

## ÉPIDÉMIOLOGIE ANALYTIQUE

# Le développement de l'épidémiologie analytique dans la seconde moitié du XX<sup>e</sup> siècle

**Pierre DUCIMETIÈRE**

Directeur de recherche honoraire à l'Inserm

### Résumé

Le rôle de l'épidémiologie analytique dans le développement des connaissances sur la santé humaine n'a cessé de s'affirmer à partir du milieu du XX<sup>e</sup> siècle. Fondée sur les tests d'hypothèses et la recherche de la causalité, elle devrait bénéficier à l'avenir des nouveaux outils offerts par le calcul numérique et l'accès aux bases de données de grande taille.

**Mots-clés :** épidémiologie analytique, épidémiologie descriptive, facteur de risque, test d'hypothèses, causalité.

### Abstract

***The development of analytical epidemiology in the second half of the 20<sup>th</sup> century***

*The role of analytical epidemiology in the development of human health knowledge increased sharply from the mid 20th century. Based upon hypothesis testing and causality research, it should profit by new tools including computing capacities and access to large health data bases.*

**Keywords :** *Analytical epidemiology, descriptive epidemiology, risk factor, hypothesis testing, causality.*

La distinction entre les aspects descriptifs et analytiques de l'épidémiologie a structuré le développement de la discipline au cours de la seconde moitié du XX<sup>e</sup> siècle. Cette distinction est pourtant très schématique et réductionniste, car en pratique, les deux aspects sont le plus souvent liés : un recueil (descriptif) de données de santé peut à l'évidence contribuer à l'établissement d'un risque (analytique) et vice-versa.

### **Distinction entre épidémiologie descriptive et analytique : le poids des logiques institutionnelles**

La distinction entre une épidémiologie qui serait destinée à guider l'action de santé publique à partir principalement de données descriptives et une épidémiologie « pour la connaissance » mettant en évidence des déterminants ou des facteurs de risque est sans doute plus éclairante, même si là encore, ces aspects sont loin d'être exclusifs dans la vie réelle. Forçant sans doute le trait, on a ainsi vu apparaître dans les années 1980, un véritable conflit entre une épidémiologie « de terrain » qui serait dévolue à l'action de santé et une épidémiologie « académique » qui serait vouée au progrès des connaissances. Sur le plan conceptuel, les deux approches ont évidemment en commun de s'attacher à étudier les phénomènes de santé tels qu'ils peuvent être observés à l'échelle des populations humaines, comme y renvoie l'étymologie du terme épidémiologie.

Les logiques institutionnelles, poussées par l'essor prodigieux de la science et de la technologie dans les pays occidentaux, après la Seconde Guerre mondiale, ont conduit à créer des institutions spécifiques dans le champ de l'épidémiologie. C'est ainsi que, dès les années 1945, l'existence aux États Unis de deux agences fédérales distinctes, les *Centers for Disease Control* (CDC) et les *National Institutes of Health* (NIH) a servi de modèle pour de nombreux pays, soucieux de développer en parallèle les deux approches de l'épidémiologie. En France, les missions de l'Institut National d'Hygiène (INH), redéfinies après 1945, visaient clairement l'objectif d'aide à la décision de santé par l'intermédiaire de sa division de la Recherche médico-sociale (DRMS) tout en lui donnant un rôle moteur dans le développement d'une recherche biologique et médicale moderne et compétitive. Cette seconde orientation a pris une place prépondérante dans l'activité de l'organisme, en particulier en lien avec la réforme hospitalo-universitaire. Mais il faut noter que lors de la transformation de l'INH en Institut national de la santé et de la recherche médicale (Inserm), en 1964, l'objectif d'appui à la décision en matière de santé publique, assurée par la DRMS, a été maintenu, au moins jusque dans les années 1980. Il faudra attendre 1998 pour que soit créée une agence sanitaire indépendante, spécifiquement chargée de cette mission, l'Institut national de Veille Sanitaire (InVS), devenue Santé publique France, en 2016.

## Les développements de l'épidémiologie analytique

Le nom de Daniel Schwartz est évidemment associé à l'essor de l'épidémiologie analytique en France. Son enseignement des méthodes statistiques appliquées à la médecine a fortement contribué, à la suite des écoles anglo-saxonnes, à la formalisation d'une méthodologie de la recherche dans ce domaine<sup>1</sup>. Les travaux entrepris dans les années 1950-1970 pour mettre en évidence des « facteurs de risque » des maladies, qu'ils soient de nature exogène comme la consommation de tabac ou endogène comme le niveau de cholestérolémie, ont fait appel à l'un des deux principaux modèles d'études : l'étude cas-témoins rétrospective (tabac et cancer pulmonaire) et l'étude de cohorte prospective (cholestérol et maladie coronaire). Par la suite, de très nombreuses variantes méthodologiques ont été mises au point, adaptées, en particulier, au type de recrutement des sujets, mais elles ne modifient aucunement les deux caractères propres de l'observation épidémiologique : la formulation a priori des hypothèses devant être testées et la difficile question de la « causalité ».

