

Bias in observational studies on the effectiveness of in hospital use of hydroxychloroquine in COVID-19

Mirjam Hempenius¹  | Sophie H. Bots¹ | Rolf H. H. Groenwold² |
 Anthonius de Boer¹ | Olaf H. Klungel^{1,3} | Helga Gardarsdottir^{1,4,5}

¹Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Utrecht, The Netherlands

²Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands

³Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands

⁴Department of Clinical Pharmacy, Division Laboratory and Pharmacy, University Medical Center Utrecht, Utrecht, The Netherlands

⁵Faculty of Pharmaceutical Sciences, University of Iceland, Reykjavik, Iceland

Correspondence

Helga Gardarsdottir, Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, PO Box 80 082, 3508 TB Utrecht, The Netherlands.
 Email: h.gardarsdottir@uu.nl

Funding information

Leids Universitair Medisch Centrum fellowship; Netherlands Organization for Scientific Research (ZonMW-Vidi), Grant/Award Number: 917.16.430

[Correction added on 16 June 2023, after first online publication: the author names have been spelled out for clarity.]

Abstract

Purpose: During the first waves of the coronavirus pandemic, evidence on potential effective treatments was urgently needed. Results from observational studies on the effectiveness of hydroxychloroquine (HCQ) were conflicting, potentially due to biases. We aimed to assess the quality of observational studies on HCQ and its relation to effect sizes.

Methods: PubMed was searched on 15 March 2021 for observational studies on the effectiveness of in-hospital use of HCQ in COVID-19 patients, published between 01/01/2020 and 01/03/2021 on. Study quality was assessed using the ROBINS-I tool. Association between study quality and study characteristics (journal ranking, publication date, and time between submission and publication) and differences between effects sizes found in observational studies compared to those found in RCTs, were assessed using Spearman's correlation.

Results: Eighteen of the 33 (55%) included observational studies were scored as critical risk of bias, eleven (33%) as serious risk and only four (12%) as moderate risk of bias. Biases were most often scored as critical in the domains related to selection of participants ($n = 13$, 39%) and bias due to confounding ($n = 8$, 24%). There were no significant associations found between the study quality and the characteristics nor between the study quality and the effect estimates.

Discussion: Overall, the quality of observational HCQ studies was heterogeneous. Synthesis of evidence of effectiveness of HCQ in COVID-19 should focus on RCTs and carefully consider the added value and quality of observational evidence.

KEYWORDS

bias, COVID-19, efficacy, hydroxychloroquine, methodology, pharmacoepidemiology, post market, regulatory

Key Points

- Conflicting results were reported in observational studies on hydroxychloroquine (HCQ) and COVID-19.
- This study assessed the quality of 33 articles on the effectiveness of in-hospital use of hydroxychloroquine in COVID-19 by using the ROBINS-I tool.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Pharmacoepidemiology and Drug Safety* published by John Wiley & Sons Ltd.

- More than half of the included observational studies (18/33, 55%) were at critical risk of bias, eleven (33%) were at serious risk and only four (12%) were at moderate risk of bias.
- The added value and quality of observational evidence regarding HCQ for COVID-19 patients should therefore be carefully considered.

Plain Language Summary

During the first waves of the coronavirus pandemic, evidence was urgently needed about potential effective treatments, such as hydroxychloroquine (HCQ). However, results from observational studies on the effectiveness of HCQ were conflicting, probably due to biases. In this study, we assessed the quality of observational studies that assessed effectiveness of HCQ treatment for COVID-19 patients with the ROBINS-I tool. This tool assesses the risk on seven domains of bias: confounding bias, selection bias due to misclassification of interventions and of outcomes, deviation from intended intervention, missing data, and completeness of reporting. For each of the domains, the risk of bias was judged independently by two reviewers to be “Low”, “Moderate”, “Serious”, or “Critical”. The overall score per study was determined by the highest score across all the domains. We found that more than half of the included observational studies (18/33, 55%) were at critical risk of bias, eleven (33%) were at serious risk and only four (12%) were at moderate risk of bias. In addition, we studied the association between quality of the studies and estimates of effect sizes for which effects sizes found in the observational studies were compared to those from randomized clinical trials (RCTs). We did not find an association between the effect sizes and the study design. Because of the heterogeneity in the quality of observational HCQ studies, we recommend that synthesis of evidence of effectiveness of HCQ in COVID-19 patients should focus on RCTs and that the added value and quality of observational evidence should be carefully considered.

1 | INTRODUCTION

During the first waves of the pandemic with severe acute respiratory syndrome coronavirus (SARS-CoV-2) and its associated 2019 coronavirus disease (COVID-19), effective treatments were urgently needed to reduce mortality, the severity of symptoms, and hospitalization rates. Doctors were forced to make choices regarding which treatments were likely to save the lives of critically ill patients, without sufficient evidence-based knowledge about effective treatments for this new disease.¹

Hydroxychloroquine (HCQ) was one of the drugs that caught the attention of researchers, clinicians, and the public early in the pandemic. Preclinical studies indicated HCQ as a potentially effective treatment for the symptoms of COVID-19 because of its *in vitro* antiviral effects.² Researchers conducted many observational studies while waiting for results from randomized controlled trials (RCTs). As HCQ was already used early during the pandemic in the treatment of COVID-19, there was an opportunity to perform observational studies using routinely collected patient data.³ Indeed, well-designed observational studies can be helpful in the generation of hypotheses about the potential effects of drugs; however, observational studies can also provide biased results when not properly designed and analyzed.⁴

The observational studies on the effectiveness of HCQ reported divergent results, from a five-fold reduction in mortality risk to an eight-fold increased risk of intensive care unit (ICU) admission,^{5,6} leading to a heated debate about the effectiveness of HCQ.^{6–10} Today,

RCTs have convincingly shown that HCQ has no benefit in the treatment of COVID-19 patients and may even be harmful.^{11,12}

Variation in the estimated effects of HCQ treatment and divergence from the RCT results may arise from differences between the included patient populations. Subgroup analyzes of a meta-analysis on the effectiveness of HCQ in COVID-19 patients showed however, that disease severity or age did not lead to different outcomes; in all these subgroups, no effect of HCQ on COVID-19 outcomes was found.¹³ The differences in estimated effect have therefore led to discussions on how the lack of a proper study design, short review times, and reviewers' possible lack of expertise might have led to the unjustified conclusions.^{14,15} For example, many studies suffered from immortal time bias, confounding bias, or bias due to inadequately accounting for competing risks.^{16,17}

The aim of the current study was to provide a comprehensive overview and assess the overall quality of observational studies on HCQ conducted during the first wave of the COVID-19 pandemic, and to relate this quality to the observed effect sizes reported in completed trials.

