nature communications

Matters arising

https://doi.org/10.1038/s41467-025-60478-x

On meta-analytic models and the effect of hydroxychloroquine use in COVID-19

Received: 15 January 2024

Accepted: 19 May 2025

Published online: 11 July 2025

Check for updates

Diego Pasquier

ARISING FROM C. Axfors et al. *Nature Communications* https://doi.org/10.1038/s41467-021-22446-z (2021)

Hydroxychloroquine (HCQ) has been widely tested as a potential treatment in COVID-19. The largest randomized trial in hospitalized patients (the RECOVERY trial) found a numerically larger rate of all-cause mortality in the HCQ arm compared with standard of care alone (RR, 1.09, 95% CI, 0.96 −1.23)[1]. These data clearly exclude a clinical benefit and raise the question of a possible increased fatality rate caused by widespread HCQ use in the early phase of the pandemic. Since the RECOVERY trial was not powered to detect such ~10% increase in mortality, it is natural to pool the data with those of similar trials to obtain more precise estimates. The study of Axfors et al.[2] is a high-quality systematic review that relevantly addresses this question and leads to the conclusion of a significant increase of mortality associated with HCQ (OR, 1.11, 95% CI, 1.02 −1.20, $p = 0.02$). This is in contrast with other meta-analyses based on similar sets of trials, which reported wider confidence intervals[3–5]. In a more recent meta-analysis, including more trials and with the final Solidarity results, the trend was weaker[6]. This eleven percent increase in mortality is now being used to estimate the number of deaths caused by HCQ in various countries[7].

The shorter confidence intervals reported in Axfors et al. originate mostly from the meta-analytic model used. Axfors et al. also included some extra unpublished studies, but these turned out to carry only 7.8% of the total weight, which cannot explain such tighter confidence intervals. Here, we point out some difficulties related to the use of the Hartung–Knapp random-effect model in this dataset. First, the rationale for the Hartung–Knapp approach given by Axfors et al. is inconsistent with the nature of the adjustment. Second, the Hartung–Knapp adjustment may result in an effectively increased precision in certain cases[8]. This is a known problematic feature of the method (that can be fixed in several ways). Some of the results reported here are a good example of a rather dramatic effect of the adjustment in real datasets. Finally, we conclude that there remains uncertainty regarding a potential adverse HCQ effect, in particular in light of the most recent meta-analysis. To further illustrate our point on the choice of model, we also discuss the case of HCQ effect on COVID-19 hospitalization in outpatients[9].

Results

Random-effect models and Hartung–Knapp adjustment

In ref. 2, it is stated that "In our protocol, we prespecified a random-effects model of the Hartung–Knapp–Sidik–Jonkman (HKSJ) approach, in order to provide more equality of weights between trials with moderate to large size (than, e.g., the DerSimonian–Laird approach)." This is a rather surprising justification because, by construction, the weights in the two approaches are exactly the same, meaning that the point estimates of the meta-analyses always agree and only the confidence intervals and p-values can differ.

In either case the average effect is given by

$$\hat{\mu} = \frac{\sum_i w_i x_i}{\sum_i w_i} \,, \tag{1}$$

where x_i is the estimate of study i and w_i is its weight, given by the inverse of its variance, i.e. $w_i = 1/s_i^2$. It is assumed that x_i is normally distributed, i.e $x_i \sim \mathcal{N}(\mu, s_i^2)$, with $s_i^2 = \sigma_i^2 + \tau^2$, where σ_i^2 is the within-study sampling variance and τ^2 is the between-study variance to be estimated.

The standard way ("DL" meta-analysis) of calculating confidence intervals for μ consists in treating the weights as known parameters, and using a normal distribution. One can estimate the variance of $\hat{\mu}$ as $\hat{\mathrm{var}}(\hat{\mu}) = (\sum_i w_i)^{-1}$, and calculate confidence intervals

$$\hat{\mu} \pm z_{\alpha/2} \sqrt{\hat{\mathrm{var}}(\hat{\mu})} \,, \tag{2}$$

where α is the significance threshold and z are the quantiles of a normal distribution. In reality, there exists uncertainty on the parameters, because one uses point estimates for the variances rather than true values, which can lead to an increase of Type I error in many scenarios.

