HAL
open science

Comment le hasard se fait scientifique

Leila Schneps

To cite this version:

Leila Schneps. Comment le hasard se fait scientifique. Pourquoi Moi?, Jul 2021, Paris, France. hal-02994296

HAL Id: hal-02994296

https://hal.science/hal-02994296v1

Submitted on 7 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Comment le hasard se fait scientifique

Leila Schneps

Selon un malentendu très répandu science et hasard se situent en opposition, quasiment contraire l'une de l'autre. C'est ce même malentendu qui réduit la science à un ensemble de faits avérés, démontrés, bien compris. Devant une question posée, une réponse "scientifique" sera perçue comme l'opposé d'une une réponse aventurée, "au hasard". Un événement survenu par hasard sera tout le contraire d'un événement qui est l'effet inévitable d'une cause répertoriée, connue. Cette vision de la science et du hasard comme des opposés domine le discours moderne, reflétant sans doute dans le langage de notre époque l'opposition fondamentale entre ordre et désordre.

Malheureusement, ce malentendu premier, interprété et mis en œuvre dans des situations réelles, peut mener à des conséquences concrètes souvent tragiques. Dans les cours de justice, les experts qui énoncent des résultats probabilistes absolument corrects de leurs analyses convainquent infiniment moins les jurés que ceux qui se déclarent absolument certains. En effet, on a souvent du mal à admettre qu'un résultat scientifique, particulièrement dans le cadre judiciaire, puisse s'exprimer très souvent comme un résultat probabiliste, apportant le renseignement que tel événement est plus probable que tel autre, plutôt qu'une certitude. Le calcul des probabilités formant une branche à part entière des mathématiques, il est malheureux d'entendre l'expression "C'est mathématique" utilisée, dans le langage quotidien, comme synonyme de "C'est un fait établi". Il faut comprendre qu'un énoncé probabiliste est lui-même "mathématique", qu'un calcul de probabilité est un résultat mathématique aussi valable que n'importe quel autre, et que la science ne se borne pas à établir des faits de manière absolue.

Il arrive très souvent que la publication des résultats d'une étude observationnelle déclenche une controverse énorme, pendant laquelle on aperçoit très clairement que la compréhension de ce que c'est qu'un résultat probabiliste, et l'idée même qu'un résultat scientifique peut être purement probabiliste, manque non seulement aux lecteurs mais souvent même aux auteurs de l'étude en question. Les erreurs d'interprétation d'un ensemble de données sont nombreuses et variées : erreurs logiques (p. ex. confondre "ce fait n'apporte aucun élément de confirmation de telle hypothèse" avec "ce fait montre que la réalité de telle hypothèse n'est pas probable"), erreurs de transposition (confondre "il y a des erreurs dans l'analyse des données" avec "les résultats sont faux"), erreurs de statistiques (baser des conclusions sur un pourcentage sans tenir compte de la taille de la population, oublier les populations non observées, etc.) Pire encore, quand l'étude porte sur un sujet d'intérêt général, présentant éventuellement un caractère d'urgence, il arrive que la controverse se développe jusqu'au point de devenir une querelle publique avec insultes, mépris et opinions politiques à la clé, dans laquelle tous les côtés brandissent à qui mieux mieux la bannière de la "science", sans que la vraie voix de la science ne soit jamais entendue.

Confronté à une étude statistique, que nous dit en réalité "la véritable voix de la science" ? Il faut comprendre avant tout qu'en posant cette question, nous ne parlons pas d'autre chose que de hasard : la question toute entière par rapport à n'importe quelle étude est de savoir si les observations proviennent d'une cause particulière, ou s'ils sont purement un effet du hasard - ou une combinaison des deux. Du moment où les observations collectées sont considérées comme fiables (et ce n'est pas toujours le cas ; voir la récente débâcle survenue au journal médical britannique très réputé, *The Lancet*), la seule

question reste de savoir si l'on est en droit de conclure que les observations proviennent d'une cause particulière ou non, et si oui avec quelle probabilité. C'est-à-dire que la science sera en mesure de déterminer une réponse fiable à la question mais, très souvent, cette réponse ne sera pas une réponse définitive à la question de savoir si oui ou non telle cause hypothétique produit les résultats observés. C'est précisément cette idée qui est le plus souvent mal comprise.