2 | METHODS

2.1 | Search strategy and inclusion criteria

On March 15, 2021, we searched PubMed for all observational studies on COVID-19 and the use of HCQ published between 01/01/2020 and 01/03/2021, using the search string “hydroxychloroquine AND

covid-19.” The included studies were peer-reviewed primary research articles, were published in English, and used an observational design to investigate in-hospital treatment with HCQ. Moreover, studies were included if they measured one of the following clinical outcomes: mortality, duration of hospitalization, need for mechanical ventilation, or time to clinical improvement. We included only studies in which HCQ (with or without azithromycin) as treatment for COVID-19 was compared to standard care. All studies focusing on the use of HCQ as prophylaxis for COVID-19 were excluded.

For each included observational study, we extracted the journal name, the impact factor of the scientific journal, and the journal ranking according to the InCites Journal Citation Reports (Q1, Q2, Q3, or Q4).¹⁸ We also extracted the date of first submission, acceptance, and first publication; the geographical region of the study population (Africa, Asia, North America, and Europe); the study design (cohort, case-control, or other); the number of included subjects; and the proportion of study subjects treated with HCQ.

2.2 | Quality of studies

We used the ROBINS-I tool to assess the risk of bias for each of the included studies. This tool is often used for the quality assessment of observational studies.¹⁹ It defines seven domains in which bias could occur: bias due to confounding, selection of participants, classification of interventions, deviations from intended interventions, missing data, measurement of outcomes, and selective reporting. For each of the domains, the risk of bias could be “Low”, “Moderate”, “Serious”, or “Critical”, based on the decision support table provided with the tool.¹⁹

In order to assess the risk of bias due to confounding, we prespecified a minimal set of potential confounders that play a role in the association between in-hospital treatment with HCQ and clinical outcomes in COVID-19 patients (see supplementary materials, section A). Since it is very likely that treatment decisions were related to factors that also influenced the outcome risk, the lowest score for this domain was “Moderate.”¹⁹ The quality assessment was performed in duplicate (MH and SB), and discrepancies were discussed. The overall score per study was determined by the highest score across the domains.

2.3 | Effect size

The effect estimates reported in the observational studies were compared to the effect estimates found in a meta-analysis of RCTs; the latter served as a reference. Since effect estimates might also differ between RCTs, we selected benchmark estimates from the meta-analysis by Siemieniuk et al. This meta-analysis, published in *the BMJ*, is a living review, with the last update (at the time of study) on the 6th of April 2021.¹³ The outcomes included in this meta-analysis were mortality, need for mechanical ventilation, duration of hospital stay, ventilator-free days, and time to clinical improvement. Clinically relevant effects of HCQ were not observed for any of the outcomes (Table 1). For each of the outcomes, we extracted the point estimates and corresponding confidence intervals.

TABLE 1 Estimates of the effect of hydroxychloroquine (HCQ) in COVID-19 patients from randomized controlled trials (RCTs).

Outcome	Estimate from meta-analysis/RCTs
Mortality	OR 1.09 (0.93–1.27) ¹³
Mechanical ventilation	OR 1.15 (0.92–1.46) ¹³
Duration of hospital stay	Mean difference 0.1 (–1.8–2.0) ¹³
Ventilator-free days	Mean difference –1.4 (–4.9–2.2) ¹³
Time to clinical improvement	Mean difference –0.9 (–2.9–2.1) ¹³
Composite outcome ^a	OR 1.13 (0.60–2.14) ^{b41}
Intensive care unit (ICU) admission	OR 1.13 (0.60–2.14) ^c
Survival	OR 0.92 (0.79–1.09) ^d

^aComposite outcome indicating disease aggravation, for example ICU admission, need for mechanical ventilation, or mortality.

^bThe meta-analysis by Siemieniuk did not provide an estimate for a composite outcome of disease aggravation, ICU admission, and death, or for ICU admission alone. The odds ratio of the composite outcome was found to be 1.13 (0.60–2.14) in the RCT by Self et al.

^cIn other studies, the occurrence of ICU admission was categorized as disease aggravation. We assume that the effect estimate of ICU admission is in line with the estimates for mortality, ventilation, and the composite outcome, and it was set at 1.13.

^dInverse of estimate for mortality.

We note that the effect estimates obtained from RCTs are also estimates and hence do not necessarily reflect the *true* effect of HCQ. In addition, this approach cannot distinguish between deviations between effect estimates from observational studies and RCTs that arise due to bias, effect modification by patient characteristics, or (random) sampling variability. Moreover, different relative measures were being estimated, such as odds ratios (ORs), hazard ratios (HRs) and relative risks (RR). While these estimates differ in the way they are calculated, for all estimates closer to 1, no effect of HCQ is assumed, while estimates further from the null effect indicate a positive or negative effect of HCQ. Therefore, we hypothesize that in observational studies where effect estimates deviate more from those found in this meta-analysis of RCTs, the potential for bias is larger.

For each observational study, we extracted the point estimates for the primary outcome. If a study included multiple primary outcomes, we included the effect estimate for mortality, if present. For all relative measures, we subsequently calculated the extent to which this effect deviated from the benchmark estimates. This deviation was calculated as $\text{abs}(\log(\text{HR}_{\text{obs}}) - \log(\text{HR}_{\text{RCT}}))$.

2.4 | Data analysis

Publication and study characteristics were described using descriptive statistics. The publication date was dichotomized as before or in June 2020, or after June 2020, which was the month in which the interim results of the RECOVERY trial were published and the FDA decided to revoke the emergency use authorization for HCQ.^{20,21} For each of the included studies, we described the score for each of the domains and the overall score. In addition, we summarized the number of

studies within each of the domains, that were considered to be at “Low”, “Moderate”, “Serious” or “Critical” risk of bias, or that were scored as having “No information.” The relation between publication details (journal ranking [defined as Q1, Q2, Q3, or Q4], publication date, and time between submission and publication, as proxy for the peer-review process) and the overall quality of the studies as well as the relation between the effect size and the overall quality were assessed using Spearman's correlation.

3 | RESULTS

Our search strategy yielded 2331 hits in PubMed, 79 of which were selected on the basis of title and abstract. Of those studies, 33 were

included in this review. The reasons for inclusion and exclusion are depicted in Figure 1. A list of all included studies is presented in the supplementary materials (Section B).

