
Brief Communications

COVID-19 Insights Partnership: Leveraging big data from the Department of Veterans Affairs and supercomputers at the Department of Energy under the public health authority

Rachel Ramoni¹, Molly Klote², Sumitra Muralidhar¹, Cynthia Brandt^{3,4,5},
Maya A. Bernstein⁶, Benjamin H. McMahon⁷, Daniel A. Jacobson^{8,9,10,11}, and
Amy C. Justice^{3,12,13} 

¹Office of Research and Development, Veterans Health Administration, U.S. Department of Veterans Affairs, Washington DC, USA, ²Office of Research Protections, Policy, and Education, Veterans Health Administration, U.S. Department of Veterans Affairs, Washington DC, USA, ³VA Connecticut Healthcare System, U.S. Department of Veterans Affairs, West Haven, Connecticut, USA, ⁴Department of Emergency Medicine, Yale University School of Medicine, New Haven, Connecticut, USA, ⁵Yale Center for Medical Informatics, Yale University School of Medicine, New Haven, Connecticut, USA, ⁶Office of the Assistant Secretary Planning and Evaluation, U.S. Department of Health and Human Services, Washington DC, USA, ⁷Los Alamos National Laboratory, U.S. Department of Energy, Los Alamos, New Mexico, USA, ⁸Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA, ⁹Bredesen Center for Interdisciplinary Research and Graduate Education, University of Tennessee, Knoxville, Tennessee, USA, ¹⁰Department of Psychology, University of Tennessee, Knoxville, Tennessee, USA, ¹¹NeuroNet Research Center, University of Tennessee, Knoxville, Tennessee, USA, ¹²Department of Internal Medicine, Yale University School of Medicine, New Haven, Connecticut, USA, and ¹³Department of Health Policy and Management, Yale University School of Public Health, New Haven, Connecticut, USA

Corresponding Author: Amy C. Justice, MD, PhD, VA Connecticut Healthcare System, 950 Campbell Avenue, 11-ACSLG, West Haven, CT 06516-2770, USA; amy.justice2@va.gov

Received 10 December 2020; Revised 17 February 2021; Editorial Decision 21 March 2021; Accepted 25 March 2021

Key words: COVID-19, Big Data, Public Health Authority, Department of Energy, Department of Veterans Affairs

In March 2020, the U.S. Department of Veterans Affairs (VA), Department of Energy (DOE), Department of Health and Human Services (HHS), and National Security Council (NSC) formed the COVID-19 Insights Partnership. Initially suggested by the VA Office of Research and Development, the partnership represents a major milestone in cooperation among U.S. federal agencies to advance public health. The goal is to establish a secure computing enclave at the DOE's Oak Ridge National Laboratory (ORNL), where datasets from multiple federal agencies can be brought together and combined for public health research purposes. With appropriate regulation, data governance, and data access policies in place, qualifying investigators from multiple federal agencies will be able to combine the DOE's high-performance computing and artificial intelligence

expertise with national longitudinal electronic health record data from veterans as well as health data from nonveterans diagnosed with coronavirus disease 2019 (COVID-19).^{1–5} This research will bolster the nation's ability to understand, prevent, and treat COVID-19 and will better position the nation to respond to future public health emergencies.

This advance would not have been possible without 2 inter-agency programs. The first began before COVID-19. In 2016, the VA and DOE formed an interagency collaboration known as Computational Health Analytics for Medical Precision to Improve Outcomes Now (CHAMPION) to demonstrate the power of combining the VA's health record system, Million Veteran Program (MVP) genetic data, and clinical research expertise with the DOE's high-

performance computing infrastructure and artificial intelligence expertise. The second occurred as part of the Insights Partnership in direct response to the pandemic. This was a novel application of the public health authority already invested within HHS to allow for sharing of health data among federal agencies for purposes that support public health. The use of this public health authority set the stage for collaboration not only for the current pandemic, but also for future national health emergencies, and for ongoing concerns such as suicide prevention or cancer care.

