VA Big Data Science: A Model for Improved National Pandemic Response Present and Future

Yinong Young-Xu, ScD, MA, MS[a,b]; Victoria Davey, PhD, MPH[c]; Vincent C. Marconi, MD[d,e]; Francesca E. Cunningham, PharmD[f]

Background: The US Department of Veterans Affairs (VA) enterprise approach to research (VA Research) has built a data-sharing framework available to all research teams within VA. Combined with robust analytic systems and tools available for investigators, VA Research has produced actionable results during the COVID-19 pandemic. Big data science techniques applied to VA's health care data demonstrate that medical research can be performed quickly and judiciously during nationwide health care emergencies.

Observations: We envision a common framework of data collection, management, and surveillance implemented in partnership with other health care agencies that would capture even broader, actionable, and timely observational data on populations, while providing opportunities for enhanced collaborative research across agencies. This model should be continued and expanded through the current COVID-19 and future pandemics.

Conclusions: Extending the achievements of VA Research in the COVID-19 pandemic to date, we advocate national goals of open science by working toward a synergistic national framework of anonymized, synchronized, shared health data that would provide researchers with potent tools to combat future public health crises.

The COVID-19 pandemic emphasized the need for rapid response research in health care. The robust enterprise approach used by the US Department of Veterans Affairs (VA), termed VA Research, is meeting these needs by using existing outstanding data resources and interdisciplinary collaborations.[1] In the first 7 months of 2021 alone, while many US health care systems struggled with limited data, VA Research published more than 300 unique and instrumental research papers addressing urgent questions about transmission, vaccination, therapeutics, and health impacts of COVID-19 on its high-risk population.[1] The ability to leverage the VA electronic health record (EHR) and Corporate Data Warehouse (CDW)—a fully established data system bringing together test results, prescriptions, and complete patient health records, readily accessible and updated daily—was substantial.

With more than 9 million veterans enrolled in care at 171 medical centers and 1113 outpatient facilities across the US and its territories, the CDW provides an unprecedented opportunity to examine outcomes in real time. This allowed research groups such as the VA St Louis Health Care System Research and Education Service to build a cohort of 181,280 veterans with diabetes and positive COVID-19 test results within a 6-month period in 2021 to study the incidence of new diagnoses of diabetes after COVID-19 infection.[2] Similarly, the Clinical Epidemiology Program (CEP) at VA White River Junction Health Care System built a cohort of 1,363,180 veterans who received at least 1 COVID-19 vaccine by March 7, 2021, to analyze coverage and effectiveness of those vaccines.[3] This time-sensitive research was possible because the VA had the data and tools in place. Moreover, the the CEP quickly built an infrastructure to make its cohort and programming codes available to researchers in and outside the VA, resulting in additional significant research.[4]

The innovation and speed of COVID-19 vaccine development and distribution in the US were unprecedented. The rapid discovery and implementation of multiple preventives and therapeutics for COVID-19 could not have been possible without shared information within a competitive industry. VA studies added significantly to understanding the clinical performance of the messenger RNA (mRNA) COVID-19 vaccines, antivirals, and monoclonal treatments in a real-world setting. For example, a vaccine coverage study by VA Research illustrated how successful vaccination for COVID-19 at the VA has been in protecting a diverse community of patients from hospitalization and death, particularly the highly comorbid, racial and ethnic minorities, and other high-risk populations.[3] The study demonstrated the power of the VA system to generate robust and compelling clinical endpoint effectiveness data across a broad range of high-risk groups.

This success is promising. However, the COVID-19 pandemic is not over, and the next could prove even more challenging. For example, through a recent partnership with the US Department of Defense (DoD), the VA was able to rapidly analyze the effectiveness of previous smallpox vaccination efforts in the

Author affiliations can be found at the end of this article.
Correspondence: Yinong Young-Xu (yinong.young-xu@va.gov)

Fed Pract. 2023;40(suppl 5). Published online November 1. doi:10.12788/fp.0412

military for preventing mpox infections.[5] We should take this opportunity to think creatively about ways to improve our existing infrastructure based on what we have learned.

A ROLE FOR VA RESEARCH IN EFFICACY

The US Food and Drug Administration (FDA) Reauthorization Act of 2017 requires that manufacturers submit evidence establishing a product's benefits (effectiveness) outweigh its risks (safety) before it can be promoted and distributed.[6] As such, the FDA has been obligated by external stakeholders and Congress to be more explicit and transparent about benefit-risk profile supporting its decisions on licensure. This process led to requiring more phase 4 postmarketing observational studies for safety and effectiveness.[7] Although the FDA postlicensure system remains vigilant toward safety, effectiveness information is limited due to insufficient reporting (with exceptions of manufacturer studies for new indications or to exhibit superior comparative effectiveness). The agency typically relies on a static set of efficacy data generated prelicensure with a dynamic and evolving set of safety data accrued postlicensure to support its assessment that benefits outweigh risks.

