

# Sciences sociales computationnelles

Analyse de données Twitter

Émilien Schultz - SESSTIM (AMU)  
[emilien.schultz@univ-amu.fr](mailto:emilien.schultz@univ-amu.fr)

# Une définition rapide

“Computational social science is an interdisciplinary field that advances theories of human behavior by applying computational techniques to large datasets from social media sites, the Internet, or other digitized archives such as administrative records”

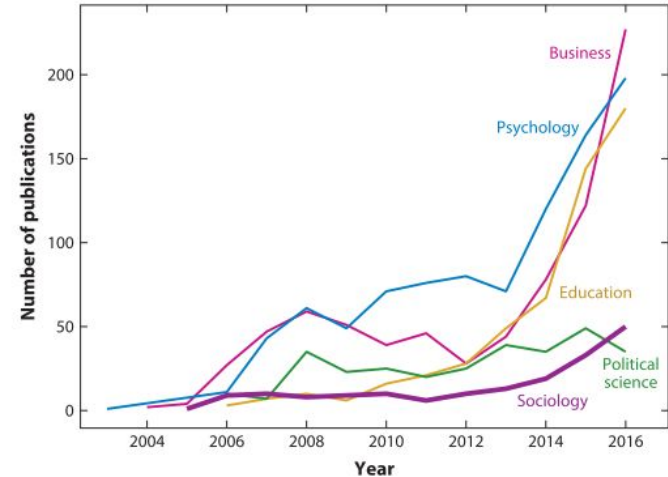


Figure 1

Number of computational social science publications by year—2003–2016—across four scholarly disciplines.

[Computational Social Science and Sociology](#), Achim Edelmann, Tom Wolff, Danielle Montagne, Christopher A. Bail Annual Review of Sociology 2020 46:1, 61-81

# Les sciences sociales computationnelles ?

## Rencontre entre deux spécialités

**Au-delà des *big data***  
Les sciences sociales et la multiplication des données numériques

***Beyond big data***  
*Social sciences and the proliferation of digital data*

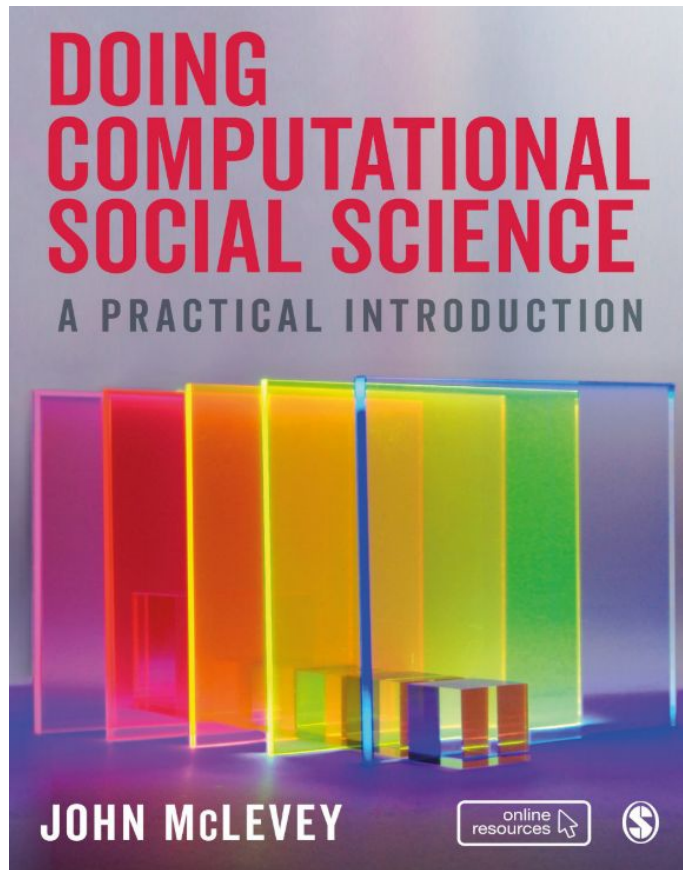
par Étienne Ollion\* et Julien Boelaert\*\*

