Remdesivir for COVID-19: challenges of underpowered studies

Published Online
April 29, 2020
https://doi.org/10.1016/S0140-6736(20)31023-0

See Online/Articles
https://doi.org/10.1016/S0140-6736(20)31022-9

In *The Lancet*, Yeming Wang and colleagues[1] report a randomised trial of remdesivir (200 mg on day 1 followed by 100 mg on days 2–10, in single daily infusions) versus placebo for adults with severe coronavirus disease 2019 (COVID-19) in ten hospitals in Wuhan, China. The authors report on 236 patients (140 [59%] men and 96 [41%] women; median age 65 years [IQR 56–71]), with inconclusive findings on the primary outcome of time to clinical improvement, defined as a two-point improvement on a 6-point ordinal scale,[2] a hazard ratio of 1·23 (95% CI 0·87–1·75; favouring remdesivir), and median observation times of 21 days (IQR 13–28) in the remdesivir group versus 23 days (15–28) in the placebo group (a non-significant difference).

The study was well designed—a double-blind, placebo-controlled, multicentre, randomised trial—and well conducted, with high protocol adherence and no loss to follow up. Randomised evidence was needed following high-profile publications on the first US COVID-19 case[3] and the subsequent compassionate use of remdesivir in a 53-patient case series,[4] which, coupled with in-vitro and animal model evidence, had generated high expectations of remdesivir efficacy.

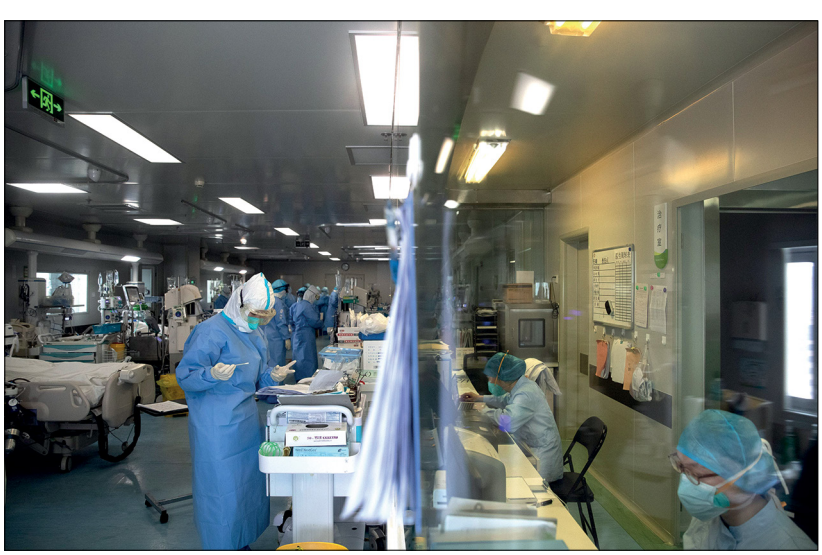
Promising signals from observational data must be rigorously confirmed or refuted in high-quality randomised trials—particularly given that for COVID-19 no proven safe and effective treatments yet exist. Ideally, efficacy-based trials, including proof-of-mechanism studies, should precede larger pragmatic effectiveness trials.[5] That is additionally challenging in a pandemic, and the temptation to lower the threshold of convincing evidence must be resisted, because adopting ineffective and potentially unsafe interventions risks only harm without worthwhile benefit, while making it even harder to undertake trials to find truly effective and safe interventions. We have already seen other drugs, repurposed for COVID-19, including hydroxychloroquine[6] and lopinavir–ritonavir,[7] report disappointing findings so far in randomised trials after early promise.

Wang and colleagues' study[1] stopped early after 237 of the intended 453 patients were enrolled, because by March 12 there were no further patients meeting eligibility criteria admitted in Wuhan. The study closed on March 29, having begun on Feb 6.

Here, stopping early gives an underpowered trial, which taken alone, gives inconclusive findings. The study has not shown a statistically significant finding that confirms a remdesivir treatment benefit of at least the minimally clinically important difference, nor has it ruled such a benefit out. The study sought a treatment effect of hazard ratio (HR) 1·40, translating to reducing median time to clinical improvement to 15 days (remdesivir) versus 21 days (placebo). The observed HR of 1·23 suggests that a benefit, if it exists, might be smaller than anticipated. This study is the first randomised trial of intravenous remdesivir in patients with severe COVID-19, so it is difficult to know what the minimally clinically important difference is.[8] That will depend on a complex reckoning of evidence for effectiveness, safety, acceptability, access, and cost. It is possible that even if the 453-patient target was reached, the study would have still been underpowered if a minimally clinically important difference of less than an HR of 1·4 was warranted.

However, likewise, a larger benefit might exist, or remdesivir might actually do harm. It is unknown—more data are needed. Fortunately, ClinicalTrials.gov indicates that five randomised trials involving remdesivir are recruiting globally, with one in severe COVID-19 from Gilead (NCT04292899), the drug manufacturer, with a target of 6000 participants; naively, this trial should be adequately powered.

In the meantime, how can the findings of Wang and colleagues be interpreted? The statistical reporting is



STR/Getty Images

clear, stating that the main findings were not statistically significant and acknowledging that the trial was underpowered (their post-hoc calculation indicated a power of 58% given the 236 participants with available data). However, a trial is not just its primary clinical outcome—there are important data on safety, viral load, and secondary outcomes. 22 (14%) of 158 patients on remdesivir died versus ten (13%) of 78 on placebo, and there was no signal that viral load decreased differentially over time between remdesivir and placebo groups. Furthermore, there were no differential signals on safety. Analyses were very similar under both the intention-to-treat and per-protocol principles.

