

4

Quand les statistiques sauvent des vies

L'épidémiologie humaine au service de la médecine

Myriam Alexander (2008), Epidémiologiste post-doctorante, Roche Pharmaceuticals,
Camille Charbonnier (2009), Assistante de Recherche au CHU de Rouen

Le domaine de la santé est un vaste domaine qui inclut un très grand nombre de professionnels. Il s'étend du chercheur-biologiste qui cherche à comprendre les mécanismes à l'œuvre dans le développement d'une maladie, au pharmacien industriel qui se concentre sur le développement de médicaments, au médecin qui administre le traitement, à l'administrateur en charge des politiques de prévention et protection de la santé, etc.

Quel peut être l'apport du statisticien dans le domaine de la santé ?

Il est multiple et intervient à de nombreux niveaux. Dénommé épidémiologiste, ou encore biostatisticien, un tel statisticien a pour priorité de promouvoir une médecine basée sur la preuve scientifique, plutôt que sur la tradition, la plausibilité biologique ou encore la superstition. La distinction entre épidémiologie et bio-statistique est parfois ténue. Tandis que le biostatisticien est souvent compris comme celui qui développe les modèles adaptés à l'analyse des bases de données biologiques ou médicales, l'épidémiologiste est celui qui applique ces modèles à des cas concrets et produit des statistiques interprétables. Nous nous restreindrons ici au terme épidémiologiste par simplicité.

Nous allons illustrer dans cet article comment l'épidémiologiste est devenu un compagnon incontournable de la recherche médicale et pharmaceutique. Son apport sur le plan médical se fait en proposant de nouvelles hypothèses biologiques, en particulier génétiques, et en testant la validité de ces hypothèses. Son apport dans le domaine pharmaceutique est d'accompagner le développement d'un médicament en quantifiant les différents risques médicaux encourus par les populations de cas et de témoins sur lesquelles sont testées les nouvelles molécules. Par souci de concision, nous n'aborderons pas ici le rôle de l'épidémiologiste qui travaille en santé publique

et évalue la qualité des politiques de santé et les structures médicales et hospitalières. Son rôle est crucial pour la bonne mise en œuvre de politiques de prévention et le bon fonctionnement du système de santé, et il se trouve souvent en interaction avec l'économiste de santé publique.

L'épidémiologie pour comprendre une maladie

La première étude épidémiologique moderne a permis de comprendre les causes d'une maladie qui décimait les populations européennes au XIX^{ème} siècle, le choléra. Dr John Snow, plutôt que de proposer une nouvelle théorie expliquant le processus biologique à l'œuvre dans la propagation du choléra, entreprit de noter systématiquement la localisation géographique des cas de choléra dans la ville de Londres [1]. Il se rendit ainsi compte de l'existence d'une concentration anormalement élevée de cas dans deux quartiers et en déduisit que l'approvisionnement de ces quartiers à une seule pompe à eau était la source de contagion. John Snow est considéré comme un précurseur, qui sut déduire d'observations statistiques le mécanisme de transmission d'une maladie et ainsi permettre d'agir rapidement pour stopper l'épidémie. De nos jours, l'épidémiologie de surveillance utilise des techniques statistiques variées pour identifier et prédire le développement d'épidémies. Le mot « épidémie » est ici compris au sens large, que la maladie soit transmissible (par exemples maladies virales comme le SIDA ou H1N1) ou non (diabète, maladie d'Alzheimer, maladies cardiovasculaires), mais dont le nombre de cas est en augmentation rapide dans une population donnée. Les techniques dites « de *cluster* » établissent la présence d'une concentration statistiquement élevée de malades dans une région donnée. Les modèles de transmission, markoviens ou autres, reconstituent la chaîne de propagation de la maladie. Les données utilisées sont régionales, nationales, et de plus en plus internationales,

depuis la mise en place d'organisations telles que l'Organisation Mondiale de la Santé.

Rechercher les causes d'une maladie pour pouvoir agir préventivement, sans forcément chercher à comprendre comment ces causes produisent la maladie, c'est ce que permet l'épidémiologie. Les exemples sont nombreux auxquels s'applique la stratégie « mieux vaut prévenir que guérir », en voici deux bien connus.

