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.
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. 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. 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). 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.
steering committee composed of the MVP executive committee, staff from computational environments, working group co-chairs, 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; under-
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.

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. 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.\textsuperscript{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.\textsuperscript{12,15} A recent study demonstrated the benefit of prophylactic anticoagulation at initial hospitalization.\textsuperscript{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, syndromatology, 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
\textsuperscript{a}Veterans Affairs Boston Healthcare System, Massachusetts
\textsuperscript{b}Brigham and Women’s Hospital, Boston, Massachusetts
\textsuperscript{c}Harvard Medical School, Boston, Massachusetts
\textsuperscript{d}Office of Research and Development, Department of Veterans Affairs, Washington, DC
\textsuperscript{e}Veterans Affairs Palo Alto Healthcare System, California
\textsuperscript{f}Veterans Affairs Connecticut Healthcare System, West Haven
\textsuperscript{g}Yale University School of Medicine and School of Public Health, New Haven, Connecticut
\textsuperscript{h}Novartis Institute for Biomedical Research, Cambridge, Massachusetts
\textsuperscript{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.

Disclaimer
The opinions expressed herein are those of the authors and do not necessarily reflect those of Federal Practitioner, Frontline Medical Communications Inc., the US Government, or any of its agencies. This article may discuss unlabeled or investigational use of certain drugs. Please review the complete prescribing information for specific drugs or drug combinations—including indications, contraindications, warnings, and adverse effects—before administering pharmacologic therapy to patients.

Ethics and consent
Not applicable.

References
2. Song RJ, Ho YL, Schubert P, et al. Phenome-wide as-


