Rapport interne sur la qualité méthodologique de l'étude « Efficacy of hydroxychloroquine in patients with COVID-19: results of a randomized clinical trial»

Antoine Vanier, MD (Santé Publique), PhD (Biostatistique)

Méthodologiste, Chercheur en Biostatistique

Inserm UMR U1246 Sphere, Université de Nantes

02 Avril 2020

Ce document est disponible sous licence Creative Commons BY NC ND 4.0.

I/ CONTEXTE D'ÉVALUATION

En Décembre 2019 est constaté l'émergence d'une nouvelle maladie (COVID-19) due à un nouveau coronavirus (SARS-CoV-2). Le point de départ étant Wuhan en Chine. On assiste ensuite à une diffusion en Chine puis internationale. L'OMS a qualifié l'épidémie de COVID-19 en pandémie le 12 Mars 2020. La France est confrontée à la vague épidémique depuis fin Février - début Mars 2019.

Au 1 Avril 2020, Santé Publique France recensait 853 200 cas de COVID-19 confirmés dans le monde, et 41 887 décès. En France, on dénombre 56 989 cas confirmés pour 4032 décès.

Dans ce contexte, la nécessité de disposer d'un traitement curatif est évidente.

J'analyserai ici les résultats d'une étude publiée par Chen et al. sous forme de pré-print. Ces résultats ont été mis à disposition sur medRxiv le 22 Mars 2020. La version analysée est un pre-print mise à disposition par les auteurs sous licence CC BY-NC-ND 4.0 internationale antérieure à tout processus de peer-review externe. Je ne sais pas si cette publication fait actuellement l'objet d'une soumission pour publication dans une revue à comité

II/ RATIONNEL AVANCÉ PAR L'ÉTUDE

L'essai ici considéré se base concernant le choix de l'hydroxychloroquine comme molécule à tester sur les arguments suivants. Premièrement, une étude publiée en 2019 montrant une activité inhibitrice *in-vitro* de la chloroquine sur la réplication virale du SARS-CoV-2. Puis, l'existence de recommandations d'experts chinois ou français quant à l'utilisation de la chloroquine dans le traitement du Covid-19. Enfin, les auteurs mentionnent que 80 patients du Renmin Hospital University of Wuhan (affiliation principale du premier et dernier auteur de l'étude) traités par hydroxychloroquine au long cours pour lupus érythémateux disséminé n'ont pas eu d'infection confirmée au SARS-CoV-2, et qu'aucun des 178 patients diagnostiqués Covid-19 au sein de cet hôpital ne recevaient d'hydroxychloroquine au long cours.

Les auteurs du présent essai justifient leur préférence envers l'hydroxychloroquine et non la chhloroquine sur la tolérance meilleure.

III/ ÉLÉMENTS GÉNÉRAUX DE MÉTHODOLOGIE

Promotion : ?

Centres investigateurs : Renmin Hospital University of Wuhan, Chine. Essai enregistré sur le Chinese Clinical Trial Registry (identifiant : ChiCTR2000029559).

Objectif principal de l'étude : Première évaluation précoce de l'efficacité de l'hydroxychloroquine chez des patients infectés par le virus SARS-CoV-2 sur le « Time To Clinical Revovery » (TTCV, délai de guérison clinique).

Type de l'étude : Essai de phase 2, unicentrique ou multicentrique ?, contrôlé, randomisée, supposément en double aveugle, en bras parallèles versus absence d'hydroxychloroquine (traitement symptomatique usuel, pas de placebo d'hydroxychloroquine).

Date et durée de l'étude : Inclusion du 4 au 28 Février 2020. Durée de 5 jours pour les patients inclus.

Critères d'inclusion : Patients hospitalisés avec diagnostic de COVID-19 de 18 ans et plus avec confirmation par PCR d'un portage nasopharyngé de SARS-CoV-2 à l'admission, des signes évocateurs de pneumonies sur image tomodensitométrique, un ratio SaO2/SPO2 supérieur à 93% et un ratio PaO2/FiO2 supérieur à 300 mmHg an air ambiant.

Critères de non-inclusion :

- Pathologie sévère ou critique ne permettant pas un suivi efficace ou étant une perte de chance pour le patient
- Rétinopathies et autres pathologies rétiniennes
- Insuffisance hépatique sévère (Child-Pugh C ou plus ou ASAT/ALAT supérieures à deux fois la limite supérieure de la normale)
- Insuffisance rénale sévère (débit de filtration glomérulaire estimé inférieur ou égal à 30 mL/min/1.73m^2)
- Femmes enceintes (confirmées par test de grossesse positif ou déclaratif)
- Femmes en situation d'allaitement maternel (déclaratif)
- Possibilité de transfert dans un autre hôpital sous 72 heures après admission

Schéma de l'étude : En bras parallèles. Visite baseline pour le recrutement, la collecte des données baseline et l'initiation du traitement à J0 de l'admission hospitalière. Visite d'évaluation finale 5 jours après l'admission. Évaluation finale par les médecins investigateurs qui suivent les patients. Les traitements symptomatiques et une antibiothérapie en prévention d'une sur-infection bactérienne étaient laissés à l'appréciation des investigateurs sur jugement clinique.

