



# Inserm

La science pour la santé  
From science to health

**le comité pour l'histoire de l'Inserm**

# LES CAHIERS N° 1

**Histoire  
de l'épidémiologie**  
Enjeux passés,  
présents et futurs





# LES CAHIERS DU COMITÉ POUR L'HISTOIRE DE L'INSERM

## HISTOIRE DE L'ÉPIDÉMIOLOGIE ENJEUX PASSÉS, PRÉSENTS ET FUTURS

N° 1, janvier 2020

**Direction de la publication** Gilles Bloch  
**Directrice adjointe de la communication** Priscille Rivière  
**Direction éditoriale et scientifique** Pascal Griset  
**Secrétariat de rédaction** Céline Paillette  
avec la collaboration des membres du Comité pour l'histoire de l'Inserm  
**Création couverture** Myriem Belkacem Inserm  
**Impression** Inserm ADS/reprographie  
**N°ISSN** *en cours*  
**Dépôt légal** *en cours*

Inserm, 101, rue de Tolbiac, 75654, Paris Cedex 13 [www.inserm.fr](http://www.inserm.fr)  
Version électronique en libre accès sur <http://www.ipubli.inserm.fr/handle/10608/10030>

---



## Sommaire

### Éditorial

Pascal GRISET .....	5
<b>L'épidémiologie, la connaissance et l'action en santé publique</b>	
Philippe LAZAR .....	9
<b>L'Inserm, l'épidémiologie et l'orientation de la santé publique</b>	
Corinne ALBERTI.....	13
<b>Epidemiology, statistics, and public health in the victorian britain, 1837-1902</b>	
Anne HARDY .....	21
<b>Mort et résurrection du virus de la grippe espagnole</b>	
Patrick BERCHE.....	27
<b>Le développement de l'épidémiologie analytique dans la seconde moitié du XX<sup>e</sup> siècle</b>	
Pierre DUCIMETIÈRE.....	33
<b>La réflexion épistémologique en épidémiologie dans la seconde moitié du XX<sup>e</sup> siècle</b>	
Joël COSTE .....	39
<b>L'étude de Framingham. Quel rôle dans l'histoire de l'épidémiologie des facteurs de risque ?</b>	
Élodie GIROUX.....	47
<b>Un monde moderne. L'épidémiologie des maladies infectieuses avant sa réinvention des années 1980-1990</b>	
Gaëtan THOMAS .....	57
<b>Données massives et médico-administratives. Nouveaux enjeux pour l'épidémiologie</b>	
Grégoire REY .....	65
<b>Santé globale et mesure. Un nouveau paradigme pour de nouveaux acteurs</b>	
Grégoire LURTON .....	73
<b>Quelques considérations personnelles. Sur l'évolution de l'épidémiologie ces dernières années en France</b>	
Marcel GOLDBERG .....	81



## Éditorial

### Pascal GRISET

Président du Comité pour l'histoire de l'Inserm  
Professeur d'histoire contemporaine, Lettres-Sorbonne Université

**Mots-clés :** Comité pour l'histoire, histoire, Inserm, épidémiologie

**Keywords :** *History Committee, History, French National Institute for Health and Biomedical Research, Epidemiology*

Le Comité pour l'histoire de l'Inserm est un espace de rencontres entre les historiens, les chercheurs en sciences humaines et sociales, les acteurs contemporains de la santé publique et de la recherche biomédicale. À l'image des Comités pour l'histoire, créés à l'initiative de François Caron à partir du début des années 1980 – dans les champs de l'électricité, de l'aluminium, des chemins de fer, des postes<sup>1</sup>... –, le Comité pour l'histoire de l'Inserm favorise le « contact entre des spécialistes venus d'horizons différents<sup>2</sup> » et encourage les échanges entre historiens, témoins et acteurs du temps présent.