3.1 | Publication and study characteristics

The characteristics of the included studies are summarized in Table 2. Out of 33 studies, 14 (42%) were published in journals that were ranked in the first quartile (Q1), according to the InCites Journal Citation Reports. Four of those 14 were published in a journal with an impact factor >30. The first observational study was published on the 7th of May 2020, and the last on the 18th of February 2021. Seven studies (21%) were published before or in

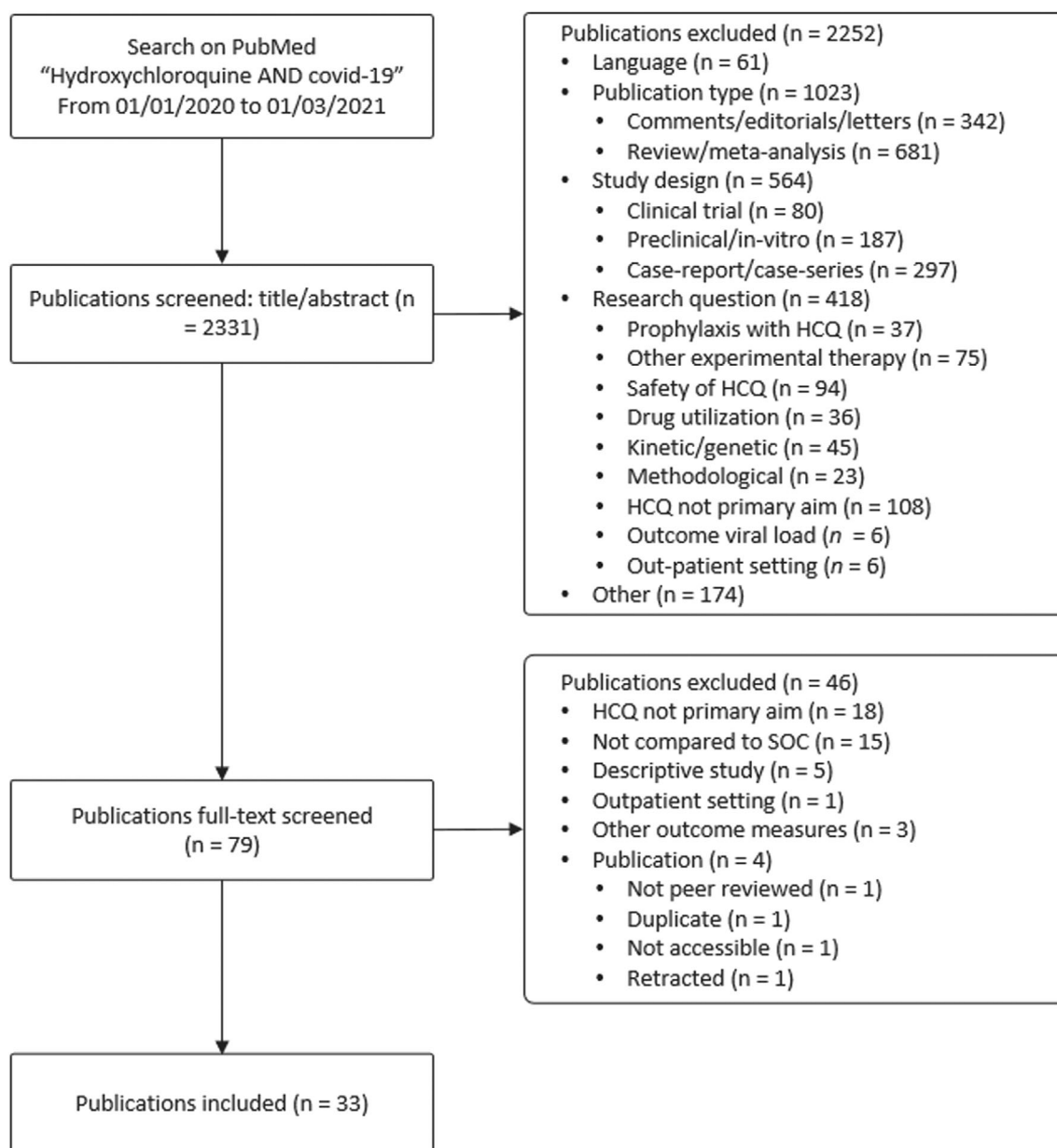


FIGURE 1 Flow chart of the search and screening process according to inclusion and exclusion criteria.

TABLE 2 Characteristics of the 33 observational studies of HCQ in COVID-19 that were included in the systematic review.

Journal ranking ^a	n (%)
Journals ranked as "Q1"	14 (42)
Journals ranked as "Q2"	12 (36)
Journals ranked as "Q3"	4 (12)
Journals ranked as "Q4"	0 (0)
No journal ranking available	3 (9)
Publication date^b	
Before or in June 2020	7 (21)
After June 2020	26 (79)
Time from submission to publication	
0–45 days	10 (30)
46–90 days	13 (39)
>90 days	4 (12)
Insufficient information for estimation	6 (18)
Geographical area	
Africa	1 (3)
Asia	4 (12)
North America	13 (39)
Europe	15 (45)
Study design	
Cohort	33 (100)
Case-control	0 (0)
Other design	0 (0)
Number of subjects included (median [IQR])	
1–250	8 (24)
251–1,000	10 (30)
1,001–2,500	9 (27)
>2,500	6 (18)
% of cohort treated with HCQ (median [IQR])	
<20%	1 (3)
20%–80%	25 (76)
>80%	6 (18)
Unclear	1 (3)

Abbreviations: HCQ, hydroxychloroquine; IQR, interquartile range.

^aAccording to the InCites Journal Citation Reports.¹⁸

^bIn June 2020, the interim results of the RECOVERY trial were published and the FDA decided to revoke the emergency use authorization for HCQ.^{20,21}

June 2020. The median time from submission to first publication was 64 days (interquartile range [IQR] 34.5–83.5 days). All studies were performed during the first wave of the COVID-19 pandemic (Supplementary materials, Section D). Studies were most often performed in Europe (45%) and the US (39%), and all studies used a cohort design. The median number of subjects included in the observational studies was 807 (IQR 307–1949), and the median proportion of the study population treated with HCQ was 54.6% (IQR 39.6%–74.8%).

3.2 | Quality of studies

The quality of the studies is depicted in Figure 2 and summarized per domain in Figure 3. A substantiation for the assigned scores is given in the Supplementary materials (Section C). Overall, 18 (55%) studies were considered to be at critical risk of bias, 11 (33%) were considered to be at serious risk of bias, and 4 (12%) to be at moderate risk of bias.

The domains that were most often scored as critical risk of bias were bias due to selection of participants ($n = 13$, 39%) and bias due to confounding ($n = 8$, 24%). Bias due to selection of participants was most often caused by the introduction of immortal time bias, which was present in 22 studies (67%), although in varying degree. Bias due to confounding was to some extent present in all studies due to the observational nature of the studies, but some studies had a greater risk of bias due to confounding than others. For example, studies were considered to be at critical risk of bias when there was no adjustment for confounders, if there were major differences in disease severity between the treatment groups for which no adjustment was made, or if the study adjusted for intermediates, such as the use of mechanical ventilation during hospital stay.