For example, operating in near real time, postauthorization safety monitoring systems, led by the Centers for Disease Control and Prevention and other federal systems, identified a safety signal for thrombosis following the Janssen COVID-19 vaccination. Distribution was quickly paused, the safety signal was investigated, the magnitude of the risk was characterized, new language describing the risk and providing guidance regarding clinical management was included in labeling, and distribution was resumed, all within a few weeks. This remarkable success demonstrated how timely the safety system can operate to evaluate risk.

In contrast, the duration and extent of protection against COVID-19 variants are largely limited to the assessment of immune biomarker surrogates. Such clinical effectiveness data are urgently needed for the FDA's Center for Biologics Evaluation and Research and Center for Drug Evaluation and Research to make accurate benefit-risk assessments and continue to conclude the balance is favorable. As we prepare for the next pandemic, we must consider plans for monitoring postauthorization/postlicensure effectiveness as well as safety in real time. VA Research is ideally situated for this task.

Published studies on effectiveness at the VA serve as a prototype and could lead the way to initiating those preparations.[4,8-11] One of the striking features of the VA system that became apparent in the preparation of the mRNA vaccine study was the speed at which an enormous volume of COVID-19 testing data were produced. This enabled implementation of methodologically sound test-negative and case-control analysis. Analyses sufficiently powered to conclude mRNA vaccines were highly effective when used in real-world conditions among a diverse population from nearly every state and territory during a period in which multiple COVID-19 variants were already circulating.[3] This is unique to the VA and would not be possible for any other US health care system. With planning, the VA system could produce product-specific, real-world evidence of effectiveness comparable to the timeliness and quality of the safety data currently produced to support regulatory benefit-risk assessments. For example, the VA conducted an effectiveness study of tixagevimab/cilgavimab for preventing COVID-19 during the initial Omicron surge, which is continually updated while Omicron circulates and repeatable for different subvariants.[12]

The FDA continues to collaborate with the VA on demonstration projects to evaluate the impact of available vaccines and treatment against COVID-19 variants. The VA has also initiated several large-scale sequencing programs for COVID-19 specimens that will support these efforts, including VA Science and Health Initiative to Combat Infectious and Emerging Life-Threatening Diseases (VA SHIELD), VA Sequencing for Research Clinical and Epidemiology (SeqFORCE), and VA Sequencing Collaborations United for Research and Epidemiology (SeqCURE).[13,14] Successful proof-of-concept studies using these data could provide a template for VA and other medical systems/databases to report effectiveness in near real time.

INTERAGENCY COLLABORATION

The potential advantages of federal agencies working with the VA to build an infrastructure capable of generating real-world evidence effectiveness analyses in near real time is not limited to needs that will arise in the next pandemic. For example, generating randomized, placebo-controlled, clinical trial endpoint data on the effectiveness of new variant vaccines will be difficult from a feasibility and ethical standpoint. Combining the VA's robust virus sequencing program with preexisting

mechanisms, such as expanded access studies (allowed under FDA Investigational New Drug regulations), researchers could enable a large-scale effective evaluation program of vaccination with variant or universal COVID-19 vaccines, using rapidly accruing effectiveness data.

The pandemic created opportunities to advance innovative approaches to medical product development. Some have advocated these innovative approaches should proceed together toward a seamless convergence between the domains of medical research and clinical care. A shift toward expecting, as a matter of routine, effectiveness data to be generated in near real time and made available for benefit-risk assessment would be a useful step in that direction.

Expanding and sharing analytical platforms, including methodology and programming codes, will allow increased access to rapidly refreshed real-world data. A common adaptive platform of complete and continuously updated data will also enable a wider community of researchers to create multiple investigatory groups simultaneously accessing fully de-identified data for concurrent observational studies. In turn, researchers need to have programming, study design, and methodology ready in an open-source platform. An efficient platform would also require the adoption of artificial intelligence, natural language processing, imaging processing, and quantum computing for validation and improved data quality.

COVID-19 has demonstrated the need for open science data synchronization with universal access for faster action and improved outcomes able to gain public confidence. OpenSafely (UK), a software platform for analysis of EHR data that is shared automatically and openly for scientific review and efficient reuse, created a cohort of about 23.4 million records for observational review of monoclonal COVID-19 treatments. To keep pace with the UK, Israel, and other nationalized systems, the US would benefit from duplicating this example of coordination between federal agencies and their data repositories. For example, combining data between the DoD, which captures active military health care data through TRICARE, and VA, which follows postmilitary discharge, would create datasets encompassing complete life spans. Additionally, expanding the National COVID Cohort Collaborative (N3C) program—one of the largest collections of clinical data related to COVID-19 symptoms and patient outcomes in the US—to include EHR data from DoD, VA, Medicare, and Test to Treat initiative partners would further expand research capabilities. This could be accomplished through a framework of anonymized, readily available, harmonized data. EHRs with synchronized datasets from every health care practitioner—independent pharmacies, primary care physicians, and hospitals—could all work to create a de-identified, comprehensive, continuously updated, near real-time dataset accessible to all federal researchers.

