Leveraging the Million Veteran Program Infrastructure and Data for a Rapid Research Response to COVID-19

Stacey B. Whitbourne, PhD[a,b,c]; Jennifer Moser, PhD[d]; Kelly Cho, PhD, MPH[a,b,c]; Jennifer Deen[d]; Lori L. Churby[e]; Amy C. Justice, MD, PhD[f,g]; Juan P. Casas, MD, PhD[h]; Saiju Pyarajan, PhD[a]; Phil S. Tsao, PhD[e,i]; J. Michael Gaziano, MD, MPH[a,b,c]; Sumitra Muralidhar, PhD[d]

Background: The Veterans Health Administration Office of Research and Development (ORD) played a key role in the federal government's response to the COVID-19 pandemic. The ORD effectively leveraged existing resources to answer questions related to the SARS-CoV-2 virus and COVID-19.
Observations: When the COVID-19 pandemic hit in 2020, the Million Veteran Program (MVP), one of the largest genomic cohorts in the world, extended the centralized recruitment and enrollment infrastructure to develop a COVID-19 research volunteer registry to assist enrollment in the vaccine and treatment trials in which the US Department of Veterans Affairs (VA) participated. In addition, the MVP allowed for new data collection and a large genomic cohort to understand host contributions to COVID-19. This article describes ways the MVP contributed to the VA's rapid research response to COVID-19. Several host genetic factors believed to play a role in the development and severity of COVID-19 were identified. Furthermore, existing MVP partnerships with other federal agencies, particularly with the Department of Energy, were leveraged to improve understanding and management of COVID-19.
Conclusions: A previously established enterprise approach and research infrastructure were essential to the VA's successful and timely COVID-19 research response. This infrastructure not only supported rapid recruitment in vaccine and treatment trials, but also leveraged the unique MVP and VA electronic health record data to drive rapid scientific discovery and inform clinical operations. Extending the models that VA research applied to the federal government at large and establishing centralized resources for shared or federated data analyses across federal agencies will better equip the nation to respond to future public health crises.

The Million Veteran Program (MVP) was launched in 2011 by the US Department of Veterans Affairs (VA) to enroll at least 1 million veterans in a longitudinal cohort to better understand how genes, lifestyle, military experience, and environmental exposures interact to influence health and illness and ultimately enable precision health care. The MVP has established a national, centralized infrastructure for recruitment and enrollment, biospecimen and data collection and storage, data generation and curation, and secure data access. When the COVID-19 pandemic hit in 2020, the MVP was leveraged to support research utilizing the following key infrastructure components: (1) MVP recruitment and enrollment platform to provide support for COVID-19 vaccine and treatment trials and to collect COVID-19 data from MVP participants; (2) using MVP Phenomics for COVID-19 research data cleaning and curation, assisting with the development of a VA Severity Index for COVID-19, and forming 6 scientific working groups to coordinate COVID-19 research questions; and (3) the VA/MVP and US Department of Energy (DOE) partnership to assist in responding to COVID-19 research questions identified by the US Food and Drug Administration (FDA). This article describes these infrastructure components in more detail and highlights key findings from the MVP COVID-19 research efforts.

MVP INFRASTRUCTURE

The Veterans Health Administration (VHA) Office of Research and Development (ORD) oversaw efforts to develop the VA Coronavirus Research Volunteer List (the COVID-19 registry). To support the registry, the MVP leveraged its infrastructure to facilitate a rapid response. The MVP is designed as a full-service and centralized recruitment and enrollment platform. This includes MVP office oversight; MVP coordinating centers that manage the centralized platform; an information center that handles inbound and outbound calls; an informatics system built for recruitment and enrollment monitoring and tracking; and a network of more than 70 participating MVP sites with dedicated staff to conduct recruitment and enrollment activities. The MVP used its informatics infrastructure to support secure data storage for the registry volunteer information. MVP coordinating center staff worked with the COVID-19 registry to invite > 125,000 MVP participants from approximately 20 MVP sites. Additionally, MVP information center staff made > 4000 calls to prospective registry volunteers. This work resulted in 1300 volunteers agreeing to be contacted by COVID-19 vaccine clinical trial study teams (including Moderna, Janssen, AstraZeneca, and Novavax). About 20 MVP site staff (spanning 14 MVP sites) also were deployed to support COVID-19 work for clinical care capabilities or vaccine trials.

