

CARDIOVASCULAR PERSPECTIVE

American Heart Association Precision Medicine Platform Addresses Challenges in Data Sharing

Laura M. Stevens¹ ID, BS; James A. de Lemos² ID, MD; Sandeep R. Das, MD, MPH; Christine Rutan³ ID, BS; Heather M. Alger, PhD, MPH; Mitchell S.V. Elkind⁴ ID, MD; Juan Zhao, PhD; Kritika Iyer⁵ ID, MSE; C. Alberto Figueroa, PhD; Jennifer L. Hall⁶ ID, PhD

GOALS OF THE AMERICAN HEART ASSOCIATION PRECISION MEDICINE PLATFORM

The cloud-based American Heart Association (AHA) Precision Medicine Platform (PMP; <https://precision.heart.org/>)¹ was designed to address and overcome major challenges faced by researchers. The first challenge to overcome was sharing data. We have tested several data sharing options with researchers in the past 2 years. When the coronavirus disease 2019 (COVID-19) pandemic hit, we were prepared to launch a process that researchers supported. In short, we opened up our own COVID-19 registry data powered by Get With the Guidelines (GWTG). Increasing access to data for all researchers versus keeping it walled off to a select few had the ability to improve the quality, reproducibility, and validity of scientific findings during a time when the scientific process suffered a major setback. Through our initial tests, we also learned that researchers are not willing to invest time/effort in an access and reuse process that was complicated. Thus, we eliminated several steps and provided researchers a ready-to-run, cloud-based, virtual workspace that included (1) the necessary data documentation and data files, (2) statistical and visualization software, as well as machine learning and deep learning analysis tools, and (3) the computational power necessary to perform the analyses.¹⁻⁷

STRENGTHS AND CHALLENGES

The PMP is unique among cloud-based academic platforms in that researchers may access data along with

comprehensive data documentation from multiple sources including real-world patient data, longitudinal epidemiological studies, electronic health record data, and more. The long-standing reputation of the AHA and the trust it has earned in the community allow our organization to serve as a neutral broker for housing many rich data sources. Through our testing with researchers through the years, we have learned that a critical factor in sharing data has been allowing data owners to provision access. We do this through a process called Data Use Operating System. The open-source code (to which we made slight modifications) can be found on GitHub. This process involves researchers requesting access to answer a few short questions that are then emailed directly to the Data Access Committee assigned by the owner of the data. This committee votes on approval/revision/no approval to the data. The data requester is informed by email of the final decision, and if approved, the data and documentation are deposited into their workspace. For some data sets, like the COVID-19 registry data, the Data Access Committee requests and reviews a manuscript proposal with a statistical analysis plan. A key learning from this process for the AHA and researchers who thought their data might be published with errors or felt that all their hard work would now simply be taken by others, was that in fact, researchers requesting access wanted to collaborate and learn from the data owners. In many cases, the quality, validation, and replication of the research has improved.

Finally, the cloud-based AHA PMP allows all researchers at any University (with or without large data sets or resources) equal opportunities. Along these same lines,

Key Words: clinical study ■ computational biology ■ database ■ machine learning ■ registries

The opinions expressed in this article are not necessarily those of the editors or of the American Heart Association.

Correspondence to: Jennifer L. Hall, PhD, American Heart Association, National Center, 7272 Greenville Ave, Dallas, TX 75231. Email jennifer.hall@heart.org

This manuscript was sent to Dennis T. Ko, MD, Senior Guest Editor, for review by expert referees, editorial decision, and final disposition.

For Sources of Funding and Disclosures, see page 943.

© 2021 American Heart Association, Inc.

Circulation: Cardiovascular Quality and Outcomes is available at <http://www.ahajournals.org/journal/circoutcomes>

the AHA collaborates with cloud providers to allow academic researchers to use the secure workspaces at a reduced cost.

The 3 major challenges from our experience that researchers face with data sharing include (1) the data governance process including data use agreements, (2) access to critical standardized information accompanying the data sets including data dictionaries and case report forms, and (3) lack of flexibility in many cloud-based environments to scale resources to meet performance needs for analyses of shared data including images and electronic health record data.