### R É S U M É

Dans le débat public comme dans le monde académique, l'enthousiasme pour les *big data* n'a eu d'égal que les critiques que ce phénomène a suscité. « Opportunité empirique inouïe » vs « données pauvres » ; « révolution méthodologique » vs « fascination pour le nombre » ; « révolution scientifique » vs « dégradation du savoir produit » : les positions sont tranchées. À partir d'une lecture de ces débats et des travaux en sciences sociales souvent regroupés sous ce label, l'article soutient que cette situation polarisée a de fortes chances de perdurer tant que la discussion s'organise autour du concept mal défini de *big data*. Il propose de distinguer différents types de données souvent regroupées sous ce terme. Il montre ce faisant que les *big data* souvent évoquées ne sont qu'un aspect limité d'une transformation bien plus importante : la disponibilité croissante et massive de données numériques, qui pose des questions nouvelles à nos disciplines. Quatre aspects sont plus particulièrement explorés : les réorganisations disciplinaires, les transformations des méthodes quantitatives, l'accès et la gestion des données, les objets des sciences sociales et leur rapport à la théorie.

### A B S T R A C T

*In the public sphere as in academia, the frenzy over big data has been met with equally high levels of enthusiasm and criticism. "Unprecedented empirical opportunity" vs. "poor data"; "methodological revolution" vs. "fascination for large N"; "scientific revolution" vs. "debasement in the production of knowledge": the positions are polarized. Based on a review of the debates in the social sciences, this paper shows that this situation is likely to persist as long as the discussion is organized around the fuzzy term of "big data." Instead, this paper proposes to distinguish between different types of data, each of which raises specific questions. By so doing, it shows that big data is but one aspect of a much wider transformation – the massive availability of digital data – which in turn raises new questions for our disciplines. Four aspects are subsequently explored: disciplinary reorganizations, the evolutions of statistical methods, the access and management of data, as well as the objects of social sciences and their relationship to theory.*



# Rien de neuf, mais des données/outils plus accessibles

- Analyse de réseaux dès les années 80
- Agent-based models
- Statistiques “exploratoires” (e.g. Benzecri)

*Journal of Mathematical Sociology*  
1995, Vol. 20(2-3), pp. 79-87  
Reprints available directly from the publisher  
Photocopying permitted by license only

© 1995 OPA (Overseas Publishers Association)  
Amsterdam B.V. Published under license by  
Gordon and Breach Science Publishers SA.  
Printed in Malaysia

## THE EMERGENCE OF COMPUTATIONAL SOCIOLOGY

NORMAN P. HUMMON\* and THOMAS J. FARARO\*

*Department of Sociology, University of Pittsburgh, Pittsburgh, PA 15260, USA*

The world of science has undergone a major transformation by virtue of technological innovations in computing and information processing. Sociology is one site in which this change is being played out. Our basic aim is to set out a revised image of any modern science, within which we can conceptualize and discuss the role of a newly emergent subfield we term computational sociology. Specifically, we expand the familiar two-component model of a science, featuring a theoretical and an empirical side, to include a computational component. We show how the three components interrelate in a triangular system in which empirical data analysis, theoretical explanation and computer simulation link the three components. We close our paper with a brief discussion of how one new development in computation relates to concepts of sociology, an instance of the hybrid character of computational sociology.

# Quelques lieux des SSC en France

## Le médialab

Laboratoire de recherche interdisciplinaire réunissant sociologues, ingénieurs et designers, le médialab mène des recherches thématiques et méthodologiques exploitant et interrogeant la place prise par le numérique dans nos sociétés.

### La recherche au médialab

Le numérique transforme nos sociétés mais aussi la production de la connaissance scientifique. Centrale dans l'approche développée au médialab, la transition numérique invite à articuler trois approches fondées sur les sciences sociales, l'ingénierie et le design. Par la multitude de données qu'elle génère, la transition numérique contribue à élargir les connaissances que nous avons des différents mondes sociaux. En parallèle, elle offre des instruments d'investigation favorisant de nouveaux modes de production des savoirs à travers la modélisation, la visualisation et l'exploration interactive des corpus. Enfin elle insiste à expérimenter avec les publics et les étudiants en encourageant de nouvelles formes de réflexivité et de partage des résultats de la recherche.

Ces approches sont développées conjointement au sein de quatre axes thématiques :

## CSS @ IPP

The Computational Social Science group of the Institut Polytechnique de Paris



[Home](#) [News](#) [Resources](#) [Team](#) [Join us](#) [Contact](#)

## Tutorial: The Augmented Social Scientist

Presentation at IC2S2 in Copenhagen, July 2023

[Read more](#)

**Computational Social Sciences** is an interdisciplinary endeavor that uses digital data in all its forms to generate fundamental and policy-oriented insights about society.