The Hartung–Knapp adjustment (ref. 10, also proposed by Sidik and Jonkman in ref. 11) usually improves the situation and generally gives confidence intervals with better coverage properties. The approach differs from the standard one in two ways: a rescaled variance is used ($\hat{\mathrm{var}}_{\mathrm{HK}}(\hat{\mu}) = q\hat{\mathrm{var}}(\hat{\mu})$) and a Student distribution is assumed. Within the Hartung–Knapp approach, the confidence interval for a meta-analysis of k studies is given by

$$\hat{\mu} \pm t_{k-1,\alpha/2} \sqrt{q\hat{\mathrm{var}}(\hat{\mu})} \,, \tag{3}$$

where $t_{k-1,\alpha/2}$ is the quantile of the Student distribution with $k - 1$ degrees of freedom. The p-value can be evaluated from the

Institute of Physics, Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland. e-mail: pasquierdiego@gmail.com

approximate pivot $\mu/\sqrt{\mathrm{var}_{\mathrm{HK}}(\mu)} \sim t_{k-1}$. The scale factor q is given by

$$q = \frac{1}{k-1}\sum_i w_i(x_i - \hat{\mu})^2 \,. \quad (4)$$

Using q calculated from Eq. (4) does not always come without difficulties, especially if there is great variability in study sizes. This is because small studies contribute equally to q, which tends to dilute the signal of the larger ones. In cases where $\hat{\tau}^2$ is estimated to be negative from either the DL or PM scheme, q derived from Eq. (4) with $\hat{\tau}^2$ truncated to zero will be less than unity and possibly arbitrarily small. Examining Eq. (3), it is clear that it can then give an unnaturally small variance, with an undesired increase of Type I error as a consequence. This is confirmed by simulation studies, that showed too short confidence intervals associated with this scheme for most of the scenarios that are relevant for practical purpose[12], even though the model is exact in certain extreme limits (e.g. if the studies all have the exact same variance). For studies with few events, which are numerous in the HCQ meta-analysis, the assumption of a normal distribution for the log odds ratio is violated, which will skew the distribution of q towards small values and affect the coverage properties of the HKSJ model.

Several fixes have been proposed. In ref. 12, it was shown that simply substituting q by $q' = \mathrm{Max}(1, q)$ in Eq. (3) (i.e. truncating q similarly to $\hat{\tau}^2$) gives satisfactory results, although there was a loss of power in some scenarios. Another possibility would be to present both standard and HKSJ confidence intervals, and consider the widest of the two as the main result.

HCQ effect in meta-analysis

Inspecting the cumulative meta-analyses of ref. 2 (Fig. 3 of Axfors et al.) reveals some startling results. It is rather surprising that adding only one (with at least one event) small study to the RECOVERY trial substantially reduces the confidence interval (1.01–1.20), even though this trial (NO COVID-19) carries negligible weight and adds only one event per arm. In Fig. 1a, we reproduced Fig. 3b of Axfors et al. to emphasize an even more striking result for the subgroup of published studies (which constitute more than 90% of the total weight). One can see that adding one more small trial (COVID-PEP, adding only one event per arm) leads to an exceptionally short 95% confidence interval (1.08–1.13) despite virtually no information being added. These results are conceptually problematic because one would normally require a much larger number of events to obtain such tight confidence intervals.

In Table 1, we report the calculated scale factors q, for several subgroups of studies that were associated with a significant difference in Axfors et al., alongside the calculated p-values for both $q < 1$ (calculated from Eq. (4)) and $q = 1$. In all cases, the point estimate is close to OR ≈ 1.11. As can be seen, the smaller p-values originate from the small scale factor rather than from the accumulated totality of evidence. The modified Hartung–Knapp approach with truncated q therefore gives results that are similar to other meta-analyses[3–5]. The 95% confidence interval for all studies is (0.97 –1.26, $p = 0.11$), which is also similar to that of the fixed-effect model (0.98– 1.25, $p = 0.09$).