### *Épidémiologie et tests d'hypothèses*

La définition précise des objectifs de l'étude analytique représente une étape essentielle de la constitution du protocole. Elle impose de tenir compte soigneusement des connaissances déjà établies, de nature le plus souvent biologique, concernant les mécanismes de la maladie d'une part, des progrès technologiques permettant de mesurer, avec précision, chez l'homme, les « facteurs d'exposition » à étudier d'autre part. Cette nécessité de poser un ensemble d'hypothèses scientifiques auxquelles l'étude doit répondre a pu être considérée, particulièrement en France, dans les années 1980, comme une limitation qui pourrait être dépassée en ayant recours aux « techniques d'analyse de données », dont « l'analyse factorielle des correspondances ». Schématiquement, elles proposent des représentations visuelles des relations entre les variables recueillies (et corollairement des « distances » entre les sujets), indépendamment de la nature et de la signification des données et sans qu'aucune inférence statistique ne soit possible.

Ces techniques sont aptes à réduire la dimension de l'information présente dans des tableaux de données de grande taille, mais il n'existe guère d'exemple dans la littérature montrant qu'elles aient pu permettre de mettre en évidence de nouveaux facteurs de risque. Le développement en cours de gigantesques bases de données de santé conduit à relancer le débat. Toutefois, l'apport du *Big Data* à l'épidémiologie analytique apparaît plus lié à la possibilité de constituer rétrospectivement des études cas-témoins ou de cohortes de grande taille qu'à des exploitations sans protocole précis d'un ensemble de données généralement hétérogène. C'est ce que montrent par exemple, en génétique les études GWAS<sup>2</sup> ou, en pharmaco-épidémiologie, l'étude des effets secondaires des médicaments, à partir des bases de données de l'Assurance Maladie.

### *Épidémiologie et causalité*

Il est, de nos jours, couramment admis que la mise en évidence d'une association statistique (« corrélation ») à partir de données d'observation ne peut pas recevoir d'emblée d'interprétation causale. En effet, la corrélation du facteur mesuré (« facteur de risque ») avec la maladie pourrait être la conséquence de son association avec un ou plusieurs autres facteurs, qui eux sont causalement associés à la maladie, appelés classiquement « facteurs de confusion ». Dans un tel cas, il ne faut pas s'attendre, du point de vue de la prise en charge des patients, qu'une modification du facteur de

<sup>1</sup> Schwartz D. *Méthodes statistiques à l'usage des médecins et des biologistes*. Paris : Flammarion, 1963.

<sup>2</sup> GWAS (*Genome Wide Association Study*) : étude d'association entre les variations de la structure de l'ADN au niveau d'un grand nombre de codons et la présence d'une pathologie

risque puisse modifier le cours de la maladie, même si, dans certains cas, sa mesure peut se révéler utile, dans le cadre d'un dépistage par exemple.

Bien qu'intuitive a priori, la notion de causalité est particulièrement délicate à définir dans sa généralité et les critères bien connus d'Austin Bradford Hill, publiés en 1965<sup>3</sup>, ne constituent pas une liste de conditions nécessaires et/ou suffisantes qui conduiraient à conclure dans chaque situation. La liste de critères qu'il a proposée permet, au mieux, de préciser les points devant être systématiquement discutés dans le but de cerner la plausibilité d'une relation causale. Ils peuvent être regroupés en trois ensembles : les caractéristiques propres de l'association statistique étudiée ; la connaissance de mécanismes biologiques pouvant être impliqués, à partir, notamment de modèles animaux ; les résultats expérimentaux obtenus chez l'Homme, éventuellement disponibles.

### *Les caractéristiques de l'association*

Plusieurs critères d'Austin Bradford Hill concernent l'association observée dans l'étude : son intensité, la reproductibilité des résultats de la littérature selon les populations étudiées et, tout particulièrement, leur robustesse après la prise en compte des facteurs de confusion potentiels... Ce dernier critère est central d'un point de vue théorique, mais son application pose en pratique de nombreuses difficultés puisque les facteurs de confusion réels ne peuvent être que très partiellement connus et/ou mesurés dans l'étude. Il convient de remarquer qu'une méthodologie élégante (« randomisation mendélienne »), utilisant le polymorphisme de l'ADN comme variable instrumentale dans la relation entre le facteur de risque et la maladie, permet, lorsqu'elle est possible et au prix d'un certain nombre d'hypothèses, de dépasser ces difficultés et de mieux cerner la possibilité d'un effet causal.