Author affiliations can be found at the end of this article.
Correspondence: Sumitra Muralidhar (sumitra.muralidhar@va.gov)

Fed Pract. 2023;40(suppl 5). Published online November 1. doi:10.12788/fp.0416

New Data Collection

The MVP protocol was approved by the VA Central Institutional Review Board (IRB) in 2011. As part of initial enrollment in MVP, participants consented to recontact for additional self-report information along with access to their electronic health record (EHR). This allows for the linkage of EHR and survey response data, thus providing a comprehensive understanding of health history before and after a self-reported COVID-19 diagnosis. Between May 2020 and September 2021, the MVP COVID-19 survey was distributed to existing MVP participants via mail, telephone, and email with the ability to complete the survey by paper and pencil or through the MVP online system. Dissemination of the survey was approved by the VA Central IRB in 2020, with nearly 730,000 eligible MVP participants contacted. As of June 2022, 255,737 MVP participants (35% of the eligible cohort) had completed the survey; 86% completed a paper survey while 14% completed it online. Respondents were primarily older (≥ 65 years); 90% were male; close to 7% reported Hispanic ethnicity, and 11% reported Black race.

Findings from this survey provide insight into pandemic behaviors not consistently captured in EHRs, such as psychosocial aspects, including social and emotional support, loss of tangible and intangible resources, as well as COVID-19–related behaviors, such as social distancing and self-protective practices.[1] MVP COVID-19 survey data combined with veteran EHRs, responses to other MVP surveys, and genetic data enable MVP researchers to better understand epidemiological, clinical, and psychosocial aspects of the disease. Future COVID-19 studies may use self-reported survey responses to enrich understanding about the effects of the disease on a veteran's daily life, and possibly validate existing EHR COVID-19 diagnoses and hospitalization findings. This comprehensive data resource provides a unique opportunity to identify new targets for disease prevention, treatment, and management with an emphasis on individual variability in genes, environment, and lifestyle.

COVID-19 RESEARCH

In early 2020, the burden of COVID-19 on the US was unprecedented, and little was known about risk factors for severe COVID-19 and deaths. The MVP Phenomics team quickly responded with a large-scale phenome-wide association study (PheWAS) of > 1800 phenotypes (physical and biochemical traits) and COVID-19 progression. Its goal was to characterize risk factors and outcomes associated with COVID-19 disease progression.[2] Data curation and assembly occurred rapidly through integrated efforts led by MVP and VA COVID-19 initiatives. The MVP utilized its phenomics core resource to understand the progression of COVID-19 defined by SARS-CoV-2 infection, hospitalization, intensive care unit admission, and 30-day mortality using VA EHR data.

To broaden disease progression data curation and fit the specific needs of the VA, we operationalized and validated the World Health Organization clinical severity scale and used VA EHR data to create the VA Severity Index for COVID-19 (VASIC).[3] The VASIC category is now part of the MVP core data repository, where volumes of data from multiple activities are integrated through an automated process to create monthly research-ready data cubes. These activities include extensive data curation, mapping, phenotyping, and adjudication that are performed to curate oxygen supplementation status and other procedures related to treatment that are processed and understood in real time. The data cubes were provisioned to MVP COVID-19 researchers. In addition, the VASIC scale variable is now integrated within the larger VA system for all researchers to use as part of its wider COVID-19 initiative. The VA Centralized Interactive Phenomics Resource (CIPHER) phenomics library now hosts the details of VASIC, codes, metadata, and related COVID-19 data products for all VA communities. In partnership with CIPHER and other internal and external COVID-19 initiatives, the MVP continues to play an integral part for the VA and beyond in the development of a phenomics algorithm for long COVID, or post-acute COVID-19 syndrome (PACS).