The [CSS team at Institut Polytechnique de Paris](#) is invested in tackling the challenges raised by the current computational revolution, at various levels:

- Leveraging the empirical power of these new masses of data
- Analyzing how the new data abundance affects our daily lives and societies altogether
- Reflecting on how artificial intelligence and statistical learning can be used for social scientific research
- Studying the individuals who operate digital infrastructures

[Each of our team members](#) is invested in one of these fastly developing areas of CSS. This diversity in approaches and methods allows us to meet the challenges raised by these deep-seated transformations.

# Quelques exemples d'études

Humanities & Social Sciences  
Communications



LES SOMMETS

## DU PALAIS

ANALYSER L'ESPACE PARLEMENTAIRE  
AVEC DES CARTES AUTO-ORGANISATRICES<sup>1</sup>

Julien Boelaert, Étienne Ollion

Il est un peu plus de 13 h, ce 20 juin 2017, à l'Assemblée nationale<sup>2</sup>. En ce jour de rentrée, les députés élus ou réélus quelques jours auparavant viennent procéder aux formalités d'usage. Successivement, ils passent par les diverses étapes de l'enregistrement. D'un bureau à l'autre, ils déclinent leur identité et posent pour la photo officielle qui ornera leur page Internet. Ils se voient remettre les identifiants de leurs différents comptes, collectent les insignes du parlementaire (l'écharpe tricolore, la cocarde pour la voiture ou encore le « baromètre »).

À la suite de l'élection d'Emmanuel Macron le mois précédent, c'est une législature profondément renouvelée qui prend ce jour possession des lieux. Avec 72 % de primo-députés (contre environ un tiers d'habitude), 28 % de parlementaires jamais élus à quelque mandat politique auparavant (contre 5 % lors de la mandature précédente), l'élection a été l'occasion d'un grand renouvellement des visages. Elle a non seulement vu nombre de responsables politiques historiques être éliminés, mais ils ont parfois été remplacés par de parfaits inconnus en politique. Plus qu'une alternance, la XV<sup>e</sup> législature est marquée par le nombre record de novices dans ses rangs. Près d'un cinquième des nouveaux élus ont en effet une expérience politique nulle ou limitée.

### ARTICLE

<https://doi.org/10.1057/s41599-021-00615-9>

OPEN

## Computational appraisal of gender representativeness in popular movies

Antoine Mazières<sup>1</sup> , Telmo Menezes<sup>1</sup> & Camille Roth<sup>1,2</sup>

Gender representation in mass media has long been mainly studied by qualitatively analyzing content. This article illustrates how automated computational methods may be used in this context to scale up such empirical observations and increase their resolution and significance. We specifically apply a face and gender detection algorithm on a broad set of popular movies spanning more than three decades to carry out a large-scale appraisal of the on-screen presence of women and men. Beyond the confirmation of a strong under-representation of women, we exhibit a clear temporal trend towards fairer representativeness. We further contrast our findings with respect to a movie genre, budget, and various audience-related features such as movie gross and user ratings. We lastly propose a fine description of significant asymmetries in the *mise-en-scène* and *mise-en-cadre* of characters in relation to their gender and the spatial composition of a given frame.



# Récemment, beaucoup d'études dans le contexte COVID sur les médias sociaux

masques, confinement, chloroquine, vaccins, covid long, etc.

PLOS ONE

## RESEARCH ARTICLE

### Assessing the influence of French vaccine critics during the two first years of the COVID-19 pandemic

Mauro Faccin<sup>1\*</sup>, Floriana Gargiulo<sup>2</sup>, Laëtitia Atlani-Duault<sup>1,3,4,5</sup>, Jeremy K. Ward<sup>6</sup>

