

**Rapport interne sur la qualité méthodologique de l'étude
« Hydroxychloroquine and azithromycin as a treatment of COVID-19:
results of an open-label non-randomized clinical trial »**

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

Méthodologiste, Chercheur en Biostatistique

Inserm UMR U1246 Sphere, Université de Nantes

24 Mars 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 23 Mars 2020 à 19h15, Santé Publique France reportait 19 856 cas de COVID-19 confirmés par diagnostic virologique, 2082 hospitalisations en réanimation et 860 décès.

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

J'analyserai ici les résultats d'une étude publiée par Gautret et al. dans l'International Journal of Antimicrobial Agents (IJAA) sur l'efficacité de l'hydroxychloroquine et de l'azithromycine dans le traitement du COVID-19. Cette étude a d'abord circulé sous-forme de pré-print au début du mois de Mars via Google Drive, puis a été acceptée pour publication et mise en ligne (sous forme de pre-proof) sur le site de l'IJAA le 20 Mars 2020 ainsi que sur MedRXiv.org. La version analysée dans ce rapport est le pre-proof du site de l'IJAA du 20 Mars. Il semble néanmoins qu'elle ne diffère pas ou quasi pas du pré-print qui circulait auparavant ce qui peut suggérer un processus de peer-review externe très accéléré.

II/ RATIONNEL AVANCÉ PAR L'ÉTUDE

L'étude met en avant deux objectifs de santé vis-à-vis de l'intérêt de disposer de molécules efficaces dans le contexte du COVID-19 : le traitement symptomatique des patients ainsi qu'un potentiel intérêt de réduction de la durée de portage dans la population afin de réduire la probabilité de transmissions entre sujets.

Il est rappelé qu'une stratégie visant à investiguer le potentiel thérapeutique de molécules connues présente comme avantages de travailler sur des molécules dont on connaît le profil de tolérance, les effets secondaires, la posologie plausible ainsi que le profil d'interactions médicamenteuses.

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, ainsi qu'une étude publiée en 2006 montrant une activité inhibitrice *in-vitro* de l'hydroxychloroquine sur le virus SARS-CoV (et non SARS-CoV-2). Puis, une supposée existence de résultats d'essais cliniques Chinois suggérant un effet de la chloroquine, ayant mené en Chine à de potentielles recommandations d'expert préconisant de traiter le Covid-19 symptomatique par chloroquine.

Les auteurs du présent essai justifient leur préférence envers l'hydroxychloroquine et non la chloroquine sur la tolérance, la possibilité de l'utiliser à plus haute dose et un profil d'interactions médicamenteuses plus favorable.

III/ ÉLÉMENTS GÉNÉRAUX DE MÉTHODOLOGIE

Promotion : Assistance Publique-Hôpitaux de Marseille ? Financement par le programme « Investissements d'Avenir » de l'Agence Nationale de la Recherche française

Centres investigateurs : Institut Universitaire d'infectiologie de l'AP-HM de Marseille (principal et recrutement des patients traités par hydroxychloroquine). Les patients du groupe témoins ont aussi été recrutés en plus de Marseille à Nice, Avignon et Briançon.

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 la réduction du portage viral.

Type de l'étude : Essai de phase 2 précoce, unicentrique (groupe hydroxychloroquine) ou multicentrique (groupe témoins), contrôlé, non randomisée, en ouvert, en bras parallèles versus absence d'hydroxychloroquine (traitement symptomatique usuel, pas de placebo d'hydroxychloroquine).

Date et durée de l'étude : Réalisée en Mars 2020. Durée de 14 jours pour les patients inclus.

Critères d'inclusion : Patients hospitalisés avec diagnostic de COVID-19 de 12 ans et plus avec confirmation par PCR d'un portage nasopharyngé de SARS-CoV-2 à l'admission, indépendamment de l'état clinique.

