




Demonstration of Interoperability Between MIDRC and N3C: A COVID-19 Severity Prediction Use Case

Heather M. Whitney^{1,2}  · Rachel Baccile³ · Hui Li^{1,2} · Karen Drukker^{1,2} · Christopher Meyer^{2,4} · Nicholas P. Gruszaszkas^{1,2} · Weijie Chen^{2,5} · Diane S. Lauderdale^{2,6} · Sandy Napel^{2,7} · Seyed Kahaki^{2,5} · Rui Carlos Sá⁸ · Chris Beesley⁹ · Brandy Phalora⁹ · Sam Michael¹⁰ · Robert L. Grossman^{2,4} · Ken Gersing¹⁰ · Maryellen L. Giger^{1,2} · and on Behalf of the National COVID Cohort Collaborative (N3C) Consortium

Received: 4 December 2024 / Revised: 10 June 2025 / Accepted: 28 June 2025
© The Author(s) 2025

Abstract

Interoperability between data sources, one of the FAIR (Findability, Accessibility, Interoperability, and Reusability) principles for scientific data management, can enable multi-modality research. The purpose of our study was to investigate the potential for interoperability between an imaging resource, the Medical Imaging and Data Resource Center (MIDRC), and a clinical record resource, the National COVID Cohort Collaborative (N3C). The use case was the prediction of COVID-19 severity, defined as evidence for invasive ventilatory support, extracorporeal membrane oxygenation, death, or discharge to hospice in the N3C clinical record. Patient-level matching between MIDRC and N3C was identified using Privacy Preserving Record Linking via an honest broker. We identified positive COVID-19 tests and chest radiograph procedures in N3C and used the interval between them to identify images with matching intervals in MIDRC. Of the 236 patients (306 unique images) meeting initial inclusion criteria in MIDRC, 117 patients (and 139 unique images) remained after date interval matching between repositories and exclusion of patients with multiple potential matches. The Charlson Comorbidity Index (CCI) and the minimum mean arterial pressure (MAP) on the day of the chest radiograph were used as clinical indicators. The AUC in the task of predicting severe COVID-19 was evaluated using the computer-extracted imaging index alone (MIDRC), clinical indicators alone (N3C), and both together. Our model combining imaging and clinical indicators (CCI over 2 and MAP below 70) to predict severe COVID had an AUC of 0.73 (95% CI 0.62–0.84), and the models including imaging or clinical indicators alone were 0.67 (95% CI 0.56–0.79) and 0.69 (95% CI 0.59–0.80), respectively. This study highlights the potential for cross-platform data sharing to facilitate future multi-modality research and broader collaborative studies.

Keywords Interoperability · COVID-19 · Comorbidity · Multi-modality

Introduction

Data commons and repositories are an important resource for big-data analysis that can lead to advances in medicine. They have the potential to contribute multi-modal data,

bringing together data derived from medical images and from clinical records. The National Institutes of Health (NIH) launched several data initiatives in response to the coronavirus-19 (COVID-19) pandemic, including the Medical Imaging and Data Resource Center (MIDRC) [1] and the National COVID Cohort Collaborative (N3C) [2]. These massive data resources have the potential to support retrospective evaluation of data that can then serve as use cases for future medical advancements, including use of artificial intelligence and machine learning methods [3].

Interoperability between data resources can be particularly high impact for medicine [4, 5] and is part of the FAIR [6] principles for data (Findability, Accessibility, Interoperability, and Reusability). In the context of FAIR, interoperability is defined as “the ability of data or

Heather M. Whitney and Rachel Baccile are considered co-first authors.

Robert L. Grossman, Ken Gersing, and Maryellen L. Giger are considered co-senior authors.

Rui Carlos Sá is formerly from the Department of Medicine, University of California, San Diego, CA, USA.

Extended author information available on the last page of the article

tools from non-cooperating resources to integrate or work together with minimal effort” [6]. Governance models, which oversee data ingestion, storage, maintenance, and disposal, have a substantial influence on interoperability [5]. Challenges for data interoperability for COVID-19 research were identified early on in the pandemic [7] and addressing them has been an ongoing area of effort, for purposes such as integrating genomic and clinical data [8] and vaccine response data [9]. A recent review highlighted the lack of interoperable datasets for COVID-19, especially for imaging-related data [10].

MIDRC and N3C operate under two different data privacy governance models. Images in the open MIDRC data commons are fully de-identified during the ingestion process. The MIDRC de-identification process includes randomly shifting dates while maintaining the relative sequence of events for each patient, ensuring the integrity of their longitudinal timeline. This means that only *relative* data with respect to imaging exam dates are available, such as the interval between imaging acquisition and an associated measurement (such as a COVID-19 test). On the other hand, data in N3C exist as a Limited Dataset [11, 12], which means that actual dates are associated with patients (e.g., dates of imaging procedures and dates of laboratory tests). Integrating imaging data from MIDRC and clinical data from N3C requires coordination and collaboration between the two

data resources via an honest broker [13] to accurately match patients, align dates, and achieve interoperability capabilities that may enhance the use of each dataset (Fig. 1).