Le Comité pour l'histoire de l'Inserm est en effet ouvert à toutes les approches pouvant contribuer à l'écriture d'une histoire globale de la santé publique et de la recherche biomédicale. « L'histoire globale, écrivait François Caron en 1983, est l'inverse d'une histoire simplificatrice. Elle prend en compte l'infinie diversité du réel tout en cherchant à comprendre les influences réciproques et les liens qui unissent les différentes sphères d'activités<sup>3</sup> ». Nos *Cahiers* s'inscrivent dans cette lignée et ils refléteront l'activité du Comité, devenant ainsi un « carrefour<sup>4</sup> » favorisant l'émergence de nouveaux projets.

En janvier 2019, la deuxième journée d'étude du Comité a été consacrée à l'histoire de l'épidémiologie, sur une proposition de Pierre Corvol<sup>5</sup>. Environ 30 ans après le colloque Inserm

---

<sup>1</sup> Merger M. *François Caron, 1934-2014*. Site Association française d'Histoire économique, [https://afhe.hypotheses.org/5686] ; programme de la première journée d'étude du Comité pour l'histoire de l'Inserm, 18 janv. 2018, site Inserm [https://www.inserm.fr/actualites-et-evenements/actualites/comite-pour-histoire-inserm-organise-sa-premiere-journee-etude] (consultés le 15 déc. 2019).

<sup>2</sup> Éditorial, La Rédaction, *Bulletin d'histoire de l'électricité*. 1983, 1 : 4.

<sup>3</sup> Caron F. Ouverture du colloque. In *L'électricité dans l'histoire. Problèmes et méthodes*. Actes du colloque de l'Association pour l'histoire de l'électricité en France, Paris, 11-13 oct. 1983 : 2.

<sup>4</sup> Éditorial, La Rédaction, *Bulletin d'histoire...*, *op. cit.*

<sup>5</sup> Corvol P, Griset P, Paillette C. L'épidémiologie entre le terrain des épidémies et l'approche populationnelle, XIX-XX<sup>e</sup> siècle. *Médecine/Sciences*, nov. 2019, 35 : 886-890.

organisé en hommage à Daniel Schwartz, considéré comme le père fondateur de la discipline en France, des historiens, des personnalités et des grands témoins de l'histoire de l'Inserm, et plus largement de la discipline, ainsi que des acteurs du temps présent, échangeaient autour de l'histoire de l'épidémiologie et de ses enjeux passés, présents et futurs<sup>6</sup>. Plusieurs voix se sont alors élevées pour nous inciter à ne pas laisser perdre les apports de cette rencontre. C'est donc bien la richesse de cette journée qui nous a amenés à lancer ces *Cahiers du Comité pour l'histoire de l'Inserm*.

Ce premier numéro vient éclairer certains aspects de l'histoire de l'épidémiologie depuis le XIX<sup>e</sup> siècle autour de quatre axes principaux débattus lors de la journée d'étude : l'épidémiologie et le gouvernement de la santé publique ; les nouvelles fondations épistémologiques de l'épidémiologie au cours du XX<sup>e</sup> siècle ; l'épidémiologie et les crises internationales ; les défis contemporains de l'information et de l'innovation.

La question des relations entre l'épidémiologie et le gouvernement de la santé publique est posée sous différents angles. Philippe Lazar, directeur général de l'Inserm (1982-1996) souligne tout le poids des connaissances issues de la recherche en matière de passage à l'action en santé publique et témoigne sur un aspect crucial de l'histoire de la décision. Corinne Alberti, directrice de l'ITMO Santé publique relie les différentes écoles de l'épidémiologie française (celle de Daniel Schwartz, celle de François Grémy). Elle met en lumière les liens forts et structurants qui associent, dans l'histoire de l'Inserm et aujourd'hui encore, l'épidémiologie et la santé publique, au sein de dynamiques sans cesse évolutives. L'historienne Anne Hardy retrace les liens entre le travail d'enquête sur le terrain des épidémies, les approches statistiques de l'épidémiologie, et son institutionnalisation au sein d'enceintes expertes, à l'époque de l'Angleterre victorienne. Les épidémies du passé font parfois encore l'objet de recherches contemporaines. Comme le montre Patrick Berche, c'est le cas de la « grippe espagnole », fléau mondial dont le bilan humain dépasse celui de la Grande Guerre, et dont l'impact demeure inexplicable.

