dubiel¹⁸, Scott L. DuVall²³, Loic Esteve⁷⁶, Hossein Estiri⁷⁷, Shirley Fan⁷⁸, Robert W. Follett¹⁷, Thomas Ganslandt⁷⁹, Noelia García-Barrio²⁴, Lana X. Garmire⁸⁰, Nils Gehlenborg¹, Emily J. Getzen⁸¹, Alon Geva⁸², Tobias Gradinger⁷⁹, Alexandre Gramfort⁶³, Romain Griffier⁷⁰, Nicolas Griffon⁷³, Olivier Grisel⁶³, Alba Gutiérrez-Sacristán¹, Larry Han⁷⁵, David A. Hanauer⁸³, Christian Haverkamp⁸⁴, Derek Y. Hazard⁸⁵, Bing He⁸⁰, Darren W. Henderson⁴², Martin Hilka¹⁸, Yuk-Lam Ho²¹, John H. Holmes^{26,27}, Chuan Hong^{1,2,152}, Kenneth M. Huling¹, Meghan R. Hutch³⁰, Richard W. Issitt²⁸, Anne Sophie Jannot⁸⁶, Vianney Jouhet⁷⁰, Ramakanth Kavuluru⁸⁷, Mark S. Keller¹, Chris J. Kennedy⁸⁸, Kate F. Kernan⁸⁹, Daniel A. Key²⁸, Katie Kirchoff⁹⁰, Jeffrey G. Klann⁷⁷, Isaac S. Kohane¹, Ian D. Krantz⁹¹, Detlef Kraska⁹², Ashok K. Krishnamurthy⁹³, Sehi L'Yi¹, Trang T. Le²⁶, Judith Leblanc⁹⁴, Guillaume Lemaître⁶³, Leslie Lenert⁷¹, Damien Leprovost⁹⁵, Molei Liu⁹⁶, Ne Hooi Will Loh¹¹, Qi Long⁹⁷, Sara Lozano-Zahonero⁹⁸, Yuan Luo³⁰, Kristine E. Lynch²³, Sadiqa Mahmood⁴³, Sarah E. Maidlow⁹⁹, Adeline Makoudjou⁹⁸, Alberto Malovini³², Kenneth D. Mandl³³, Chengsheng Mao³⁰, Anupama Maram¹⁰⁰, Patricia Martel¹⁰¹, Marcelo R. Martins¹⁰², Jayson S. Marwaha¹⁰³, Aaron J. Masino¹⁰⁴, Maria Mazzitelli⁴¹, Arthur Mensch¹⁰⁵, Marianna Milano¹⁰⁶, Marcos F. Minicucci¹⁰⁷, Bertrand Moal¹⁰⁸, Taha Mohseni Ahooyi¹⁰⁹, Jason H. Moore¹¹⁰, Cinta Moraleda¹¹¹, Jeffrey S. Morris¹¹², Michele Morris⁸, Karyn L. Moshal¹¹³, Sajad Mousavi¹, Danielle L. Mowery²⁶, Douglas A. Murad¹⁷, Shawn N. Murphy¹¹⁴, Thomas P. Naughton¹¹⁵, Carlos Tadeu Breda Neto⁴⁵, Antoine Neuraz⁶, Jane Newburger⁶⁰, Kee Yuan Ngiam³⁶, Wanjiku F. M. Njoroge¹¹⁶, James B. Norman¹, Jihad Obeid⁷¹, Marina P. Okoshi¹⁰⁷, Karen L. Olson¹¹⁷, Gilbert S. Omenn¹¹⁸, Nina Orlova¹⁸, Brian D. Ostasiewski¹¹⁹, Nathan P. Palmer¹, Nicolas Paris¹⁸, Lav P. Patel¹²⁰, Miguel Pedrera-Jiménez²⁴, Ashley C. Pfaff¹²¹, Emily R. Pfaff¹²², Danielle Pillion¹, Sara Pizzimenti⁵⁷, Hans U. Prokosch¹²³, Robson A. Prudente¹²⁴, Andrea Prunotto⁹⁸, Víctor Quirós-González²⁴, Rachel B. Ramoni¹²⁵, Maryna Raskin⁴³, Siegbert Rieg¹²⁶, Gustavo Roig-Domínguez²⁴, Pablo Rojo¹¹¹, Paula Rubio-Mayo²⁴, Paolo Sacchi⁵⁹, Carlos Sáez¹²⁷, Elisa Salamanca¹⁸, Malarkodi Jebathilagam Samayamuthu⁸, L. Nelson Sanchez-Pinto¹²⁸, Arnaud Sandrin¹⁸, Nandhini Santhanam⁷⁹, Janaina C. C. Santos¹²⁹, Fernando J. Sanz Vidorreta¹⁷, Maria Savino¹³⁰, Emily R. Schriver³⁹, Petra Schubert²¹, Juergen Schuettler¹³¹, Luigia Scudeller⁵⁷, Neil J. Sebire¹³², Pablo Serrano-Balazote²⁴, Patricia Serre¹⁸, Arnaud Serret-Larmande¹³³, Mohsin Shah²⁸, Zahra Shakeri Hossein Abad¹, Domenick Silvio¹³⁴, Piotr Sliz¹³⁵, Jiyeon Son¹³⁶, Charles Sondag¹³⁷, Andrew M. South¹³⁸, Francesca Sperotto⁶⁰, Anastasia Spiridou²⁸

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41746-022-00601-0>.