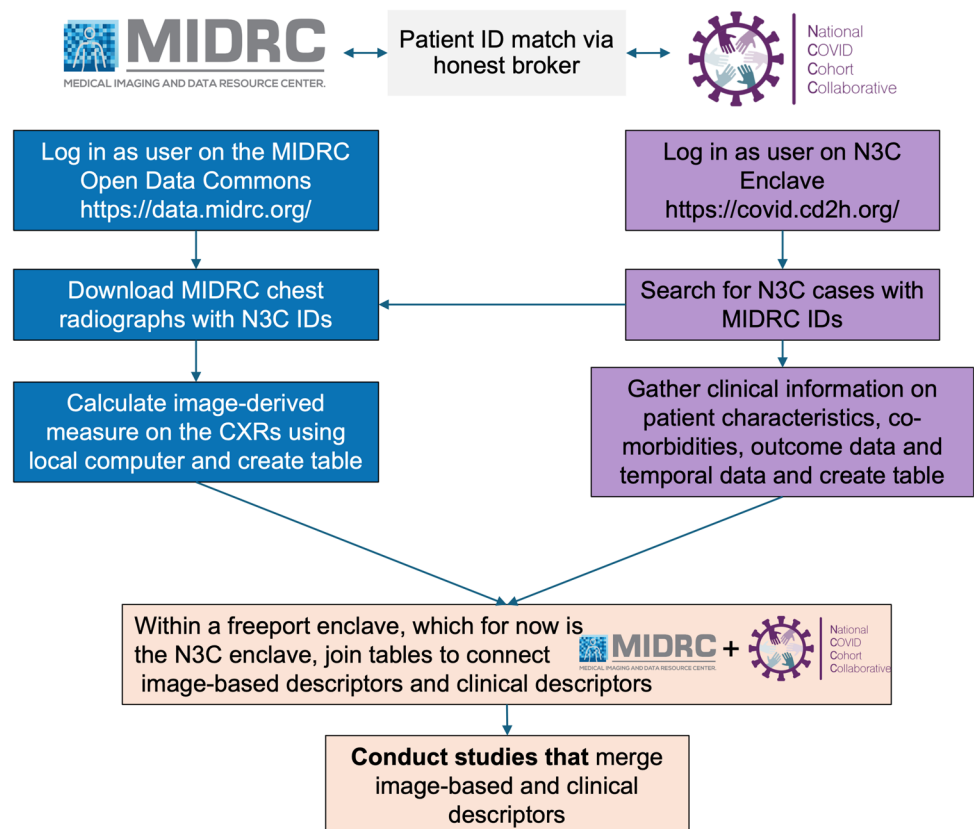
The purpose of our study was to demonstrate the interoperability between MIDRC and N3C via a use case of predicting severe outcomes for COVID-19 patients. We describe the logistics needed for the interoperability use case as well as example models. Three measures of predicting severe outcomes were investigated: (a) a previously-published computer-extracted measure using imaging alone (using images from MIDRC), (b) a measure derived from the clinical record alone (using data from N3C), and (c) a measure that combined data from both (i.e., imaging + clinical data).

Materials and Methods

Study Population

Privacy-preserving software services were established between MIDRC, N3C, and the honest broker at Regenstrief Institute [14]. Regenstrief acted as the honest broker for both MIDRC and N3C. When a MIDRC Globally Unique Identifier (“MIDRC ID”), which is privacy preserving, was passed to Regenstrief via a MIDRC Gen3

Fig. 1 Conceptual overview of interoperability between MIDRC and N3C



data mesh service (gen3.org), Regenstrief, as the honest broker, could perform a match to see if there was matching data within N3C. If N3C passed an N3C Identifier (“N3C ID”) to Regenstrief, Regenstrief could perform a match and return privacy-preserving MIDRC IDs to N3C. Due to the sensitivity of the data and the agreements used to collect the data, in all cases, N3C data (including N3C patient IDs) never left its secure enclave. The matched and integrated data was always analyzed within the N3C secure enclave, which is approved to manage controlled access data and authorized to interoperate with MIDRC [15]. With these services, a study population of subjects with data in both MIDRC (identified with a MIDRC ID) and N3C (identified with a N3C ID) was created. For these patients, the following steps were taken.

Imaging Index from Images in MIDRC

Unique de-identified patient MIDRC IDs were downloaded from N3C via a data download request. These MIDRC IDs were used to query the MIDRC open data commons for imaging studies relevant for this work according to two stages of inclusion criteria: (a) chest radiographs that were acquired no later than 7 days from a positive COVID-19 test result (“C19 + test result”) and (b) acquired in anterior–posterior view. A computer-extracted imaging-derived measure of COVID-19 severity was determined for each chest radiograph using a previously published deep learning model [16] that was developed to predict the development of severe COVID-19 within 24 h of the C19 + test result. The model uses a sequential learning framework that was successively trained in four phases; full technical details can be reviewed at the original publication [16]. In that study, severe COVID-19 was defined as patient admission to intensive care and/or evidence of intubation from the clinical record. Using the existing algorithm, the “imaging index” was generated for each eligible chest radiograph. Note that in the present study, severe COVID-19 was defined differently, as described below. Thus, the imaging index served as a surrogate measure of severe COVID-19 for this study.

Note that because the focus of this study is to demonstrate interoperability and an existing algorithm was used, this study did not develop, train, or validate a new imaging-based algorithm to predict COVID-19 severity.

Matching of the Specific Imaging Exam (MIDRC) and Specific Clinical Procedures (N3C)