La réflexion autour de l'épistémologie de l'épidémiologie au XX<sup>e</sup> siècle occupe une place importante : l'analyse de Pierre Ducimetière permet de mieux saisir les enjeux qui présidèrent au développement de l'épidémiologie analytique dans la seconde moitié du XX<sup>e</sup> siècle. Joël Coste propose une réflexion sur l'épistémologie de l'épidémiologie autour des aspects fondamentaux de la causalité et de la validité des enquêtes tandis qu'Élodie Giroux livre une histoire de l'incontournable enquête de Framingham. Gaëtan Thomas s'intéresse à l'histoire méconnue de la collaboration de Daniel Schwartz avec le Centre international de l'enfance et sa Station pilote, et aux relations qui s'y nouaient entre épidémiologie statistique et épidémiologie de terrain.

Enfin, la contribution de Grégoire Rey sur les données massives et médicaux-administratives et celle de Grégoire Lurton sur le nouveau paradigme de la mesure de la santé globale montrent bien toute la complexité (processus de collecte, risques de biais, etc.) associée à un traitement de l'information dont l'ampleur est devenue colossale et dont l'impact sociétal et éthique n'est pas encore réellement mesuré. La conclusion de ce premier numéro revient à Marcel Goldberg qui livre un regard personnel, riche de son expérience et tourné vers les enjeux futurs de l'épidémiologie.

Ce numéro, par ses analyses et ses témoignages, est un apport à l'écriture d'une histoire de l'épidémiologie. Il tente de montrer les approches qui permettront de prendre en compte les oscillations et les structurations complémentaires et, parfois, concurrentes entre l'épidémiologie de terrain et l'approche populationnelle, les relations de l'épidémiologie à l'hygiène publique, puis à la santé publique, et cela à de multiples échelles, du local à l'international. Nous espérons qu'il pourra modestement contribuer à éclairer les débats actuels. Ces derniers interrogent le politique et les

---

<sup>6</sup> Livret de présentation et captation audio de la journée 2019 disponibles sur le site Inserm [<https://www.inserm.fr/actualites-et-evenements/actualites/histoire-epidemiologie-enjeux-passes-presentes-et-futurs>] (consulté le 15 déc. 2019).

décisions en matière de santé publique à l'aune d'une « épidémiologie populaire<sup>7</sup> » et de la mobilisation des individus tant sur les enjeux de la vaccination qu'autour des crises sanitaires.

Avant toute chose cependant, nous espérons que cette première livraison des *Cahiers* incitera historiens et acteurs, analystes et témoins, à continuer à échanger et à s'inscrire dans une démarche collective dédiée à une histoire vivante, visible et présente, au sein même des communautés engagées aujourd'hui dans la santé. Le Comité est à leur écoute et sera totalement disponible pour favoriser et aider les initiatives allant dans ce sens.

## Remerciements

*Nous remercions la Présidence et la Direction générale de l'Inserm qui assure le Comité pour l'histoire de l'Inserm de son soutien et l'équipe du département de l'Information scientifique et de la Communication – Carine Delrieu et Catherine d'Astier qui ont soutenu ce projet, Yann Cornillier pour ses conseils éditoriaux et Priscille Rivière pour son lancement – ainsi que l'équipe de l'Information scientifique et technique autour de Michel Pohl, qui a rendu possible la publication de cette nouvelle collection sur la plateforme iPubli.*

## Appel à témoignages

*Le Comité pour l'histoire de l'Inserm encourage les témoins et les acteurs de l'histoire de l'Inserm, de la santé publique et de la recherche biomédicale à soumettre des textes livrant leurs expériences et leurs itinéraires professionnels, individuels et collectifs. Les textes doivent s'inscrire dans un format de 25 000 signes (espaces et notes de pages inclus). Pour toute question et information, contactez le secrétariat scientifique : [celine.paillette@ext.inserm.fr](mailto:celine.paillette@ext.inserm.fr)*

---

<sup>7</sup> Selon l'expression forgée à la fin des années 1980 par Phil Brown. Brown P. Retour sur l'épidémiologie populaire. In Akrich M, Barthe Y, Rémy C, dir. *Sur la piste environnementale. Menaces sanitaires et mobilisations profanes*. Paris : Presses des Mines : 53-85.