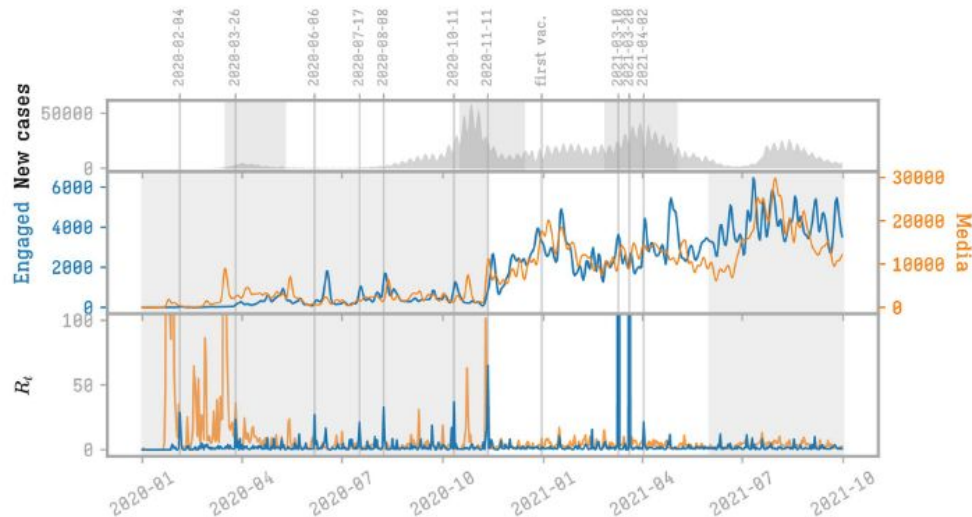
**1** IRD and CEPED, Université de Paris, Paris, France, **2** GEMASS, CNRS, Paris, France, **3** Institut COVID-19 Add Memoriam, Université de Paris, Paris, France, **4** WHO Collaborative Center for Research on Health and Humanitarian Policies and Practices, Université de Paris, Paris, France, **5** Mailman School of Public Health, Columbia University, New York, NY, United States of America, **6** CERMEPS, INSERM, CNRS, EHESP, Université de Paris, Villejuif, France

\* mauro.faccin@gmail.com



## Abstract

When the threat of COVID-19 became widely acknowledged, many hoped that this pandemic would squash "the anti-vaccine movement". However, when vaccines started arriving in rich countries at the end of 2020, it appeared that vaccine hesitancy might be an issue even in the context of this major pandemic. Does it mean that the mobilization of vaccine-critical activists on social media is one of the main causes of this reticence to vaccinate against COVID-19? In this paper, we wish to contribute to current work on vaccine hesitancy during the COVID-19 pandemic by looking at one of the many mechanisms which can cause reticence towards vaccines: the capacity of vaccine-critical activists to influence a wider public on social media. We analyze the evolution of debates over the COVID-19 vaccine on the French Twittersphere, during two first years of the pandemic, with a particular attention to the spreading capacity of vaccine-critical websites. We address two main questions: 1) Did vaccine-critical contents gain ground during this period? 2) Who were the main actors in the diffusion of these contents? While debates over vaccines experienced a tremendous surge during this period, the share of vaccine-critical contents in these debates remains stable except for a limited number of short periods associated with specific events. Secondly, analyzing the community structure of the re-tweets hyper-graph, we reconstruct the mesoscale structure of the information flows, identifying and characterizing the major communities of users. We analyze their role in the information ecosystem: the largest right-wing community has a typical echo-chamber behavior collecting all the vaccine-critical tweets from outside and recirculating it inside the community. The smaller left-wing community is less permeable to vaccine-critical contents but, has a large capacity to spread it once adopted.



**Fig 3. Engagement of users with vaccine-critical content.** Above: evolution of the total number of users engaging with news media (orange) or with vaccine-critical contents (blue). Below: the reproduction number shows peaks of engagement around some events. The greyed area on the top panel is proportional to the daily new cases in France.

<https://doi.org/10.1371/journal.pone.0271157.g003>

<https://pubmed.ncbi.nlm.nih.gov/35925962/>

## OPEN ACCESS

**Citation:** Faccin M, Gargiulo F, Atlani-Duault L, Ward JK (2022) Assessing the influence of French vaccine critics during the two first years of the COVID-19 pandemic. PLOS ONE 17(8): e0271157. <https://doi.org/10.1371/journal.pone.0271157>