Host Genetics in COVID-19

As the SARS-CoV-2 virus continued to spread globally, it became clear that the symptoms and severity of infection experienced by patients varied across a broad spectrum, from being asymptomatic carriers to experiencing severe symptoms in 1 or more organ systems in the body, resulting in death. This variability suggested that host genetics and other host factors may play a role in determining the severity of COVID-19 infection. The MVP dataset, with genetic and health information on > 600,000 MVP participants, provided an ideal dataset to explore host contributions to COVID-19.

In late spring 2020, the MVP executive committee issued a call to the MVP research community to propose study aims around the COVID-19 pandemic that could leverage the phenotypic and genetic data and resources. The MVP quickly formed 6 rapid-response scientific working groups. Their mission was to cultivate collaboration and inclusivity and to coordinate COVID-19 research questions. A

TABLE MVP COVID-19 Working Group Primary Objectives, Key Findings, and Implications

Working group	Primary objectives	Key findings	Implications
Druggable genome	Examine drug-repurposing opportunities for COVID-19 prevention; hospitalized patient severity/complications	Angiotensin-converting enzyme 2, IFNAR2, and IL10RB identified as potential targets for early COVID-19 management[4]	Among the first to determine the role of IFN as part of the immune response to COVID-19
Pharmaco-genomics	Risk variants of AKI and mortality among Black veterans with COVID-19; remdesivir-induced liver chemistry abnormalities pharmacogenetic analysis	MVP Black participants hospitalized for COVID-19 with specific APOL1 variants associated with higher risks of AKI, AKI severity, and death, even for individuals with prior normal kidney function[5]	APOL1 variants associated with increased risk of AKI among Black veterans hospitalized with COVID-19
Disease mechanisms	Specific genetic markers and effects on COVID-19 including: polygenic predisposition to VTE; COVID-19–positive SCT carrier outcomes; mucin 5B, oligomeric mucus/gel-forming gene polymorphism; COVID-19 infection protective effects	SCT associated with increased death following COVID-19[6]; common variant in a mucoprotein gene associated with idiopathic pulmonary fibrosis (present in about 20% of the population) associated with fewer COVID-19 infection hospitalizations and pneumonia events[7]	SCT (present in 8% of Black individuals) is complicating factor in COVID-19 infection and associated with increased death following COVID-19; common variant in a mucoprotein gene associated with fewer hospitalizations and pneumonia events COVID-19 infection and may be protective
Genomics for risk prediction, PRS, and Mendelian randomization	How genetic information can enhance COVID-19 risk prediction including: blood group typing and protective effects on COVID-19 infection; PRS/human leukocyte antigen typing and COVID-19 outcomes; transcriptome-wide association study of COVID-19–positive participants	In 20,000 COVID-19 cases of a racially diverse group of MVP participants, 4 independent genetic variants identified that contributed to COVID-19 positivity, including a novel locus exclusively among Hispanic patients[8]; confirmation that A, AB, and B blood types associated with COVID-19 positivity	Evidence of racial and ethnic differences in COVID-19 positivity rates; blood types are associated with COVID-19 risk
GWAS and downstream analysis	GWAS of the main COVID-19 outcomes	New genetic loci to suggest further investigation on these candidate genes[8]	New genetic loci used by other working groups/external collaborations; workgroup combined with other workgroups to explore racial/ethnic differences in COVID-19 positivity rates
COVID-19–Related Phenome-Wide Association Study	Develop a resource to understand genetic variants associated with susceptibility to or outcomes of COVID-19 infection	Using MVP and COVID-19 Host Genetics Initiative[9] data: identified traits that share genetic variants associated with severe COVID-19; most COVID-19 severity variants associated with risk factors or COVID-19 complications, such as venous embolism and thrombosis; subgroup of variants was inversely associated with psoriasis and inflammatory lung conditions; variant linked with severe COVID-19 was associated with neutropenia among veterans of African ancestry; older age, higher body mass index, male patients, and patients with a history of respiratory, kidney, bacterial, or metabolic comorbidities experienced greater COVID-19 severity	COVID-19 severity variants associated with risk factors or complications of COVID-19, like venous embolism and thrombosis; multiple genetic variants linked with both COVID-19 severity and other conditions such as circulatory issues, inflammatory conditions, and neutropenia; future studies needed to investigate the underlying mechanisms associated with these phenotype clusters and COVID-19