Nous nous proposons d'examiner ici un cas particulier, récent et frappant: celui de la parution du premier article de l'équipe du Dr Didier Raoult de l'IHU de Marseille au sujet des résultats d'un traitement à base d'hydroxychloroquine (HCQ) sur des patients atteints de Covid-19. Cet article a été mis en ligne le 17 mars, au moment où la France commençait tout juste à comprendre la gravité de l'épidémie.

Un mois plus tard, le tollé déclenché par ce travail avait traversé les mers, divisant le public et les médecins et menant des chefs d'état à prendre des décisions ayant des conséquences gravissimes sur la santé de leurs citoyens. Bien sûr, une étude importante peut et doit avoir de grandes conséquences mais, en l'occurrence, toute la discussion s'orientait autour du fait que certains croyaient les conclusions de l'article justes, une opinion fondée sur la carrière du Dr Raoult et les années d'expérience de son équipe et, par la suite, sur des études supplémentaires et l'observation générale d'autres populations chez lesquelles l'hydroxychloroquine a été massivement préscrite. D'autres au contraire considéraient ces mêmes conclusions comme fausses ; cette opinion se trouvait étayée par la présence de plusieurs erreurs d'analyse statistique dans le texte et, surtout par la suite, par de nombreuses autres études.

L'article présentait les résultats d'un traitement à base d'hydroxychloroquine (HCQ), combinée dans certains cas avec l'antibiotique azithromycine (AZT), et annonçait des conclusions incroyablement positives : 6 patients sur 6 guéris au bout de 6 jours parmi ceux traités avec HCQ et AZT et 8 sur 14 parmi ceux traités à l'HCQ seul, comparé à 2 sur 16 guéris parmi ceux traités par les méthodes dites "standards". Au tout début, ces conclusions ont été accueillies avec un enthousiasme dû en partie à l'espoir de chacun qu'un traitement efficace contre le Covid-19 soit rapidement trouvé, en partie à la réputation de l'institut hospitalo-universitaire de Marseille et de son directeur le Dr Raoult, personnage charismatique qui, de quasiment inconnu qu'il était en janvier 2020, est devenu rapidement une vedette et le centre d'une controverse mondiale.

Toute personne connaissant un tant soit peu les statistiques est capable de constater qu'il y a bien des erreurs de procédure et d'analyse dans l'article. Les auteurs ont exclu de l'étude certains patients présents au début mais qui ont changé de traitement en cours de route, soit parce qu'ils ont été transportés en réanimation, soit parce qu'ils sont partis de leur propre gré (rebutés par les effets secondaires ou se sentant déjà guéris). Les conditions n'étaient pas identiques pour tous les patients ; non seulement deux traitements différents étaient proposés (avec ou sans AZT), mais le groupe contrôle consistait en des patients d'un autre hôpital (Nice), puisque l'équipe du Dr Raoult a traité tous les patients qui se présentaient. Les auteurs ont considéré qu'un patient était guéri s'il testait négatif au sixème jour, alors qu'il arrive assez souvent qu'un test soit négatif un jour et positif le lendemain à cause de variations dans la fiabilité du test et dans la quantité

de virus présent. Enfin les données publiées dans l'article étaient incomplètes[1].

Les journalistes ont veillé à ce que la liste des erreurs d'analyse dans l'article fasse rapidement le tour du monde. Les réfutations furent immédiates et nombreuses. En voici un exemple typique, par le chercheur néerlandais Frits Rosendaal:

This study suffers from major methodological shortcomings which make it nearly if not completely uninformative. Hence, the tone of the report, in presenting this as evidence of an effect of hydroxychloroquine and even recommending its use, is not only unfounded, but, given the desperate demand for a treatment for Covid-19, coupled with the potentially serious side-effects of hydroxychloroquine, fully irresponsible[2].

