1 Department of Environmental Health Sciences, Yale School of Public Health, New Haven, CT, USA

2 Yale University School of Medicine, New Haven, CT, USA

3 Section of General Medicine and the National Clinician Scholars Program, Department of Internal Medicine, Yale School of Medicine; Center for Outcomes Research and Evaluation, Yale-New Haven Health System; Department of Health Policy and Management, Yale School of Public Health, New Haven, CT, USA

Cite this as: *BMJ* 2022;377:o1229
http://dx.doi.org/10.1136/bmj.o1229
Published: 16 May 2022

Generating evidence during a pandemic: what's reliable?

Joshua D Wallach, [1] Osman Moneer, [2] Joseph S Ross[3]

The urgency of the covid-19 pandemic highlighted the need to leverage any and all available data that may potentially inform clinical, public health, and regulatory decision making.[1] Although randomised controlled trials (RCTs) are generally considered the gold standard for determining therapeutic safety and efficacy,[2] these studies are often logistically challenging, expensive, and can sometimes take years to plan and complete. In contrast, observational studies can be undertaken in real time, taking advantage of existing data sources that reflect real world care.[3] If clinicians are treating their patients with a drug, even if regulators have not approved its use for that indication,[4] researchers can generate evidence to characterise its benefits and harms. However, concerns have consistently been raised about whether observational studies offer valid conclusions given the potential impact of selection bias and confounding.

Numerous empirical assessments have compared the treatment effects reported in observational studies and RCTs evaluating the same questions. Several found relatively high levels of agreement between observational studies and RCTs across different topics,[5 6] while others suggested that using different study designs can lead to conflicting findings.[7] As a result, decision makers are often left uncertain about whether observational research can be a reliable proxy for RCTs. But what about in the context of a pandemic, when there is a need for the rapid generation of evidence?

In our linked study in *The BMJ*,[8] we systematically identified, matched, and compared study demographics and treatment effects from individual or meta-analysed observational studies and RCTs evaluating the same covid-19 therapeutics, comparators, and outcomes. We conducted 17 new, independent meta-analyses of observational studies of hydroxychloroquine, lopinavir-ritonavir, or dexamethasone —which were the most frequently studied therapeutic interventions for covid-19 at the time we started our study—in comparison with an active or placebo comparator. These meta-analyses were matched and compared with 17 meta-analyses of RCTs reported in a landmark living review in *The BMJ*.[9] We also compared 10 matched pairs for which only one observational study and/or only one RCT were identified. Matched observational studies and RCT pairs were considered to be in agreement if both observational and RCT treatment effects were statistically significantly increasing/decreasing ($P<0.05$) or if both treatment effects were non-statistically significant ($P>0.05$).

We found that all matched pairs with adequate reporting of demographic and clinical data had overlapping distributions of sex, age, and disease severity—in other words, the patient populations included in both the observational studies and RCTs appeared similar. This is critical because previous studies comparing the results derived from these two study designs have determined that patient populations in observational studies are frequently different from those in RCTs.

Next, comparing the treatment effects between observational studies and RCTs, we found that over three quarters were in agreement—in other words, the two different study designs were consistent in determining whether hydroxychloroquine, lopinavir-ritonavir, and dexamethasone were effective for the treatment of covid-19. In fact, comparing treatment effects using dichotomous outcomes, such as hospitalisation or death, had even higher levels of agreement.

A number of important lessons are illustrated by our study. Firstly, in future pandemics, neither RCTs nor observational studies should be automatically assumed to serve as a gold standard. The vast majority of individual RCTs and observational studies had at least one methodological limitation, which can affect the direction and magnitude of the observed findings. Instead, these two study designs can complement one another as evidence accumulates and a deeper understanding emerges.

Secondly, the vast majority of treatment effects across all interventions and comparisons were null—many therapies being repurposed for covid-19 were simply not effective. Our findings may have differed if studies had consistently identified a positive treatment effect.