Traitements étudiés : Hydroxychloroquine orale, Deux prises de 200 mg orale par jour + traitement usuel *versus* traitement usuel (pas de placebo d'hydroxychloroquine).

Critère de jugement principal : Différence moyenne de TTCR à J5 post-inclusion : défini comme le délai de retour à une température corporelle normale (en dessous de 36,6°C, en dessous de 37;2 °C sous les aisselles ou la langue, en dessous de 37,8°C en intra-rectale ou intra auriculaire) pendant au moins 72 heures. Ou le délai de retour à une toux faible à absente selon le patient depuis plus de 72 heures. Les deux critères ont été analysés séparément. La définition de ce qu'est un succès n'est pas explicitée (les deux critères étaient ils considérés comme de multiples critères de jugement principaux ou comme co-critères de jugement principaux ?)

Critères de jugement secondaires : Température corporelle moyenne journalière et niveau de toux journalier moyen (moyenne sur trois évaluations par jour). Évolution du résultat tomodensitométrique entre J0 et J6 : classé selon 4 niveaux (exacerbé, inchangé, amélioration de « moins de 50% de la pneumonie », amélioration de « plus de 50% de la pneumonie »). Aucune évaluation de tolérance correcte.

Taille de l'échantillon : Aucune hypothèse de travail et aucune justification du nombre de sujets nécessaire.

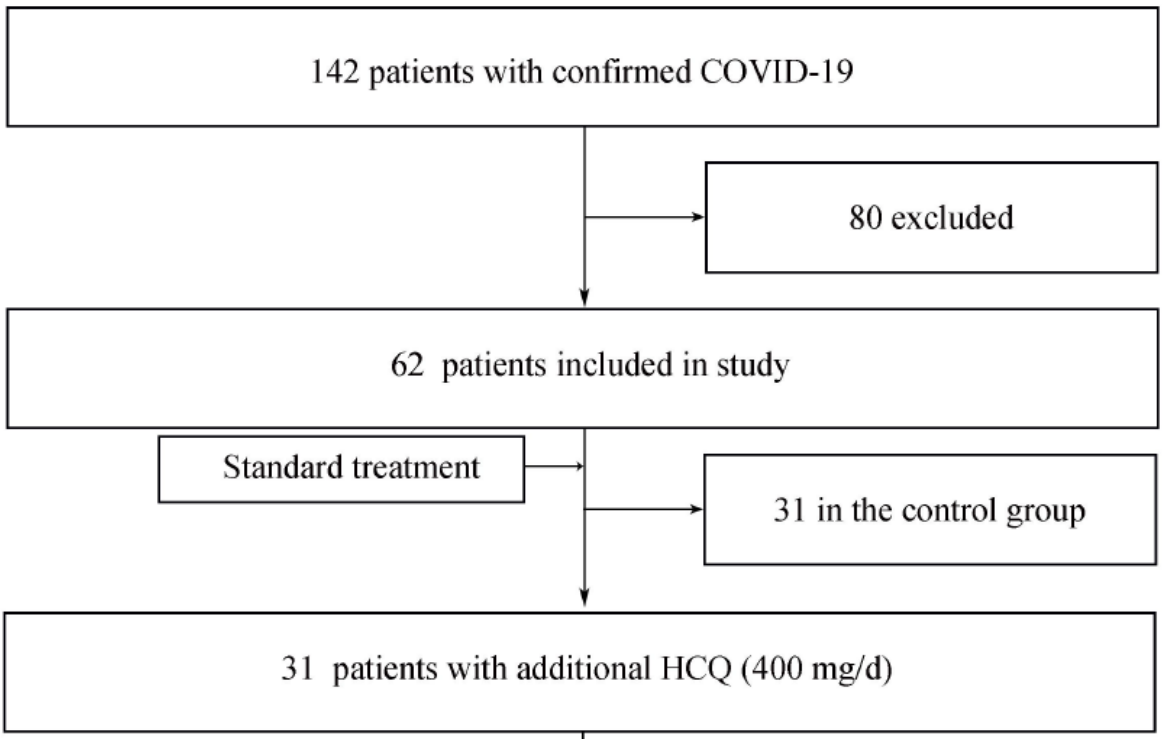
Méthode d'analyse des résultats : L'existence d'une différence statistiquement significative sur le critère de jugement principal entre les deux groupes de comparaison par test de student (de même pour les autres variables quantitatives). Concernant les variables qualitatives, la comparaison a été réalisée par tests de χ^2.