Correspondence and requests for materials should be addressed to Tianxi Cai or Gabriel A. Brat.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

Zachary H. Strasser⁷⁷, Amelia L. M. Tan¹, Bryce W. Q. Tan⁴⁰, Byorn W. L. Tan⁴⁰, Suzana E. Tanni¹⁰⁷, Deanne M. Taylor¹³⁹, Ana I. Terriza-Torres²⁴, Valentina Tibollo³², Patric Tippmann⁸⁵, Emma M. S. Toh¹⁴⁰, Carlo Torti⁴¹, Enrico M. Trecarichi⁴¹, Yi-Ju Tseng¹⁴¹, Andrew K. Vallejos¹⁴², Gael Varoquaux¹⁴³, Margaret E. Vella¹, Guillaume Verdy¹⁰⁸, Jill-Jënn Vie¹⁴⁴, Shyam Visweswaran⁸, Michele Vitacca¹⁴⁵, Kavishwar B. Wagholikar⁹, Lemuel R. Waitman¹⁴⁶, and Xuan Wang¹, Demian Wassermann⁶³, Griffin M. Weber¹, Martin Wolkewitz⁸⁵, Scott Wong⁴⁰, Zongqi Xia¹⁴⁷, Xin Xiong⁷⁵, Ye Ye^{148,149}, Nadir Yehya¹⁵⁰, William Yuan¹, Alberto Zambelli¹⁵¹, Harrison G. Zhang¹, Daniela Zöller⁹⁸, Valentina Zuccaro⁵⁹ and Chiara Zucco¹⁰⁶

⁴²Department of Biomedical Informatics, University of Kentucky, Lexington, USA. ⁴³Health Catalyst, INC., Cambridge, USA. ⁴⁴BIOMERIS (BIOMedical Research Informatics Solutions), Pavia, Italy. ⁴⁵Clinical Research Unit of Botucatu Medical School, São Paulo State University, Botucatu, Brazil. ⁴⁶Pediatric emergency Department, Hôpital Necker-Enfants Malades, Assistance Public-Hôpitaux de Paris, Paris, France. ⁴⁷National Center for Infectious Diseases, Tan Tock Seng Hospital, Singapore, Singapore. ⁴⁸BIG-ARC, The University of Texas Health Science Center at Houston, School of Biomedical Informatics, Houston, USA. ⁴⁹Department of Pediatrics, Medical University of South Carolina, Charleston, USA. ⁵⁰Internal Medicine Department, Botucatu Medical School, São Paulo State University, Botucatu, Brazil. ⁵¹Department of Surgery, St. Luke's University Health Network, Bethlehem, USA. ⁵²Department of Medicine, Division of Nephrology, Ente Ospedaliero Cantonale, Lugano, Switzerland. ⁵³IT Department, Innovation & Data, APHP Greater Paris University Hospital, Paris, France. ⁵⁴IT Department, ASST Pavia, Voghera, Italy. ⁵⁵Strategy and Transformation Department, APHP Greater Paris University Hospital, Paris, France. ⁵⁶Technical University of Munich, Munich, Germany. ⁵⁷Scientific Direction, IRCCS Ca' Granda Ospedale Maggiore Policlinico di Milano, Milan, Italy. ⁵⁸North Carolina Translational and Clinical Sciences (NC TraCS) Institute, UNC Chapel Hill, Chapel Hill, USA. ⁵⁹Division of Infectious Diseases I, Fondazione I.R.C.C.S. Policlinico San Matteo, Italy, Pavia, Italy. ⁶⁰Department of Cardiology, Boston Children's Hospital, Harvard Medical School, Boston, USA. ⁶¹Department of Biomedical Informatics, HEGP, APHP Greater Paris University Hospital, Paris, France. ⁶²Department of Anesthesia, St. Luke's University Health Network, Bethlehem, USA. ⁶³Université Paris-Saclay, Inria, CEA, Palaiseau, France. ⁶⁴INRIA Sophia-Antipolis – ZENITH team, LIRMM, Montpellier, France. ⁶⁵Computational Health Informatics Program, Boston Children's