Weighing the Benefits and Risks of Proliferating Observational Treatment Assessments Observational Cacophony, Randomized Harmony

Amid the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic, substantial effort is being directed toward mining databases and publishing case series and reports that may provide insights into the epidemiology and clinical management of coronavirus disease 2019 (COVID-19). However, there is growing concern about whether attempts to infer causation about the benefits and risks of potential therapeutics from non-randomized studies are providing insights that improve clinical knowledge and accelerate the search for needed answers, or whether these reports just add noise, confusion, and false confidence. Most of these studies include a caveat indicating that "randomized clinical trials are needed." But disclaimers aside, does this approach help make the case for well-designed randomized clinical trials (RCTs) and accelerate their delivery? Or do observational studies reduce the likelihood of a properly designed trial being performed, thereby delaying the discovery of reliable truth?

The growth of structured registries and organization of claims and electronic health record data have greatly expedited sophisticated comparisons of therapies provided in clinical practice settings (ie, observational "real-world" evidence). Large troves of administrative and clinical data can be accessed and tabulated, using programs to construct propensity scores and inverse-weighted probability estimates. The benefits of this approach, if well done, are obvious: by sifting potential treatments and measuring outcomes and safety signals, qualified investigators and funding agencies can choose the most promising therapies for testing in rigorous RCTs. Sample sizes and expected event rates can be calculated, and communities and health care systems with relevant patient populations identified. The risks, however, are also clear: aggregating information about diagnosis, comorbidities, treatment, and outcomes can lend a patina of technical excellence that obscures the influence of systematic bias (patients who receive a given treatment are not the same as those who do not), leading to erroneous estimates of treatment effects. These risks are often unclear to the public when observational findings are widely disseminated by the lay media.

Anxious, frightened patients, as well as clinicians and health systems with a strong desire to prevent morbidity and mortality, are all susceptible to cognitive biases. Furthermore, profit motives in the medical products industry, academic hubris, interests related to increasing the valuation of data platforms, and revenue generated by billing for these products in care delivery can all tempt investigators to make claims their methods cannot fully support, and these claims often are taken up by traditional media and further amplified on social media. Politicians have been directly involved in discourse about treatments they assert are effective. The natural desire of all elements of society to find effective therapies can obscure the difference between a proven fact and an exaggerated guess. Nefarious motives are not necessary for these problems to occur.

The role of regulators in this context is crucial. In the United States, the 21st Century Cures Act and user fee agreements require industry, academia, and regulators to advance the use of data and evidence from clinical settings. This legislation directed the US Food and Drug Administration (FDA) and the National Institutes of Health (NIH) to work with the clinical research ecosystem to develop robust methods for generating such evidence and clear guidance for applying it. Historically, the FDA has insisted on high-quality evidence as a condition for granting marketing approval for drugs and devices, and for specific marketing claims.

Considerable progress has been made in defining appropriate methods for improving the quality of observational treatment comparisons. Both NIH- and FDA-funded work fosters transparency by publishing study protocols, reporting results, and ensuring methodological rigor in this treacherous field. Methods for ensuring data quality are also evolving rapidly. In the context of COVID-19, the FDA has worked through the Evidence Accelerator to advance observational research methods and characterize quality and bias of newly available data sets.

This approach addresses valid concerns about veracity and data quality in observational research. However, this approach also should accelerate and prioritize the development and delivery of RCTs, not be viewed as a substitute for them. In fact, the most important data and evidence will accrue from applying randomized designs (individual, cluster, adaptive) within the context of data from clinical practice settings. The exigencies of the pandemic have created an understandable temptation to rush toward therapeutic options without the usual rigor, but the conclusions of reports must include appropriate caveats about the degree of uncertainty. Care must be taken to eschew "pandemic exceptionalism" to produce reliable evidence to guide intervention.

Academic leaders and clinicians also have critical responsibilities. The pressure to issue newsworthy pronouncements often fuels communications efforts by universities and companies that can promote unwarranted expectations in an era of social media "virality." Clinicians must find the balance between supporting optimism in their patients and being truthful about the quality and uncertainties of therapeutic evidence. When given the option of using an unproven treatment or enrolling patients in appropriate, well-designed trials, the choice of advancing reliable knowledge should be far preferable.
Several recent experiences in the public arena exemplify concerns about a cacophony of scientific claims regarding candidate therapeutics. In the case of hydroxychloroquine, initial reports of benefit were followed by the initiation of multiple clinical trials using randomization across the spectrum of relevant populations. While these trials were accruing, multiple observational studies were published, claiming to show either no benefit or harm, and one very large published study received sharp criticism from experts and immediate calls for retraction due to methodological flaws and concerns about data provenance.6