**Editor:** Constantine Dvorolis, Georgia Institute of Technology, UNITED STATES

**Received:** February 24, 2022

**Accepted:** June 27, 2022

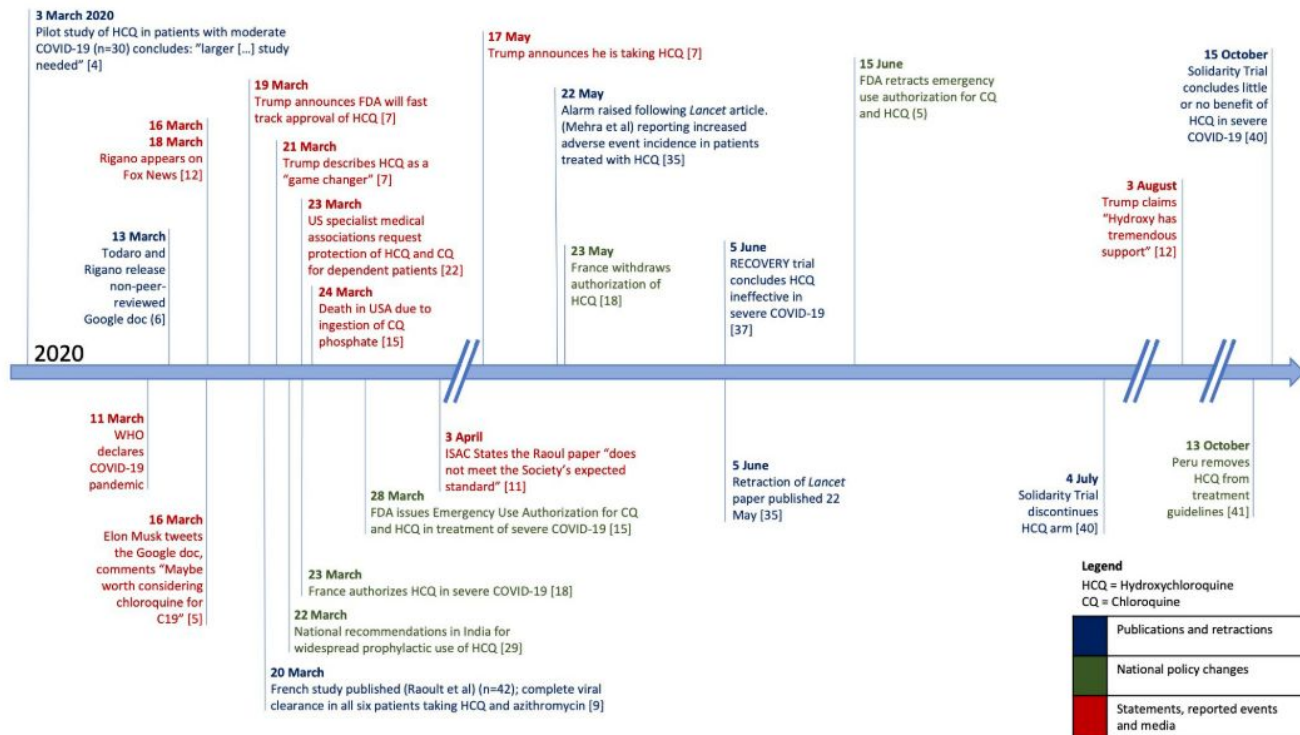
**Published:** August 4, 2022

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0271157>

**Copyright:** © 2022 Faccin et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original

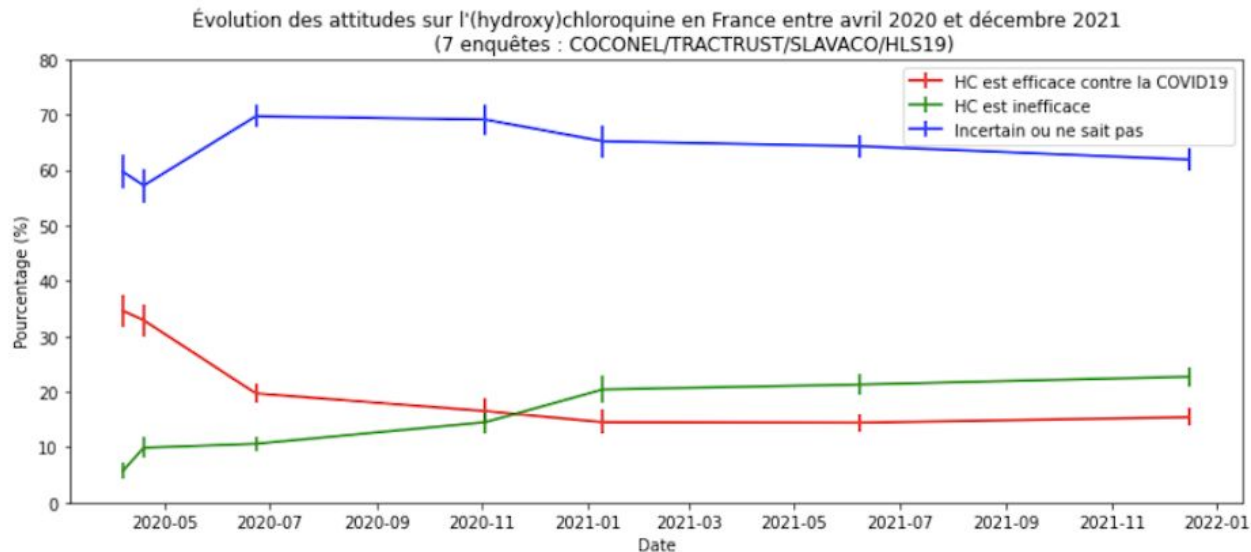
# Ce TP : je vous embarque dans une recherche en cours