The authors also report primary outcome subgroup analyses. Only patients who were 12 days or less from illness onset were eligible overall, so a prespecified subgroup analysis investigated those who started study treatment up to 10 days versus more than 10 days (up to 12 days) from illness onset. Of course, even with an adequately powered study, subgroup analyses are generally not powered (and here, the 2:1 allocation further reduced power). There was no significant interaction of 10 days or less versus more than 10 days—ie, little support statistically of treatment effect moderation by time of initiation. Nor was either the 10 days or less or the more than 10 days within-subgroup treatment effects significant. Nonetheless, the authors give prominence to the 10 days or less subgroup, reporting a non-statistically significant HR of 1·52 (95% CI 0·95 to 2·43), median 18 days (IQR 12 to 28) versus 23 days (15 to 28), and a non-significant reduction in mortality (difference –3·6% [95% CI –16·2 to 8·9]). There was a possible baseline imbalance with 71 (45%) remdesivir patients versus 47 (60%) placebo patients in the 10 days or less subgroup, and possibly more patients with hypertension, diabetes, and coronary heart disease allocated to remdesivir than placebo, making interpretation even more difficult. Subgroup analyses, particularly for phase 3 confirmatory effectiveness trials, have justifiably been criticised[9] and even ridiculed.[10] Giving a subgroup analysis prominence over the primary analysis is unfortunately common. In early phase studies in a pandemic, little is known for certain, and it seems biologically plausible that treating patients earlier could be more effective. Nonetheless, as well as being vigilant against overinterpretation, we need to ensure that hypotheses generated in efficacy-based trials, even in subgroups, are confirmed or refuted in subsequent adequately powered trials or meta-analyses.

We have already seen how different interpretations will be put on these results, with the unintended early release of this study's results on the WHO website.[11] This underlines how labelling of trials is mistaken as positive or negative—equating a p>0·05 with no evidence of benefit. There has been a welcome discussion of p value limitations recently.[12] An absence of statistical significance in an underpowered trial means that the findings are inconclusive. The particular challenges of delivering pandemic trials underline the importance of data sharing, allowing rapid curation of relevant datasets for individual patient data meta-analyses.[13] With each individual study at heightened risk of being incomplete, pooling data across possibly several underpowered but high-quality studies looks like our best way to obtain robust insights into what works, safely, and on whom. We eagerly await the ongoing trials.

I am employed by University of Edinburgh and by the UK Medical Research Council/National Institutes of Health Research as Chair of the Efficacy and Mechanisms Evaluation Funding Committee.

John David Norrie
j.norrie@ed.ac.uk

Edinburgh Clinical Trials Unit, Usher Institute, Edinburgh EH16 4UX, UK

1 Wang Y, Zhang D, Du G, et al. Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. *Lancet* 2020; published online April 29. https://doi.org/10.1016/S0140-6736(20)31022-9.
2 WHO. COVID-19 therapeutic trial synopsis. Geneva: World Health Organization, Feb 18, 2020. https://www.who.int/blueprint/priority-diseases/key-action/COVID-19_Treatment_Trial_Design_Master_Protocol_synopsis_Final_18022020.pdf (accessed 25 April 2020).
3 Holshue ML, DeBolt C, Lindquist S, et al. First case of 2019 novel coronavirus in the United States. *N Engl J Med* 2020; **382:** 929–36.
4 Grein J, Ohmagari N, Shin D, et al. Compassionate use of remdesivir for patients with severe COVID-19. *N Engl J Med* 2020; published online April 10. DOI:10.1056/NEJMoa2007016.
5 Ford I, Norrie J. Pragmatic Trials. *N Engl J Med* 2016; **375:** 454–63.
6 Tang W, et al. Hydroxychloroquine in patients with COVID-19: an open-label, randomised, controlled trial. medRxiv 2020; published online April 14. https://doi.org./10.1101/2020.04.10.20060558 (preprint).
7 Cao B, Wang Y, Wen D, et al. A trial of lopinavir–ritonavir in adults hospitalised with severe COVID-19. *N Engl J Med* 2020; published March 18. DOI:10.1056/NEJMoa2001282.
8 Cook JA, Julious SA, Sones W, et al. DELTA^2 guidance on choosing the target difference and undertaking and reporting the sample size calculation for a randomised controlled trial. *BMJ* 2018; **363:** k3750.
9 Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. *Lancet* 2000; **355:** 1064–69.
10 Sleight P. Debate: subgroup analyses in clinical trials: fun to look at—but don't believe them! *Curr Control Trials Cardiovasc Med* 2000; **1:** 25–27.
11 BBC. Hopes dashed as coronavirus drug remdesivir 'fails first trial'. April 23, 2020. www.bbc.co.uk/news/world-52406261 (accessed April 25, 2020).
12 Wasserstein RL, Schirm AL, Lazar NA. Moving to a world beyond "p<0.05". *Am Stat* 2019; **73:** 1–19.
13 Simmonds MC, Higgins JPT, Stewart LA, Tierney JF, Clarke MJ, Thompson SG. Meta-analysis of individual patient data from randomized trials: a review of methods used in practice. *Clin Trials* 2005; **2:** 209–17.