Hospital, Boston, USA. ⁶⁶Department of Internal Medicine, University of Kentucky, Lexington, USA. ⁶⁷UOC Ricerca, Innovazione e Brand reputation, ASST Papa Giovanni XXIII, Bergamo, Bergamo, Italy. ⁶⁸Population Health and Data Science, MAVERIC, VA Boston Healthcare System, Boston, USA. ⁶⁹Informatics Institute, University of Alabama at Birmingham, Birmingham, USA. ⁷⁰IAM unit, INSERM Bordeaux Population Health ERIAS TEAM, Bordeaux University Hospital/ERIAS - Inserm U1219 BPH, Bordeaux, France. ⁷¹Biomedical Informatics Center, Medical University of South Carolina, Charleston, USA. ⁷²Clinical Research Informatics, Boston Children's Hospital, Boston, USA. ⁷³IT department, Innovation & Data (APHP), UMRS1142 (INSERM), APHP Greater Paris University Hospital, INSERM, Paris, France. ⁷⁴Department of Biomedical and Health Informatics, Children's Hospital of Philadelphia, Philadelphia, USA. ⁷⁵Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, USA. ⁷⁶SED/SIERRA, Inria Centre de Paris, Paris, France. ⁷⁷Department of Medicine, Massachusetts General Hospital, Boston, USA. ⁷⁸Health Information Technology & Services, University of Michigan, Ann Arbor, USA. ⁷⁹Heinrich-Lanz-Center for Digital Health, University Medicine Mannheim, Heidelberg University, Mannheim, Germany. ⁸⁰Department of Computational Biology and Bioinformatics, University of Michigan, Ann Arbor, USA. ⁸¹Biostatistics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, USA. ⁸²Department of Anesthesiology, Critical Care, and Pain Medicine and Computational Health Informatics Program, Boston Children's Hospital, Boston, USA. ⁸³Department of Learning Health Sciences, University of Michigan Medical School, Ann Arbor, MI, USA. ⁸⁴Institute of Digitalization in Medicine, Faculty of Medicine and Medical Center, University of Freiburg, Germany, Freiburg, Germany. ⁸⁵Institute of Medical Biometry and Statistics, Institute of Medical Biometry and Statistics, Medical Center, University of Freiburg, Freiburg, Germany. ⁸⁶Department of Biomedical Informatics, HEGP, APHP Greater Paris University Hospital, Paris, France. ⁸⁷Division of Biomedical Informatics (Department of Internal Medicine), University of Kentucky, Lexington, USA. ⁸⁸Center for Precision Psychiatry, Massachusetts General Hospital, Boston, USA. ⁸⁹Department of Critical Care Medicine, Children's Hospital of Pittsburgh, Pittsburgh, USA. ⁹⁰Medical University of South Carolina, Charleston, USA. ⁹¹Department of Pediatrics, Division of Human Genetics, The Children's Hospital of Philadelphia and the Perelman School of Medicine at the University of Pennsylvania, Philadelphia, USA. ⁹²Center for Medical Information and Communication Technology, University Hospital Erlangen, Erlangen, Germany. ⁹³Renaissance Computing Institute/Department of Computer Science, University of North Carolina, Chapel Hill, USA. ⁹⁴Clinical Research Unit, Saint Antoine