### *La connaissance des mécanismes biologiques*

Quelle que soit la nature exogène ou endogène du facteur de risque, la vraisemblance de son implication causale dépend bien entendu de la connaissance des mécanismes physiopathologiques sous-jacents, une attention particulière devant être accordée à la place du facteur dans les chaînes de causalité conduisant à la maladie. L'expérimentation sur modèles animaux fournit des éléments de discussion indispensables, même s'ils ne permettent pas à eux seuls de conclure sur la causalité de l'association. Dans cette direction, le développement du « calcul intensif » laisse entrevoir la possibilité de confronter, à l'avenir, les risques empiriques établis par l'épidémiologie, aux résultats de modélisations mathématiques, suite à de véritables expérimentations réalisées *in silico*.

### *L'expérimentation chez l'Homme*

Dans certaines conditions, une expérimentation est possible chez l'Homme. Elle représente, dans ce cas, une voie privilégiée pour juger de la causalité d'une association. La méthodologie de « l'essai clinique contrôlé », développée et popularisée en France par Daniel Schwartz et son équipe<sup>4</sup>, a joué un rôle essentiel dans l'essor de la recherche clinique et plus précisément de l'évaluation scientifique des thérapeutiques, à partir des années 1970. Son application à la recherche épidémiologique analytique s'est révélée limitée par l'impossibilité, pour des raisons avant tout éthiques, d'attribuer ou non à des sujets humains, un facteur de risque considéré comme possiblement délétère. Pour reprendre l'exemple du tabagisme, alors que sa responsabilité dans l'étiologie du cancer pulmonaire est établie depuis de nombreuses années, certaines voix persistent à la mettre en doute parce que le

<sup>3</sup> Hill AB. The environment and disease : association or causation. *Proc R Soc Med*, 1965 ; 58, 5 : 295-300.

<sup>4</sup> Schwartz D, Flamant R, Lellouch J, Rouquette C. *Les essais thérapeutiques cliniques*. Paris : Masson, 1960.

tabagisme n'a pas fait l'objet d'une véritable expérimentation chez l'Homme. Il en est ainsi de la plupart des facteurs d'exposition en épidémiologie environnementale, pour lesquels seule l'expérimentation animale peut être envisagée. Toutefois, l'abaissement du niveau d'un facteur de risque chez l'Homme, par des interventions médicamenteuses ou comportementales, qui conduit à retarder la survenue de la maladie (prévention primaire) ou diminuer la probabilité de la récurrence (prévention secondaire), est une indication forte d'association causale comme l'illustrent bien les relations établies historiquement entre cholestérolémie et pathologie coronaire.

### *La complémentarité des domaines de recherches*

La mise en évidence de l'implication du niveau de cholestérol circulant (plus précisément du cholestérol transporté par les lipoprotéines de faible densité) dans le développement de la maladie coronaire a résulté de l'accumulation, pendant plus de cinquante années de données convergentes, issues de tous les domaines de la science : la biologie fondamentale (élucidation des mécanismes biochimiques de régulation endogène, modèles animaux d'athérogénèse...), la recherche clinique (identification des particules athérogènes circulantes, classification des hyperlipidémies, prévention secondaire de l'infarctus du myocarde en grande partie indépendante de la voie métabolique concernée...) et la recherche épidémiologique (valeur prédictive du facteur comparable dans la plupart des populations humaines, prévention primaire observée chez les sujets à risque coronaire élevé...). La grande cohérence des résultats obtenus montre qu'une complémentarité des différents domaines de recherche peut permettre d'établir un socle solide de connaissances sur lesquelles des consensus peuvent s'établir.

Le développement, au cours de la seconde moitié du xx<sup>e</sup> siècle, d'une épidémiologie analytique fondée sur l'étude des populations humaines, est inséparable des progrès de la recherche biologique faite dans les laboratoires d'une part, de la recherche clinique effectuée auprès des malades d'autre part. Plus que jamais les trois approches apparaissent nécessaires afin que la santé individuelle et collective bénéficie au mieux des progrès obtenus<sup>5</sup>. Les « révolutions numériques » annoncées devraient, quant à elles, être en mesure d'indiquer de nouvelles pistes de recherche en proposant de nouveaux outils, à la condition que le « caractère exploratoire » de leur démarche soit réaffirmé.

L'auteur déclare n'avoir aucun lien d'intérêt.

---

<sup>5</sup> Ducimetière P. Recherche : la triple approche. *Revue d'épidémiologie et de santé publique*, 2005 : 53 : 233-4.