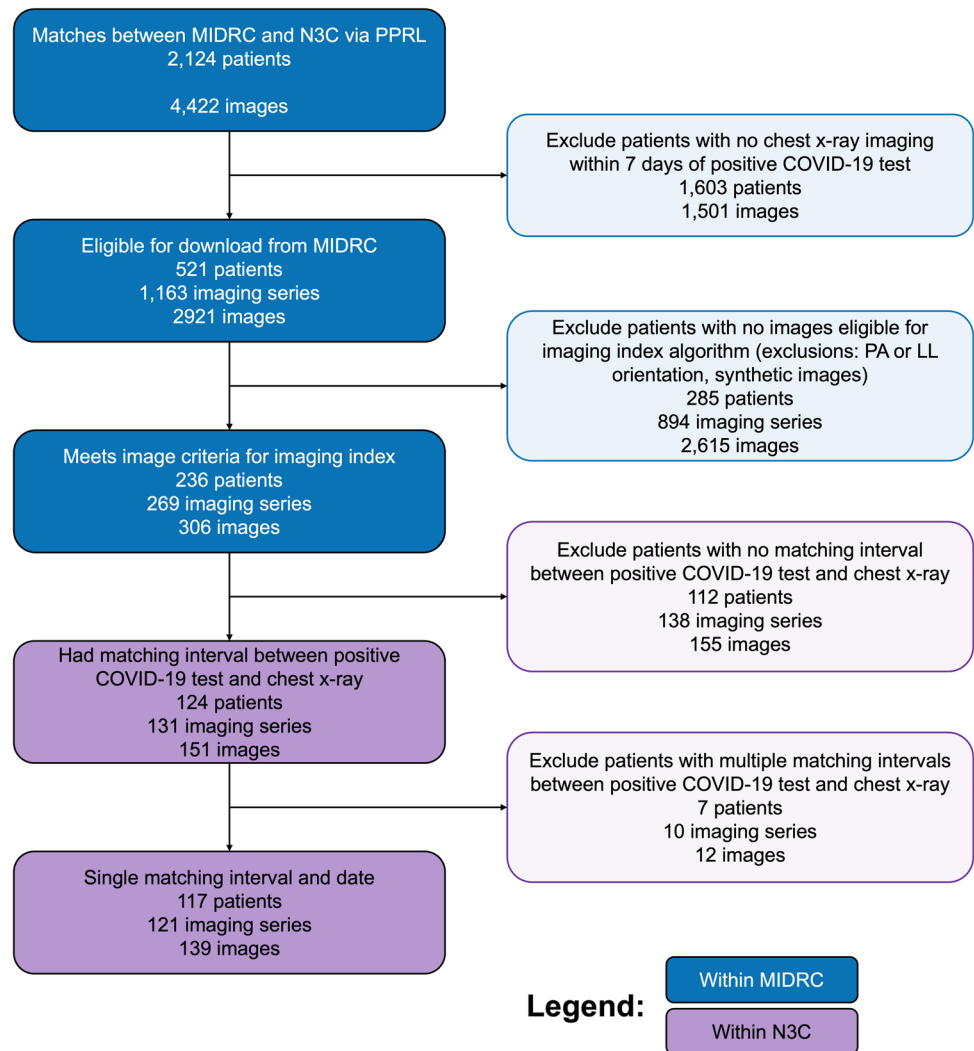
To match data between MIDRC and N3C, the time interval between a patient’s positive C19 + test and a chest radiograph procedure in N3C was calculated and matched exactly to the MIDRC interval. Patients were excluded if there was no exact matching date interval, either due to missing C19 + test results, missing a documented chest radiograph procedure, or the time interval between the two in N3C did not match the time interval in MIDRC. Because some patients had multiple C19 + tests and chest radiograph procedures recorded in N3C, the patient cohort was separated into subcohorts based upon the number of imaging studies available in MIDRC, the number of intervals between a C19 + test and imaging studies in MIDRC, and the number of matching intervals between a C19 + test and image procedure in N3C (Table 1). Patients with one imaging study, one interval between C19 + test and imaging study in MIDRC, and one matching interval within N3C were used in the final cohort.

Clinical Indicators from Data in N3C

Within N3C, the Charlson Comorbidity Index (CCI) [17, 18] at the admission date of the associated hospitalization and the minimum mean arterial pressure (MAP) on the day of the chest radiograph recorded in N3C were used as the clinical indicators in modeling. These indicators were identified for their use in other studies of COVID-19 outcomes using data from N3C [19–22]. The CCI was determined using the CCI Logic Liaison template available within the N3C enclave. For each chest radiograph, the patient’s CCI was classified into two categories: (a) CCI score between 0 and 2 and (b) CCI score 3 or greater. The minimum MAP was classified as either low (< 70) or not low (≥ 70) [23].

Table 1 Subcohorts identified through combinations of imaging studies in MIDRC and number of intervals between COVID-19 tests in MIDRC and in N3C. Patients in Subcohort 1 were used in the final analysis of this study. C19 +, COVID-19 positive

Sub cohort	Number of imaging studies in MIDRC	Number of intervals between C19 + test and imaging studies in MIDRC	Number of intervals between C19 + test and image procedure in N3C
1	One imaging study	One interval	One matching interval
2	One imaging study	One interval	One matching interval on multiple calendar days
3	One imaging study	Multiple intervals	Multiple matching intervals

Fig. 2 Consort figure of patients included in the study

Predictive Models for Severe COVID-19

Three predictive models for severe COVID-19 were evaluated for their potential to demonstrate interoperability between MIDRC and N3C: (a) using the radiographic AI-predictive model (imaging index alone), (b) using clinical indicators alone, and (c) using the combination of the imaging index and clinical indicators. In the present study, severe COVID-19 was defined as evidence for invasive ventilatory support, extracorporeal membrane oxygenation, death, or discharge to hospice [22] included within the clinical record held at N3C via the invasive respiratory support Logic Liaison template available within the N3C enclave. A logistic regression model was used to predict severe COVID-19, and standard errors were calculated at the patient level to account for multiple images within one imaging study (i.e., clustered data).

Statistical Analysis

The odds ratio for the factors in each logistic model were determined, along with their 95% confidence intervals (CI) using standard errors for clustered data to account for multiple images in one imaging exam for some patients. Specifically, we applied the Huber-White (HC0) sandwich estimator with clustering at the individual level to obtain standard errors using the Imtest R package [24]. The imaging index ranges from 0 to 1 and was represented in the model as tenths (0.1 units), so that the odds ratio represents the difference in odds of severe COVID-19 associated with a 0.1 unit change in the index. Receiver operating characteristic (ROC) analysis [25] was conducted in the task of predicting severe COVID-19. The area under the ROC curve (AUC) and 95% CIs were determined empirically using the R pROC package [26] for each of the three predictive models and served as the figure of merit.