THE NEED FOR DATA INTEGRATION

As of this writing, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) vaccines have been brought from conception to mass deployment in under 12 months. Similarly, the case fatality rate (confirmed COVID deaths/confirmed COVID cases) has decreased from ~7% to ~1%, through a combination of caseload reduction and improved clinical practice. Yet many critical public health questions remain. These concern epidemiological situational awareness and disease transmission and manifestation that can best be answered through data integration across clinical, pathophysiological, disease mechanism, epidemiological, and decision-support perspectives. Some examples include the following:

- Characterizing and determining the extent of long-hauler syndrome
- Understanding the impact of SARS-CoV-2 strain evolution
- Rapidly identifying and characterizing patients in whom reinfection may have occurred
- Understanding impact on reduced transmission and morbidity and mortality of various vaccination protocols
- Characterizing the role of viral exposure, genetic factors (like numerous other viral infections), comorbidities, and prescribed medications on outcomes
- Combining mechanistic and clinical studies to better understand COVID-19 pathogenesis and to identify potential new therapies
- Situational awareness of the accurate costs of COVID morbidity and mortality, in the context of clinical history and improved understanding of pathophysiology
- Expediting design, development, and validation of diagnostics in a real-world setting

Common to each of the previous items is a targeted computational analysis of appropriate clinical data in an epidemiological context. Each of these attributes falls under the purview of a different agency and requires input from both policymakers and scientists. The COVID-19 Insights Partnership brings these capabilities together in support of an improved national response.

LEVERAGING THE PUBLIC HEALTH AUTHORITY

Drawing on the expertise of the regulatory and privacy community, we and our collaborators came to the conclusion that use of the HHS public health authority would make such sharing possible with the least burden and highest data quality while still preserving the privacy of the record subjects. While many researchers had previously considered combining individually identifiable health information covered by the Health Insurance Portability and Accountability Act (HIPAA) across federal agencies, no group had ever found a viable pathway. This partnership represents a first in that regard. The heart of the partnership is a memorandum of understanding signed

by HHS, VA, and DOE. The Partnership plans additional data use agreements for the future.

Each agency plays a unique and critical role: HHS will establish the COVID-19 Insights Collaboration Records Database to support understanding and tracking of SARS-CoV-2 and responses to COVID-19 outbreaks. The DOE will maintain this database for HHS at ORNL in Tennessee. The Food and Drug Administration (FDA) and the Centers for Disease Control and Prevention within HHS are identifying research questions that can be answered through analyses by this partnership, and the NSC is providing logistics support and oversight.

The VA, in addition to clinical research expertise, will contribute data from patients with positive COVID-19 test results to this database. As of February 2021, the Veterans Health Administration (VHA) Corporate Data Warehouse had records of approximately 1 million patients tested for COVID-19 and over 110 000 who have tested positive. Initial analyses with VA data have begun using the existing VA computing resources at ORNL under previously established agreements.

Veterans who receive VHA healthcare services may also obtain services from other providers, including from the Department of Defense and from providers reimbursed by other health insurance programs, such as Medicare. Indeed, the VA has identified an additional 60 000 individuals as COVID-19 cases who do not have tests recorded within the VA Healthcare System. Negotiations are underway to add records for these patients to the COVID-19 Insights Collaboration Records Database from the Centers for Medicare and Medicaid Services and the Department of Defense's Military Health Information System.

Relying on the public health authority permitted the integration of records sets and analytic tools in a privacy appropriate environment. Protected health information that is collected by the federal government is restricted from sharing without patient authorization by privacy laws and regulations except in limited circumstances. Use of the public health authority is one of those exceptions. The agencies are also establishing a new "system of records" under the Privacy Act with a public notice in the *Federal Register*, preparing a privacy impact assessment under the E-Government Act, and have concluded data use agreements to set the rules for data sharing and retention.