**Figure 1. Timeline of key events in the story of chloroquine/hydroxychloroquine and COVID-19**





# Un débat au long terme



<https://theconversation.com/les-francais-et-la-chloroquine-une-defaite-de-la-culture-scientifique-174482>

# Une affaire pas finie pour tous...



**Didier Raoult** ✓  
@raoult\_didier

...

Hydroxychloroquine et mensonges

L'hydroxychloroquine dans le covid: cela tue (clamé par le Lancet dans un article frauduleux et rétracté mais entériné par Mr Veran et l'OMS en 24 H).

L'hydroxychloroquine cela ne marche pas, nos chiffres sont truqués (clamé par "complément d'enquête"), un huissier confirme nos données. Notre dernier travail, contrôlé par huissier, rapporte une diminution de 70% de morts chez les patients traités précocement et 40% sur 30 000 patients hospitalisés entre 2020 et 2021, soit 800 à 1000 morts évités à Marseille (étude en ligne).

Oui mais c'est illégal. Cela n'a pas été noté par la visite à charge de 8 inspecteurs pendant 7 mois (ANSM, IGAS, IGAESRI) ni par l'avis du directeur général de la santé (Pr Salomon) consulté en 2022.

Ces résultats sont comparables à ceux de la synthèse du site "early C19" (visible sur internet) basé sur 393 études incluant 500 000 personnes. Est-ce illégal de sauver des gens? Il devient urgent qu'une commission indépendante examine ces résultats qui sont tous disponibles sur internet.



**La Provence** @laprovince · 21h

...

Didier Raoult, l'éternel retour : l'ex-directeur de l'IHU publie à nouveau une étude sur l'hydroxychloroquine

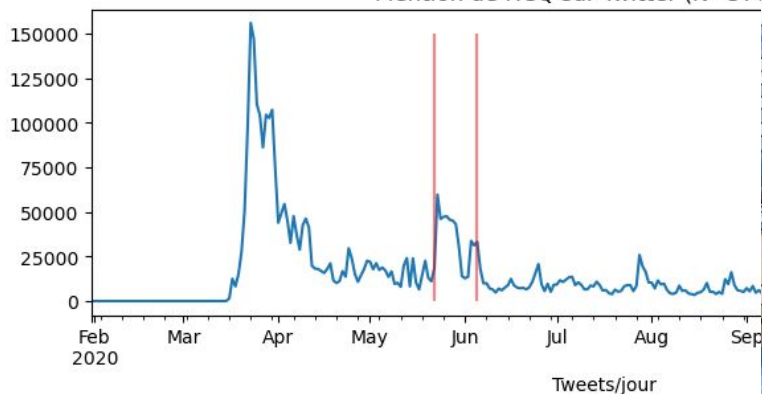
👉 [l.laprovence.com/TTj2](https://l.laprovence.com/TTj2)



# Médias sociaux & le #Lancetgate en France : un moment de "l'affaire HCQ"

Un rebond dans les débats publics sur l'efficacité de HCQ largement initié par la visibilité de D. Raoult (et sa chaîne Youtube)

Mention de HCQ sur Twitter (N=576)



Article 19 (abrogé)

Version révisée de l'article 19 du décret n°2020-630 du 26 mai 2020 - art. 1



## Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis

Mandeep R Mehra, Sapan S Desai, Frank Ruschitzka, Amit N Patel

### Summary

**Background** Hydroxychloroquine or chloroquine, often in combination with a second-generation macrolide, are being widely used for treatment of COVID-19, despite no conclusive evidence of their benefit. Although generally safe when used for approved indications such as autoimmune disease or malaria, the safety and benefit of these treatment regimens are poorly evaluated in COVID-19.