Table 2 Patient characteristics in the cohort used in this study (by image)

Characteristic	N (%) or median [IQR]
Demographics	
Race/ethnicity	
Black not Hispanic	105 (75.54)
Hispanic or Latino	< 20
White not Hispanic or Latino	< 20
Asian not Hispanic or Latino	< 20
Other race not Hispanic or Latino	< 20
Sex	
Male	78 (56.12)
Female	61 (43.88)
Age, median [IQR]	60 [47.5, 74]
Clinical indicators	
CCI category	
0–2	77 (55.40)
3+	62 (44.60)
Low MAP (< 70)	36 (25.90)
Imaging indicator	
Image index, median [IQR]	0.24 [0.04, 0.62]
Outcomes	
Severe COVID-19	22 (15.83)

Severe COVID-19 was identified for patients who had *any* clinical indicator

Results

Study Population Characteristics

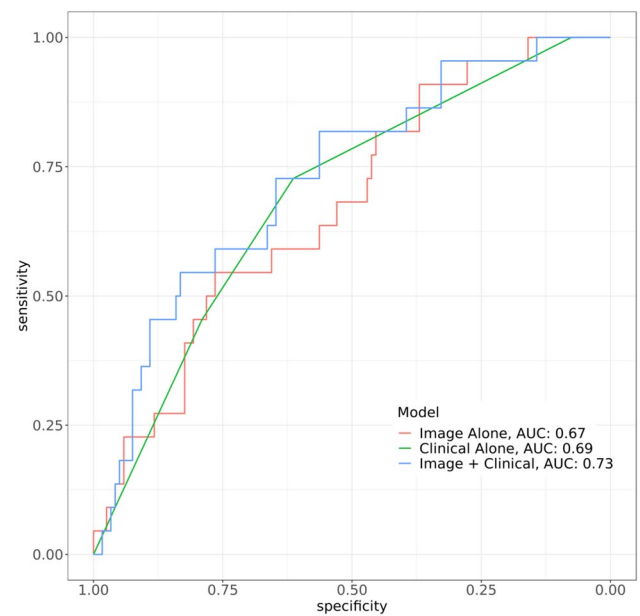
The final data set included 139 unique images from 121 imaging studies of 117 patients that had chest radiographs between June 3, 2020, and April 24, 2023. The consort figure for the inclusion criteria in the study is shown in Fig. 2.

Of the 139 images, 22 (15.83%) were associated with severe COVID, 62 (44.60%) had a CCI of 3 or greater, and the median minimum MAP was 78.30 [IQR 69.17, 89.75] (Table 2). The cohort was primarily Black (105, 75.54%) and male (78, 56.12%) and had a median age of 60 [IQR 47.50, 74].

Table 3 Regression results for the prediction of severe COVID-19 using the imaging index alone and the clinical indicators. CCI, Charlson comorbidity index; MAP, mean arterial pressure; OR, odds ratio; CI, confidence interval

	Model 1: Imaging index alone		Model 2: Clinical indicators alone		Model 3: imaging index + clinical indicators	
	OR	95% CI	OR	95% CI	OR	95% CI
Imaging index	1.19	(1.00, 1.42)	-	-	1.12	(0.95, 1.32)
CCI: 3+	-	-	4.94	(0.95, 25.56)	3.89	(0.77, 19.73)
Minimum MAP: < 70	-	-	0.61	(0.19, 1.98)	0.61	(0.18, 2.02)

The regression factors for the imaging index were multiplied by 10 to support interpretability, such that the OR is associated with an increase in the imaging index of 0.1

**Fig. 3** Receiver operating characteristic curves in the task of predicting severe COVID-19 from imaging index alone, clinical indicators alone, or the combination of imaging index and clinical indicators. AUC, area under the receiver operating characteristic curve

Statistical Analysis

In the model predicting severe COVID-19 using only the imaging index, there was an estimated odds ratio of 1.19 per tenth imaging unit (95% CI 1.00–1.42), indicating a marginal association with greater odds of severe COVID-19. In the model using only clinical indicators (CCI over 2 and MAP below 70), the odds ratio for severe COVID-19 was 4.94 (95% CI 0.95–25.56) for a high CCI and 0.61 (95% CI 0.19–1.98) for low MAP. When the imaging index and clinical indicators were combined, odds ratios for all factors remained similar, though none reached statistical significance at the 0.05 level. However, these results should be interpreted with caution, as the study is underpowered to detect significant associations. Regression results are shown in Table 3.

The AUC in the task of predicting severe COVID-19 using the imaging index alone was 0.67 (95% CI 0.56–0.79), and using clinical indicators (CCI over 2 and MAP below

70) alone was 0.69 (95% CI 0.59–0.80). The model combining imaging and clinical indicators to predict severe COVID had a greater median AUC (0.73, 95% CI 0.62–0.84) than the models including imaging or clinical indicators alone. ROC curves for the three models are shown in Fig. 3.

Discussion

Interoperability between different data repositories allows matching different types of data available for the same patient and plays a critical role in realizing the great potential of multimodal data analysis to improve healthcare. In particular, data commons interoperability enables curation of multimodal datasets that are foundational for the development of data-fusion AI, which has emerged as a promising technology to combine multiple types of data (e.g., radiology, pathology, clinical lab measurements, genomics) in various clinical tasks. In this study, interoperability between MIDRC (imaging data) and N3C (clinical data) has been demonstrated for the use case of prediction of severe COVID-19 using three predictive models. Two of them used the data from each repository alone and one used it in combination, the latter specifically facilitated by interoperability coordination efforts. The results demonstrate the potential for future discovery and hypothesis-driven studies, such as optimizing clinical data used for the task and subcohort analysis, or using other summary measures of comorbidity (such as the Elixhauser [27, 28] or Quan-Charlson indices [29]).