Critères de non-inclusion :

- Allergie connue à l'hydroxychloroquine, la chloroquine

- Contre-indication à un traitement par les molécules de l'étude, y compris les rétinopathies
- Déficit en G6PD
- Allongement du QT
- Femmes enceintes (confirmées par test de grossesse positif ou déclaratif)
- Femmes en situation d'allaitement maternel (déclaratif)

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. Visites de suivi journalières pendant 14 jours (examen clinique standardisé et prélèvement nasopharyngé journalier si possible pour PCR). Tous les patients hospitalisés à Marseille se voyaient proposer un traitement oral par hydroxychloroquine, à la dose de totale de 600 mg par jour, en trois prises de 200 mg. Les patients de Marseille refusant le traitement ou ne pouvant être inclus pouvaient faire office de patients témoins. Les patients des autres centres faisaient office de témoins. 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, trois prises de 200 mg par jour + traitement usuel *versus* traitement usuel.

Critère de jugement principal : Négativation du portage viral à J6 post-inclusion.

Critères de jugement secondaires : Négativation du portage viral à tout moment au cours du suivi. Suivi clinique (Température corporelle, Fréquence respiratoire, Durée de séjour et Taux de mortalité). Survenue d'effets secondaires.

Taille de l'échantillon : L'étude indique que le calcul du nombre de sujets nécessaire s'est basé sur une hypothèse de 50% d'efficacité supplémentaire de l'hydroxychloroquine sur le critère de jugement principal, avec un risque alpha fixé à 5% et une puissance souhaitée de 85%. Les investigateurs ont aussi supposé 10% de perdus de vue. Un recrutement de 24 sujets par groupe a ainsi été envisagé.

Méthode d'analyse des résultats : L'existence d'une différence statistiquement significative sur le critère de jugement principal entre les deux groupes de comparaison a été analysée par test exact de Fisher ou test de χ^2 , selon vérification des conditions de validité du test de χ^2 (de même pour les autres variables catégorielles d'intérêt). Concernant les variables quantitatives, la comparaison a été réalisée par tests de Student.

Population d'analyse : Analyse « per-protocol » après exclusion des patients présentant des données incomplètes concernant le critère de jugement principal à J6 post-inclusion. Aucun report de critères définissant les déviations mineures ou majeures au protocole.

IV/ SYNTHÈSE DES PRINCIPAUX RÉSULTATS

L'étude comporte 26 patients dans le groupe traité par hydroxychloroquine et 16 témoins. Six patients initialement traités par hydroxychloroquine sont perdus de vue et ont été exclus des analyses pour données manquantes sur le critère de jugement principal : 3 patients pour transferts en réanimation (tous PCR positifs le jour précédant le transfert, un patient décédé à J3 avec PCR négative à J2, un arrêt de traitement à J3 pour nausées avec PCR positive à J2, un patient a pris la décision de quitter l'hôpital à J3 et était PCR négative à J2).

Voici les caractéristiques à l'inclusion des deux groupes tels que présentés dans le pré-proof (Note. Le statut clinique ci-dessous correspond à trois groupes : 1/ asymptomatiques, 2/ symptômes respiratoires hauts ou syndrome pseudo-grippal, 3/ symptômes respiratoires bas) :

	Age (years)			Male gender		Clinical status				Time between onset of symptoms and inclusion (days)		
	Mean ± SD	t	p-value	n (%)	p-value	Asymptomatic	URTI	LRTI	p-value	Mean ± SD	t	p-value
Hydroxychloroquine treated patients (N=20)	51.2 ± 18.7	-1.95	0.06	9 (45.0)	0.65	2 (10.0)	12 (60.0)	6 (30.0)	0.30	4.1 ± 2.6	-0.15	0.88
Control patients (N=16)	37.3 ± 24.0			6 (37.5)		4 (25.0)	10 (62.5)	2 (12.5)		3.9 ± 2.8		
All patients (36)	45.1 ± 22.0			15 (41.7)		6 (16.7)	22 (61.1)	8 (22.2)		4.0 ± 2.6		

URTI: upper tract respiratory infection, LRTI: lower tract respiratory infection

Parmi les 26 patients traités sous hydroxychloroquine, 6 ont reçu de l'azithromycine pour prévention de sur-infection bactérienne (500 mg le premier jour, suivi de 250 mg les 4 jours suivants).