The domains that were most often scored as low risk of bias were bias due to deviation from intended intervention ($n = 31$, 94%), because any deviations observed were not beyond what would be expected in usual practice. Bias due to measurement of outcomes was also often scored as low ($n = 23$, 70%), because clinical outcomes like mortality are unlikely to suffer from measurement errors, although we did observe competing risk bias in two studies (6%).

The domain "bias due to missing data," was the domain which was most often scored as "No information"; 11 studies (33%) did not report how missing data were handled.

3.3 | Outcome measure and effect size

Twenty-one of 33 studies (64%) measured the effect of HCQ on mortality. Other outcome measures that were used were survival ($n = 2$, 6%); ICU admission ($n = 2$, 6%); hospital length of stay ($n = 2$, 6%); or a composite outcome of mortality, ventilation, and/or ICU admission ($n = 6$, 18%). Since the outcomes for mortality and ICU admission as well as the composite outcome had comparable benchmark estimates for the use of HCQ compared to standard care (OR 1.09, 1.13, and 1.13, respectively), these effect estimates are depicted in Figure 4. The effect estimates for survival were OR 0.96 (95% confidence interval [CI] 0.47 to 2.02) and OR 0.9 (0.4 to 2.1), and for hospital length of stay, they were -2.12 days (95% CI 0.47 to 4.50) and -5.41 days (95% CI -10.49 to 0.32) when comparing HCQ to standard care.

Almost all included studies (30 of 33) reported relative measures of the effectiveness of HCQ for which the deviation from the estimate of the meta-analysis of RCTs could be calculated on a log scale. Of the three remaining studies, two reported differences in length of stay, and one presented only p -values. The estimate in the study by Peters et al. was closest to the benchmark estimates (deviation 0.00). The authors found an HR of 1.09 (95% CI 0.81–1.47) for mortality,

Study	Risk of bias domains							Overall
	D1	D2	D3	D4	D5	D6	D7	
Abdulrahman A	⊗	⊗	+	+	?	+	-	⊗
Albani F	⊗	⊗	+	+	-	+	-	⊗
Almazrou SH	⊗	⊗	-	+	⊗	-	-	⊗
Annie FH	⊗	⊗	⊗	+	?	+	-	⊗
Arshad S	⊗	⊗	-	+	⊗	+	-	⊗
Ashinyo ME	⊗	⊗	-	+	⊗	+	-	⊗
Awad N	⊗	⊗	-	-	?	+	⊗	⊗
Ayerbe L	⊗	-	⊗	+	⊗	+	-	⊗
Catteau L	-	-	-	+	-	+	-	-
CORIST Collaboration.	⊗	⊗	-	+	-	+	-	⊗
Dubernet A	⊗	⊗	+	+	?	+	⊗	⊗
Geleris J	-	⊗	+	+	+	+	-	⊗
Grimaldi D	⊗	⊗	+	+	⊗	+	-	⊗
Huang HD	⊗	⊗	+	+	⊗	⊗	⊗	⊗
Ip A	⊗	⊗	⊗	+	-	-	-	⊗
Kalligeros M	-	?	⊗	+	?	+	⊗	⊗
Kim EJ	⊗	⊗	+	+	-	+	-	⊗
Lagier JC	⊗	⊗	⊗	+	?	⊗	⊗	⊗
Lammers AJJ	⊗	⊗	+	⊗	+	-	⊗	⊗
Lauriola M	⊗	⊗	+	+	+	+	-	⊗
Lecronier M	⊗	-	+	+	+	⊗	-	⊗
Magagnoli J	-	+	-	+	?	+	-	-
Mahevas M	-	-	-	+	+	+	-	-
Paccoud O	⊗	⊗	+	+	+	+	-	⊗
Peters EJ	-	-	+	+	+	+	-	-
Rivera DR	⊗	⊗	-	+	⊗	+	⊗	⊗
Rivera-Izquierdo M	⊗	⊗	-	+	⊗	-	⊗	⊗
Roomi S	⊗	⊗	⊗	+	?	+	⊗	⊗
Rosenberg ES	-	⊗	⊗	+	⊗	+	⊗	⊗
Sands K	⊗	⊗	+	+	⊗	-	⊗	⊗
Su Y	⊗	⊗	+	+	?	⊗	-	⊗
Vernaz N	⊗	⊗	+	+	?	⊗	⊗	⊗
Yu B	⊗	⊗	⊗	+	?	+	⊗	⊗

Domains:
 D1: Bias due to confounding.
 D2: Bias due to selection of participants.
 D3: Bias in classification of interventions.
 D4: Bias due to deviations from intended interventions.
 D5: Bias due to missing data.
 D6: Bias in measurement of outcomes.
 D7: Bias in selection of the reported result.

