

THE LANCET

LES TROMPERIES DE LA RENOMMÉE

Par Ariel Beresniak

Patatras, quelques heures avant d'envoyer Causeur à l'imprimerie, nous avons appris que *The Lancet* prenait ses distances avec l'étude qu'elle avait publiée sur la dangerosité de la chloroquine. Il faut dire que malgré sa haute réputation, cette revue scientifique a déjà publié des études fausses. Reste à comprendre pourquoi l'OMS et notre ministère de la Santé lui ont si vite emboîté le pas.

« **N**ous n'avons pas été capables de confirmer l'efficacité de l'hydroxychloroquine ou de la chloroquine, utilisée seule ou avec un macrolide. Chacune de ces stratégies thérapeutiques utilisées dans le Covid-19 est associée avec une baisse de la survie à l'hôpital et une augmentation des troubles du rythme cardiaque. » Cette phrase, qui conclut l'étude publiée le 22 mai par Mandeep R. Mehra, cardiologue à Boston et professeur de médecine à Harvard, dans la prestigieuse revue *The Lancet*, n'est pas la plus simple ni la plus élégante qui ait jamais été publiée, mais elle a eu l'effet d'une bombe. En résumé, elle signifie : le professeur Raoult a tort. Définitivement. Et avec lui, bien sûr, tous ceux qui, de Donald Trump à

Michel Onfray, l'ont érigé en héros du peuple, chevalier du bon sens en croisade contre l'establishment et les « élites ».

L'étude, selon laquelle le traitement à l'hydroxychloroquine associée à un antibiotique accroît le risque de mortalité pour les malades du coronavirus, ne fait pas que conforter le vaste camp des anti-Raoult, elle entraîne la conviction de ceux qui n'ont pas – ou ne sont pas supposés avoir – d'a priori idéologiques ou politiques. Or, moins de quinze jours après avoir publié cette étude, la revue *The Lancet* a elle-même mis en garde ses lecteurs – « *d'importantes questions scientifiques ont été soulevées concernant les données rapportées dans l'article* » – quant à la validité de ses conclusions. Ce ne sont pas les doutes exprimés le 2 juin et le début de revirement du *Lancet* qui laissent songeur mais plutôt la vitesse avec laquelle la publication du 22 mai a recueilli l'approbation de l'OMS et du ministère français de la Santé. Ce dernier a quasi immédiatement adopté des mesures en accord avec les conclusions de l'article. Pourquoi une telle précipitation des autorités sanitaires à interdire la prescription de l'hydroxychloroquine sans attendre les audits et éclaircissements méthodologiques demandés ? On peut expliquer ce pouvoir de persuasion par deux facteurs : le nombre de sujets étudiés (96 032 patients !) et la réputation du journal qui l'a publiée.

Pour la première fois, une analyse reposant sur les « big data », l'exploitation d'un nombre considérable de données informatiques, et non sur un véritable essai clinique, aboutit en quelques heures à l'interdiction de prescription d'une molécule présente sur le marché depuis 1949 et à l'arrêt des essais cliniques de cette molécule dans le Covid-19. Ce qui conduit à s'interroger à la fois sur l'effet de sidération produit par les grands chiffres et sur l'effet d'intimidation, exercé tant par ces mêmes grands chiffres que par le prestige de la revue. On croit spontanément que plus l'échantillon d'une étude est vaste, plus il est représentatif, donc, plus l'étude →

Docteur en médecine, Ariel Beresniak est spécialiste en santé publique et médecine sociale et docteur en mathématiques appliquées, habilité à diriger des recherches. Auteur d'ouvrages de référence tels que : *Dictionnaire raisonné des termes des entreprises du médicament* (Flammarion), *Dictionnaire commenté d'économie de la santé* (Masson).



Un homme tient un portrait de Didier Raoult lors des applaudissements pour remercier le personnel soignant, Saint-Mandé, 11 mai 2020.

est sérieuse. Cela est loin d'être vrai. En 1936, le magazine américain *The Literary Digest* procéda à un vaste sondage sur l'élection du président des États-Unis. Il interrogea par courrier 10 millions d'électeurs issus de ses listes d'abonnés, des annuaires téléphoniques et des listes de propriétaires d'automobiles. L'analyse de 2,3 millions de réponses permit de prédire l'élection d'Alfred M. Landon avec 55 % des voix. Franklin D. Roosevelt fut pourtant élu président avec 61 % des voix ! L'erreur était d'avoir utilisé en toute confiance une base de données très volumineuse, mais non représentative de la population des électeurs : la probabilité qu'un citoyen américain pauvre ne disposant ni d'un abonnement, ni d'un téléphone, ni d'une automobile figure dans l'échantillon était quasi nulle. Ce qui a abouti à l'erreur de sondage et incidemment à la faillite du magazine...