CONCLUSIONS

The VA has been lauded for its rapid, effective response to the current pandemic. The successful management and prescription of vaccines and treatment to the largely high-risk veteran population was possible because of the existing data framework within the VA. VA Research continues to build and refine infrastructure to improve speed, quality, and value of data analytics. We can do more. Expanding partnerships to use existing VA data strategies in designing a cooperative national data alliance would deliver necessary progress to research and public health.

Acknowledgments

The authors thank Jeff Roberts, MD, for his insight on the US Food and Drug Administration, its responsibilities, and the potential benefit of real world data to its missions.

Author affiliations

[a]White River Junction Veterans Affairs Medical Center, Vermont
[b]Geisel School of Medicine at Dartmouth, Hanover, New Hampshire
[c]Office of Research and Development, Department of Veterans Affairs, Washington, DC
[d]Atlanta Veterans Affairs Medical Center, Decatur, Georgia
[e]Emory University School of Medicine, Atlanta, Georgia
[f]Center for Medication Safety, Pharmacy Benefits Management Services, Department of Veterans Affairs, Hines, Illinois

Author disclosures

Vincent C. Marconi received investigator-initiated research grants (to Emory University) and consultation fees from Eli Lilly, Bayer, Gilead Sciences and ViiV. The grants and fees were unrelated to the work discussed here.

Disclaimer

The opinions expressed herein are those of the authors and do not necessarily reflect those of *Federal Practitioner*, Frontline Medical Communications Inc., the U.S. Government, or any of its agencies.

References

1. US Department of Veterans Affairs, Veterans Health Administration. Third report details VA's continued efforts addressing COVID-19 pandemic. Accessed August 15, 2023. https://www.va.gov/opa/pressrel/pressrelease.cfm?id=5748
2. Xie Y, Ziyad A. Risks and burdens of incident diabetes in long COVID: a cohort study. *Lancet Diabetes Endocrinol*. 2022;10(5):311-321. doi:10.1016/S2213-8587(22)00044-4

3. Young-Xu Y, Korves C, Roberts J, et al. Coverage and estimated effectiveness of mRNA COVID-19 vaccines among US veterans. *JAMA Netw Open*. 2021;4(10):e2128391. doi:10.1001/jamanetworkopen.2021.28391
4. Dickerman BA, Gerlovin H, Madenci AL, et al. Comparative effectiveness of BNT162b2 and mRNA-1273 vaccines in U.S. veterans. *N Engl J Med*. 2022;386(2):105-115. doi:10.1056/NEJMoa2115463
5. Titanji BK, Eick-Cost A, Partan ES, et al. Effectiveness of smallpox vaccination to prevent mpox in military personnel. *N Engl J Med.* 2023;389(12):1147-1148. doi:10.1056/NEJMc2300805
6. Sarata AK, Dabrowska A, Johnson JA, Thaul S. FDA Reauthorization Act of 2017. Accessed August 15, 2023. https://sgp.fas.org/crs/misc/R44961.pdf
7. US Food and Drug Administration. FDA's sentinel initiative–background. February 2, 2022. Updated February 4, 2022. Accessed August 15, 2023. https://www.fda.gov/safety/fdas-sentinel-initiative/fdas-sentinel-initiative-background
8. Bajema KL, Dahl RM, Prill MM, et al; SUPERNOVA COVID-19; Surveillance Group. Effectiveness of COVID-19 mRNA vaccines against COVID-19–associated hospitalization—five Veterans Affairs medical centers, United States, February 1–August 6, 2021. *MMWR Morb Mortal Wkly*. 2021;70(37):1294-1299. doi:10.15585/mmwr.mm7037e3
9. Sharma A, Oda G, Holodniy M. COVID-19 vaccine breakthrough infections in Veterans Health Administration. *medRxiv*. Posted September 26, 2021. doi:10.1101/2021.09.23.21263864
10. Dickerman BA, Gerlovin H, Madenci AL, et al. Comparative effectiveness of third doses of mRNA-based COVID-19 vaccines in US veterans. *Nat Microbiol*. 2023;8(1):55-63. doi:10.1038/s41564-022-01272-z
11. Tang F, Hammel IS, Andrew MK, Ruiz JG. Frailty reduces vaccine effectiveness against SARS-CoV-2 infection: a test-negative case control study using national VA data. *J Nutr Health Aging*. 2023;27(2):81-88. doi:10.1007/s12603-023-1885-1
12. Young-Xu Y, Epstein L, Marconi VC, et al. Tixagevimab/cilgavimab for preventing COVID-19 during the Omicron surge: retrospective analysis of National Veterans Health Administration electronic data. *mBio*. 2023;14(4):e0102423. doi:10.1128/mbio.01024-23
13. US Department of Veterans Affairs. VA science and health initiative to combat infectious and emerging life-threatening diseases. *Open Forum Infect Dis*. 2022;9(12):ofac641. doi:10.1093/ofid/ofac64
14. Bilal MY. Similarity index–probabilistic confidence estimation of SARS-CoV-2 strain relatedness in localized outbreaks. *Epidemiologia*. 2022;3(2):238-249. doi:10.3390/epidemiologia3020019