Hospital, APHP Greater Paris University Hospital, Paris, France. ⁹⁵Clevy.io, Paris, France. ⁹⁶Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, USA. ⁹⁷Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, USA. ⁹⁸Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center, University of Freiburg, Freiburg, Germany, Freiburg, Germany. ⁹⁹Michigan Institute for Clinical and Health Research (MICHR) Informatics, University of Michigan, Ann Arbor, MI, USA. ¹⁰⁰Harvard Catalyst, Harvard Medical School, Boston, USA. ¹⁰¹Clinical Research Unit, Paris Saclay, APHP Greater Paris University Hospital, Boulogne-Billancourt, France. ¹⁰²Medical Informatics Center, Hospital das Clínicas, Faculty of Medicine of Botucatu, Clinics hospital of the Botucatu Medical School, São Paulo State University, Botucatu, Brazil. ¹⁰³Department of Surgery, Beth Israel Deaconess Medical Center, Boston, USA. ¹⁰⁴Department of Anesthesiology and Critical Care, Children's Hospital of Philadelphia, Philadelphia, USA. ¹⁰⁵ENS, PSL University, Paris, France. ¹⁰⁶Department of Medical and Surgical Sciences, University Magna Graecia of Catanzaro, Italy, Catanzaro, Italy. ¹⁰⁷Internal Medicine Department of Botucatu Medical School, São Paulo State University, Botucatu, Brazil. ¹⁰⁸IAM unit, Bordeaux University Hospital, Bordeaux, France. ¹⁰⁹Department of Biomedical Health Informatics, Children's Hospital of Philadelphia, Philadelphia, USA. ¹¹⁰Department of Computational Biomedicine, Cedars-Sinai Medical Center, West Hollywood, USA. ¹¹¹Pediatric Infectious Disease Department, Hospital Universitario 12 de Octubre, Madrid, Spain. ¹¹²Department of Biostatistics, Epidemiology, and Informatics (dept), Institute for Biomedical Informatics, University of Pennsylvania Perelman School of Medicine, Berwyn, USA. ¹¹³Department of Infectious Diseases, Great Ormond Street Hospital for Children, UK, London, UK. ¹¹⁴Department of Neurology, Massachusetts General Hospital, Boston, USA. ¹¹⁵Harvard Catalyst | The Harvard Clinical and Translational Science Center, Harvard Medical School, Boston, USA. ¹¹⁶Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, USA. ¹¹⁷Computational Health Informatics Program and Department of Pediatrics, Boston Children's Hospital and Harvard Medical School, Boston, USA. ¹¹⁸Department of Computational Medicine & Bioinformatics, Internal Medicine, Human Genetics, and School of Public Health, University of Michigan, Ann Arbor, USA. ¹¹⁹CTSI, WFBMI, Wake Forest School of Medicine, Winston Salem, USA. ¹²⁰Department of Internal Medicine, Division of Medical Informatics, University of Kansas Medical Center, Kansas City, USA. ¹²¹Department of Surgery, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, USA. ¹²²NC TraCS Institute, UNC Chapel Hill, Chapel Hill, USA. ¹²³Department of Medical Informatics, University of Erlangen-Nürnberg, Erlangen, Germany. ¹²⁴Clinical