## ÉPIDÉMIOLOGIE ET SANTÉ PUBLIQUE

## L'épidémiologie, la connaissance et l'action en santé publique

**Philippe LAZAR**

Directeur général de l'Inserm, 1982-1996

**Résumé**

Philippe Lazar, ancien élève de l'École Polytechnique, oriente sa carrière vers les statistiques appliquées aux sciences du vivant et vers l'épidémiologie. Il entre à l'Institut National d'Hygiène en 1960 et fait donc partie de l'Inserm dès sa création en 1964. Il y dirigera l'unité « Recherches épidémiologiques et statistiques sur l'environnement et la santé » puis devient le Directeur général de l'Inserm de 1982 à 1996. Il présidera par la suite le conseil d'administration de l'Institut de recherche pour le développement de 1997 à 2001. Ses travaux scientifiques ont porté principalement sur la cancérologie (clinique et expérimentale), la prématurité et les risques d'avortement spontané précoce. Il est l'auteur, hormis ses publications scientifiques, de quatre ouvrages destinés au grand public : *Les explorateurs de la santé*, en 1989, chez Odile Jacob, *L'éthique biomédicale en question*, en 1996, *Autrement dit laïque*, en 2003, chez Liana Lévi, et *Court traité de l'âme*, en 2008, chez Fayard. Il dirige actuellement la revue interculturelle *Diasporiques/Cultures en mouvement*. Le 23 janvier 2019, lors de la journée d'étude organisée par le Comité pour l'histoire de l'Inserm, il introduisit et présida la session « L'épidémiologie et le gouvernement de la santé publique ».

**Mots-clés** : épidémiologie, biostatistique, santé publique, décision, expertise collective.

**Abstract*****Epidemiology, knowledge and action in public health***

*Philippe Lazar, a former student of the Ecole Polytechnique, directs his career towards statistics applied to life sciences and epidemiology. He joined the French National Institute of Hygiene in 1960. Therefore, he was part a few years later of the creation of the French National Institute for Health and Medical Research (Inserm) in 1964. There, he headed the "Epidemiological and statistical research on environment and health" unit and then became the Director General of Inserm from 1982 to 1996. He then chaired the Board of Directors of the Research Institute for Development from 1997 to 2001. His scientific work focused mainly on oncology (clinical and experimental), prematurity and the risks of early spontaneous abortion. In addition to his scientific publications, he is the author of four books for the general public : Les explorateurs de la santé (1989, Odile Jacob), L'éthique biomédicale en question, (1996), Autrement dit laïque (2003, Liana Lévi), and Court traité de l'âme, (2008, Fayard). He is currently the editor of the intercultural journal Diasporiques/Cultures en mouvement. On 23 January 2019, during the meeting organised by the Committee for the History of Inserm, he introduced and chaired the session "Epidemiology and the government of public health".*

**Keywords** : epidemiology, biostatistics, public health, decision-making, collective expertise.

Je suis particulièrement honoré et heureux de vous dire quelques mots d'introduction pour cette session de l'histoire de l'Inserm consacrée à l'épidémiologie<sup>1</sup>. Quand j'ai commencé avec nombre de mes collègues ici présents à travailler sous la direction de Daniel Schwartz à Villejuif, nous n'employions pas ce terme : le laboratoire dans lequel nous exerçons notre activité – l'unité 21 – était une unité de recherche en « statistique médicale ». Il a fallu plusieurs années pour qu'à l'initiative de Claude Rumeau-Rouquette – à qui je voudrais rendre hommage pour sa judicieuse audace en la matière –, nous nous décidions à introduire le terme « épidémiologie » comme déterminant l'ensemble de nos activités, alors en fait plus spécifiquement consacrées aux maladies chroniques. Et il est vrai qu'il s'agissait là d'une prise de position originale puisque, par exemple, à l'École de santé publique de Harvard, il existe une distinction nette entre les laboratoires de biostatistique et d'épidémiologie, les premiers étant placés sous la responsabilité de statisticiens et les seconds de médecins.