However, despite the refrain that RCTs are needed, the lay and scientific press amplified various estimates of treatment effect, while at the same time hydroxychloroquine was promoted in the global political arena. The fact that a high-profile study incorporating observational data was later retracted6 is in some ways less relevant: during the brief interval when the study data were thought to be valid, many (including some international regulators) were duped by the method, turning the conclusion of “evidence from RCTs is needed” into a movement of “RCTs should cease.” However, several pragmatic RCTs were conducted, and definitive findings of no benefit for hydroxychloroquine in hospitalized patients with COVID-19 have been announced.7,8

Meanwhile, a venerable candidate for treating acute lung injury, the corticosteroid dexamethasone, was also being examined. In a preliminary report, low-dose dexamethasone resulted in a mortality reduction in patients with COVID-19 requiring oxygen or ventilator support,9 showing that this inexpensive, generic, lifesaving treatment is beneficial for relevant patients.

Another recent study of 20,000 patients treated with plasma infusions from recovering COVID-19 patients10 claimed evidence of safety and expressed optimism for benefit based on low reported event rates, although there was no control group to anchor the observed event rates. If a fraction of these patients had been enrolled in RCTs, the answer for whether this intervention was effective would now be known. Ongoing US RCTs are slowly accruing patients in the face of massive public plasma donation and uncertainty regarding benefits or risk in the treatment of COVID-19 patients.

Ideally, robust ongoing evaluation would be applied to the use of treatments and clinical outcomes. Continuing quality improvement in electronic health record and claims data; development of multiple registries to evaluate technologies, medical procedures, and quality of care; and ongoing methodological refinements all contribute to making a system of continuous learning feasible. In some situations, observational findings about treatment effects associated with specific interventions merit adoption in practice, but in most cases this learning system should identify promising treatments and approaches for designing proper large-scale trials or should supplement RCT findings by modeling effects seen in RCTs in broader populations. Rather than promoting inconclusive observational findings in medical journals and the press, a repository could be created to register results in a manner less apt to inappropriately influence practice. In addition, it seems prudent to place a moratorium on reporting observational studies that could mislead the public.

Once promising treatments are identified, the system should be aligned to optimize enrollment in well-designed RCTs with sufficient power to provide definitive answers. This will require reimagining the entire system to remove unjustified barriers, such as onerous bureaucratic steps, excessive costly monitoring, and data collection that is cumbersome and far exceeds the needs of the trial.4 Mechanisms must be in place to make it as easy or preferable for potential participants to enroll in a trial of a potentially worthwhile treatment as it is to prescribe the same unproven treatment. The former approach ensures the rapid advance of reliable clinical knowledge and benefits future patients; the latter means clinicians and researchers will remain ignorant. But if leaders, commentators, academics, and clinicians cannot restrain the rush to judgment in the absence of reliable evidence, the proliferation of observational treatment comparisons will hinder the goal of finding effective treatments for COVID-19—and a great many other diseases.

REFERENCES


ARTICLE INFORMATION

Published Online: July 31, 2020. doi:10.1001/jama.2020.13319

Conflict of Interest Disclosures: Dr Califf reported being head of clinical policy and strategy at Verily Life Sciences and Google Health, an adjunct professor of medicine at Duke University and Stanford University, a board member for Cytokinetiks, and former commissioner for the FDA. Dr Hernandez reported receipt of grants and personal fees from AstraZeneca, Amgen, Boehringer Ingelheim, Novartis, and Merck, personal fees from Bayer, and grants from Janssen and Verily, as well as being the principal investigator for the Healthcare Worker Exposure & Outcomes Research (HEROES) Program funded by the Patient-Centered Outcomes Research Institute. Dr Landray reported receipt of grants from Boehringer Ingelheim, Novartis, The Medicines Company, Merck, Sharp & Dohme, and UK Biobank and being co-chief investigator for the RECOVERY trial of potential treatments for hospitalized patients with COVID-19, funded by UK Research & Innovation and the National Institute for Health Research (NIHR). Dr Landray is supported by Health Data Research UK, the NIHR Oxford Biomedical Research Centre, and the Medical Research Council Population Health Research Unit.

Role of the Funder/Sponsor: Supporters had no role in the preparation, review, or approval of the manuscript or decision to submit the manuscript for publication.

Additional Contributions: We thank Jonathan McCall, MS (Duke Forge, Duke University), for editorial assistance. No compensation other than usual salary was received.