HCQ effect in COVID-19 outpatients

It is also instructive to investigate the effect of meta-analytic choices on the study of HCQ use in COVID-19 outpatients with uncomplicated disease. Given that HCQ was repurposed as an antiviral drug, there was probably a higher chance a priori to observe a benefit in early disease than for hospitalized patients. The largest phase 3 trial reported numerically less hospitalizations in the HCQ group (RR, 0.77, 95% CI, 0.52–1.12, $p = 0.16$)[9]. The paper contained a meta-analysis as well, pooling the results with other randomized trials in the same population, and the result was (RR, 0.77, 95% CI, 0.57–1.04, $p = 0.09$). All but one small trial were double blind.

In Fig. 1b, we reproduce the meta-analysis using the same data as in ref. 9, testing both a standard approach (as done in the original paper) and the Hartung–Knapp one. The result is also a small scale factor ($q = 0.28$), so that the HKSJ model leads to a shorter confidence interval (95% CI, 0.62 –0.95, $p = 0.03$) and an association, but this time favouring HCQ use. Therefore, if one accepts the demonstration of harm in inpatients, then one should in principle also accept the demonstration of a benefit of HCQ as an early outpatient therapy.

Discussion

In this note, we have pointed out some conceptual difficulties associated with a crude application of the Hartung–Knapp adjustment in this dataset of HCQ COVID-19 trials. We stress that we do not dispute the included studies, nor even the possibility that HCQ may indeed have had a non-zero harmful effect on the fatality rate of hospitalized COVID-19 patients, especially at the highest doses. We also emphasize that the systematic review of Axfors et al. is a high-quality and very useful one as it included unpublished data from ongoing trials as well.

However, the demonstration of a harmful effect depends on a statistical method that was, as we have argued here, most likely not best suited for the dataset at hand. Therefore, the conclusions of

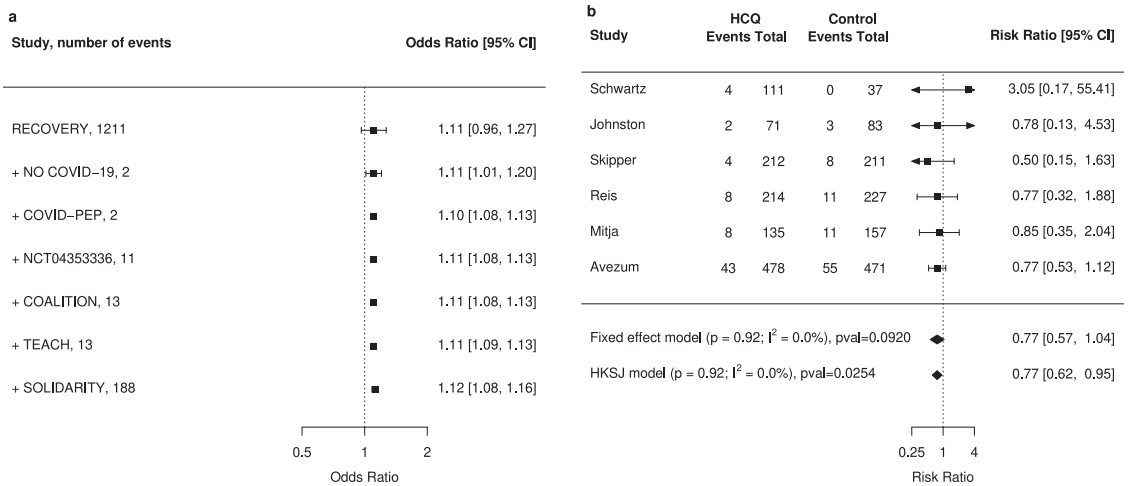

Fig. 1 | Forest plots for the meta-analyses. a Cumulative meta-analysis (in chronological order) for published studies, using the Hartung–Knapp method as in ref. 2. The total number of events added by each study is also highlighted. Studies given zero weight are not shown on the plot. **b** Meta-analysis of HCQ effect on COVID-19 hospitalization, using the same data as in ref. 9.