Judgement
 ⊗ Critical
 ⊗ Serious
 - Moderate
 + Low
 ? No information

FIGURE 2 Risk of bias for each of the domains per included study.⁴²

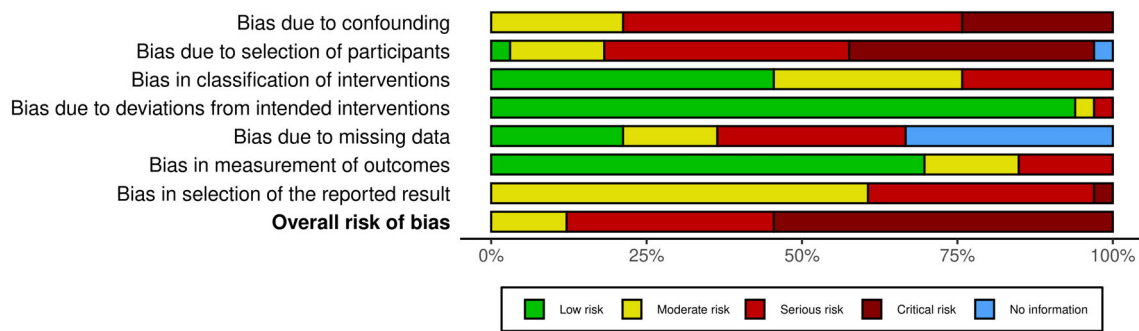


FIGURE 3 Summary of the risk of bias per domain.⁴²

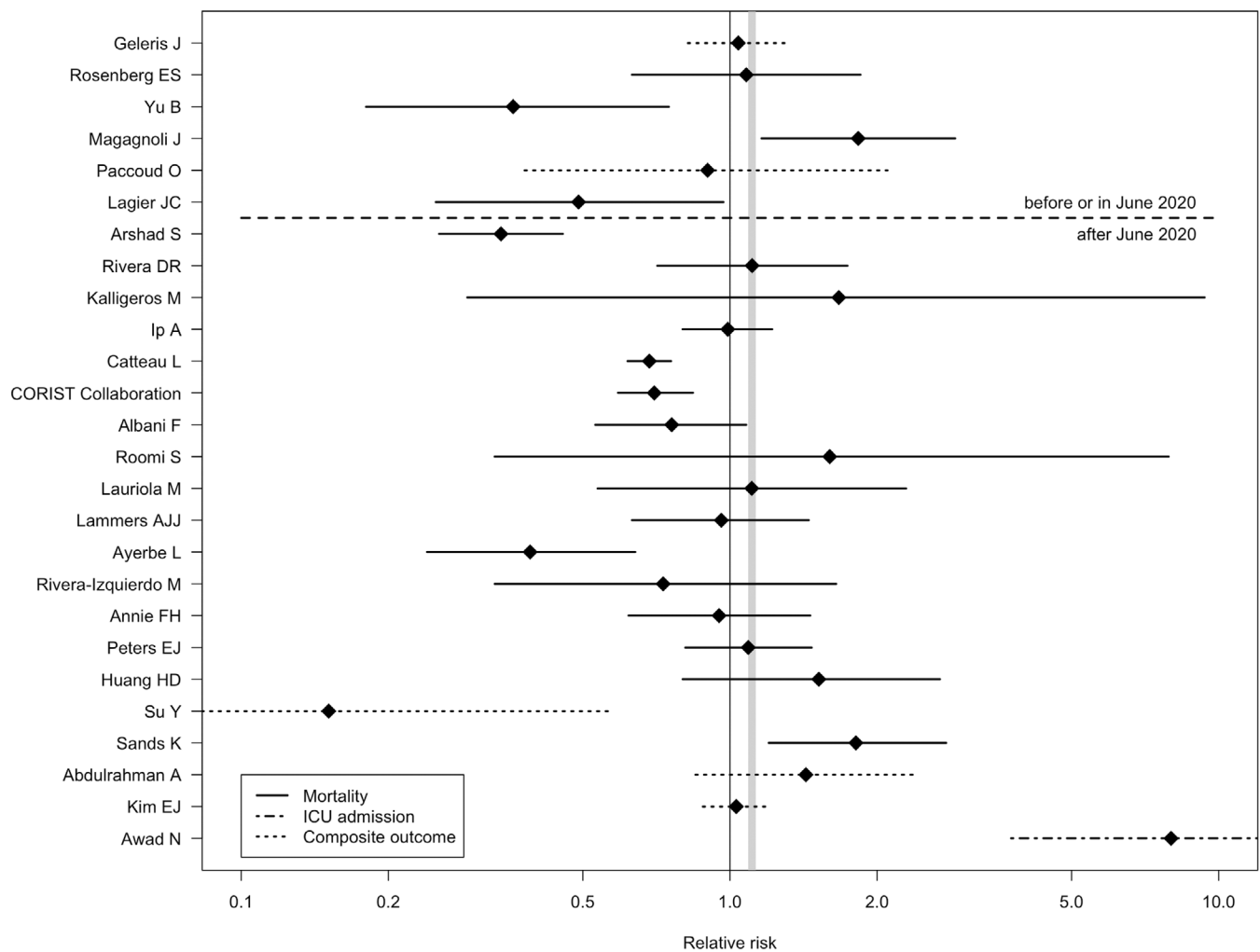


FIGURE 4 Effect estimates and 95% confidence intervals of the effect of HCQ on outcomes mortality, ICU admission, and a composite outcome of disease aggravation. Estimates are in chronological order. The dashed line indicates June 2020, in which both the interim results of the RECOVERY trial were published and the FDA decided to revoke the emergency use authorization for HCQ.^{20,21} Benchmark estimates for mortality, ICU admission, and the composite outcome are OR 1.09, 1.13, and 1.13, respectively. The grey area indicates estimates between 1.09 and 1.13. Two studies were excluded from this plot because no CIs were presented.

Note: The outcomes for mortality and ICU admission as well as the composite outcome had comparable benchmark estimates for the use of HCQ compared to standard care (OR 1.09, 1.13, and 1.13, respectively). The estimates for survival were directed in the opposite direction and the absolute estimates could not be depicted on this scale. Therefore, these effect estimates are not depicted in this figure.

which is equal to the odds ratio of the benchmark estimate.²² The study with the most deviating estimate (deviation 0.87 on the log scale) was by Su et al., where HR 0.15 (95% CI 0.040–0.575) was reported for disease aggravation.²³ Of all 31 studies that reported an effect estimates with confidence intervals, 24 studies (77%) reported CIs that had overlap with the CIs estimated in the RCTs and 7 studies (23%) reported CIs that had no overlap with the CIs from the RCTs.

3.4 | Relation between publication details and overall quality

The results of the Spearman's correlation test estimating the association between publication details and the overall quality are summarized in Table 3. Although studies scored to be at moderate risk of bias were more often published in journals ranked as "Q1" compared to studies scored to be at serious or critical risk of bias (75% of the studies scored as moderate risk were published in a journal ranked as "Q1", for studies scored serious or critical these percentages were 45.5% and 33.3%, respectively), this relation was not significant (Spearman's rho 0.27, *p*-value 0.13). There were also no significant relations found between the overall quality and the length of the reviewing process or the publication date.

3.5 | Association between effect size and overall quality

The association between the effect size and the overall quality is given in Table 3. There was no significant association between the

effect size and the study quality (Spearman's rho 0.20, *p*-value 0.31), although the deviation from the effect estimates found in RCTs was lower in studies scored at moderate risk of bias.

4 | DISCUSSION

4.1 | Summary of findings

In this review of observational studies of HCQ treatment for COVID-19, we observed that more than half of all included observational studies were scored as critical risk of bias (18/33, 55%), eleven (33%) at serious risk and only four (12%) at moderate risk of bias. No significant associations were observed between study quality and the estimated effect or between the study quality and the journal impact factor in which the study was published. In addition, we did not find an association between the study quality and the time between submission and publication date. The time between submission and publication date was used as a proxy for the peer-review process, although time to publication may be misclassified when a manuscript has been submitted to multiple journals before acceptance. Lastly, we did not find a relation between publication date (before or in June 2020 versus after June 2020) and study quality.