SOLUTIONS TO CHALLENGES

Data Governance

In 2020, the global COVID-19 pandemic provided a valuable opportunity to overcome the challenges of data governance and access to standardized information including accompanying data dictionaries and case report forms. Clinicians had a need for generalizable real-world data to inform their understanding of COVID-19. Since no preexisting data sources or workflows were available, the AHA launched the COVID-19 CVD registry powered by GWTG and opted to make these data available on the AHA PMP.¹ This voluntary registry (described previously)¹ was designed to fill the unique gap in understanding cardiovascular risk and outcomes in patients with COVID-19 and is open to all hospitals and health systems in the United States treating adult patients with acute COVID-19 infections.

In the past, access to AHA GWTG data sets involved coordination across a variety of stakeholders. Although this process resulted in >600 publications over the last 17 years, the COVID-19 pandemic necessitated more

rapid collaboration, evaluation, and publication of findings to expedite the pace of science.

The Figure highlights the key implementation phases of this initiative. The initial phase of the initiative was open to all 104 hospitals that were actively enrolling records for the COVID-19 CVD registry, as well as the steering committee members. During this initial phase, we opened PMP workspaces for researchers with approved manuscript proposals to complete their analyses. A data use agreement was necessary before end users receiving a secure workspace equipped with aggregate data containing records from all registry sites, data documentation, and analysis tools. The data use agreement was simplified to enhance efficiency. In short, (1) we implemented a no-redline policy for end users, (2) added project data fields allowing for more specified data files to be applied to individual workspaces (based on modules and year), (3) required a signature from only the investigator instead of the institution, and (4) removed unnecessary language involving data use and disclosures that are now under the control of the AHA and permissible since the data are accessed on a secure cloud-based workspace. All changes are Health Insurance Portability and Accountability Act of 1996 compliant and improved turnaround time and overall completion of data use agreements. We are moving toward a data use agreement that will be moved to the PMP, accessible online, and able to be executed upon receipt.

Several approaches were used to support new investigators on the PMP. A member of the COVID-19 research and publications committee was assigned as a liaison to each project, to help with analysis planning and data questions. Weekly office hours were held virtually to listen to the needs and suggestions from researchers. Together, these approaches improved communication, addressed many questions, and significantly accelerated the process. As of April 20, 2021, 40 proposals

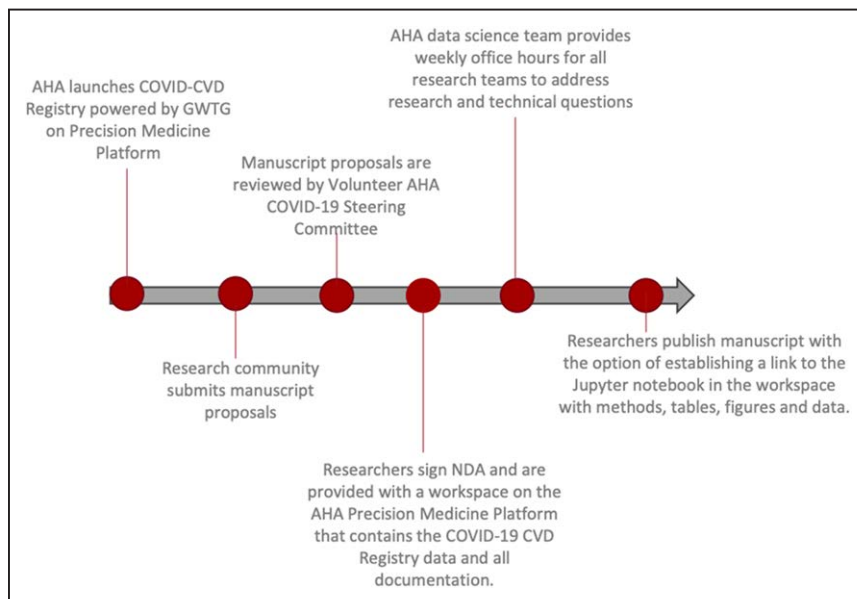


Figure. Design of open data initiative on the American Heart Association (AHA) Precision Medicine Platform (PMP).