Table 1 | Calculated scale factors and two-sided *p*-values (for both untruncated and modified HKSJ) for various sets of studies from ref. 2

Studies	*q*	p(*q* < 1)	p(*q* = 1)
All studies	0.39	0.017	0.114
Published studies	0.05	<0.001	0.114
High-dose studies	0.16	0.045	0.215
Open-label studies	0.32	0.007	0.091

Axfors et al. should be taken with some caution when estimating potential deaths caused by HCQ use during the COVID-19 pandemic. A small Bayesian trial (REMAP-CAP) also suggested a high probability of harm associated with HCQ[13]. This result may have been confounded by the use of a non-concurrent control group and complex modelling. To estimate deaths caused by HCQ, one should ideally use the most recent and complete meta-analysis available, which however showed a weaker rather than stronger trend when more data were added[6].

To illustrate further our point, we also revisited the meta-analysis of HCQ effect in COVID-19 outpatients. We showed that the Hartung–Knapp adjustment also led to a shorter confidence interval and an association, but this time favouring HCQ. This exemplifies how dicey a misguided use of these meta-analytic models might be if the result is a premature claim of efficacy.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Data were extracted from the meta-analyses by Axfors and colleagues and Avezum and colleagues. Source data are provided with this paper.

Code availability

The R package 'metafor'[14], used to generate the forest plots, is available at https://www.metafor-project.org. Scripts to reproduce the figure are available at https://github.com/dpasquie/HCQMattersArising.

References

1. RECOVERY collaborative group. Hydroxychloroquine in hospitalized patients with Covid-19. *N. Engl. J. Med.* **384**, 881–882 (2021).
2. Axfors, C. et al. Mortality outcomes with hydroxychloroquine and chloroquine in COVID-19 from an international collaborative meta-analysis of randomized trials. *Nat. Commun.* **12**, 2349 (2021).
3. WHO Solidarity trial Consortium. Repurposed antiviral drugs for Covid-19—Interim WHO solidarity trial results. *N. Engl. J. Med.* **384**, 497–511 (2021).
4. Singh, B., Ryan, H., Kredo, T., Chaplin, M. & Fletcher, T. Chloroquine or hydroxychloroquine for prevention and treatment of COVID 19. *Cochrane Database Syst. Rev.* **2**, CD013587 (2021).
5. Di Castelnuovo, A. et al. Hydroxychloroquine and mortality in COVID-19 patients: a systematic review and a meta-analysis of observational studies and randomized controlled trials. *Pathog. Glob. Health* **115**, 456–466 (2021).
6. Pan, H. et al. Remdesivir and three other drugs for hospitalised patients with COVID-19: final results of the WHO Solidarity randomised trial and updated meta-analyses. *Lancet* **399**, 1941–1953 (2022).
7. Pradelle, A. et al. Deaths induced by compassionate use of hydroxychloroquine during the first COVID-19 wave: an estimate. *Biomed. Pharmacother.* **171**, 116055 (2024).
8. Wiksten, A., Rücker, G. & Schwarzer, G. Hartung-Knapp method is not always conservative compared with fixed-effect meta-analysis. *Stat. Med.* **35**, 2503–2515 (2016).
9. Avezum, À. et al. Hydroxychloroquine versus placebo in the treatment of non-hospitalised patients with COVID-19 (COPE–Coalition V): a double-blind, multicentre, randomised, controlled trial. *Lancet Reg. Health Am.* **11**, 100243 (2022).
10. Hartung, J. & Knapp, G. A refined method for the meta-analysis of controlled clinical trials with binary outcome. *Stat. Med.* **20**, 3875–3889 (2001).
11. Sidik, K. & Jonkman, J. N. A simple confidence interval for meta-analysis. *Stat. Med.* **21**, 3153–3159 (2002).
12. Röver, C., Knapp, G. & Friede, T. Hartung-Knapp-Sidik-Jonkman approach and its modification for random-effects meta-analysis with few studies. *BMC Med. Res. Methodol.* **15**, 99 (2015).
13. Florescu, S. et al. Long-term (180-day) outcomes in critically Ill patients with COVID-19 in the REMAP-CAP randomized clinical trial. *JAMA* **329**, 39–51 (2023).
14. Viechtbauer, W. Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* **36**, 1–48 (2010).

Acknowledgements

I thank Guido Knapp for discussion.

Author contributions

D.P. performed all the work.

Competing interests

The author declares no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-60478-x.

Correspondence and requests for materials should be addressed to Diego Pasquier.