Journal of Antimicrobial Chemotherapy

Concordance between the results of randomized and non-randomized interventional clinical trials assessing the efficacy of drugs for COVID-19: a cross-sectional study

Daniel Shepshelovich[1,2]*, Dafna Yahav[2,3], Ronen Ben Ami[2,4], Hadar Goldvaser[5,6] and Noam Tau [ID] [2,7]

[1]*Medicine T, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel;* [2]*Sackler Faculty of Medicine, Tel-Aviv University, Tel Aviv, Israel;* [3]*Infectious Diseases Unit, Rabin Medical Center, Beilinson Hospital, Petach Tikva, Israel;* [4]*Infectious Diseases Unit, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel;* [5]*Department of Medical Oncology, Shaare Zedek Medical Center, Jerusalem, Israel;* [6]*Faculty of Medicine, Hebrew University, Jerusalem, Israel;* [7]*Department of Diagnostic Imaging, Chaim Sheba Medical Center, Ramat Gan, Israel*

*Corresponding author. E-mail: shepshelovich@yahoo.com

Objectives: To assess whether results of observational studies of potential anti-COVID-19 drugs were reproduced in subsequent randomized controlled trials (RCTs).

Methods: This was a retrospective cross-sectional study, including studies published online between 1 January and 27 October 2020 that evaluated potential COVID-19 treatments and reported all-cause mortality.

Results: Of 133 comparisons included in 117 studies, most were non-randomized (104/133, 78%). Hydroxychloroquine was the most common drug type, combined with azithromycin ($n = 27$, 20%) or alone ($n = 22$, 16%), followed by IL-6 inhibitors ($n = 36$, 27%) and corticosteroids ($n = 26$, 20%). Seventy-one percent (74/104) of non-randomized studies reported adjusted survival results and only 8% (8/104) adjusted for immortal time bias. Only two RCTs (2/29, 7%) reported significant survival benefit, both reporting treatment with corticosteroids, while 32/104 (31%) non-randomized studies showed statistically significant survival benefit associated with the intervention arm. The results of the majority (28/32, 88%) of non-randomized studies reporting survival benefit were not replicated by large-scale RCTs.

Conclusions: The results of most non-randomized studies reporting survival benefit of potential anti-COVID-19 drugs were not replicated by large RCTs. Regulators and healthcare professionals should exercise caution and resist the pressure to approve and prescribe drugs of unproven efficacy and potential toxicity to optimize patient care and maintain public trust in medical science.

Introduction

Randomized controlled trials (RCTs) are recognized as the gold standard for assessing the efficacy and safety of novel medical interventions. However, RCTs are expensive and time consuming, limiting their effectiveness for delivering immediate results for urgent medical issues.[1]

Facing the exponentially rising tolls of COVID-19-related hospitalizations and death, the medical and scientific communities faced an urgent need to identify effective treatments. Many early reports of potentially effective anti-COVID drugs were observational and some showed survival advantages associated with investigational interventions.[2–4] Some of those reported findings were not reproduced by subsequent RCTs, reflected by rapid changes in treatment recommendations.[5,6]

We aimed to describe the concordance between the reported survival benefit of RCTs and non-randomized studies of potential COVID-19 drugs. We hypothesized that the results of most observational trials will not be replicated by subsequent RCTs.

Methods

Data source and eligibility criteria

We searched PubMed for studies published online between 1 January and 27 October 2020 reporting survival results of medical interventions in COVID-19 patients. The search term combined 'COVID-19 OR SARS-CoV-2' with interventions considered as promising anti-COVID-19 agents at the time of the study's design: hydroxychloroquine±azithromycin, corticosteroids, remdesivir, convalescent plasma, IL-6 inhibitors, antiretroviral drugs and anti-influenza drugs. Search terms included both drug categories and

specific drug names. Three authors (N.T., D.Y. and D.S.) independently reviewed the identified studies and included only those comparing interventional drugs with placebo or with standard-of-care therapy. Studies reporting outcomes of patient groups treated with several anti-COVID-19 interventions were excluded. In studies including several treatment arms compared with placebo, each arm was considered separately for our final analysis.