Abbreviations: AKI, acute kidney injury; APOL1, apolipoprotein L1; GWAS, Genome-Wide Association Study; IFN, interferon; IFNAR2, interferon α/β receptor 2; IL10RB, interleukin 10 receptor β subunit; MVP, Million Veteran Program; PRS, Polygenic Risk Scores; SCT, sickle cell trait; VTE, venous thromboembolism.

steering committee composed of the MVP executive committee, staff from computational environments, working group cochairs, and an administrator, who was responsible for daily oversight of the working groups. In addition, the ORD COVID-19 steering committee reviewed and approved research activities to ensure scientific rigor, as well as alignment with overall ongoing research activities.

The MVP COVID-19 working groups included dozens of researchers who used MVP data to identify disease mechanisms; understand the impact of host genetics on susceptibility, morbidity, and mortality; and identify potential targets for treatments and therapies. The working groups were further supported by MVP analysts to work cross-functionally on genomics, phenomics, statistical genetics, and PheWAS. Each working group chair was responsible for prioritizing concepts and moving them forward in coordination with the MVP and ORD COVID-19 steering committees. An overview of the MVP COVID-19 working groups follows (Table).[4-9]

Druggable genome. This working group researched drug-repurposing opportunities to prevent severe COVID-19, defined as hospitalization with oxygen therapy (high flow), intubation, mechanical ventilation, vasopressors, dialysis, or death from COVID-19; and prevent complications in patients hospitalized by COVID-19.

Pharmacogenomics. This working group focused on 2 main aims: the impact of apolipoprotein L1 risk variants on acute kidney injury (AKI) and death in Black veterans with COVID-19; and pharmacogenetic analysis of remdesivir-induced liver chemistry abnormalities.

Disease mechanisms. Understanding the underlying pathways and mechanisms behind COVID-19 has been a difficult but important challenge overall in the scientific community. This working group investigated specific genetic markers and effects on COVID-19, including polygenic predisposition to venous thromboembolism associated with increased COVID-19 susceptibility; renal comorbidities and new AKI and unfavorable outcomes among COVID-19–positive sickle cell trait carriers; and mucin 5B, oligomeric mucus/gel-forming gene polymorphism, and protective effects in COVID-19 infection.

Genomics for risk prediction, polygenic risk scores, and mendelian randomization. Risk prediction for COVID-19 has been widely studied mostly aiming at comorbidities and preexisting conditions. The MVP cohort provided a unique opportunity to understand how genetic information can enhance our understanding of COVID-19 risk. This working group focused on: (1) ABO blood group typing and the protective effects of the O blood group on COVID-19 infection; (2) polygenic risk scores and COVID-19 outcomes; (3) human leukocyte antigen typing and COVID-19 outcomes; and (4) a transcriptome-wide association study of COVID-19–positive MVP participants.

Genome-Wide Association Study (GWAS) and Downstream Analysis. This working group performed GWAS of the main COVID-19 outcomes. Results from GWAS unveiled new genetic loci to suggest further investigation on these candidate genes. The results were used by other MVP COVID-19 working groups for their activities. The results also contributed to external collaborations, such as the COVID-19 Host Genetics Initiative.

COVID-19–Related PheWAS. This working group focused on understanding the potential clinical significance of genetic variants associated with susceptibility to, or outcomes of, COVID-19 infection. They worked to identify traits that share genetic variants associated with severe COVID-19 from the Host Genetics Initiative. The group also studied the phenotypic consequences of acquired mosaic chromosomal alterations with early data linking to COVID-19 susceptibility.