Limitations of this study included that the interoperability demonstration incorporated somewhat ad hoc choices for the intervals between C19 + test and imaging and for the clinical data windows used. During the development of the study, we reviewed the data and saw negligible change in, e.g., the CCI if the window was different. However, other studies may be more sensitive and may need to rely on more strict definitions of intervals. We also limited the study to straightforward matches in interval between imaging and clinical variables in order to simplify the demonstration of interoperability. In the future, more improvements are needed for series-based matches (as opposed to image-based matches, which was the focus here). The interval check was not optimal but was what was possible now because of date shifting (including potentially unknown date shifting between two commons). This limitation impacted the size of the final dataset, which limited the power to detect statistical significance. Expansion of interoperability will allow for research on larger cohorts by increasing the confidence in the matching of imaging data in MIDRC to clinical data in N3C.

The comparison of models was also limited in terms of statistical power due to the small sample size. Nonetheless,

there was evidence of effects worth exploring with a larger case series. Different interoperability procedures are needed to expand the dataset to additional patients and to extend the analysis to other subcohorts. Thus, this work shows the potential for interoperability and future hypothesis-driven work, including at specific operating points.

Conclusion

Interoperability between MIDRC and N3C has been demonstrated for the first time via the use case of predicting severe COVID-19 by combining both imaging-derived measures and data from the clinical record. Future work will investigate the incorporation of other clinical indicators and use tasks as well as an in-depth evaluation of how integrating imaging and clinical data improves clinical decision-making in this task and other tasks.

Acknowledgements The authors are grateful to Saad Ljazouli and Gianna Beck at Palantir for their contributions to this work.

The N3C Publication committee confirmed that this manuscript msid: 2189.754 is in accordance with N3C data use and attribution policies; however, this content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the N3C program.

The analyses described in this publication were conducted with data or tools accessed through the NCATS N3C Data Enclave <https://covid.cd2h.org> and N3C Attribution & Publication Policy v 1.2-2020-08-25b supported by NCATS Contract No. 75N95023D00001, and Axle Informatics Subcontract: NCATS-P00438-B. This research was possible because of the patients whose information is included within the data and the organizations (<https://ncats.nih.gov/n3c/resources/data-contribution/data-transfer-agreement-signatories>) and scientists who have contributed to the ongoing development of this community resource [<https://doi.org/10.1093/jamia/ocaa196>]. The N3C Enclave is available for public research use. To access data including that used in this manuscript, institutions must have a signed Data Use Agreement executed with the U.S. National Center for Advancing Translational Sciences (NCATS) and their investigators must complete mandatory training and must submit a Data Use Request (DUR) to N3C. To request N3C data access, researchers must follow instructions at <https://covid.cd2h.org/onboarding>. More than 4000 researchers currently have access to data in N3C; together they represent more than 300 US research institutions. Reviewers can confidentially request access by these same means.

We gratefully acknowledge the following core contributors to N3C:

Adam B. Wilcox, Adam M. Lee, Alexis Graves, Alfred (Jerrod) 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 MadlockBrown, 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, 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, J.W. Awori Hayanga, Jami Pincavitch, Jaylyn Clark, Jeremy Richard Harper, Jessica Islam, Jin Ge, Joel Gagnier, Johanna Loomba, John Buse, Jomol Mathew, Joni L. Rutter, Julie A. McMurtry, 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, Lesley Cottrell, Lili Portilla, Mariam Deacy, Mark M. Bissell, Marshall Clark, Mary Emmett, Matvey B. Palchuk, Melissa A. Haendel, Meredith Adams, Meredith Temple-O'Connor, Michael G. Kurilla, Michele Morris, Nasia Safdar, Nicole Garbarini, Noha Sharafeldin, Ofer Sadan, Patricia A. Francis, Penny Wung Burgoon, Philip R.O. Payne, 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, Umit Topaloglu, 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.

The following Data Partners with Released Data acknowledgment is required by the N3C Publications review.

Available: Advocate Health Care Network — UL1TR002389: The Institute for Translational Medicine (ITM) • Aurora Health Care Inc — UL1TR002373: Wisconsin Network For Health Research • Boston University Medical Campus — UL1TR001430: Boston University Clinical and Translational Science Institute • Brown University — U54GM115677: Advance Clinical Translational Research (Advance-CTR) • Carilion Clinic — UL1TR003015: iTHRIV Integrated Translational Health Research Institute of Virginia • Case Western Reserve University — UL1TR002548: The Clinical & Translational Science Collaborative of Cleveland (CTSC) • Charleston Area Medical Center — U54GM104942: West Virginia Clinical and Translational Science Institute (WVCTSI) • Children's Hospital Colorado — UL1TR002535: Colorado Clinical and Translational Sciences Institute • Columbia University Irving Medical Center — UL1TR001873: Irving Institute for Clinical and Translational Research • Dartmouth College — None (Voluntary) Duke University — UL1TR002553: Duke Clinical and Translational Science Institute • George Washington Children's Research Institute — UL1TR001876: Clinical and Translational Science Institute at Children's National (CTSA-CN) • George Washington University — UL1TR001876: Clinical and Translational Science Institute at Children's National (CTSA-CN) • Harvard Medical School — UL1TR002541: Harvard Catalyst • Indiana University School of Medicine — UL1TR002529: Indiana Clinical and Translational Science Institute • Johns Hopkins University — UL1TR003098: Johns Hopkins Institute for Clinical and Translational Research • Louisiana Public Health Institute — None (Voluntary) • Loyola Medicine • Loyola University Medical Center • Loyola University Medical Center — UL1TR002389: The Institute for Translational Medicine (ITM) • Maine Medical Center — U54GM115516: Northern New England Clinical & Translational Research (NNE-CTR) Network • Mary Hitchcock Memorial Hospital & Dartmouth Hitchcock Clinic — None (Voluntary) • Massachusetts General Brigham — UL1TR002541: Harvard Catalyst • Mayo Clinic Rochester — UL1TR002377: Mayo Clinic Center for Clinical and Translational Science (CCaTS) • Medical University of South Carolina — UL1TR001450: South Carolina Clinical & Translational Research Institute (SCTR) • MITRE Corporation — None (Voluntary) • Montefiore Medical Center — UL1TR002556: Institute for Clinical and Translational Research at Einstein and Montefiore • Nemours — U54GM104941: Delaware CTR ACCEL Program • NorthShore University HealthSystem — UL1TR002389: The Institute for Translational Medicine (ITM) • Northwestern University at Chicago — UL1TR001422: Northwestern University Clinical and Translational Science Institute (NUCATS) • OCHIN — INV-018455: Bill and Melinda Gates Foundation grant to Sage Bionetworks • Oregon Health & Science University — UL1TR002369: Oregon Clinical and Translational Research Institute • Penn State Health Milton S. Hershey Medical Center — UL1TR002014: Penn State Clinical and