Data collection

For each study, the following data were collected: journal impact factor (IF), study design (RCT versus non-randomized studies), single centre versus multicentre, patients' COVID-19 disease severity according to the NIH classification (mild, moderate, severe and critical),[7] sample size, patient age and gender, primary study endpoint (overall survival versus other), survival follow-up duration and funding source (industry versus other/none). For non-randomized studies reporting adjusted survival analysis, the adjustment method (propensity score weighting, patient matching or regression) was noted. Adjustments for immortal time bias were also noted. For studies reporting crude mortality rates without adjustments, the relative risk of death was extracted or calculated.

Statistical analysis

Data were reported descriptively for all studies as well as separately for RCTs and non-randomized studies. Associations between study characteristics and study design were assessed using Fisher's exact test for categorical variables and the Mann–Whitney test for continuous variables. Pre-planned sensitivity analyses of survival benefit according to study design included: exclusion of studies unlikely to show statistically significant survival benefit (i.e. studies with patients with mild COVID-19 and those

including <50 patients), exclusion of multi-armed trials and exclusion of studies published in journals with an IF <2. All statistical analyses were conducted using IBM SPSS Statistics v.25.0 (IBM Corp., Armonk, NY, USA).

Results

Trial characteristics

The literature search yielded 3528 unique titles. Exclusion reasons are shown in Figure S1 (available as Supplementary data at *JAC* Online). A total of 117 studies comprising 133 interventional arms were included (Table 1). Eight studies included two intervention arms: six assessed hydroxychloroquine and a combination of hydroxychloroquine and azithromycin, and two evaluated hydroxychloroquine and lopinavir/ritonavir. Four studies included three intervention arms (hydroxychloroquine, azithromycin and a combination of both). Hydroxychloroquine was the most common drug type, combined with azithromycin ($n = 27$, 20%) or alone ($n = 22$, 16%), followed by IL-6 inhibitors ($n = 36$, 27%) and corticosteroids ($n = 26$, 20%).

Most studies ($n = 104$, 78%) were non-randomized. Drug-type distribution was comparable in randomized and non-randomized studies, except for IL-6 inhibitors, which were more commonly represented in non-randomized studies [3/29 (10%) for RCTs versus 33/104 (34%) for non-randomized studies; $P = 0.03$]. RCTs had a larger median patient sample size [243 (IQR = 111–464) for RCTs versus 191 (IQR = 80–550) for non-randomized studies; $P < 0.001$] and were published in journals with significantly higher IFs [median = 45.5 (IQR = 17.1–64.0) for RCTs versus median = 4.3

Table 1. Characteristics of included studies

Study characteristics	All studies ($n = 133$)	Randomized studies ($n = 29$)	Non-randomized studies ($n = 104$)	P
Study drug, *n* (%)				
hydroxychloroquine+azithromycin	27 (20)	4 (14)	23 (22)	0.43
hydroxychloroquine	22 (17)	7 (24)	15 (14)	0.26
IL-6 inhibitors	36 (27)	3 (10)	33 (32)	0.03
corticosteroids	26 (20)	6 (21)	20 (19)	0.80
convalescent plasma	10 (8)	3 (10)	7 (7)	0.45
remdesivir	5 (4)	3 (10)	2 (2)	0.07
other[a]	7 (5)	3 (10)	4 (4)	0.17
IF, median (IQR)	4.8 (3.3–14.8)	45.5 (17.1–64.0)	4.3 (3.2–7.2)	<0.001
Multicentred, *n* (%)	61 (46)	20 (69)	41 (39)	0.006
Sample size, median (IQR)	199 (83–513)	243 (111–464)	191 (80–550)	<0.001
Clinical severity, *n* (%)				
mild	5 (4)	2 (7)	3 (3)	0.30
moderate	29 (22)	7 (24)	22 (21)	0.80
severe	78 (59)	16 (55)	62 (60)	0.68
critical	21 (16)	4 (14)	17 (16)	>0.999
Age (years), median (IQR)	62 (51.5–71)	58.7 (47.6–68.6)	63 (52.5–72.9)	0.31
Percentage male, median (IQR)	62 (56.7–69.5)	61.1 (58.3–66)	62 (56.3–70.5)	0.07
Primary endpoint overall survival, *n* (%)	43 (32)	3 (10)	40 (38)	0.003
Overall survival measurement timeframe (days), median (IQR)	28 (21–30)	28 (21–28)	28 (21–30)	0.36
Industry sponsorship, *n* (%)	9 (7)	6 (21)	3 (3)	0.003

