Hydroxychloroquine for the treatment of COVID-19: the importance of scrutiny of positive trials

Ronald Chow[1], Charles B. Simone II[2], Michael Lock[1]

[1]Schulich School of Medicine & Dentistry, University of Western Ontario, London, ON, Canada; [2]Department of Radiation Oncology, New York Proton Center, New York, NY, USA

Correspondence to: Ronald Chow, BMSc. Schulich School of Medicine & Dentistry, University of Western Ontario, London, ON, Canada. Email: ronald.chow@uwo.ca.

Coronaviruses primarily target the human respiratory system. Over the past two decades, these viruses were responsible for three epidemics. Prior to 2019, the two coronavirus public health threats were caused by the Severe Acute Respiratory Syndrome (SARS-CoV) and Middle East Respiratory Syndrome (MERS-CoV) coronaviruses (1). In December 2019, an outbreak of a novel coronavirus was reported in Wuhan, China (2) with subsequent rapid spread around the world. On March 12, 2020, the World Health Organization (WHO) officially declared COVID-19 a pandemic. At the time of this writing in July 2020, there have been over 617,000 deaths worldwide (3), and the virus has required a significant role from palliative medicine providers to provide supportive care to patients and families impacted by COVID-19.

Although the scientific community is racing to find possible active agents for COVID-19, there currently is no effective treatment (4). With no effective treatment, supportive care is the only option for patients, caregivers and family; as COVID-19 is known to produce significant symptom burden across multiple organ systems, multidisciplinary care will need to be expended while a treatment is being developed. Dexamethasone may be effective, but only in a subset of hospitalized patients (5). Remdesivir, which was developed to treat other viral diseases including Ebola, may also be effective. One of the first purportedly positive trials was published by Gautret *et al.* (6), a small, single-arm study reporting hydroxychloroquine to be a potentially-promising treatment to significantly reduce the viral carriage. Earlier versions of the results were released previously due to the urgent need to provide data

to health care providers as the virus was spreading rapidly across continents. This includes the publication in medRxiv, which had slightly different results denoted; as these were not peer reviewed, we only considered the data from the final peer reviewed publication in the *International Journal of Antimicrobial Agents* (6). Gautret *et al.* treated 20 cases with hydroxychloroquine in Marseille as the first large surge of cases reached the critical care departments of Europe; this study found that the drug was significantly associated with viral load reduction in COVID-19 patients (6). After 6 days of treatment, 100%, 57.1% and 12.5% of the hydroxychloroquine plus azithromycin group, hydroxychloroquine alone group, and control group had negative viral reverse transcriptase-PCR (RT-PCR) tests, respectively. This observation was statistically significant at P<0.01.

The dissemination of early positive findings is especially important during a public health crisis, where scientific discoveries need to be timely. Positive trial results could motivate the scientific community to use the agents in infected patients, or commit fast-track funding and clinical resources to explore this drug regimen further. However, it is important to pause for a moment to ensure that the purportedly positive trial results are indeed robust enough to pursue further. Making inferences about this regimen without robust results may take precious time and resources away from pursuing other promising therapies. Furthermore, the benefits of a new regimen need to be weighed against its potential harms; therefore, hydroxychloroquine use needs to be carefully weighed against its potential harms such as prolonging of the QTc

Table 1 FI example

Study	Experimental arm	Control arm	P
Original study			0.033
Number of events	5	0	
Total number of patients	10	10	
Fragile study			0.141
Number of events	5	0+1	
Total number of patients	10	10	

FI, fragility index.

interval that can be life threatening.

Today, the U.S. Food and Drug Administration has released cautionary statements around the use of hydroxychloroquine for the treatment of COVID-19, warning about the risk of heart rhythm problems (7). In hindsight, subsequent trial resources used to investigate the efficacy and safety of hydroxychloroquine might have been more prudently diverted to other treatment possibilities.

Therefore, it is important to assess if the first positive trial results of hydroxychloroquine by Gautret *et al.* were robust enough to warrant further investigation. There have been several critical appraisals of this trial, but appraisals are in many ways subjective with no rigorous assessment performed to date. Therefore, in this urgent and grave subject, we sought to more quantitatively assess the value of this trial. A fragility index (FI) was employed to assess for robustness of this positive trial. As described by Walsh *et al.* (8), this index complements the P value by informing the robustness of results. The index was developed by leaders in the field of evidence-based medicine, including Guyatt and Sackett, to provide information beyond the much criticized, but commonly reported P value. The FI is most applicable and useful in trials with small numbers of events where the P values can often erroneously suggest that a trial is positive; this may be the case for the small-sample trial published by Gautret *et al.* The index adds information beyond that provided by historically used measures such as P value, number of events, confidence interval and sample size. The aim of this editorial is to quantitatively assess the first positive trial reporting on hydroxychloroquine for COVID-19 to serve as a model for rigorous assessments that can be performed for future COVID-19 research involving small sample sizes of promising treatment options.

