The NEW ENGLAND JOURNAL of MEDICINE

Perspective

Leveraging Open Science to Accelerate Research

Kushal T. Kadakia, M.Sc., Adam L. Beckman, B.S., Joseph S. Ross, M.D., M.H.S., and Harlan M. Krumholz, M.D.

The United States has mobilized the full force of its clinical research enterprise to address the Covid-19 pandemic, allocating billions of dollars to support timely research. As of January 2021, for example, the National Institutes of Health (NIH) had issued nearly a thousand awards cumulatively worth roughly $2 billion to support Covid-19 projects ranging from the development of medical products (including diagnostics and vaccines) to evaluations of population-specific risk factors and outcomes.[1] Such initiatives, which have yielded new technologies and important evidence, illustrate the value of robust scientific infrastructure.

Society's challenge now involves maximizing the return on these research investments. Studies have shown that investigators don't report results or share data from many federally supported trials, making replication impossible.[2] Often, protocols aren't made public and data-collection tools aren't shared. We believe that policymakers should incorporate open-science principles into research policies and programs to optimize the return on federal investment in clinical research, which could have benefits beyond the pandemic.

The idea of embracing open science represents a vision for research conduct that promotes standard processes for sharing protocols and registering studies, reporting and disseminating results, and sharing data, biospecimens, and code. The advancement of science — an intrinsically iterative process — is contingent on reporting practices that enable data to be findable, accessible, interoperable, and reusable to permit independent scrutiny, replication, and follow-on investigations. Realizing the value of research and fostering trust in science requires study information to be readily available to the public and the scientific community, including in open-access journals and on preprint platforms. Over the past 20 years, policymakers and investigators have promoted open science to counteract clinical researchers' tendency to sequester data. Such efforts have included the recent release of NIH data-sharing guidelines and public–private partnerships for data sharing, such as the Yale University Open Data Access Project (which two of us help to lead).

The urgency associated with the pandemic has created an imperative to accelerate the adoption of open science. Researchers uploaded the initial genome sequence of SARS-CoV-2 in an open-access database in January 2020, creating a data-sharing precedent and metadata that would later enable insights about new Covid-19 variants. The NIH developed a dedicated platform for sharing research tools for Covid-19 and encouraged investigators to ex-

pedite reporting to ClinicalTrials.gov ahead of requirements.[3] Open-science publishing agreements supporting evidence dissemination have complemented these practices and policies. The day after the World Health Organization declared Covid-19 a public health emergency, more than 50 academic publishers issued a joint statement committing to open-access policies for Covid-19 research.[4] Support for preprint servers has promoted awareness of research successes and failures, and journals have helped accelerate the distribution of actionable information, including by means of dedicated Covid-19 Web pages, endorsement of preprints, and an emphasis on sharing data with public health authorities.

Yet scientists haven't entirely embraced these steps, and many process changes have been voluntary or lacked enforcement. The results of only 8% of completed or terminated Covid-19 studies have been published on ClinicalTrials.gov, and gaps in coordination — reflected in the existence of redundant studies and studies with misaligned end points — have hindered researchers' ability to synthesize and learn from emerging evidence. The pandemic, therefore, provides an opening for policymakers to adopt mandates and incentives to help make open science the norm.

Policymakers could use several strategies to harness open science and accelerate Covid-19 research. First, the government could encourage transparency among entities that receive Covid-19 research awards. The Department of Health and Human Services, for example, could include data on adherence to open-science requirements in its Tracking Accountability in Government Grants System (TAGGS) for Covid-19. Specifically, regulators could provide funding for fulfilling data-sharing requirements to remove the onus on investigators and could publish data on institutional compliance with transparency expectations, such as on rates of timely reporting and data accessibility, on the TAGGS dashboard. Policymakers could also condition federal-award funding on adherence to open-science principles, including the timely dissemination of study protocols, reports, and participant-level data.

Second, policymakers could promote open science in their expectations for federally funded projects. The NIH has advanced transparency during the pandemic, including by expediting processing of study registration and results reporting on ClinicalTrials.gov and launching a preprint pilot through PubMed Central for NIH-funded research.[3] Truly elevating open-science research practices, however, would require addressing the long-standing inconsistencies in ClinicalTrials.gov compliance and sharing of protocols and participant-level data and extending pandemic-era transparency policies beyond Covid-19. To start, policymakers could develop standards that mandate timelines for study registration, protocol posting, results reporting, and data and code sharing for all Covid-19 grantees — standards that could be integrated into research protocols and implemented with minimal disruption. There are more than 2500 registered clinical trials related to Covid-19, many of which are supported by federal funds. The NIH and the Food and Drug Administration (FDA) have the power to penalize parties that violate legal requirements for registration and results reporting. Although neither agency has exercised this authority, two recent developments — the release of FDA guidance on civil monetary penalties for responsible parties that don't submit required ClinicalTrials.gov data and a 2020 federal court ruling for the NIH requiring sponsors to submit missing data from 2007 onward — provide momentum for policymakers to address inconsistencies in study transparency. The federal government could further strengthen enforcement by publicly identifying noncompliant parties and imposing sanctions, such as holds on grant funding.