Je suis reconnaissant à Pierre Corvol de nous avoir raconté la longue histoire du calcul des probabilités et de la statistique dans notre pays, et j'en profite pour rappeler le rôle majeur qu'il a personnellement joué pour introduire à l'Inserm la recherche clinique dans son sens étymologique, c'est-à-dire « au lit du malade », ce complément majeur de la recherche biomédicale au sens large.

Le thème de cette rencontre porte plus spécifiquement sur les relations entre l'épidémiologie et la santé publique : une question à l'évidence majeure. Pour éviter toute interprétation réductrice de ces relations, il me semble toutefois essentiel de commencer par rappeler que l'épidémiologie est d'abord une science, c'est-à-dire un outil de connaissance rationnelle, et que ce n'est pas la même chose que de comprendre et d'agir ! Passer de la connaissance à l'action est loin d'aller de soi, pour de multiples raisons. Cette remarque vaut pour toutes les sciences, et en particulier les sciences sociales qui sont, vous le savez, très souvent instrumentalisées au travers de l'idée que leur utilité essentielle est de faciliter l'acceptabilité sociale des mesures qu'on est en train de prendre ! Et il en est de même pour l'épidémiologie tant, par nature, elle a affaire à des questions directement relatives à la santé humaine.

On a bien sûr besoin de connaissances aussi solides que possible pour agir en toute « connaissance » de cause. Mais les procédures d'acquisition de connaissances et d'action sont différentes. Comme toute recherche, une recherche épidémiologique est une procédure à long terme et qui porte sur des questions spécifiques. Agir en fonction des résultats d'une telle recherche est un processus social qui entraîne nécessairement toute une série de modifications des milieux sur lequel porte l'action. Comprendre et agir sont des processus l'un et l'autre complexes mais d'une complexité radicalement distincte. Quitte à prendre le risque de m'exprimer de façon trop simplificatrice, je dirai volontiers qu'une recherche est beaucoup plus souvent monodisciplinaire qu'on a de nos jours tendance à le dire alors qu'une action est impérativement multidisciplinaire.

---

<sup>1</sup> La session est disponible en audio sur le site Inserm [<https://www.inserm.fr/actualites-et-evenements/actualites/histoire-epidemiologie-enjeux-passes-presentes-et-futurs>] (consulté le 26 nov. 2019).

Soit dit en passant, c'est la raison pour laquelle j'avais introduit à l'Inserm le concept et les modalités potentielles d'action qu'exprime l'appellation « expertise collective » avec son implication majeure : être l'objet de la synthèse critique de l'ensemble des informations disponibles sur un sujet donné avant de tenter d'agir à son propos, ce qui suppose de réunir à cette fin des experts relevant de disciplines différentes et donc capables de l'aborder dans toutes ses dimensions.

Encore faut-il évidemment, qu'on ait, pour agir, du grain à moudre. Et, comme les exemples qui vont être exposés le montrent, l'épidémiologie est un pourvoyeur privilégié d'informations utilisables dans le domaine de la santé publique.

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



## SANTÉ PUBLIQUE

# L'Inserm, l'épidémiologie et l'orientation de la santé publique

**Corinne ALBERTI**

Directrice de l'Institut thématique Inserm Santé Publique

### Résumé

Médecin et Professeure d'épidémiologie et de santé publique, à l'Université de Paris, Corinne Alberti est, depuis 2017, directrice de l'Institut de Santé publique de l'Inserm. Elle est aussi directrice du Centre d'investigation clinique - Épidémiologie clinique (Inserm CIC-EC 1426) de l'hôpital universitaire pédiatrique Robert Debré et codirectrice de l'UMR 1123 Épidémiologie clinique et évaluation économique appliquées aux populations vulnérables (ECEVE). Ses travaux concernent les problématiques spécifiques de la recherche en pédiatrie qu'elles soient méthodologiques, épidémiologiques, ou en santé publique. Dans le champ de la santé publique, Corinne Alberti s'intéresse notamment aux parcours de soins chez les adolescents atteints de maladie chronique et au développement d'interventions pour le bien-être des adolescents et jeunes adultes. Elle propose ici une synthèse sur l'articulation entre épidémiologie et santé publique au sein de l'Inserm.