[a]Including five antiretroviral drugs and two anti-influenza drugs.

(IQR = 3.2–7.2) for non-randomized studies; P < 0.001]. More RCTs were multicentred [20/29 (69%) for RCTs versus 41/104 (39%) for non-randomized studies; P = 0.006]. Compared with non-randomized interventions, RCTs were considerably more likely to be sponsored by industry [6/29 (21%) for RCTs versus 3/104 (3%) for non-randomized studies; P = 0.003] and less likely to include survival as a primary outcome [3/29 (10%) for RCTs versus 40/104 (38%) for non-randomized studies; P = 0.003]. Patient clinical severity, age and gender distributions were similar in both groups (Table 1).

Non-randomized study adjustments to survival analysis

Of 104 non-randomized study arms included, 74 (71%) reported adjusted survival analysis between the intervention and control arms, and 30 studies (29%) reported only crude mortality rates. Forty-one studies (55%) used propensity adjustment analysis, 24 (32%) used regression and 9 (12%) used matched cases–controls. Studies adjusted for a median of 8 parameters (IQR = 6–14). Only eight non-randomized studies (8%) contained any type of adjustment for immortal time bias.

Concordance of survival analysis between randomized and non-randomized studies

Statistically significant survival results of the included studies according to drug type and study design are shown in Table 2. Only two (2/29, 7%) RCTs reported significant survival benefit associated with the intervention arm, both testing corticosteroid efficacy. Thirty-two (32/104, 31%) non-randomized studies showed statistically significant survival benefit associated with the intervention arm, mostly (28/32, 88%) not replicated by large RCTs. Results remained unchanged in multiple pre-planned sensitivity analyses (Table S1).

Discussion

Healthcare professionals have published their experience with potential anti-COVID-19 drugs in real-world settings, without randomization of patients to treatment and control arms. Here, we compared the survival benefit reported by these studies to findings from large RCTs evaluating the same drugs. Most reports of survival benefit by non-randomized studies were not reproduced by subsequent large RCTs.

The lack of concordance between the positive results of non-randomized studies and those reported by large RCTs is concerning. First, some non-randomized studies have led to the widespread prescribing of ineffective drugs, with potential adverse effects.[8–10] Second, massive prescribing of some of these drugs led to market shortages, denying effective treatments from patients with medical conditions known to respond to these drugs.[11] Third, the hype and hope created by non-randomized studies,[12] sometimes supported by official agencies,[13,14] with treatments later reported to be ineffective,[15,16] might have created confusion and mistrust in the soundness of medical science, even more so when the excitement–disappointment cycle was repeated in a short time span.[17]

There are several potential explanations for the large survival benefit reported by many non-randomized studies not replicated by subsequent RCTs. Physicians may have reserved expensive, experimental drugs, or those in limited quantity, for patients thought to have better survival chances, introducing selection bias into the study results. Observational studies are also prone to immortal time bias, as patients who die early will not receive the investigational drug, leading to a biased positive treatment effect measure.[18] Publication bias might cause the rapid publication in prestigious journals of studies reporting survival benefit of experimental COVID-19 therapies.