The FI is the minimum number of patients that needs to change from a negative to a positive event, counterfactually, in the non-favored group for a study's positive result to become negative/non-significant. The higher the FI, the more robust a positive trial's results. The median FI of published positive trials is reported to be between 2 and 3 (9-12).

For example, in a trial where there are 10 patients in each of the control and experimental arms and five patients with a positive event in the experimental arm, the results of one patient in the control arm needs to change from a negative to a positive event for a non-significant difference (*Table 1*). As the number of events that needs to be changed is 1, the FI is therefore 1.

For the Gautret *et al.* trial, the number of patients with negative nasopharyngeal RT-PCR COVID-19 results among the number of patients per hydroxychloroquine and control arms, per day, were analyzed. To calculate the FI, the event number was incrementally increased by positive integer values until there was no statistical difference between the two trial arms. Fisher's exact test was calculated for each counterfactual simulation, using Stata 15 (StataCorp, College Station, TX, USA).

At day 3, 10 of 20 hydroxychloroquine patients had a virological cure, compared to 1 of 16 patients receiving placebo. A minimum of 2 additional control patients theoretically needed to have a virological cure in order to observe a nonsignificant difference. As such, the FI for day 3 was 2, increasing the P value from 0.009 to 0.083 (*Table 2*). The FI for virological cure is 1 for day 4, 2 for day 5, and 4 for day 6.

This analysis demonstrates that the FI in the study by Gautret *et al.* ranged from 1 to 4 across its study duration. This would suggest that if as little as one event changed, the hydroxychloroquine trial would have become statistically non-significant. Though the Gautret *et al.* study has low indices, these indices are comparable to many other

3718

Chow et al. The importance of scrutiny of positive trials

Table 2 FI—virological cure (RT-PCR-negativity) by day 3

Study	Hydroxychloroquine treated patients	Control patients	P
Original study			0.009
Number of patients with virological cure	10	1	
Total number of patients	20	16	
Fragile study			0.083
Number of patients with virological cure	10	1+2	
Total number of patients	20	16	

FI, fragility index; RT-PCR, reverse transcriptase-PCR.

published trials. Evaniew *et al.* reviewed 40 orthopedic spine surgery trials and reported a median FI of 2 (9). Pediatric orthopedic trials have a median FI of 3, as reported in a review of 17 trials by Khormaee *et al.* (12). In the field of ophthalmology, the median index was 2 across 156 published trials (11). Oncology trials are reported to have a median index of 2, according to a 36-trial review by Del Paggio and Tannock (13). Perhaps the most pertinent assessment with respect to the COVID-19 trial by Gautret *et al.* is a calculation of the FI in 57 critical care trials by Ridgeon *et al.* that found a median FI of 2 (9). The results of Gautret *et al.* are as robust, statistically, as half of published trials, and therefore did warrant further trial resources.

As a result of this promising first study, many trials were subsequently initiated. Serious adverse events were noted in the hydroxychloroquine arm, and currently the United States Food and Drug Administration urges caution around using hydroxychloroquine (13). Such significant reversals unfortunately confirm that a better allocation of study resources could have been employed, but it does not mean that conducted trials were unwarranted. In this setting of hydroxychloroquine, the Gautret *et al.* trial provided statistical evidence as robust as the majority of the published literature to encourage the investigation of hydroxychloroquine.

In this case, methodological concerns may be a reason for significant reversal on the view of hydroxychloroquine. Gautret *et al.* published essentially preliminary data from a practical and ongoing trial as they treated their first patients in Marseille, France. Thirty-six hospitalized patients with positive nasopharyngeal sample positive PCR results were included (20 treatment group and 16 untreated controls). Patients received hydroxychloroquine sulfate 200 mg three times per day for 10 days, while six patients in the treatment group also received azithromycin. A comparison group was collected primarily from patients in Marseille, Avignon, Briancon, and Nice. The primary endpoint was clearance of the virus on RT-PCR analysis of nasopharyngeal specimens on day 6. Six patients were not included as they were lost to follow-up. This is pertinent given that the FI indicates that the number of patients required to change the conclusion is 4. However, Walsh *et al.* in their review of randomized trials published in high impact journals found that 52.9% of trials had higher loss to follow-up levels than their Fragility Indices. The results of Gautret *et al.* in this respect are similar to other published trials.