Researchers submitted manuscript proposals, which were reviewed by a Research and Publications Committee. Once approved, researchers analyzed the data in a secure workspace on the PMP and published the manuscript. COVID-19 indicates coronavirus disease 2019; GWTG, Get With The Guidelines; and NDA, non-disclosure agreement.

have been accepted for investigator-led analyses, 15 analyses have been drafted for submission, and 9 manuscripts have been published.^{1,3,6,8-13}

Data Documentation

We were committed to the FAIR (Findable, Accessible, Interoperable, and Reusable) guiding principles that include uniform definitions, data dictionaries, and data documentation. This interactive data documentation for our COVID-19 GWTG data (<https://precision.heart.org/documentation/AHA-COVID19-CVD-GWTG/index.html>) improved the understanding of data definitions and derived variables, thereby reducing inconsistencies across manuscripts and frustrations in dealing with open data. In particular, the explore and discover section of the interactive data documentation in the manuscript illustrates how researchers are able to access all data documentation files, missingness of data, data distribution, and more. The data documentation is not multiple flat PDF files that require users to toggle between files without the clear understanding of what the variables mean. We worked in coordination with the AHA COVID-19 steering committee, made up of clinical, statistical, and epidemiological experts, to arrive at data standards that were based on previous GWTG data standards, as well as data from European registries. To further address this challenge, we piloted the use of usage examples and tutorials written by our AHA data science team in multiple languages that allowed users to reuse shared code in their workspace resulting in final products like a demographic profile of the data. Thus, all researchers using the same data set in their workspace and the code that was verified and approved would end up with the same demographic profile for their manuscripts. This improved consistency across manuscripts. Finally, members of the COVID Research and Publications Committee and AHA leadership reviewed manuscripts before submission to provide quality oversight and conformity with the original proposal.

Solutions for data governance and data documentation on the PMP were also tested by other nonprofit groups, including the American Society of Clinical Oncology, which licenses the underlying technology of the PMP to deliver CancerLinQ Registry data to academic users for analysis. The AHA also works with the Society of Critical Care Medicine to map variables between our registries with the end goal of increasing our understanding of COVID-19 and its impact on patient lives. Both opportunities provided additional learnings and solutions from end users with respect to data use agreements and data documentation.

Scaling Resources and Improving Flexibility

To overcome challenges in scaling resources for performance needs, we worked closely with researchers training

neural networks for medical image segmentation and large-scale simulations. In the cardiovascular field, diagnostic decisions can be improved using algorithms to segment coronary vessels in angiograms. The scalability, larger memory, and computing power of the PMP paired researchers with 4 NVIDIA Tesla K80 graphics processing unit to train a custom pipeline, *AngioNet*—a neural network for coronary segmentation.⁴ By doing so, the research team was able to increase the number of images used to train each iteration of the network, improving accuracy and generalizability compared with training on a single graphics processing unit. Another computationally intensive application deployed to the PMP is *CRIMSON*,² an open-source hemodynamic modeling software that has been used in a wide range of applications, from cardiovascular disease research to surgical planning.² The finite element based flowsolver of this software has been compiled on the PMP in a Docker container, allowing researchers to perform large-scale hemodynamic simulations on patient-specific anatomic models using the high-performance computing resources of the PMP.