The challenges of conducting large RCTs in settings overwhelmed by COVID-19 are formidable, including the recruitment of many medical centres committed to a single protocol, patient recruitment in the face of experimental uncertainty and a potentially life-threatening infection, drug delivery, blinding and seamless data collection. However, even in this state of emergency, conducting large RCTs is both feasible and rewarding. Large, multiarm RCTs, such as the RECOVERY and SOLIDARITY trials, have provided the highest-quality data regarding the efficacy, or lack thereof, of potential COVID-19 treatments.[15,16] Until such high-quality data are available, observational data should be used with caution

Table 2. Significant survival benefit of included trials

Drug type	Randomized studies (n = 29)	Non-randomized studies (n = 104)		
		all non-randomized trials	trials reporting adjusted survival analysis (n = 49)	trials reporting unadjusted survival analysis (n = 55)
All drugs	2/29 (7%)	32/104 (31%)	23/49 (47%)	9/55 (16%)
Hydroxychloroquine+ azithromycin (n = 27)	0/4 (0%)	6/23 (26%)	5/16 (31%)	1/7 (14%)
Hydroxychloroquine (n = 22)	0/7 (0%)	5/15 (33%)	4/10 (40%)	1/5 (20%)
IL-6 inhibitors (n = 36)	0/3 (0%)	13/33 (39%)	8/10 (80%)	5/23 (22%)
Corticosteroids (n = 26)	2/6 (33%)	4/20 (20%)	3/8 (37%)	1/12 (8%)
Convalescent plasma (n = 10)	0/3 (0%)	1/7 (14%)	1/3 (33%)	0/4 (0%)
Remdesivir (n = 5)	0/3 (0%)	2/2 (100%)	1/1 (100%)	1/1 (100%)
Other (n = 7)	0/3 (0%)	1/3 (33%)	1/1 (100%)	0/3 (0%)

to guide patient care. Healthcare professionals should consider only high-quality observational studies in their decision-making; those using a quasi-randomized design (e.g. drugs available in one hospital but not in another serving similar patients), following large patient cohorts for long durations, adjusting for patient characteristics and potential biases, using patient-centred endpoints and reporting results according to the STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) guidelines. Results reported by such high-quality observational studies can be very similar to results reported by RCTs.[19,20] Large clinical datasets worldwide can be analysed in real-time using predetermined methods to guide clinicians and regulators in uncertain times, until higher-quality data are available. Comparable data collection, analysis of different datasets, data sharing and transparency can facilitate rapidly available, high-quality observational data in times of crisis. Ultimately, different data types from variable sources should be integrated to optimize clinical and regulatory decision-making.

This study has several limitations. Inclusion of different anti-COVID-19 drug classes, or of endpoints other than overall survival, might have led to different results. Multivariable analysis of the association between study characteristics and concordance of the results of non-randomized studies and large RCTs was not undertaken due to an insufficient number of RCTs to properly fit a multivariable model. Lastly, the evaluation of some anti-COVID-19 drugs is evolving and some included non-randomized studies might yet be supported by future RCTs.

In conclusion, the results of most non-randomized studies reporting survival benefit of potential anti-COVID-19 drugs were not replicated by large RCTs. Until high-quality evidence is available, regulators and healthcare professionals should exercise caution and resist the pressure to approve and prescribe drugs of unproven efficacy and potential toxicity to optimize patient care and maintain public trust in the quality of medical science.

Funding

Transparency declarations

H.G. reports personal fees from Roche (honorarium), Pfizer (honorarium), Novartis (honorarium and consulting) and Oncotest (honorarium), all outside of the submitted work. All other authors: none to declare.

Author contributions

All authors participated in the research and preparation of the manuscript.

Supplementary data

Figure S1 and Table S1 are available as Supplementary data at *JAC* Online.

References

1 DiMasi JA, Grabowski HG, Hansen RW. Innovation in the pharmaceutical industry: new estimates of R&D costs. *J Health Econ* 2016; **47**: 20–33.