Population d'analyse : Non définie. A priori, « intention de traiter » ?

IV/ SYNTHÈSE DES PRINCIPAUX RÉSULTATS

Pendant la période d'inclusion, 142 patients présentaient un diagnostic confirmé de COVID-19, 62 ont été inclus dans l'étude (31 par bras).

Voici le « flow-chart » de l'étude.



Voici les caractéristiques à l'inclusion des deux groupes tels que présentés dans le pré-print (qui contient aussi l'analyse du ou des critères principaux…).

Characteristics	All	Control	HCQ	P value
Cases, n	62	31	31	
Age, mean (SD)	44.7 (15.3)	45.2 (14.7)	44.1 (16.1)	0.8809
Sex, n (%)				0.7991
Male	29 (46.8%)	15 (48.3%)	14 (45.2%)	
Female	33 (53.2%)	16 (51.7%)	17 (54.9%)	
Fever, day (SD)[a]	2.6 (1.0)	3.2 (1.3)	2.2 (0.4)	0.0008
Cough, day (SD)[b]	2.4 (1.1)	3.1 (1.5)	2.0 (0.2)	0.0016
Progressed to severe illness	4 (6.5 %)	4 (12.9 %)	0	
Adverse effects	2 (3.2 %)	0	2 (6.4 %)	

Table 1: Characteristics of patients in this trial.

[a]22 patients in the HCQ treatment group,17 patients in the control group with a fever one day before the intervention. [b]22 patients in the HCQ treatment group,15 patients in the control group with a cough one day before the intervention. Abbreviations: SD, standard deviation; HCQ, hydroxychloroquine; CT, computed tomography.

Le délai moyen de retour à température corporelle normale a été estimé significativement plus court dans le groupe traité par rapport au groupe témoins (cf tableau au-dessus). De même pour le délai de retour à une toux faible à inexistante sur base déclarative (cf tableau au-dessus).

Concernant l'évolution entre J0 et J6, voici les résultats (tableau du pré-print) :

Group	All	Exacerbated	Unchanged	Improved		
				Moderate	Significant	Total
All	62	11 (17.7 %)	9 (14.5 %)	18 (29.0 %)	24 (38.7 %)	42 (67.7 %)
Control, n (%)	31	9 (29.0 %)	5 (16.1%)	12 (38.7 %)	5 (16.1%)	17 (54.8%)
HCQ, n (%)	31	2 (6.5 %)	4 (12.9 %)	6 (19.4%)	19 (61.3%)	25 (80.6%)
P value	0.0476					

Table 2: Absorption of pneumonia on chest CT.

Abbreviations: HCQ, hydroxychloroquine.

V/ ÉLÉMENTS DE CRITIQUES MÉTHODOLOGIQUES

A/ Sur le rationnel

L'étude citée par les auteurs attestant de l'efficacité *in-vitro* de la chloroquine est une *letter*, et non un article original (référence 8 dans le papier).

Le fait que selon les auteurs il existe des recommandations d'expert proposant le traitement du Covid-19 par hydroxychloroquine est basé sur deux citations (référence 9 et 10 dans le papier). La première est une *letter* de Février 2020 mentionnant des essais cliniques en cours, suspendus ou arrêtés se déroulant en Chine, cette *letter* n'est en réalité pas une recommandation d'experts. La seconde est une *letter* publiée par Colson et al en Mars 2020 synthétisant les résultats d'études *in vitro* sur l'activité antivirale de la chloroquine et hydroxychloroquine.

B/ Considérations méthodologiques

1. **Les critères d'inclusion et d'exclusion ne sont ni pertinents, ni valides, ni standardisés, ni fiables ni reproductibles : que représente cet échantillon ?**

Une part des critères d'inclusion imposent en plus de la présence d'un portage objectivé de SARS-CoV-2 par PCR des critères paracliniques de constantes respiratoires et des critères radiologiques de signes tomodensitométriques d'une pneumonie. Outre la définition extrêmement lâche du critère radiologique, qui du fait de son caractère « opérateur-dépendant » a probablement été jugée par les différents investigateurs avec un faible degré de concordance, la focalisation sur des patients présentant une forme de la maladie que l'on pourrait qualifier de modérée est incompréhensible et non motivée par le rationnel de l'étude.

De plus, le premier critère de non-inclusion mentionné est lui aussi défini sans aucune standardisation, et a très probablement mené à des non-inclusions arbitraires massives d'autant plus que le flow-chart de l'étude ne mentionne pas les raisons de non-inclusions de 80 patients (56,3% des patients admissibles !).