In line with what others found,^{16,17} we also observed that many studies suffered from immortal time bias, confounding bias, or bias due to inadequately accounting for competing risks. Immortal time bias was found in 67% of all included studies and competing risk bias in 6% of all included studies. Both immortal time bias and competing risk bias may impact the effect estimates,^{24,25} which the sensitivity

TABLE 3 Association between publication characteristics, effect size, and overall quality of the studies.

	Overall (<i>n</i> = 33)	Moderate (<i>n</i> = 4)	Serious (<i>n</i> = 11)	Critical (<i>n</i> = 18)	rho	<i>p</i> -value
Journal ranking (<i>n</i>, %)						
Q1	14 (42.4)	3 (75)	5 (45.5)	6 (33.3)	0.27	0.13
Q2	12 (36.4)	0 (0)	6 (54.5)	6 (33.3)		
Q3	4 (12.1)	0 (0)	0 (0)	4 (22.2)		
No journal ranking available	3 (9.1)	1 (25)	0 (0)	2 (11.1)		
Time from submission to publication (<i>n</i>, %)						
0–45 days	10 (30.3)	2 (50)	2 (18.2)	6 (33.3)	–0.04	0.85
46–90 days	13 (39.4)	0 (0)	6 (54.5)	7 (38.9)		
>90 days	4 (12.1)	1 (25)	1 (9.1)	2 (11.1)		
Insufficient information for estimation	6 (18.2)	1 (25)	2 (18.2)	3 (16.7)		
Publication date^b (<i>n</i>, %)						
Before or in June 2020	7 (21.2)	2 (50)	3 (27.3)	2 (11.1)	0.30	0.08
After June 2020	26 (78.8)	2 (50)	8 (72.7)	16 (88.9)		
Effect size (deviation from RCTs on the log scale)^a (mean, SD)	0.20 (0.23)	0.11 (0.10)	0.16 (0.16)	0.26 (0.28)	0.20	0.31

Note: Results of the univariate Spearman correlation analysis of the association between publication characteristics and effect size and the study quality.

^aCalculated as $\text{abs}(\log(\text{HR}_{\text{obs}}) - \log(\text{HR}_{\text{RCT}}))$.

^bIn June 2020, both the interim results of the RECOVERY trial were published and the FDA decided to revoke the emergency use authorization for HCQ.^{20,21}

analyzes in some of the included studies also suggested. Some studies assessed, for example, the impact of immortal time bias in sensitivity analyses and found large differences in the effect estimates, with estimates changing, for instance, from 1.08 to 1.46 (deviation 0.13 on the log scale)²⁶ or from 0.68 to 0.82 (deviation 0.08 on a log scale).²⁷ Interestingly, these differences in effect estimates were neither presented nor discussed in the main text, but only presented in the supplementary materials. In contrast, we found that different methods to handle missing data or different approaches for confounder adjustment (e.g., multivariable analysis and propensity score adjustment) had a more limited impact on the effect estimates in this selection of studies; the largest deviation from the primary effect estimate was 0.07 on the log scale.²⁷⁻²⁹

Furthermore, in one third of all included studies, insufficient information was reported in the article to fully comprehend all methodological choices. We observed this most often in the assessment of bias due to missing data. This poor reporting is not specific to COVID-19 research and has also been observed for pharmacoepidemiologic studies in general.^{30,31} Understandably, due to word limits, authors are unable to elaborate on all methodological decisions in their manuscript. However, in situations where methods deviate from generally accepted methods, substantiation of choices that were made is needed for correct interpretation of the results of a study as well as for assessment of its validity.

4.2 | Strengths and limitations

One of the strengths of our study were the systematic and duplicate assessment analysis of all biases defined in the ROBINS-I tool and our transparent substantiation of the assigned scores in the Supplementary materials. Another strength was the use of strict inclusion criteria to define our study sample. As a result, a relatively homogeneous set of included studies was obtained, which enabled us to compare effect estimates and relate these to the study quality.

One of the limitations of our study were that our search strategy was limited to PubMed (Medline), and we did no search Embase. However, the aim of our review was not to give a comprehensive overview of all studies of HCQ and COVID-19, but rather to give an impression of the quality of such studies. Another limitation was that the quality assessment was performed on the basis of information that was reported in the publication. This is actually an indirect way of assessing the risk of bias, as the extent to which a correct assessment is possible depends on the quality of reporting. A third limitation was the fact that we used the effect estimates from the meta-analysis of RCTs as benchmark, while there can be more reasons besides the presence of bias that could explain the deviations between the estimates from the observational studies and the RCTs. Yet, we hypothesize that in observational studies where effect estimates deviate more from those found in RCTs, the potential for bias is larger.

4.3 | Implications and recommendations

During the COVID-19 pandemic, with a high need for evidence-based therapy decisions, the results of observational studies formed the

basis for clinical guidelines when RCT results were not yet available. Although more than half of all included observational studies were scored as critical risk of bias, the promising results of some early observational studies did lead to strong recommendations to treat with HCQ in some countries or healthcare organizations, instead of waiting for the results of RCTs or high-quality observational studies.^{6,8-10} These strong recommendations favoring HCQ treatment made it difficult to make a valid comparison of the outcome risk between treated and untreated patients, as there might be underlying reasons why subjects are not treated, for example contra-indications for the use of HCQ.^{6,9,10} Since subjects without contra-indications are likely to be healthier than those with contra-indications, the results of these studies are likely to confirm the positive results of previous studies that led to these strong recommendations, due to confounding by (contra-) indication. In circumstances with such strong beliefs in the effectiveness of a potential treatment, random allocation of treatments may facilitate a true assessment of the effect of potential drugs. Therefore, the appropriateness of an observational design should always be assessed, taking into account the potential risks and benefits of the medicine of interest, and the feasibility of a valid non-randomized comparison, which may also lead to the conclusion that a randomized trial is needed to elucidate the uncertainties about the effectiveness and safety.