Working with researchers performing unsupervised machine learning across electronic health record data also provided innovative solutions to improve agility in workspaces for end users.⁷ Zhao et al⁷ used constrained nonnegative tensor factorization to extract phenotypic topics across time scales in a study cohort derived from a deidentified copy of the electronic health record for patients in the Vanderbilt University Medical Center on the PMP. This study identified previous risk factors associated with cardiovascular disease, as well as new potential factors including vitamin D deficiency and depression, as well as urinary infection.⁷

SUMMARY AND CONCLUSIONS

AHA's PMP has enabled secure delivery of data through agile workspaces that scale with the high-performance compute needs of researchers and allow flexibility in the ready-to-run analysis tools. By listening and partnering with end users, we have overcome many of the hurdles facing researchers today including outdated data governance policies, insufficient data documentation, and inability of cloud-based environments to scale up resources for performance and allow researchers to personalize their workspaces with their own tools and pipelines.

ARTICLE INFORMATION

Affiliations

Department of Cardiology and Computational Bioscience Program, University of Colorado, Aurora (L.M.S.). Quality, Outcomes Research and Analytics, American Heart Association, Dallas, TX (L.M.S., C.R., H.A., J.L.H.). UT Southwestern Medical Center, Dallas, TX (J.A.d.L., S.R.D.). Center for Innovation and Value at Parkland, Dallas, TX (S.R.D.). Vagelos College of Physicians and Surgeons and Mailman School of Public Health, Columbia University, NewYork-Presbyterian Hospital (M.S.V.E.). Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN (J.Z.). Departments of Biomedical Engineering (K.I. and

C.A.F.) and Surgery (C.A.F.), University of Michigan, Ann Arbor. Department of Medicine, Lillehei Heart Institute, University of Minnesota, Minneapolis (J.L.H.).

Acknowledgments

We would like to thank all of the members of the American Heart Association (AHA) COVID-19 Steering Committee for volunteering their time and expertise to this initiative. The Get With The Guidelines programs are provided by the AHA. The Precision Medicine Platform was established by the AHA, is powered by Amazon Web Services, and is supported by Hitachi Vantara.

Sources of Funding

This project was supported by the American Heart Association (AHA). AHA's COVID-19 CVD Registry is partially supported by generous funds from the Gordon and Betty Moore Foundation.

Disclosures

L.M. Stevens, Dr Alger, Dr Hall, and C. Rutan are employees of the American Heart Association. Dr Hall is an adjunct professor at the University of Minnesota. Drs de Lemos and Elkind are unpaid officers of the American Heart Association. Dr de Lemos discloses receiving grant support from Abbott Diagnostics and Roche Diagnostics and consulting income from Abbott Diagnostics, Ortho Clinical Diagnostics, Quidel Cardiovascular, Amgen, Regeneron, Eli Lilly and Novo Nordisk, and Janssen. Dr Elkind discloses receiving study drug in-kind from the BMS-Pfizer Alliance for Eliquis and ancillary research funding from Roche for an NIH-funded trial of stroke prevention; receiving royalties from UpToDate for chapters related to stroke; and receiving funding from NINDS, NHLBI, and the Leducq Foundation. Dr Figueroa discloses he is a cofounder of AngioInsight, Inc. The other authors report no conflicts.