Research Unit São Paulo State University, Clinical Research Unit São Paulo State University, Botucatu, Brazil. ¹²⁵Department of Veterans Affairs Department of Veterans Affairs, Office of Research and Development, Washington, DC, USA. ¹²⁶Division of Infectious Diseases, Department of Medicine II, Medical Center – University of Freiburg, Faculty of Medicine, Freiburg, Germany. ¹²⁷Biomedical Data Science Lab, ITACA Institute, Universitat Politècnica de València, Valencia, Spain. ¹²⁸Department of Pediatrics (Critical Care), Northwestern University Feinberg School of Medicine, Chicago, USA. ¹²⁹Nurse Department of FMB - Medicine School of Botucatu, Clinical Research Unit of Botucatu Medical School, São Paulo State University, Botucatu, Brazil. ¹³⁰ASST Pavia, Lombardia Region Health System, Management Engineer, Direction, Pavia, Italy. ¹³¹Department of Anesthesiology, University Hospital Erlangen, FAU Erlangen-Nürnberg, Germany, Erlangen, Germany. ¹³²Digital Research, Informatics and Virtual Environments (DRIVE), Great Ormond Street Hospital for Children NIHR BRC, UK, London, UK. ¹³³Hôpital Saint Louis, Department of Biostatistics and Bioinformatics, APHP Greater Paris University Hospital, Paris, France. ¹³⁴MICHR Informatics, University of Michigan, Ann Arbor, USA. ¹³⁵CHIP, Boston Children's Hospital, Boston, USA. ¹³⁶Department of Neurology, University of Pittsburgh Medical Center, Pittsburgh, USA. ¹³⁷Critical Care Medicine, Department of Medicine, St. Luke's University Health Network, Bethlehem, PA, USA. ¹³⁸Department of Pediatrics-Section of Nephrology, Brenner Children's, Wake Forest University School of Medicine, Winston Salem, USA. ¹³⁹Department of Biomedical Health Informatics and the Department of Pediatrics, The Children's Hospital of Philadelphia and the University of Pennsylvania Perelman Medical School, Philadelphia, USA. ¹⁴⁰Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. ¹⁴¹Department of Information Management, National Central University, Taoyuan, Taiwan. ¹⁴²Clinical & Translational Science Institute, Medical College of Wisconsin, Milwaukee, USA. ¹⁴³Université Paris-Saclay, Inria, CEA, Montréal Neurological Institute, McGill University, Palaiseau, France. ¹⁴⁴Sequel, Inria Lille, Villeneuve-d'Ascq, France. ¹⁴⁵Respiratory Department, ICS S. Maugeri IRCCS Pavia Italy, Lumezzane (Bs), Italy. ¹⁴⁶Department of Health Management and Informatics, University of Missouri, MO, Columbia, USA. ¹⁴⁷Department of Neurology, University of Pittsburgh, Pittsburgh, USA. ¹⁴⁸Department of Veterans Affairs, 1100 First Street, NW, Washington, DC 20420, USA. ¹⁴⁹University of Pittsburgh, Pittsburgh, USA. ¹⁵⁰Department of Anesthesiology and Critical Care Medicine, Children's Hospital of Philadelphia and University of Pennsylvania, Philadelphia, USA. ¹⁵¹Department of Oncology, ASST Papa Giovanni XXIII, Bergamo, Bergamo, Italy.