COVID-19 Research Partnerships

In 2016, the VA and DOE formed an interagency partnership known as Computational Health Analytics for Medical Precision to Improve Outcomes Now (CHAMPION) to demonstrate the power of combining the VA EHR system, MVP genetic data, and clinical research expertise with DOE high-performance computing infrastructure and artificial intelligence expertise. The VA EHR captures longitudinal care information on veterans with records that go back decades. Furthermore, the VA covers the costs of medications and provides a variety of services through the Veterans Benefits Administration. As a result, VA data include medications used by patients before, during, and after COVID-19. Similarly, the VA has comprehensive vital records, whereas other large health systems do not capture events such as death after patients leave the hospital.

The DOE Oak Ridge National Laboratory (ORNL) in Tennessee securely maintains this rich database for the VA. The ORNL Summit supercomputer can complete trillions of calculations per second to provide critical and timely analyses, applying the most advanced and powerful artificial intelligence methods, which would not be possible in more conventional research settings. CHAMPION taught the VA and DOE how to bring their disparate research cultures together for innovative collaborative investigation. Moreover, this collaboration produced a cadre of VA and DOE scientists familiar with VA patient data and experienced in conducting joint research successfully and integrating omics data with clinical data for a better mechanistic understanding. Because of this preexisting collaboration between the VA and DOE, interagency teams were prepared at the start of the COVID-19 pandemic.[10-15]

During the pandemic, the FDA and VA conducted research together. One joint study found that the bradykinin storm is likely to play a role in many COVID-19 symptoms. Using VA data, researchers compared COVID-19 testing patterns, positive test results, and 30-day mortality rates by race and ethnicity among VA patients.[10,11] These findings demonstrated the higher burden COVID-19 placed on Black and Hispanic communities, not fully explained by underlying

health conditions, access to medical care, or geographic locale.[11]

Other recently completed studies have developed and validated short-term mortality indices in individuals with COVID-19 based on their preexisting conditions, assessed the generalizability of VA COVID-19 experiences to the US population, and evaluated the effectiveness of hydroxychloroquine with and without azithromycin in VA patients with COVID-19.[12,15] A recent study demonstrated the benefit of prophylactic anticoagulation at initial hospitalization.[14]

The VA also provided 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 regularly sent the FDA aggregated data showing patterns of medication use and retrospective analyses of the effectiveness of certain medications (including remdesivir and some antithrombotic agents). The FDA used these data along with other 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.

Limitations

For the most part, MVP infrastructure and partnerships were efficiently leveraged to significantly advance our understanding of the biological basis of COVID-19 and to develop treatments and vaccines. However, there were a few limitations that may have slowed timely and optimal outcomes. An issue not limited to the MVP or VA was the continual evolution of the pandemic and its response. This included evolving definitions of disease, symptomatology, testing, vaccines, and public health recommendations. Keeping pace with the emerging knowledge from these domains was a struggle for the entire scientific community. A more discrete limitation was the number of participants in the MVP with positive COVID-19 test results and positive symptoms; however, this was mitigated by partnering with other groups like the COVID-19 Host Genetics Initiative to increase study participant numbers. Finally, there were logistical and regulatory challenges associated with coordination of national clinical trial recruitment across a VA system with > 100 discrete hospitals.

CONCLUSIONS

Having a centralized infrastructure for recruitment and enrollment, including a national research volunteer registry, information center, research staff, and coordinating centers, can allow for expedited enrollment in vaccine and treatment trials in the face of future public health emergencies. VA assets, including its rich EHR and MVP, the world's largest genomic cohort, have contributed to improving our understanding and management of COVID-19. MVP's ready-to-respond research infrastructure embedded within the country's largest national health care system allows for both the facilitation of the research work and applications of the research findings into practice. Findings from the MVP COVID-19 working groups have yielded compelling results, particularly around genetic variants among various racial and ethnic groups. Looking ahead, the VA and DOE are launching a new joint project on long COVID that will include developing a gold-standard definition for long COVID. The ORD has established a Partnered Research Program to facilitate collaborations with industry to speed up clinical trials, and the MVP will continue to contribute toward expanding scientific knowledge to improve the management of COVID-19.