The NSC has formed the COVID Insights Task Force to oversee the partnership's work. The task force has representatives of the VA, HHS, DOE, and NSC. The task force initially met daily and now meets twice a week to review requests from the FDA for analyses to address related public health questions of national importance. The task force determines the priority level of each request based on the latest trends in the COVID-19 pandemic, and it has the authority to determine which analyses will be conducted.

SECURE HIGH-PERFORMANCE COMPUTING ON IDENTIFIABLE HEALTH INFORMATION

The COVID-19 Insights Partnership leverages the experience gained from a collaboration formed in 2016 to combine the VA's health record system and MVP data with the DOE's high-performance computing infrastructure and expertise. Through this collaboration, known as MVP CHAMPION, a copy of the VA's corporate data warehouse data, consisting of billions of health, medication, and laboratory test records; physicians' notes; and genetic and survey data from the MVP have been housed at ORNL since MVP CHAM-

PION began. Access to the MVP CHAMPION resource has been granted to approved VA and DOE researchers for studies in high-priority research areas with clinical translational potential. For example, MVP CHAMPION is using VA and MVP data to develop risk-prediction tools for suicide, metastatic prostate cancer, and cardiovascular disease for use by VA clinicians along with clinical decision support tools.

MVP CHAMPION taught the VA and DOE how to bring their disparate research cultures together for innovative collaborative investigation. Furthermore, this collaboration has produced a cadre of VA and DOE scientists familiar with VA patient data and experienced in conducting joint research successfully. It is also leading to the integration of omics data with clinical data for a better mechanistic understanding of COVID-19 pathogenesis and potential therapies. Because of this collaboration between the VA and DOE, interagency teams were prepared at the start of the COVID-19 pandemic to develop the required agreements.^{1,3-5}

The COVID-19 Insights Collaboration Records Database will only have data deemed necessary by the Task Force to address COVID-19 pandemic questions. Although HHS will have ultimate responsibility for the dataset, the VA and any other agencies contributing data expect that the data will not be used for purposes beyond SARS-CoV-2 or COVID-19 public health activities.

The database resulting from the combination of records may contain personally identifiable information but is not designed to maintain direct identifiers such as names or social security numbers. The HIPAA Privacy Rule allows protected health information to be disclosed from a covered entity, such as the VHA, for public health activities to a public health authority, such as HHS. In addition, while in the custody of VHA, the records are covered by the Privacy Act of 1974, and require a separate authority to disclose under that law. The Privacy Act permits a federal agency to disclose Privacy Act records to parties outside the agency without the consent of the individual record subjects, for purposes compatible with the records' original collection purpose by establishing such a disclosure through an administrative process. The resulting disclosure is known by a term of art in the Privacy Act as a "routine use."

For this reason, as the nation's public health authority, the HHS Office of the Assistant Secretary for Health published a system of records notice specifically for this project, 09-90-2002 COVID-19 Insights Collaboration Records, at 85 FR 43242 (July 16, 2020), describing the HHS Privacy Act records that will be involved and the routine uses for which HHS may disclose them to non-HHS parties without the subject individuals' consent. This notice explains that the department has the authority to maintain the records under 42 U.S.C. §§ 214 and 247d.

To ensure the safety and confidentiality of the database, the VA is storing its data in a high-security computing facility in the United States. Other federal agencies that contribute data will also have this option. An institutional review board must review and approve all partnership studies to ensure that potential ethical and regulatory concerns are addressed. The uses of the partnership data are similar to the ways in which healthcare systems use their own data to conduct chart reviews and other retrospective studies with institutional review board approval and a waiver of HIPAA authorization.

The VA has recommended to HHS that all data released from the COVID-19 Insights Collaboration Records Database be de-identified (per the HIPAA definition) and that investigators from federal agencies (other than the DOE) or other institutions that have not provided data never have an opportunity to view identifiable forms of the data.

ENABLING DIVERSE STUDIES ON COVID-19

The COVID-19 Insights Partnership hopes to combine data on care provided to veterans inside and outside the VA. This enables investigators from any agency contributing data to understand the pandemic in new ways by bringing together data on health care received by veterans before, during, and after each COVID-19 episode. The partnership can also more efficiently identify modifiable mediators of COVID-19 healthcare outcomes and design appropriate interventions to address them. An important goal is to create generalizable knowledge from the data for translation beyond the VA population.