Also of note is the lack of systematic RT-PCR testing between the groups, which represents a time bias. The hydroxychloroquine group was, on average, more severely ill with COVID-19. It is possible that their RT-PCR tests would become negative quicker, even without treatment, due to a time bias. Furthermore, RT-PCR cannot even be used as a test of virological cure, as it measures viral RNA rather than the live virus. Given this methodological flaw, more trials are required to verify this purportedly positive trial.

While this analysis adds quantification to study assessment, with the FI providing insight into the value of a trial, its use is not without limitations. In fact, its application to non-randomized trials may be more limited, as there is limited control of confounding variables. To further understand the results of Gautret *et al.* and its applicability to future research and practice, the calculation of the FI should ideally be conducted alongside appraisals on epidemiological aspects (i.e., trial design to control for confounder, inclusion/exclusion criteria, and choice of outcome), such as the appraisal by Kim *et al.* (14).

In conclusion, the FI may be a simple tool to assess trials, particularly when the issue may be clouded by non-clinical or political agendas. For the first positive trial result

by Gautret *et al.*, the FI calculations indicated that the data were not robust, but they did have sufficient evidence to support the subsequent level of investigation and resource utilization that occurred worldwide. Fortunately, in this scenario, the indices support the past course of action. FIs should be used more routinely and considered a priori, in the interest of preserving scarce clinical trial resources and to provide greater clarity in determining which agents warrant additional investigation.

Footnote

Provenance and Peer Review: This article was commissioned by the editorial office, *Annals of Palliative Medicine*. The article did not undergo external peer review.

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/apm-20-1538). CBS serves as the Editor-in-Chief of *Annals of Palliative Medicine*. ML reports consulting fees from Ferring, Abbvie, Sanofi, and AstraZeneca in the past 10 years outside the submitted work. The other author has no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. Rothan HA, Byrareddy SN. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. J Autoimmun 2020;109:102433.

2. Lai CC, Shih TP, Ko WC, et al. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. Int J Antimicrob Agents 2020;55:105924.

3. World Health Organization. Coronavirus disease (COVID-19) pandemic. Geneva: World Health Organization, 2020. Available online: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 (cited 2020 Jun 20).

4. Baden LR, Rubin EJ. Covid-19 - the search for effective therapy. N Engl J Med 2020;382:1851-2.

5. Horby P, Lim WS, Emberson J, et al. Effect of dexamethasone in hospitalized patients with COVID-19: preliminary report. medRxiv 2020. doi: 10.1101/2020.06.22.20137273.

6. Gautret P, Lagier JC, Parola P, et al. Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. Int J Antimicrob Agents 2020;56:105949.

7. U.S. Food & Drug Administration. FDA cautions against use of hydroxychloroquine or chloroquine for COVID-19 outside of the hospital setting or a clinical trial due to risk of heart rhythm problems. 2020. Available online: https://www.fda.gov/drugs/drug-safety-and-availability/fda-cautions-against-use-hydroxychloroquine-or-chloroquine-covid-19-outside-hospital-setting-or (cited 2020 Jun 22).

8. Walsh M, Srinathan SK, McAuley DF, et al. The statistical significance of randomized controlled trial results is frequently fragile: a case for a Fragility Index. J Clin Epidemiol 2014;67:622-8.

9. Evaniew N, Files C, Smith C, et al. The fragility of statistically significant findings from randomized trials in spine surgery: a systematic survey. Spine J 2015;15:2188-97.

10. Ridgeon EE, Young PJ, Bellomo R, et al. The fragility index in multicenter randomized controlled critical care trials. Crit Care Med 2016;44:1278-84.

11. Shen C, Shamsudeen I, Farrokhyar F, et al. Fragility of results in ophthalmology randomized controlled trials: a systematic review. Ophthalmology 2018;125:642-8.

12. Khormaee S, Choe J, Ruzbarsky JJ, et al. The fragility of statistically significant results in pediatric orthopaedic randomized controlled trials as quantified by the fragility index: a systematic review. J Pediatr Orthop 2018;38:e418-23.

13. Del Paggio JC, Tannock IF. The fragility of phase 3 trials supporting FDA-approved anticancer medicines: a

retrospective analysis. Lancet Oncol 2019;20:1065-9.

14. Kim AHJ, Sparks JA, Liew JW, et al. A rush to judgment? Rapid reporting and dissemination of results and its consequences regarding the use of hydroxychloroquine for COVID-19. Ann Intern Med 2020;172:819-21.