Concernant l'analyse du critère de jugement principal (population « per-protocol », après exclusion de 6 patients perdus de vue dans le groupe hydroxychloroquine), il est indiqué que 14/20 patients (70%) présentaient un portage viral négatif à J6 *versus* 2/16 patients (12,5%), différence statistiquement significative au risque alpha de 5% ($p < 0,001$).

Sur les 20 patients analysés du groupe hydroxychloroquine, 6 ont reçu en plus de l'azithromycine. Une analyse complémentaire a été effectuée comparant les trois groupes (hydroxychloroquine + azithromycine *versus* hydroxychloroquine seule *versus* absence de traitement). Les résultats sont 100% (6/6) de négativation du portage viral à J6 dans le groupe hydroxychloroquine + azithromycine *versus* 57% (8/14) dans le groupe hydroxychloroquine seule *versus* 12,5% (2/16) dans le groupe absence de traitement. Il est reporté une différence statistiquement significative au risque alpha de 5% entre ces trois groupes ($p < 0,001$).

L'effet des traitements médicamenteux sur la réduction du portage viral serait statistiquement plus important chez les patients présentant des symptômes que chez les patients asymptomatiques ($p < 0,05$). Toutefois les résultats quantitatifs ne sont pas reportés dans la publication.

Les résultats concernant les autres critères présentés ne sont pas reportés dans la publication actuelle.

En conclusion, les auteurs concluent à l'efficacité de l'hydroxychloroquine et supposent un effet synergique de l'azithromycine en combinaison avec l'hydroxychloroquine.

VI/ ÉLÉMENTS DE CRITIQUES MÉTHODOLOGIQUES

A/ Sur le rationnel

Les auteurs basent une partie de leur rationnel sur l'existence de preuves cliniques de l'efficacité de l'hydroxychloroquine, et citent deux références en provenance de Chine (référence 9 et 10 dans la publication). La première est une *letter*, ne contenant pas de données d'essais en soit, mais faisant référence à une conférence tenue en Chine en Février. La seconde citation renvoie à plusieurs essais cliniques tenus en Chine, soient suspendus ou arrêtés, soit

toujours en cours de recrutement. Aussi, à ma connaissance, il n'y a pour le moment pas de données publiées sous forme d'article original d'essais clinique testant l'efficacité de la chloroquine dans le traitement du Covid-19.

Les auteurs basent aussi une partie de leur rationnel sur des recommandations de pratique proposés par des experts chinois suggérant le traitement des formes symptomatiques de Covid-19 par la chloroquine (référence 11). Le document cité est indexé dans Medline et accessible via PubMed. Il n'est cependant disponible qu'en mandarin.

B/ Considérations éthiques

Un aspect interroge particulièrement dans cette étude : l'inclusion de patients ayant refusé le traitement actif comme témoins. A priori, si ces patients ont refusé le traitement actif, c'est un refus de participation à l'étude. Certes, en termes de stratégie thérapeutique, ces patients ont donc reçu les soins standards. Néanmoins, ils n'auraient pas dû être inclus dans l'étude, ayant refusé. Ceci pose tout de même la question des conditions de délivrance des informations liées au protocole et de l'obtention d'un consentement libre et éclairé pour tous les patients de cette étude.

De plus, même si c'est mineur (et moins important que l'aspect éthique), en termes méthodologiques, cette inclusion peut être considéré comme une rupture de la clause d'équivalence (les patients doivent pouvoir en théorie recevoir les deux stratégies comparées avant d'être inclus) dans un essai contrôlé. Si les raisons du refus de participation ne sont pas indépendantes du pronostic, cette inclusion peut être source de biais.