**Methods** We did a multinational registry analysis of the use of hydroxychloroquine with or without a macrolide for treatment of COVID-19. The registry comprised data from 671 hospitals in 35 continents. We included patients hospitalised between Dec 20, 2019, and April 14, 2020, with a positive laboratory finding for SARS-CoV-2. Patients who received one of the treatments of interest within 48 h of diagnosis were included in one of four treatment groups (chloroquine alone, chloroquine with a macrolide, hydroxychloroquine alone, or hydroxychloroquine with a macrolide), and patients who received none of these treatments formed the control group. Patients for whom one of the treatments of interest was initiated more than 48 h after diagnosis or while they were on mechanical ventilation, as well as patients who received remdesivir, were excluded. The main outcomes of interest were in-hospital mortality and the occurrence of de-novo ventricular arrhythmias (as defined as sustained or transient ventricular tachycardia or ventricular fibrillation).

**Findings** 96 032 patients (mean age 53·8 years, 46·3% women) with COVID-19 were hospitalised during the study period and met the inclusion criteria. Of these, 3016 patients were in the treatment groups (1868 received chloroquine, 3783 received chloroquine with a macrolide, 3016 received hydroxychloroquine, and 6221 received hydroxychloroquine with a macrolide) and 95 016 patients were in the control group. 10 698 (11·1%) patients died in hospital. After controlling for multiple potential confounding factors (age, sex, race or ethnicity, body-mass index, underlying cardiovascular disease and its risk factors, diabetes, underlying lung disease, smoking, immunosuppressed condition, and baseline disease severity), when compared with mortality in the control group (9·3%), hydroxychloroquine (18·0%; hazard ratio 1·335, 95% CI 1·238–1·457), hydroxychloroquine with a macrolide (23·8%; 1·447, 1·368–1·531), chloroquine (16·4%; 1·365, 1·218–1·531), and chloroquine with a macrolide (22·2%; 1·368, 1·273–1·469) were each independently associated with an increased risk of in-hospital mortality. Compared with the control group (0·3%), hydroxychloroquine (6·0%; 2·360–4·935), hydroxychloroquine with a macrolide (8·1%; 5·106, 4·106–5·983), chloroquine (4·3%; 1·717–10·439), and chloroquine with a macrolide (6·5%; 4·011, 3·344–4·812) were independently associated with an increased risk of de-novo ventricular arrhythmia during hospitalisation.

**Interpretation** We were unable to confirm a benefit of hydroxychloroquine or chloroquine, when used alone or with a macrolide, on in-hospital outcomes for COVID-19. Each of these drug regimens was associated with decreased in-hospital mortality and increased frequency of ventricular arrhythmias when used for treatment of COVID-19.

**Funding** William Grey Distinguished Chair in Advanced Cardiovascular Medicine at Brigham and Women's Hospital.

**Copyright** © 2020 Elsevier Ltd. All rights reserved.



Published Online  
22, 2020  
[https://doi.org/10.1016/S0140-6736\(20\)31180-6](https://doi.org/10.1016/S0140-6736(20)31180-6)  
This online publication has been corrected. The corrected version first appeared at thelancet.com on May 29, 2020.  
See Online/Comment  
[https://doi.org/10.1016/S0140-6736\(20\)31180-6](https://doi.org/10.1016/S0140-6736(20)31180-6)  
Brigham and Women's Hospital  
Heart and Vascular Center and  
Harvard Medical School,  
Boston, MA, USA  
(Prof M R Mehra MD);  
Sunigpharm Corporation,  
Chicago, IL, USA (S S Desai MD);  
University Heart Center,  
University Hospital Zurich,  
Zurich, Switzerland  
(Prof F Ruschitzka MD);  
Department of Biomedical  
Engineering, University  
of Utah, Salt Lake City, UT, USA  
(A N Patel MD); and ICA  
Research Institute, Nashville,  
TN, USA (A N Patel)  
Correspondence to:  
Prof Mandeep R Mehra, Brigham  
and Women's Hospital Heart and  
Vascular Center and Harvard  
Medical School, Boston,  
MA 02115, USA  
mmehra@bwh.harvard.edu