Thirdly, if observational studies are intended to replicate or predict RCTs evaluating specific questions, they should be designed in line with the target trial framework[10] and follow structured templates for planning and reporting.[11] These frameworks focus on ensuring that observational studies consider the types of patients, follow-up periods, outcomes of interests, and analyses methods that will answer the causal question of interest and heighten reproducibility while minimising bias. Several ongoing efforts are focused on emulating RCTs using the target framework.[12 13] Well designed studies such as these can increase the likelihood that findings from observational studies will be reliable and complement evidence generated from RCTs, whether in the context of a pandemic or not.

Competing interests: Dr Wallach currently received research support through Yale University from Johnson and Johnson to develop methods of clinical trial data sharing, from the Food and Drug Administration, and from the National Institute on Alcohol Abuse and Alcoholism under award K01AA028258. Mr Moneer received support from the Yale-Mayo Clinic Center of Excellence in Regulatory Science and Innovation (CERSI) Scholars Program. Dr Ross currently receives research support through Yale University from Johnson and Johnson to develop methods of clinical trial data sharing, from the Medical Device Innovation Consortium as part of the National Evaluation System for Health Technology (NEST), from the Food and Drug Administration for the Yale-Mayo Clinic CERSI program (U01FD005938), from the Agency for Healthcare Research and Quality

(R01HS022882), from the National Heart, Lung and Blood Institute of the National Institutes of Health (NIH) (R01HS025164, R01HL144644), and from the Laura and John Arnold Foundation to establish the Good Pharma Scorecard at Bioethics International. In addition, Dr Ross is an expert witness at the request of Relator's attorneys, the Greene Law Firm, in a qui tam suit alleging violations of the False Claims Act and Anti-Kickback Statute against Biogen Inc.

Provenance and peer review: Not commissioned; not peer reviewed.

1 US Food and Drug Administration. Framework for FDA's Real-World Evidence Program. December 2018. https://www.fda.gov/media/120060/download
2 Jones DS, Podolsky SH. The history and fate of the gold standard. *Lancet* 2015;385:1502-3. doi: 10.1016/S0140-6736(15)60742-5. pmid: 25933270
3 US Food and Drug Administration. Real-World Evidence. https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence
4 US Food and Drug Administration. Understanding Unapproved Use of Approved Drugs "Off Label." https://www.fda.gov/patients/learn-about-expanded-access-and-other-treatment-options/understanding-unapproved-use-approved-drugs-label#:~:text=Unapproved%20use%20of%20an%20approved,a%20different%20type%20of%20cancer
5 Ioannidis JPA, Haidich AB, Pappa M, etal. Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA* 2001;286:821-30. doi: 10.1001/jama.286.7.821. pmid: 11497536
6 Anglemyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. *Cochrane Database Syst Rev* 2014;(4):MR000034. doi: 10.1002/14651858.MR000034.pub2. pmid: 24782322
7 Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JPA. Agreement of treatment effects for mortality from routinely collected data and subsequent randomized trials: meta-epidemiological survey. *BMJ* 2016;352:i493. doi: 10.1136/bmj.i493. pmid: 26858277
8 Moneer O, Daly G, Skydel JJ, etal. Agreement of treatment effects from observational studies and randomized controlled trials evaluating hydroxychloroquine, lopinavir-ritonavir, or dexamethasone for covid-19: meta-epidemiological study. *BMJ* 2022;377:e069400. doi: 10.1136/bmj-2021-069400. pmid: 35537738
9 Siemieniuk RA, Bartoszko JJ, Ge L, etal. Drug treatments for covid-19: living systematic review and network meta-analysis. *BMJ* 2020;370:m2980. doi: 10.1136/bmj.m2980. pmid: 32732190
10 Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol* 2016;183:758-64. doi: 10.1093/aje/kwv254. pmid: 26994063
11 Wang SV, Pinheiro S, Hua W, etal. STaRT-RWE: structured template for planning and reporting on the implementation of real world evidence studies. *BMJ* 2021;372:m4856. doi: 10.1136/bmj.m4856. pmid: 33436424
12 Duplicate RCT. https://www.rctduplicate.org/
13 Wallach JD, Deng Y, McCoy RG, etal. Real-world Cardiovascular Outcomes Associated With Degarelix vs Leuprolide for Prostate Cancer Treatment. *JAMA Netw Open* 2021;4:e2130587. doi: 10.1001/jamanetworkopen.2021.30587. pmid: 34677594