Using previously established interagency agreements and data governance, the partnership has already conducted research showing that the bradykinin storm is likely to play a role in many COVID-19 symptoms¹ and, using VA data, compared COVID-19 testing patterns, positive test results, and 30-day mortality rates by race and ethnicity among VA patients.² Other recently completed studies have developed and validated short-term mortality indices in individuals with COVID-19 based on their preexisting conditions,³ assessed the generalizability of VA COVID-19 experiences to the U.S. population, and evaluated the effectiveness of hydroxychloroquine with and without azithromycin in VA patients with COVID-19. The most recent study demonstrates the benefit of prophylactic anticoagulation at initial hospitalization.⁵ Ongoing studies are evaluating experiences with steroid therapies and with antibody tests, and in collaboration with Centers for Disease Control and Prevention, are beginning to explore the long-hauler syndrome. As COVID-19 infections continue, the partnership will conduct public health analyses on the longer-term complications of this infection as well as on a "steady-state" approach to curbing infection that will include vaccination programs.

The VA is also providing the FDA with daily reports on aggregate VA COVID-19 cases and their distribution across the VA system, demographics of VA patients with COVID-19, and analyses of predictive models for positive test results and death. The VA also sends to the FDA on a regular basis aggregated data showing patterns of medication use and retrospective analyses of the effectiveness of certain medications (including remdesivir and some antithrombotic agents). The FDA can use these data to understand the scope of the pandemic and to predict drug shortages or needs for additional medical equipment, including ventilators. This information was critical at the start of the pandemic, and it might become important again in the event of further waves of the pandemic.

Establishing the Insights Collaboration Records Database and integrating relevant datasets from other federal agencies will accelerate public health advancements related to COVID-19.

LAYING THE FOUNDATION FOR FUTURE FEDERAL DATA-SHARING INITIATIVES

The DOE's Summit, the fastest supercomputer in the United States, will help speed this research and make new types of research possible by running complex analyses on the vast health data in the COVID-19 Insights Collaboration Records Database. Summit, housed at ORNL, can complete 200 000 trillion calculations per second at 64-bit precision and 2.41×10^{18} calculations per second at 16-bit precision.⁴ This supercomputer's ability to analyze massive integrated datasets will make it possible to identify rapidly and advance potential COVID-19 treatment and prevention strategies and improve outcomes for patients who have the disease. Such analyses, applying the most advanced and powerful artificial intelligence

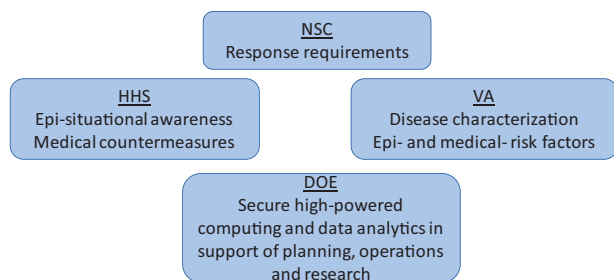


Figure 1. Schematic organization of component parts of COVID-19 Insights Partnership. DOE: U.S. Department of Energy; HHS: U.S. Department of Health and Human Services; NSC: National Security Council; VA: U.S. Department of Veterans Affairs.

methods and leveraging the partnership's shared expertise and robust infrastructure, would not be possible in more conventional research settings.

The VA electronic health record data, which capture longitudinal care information on veterans, are an equally valuable resource. The VHA has records for some patients that go back decades. Furthermore, the VA covers costs of medications and provides a death benefit. As a result, VA data include medications used by patients before, during, and after each COVID-19 episode. Similarly, the VA has complete mortality data on its patients, whereas other large health systems do not capture mortality events after patients leave the hospital.