Translational Science Institute • Rush University Medical Center — UL1TR002389: The Institute for Translational Medicine (ITM) • Rutgers, The State University of New Jersey — UL1TR003017: New Jersey Alliance for Clinical and Translational Science • Stony Brook University — U24TR002306 • The Alliance at the University of Puerto Rico, Medical Sciences Campus — U54GM133807: Hispanic Alliance for Clinical and Translational Research (The Alliance) • The Ohio State University — UL1TR002733: Center for Clinical and Translational Science • The State University of New York at Buffalo — UL1TR001412: Clinical and Translational Science Institute • The University of Chicago — UL1TR002389: The Institute for Translational Medicine (ITM) • The University of Iowa — UL1TR002537: Institute for Clinical and Translational Science • The University of Miami Leonard M. Miller School of Medicine — UL1TR002736: University of Miami Clinical and Translational Science Institute • The University of Michigan at Ann Arbor — UL1TR002240: Michigan Institute for Clinical and Health Research • The University of Texas Health Science Center at Houston — UL1TR003167: Center for Clinical and Translational Sciences (CCTS) • The University of Texas Medical Branch at Galveston — UL1TR001439: The Institute for Translational Sciences • The University of Utah — UL1TR002538: Uhealth Center for Clinical and Translational Science • Tufts Medical Center — UL1TR002544: Tufts Clinical and Translational Science Institute • Tulane University — UL1TR003096: Center for Clinical and Translational Science • The Queens Medical Center — None (Voluntary) • University Medical Center New Orleans — U54GM104940: Louisiana Clinical and Translational Science (LA CaTS) Center • University of Alabama at Birmingham — UL1TR003096: Center for Clinical and Translational Science • University of Arkansas for Medical Sciences — UL1TR003107: UAMS Translational Research Institute • University of Cincinnati — UL1TR001425: Center for Clinical and Translational Science and Training • University of Colorado Denver, Anschutz Medical Campus — UL1TR002535: Colorado Clinical and Translational Sciences Institute • University of Illinois at Chicago — UL1TR002003: UIC Center for Clinical and Translational Science • University of Kansas Medical Center — UL1TR002366: Frontiers: University of Kansas Clinical and Translational Science Institute • University of Kentucky — UL1TR001998: UK Center for Clinical and Translational Science • University of Massachusetts Medical School Worcester — UL1TR001453: The UMass Center for Clinical and Translational Science (UMCCTS) • University Medical Center of Southern Nevada — None (voluntary) • University of Minnesota — UL1TR002494: Clinical and Translational Science Institute • University of Mississippi Medical Center — U54GM115428: Mississippi Center for Clinical and Translational Research (CCTR) • University of Nebraska Medical Center — U54GM115458: Great Plains IDeA-Clinical & Translational Research • University of North Carolina at Chapel Hill — UL1TR002489: North Carolina Translational and Clinical Science Institute • University of Oklahoma Health Sciences Center — U54GM104938: Oklahoma Clinical and Translational Science Institute (OCTSI) • University of Pittsburgh — UL1TR001857: The Clinical and Translational Science Institute (CTSI) • University of Pennsylvania — UL1TR001878: Institute for Translational Medicine and Therapeutics • University of Rochester — UL1TR002001: UR Clinical & Translational Science Institute • University of Southern California — UL1TR001855: The Southern California Clinical and Translational Science Institute (SC CTSI) • University of Vermont — U54GM115516: Northern New England Clinical & Translational Research (NNE-CTR) Network • University of Virginia — UL1TR003015: iTHRIV Integrated Translational Health Research Institute of Virginia • University of Washington — UL1TR002319: Institute of Translational Health Sciences • University of Wisconsin-Madison — UL1TR002373: UW Institute for Clinical and Translational Research • Vanderbilt University Medical Center — UL1TR002243: Vanderbilt Institute for Clinical and Translational Research • Virginia Commonwealth University — UL1TR002649: C.