Un échantillon de grande taille n'est donc en aucune façon une assurance de qualité. On peut même avancer que, plus le nombre de sujets est important, plus il est difficile de vérifier la qualité des données et donc de garantir la fiabilité de l'étude. En 1926, quelques années avant l'élection de Roosevelt, le statisticien et économiste anglais Arthur Lyon Bowley avait établi le premier principe garantissant la représentativité d'un échantillon, celui des tirages aléatoires : tout sujet de la population de référence doit avoir une probabilité non nulle d'être sélectionné dans l'échantillon. En conséquence, un échantillon d'un million de personnes non tirées au sort ne sera pas représentatif de la population d'ensemble, alors qu'un échantillon de 1 000 personnes tirées au sort le sera. Cette démarche fondée sur les tirages aléatoires est suivie par les autorités locales corses pour estimer la séroprévalence au coronavirus dans la population insulaire. Curieusement, elle n'a pas été employée à l'échelle nationale ou européenne.

Les 96 032 patients provenant de 671 hôpitaux sur les six continents analysés dans l'étude de Mehra publiée par *The Lancet* n'ont pas été tirés au sort parmi les patients de tous les hôpitaux du monde, ni parmi l'ensemble des patients hospitalisés pour Covid-19. Et pour cause : seuls des hôpitaux disposant d'une plate-forme de dossiers électroniques interopérables, donc compatibles, ont pu être intégrés dans l'analyse, alors que de très nombreux établissements ne sont pas équipés pour le traitement informatique de leurs dossiers patients, ou équipés d'un système maison. Quand bien même les données seraient de bonne qualité (ce qui est peu probable, car il est impossible d'auditer les données de 671 hôpitaux en période pandémique), l'étude de Mehra n'est pas représentative.

Le deuxième problème tient à la méthode d'analyse des données collectées. La collecte d'informations sur la santé auprès des patients est probablement l'un des actes médicaux les plus anciens. Dans la Grèce antique, le Corpus hippocratique était un recueil de dossiers médicaux et l'une des premières tentatives de classement des maladies en fonction des symptômes et des observations. Avec la médecine moderne, on a créé des dossiers pour conserver des informations sur les patients, les investigations effectuées et les traitements en cours. Tandis que les soins médicaux deviennent plus complexes et personnalisés, les progrès des technologies numériques permettent le stockage et l'exploitation statistique de ces données. On peut par exemple connaître rapidement le nombre de sujets hospitalisés pour une pathologie donnée, le nombre d'exams médicaux par type de maladie, les profils de patients par type de consommation médicale, etc. En revanche, l'analyse rétrospective des bases de données hospitalières ne peut en aucune manière servir à comparer l'efficacité et la tolérance des traitements comme l'a fait Mehra.



Emmanuel Macron rencontre Didier Raoult à l'Institut hospitalo-universitaire Méditerranée infection, à Marseille, 9 avril 2020.

On illustrera ce propos par un exemple réel. Chez les femmes souffrant d'infertilité, la production d'ovules par les ovaires peut être stimulée par l'hormone FSH que l'on obtient soit par purification d'urines de femme, soit par synthèse en biotechnologie. Les essais cliniques comparant les deux formes ont toujours conclu à un léger avantage pour la FSH de biotechnologie en termes de taux de grossesse, probablement du fait que le produit d'origine naturelle n'est pas aussi pur que son cousin artificiel. L'Allemagne a rassemblé dans une base de données l'intégralité des dossiers médicaux électroniques des femmes auxquelles étaient prescrits ces produits. Et l'exploitation de ces données (plusieurs dizaines de milliers de dossiers médicaux électroniques) a conclu à une meilleure efficacité de la FSH d'origine urinaire, contredisant tous les essais cliniques réalisés dans le monde. Des investigations approfondies ont permis de comprendre cette discordance : la FSH de biotechnologie, plus chère, était plutôt achetée par des femmes plus aisées, qui étaient en moyenne plus âgées que celles qui ont acheté la FSH urinaire. En l'occurrence, c'est donc l'âge (en liaison avec le niveau de vie) qui explique la différence d'efficacité exprimée en taux de grossesse. C'est en raison de ce genre de « biais de sélection » que les big data ne peuvent pas être utilisées pour comparer des traitements. Les multiples ajustements et redressements statistiques effectués par Mehra dans son étude du *Lancet* ne sont pas du tout une garantie de qualité. Bien au contraire, ce sont des pondérations arbitraires qui ne reflètent que les préjugés de leurs auteurs. Il est donc curieux que les spécialistes chevronnés de l'Agence nationale de sécurité du médicament (ANSM) et les membres du Haut Conseil de la santé publique (HCSP) se soient laissés si facilement abuser par cette étude. La haute réputation du *Lancet* explique largement le succès médiatique de cette étude. Fondé en 1823, le journal britannique est l'un des plus anciens et plus prestigieux médias médicaux. Son facteur d'impact (IF) est de 59, ce qui signifie que chaque article est cité en moyenne 59 fois par d'autres chercheurs. Publier dans *The Lancet* confère une grande visibilité et de meilleures chances de progression de carrière. Cependant, une revue médicale, aussi prestigieuse soit elle, n'est qu'un éditeur et non pas un organisme de recherche. Elle ne dispose ainsi d'aucune ressource pour vérifier l'honnêteté des articles. Le seul contrôle est constitué par deux ou trois relecteurs extérieurs bénévoles, choisis par la revue, qui émettent leurs remarques sur la base du seul manuscrit.