Personne n'a mis en doute les données elle-mêmes ; les premières réfutations concernaient une méthodologie défectueuse, trop hâtive, peu professionnelle. Très vite s'y ajoutèrent d'autres réfutations sous formes de nouvelles études, dont certaines sur des échantillons encore plus petits que celui de l'équipe de Marseille, et d'autres rétrospectives, c'est-à-dire menées sur des patients déjà traités sans qu'une étude ait été organisée au préalable. De l'expérience de l'équipe Toulouse qui administrèrent le traitement de Raoult à 11 patients hospitalisés avec des comorbidités sérieuses[3] à l'étude rétrospective sur des centaines de vétérans hospitalisés aux États-Unis[4] ou à la vaste étude internationale du *Lancet*[5], les résultats semblent infirmer les conclusions de l'équipe du Dr Raoult. C'est une vague médiatique anti-hydroxychloroquine qui déferla alors dans les médias, nourrie par une contre-réaction au soutien enthousiaste du président américain manifestement incompétent en la matière. Le "pour-ou-contre Raoult" devint un sujet de conversation quotidienne, divisant amis, familles et même médecins, entre ceux qui croient au « gourou » (ou des médecins ayant constaté que le traitement de Raoult produisait parfois des effets positifs, sans pour autant faire des études contrôlées) et ceux qui rejettent violemment ses conclusions. Malgré l'intérêt évident du président français pour les travaux du Dr Raoult, la gouvernement interdit l'utilisation de l'hydroxychloroquine par les médecins, d'abord en ville (dès le mois de mars), et ensuite au sein même des hôpitaux (27 mai).

Mais dans toute cette clameur, et malgré une référence incessante à "la science" comme l'autorité absolue, comme juge sévère et intransigeante condamnant le Dr Raoult sans appel, la véritable voix de la science n'a jamais été entendue, car elle ne dit rien de ce qui précède. En réalité, la science ne s'intéresse qu'à une seule et unique question: *Quelle est la probabilité que les observations de l'équipe du Dr Raoult soient simplement l'effet du hasard ?*

---

[1] C'est d'ailleurs le cas de toutes les publications médicales, ce qui est la source de très nombreuses erreurs et rétractations. Il est évident que les données brutes de toute étude publiée devraient être rendues publiques et disponibles pour des analyses statistiques indépendantes. S'il y a un obstacle à ce que la médecine soit considérée réellement comme une science, d'ailleurs, c'est bien ce manque de transparence.

[2] International Journal of Antimicrobial Agents, Volume 56 Issue 1, July 2020 106063.

[3] Molina et al., Médecine et Maladies Infectieuses, Volume 50 Issue 4, June 2020, p. 384.

[4] Magagnoli et al., Med (2020), https://doi.org/10.1016/j.medj.2020.06.001

[5] Mehra et al., Lancet (2020), 10.1016/S0140-6736(20)31180-6

Il existe une méthode statistique assez simple, le test exact de Fisher, qui permet de calculer la probabilité que les résultats observés soient l'effet du pur hasard, ou plus précisément s'obtiennent sous l'hypothèse "nulle", qui affirme dans le cas présent que le traitement proposé par l'équipe de Marseille n'a aucun effet curatif. C'est la méthode qui a été utilisée dans l'article, mais en excluant les patients qui ont quitté le traitement en cours de route, erreur qui fausse les résultats et qui a été largement relevée par les critiques de l'article. Toutefois, même après rectification de cette erreur, cette méthode indique que la probabilité que le résultat soit purement l'effet du hasard est infime[6].

Malheureusement, cette méthode, même corrigée, reste beaucoup trop rudimentaire pour fonctionner correctement dans une situation dans laquelle il existe de très nombreux facteurs supplémentaires (appelés covariables) qui peuvent avoir un impact significatif sur les résultats. Comme il avait déjà été largement constaté, les covariables ayant le plus d'impact sur les patients malades du Covid-19 sont d'abord l'âge du patient, et ensuite la présence d'autres maladies éventuelles (comorbidités) comme le cancer ou le diabète, qui peuvent fragiliser le système immunitaire. De plus il faut prendre en compte le stade de la maladie au début du traitement, c'est-à-dire d'une part le temps pendant lequel les patients ont souffert des symptomes du Covid-19 avant de se présenter à l'hôpital, et la gravité de leurs symptomes, qui peuvent varier d'une simple toux avec ou sans fièvre à des lésions graves des poumons. Toutes ces données étaient bien entendu présentes dans les dossiers médicaux des patients, mais pas dans l'article, ce qui rendait impossible d'en tirer une conclusion même probabiliste qui soit fiable, car il est très possible, surtout avec un échantillon aussi petit, que la différence entre deux résultats observés soit due à une différence dans les covariables entre deux cohortes.