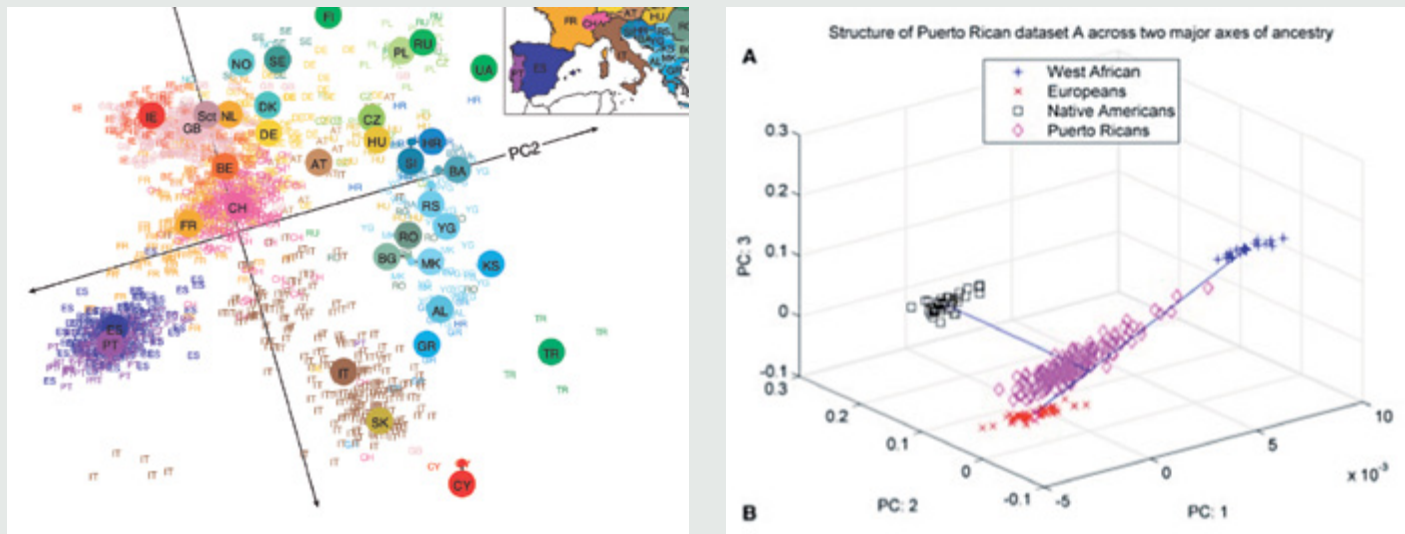
Dans les années 30, les chercheurs étaient partagés sur les causes de l'augmentation du nombre de cancers du poumon. S'agissait-il de la pollution des villes industrielles ? Du tabac ? Deux chercheurs, Richard Doll et Austin Bradford Hill, décidèrent au début des années 1950 d'une approche pragmatique pour tester l'hypothèse de la cause du tabac. Ils envoyèrent un questionnaire à tous les médecins anglais (dont à l'époque plus de 80 % fumait !), les interrogeant sur leurs habitudes de consommation de tabac, et les suivirent dans le temps, notant ceux qui développaient un cancer du poumon. Ils parvinrent en 1954 à démontrer une augmentation statistiquement significative du nombre de cancers des poumons chez les fumeurs comparés aux non-fumeurs [2]. Pour la première fois, cette démonstration était prospective, c'est-à-dire que la consommation de tabac était observée plusieurs années avant le développement de la maladie. Cela répondait aux critiques faites aux précédentes études cas-témoins dites « rétrospectives » au cours desquelles des patients atteints du cancer et des témoins sains étaient interrogés *a posteriori* sur leurs habitudes de consommation du tabac les années précédentes. Il était alors impossible de démontrer avec certitude que les patients ne se souvenaient pas à tort d'avoir plus fumé que les témoins, pressés par les questions des chercheurs et le souci de trouver une raison à leur tourment. Cette étude épidémiologique de cohorte, suivie par d'autres, fut l'argument principal qui convainquit le monde médical, politique et l'opinion publique que le tabac était la cause principale de l'épidémie de cancer des poumons dans les pays développés, à côté de bien d'autres maladies (maladies cardiovasculaires, un nombre impressionnant de cancers et maladies respiratoires), non seulement chez le fumeur, mais aussi dans son entourage par tabagisme passif. S'en est suivie l'implémentation de politiques de santé publique de prévention, et la création d'un cadre législatif restreignant l'usage du tabac dans les lieux publics. À ce jour, la preuve statistique demeure plus convaincante que la preuve biologique, car les processus à

Figure 1 : CARTE ORIGINALE DE JOHN SNOW MONTRANT LA LOCALISATION DES CAS DE CHOLÉRA PENDANT L'ÉPIDÉMIE DE LONDRES DE 1854.



l'œuvre dans les développements de cancers et maladies cardiovasculaires chez le fumeur sont encore débattus.