Second, proper design and analysis of observational studies is of utmost importance to obtain valid effect estimates, especially since the publication of results from invalid or flawed observational studies will likely create confusion, as has been seen in the debate on the effectiveness of HCQ, or may cause a shift in clinical equipoise that poses ethical issues for subsequent clinical trials. On the one hand, to improve the quality, researchers should work in multidisciplinary teams that include clinicians, methodologists, and database experts, among others, to combine their knowledge and research skills. On the other hand, guidelines for designing observational studies should be used. There are currently a number of guidelines to support the design of a pharmacoepidemiologic study and to avoid potential biases.³²⁻³⁶ In addition to these general pharmacoepidemiologic guidelines, recommendations specific to pharmacoepidemiologic COVID-19 research were published at the beginning of the pandemic (May 5th, 2020).⁴ One can also design an observational study as if it was an RCT.³⁷ This “emulated trial design” framework may be helpful in avoiding biases that can otherwise easily occur in observational studies.³⁸

Third, journals and their editors also have a responsibility to guard the quality of the studies that are published, both in the process of peer review and in their final decision regarding whether or not to publish the study results. To guarantee sufficient quality, journals should encourage or even oblige the use of checklists by authors and reviewers, such as ROBINS-I or RECORD-PE.^{19,39} Moreover, reviewers must have sufficient expertise to critically review the quality of submitted studies against the presence of potential biases. As an aid, reviewer guidelines have recently been published on how to assess and interpret real-world evidence from observational studies.⁴⁰

4.4 | Conclusions

To conclude, the overall quality of observational studies on the effectiveness of in-hospital use of HCQ for the treatment of COVID-19 symptoms was heterogeneous. The urgency of situations such as a pandemic should never be an argument for conducting and publishing observational studies that are of low quality, more so because in such situations, results quickly find their way into daily practice, and the results of biased studies can have potentially harmful consequences for patients. However, the results of this study should also not be seen as a plea against observational research; instead, particularly in times of high unmet medical needs, observational studies may provide timely evidence that could be valuable, provided that the study that provides the evidence is of sufficiently high quality.

FUNDING INFORMATION

RHHG was funded by the Netherlands Organization for Scientific Research (ZonMW-Vidi project 917.16.430) and an LUMC fellowship.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

ETHICS STATEMENT

The authors state that no ethical approval was needed.

ORCID

Mirjam Hempenius  <https://orcid.org/0000-0003-0223-4815>

REFERENCES

- Rubin EJ, Harrington DP, Hogan JW, Gatsonis C, Baden LR, Hamel MB. The urgency of care during the COVID-19 pandemic – learning as we go. *N Engl J Med*. 2020;382(25):2461-2462. doi:10.1056/nejme2015903
- Fantini J, Di Scala C, Chahinian H, Yahi N. Structural and molecular modelling studies reveal a new mechanism of action of chloroquine and hydroxychloroquine against SARS-CoV-2 infection. *Int J Antimicrob Agents*. 2020;55(5):105960.
- Gautret P, Lagier J-C, Parola P, et al. Clinical and microbiological effect of a combination of hydroxychloroquine and azithromycin in 80 COVID-19 patients with at least a six-day follow up: a pilot observational study. *Travel Med Infect Dis*. 2020;34:101663. doi:10.1016/j.tmaid.2020.101663
- Pottegård A, Kurz X, Moore N, Christiansen CF, Klungel O. Considerations for pharmacoepidemiological analyses in the SARS-CoV-2 pandemic. *Pharmacoepidemiol Drug Saf*. 2020;29(8):825-831. doi:10.1002/pds.5029
- Awad N, Schiller DS, Fulman M, Chak A. Impact of hydroxychloroquine on disease progression and ICU admissions in patients with SARS-CoV-2 infection. *Am J Heal Pharm*. 2021;78:689-696. doi:10.1093/ajhp/zxab056
- Lagier J-C, Million M, Gautret P, et al. Outcomes of 3,737 COVID-19 patients treated with hydroxychloroquine/azithromycin and other regimens in Marseille, France: a retrospective analysis. *Travel Med Infect Dis*. 2020;36:101791. doi:10.1016/j.tmaid.2020.101791
- de Lucena Alves CP, Segundo JD, da Costa GG, et al. How a few poorly designed COVID-19 studies may have contributed to misinformation in Brazil: the case for evidence-based communication of science. *BMJ Open Sci*. 2021;5(1):e100202. doi:10.1136/BMJOS-2021-100202
- Arshad S, Kilgore P, Chaudhry ZS, et al. Treatment with hydroxychloroquine, azithromycin, and combination in patients hospitalized with COVID-19. *Int J Infect Dis*. 2020;97:396-403. doi:10.1016/j.ijid.2020.06.099
- Lauriola M, Pani A, Ippoliti G, et al. Effect of combination therapy of Hydroxychloroquine and azithromycin on mortality in patients with COVID-19. *Clin Transl Sci*. 2020;13(6):1071-1076. doi:10.1111/cts.12860
- Lee TC, MacKenzie LJ, McDonald EG, Tong SYC. An observational cohort study of hydroxychloroquine and azithromycin for COVID-19: (Can't Get No) satisfaction. *Int J Infect Dis*. 2020;98:216-217. doi:10.1016/j.ijid.2020.06.095
- The RECOVERY Collaborative Group. Effect of Hydroxychloroquine in hospitalized patients with COVID-19. *N Engl J Med*. 2020;383(21):2030-2040. doi:10.1056/NEJMoa2022926
- Pan H, Peto R, Karim QA, et al. Repurposed antiviral drugs for COVID-19 – interim WHO solidarity trial results. *N Engl J Med*. 2021;384(6):497-511. doi:10.1056/nejmoa2023184
- Siemieniuk RAC, Bartoszko JJ, Ge L, et al. Drug treatments for COVID-19: living systematic review and network meta-analysis. *BMJ*. 2020;370:370. doi:10.1136/bmj.m2980
- 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(12):819-821. doi:10.7326/M20-1223
- Fiolet T, Guihur A, Rebeaud ME, Mulot M, Peiffer-Smadja N, Mahamat-Saleh Y. Effect of hydroxychloroquine with or without azithromycin on the mortality of coronavirus disease 2019 (COVID-19) patients: a systematic review and meta-analysis. *Clin Microbiol Infect*. 2021;27(1):19-27. doi:10.1016/j.cmi.2020.08.022
- Renoux C, Azoulay L, Suissa S. Biases in evaluating the safety and effectiveness of drugs for COVID-19: designing real-world evidence studies. *Am J Epidemiol*. 2021;190(8):1452-1456. doi:10.1093/aje/kwab028
- Martinuka O, von Cube M, Wolkewitz M. Methodological evaluation of bias in observational coronavirus disease 2019 studies on drug effectiveness. *Clin Microbiol Infect*. 2021;27(7):949-957. doi:10.1016/j.cmi.2021.03.003
- Clarivate. InCites Journal Citation Reports. 2021 Accessed August 4, 2021. <https://jcr.clarivate.com>
- Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*. 2016;355:355. doi:10.1136/bmj.i4919
- RECOVERY trial to stop enrolling patients to “ineffective” hydroxychloroquine arm – The Pharmaceutical Journal. Accessed September 28, 2021. <https://pharmaceutical-journal.com/article/news/recovery-trial-to-stop-enrolling-patients-to-ineffective-hydroxychloroquine-arm>
- Coronavirus (COVID-19) Update: FDA Revokes Emergency Use Authorization for Chloroquine and Hydroxychloroquine | FDA. Accessed October 6, 2021. <https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-revokes-emergency-use-authorization-chloroquine-and>
- Peters EJ, Collard D, Van Assen S, et al. Outcomes of persons with coronavirus disease 2019 in hospitals with and without standard treatment with (hydroxy)chloroquine. *Clin Microbiol Infect Off Publ Eur Soc Clin Microbiol Infect Dis*. 2021;27(2):264-268. doi:10.1016/j.cmi.2020.10.004
- Su Y, Ling Y, Ma Y, et al. Efficacy of early hydroxychloroquine treatment in preventing COVID-19 pneumonia aggravation, the experience from Shanghai, China. *Biosci Trends*. 2021;14(6):408-414. doi:10.5582/bst.2020.03340
- Suissa S, Dell'Aniello S. Time-related biases in pharmacoepidemiology. *Pharmacoepidemiol Drug Saf*. 2020;29(9):1101-1110. Accessed June 17, 2021. <https://onlinelibrary.wiley.com/doi/full/10.1002/pds.5083>