Lessons learned from the model established through the COVID-19 Insights Partnership include the following:

- The necessity of establishing interagency programs and data integration ahead of need
- The necessity of clear and rapid communication among all parties including investigators, public health authorities, regulatory and privacy communities, and policymakers
- The utility of high-performance computing and artificial intelligence for rapid, complex, real-world data analyses
- The power of multiagency interdisciplinary teams that can unite mechanistic and clinical perspectives
- The continued importance of clinical knowledge and insight to differentiate real from artifactual associations identified using machine learning, statistical, and epidemiological techniques

These lessons will help the nation prepare nimble responses to the ongoing COVID-19 pandemic as well as to the inevitable next epidemic, especially to protect those who are most vulnerable. The partnership will help determine how new artificial intelligence techniques can best complement traditional statistical and epidemiologic methods for real-world data analyses. Efforts will also be needed to ensure rapid implementation of the lessons from this partnership.

This regulatory pathway for sharing data under the HIPAA Privacy Rule's public health authority should be considered in any ongoing or future public health emergency with HHS support. The ability to leverage medical data from federal datasets to address critical public health questions developed through the COVID-19 Insights Partnerships will be valuable in the future whenever the nation faces a rapidly evolving epidemic or other public health emergency.

FUNDING

This work was supported in part by the Million Veteran Program (MVP-000), U.S. Office of Research and Development, Veterans Health Administration, U.S. Office of E-Government and Information Technology, and U.S. Department of Veterans Affairs. DAJ and BHM were supported by MVP CHAMPION and the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory and Los Alamos National Laboratory, respectively, managed by UT-Battelle, LLC, for the U.S. Department of Energy (LOIS: 10074).

AUTHOR CONTRIBUTIONS

All authors made substantial contributions to the conception of the work, participated in revising the manuscript for important intellectual content and have approved the final version of the manuscript. All questions related to the accuracy or integrity of any part of the work have been appropriately investigated and resolved. ACJ is the corresponding author.

ACKNOWLEDGMENTS

The authors acknowledge writing support provided by Deborah Berlyne and Mitchell Mirkin in the preparation of this manuscript and federal privacy expertise provided by Karen Jeans, Gi An, Stephanina Griffin, Michelle Christiano, Beth Kramer, Brooke Dickson, Kirsten Moncada, and Rahwa Keleta.

CONFLICT OF INTEREST STATEMENT

None of the authors have any competing interests to declare.

DATA AVAILABILITY STATEMENT

This manuscript has been coauthored by UT-Battelle, LLC, under contract no. DE-AC05-00OR22725 with the U.S. Department of Energy. The U.S. government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. government retains a nonexclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for U.S. government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the Department of Energy Public Access Plan (<http://energy.gov/downloads/doe-public-access-plan>).

REFERENCES

1. Garvin MR, Alvarez C, Miller JI, *et al.* A mechanistic model and therapeutic interventions for COVID-19 involving a RAS-mediated bradykinin storm. *Elife* 2020; 9: e59177.
2. Rentsch CT, Kidwai-Khan F, Tate JP, *et al.* Patterns of COVID-19 testing and mortality by race and ethnicity among United States veterans: a nationwide cohort study. *PLoS Med* 2020; 17 (9): e1003379.
3. King JT, Yoon JS, Rentsch CT, *et al.* Development and validation of a 30-day mortality index based on pre-existing medical administrative data from 13,323 COVID-19 patients: The Veterans Health Administration COVID-19 (VACO) Index. *PLoS One* 2020; 15 (11): e0241825.
4. Joubert W, Weighill D, Kainer D, *et al.* Attacking the opioid epidemic: determining the epistatic and pleiotropic genetic architectures for chronic pain and opioid addiction. In: *SC18: International Conference for High Performance Computing, Networking, Storage and Analysis*; 2018: 717–30. IEEE.
5. Rentsch CT, Beckman JA, Tomlinson L, *et al.* Early initiation of prophylactic anticoagulation for prevention of COVID-19 mortality: a nationwide cohort study of hospitalized patients in the United States. *BMJ* 2021; 372: n311