Third, policymakers could invest in open-science platforms for government-funded research. It's not enough to mandate study reporting; policymakers must also ensure that study materials are easily accessible in repositories that are sufficiently well-resourced to permit rapid processing and maintenance. An important priority is sharing participant-level data, which is necessary to bolster the evidence base for treating Covid-19 and to fulfill researchers' ethical obligation to patients who have accepted risks by participating in clinical research and consented to having their data shared. Policies should support participants' agency; they should communicate standards for reporting and how data may be used for downstream research, ideally as part of the consent process.

Protecting the interests of research participants is imperative. Policies should consider the ways in which the data type and source (for example, data from clinical trials versus data collected in real-world settings) guide the process of anonymizing and sharing participant-level information. In-

vestigators conducting prospective studies — from which data are frequently reused, such as in meta-analyses — should anticipate data sharing and secondary analyses during recruitment and look to the National Academy of Medicine's 2015 consensus report for best practices.[5] Studies using real-world data sources, such as insurance claims and medical records, should be accompanied by public postings identifying the owners of the data set, justifying its use, and documenting the governance systems (e.g., institutional review boards) and infrastructure (e.g., storage systems and methods of deidentification) that have been used for oversight.

Accelerating participant-level data sharing could also allow investigators to better engage patients as partners. People with agency over their own data have the opportunity to decide how those data can be used and can enter into more equitable relationships with researchers. Such an approach could enable participants to become authentic members of the research team and address their concerns about information security. Applying this strategy to research involving traditionally disempowered groups may be a path for empowering minority communities and engendering trust.

Covid-19 has accelerated many changes in health care and medical research. A commitment to open science during the pandemic could support the gradual transformation of the clinical research enterprise in the United States. Efforts to promote data-sharing practices for Covid-19 could reinforce the NIH's recently finalized Policy for Data Management and Sharing and provide an opportunity to clarify ambiguities. For example, the agency could create consistent data-sharing policies throughout its various institutes, centers, and offices. Similarly, the uptake of open-science practices could sustain newly established public–private partnerships beyond the pandemic to improve clinical trial transparency in general. Affirming open-science principles could also help restore public trust. Together, these steps could ensure that the pandemic's legacy is a reminder not just of what science can do, but also of how science should be done.

Disclosure forms provided by the authors are available at NEJM.org.

From Lincoln College, University of Oxford, Oxford, United Kingdom (K.T.K.); Harvard Medical School and Harvard Business School, Boston (A.L.B.); the Section of General Internal Medicine (J.S.R.), the National Clinician Scholars Program (J.S.R.), and the Section of Cardiovascular Medicine (H.M.K.), the Department of Internal Medicine, Yale School of Medicine, the Department of Health Policy and Management, Yale School of Public Health (J.S.R., H.M.K.), and the Center for Outcomes Research and Evaluation, Yale New Haven Hospital (J.S.R., H.M.K.) — all in New Haven.

This article was published on March 24, 2021, at NEJM.org.

1. Tracking Accountability in Government Grants System (TAGGS). HHS Covid-19 funding (https://taggs.hhs.gov/Coronavirus).
2. Chan AW, Song F, Vickers A, et al. Increasing value and reducing waste: addressing inaccessible research. Lancet 2014;383:257-66.
3. National Institutes of Health. NIH calls on clinical researchers to swiftly share Covid-19 results. November 10, 2020 (https://www.nih.gov/about-nih/who-we-are/nih-director/statements/nih-calls-clinical-researchers-swiftly-share-covid-19-results).
4. Wellcome. Sharing research data and findings relevant to the novel coronavirus (Covid-19) outbreak. January 31, 2020 (https://wellcome.org/coronavirus-covid-19/open-data).
5. Institute of Medicine. Sharing clinical trial data: maximizing benefits, minimizing risk. Washington, DC: The National Academies Press, 2015.

DOI: 10.1056/NEJMp2034518