Par dérogation à l'article L. 5121-8 du code de la santé publique, l'hydroxychloroquine et l'association lopinavir/ritonavir peuvent être prescrits, dispensés et administrés sous la responsabilité d'un médecin aux patients atteints par le covid-19, dans les établissements de santé qui les prennent en charge, ainsi que, pour la poursuite de leur traitement si leur état le permet et sur autorisation du prescripteur initial, à domicile. Ces prescriptions interviennent, après décision collégiale, dans le respect des recommandations du Haut Conseil de la santé publique et, en particulier, de l'indication pour les patients atteints de pneumonie oxygène-requérante ou d'une défaillance d'organe.

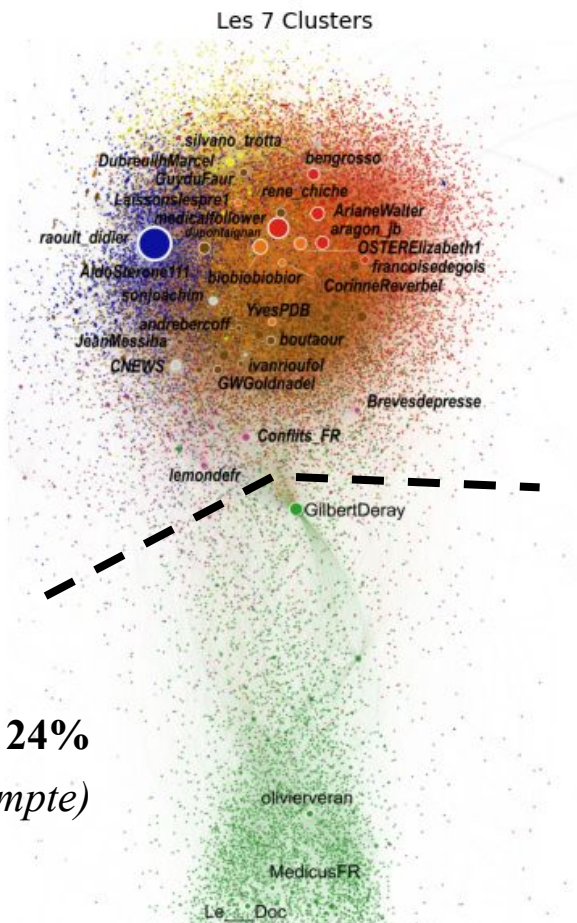


## Réseau de Retweets pendant la controverse

# 1. Les différentes communautés qui ont relayé les échanges

**Opposants:** 68%  
(nombre de compte)

**Partisans:** 24%  
(nombre de compte)



RT de @D.Raoult (Critique académique)

RT de @olivierveran (Interdiction Hcq)

RT de @GilbertDeray (RT par tous les clusters)

# Objectif

- Se saisir d'une problématique de sciences sociales
  - Controverse et polémique publique autour de la chloroquine
- Aborder des données réelles
  - Et moyennement structurées
- Utiliser différentes briques de traitement de données & différentes stratégies :
  - Analyse temporelle
  - Analyse relationnelle
  - Analyse textuelle
- Développer de nouvelles compétences





# Déroulement de la séance

1. Mise en forme de données Twitter
2. Analyse de séries temporelles
3. Analyse de réseaux sociaux
4. Contenu des tweets

Période de travail seul ou à deux

Point de restitution collective pour avancer ensemble

Tous les notebooks ici : <https://github.com/emilienschultz/ssc-twitter>

(et les slides à la fin de la séance)

# Travail de validation de la séance

Retravailler une des visualisations initiés dans la séance avec les consignes suivantes :

- Problématisée : elle doit avoir un message clair
- Finalisée : elle doit être pensée pour une diffusion
- Complexifiée : essayer d'avoir au moins 2 niveaux d'information dessus

Cela peut être : soit une distribution, soit une courbe, soit un réseau

- envoyé à [emilien.schultz@univ-amu.fr](mailto:emilien.schultz@univ-amu.fr) pour le 30 novembre avec l'objet SSC