Finalement, il est fortement probable que l'échantillon de l'étude présente des biais de sélection majeurs et ne soit en aucun cas représentatif ni de sa population source, ni de la population cible. Cette population n'étant même pas définit avec des critères reproductibles dans l'article.

2. La définition du critère de jugement principal est impropre, non pertinente, non valide, non fiable, non reproductible et incohérente

Le critère de jugement principal (ou les critères puisqu'il est impossible de comprendre quel est la définition d'un succès dans cette étude) est la différence moyenne de délai de retour à une température corporelle normale pendant au moins 72 heures, ou le retour à un niveau de toux mineur pendant au moins 72 heures selon le patient. Cette évaluation est à J5 post-inclusion.

Ces critères sont certes cliniques, mais vu de la gravité potentielle du Covid-19, ils ne sont pas pertinents pour le patient. Le Covid-19 est une pathologie mortelle, le critère d'évaluation devrait être une réduction de taux de mortalité. Quand bien même un critère intermédiaire serait envisagé, il faudrait alors envisager des critères cliniques liés à la réduction de formes graves. Ni la fièvre, ni la toux ne sont des symptômes signant la gravité de l'infection Covid-19. Ce sont des éléments de dyspnée ou de détresse respiratoire qui signent la gravité. Enfin, J5 n'est absolument pas un moment d'évaluation pertinent. D'autant plus qu'il est aujourd'hui mis en évidence que l'évolution vers une forme grave est fréquente entre J6 et J10 après le début des premiers symptômes et ce malgré fréquemment une amélioration clinique transitoire avant l'aggravation.

De plus, l'évaluation de la toux repose sur une déclaration définie de manière très lâche, il est donc fortement probable que l'évaluation de ce critère chez les patients fut l'objet d'un biais de classement majeur.

Enfin, ces critères sont basés sur le fait de repasser en dessous d'un seuil non pathologique de température ou de quantité de toux. Or, les patients n'ont pas été inclus en vérifiant qu'ils étaient au-dessus du seuil pathologique considéré pour ces deux critères ! La publication mentionne d'ailleurs bien que seule une partie des patients étaient au-dessus du seuil problématique au moment du recrutement. Donc, pour une proportion substantielle de l'échantillon, le critère de jugement principal n'était même pas évaluable ! Cette définition incohérente et illogique du critère rend son évaluation caduque.

3. L'analyse statistique du critère de jugement principal est de plus impropre

Le critère de jugement étant défini comme un évènement binaire (passage au-dessous d'un seuil pathologique de température ou de toux pendant une période minimale de temps) et l'évaluation se finissant à J5, ce critère est donc un critère censuré à droite. Les délais moyens

comparés entre les deux groupes sont donc estimés improprement et sont donc biaisés. L'analyse aurait dû être une analyse d'une fonction de survie par estimateur de Kaplan-Meier et comparaison des deux courbes par test de log-rank.

4. Il n'y a en réalité pas de double aveugle dans cette étude

Les auteurs indiquent que l'étude a été réalisée en double-aveugle. Or, le bras témoins n'a pas reçu de comprimés de placebo d'hydroxychloroquine. Dans ces conditions, il est évident que le double aveugle a été levé. Cette étude a donc été réalisée en ouvert. Les médecins investigateurs étaient aussi les évaluateurs. Le critère d'évaluation était mal défini, peu standardisé et dépendant de l'opérateur. Aussi, il est fortement probable que cette situation a mené à des biais de classement majeurs quant à l'évaluation du critère de jugement principal par les expérimentateurs.

5. L'analyse de l'évolution tomodensitométrique est aussi impropre

La définition de la classification des patients est non standardisée, ni reproductible. De plus, l'analyse statistique réalisée est une analyse globale de recherche de différence, et non un test de recherche spécifique d'amélioration supplémentaire par rapport au groupe témoin. Cette analyse est ininterprétable.

6. De nombreux autres points

Je ne citerai pas les innombrables problèmes conceptuels, de méthodologie, d'analyse ou de reporting présent dans cet article car la liste serait longue et superflue vis-à-vis des points majeurs présentés précédemment. A noter que les auteurs ont l'air satisfaits, les limites de la publication ne sont même pas discutées.

VI/ CONCLUSION

Cette étude présente des biais tellement majeurs que les données présentées ne devraient même pas être considérées. C'est une catastrophe sur le plan méthodologique. Ce serait une honte pour la communauté scientifique qu'un processus de peer-review externe aboutisse à une décision favorable de publication dans une revue à comité de lecture. L'attitude la plus utile aurait été de ne pas mettre à disposition ce pré-print.