Un autre exemple est celui de l'amiante [3]. La nocivité de l'amiante était connue depuis le début du XX^{ème} siècle, mais il a fallu attendre de larges études épidémiologiques cas-témoins et de cohortes quantifiant l'étendue de ses ravages pour que des politiques restreignant puis bannissant son usage se mettent en place en Europe. Que ce soit en faisant appel à des régressions logistiques pour les études cas-témoins, ou à des modèles de survie pour les études de cohortes, l'objectif est d'estimer l'augmentation du risque d'une maladie associée à un facteur de risque, en tenant compte des facteurs potentiels dits « de confusion » et « de médiation ». La recherche épidémiologique en France s'est intéressée à l'amiante sur le tard, bien après l'Angleterre dont les premières études ont eu lieu dans les années 1950, ce qui s'est traduit par un retard politique avec de lourdes conséquences. De nos jours, on

Figure 2 : EXEMPLES D'ACP RÉVÉLANT LA STRUCTURE ANCESTRALE D'UN ÉCHANTILLON.

A gauche, les deux premiers axes de l'ACP réalisée chez 1,387 Européens reproduisent grossièrement la carte de l'Europe [4]. A droite, trois axes d'ACP confirment la plus grande proximité ancestrale des Porto-Ricains avec les populations européennes et ouest-africaines qu'avec les populations natives d'Amérique [5].

assiste à une résorption de l'épidémie de cancers liés à l'amiante en Angleterre alors qu'elle est encore en expansion en France.

La nouvelle épidémiologie des données génétiques : source d'innovations en statistiques

Les données génétiques sont un excellent exemple de l'explosion de la taille des bases de données à analyser par des statistiques à l'ère du « Big Data ». Grâce à la diminution progressive des coûts et temps d'accès, la recherche en génétique génère aujourd'hui des quantités faramineuses de données, rendant les épidémiologistes indispensables à leur analyse. La mode est aujourd'hui à l'analyse de génomes entiers, testant simultanément l'association de chacun des gènes, ou encore à une échelle inférieure de chacune des bases génétiques portées par un individu, avec une maladie. L'épidémiologiste est alors mis à contribution pour extraire un signal statistiquement significatif de très larges bases de données contenant des dizaines de milliers de gènes, plusieurs millions de variations génétiques, ou bases de nucléotides (la fameuse suite de lettres A,C,T,G, composant la séquence génétique), l'objectif étant d'identifier les « aiguilles » d'intérêt pour le biologiste dans toutes ces « bottes de foin ». Les signaux significatifs sont ensuite interprétés biologiquement et

peuvent ouvrir la voie à l'identification de nouvelles cibles thérapeutiques ou la personnalisation des traitements.

Pour faire parler ces bases de données, des méthodes statistiques éprouvées et innovantes sont utilisées. Un exemple est le recours à l'ACP pour la correction d'effets dits « de stratification de population ». En effet, l'analyse statistique testant l'association entre une variation génétique et un risque de maladie peut être biaisée si la maladie est plus fréquente dans une sous-population dont les caractéristiques génétiques diffèrent des autres sous-populations d'un échantillon. Les axes de plus grande variance de la matrice des génotypes capturant la structure d'ancestralité d'un échantillon, la technique de l'ACP permet d'ajuster l'analyse. Deux exemples particulièrement élégants concernant les populations européennes [4] et Porto-Ricaines [5] sont décrits en Figure 2.

À côté du succès de méthodes élémentaires comme l'ACP, le déferlement de nouvelles données génétiques a participé, avec toutes les données dites « en grande dimension », à redéfinir les contours traditionnels de la statistique. En abandonnant le paradigme de la statistique asymptotique les épidémiologistes ont défini de nouveaux estimateurs prenant le relais des moindres carrés ou du maximum de vraisem-

blance quand le nombre d'observations se trouve largement dépassé par le nombre de variables. Ces méthodes sont également utilisées dans les domaines de l'imagerie médicale ou de l'astrométrie. Elles permettent entre autres de réduire drastiquement le temps d'acquisition d'une image IRM ou Scanner pour une même qualité finale d'image, ce qui n'est pas anodin quand il s'agit de réduire le temps d'immobilisation d'un enfant en bas âge ou de demander à une personne en mauvaise santé de retenir son souffle pour que l'angiographie soit nette [6].