2 Grein J, Ohmagari N, Shin D *et al.* Compassionate use of remdesivir for patients with severe Covid-19. *N Engl J Med* 2020; **382**: 2327–36.

3 Shen C, Wang Z, Zhao F *et al.* Treatment of 5 critically ill patients with COVID-19 with convalescent plasma. *JAMA* 2020; **323**: 1582–9.

4 Geleris J, Sun Y, Platt J *et al.* Observational study of hydroxychloroquine in hospitalized patients with Covid-19. *N Engl J Med* 2020; **382**: 2411–8.

5 US FDA. FDA News Release. Coronavirus (COVID-19) Update: FDA Revokes Emergency Use Authorization for Chloroquine and Hydroxychloroquine. https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-revokes-emergency-use-authorization-chloroquine-and.

6 WHO. WHO Recommends Against the Use of Remdesivir in COVID-19 Patients. https://www.who.int/news-room/feature-stories/detail/who-recommends-against-the-use-of-remdesivir-in-covid-19-patients.

7 NIH. COVID-19 Treatment Guidelines. Clinical Spectrum. https://www.covid19treatmentguidelines.nih.gov/overview/clinical-spectrum/.

8 Vaduganathan M, Meijgaard J, Mehra MR *et al.* Prescription fill patterns for commonly used drugs during the COVID-19 pandemic in the United States. *JAMA* 2020; **323**: 2524–6.

9 Danion F, Ruch Y, Fourtage M *et al.* The good, the bad, and the hoax: when publication instantaneously impacts treatment strategies for COVID-19. *Antimicrob Agents Chemother* 2020; **64**: e01127-20.

10 Elavarasi A, Prasad M, Seth T *et al.* Chloroquine and hydroxychloroquine for the treatment of COVID-19: a systematic review and meta-analysis. *J Gen Intern Med* 2020; **35**: 3308–14.

11 Mendel A, Bernatsky S, Askanase A *et al.* Hydroxychloroquine shortages among patients with systemic lupus erythematosus during the COVID-19 pandemic: experience of the Systemic Lupus International Collaborating Clinics. *Ann Rheum Dis* 2021; **80**: 1–2.

12 Liu M, Caputi TL, Dredze M *et al.* Internet searches for unproven COVID-19 therapies in the United States. *JAMA Intern Med* 2020; **180**: 1116–8.

13 US FDA. Letter to Dr Rick Bright. Re: Request for Emergency Use Authorization for Use of Chloroquine Phosphate or Hydroxychloroquine Sulfate Supplied From the Strategic National Stockpile for Treatment of 2019 Coronavirus Disease. https://www.fda.gov/media/136534/download.

14 US FDA. FDA News Release. Coronavirus (COVID-19) Update: FDA Issues Emergency Use Authorization for Potential COVID-19 Treatment. https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-issues-emergency-use-authorization-potential-covid-19-treatment.

15 Pan H, Peto R, Henao-Restrepo AM *et al.* Repurposed antiviral drugs for Covid-19 - interim WHO Solidarity Trial results. *N Engl J Med* 2021; **384**: 497–511.

16 RECOVEY. Randomised Evaluation of COVID-19 Therapy. https://www.recoverytrial.net.

17 Kreps SE, Kriner DL. Model uncertainty, political contestation, and public trust in science: evidence from the COVID-19 pandemic. *Sci Adv* 2020; **6**: eabd4563.

18 Tleyjeh IM. The misleading "pooled effect estimate" of crude data from observational studies at critical risk of bias: the case of tocilizumab in COVID-19. *Clin Infect Dis* 2020; doi:10.1093/cid/ciaa1735.

19 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; issue **4**: MR000034.

20 Fralick M, Kesselheim AS, Avorn J *et al.* Use of health care databases to support supplemental indications of approved medications. *JAMA Intern Med* 2018; **178**: 55–63.