Kenneth and Dianne Wright Center for Clinical and Translational Research • Wake Forest University Health Sciences — UL1TR001420: Wake Forest Clinical and Translational Science Institute • Washington University in St. Louis — UL1TR002345: Institute of Clinical and Translational Sciences • Weill Medical College of Cornell University — UL1TR002384: Weill Cornell Medicine Clinical and Translational Science Center • West Virginia University — U54GM104942: West Virginia Clinical and Translational Science Institute (WVCTSI) Submitted: Icahn School of Medicine at Mount Sinai — UL1TR001433: ConduITS Institute for Translational Sciences • The University of Texas Health Science Center at Tyler — UL1TR003167: Center for Clinical and Translational Sciences (CCTS) • University of California, Davis — UL1TR001860: UC Davis Health Clinical and Translational Science Center • University of California, Irvine — UL1TR001414: The UC Irvine Institute for Clinical and Translational Science (ICTS) • University of California, Los Angeles — UL1TR001881: UCLA Clinical Translational Science Institute • University of California, San Diego — UL1TR001442: Altman Clinical and Translational Research Institute • University of California, San Francisco — UL1TR001872: UCSF Clinical and Translational Science Institute NYU Langone Health Clinical Science Core, Data Resource Core, and PASC Biorepository Core — OTA-21-015A: Post-Acute Sequelae of SARS-CoV-2 Infection Initiative (RECOVER) Pending: Arkansas Children's Hospital — UL1TR003107: UAMS Translational Research Institute • Baylor College of Medicine — None (Voluntary) • Children's Hospital of Philadelphia — UL1TR001878: Institute for Translational Medicine and Therapeutics • Cincinnati Children's Hospital Medical Center — UL1TR001425: Center for Clinical and Translational Science and Training • Emory University — UL1TR002378: Georgia Clinical and Translational Science Alliance • HonorHealth — None (Voluntary) • Loyola University Chicago — UL1TR002389: The Institute for Translational Medicine (ITM) • Medical College of Wisconsin — UL1TR001436: Clinical and Translational Science Institute of Southeast Wisconsin • MedStar Health Research Institute — None (Voluntary) • Georgetown University — UL1TR001409: The Georgetown-Howard Universities Center for Clinical and Translational Science (GHUCCTS) • MetroHealth — None (Voluntary) • Montana State University — U54GM115371: American Indian/Alaska Native CTR • NYU Langone Medical Center — UL1TR001445: Langone Health's Clinical and Translational Science Institute • Ochsner Medical Center — U54GM104940: Louisiana Clinical and Translational Science (LA CaTS) Center • Regenstrief Institute — UL1TR002529: Indiana Clinical and Translational Science Institute • Sanford Research — None (Voluntary) • Stanford University — UL1TR003142: Spectrum: The Stanford Center for Clinical and Translational Research and Education • The Rockefeller University — UL1TR001866: Center for Clinical and Translational Science • The Scripps Research Institute — UL1TR002550: Scripps Research Translational Institute • University of Florida — UL1TR001427: UF Clinical and Translational Science Institute • University of New Mexico Health Sciences Center — UL1TR001449: University of New Mexico Clinical and Translational Science Center • University of Texas Health Science Center at San Antonio — UL1TR002645: Institute for Integration of Medicine and Science • Yale New Haven Hospital — UL1TR001863: Yale Center for Clinical Investigation

Author Contribution All authors contributed to the study conception and design. Material preparation, data collection, and analysis were performed by Heather Whitney and Rachel Baccile. The first draft of the manuscript was written by Heather Whitney and Rachel Baccile, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript. Authorship was determined by ICMJE requirements.

Funding This study was supported by the Institute for Translational Medicine, which is supported by the National Center for Advancing

Translational Sciences (NCATS) of the National Institutes of Health (NIH) through Grant Numbers UL1TR002389, KL2TR002387, and TL1TR002388. MIDRC is supported by the NIH, National Institute of Biomedical Imaging and Bioengineering contract 75N92020D00021 and by ARPA-H.

Data and Code Availability The imaging data are freely available at <https://data.midrc.org>. The clinical data of the N3C dataset were available to the authors by Data Use Request DUR-BB40587 and are not available to the public.

Code for processing the N3C dataset is not available due to restrictions on the Limited Dataset. In general, all code shareable by MIDRC is available at <https://github.com/MIDRC>.

Declarations

Competing interests The authors declare no competing interests.

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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.


References

1. MIDRC [Internet]. [cited 2023 Jun 1]. The Medical Imaging and Data Resource Center. Available from: <https://www.midrc.org>
2. National Covid Cohort Collaborative [Internet]. [cited 2023 Dec 29]. Available from: <https://covid.cd2h.org/>
3. Chen W, Sá RC, Bai Y, Napel S, Gevaert O, Lauderdale DS, et al. Machine learning with multimodal data for COVID-19. *Heliyon*. 2023 Jul 1;9(7):e17934.
4. Perlin JB. Health Information Technology Interoperability and Use for Better Care and Evidence. *JAMA*. 2016 Oct 25;316(16):1667–8.
5. Lehne M, Sass J, Essenwanger A, Schepers J, Thun S. Why digital medicine depends on interoperability. *npj Digit Med*. 2019 Aug 20;2(1):1–5.
6. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016 Mar 15;3(1):160018.
7. Bauer DC, Metke-Jimenez A, Maurer-Stroh S, Tiruvayipati S, Wilson LOW, Jain Y, et al. Interoperable medical data: The missing link for understanding COVID-19. *Transboundary and Emerging Diseases*. 2021;68(4):1753–60.
8. Bernasconi A, Ceri S. Interoperability of COVID-19 Clinical Phenotype Data with Host and Viral Genetics Data. *BioMed*. 2022 Mar;2(1):69–81.
9. Malik S, Pana Dorothea Z, Argyropoulos CD, Themistocleous S, Macken AJ, Valdenmaier O, et al. Data Interoperability in COVID-19 Vaccine Trials: Methodological Approach in the VACCELERATE Project. *JMIR Medical Informatics*. 2025;13:365590.