C/ Considérations méthodologiques

1. L'absence de randomisation

La critique méthodologique majeure concernant cette étude est l'absence de randomisation des patients selon les deux stratégies. Cette absence empêche de garantir la comparabilité initiale des groupes. Aussi, la différence observée sur le critère de jugement principal ne peut être attribué causalement sans ambiguïté à la présence de l'hydroxychloroquine, mais peut être la résultante de nombreux facteurs de confusion ayant un rôle explicatif pronostic. A ce sujet, il faut de plus noter que les caractéristiques à l'inclusions recueillies sont en faible quantité (âge, genre, statut clinique en trois catégories, délais entre

l'apparition des symptômes et l'hospitalisation) et ne permettent donc nullement d'apprécier un potentiel déséquilibre des caractéristiques entre les deux groupes. Enfin, certes les tests statistiques de comparaison des caractéristiques de base reportés ne permettent pas de rejeter l'hypothèse nulle d'absence de différences au risque alpha de 5%. Néanmoins il est nécessaire de rappeler que les tests statistiques font partie d'une démarche d'inférence construite pour mettre en évidence des différences et non pour prouver des égalités (« l'absence de preuves n'est pas la preuve de l'absence »). Ceci est particulièrement critique dans ce cas où les effectifs étant faibles, les tests réalisés sont peu puissants. On peut noter que sur la base des indices de position présentés, les patients du groupe contrôle sont plus jeunes, et ont un profil symptomatologique plus fréquemment asymptomatiques ou à type d'infection respiratoire haute.

De nombreuses caractéristiques possiblement pronostiques n'ont pas été documentées (les facteurs de risques d'évolution grave actuellement recherchés chez les patients suspects d'infection Covid-19).

Pour en revenir aux considérations éthiques, s'il est possible d'envisager que le caractère exceptionnel de la pandémie puisse faire discuter le caractère éthique de la randomisation en termes de pertes de chance pour les patients inclus dans les groupes témoins par aléa de la randomisation, la publication ne commente à aucun moment le rationnel de la décision de non-randomisation, pourtant un point méthodologique fondamental aidant à la démonstration d'un lien de causalité dans ce type d'études. Si la publication indique que sa mise en œuvre a été décidée après avis d'un Comité de Protection des Personnes, il semble néanmoins possible de discuter que les bénéfices d'une randomisation bien conduite en termes d'augmentation de la qualité du niveau de preuve pouvaient se justifier éthiquement.

2. L'absence de toute méthode d'ajustement sur des covariables de confusion

En l'absence de conduite d'une randomisation, la gestion des nombreux biais de confusion potentiels pouvait s'envisager via les stratégies possibles d'ajustement. Pragmatiquement, cette critique est néanmoins à tempérer car de toute façon au vu du nombre d'événements positifs sur la variable de critère de jugement principal (16 événements), les possibilités d'ajustements sur de nombreux facteurs via l'analyse multivariée étaient très restreintes et des stratégies de redressement de type score de propensions inenvisageables.

Toutefois, une analyse de sensibilité après ajustement sur un petit nombre de facteurs pronostiques essentiels aurait été un élément important conférant un peu plus de crédibilité aux résultats de l'étude vis-à-vis de la gestion des facteurs de confusion.

3. La population d'analyse

L'analyse a été réalisée sur une population que l'on peut grossièrement appeler « per-protocol » : 6 patients ayant été exclus du groupe hydroxychloroquine pour absence de données concernant le critère de jugement principal. Aussi l'analyse n'est pas une analyse en intention de traiter. Ce fait constitue un biais d'attrition non négligeable. En effet, sur ces 6 patients, 5 présentent des caractéristiques suggérant une évolution péjorative sur le plan clinique (3 transferts en réanimation, 1 décès (certes avec PCR négative, néanmoins sur le plan clinique cela ne peut évidemment pas être considéré comme un succès de prise en charge), 1 arrêt de traitement pour nausées). Ceci suggère que ces attritions sont potentiellement un biais majeur dans l'évaluation de l'efficacité de l'hydroxychloroquine dans ce contexte.