REFERENCES

- Alger HM, Rutan C, Williams JH 4th, Walchok JG, Bolles M, Hall JL, Bradley SM, Elkind MSV, Rodriguez F, Wang TY, et al. American Heart Association COVID-19 CVD registry powered by Get With The Guidelines. *Circ Cardiovasc Qual Outcomes*. 2020;13:e006967. doi: 10.1161/CIRCOUTCOMES.120.006967
- Arthurs CJ, Khlebnikov R, Melville A, Marčan M, Gomez A, Dillon-Murphy D, Cuomo F, Silva Vieira M, Schollenberger J, Lynch SR, et al. CRIMSON: an open-source software framework for cardiovascular integrated modelling and simulation. *PLoS Comput Biol*. 2021;17:e1008881. doi: 10.1371/journal.pcbi.1008881
- Hendren NS, de Lemos JA, Ayers C, Das SR, Rao A, Carter S, Rosenblatt A, Walchok J, Omar W, Khera R, et al. Association of body mass index and age with morbidity and mortality in patients hospitalized with COVID-19: results from the American Heart Association COVID-19 Cardiovascular Disease Registry. *Circulation*. 2021;143:135–144. doi: 10.1161/CIRCULATIONAHA.120.051936
- Iyer K, Najarian CP, Fattah AA, Arthurs CJ, Soroushmehr SMR, Subban V, Sankardas MA, Nadakuditi RR, Nallamothu BK, Figueroa CA. AngioNet: a convolutional neural network for vessel segmentation in X-ray angiography. MedRxiv. Preprint posted online January 26, 2021. doi: 10.1101/2021.01.25.21250488
- Kass-Hout TA, Stevens LM, Hall JL. American Heart Association precision medicine platform. *Circulation*. 2018;137:647–649. doi: 10.1161/CIRCULATIONAHA.117.032041
- Rodriguez F, Solomon N, de Lemos JA, Das SR, Morrow DA, Bradley SM, Elkind MSV, Williams JH, Holmes D, Matsouka RA, et al. Racial and ethnic differences in presentation and outcomes for patients hospitalized with COVID-19: findings from the American Heart Association's COVID-19 Cardiovascular Disease Registry. *Circulation*. 2021;143:2332–2342. doi: 10.1161/CIRCULATIONAHA.120.052278
- Zhao J, Zhang Y, Schlueter DJ, Wu P, Eric Kerchberger V, Trent Rosenbloom S, Wells QS, Feng Q, Denny JC, Wei WQ. Detecting time-evolving phenotypic topics via tensor factorization on electronic health records: cardiovascular disease case study. *J Biomed Inform*. 2019;98:103270. doi: 10.1016/j.jbi.2019.103270
- Leasure AC, Khan YM, Iyer R, Elkind MSV, Sansing LH, Falcone GJ, Sheth KN. Intracerebral hemorrhage in patients with COVID-19: an analysis from the COVID-19 Cardiovascular Disease Registry. *Stroke*. 2021;52:e321–e323. doi: 10.1161/STROKEAHA.121.034215
- Rao A, Ranka S, Ayers C, Hendren N, Rosenblatt A, Alger HM, Rutan C, Omar W, Khera R, Gupta K, et al. Association of kidney disease with outcomes in COVID-19: results from the American Heart Association COVID-19 Cardiovascular Disease Registry. *J Am Heart Assoc*. 2021;10:e020910. doi: 10.1161/JAHA.121.020910
- Bradley SM, Emmons-Bell S, Mutharasan RK, Rodriguez F, Gupta D, Roth G, Gluckman TJ, Shah RU, Wang TY, Khera R, et al. Repeated cross-sectional analysis of hydroxychloroquine deimplementation in the AHA COVID-19 CVD Registry. *Sci Rep*. 2021;11:15097. doi: 10.1038/s41598-021-94203-7
- Tehrani DM, Wang X, Rafique AM, Hayek SS, Herrmann J, Neilan TG, Desai P, Morgans A, Lopez-Mattei J, Parikh RV, et al. Impact of cancer and cardiovascular disease on in-hospital outcomes of COVID-19 patients: results from the American Heart Association COVID-19 Cardiovascular Disease Registry. *Cardiooncology*. 2021;7:28. doi: 10.1186/s40959-021-00113-y
- Roth GA, Emmons-Bell S, Alger HM, Bradley SM, Das SR, de Lemos JA, Gakidou E, Elkind MSV, Hay S, Hall JL, et al. Trends in patient characteristics and COVID-19 in-hospital mortality in the United States during the COVID-19 pandemic. *JAMA Netw Open*. 2021;4:e218828. doi: 10.1001/jamanetworkopen.2021.8828
- Daniels LB, Ren J, Kumar K, Bui QM, Zhang J, Zhang X, Sawan MA, Eisen H, Longhurst CA, Messer K. Relation of prior statin and anti-hypertensive use to severity of disease among patients hospitalized with COVID-19: findings from the American Heart Association's COVID-19 Cardiovascular Disease Registry. *PLoS One*. 2021;16:e0254635. doi: 10.1371/journal.pone.0254635