Author affiliations

[a]Veterans Affairs Boston Healthcare System, Massachusetts
[b]Brigham and Women's Hospital, Boston, Massachusetts
[c]Harvard Medical School, Boston, Massachusetts
[d]Office of Research and Development, Department of Veterans Affairs, Washington, DC
[e]Veterans Affairs Palo Alto Healthcare System, California
[f]Veterans Affairs Connecticut Healthcare System, West Haven
[g]Yale University School of Medicine and School of Public Health, New Haven, Connecticut
[h]Novartis Institute for Biomedical Research, Cambridge, Massachusetts
[i]Stanford University School of Medicine, Palo Alto, California

Author disclosures

The authors report no actual or potential conflicts of interest or outside sources of funding with regard to this article.

Ethics and consent

Not applicable.

References

1. Whitbourne SB, Nguyen XT, Song RJ, et al. Million Veteran Program's response to COVID-19: survey development and preliminary findings. *PLoS One*. 2022;17(4):e0266381. doi:10.1371/journal.pone.0266381
2. Song RJ, Ho YL, Schubert P, et al. Phenome-wide as-

sociation of 1809 phenotypes and COVID-19 disease progression in the Veterans Health Administration Million Veteran Program. *PLoS One*. 2021;16(5):e0251651. doi:10.1371/journal.pone.0251651
3. Galloway A, Park Y, Tanukonda V, et al. Impact of COVID-19 severity on long-term events in US veterans using the Veterans Affairs Severity Index for COVID-19 (VASIC). *J Infect Dis*. 2022;226(12):2113-2117. doi:10.1093/infdis/jiac182
4. Gaziano L, Giambartolomei C, Pereira AC, et al. Actionable druggable genome-wide Mendelian randomization identifies repurposing opportunities for COVID-19. *Nat Med*. 2021;27(4):668-676. doi:10.0138/s41591-021-01310-z
5. Hung AM, Sha SC, Bick AG, et al. APOL1 risk variants, acute kidney injury, and death in participants with African ancestry hospitalized with COVID-19 from the Million Veteran Program. *JAMA Intern Med*. 2022;182(4):386-395. doi:10.1001/jamainternmed.2021.8538
6. Verma A, Huffman JE, Gao L, et al. Association of kidney comorbidities and acute kidney failure with unfavorable outcomes after COVID-19 in individuals with the sickle cell trait. *JAMA Intern Med*. 2022;182(8):796-804. doi:10.1001/jamainternmed.2022.2141
7. Verma A, Tsao NL, Thomann LO, et al. A phenome-wide association study of genes associated with COVID-19 severity reveals shared genetics with complex diseases in the Million Veteran Program. *PLoS Genet*. 2022;18(4):e1010113. doi:10.1371/journal.pgen.1010113
8. Peloso GM, Tcheandjieu C, McGeary JE, et al. Genetic loci associated with COVID-19 positivity and hospitalization in White, Black, and Hispanic Veterans of the VA Million Veteran Program. *Front Genetic*. 2022;12:777076. doi:10.3389/fgene.2021.777076
9. Verma A, Minnier J, Wan ES, et al. A MUC5B gene polymorphism, rs35705950-T confers protective effects against COVID-19 hospitalization but not severe disease or mortality. *Am J Respir Crit Care Med*. 2022;182(8):796-804. doi:10.1164/rccm.202109-2166OC
10. 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;e59177. doi:10.7554/eLife.59177
11. 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. doi:10.1371/journal.pmed.1003379
12. 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. doi:10.1371/journal.pone.0241825
13. 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. SC18: International Conference for High Performance Computing, Networking, Storage and Analysis. Dallas, TX, USA, 2018:717-730. doi:10.1109/SC.2018.00060
14. 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. doi:10.1136/bmj.n311
15. Gerlovin H, Posner DC, Ho YL, et al. Pharmacoepidemiology, machine learning, and COVID-19: an intent-to-treat analysis of hydroxychloroquine, with or without Azithromycin, and COVID-19 outcomes among hospitalized US Veterans. *Am J Epidemiol*. 2021;190(11): 2405-2419. doi:10.1093/aje/kwab183