La pharmaco-épidémiologie

La pharmaco-épidémiologie est un autre domaine qui se développe à grands pas. De nos jours, on ne cherche plus seulement à évaluer l'efficacité d'un traitement pour une maladie donnée, mais à équilibrer risques et efficacité. Il est devenu crucial de quantifier précisément le risque d'effets secondaires d'un médicament, et de le comparer aux chances d'amélioration du pronostic du malade. Les compagnies pharmaceutiques se doivent d'évaluer ces risques avec une précision toujours accrue, de peur de se voir reprocher plus tard une mauvaise évaluation qui peut se traduire par une sortie du produit du marché et une très lourde amende. Les exemples sont nombreux [7] et généralement très médiatisés, pensons au cas récent du Mediator.

Le pharmaco-épidémiologiste fait appel à de très larges bases de données collectées sur les populations humaines comme par exemple les registres médicaux des patients, qui couvrent toute l'histoire médicale de millions de patients, mis à jour tous les 6 mois. Il estime par exemple le risque d'épilepsie, chez les patients affectés par la trisomie 21, d'attaques cardiovasculaires chez les diabétiques etc, pour ensuite évaluer si la détection d'une épilepsie ou d'un diabète chez un patient traité avec un nouveau médicament est une conséquence de la prise de ce nouveau médicament (effet adverse) ou n'y est pas lié et est dû à la chance.

L'épidémiologiste est aussi mis à contribution lors de la mise en place puis de l'analyse d'études cas-témoins. Il conseille par exemple sur le nombre de cas et témoins à enrôler et d'autres aspects du design de l'étude qui se doit d'être capable de détecter une amélioration statistiquement significative du pronostic du malade liée à la prise d'un nouveau médicament. C'est finalement l'épidémiologiste qui a le dernier mot : la preuve statistique que le médicament est efficace

dans une population humaine et que ses effets nocifs sont moindres que ses effets bénéfiques primera sur toute autre considération lors de la délivrance d'une autorisation de mise sur le marché d'un nouveau médicament.

Conclusion

L'épidémiologiste a un métier passionnant et varié qui implique un contact avec des professionnels de santé et des chercheurs. Son travail profite à la société de façon multiple. Les techniques statistiques qu'il utilise sont diverses et les développements de la bio-statistique se retrouvent souvent ensuite dans les modèles statistiques d'autres disciplines comme l'économie, la géographie, la démographie, l'astronomie... ■

Références :

- [1] *Doctor John Snow Blames Water Pollution for Cholera Epidemic*, by David Vachon UCLA Department of Epidemiology, School of Public Health May & June, 2005
- [2] *The mortality of doctors in relation to their smoking habits: a preliminary report*, by Doll R, Hill AB. *BMJ*, 2004;328(7455):1529-33; discussion 1533. 1954.
- [3] Le drame de l'amiante en France : comprendre, mieux réparer, en tirer des leçons pour l'avenir, Rapport d'information no 37 du 20 octobre 2005 ; « Les pathologies de l'amiante »
- [4] *Genes mirror geography within Europe*, by Novembre, J, Johnson, T, Bryc, K, Kutalik, Z, Boyko, A, Auton, A, Indap, A, King, K, Bergmann, S, Nelson, M, Stephens, M, and Bustamante, C. *Nature*. 2008; 456(7218): 98–101.
- [5] *PCA-Correlated SNPs for Structure Identification in Worldwide Human Populations*, by Paschou, P, Ziv, E, Burchard, E, Choudhry, S, Rodriguez-Cintron, W, Mahoney, M, and Drineas, P. *PLoS Genetics*, 2007; 3(9).
- [6] *Compressed Sensing MRI : A look at how CS can improve on current imaging techniques*, by Lustig, M, Donoho, DL, Santos, JM, and Pauly, JM. *IEEE Signal Processing Magazine*, 2008; 25 (2).
- [7] *Bad Pharma. How drug companies mislead doctors and harm patients*. Ben Goldacre. 2012. Published by Fourth Estate.