A minima, une analyse en intention de traiter avec différents scénarii d'imputation des résultats manquants sur le critère de jugement principal (par exemple, au minimum un « worst case scenario ») aurait été une analyse de sensibilité intéressante pour témoigner de la robustesse des résultats.

4. Le choix de l'outcome

Le critère de jugement principal choisi était la négativation du portage naso-pharyngé du virus SARS-CoV-2 objectivé par PCR à J6 post-inclusion : soit un critère de jugement intermédiaire binaire.

Premièrement, si les conditions exceptionnelles de la crise actuelle associées aux connaissances cliniques faibles dont nous disposons actuellement sur le Covid-19 peuvent faire envisager le choix d'un critère précoce, il n'est absolument pas discuté dans la publication. Il est possible que dans cette situation pathologique spécifique, la chaîne de causalité de la négativation du portage viral à la guérison clinique soit simple, et que ces deux critères soient fortement corrélés. Il faut néanmoins rappeler qu'à ce stade c'est une conjecture et que le critère choisi n'est pas un critère clinique pertinent pour le patient. D'autant plus que la corrélation entre disparition du virus et guérison clinique a été mis en défaut dans d'autres situations (par

exemple dans les formes sévères d'infection au virus de la grippe ou il est connu que malgré une activité antivirale claire de certaines molécules données en contexte de réanimation, nous n'avons aucune preuve d'une amélioration du pronostic clinique en plus des traitements purement symptomatiques de suppléance des défaillances viscérales). Il faut de plus noter que la publication ne discute pas du fait que ce critère n'a pas la même importance selon la stratégie choisie (traitement curatif après contamination ou diminution de la charge virale en population).

Ensuite, le délai à partir duquel a été mesuré le critère de jugement principal n'est pas justifié et aucun argument ne vient étayer la décision de mesurer à J6, d'autant plus que les patients sont suivis jusqu'à J14.

Par ailleurs, le critère est une discrétisation d'une mesure initialement continue (un portage viral quantitatif). Dans ce contexte de phase précoce, et au vu du peu de connaissances actuelles sur la maladie Covid-19, ne pas discrétiser cette mesure aurait pu être intéressant. D'autant plus que les investigateurs ont procédé à un prélèvement journalier. Garder toute l'information quantitative aurait pu permettre de produire une cinétique d'effet observé sur l'ensemble de la période. Outre l'avantage d'être plus riche d'un point de vue interprétatif, cette cinétique aurait de plus augmenté la quantité d'informations sur le plan statistique, permettant ainsi d'analyser la trajectoire des deux groupes via une modélisation par modèles mixtes pour mesures répétées, ce qui aurait permis de modéliser la forme de la trajectoire, d'ajuster sur le niveau de base de chaque sujet, d'étudier l'interaction temps/groupe, d'ajuster sur d'autres covariables de base, et d'inclure l'information déjà obtenue sur les patients exclus pour les points pour lesquels la PCR a été réalisée.

5. L'absence d'aveugle

Ce point est mineur mais mérite d'être mentionné. Certes, concernant l'évaluation, le critère étant un critère biologique objectif peu « opérateur dépendant » (un résultat de PCR), on peut supposer que l'absence d'aveugle n'ait pas mené à un potentiel biais d'évaluation majeur. Néanmoins les données brutes des 36 patients analysés sont disponibles dans une table (au format doc...), et on peut y constater que pour l'ensemble des patients et des temps de mesure la codification du résultat de la PCR n'est pas homogène (parfois quantitative, parfois purement binaire).