En tant que membre du public, simple observatrice (ayant toutefois l'habitude de statistiques médicales), je pensais que le tollé autour du Dr Raoult mènerait très rapidement à une ré-analyse correcte et complète basée sur l'ensemble complet des données originales. Mais les semaines passaient et il n'en était rien. En fin de compte j'ai proposé à un collègue statisticien, le Prof. Richard Gill, d'examiner nous-mêmes l'article. Devant le manque de données concernant les covariables, nous n'avons eu d'autre choix que d'écrire directement à l'équipe de Marseille pour leur demander s'ils auraient l'amabilité de nous les envoyer. Étant donné la tornade mondiale au centre de laquelle l'équipe se trouvait alors, nous n'étions pas franchement sûrs qu'ils nous répondraient, mais il faut croire qu'ils souhaitaient autant que nous que quelqu'un effectue une analyse statistique rigoureuse, qui pourrait justifier les conclusions dont ils étaient d'autant plus certains qu'ils avaient traités des centaines de patients entretemps. Ils ont accepté de nous envoyer un dossier complet contenant tous les renseignements demandés pour chacun des patients de l'étude originale (évidemment anonymisés).

Il existe plusieurs méthodes d'analyse statistique (regression logistique, méthode du lasso, bootstrap, etc.) permettant de combiner ce type de base de données, avec un échantillon assez réduit et de nombreuses covariables. Le but de ces méthodes est de mesurer l'impact de chacune des covariables (y compris le traitement) sur les résultats observés, et de classer ces covariables selon l'ordre de grandeur de leur impact. Travaillant avec mon collègue le Prof. Gill et Hoang Van Thuan, le jeune médecin épidémiologue à qui était due la partie statistique de l'article original de l'équipe marseillaise, nous avons appliqué tour à tour toutes ses méthodes à l'ensemble des données. Le résultat fut assez spectaculaire : quelle que soit la méthode utilisée, le traitement produisait un impact sur le résultat plus

---

[6] Hulme et al., MedRXiv, doi: https://doi.org/10.1101/2020.03.31.20048777

important, et de loin, que n'importe quelle autre covariable. En fait ces méthodes ont indiqué que :

* le traitement HCQ+AZT avait nettement plus d'impact que l'HCQ seule ;
* l'impact du traitement (avec ou sans AZT) était assez fort dans le premier stade de la maladie, mais se réduisait à presque rien dans un stade avancé ;
* la covariable qui avait le plus d'impact après le traitement était l'âge.

Mathématiquement, il s'agit ici d'une *indication* très forte de l'utilité du traitement, et non d'une *démonstration*. En effet nos résultats, qui présente le maximum d'informations que l'on peut tirer d'un échantillon aussi petit, ne peuvent remplacer la nécessité, à terme, d'une étude à grande échelle, contrôlée et randomisée, selon les meilleures pratiques médicales. Mais si notre calcul ne démontre pas au delà de tout doute possible l'efficacité du traitement, il démontre certainement la nécessité et l'intérêt de poursuivre les recherches et d'organiser au plus vite une telle étude. Or, cela n'a jamais été fait. Il suffit d'éplucher les nombreuses études publiées sur l'hydroxychloroquine pour constater qu'il n'y en a pas une seule qui reproduise les conditions de succès indiquées par l'analyse ci-dessus. Dans toutes les études publiées (dont beaucoup ont été proclamées dans la presse comme "la preuve définitive que l'hydroxychloroquine n'agit pas"), soit l'hydroxychloroquine est donnée sans azithromycine, soit à des doses exagérées (causant, dans certains cas, des problèmes cardiaques suffisamment grave pour avoir à interrompre l'étude) soit, et c'est de loin le cas le plus fréquent, le traitement est administré à un stade beaucoup trop avancé de la maladie, à des patients déjà hospitalisés et souvent dans un état quasi-désespéré.

L'organisation d'une étude à grande échelle du traitement HCQ+AZT dans les premiers stade de l'infection a été empêchée, en fin de compte, par deux malentendus scientifiques étonnamment répandus. Tout d'abord, il a été supposé que des erreurs d'analyse dans un article rendent les résultats faux. Or ce n'est nullement le cas - si l'analyse contient des erreurs, tout ce que l'on peut dire est que les résultats ne sont pas fiables, et l'analyse des données doit être reprise depuis le début au moyen des données brutes. D'autre part les différences de conditions dans les études publiées n'ont jamais été prises en compte : ainsi du moment où le traitement s'était révélé inefficace à un stade avancé de la maladie, on a considéré que cela démontrait son inefficacité tout court. Il s'agit là d'erreurs scientifiques majeures, dont il est possible qu'elles aient contribué à la mort de milliers de personnes supplémentaires, ce qui prête à réfléchir sur leur importance et leurs conséquences qui peuvent aller jusqu'au tragique.