10. Maxwell L, Shreedhar P, Dauga D, McQuilton P, Terry RF, Denisiuk A, et al. FAIR, ethical, and coordinated data sharing for COVID-19 response: a scoping review and cross-sectional survey of COVID-19 data sharing platforms and registries. *The Lancet Digital Health*. 2023 Oct 1;5(10):e712–36.
11. National Institutes of Health. Protecting Personal Health Information in Research: Understanding the HIPAA Privacy Rule [Internet]. 2003 [cited 2024 Oct 30]. Available from: https://privacyruleandresearch.nih.gov/pr_02.asp
12. Other requirements relating to uses and disclosures of protected health information. [Internet]. [cited 2024 Oct 30]. 164.514(e). Available from: <https://www.ecfr.gov/current/title-45/part-164/section-164.514>
13. Boyd AD, Hosner C, Hunscher DA, Athey BD, Clauw DJ, Green LA. An “Honest Broker” mechanism to maintain privacy for patient care and academic medical research. *Int J Med Inform*. 2007 Jun;76(5–6):407–11.
14. Regenstrief Institute [Internet]. [cited 2024 Sep 9]. Regenstrief Institute - Research to transform health systems and communities for a healthier, more equitable world. Available from: <https://www.regenstrief.org/>
15. Grossman RL, Boyles RR, Davis-Dusenbery BN, Haddock A, Heath AP, O'Connor BD, et al. A Framework for the Interoperability of Cloud Platforms: Towards FAIR Data in SAFE Environments. *Sci Data*. 2024 Feb 26;11(1):241.
16. Li H, Drukker K, Hu Q, Whitney HM, Fuhrman JD, Giger ML. Predicting intensive care need for COVID-19 patients using deep learning on chest radiography. *JMI*. 2023 Aug;10(4):044504.
17. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *Journal of Chronic Diseases*. 1987 Jan 1;40(5):373–83.
18. Charlson M, Szatrowski TP, Peterson J, Gold J. Validation of a combined comorbidity index. *Journal of Clinical Epidemiology*. 1994 Nov 1;47(11):1245–51.
19. Jiang S, Loomba J, Zhou A, Sharma S, Sengupta S, Brown DE. A Bayesian Hierarchical Analysis on the Disparity of Emergency Department Visits for COVID-19: A Cohort Study Using National COVID Cohort Collaborative (N3C) Data. In: 2023 International Conference on Machine Learning and Applications (ICMLA) [Internet]. 2023 [cited 2024 Oct 29]. p. 144–51. Available from: <https://ieeexplore.ieee.org/document/10459884>
20. Hadley E, Yoo YJ, Patel S, Zhou A, Laraway B, Wong R, et al. Insights from an N3C RECOVER EHR-based cohort study characterizing SARS-CoV-2 reinfections and Long COVID | *Communications Medicine*. *Communications Medicine* [Internet]. 2024 [cited 2024 Oct 29];4(129). Available from: <https://www.nature.com/articles/s43856-024-00539-2>
21. Sharafeldin N, Bates B, Song Q, Madhira V, Yan Y, Dong S, et al. Outcomes of COVID-19 in Patients With Cancer: Report From the National COVID Cohort Collaborative (N3C). *JCO*. 2021 Jul 10;39(20):2232–46.
22. Bennett TD, Moffitt RA, Hajagos JG, Amor B, Anand A, Bissell MM, et al. Clinical Characterization and Prediction of Clinical Severity of SARS-CoV-2 Infection Among US Adults Using Data From the US National COVID Cohort Collaborative. *JAMA Network Open*. 2021 Jul 13;4(7):e2116901.
23. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. *Intensive Care Med*. 1996 Jul 1;22(7):707–10.
24. Zeileis A, Hothorn T. Diagnostic Checking in Regression Relationships. *R News*. 2002;2(3).
25. Metz CE. Basic principles of ROC analysis. *Seminars in Nuclear Medicine*. 1978 Oct;8(4):283–98.
26. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011 Mar 17;12(1):77.
27. Zhou W, Qin X, Hu X, Lu Y, Pan J. Prognosis models for severe and critical COVID-19 based on the Charlson and Elixhauser comorbidity indices. *Int J Med Sci*. 2020;17(15):2257–63.
28. Sharma N, Schwendimann R, Endrich O, Ausserhofer D, Simon M. Comparing Charlson and Elixhauser comorbidity indices with different weightings to predict in-hospital mortality: an analysis of national inpatient data. *BMC Health Serv Res*. 2021 Jan 6;21(1):13.
29. Quan H, Li B, Couris CM, Fushimi K, Graham P, Hider P, et al. Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries. *Am J Epidemiol*. 2011 Mar 15;173(6):676–82.

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

Authors and Affiliations

Heather M. Whitney^{1,2}  · Rachel Baccile³ · Hui Li^{1,2} · Karen Drukker^{1,2} · Christopher Meyer^{2,4} · Nicholas P. Gruszauskas^{1,2} · Weijie Chen^{2,5} · Diane S. Lauderdale^{2,6} · Sandy Napel^{2,7} · Seyed Kahaki^{2,5} · Rui Carlos Sá⁸ · Chris Beesley⁹ · Brandy Phalora⁹ · Sam Michael¹⁰ · Robert L. Grossman^{2,4} · Ken Gersing¹⁰ · Maryellen L. Giger^{1,2} · and on Behalf of the National COVID Cohort Collaborative (N3C) Consortium

✉ Heather M. Whitney
hwhitney@uchicago.edu

¹ Department of Radiology, University of Chicago, Chicago, IL, USA

² Medical Imaging and Data Resource Center, Chicago, IL, USA

³ Institute for Translational Medicine, Chicago, IL, USA

⁴ Center for Translational Data Science, University of Chicago, Chicago, IL, USA

⁵ Division of Imaging, Diagnostics, and Software Reliability, Office of Science and Engineering Laboratories, CDRH, US FDA, Silver Spring, MD, USA

⁶ Department of Public Health Sciences, University of Chicago, Chicago, IL, USA

⁷ Department of Radiology, Stanford University School of Medicine, Stanford, CA, USA

⁸ National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health, Bethesda, MD, USA

⁹ Regenstreif Institute, Indianapolis, IN, USA

¹⁰ Office of the Director, National Center for Advancing Translational Sciences, National Institutes of Health, Bethesda, MD, USA