25. Wolkewitz M, Schumacher M. Survival biases lead to flawed conclusions in observational treatment studies of influenza patients. *J Clin Epidemiol.* 2017;84:121-129. doi:[10.1016/J.JCLINEPI.2017.01.008](https://doi.org/10.1016/J.JCLINEPI.2017.01.008)
26. Rosenberg ES, Dufort EM, Udo T, et al. Association of Treatment with Hydroxychloroquine or azithromycin with In-hospital mortality in patients with COVID-19 in New York state. *JAMA.* 2020;323(24):2493-2502. doi:[10.1001/jama.2020.8630](https://doi.org/10.1001/jama.2020.8630)
27. Catteau L, Dauby N, Montourcy M, et al. Low-dose hydroxychloroquine therapy and mortality in hospitalised patients with COVID-19: a nationwide observational study of 8075 participants. *Int J Antimicrob Agents.* 2020;56(4):106144. doi:[10.1016/j.ijantimicag.2020.106144](https://doi.org/10.1016/j.ijantimicag.2020.106144)
28. Albani F, Fusina F, Giovannini A, et al. Impact of azithromycin and/or Hydroxychloroquine on hospital mortality in COVID-19. *J Clin Med.* 2020;9(9):2800. doi:[10.3390/jcm9092800](https://doi.org/10.3390/jcm9092800)
29. Collaboration COVID-19 RISK and Treatments (CORIST). Use of hydroxychloroquine in hospitalised COVID-19 patients is associated with reduced mortality: findings from the observational multicentre Italian CORIST study. *Eur J Intern Med.* 2020;82:38-47. doi:[10.1016/j.ejim.2020.08.019](https://doi.org/10.1016/j.ejim.2020.08.019)
30. Hempenius M, Luijken K, de Boer A, Klungel O, Groenwold R, Gardarsdottir H. Quality of reporting of drug exposure in pharmacoepidemiological studies. *Pharmacoepidemiol Drug Saf.* 2020;29(9):1141-1150. doi:[10.1002/pds.5020](https://doi.org/10.1002/pds.5020)
31. Hunt NB, Gardarsdottir H, Bazelier MT, Klungel OH, Pajouheshnia R. A systematic review of how missing data are handled and reported in multi-database pharmacoepidemiologic studies. *Pharmacoepidemiol Drug Saf.* 2021;30(7):819-826. doi:[10.1002/pds.5245](https://doi.org/10.1002/pds.5245)
32. ISPE. Guidelines for good pharmacoepidemiology practices (GPP). *Pharmacoepidemiol Drug Saf.* 2008;17(2):200-208. doi:[10.1002/pds.1471](https://doi.org/10.1002/pds.1471)
33. European Medicines Agency. The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) Guide on Methodological Standards in Pharmacoepidemiology, 9th Revision. Accessed August 4, 2021. http://www.encepp.eu/standards_and_guidance
34. U.S. Department of Health and Human Services. Administration food and drug, research (CDER) Center for Drug Evaluation and, research (CBER) Center for Biologics Evaluation and Guidance for industry and FDA staff best practices for conducting and reporting pharmacoepidemiologic safety studies using electronic healthcare data. *Drug Saf.* 2013; Accessed June 18, 2018. <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm>
35. Velentgas P, Dreyer NA, Nourjah P, Smith SR, Torchia MM. Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide. 2013 Accessed May 3, 2019. <https://www.ncbi.nlm.nih.gov/books/NBK126190/>
36. European Medicines Agency. Guideline on Good Pharmacovigilance Practices (GVP) – Module VIII – Post-Authorisation Safety Studies (Rev 3). 2017 Accessed October 31, 2019. www.ema.europa.eu
37. Hernán MA, Robin J. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol.* 2016;183(8):758-764. doi:[10.1093/AJE/KWV254](https://doi.org/10.1093/AJE/KWV254)
38. Hernán MA, Sauer BC, Hernández-Díaz S, Platt R, Shrier I. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. *J Clin Epidemiol.* 2016;79:70-75. Accessed June 17, 2021. <https://pubmed.ncbi.nlm.nih.gov/27237061/>
39. Langan SM, Schmidt SA, Wing K, et al. The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). *BMJ.* 2018;363:k3532. doi:[10.1136/bmj.k3532](https://doi.org/10.1136/bmj.k3532)
40. Wang SV, Schneeweiss S. Assessing and interpreting real-world evidence studies: introductory points for new reviewers. *Clin Pharmacol Ther.* 2022;111(1):145-149. doi:[10.1002/CPT.2398](https://doi.org/10.1002/CPT.2398)
41. Self WH, Semler MW, Leither LM, et al. Effect of hydroxychloroquine on clinical status at 14 days in hospitalized patients with COVID-19: a randomized clinical trial. *JAMA.* 2020;324(21):2165-2176. doi:[10.1001/jama.2020.22240](https://doi.org/10.1001/jama.2020.22240)
42. McGuinness LA, Higgins JPT. Risk-of-bias VISualization (robvis): an R package and shiny web app for visualizing risk-of-bias assessments. *Res Synth Methods.* 2020;12:55-61. doi:[10.1002/jrsm.1411](https://doi.org/10.1002/jrsm.1411)

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Hempenius M, Bots SH, Groenwold RHH, de Boer A, Klungel OH, Gardarsdottir H. Bias in observational studies on the effectiveness of in hospital use of hydroxychloroquine in COVID-19. *Pharmacoepidemiol Drug Saf.* 2023;32(9):1001-1011. doi:[10.1002/pds.5632](https://doi.org/10.1002/pds.5632)