Il est donc arrivé que des études publiées dans de grandes revues scientifiques se soient révélées complètement fausses, voire frauduleuses. L'exemple le plus connu est l'étude publiée par *The Lancet* qui remettait en cause l'innocuité du vaccin rougeole-oreillons-rubéole (ROR). L'étude portait sur 12 enfants et concluait que le vaccin ROR était responsable d'un syndrome associant autisme et troubles digestifs. Au cours d'une conférence de presse organisée dans un hôpital londonien, le premier auteur, Wakefield, recommanda l'utilisation de trois vaccins

séparés à la place du vaccin combiné ROR qu'il présentait comme dangereux. Cet article fut immédiatement relayé par la presse britannique et internationale. La même année, beaucoup de parents refusèrent de faire vacciner leurs enfants par ce vaccin, entraînant une résurgence des cas de rougeole et de la mortalité par rougeole en Grande-Bretagne. Une enquête menée par le journaliste Brian Deer a mis au jour des tromperies : les enfants de l'étude ne présentaient pas de lésions digestives comme relatées dans l'article et leur inclusion dans l'étude résultait de la volonté de parents d'enfants autistes de poursuivre le fabricant du vaccin, et non de la constatation de signes d'autisme après la vaccination. Après quelques résistances de principe, *The Lancet* fut contraint de « retirer » l'article (en rajoutant un simple bandeau « Retracted »). Cependant, le mal était fait : de nombreux parents ont conservé une méfiance vis-à-vis du vaccin.

C'est également *The Lancet* qui publie régulièrement les résultats de l'Étude internationale sur l'impact des maladies (« Global Burden of Disease Study ») : initié par le professeur Murray de l'université de Washington, financé par la fondation de Bill et Melinda Gates, ce programme vise à comparer l'importance des maladies dans le monde grâce à un indicateur inventé par Murray lui-même en 1994, le DALY (Disability Adjusted Life Years) ou « Années de vie corrigées par l'incapacité ». Ainsi, un pays A est jugé globalement en meilleure santé qu'un pays B si le nombre d'années de vie sans handicap perdues sur une période donnée dans le pays A est inférieur à celui du pays B. Pour construire son indicateur, Murray choisit arbitrairement deux composantes qu'il se garde bien de définir : une durée de vie considérée comme « optimale » et le nombre d'années de vie avec handicap qu'on accepterait de sacrifier pour être en parfaite santé. Ses conclusions, régulièrement publiées par *The Lancet*, reposent sur de nombreuses suppositions ainsi que sur des données peu fiables dans la majorité des pays étudiés. En 2016, l'une des nombreuses études publiées par Murray dans *The Lancet* a prétendument établi que ce sont les maladies mentales qui ont le plus grand impact à l'échelle mondiale. Celles-ci étant rarement mortelles en dehors des suicides, cette étude occulte totalement les dix principales causes de décès dans le monde dont les maladies cardiovasculaires, les maladies respiratoires, le diabète, les maladies diarrhéiques, etc. L'OMS a fait auditer la méthodologie des DALY par un groupe d'experts indépendants (DALY Review Group), dont les conclusions ont été accablantes : technique opaque réduisant des informations complexes à une valeur numérique unique dans le seul but de distraire l'attention des vrais problèmes. Le nom du *Lancet* fonctionne comme un label, une garantie de sérieux scientifique. Il couvre en l'occurrence des idéologies et des pseudo-standards. Tout comme l'étiquette de la bouteille ne suffit pas à faire le bon vin, un nom, même aussi ancien et prestigieux que celui du *Lancet*, ne doit pas amadouer l'esprit critique nécessaire à toute démarche scientifique, même et surtout en période de crise. •