6. La place de l'azithromycine

Six patients du groupe traité par hydroxychloroquine ont reçu de l'azithromycine (un antibiotique) pour prévention de sur-infection bactérienne. Sur la base d'une analyse en sous-groupes, les auteurs concluent que leur étude suggère un effet synergique de l'azithromycine sur le critère de jugement principal. Il faut néanmoins noter que le test statistique réalisé est un test global de présence d'une différence entre au moins un des trois groupes hydroxychloroquine + azithromycine ou hydroxychloroquine seule ou absence de traitement. Toutefois, lorsque l'on teste spécifiquement la présence d'une différence statistiquement significative uniquement entre les groupes hydroxychloroquine+azithromycine versus l'hydroxychloroquine seule, en utilisant les données reportées dans la publication, on ne peut rejeter l'hypothèse nulle d'absence de différence statistiquement significative entre les deux groupes au risque alpha de 5% (par test exact de Fisher (conditions de validité d'un test de χ^2 non réunies) $p = 0,11$). De plus, méthodologiquement, préciser correctement cet effet aurait idéalement nécessité un autre bras parallèle dédié à cette stratégie de combinaison.

7. L'absence de discussion des limites

Il est à signaler que la publication accorde une phrase d'une ligne et demie pour discuter les limites de l'étude ce qui semble bien insuffisant au vu de l'ensemble des critiques soulevées précédemment.

8. Divers points factuels d'analyses ou de reporting

Divers points mineurs critiquables que ce soit du point de vue des analyses ou du reporting peuvent être signalés : l'absence de tout indicateur de précision sur la mesure du critère de jugement principal (au moins dans le bras chloroquine, un intervalle de confiance à 95% de la proportion de négativation de portage viral aurait pu être produit par une technique de rééchantillonnage non paramétrique) ; l'utilisation de la moyenne et de l'écart-type pour le report des variables quantitatives alors qu'au vu de la taille de l'échantillon médiane et écart inter-quartiles pouvaient être plus judicieux ; les deux figures présentant la comparaison du critère de jugement principal pour chaque jour après l'inclusion selon une stratégie d'analyse impropre pour diverses raisons ; le reporting incomplet du calcul du nombre de sujets nécessaires (les auteurs mentionnent une hypothèse de 50% d'efficacité supplémentaire, mais sans savoir de quelle base on part le calcul ne peut être reproduit)...

L'ensemble de ces détails est bien mineur et leur occurrence compréhensible au vu du timing extrêmement serré, toutefois ils entachent le reporting de l'étude. Or la répliquabilité des essais cliniques participe à l'effort de recherche clinique et cette répliquabilité est facilitée par un reporting de qualité.

VI/ CONCLUSION

Cette étude montre une négativation du portage viral pour 14 sujets traités par hydroxychloroquine parmi 20 patients 6 jours après admission à l'hôpital pour Covid-19 objectivée par une PCR positive à SARS-Cov-2 sur prélèvement nasopharyngé. Dans le même temps, 2 patients sur 16 ont présenté ce même évènement sans traitement par hydroxychloroquine.

Toutefois, au vu des nombreuses limites soulevées, le niveau de preuve que l'on peut conférer par cette étude ne peut être considéré au-dessus de celui d'une série de cas, c'est-à-dire très faible.

Les limites méthodologiques majeures sont des limites de design, aussi elles ne peuvent être modifiées a posteriori. Néanmoins, ces données existent et des analyses statistiques plus appropriées pourraient être réalisées afin de préciser un peu plus le potentiel de l'hydroxychloroquine dans le traitement du Covid-19, sous réserve que l'ensemble des données soient disponibles, y comprises celles des patients exclus à posteriori. Néanmoins, il me semble raisonnable d'affirmer que sur la base de ces données seules, cela restera au stade de génération d'hypothèses, l'étude ne présente pas les critères nécessaires pour être une étude de confirmation à fort niveau de preuve.