**Mots-clés** : épidémiologie, santé publique, recherche clinique, Inserm.

### Abstract

#### **The French National Institute for Health and Medical Research (Inserm), the epidemiology and the orientation of Public Health**

*Physician and Professor of Epidemiology and Public Health, at the University of Paris, Corinne Alberti, has been Director of the Institute of Public Health of the French National Institute for Health and Medical Research since 2017. She is also Director of the Clinical Investigation Centre - Clinical Epidemiology (Inserm CIC-EC 1426) of Robert Debré Hospital and Co-Director of UMR 1123 Clinical Epidemiology and Economic Evaluation Applied to Vulnerable Populations (ECEVE). Her work focuses on the specific issues of paediatric research, whether methodological, epidemiological or public health. In the field of public health, Corinne Alberti's research interests include the treatment pathways for adolescents with chronic disease and the development of interventions for the well-being of adolescents and young adults. She proposes here a synthesis on the articulation between epidemiology and public health within the French National Institute for Health and Medical Research.*

**Keywords** : epidemiology, public health, clinical research, French National Institute for Health and Medical Research.

Brosser l'histoire de l'épidémiologie est important pour comprendre la place de la discipline à l'Inserm, son rapport à la santé publique, son positionnement actuel et ses développements futurs.

Faire un retour en arrière sur l'histoire de la construction de l'épidémiologie et de la santé publique à l'Inserm était une gageure pour moi car ma formation est celle d'une clinicienne avec un internat, d'abord en médecine générale, puis en néphrologie et réanimation. Ce n'est que tardivement que j'ai construit mon parcours, d'abord en épidémiologie clinique et, particulièrement, en réanimation médicale, puis en pédiatrie. Je dirige l'unité de recherche clinique et le CIC-EC de l'hôpital Robert Debré depuis 2001. Depuis quelques années, j'investis le champ de la santé publique notamment en prévention, les parcours de soins et la recherche interventionnelle en santé des populations pour le bien-être et la prévention des conduites à risque chez les adolescents et les adultes jeunes. Ma construction disciplinaire s'est faite dans le champ de la recherche clinique du fait de mon double parcours médical et scientifique. Ayant effectué mon diplôme d'études approfondies (DEA) sous la direction d'Alain-Jacques Valleron, ma thèse d'épidémiologie dans l'école doctorale alors dirigée par Jean Maccario à Villejuif, j'ai été formée dans les deux maisons de l'épidémiologie, celle de François Grémy et celle de Daniel Schwartz<sup>1</sup>.

Ce papier propose une synthèse sur l'articulation de l'épidémiologie et de la santé publique au sein de l'Inserm, vue à travers les grandes étapes de son histoire, les figures pionnières et les éléments structurants la relation entre la recherche en épidémiologie et la recherche en santé publique.

## L'Inserm et les écoles de l'épidémiologie en France

### *Daniel Schwartz, un bâtisseur de l'épidémiologie et de la santé publique modernes*

L'épidémiologie à l'Inserm commença dès la création de notre Institut, en 1964, avec les travaux de Daniel Schwartz, aujourd'hui considéré comme le père de la biostatistique française et qui introduisit, en France, les méthodes de l'épidémiologie moderne<sup>2</sup>. Polytechnicien et ingénieur des tabacs, Daniel Schwartz s'intéressa à la santé humaine après avoir travaillé sur la mosaïque des plants de tabac<sup>3</sup>. Il fut l'un des premiers à avoir montré le rôle du tabac dans l'étiologie des cancers du poumon et des cancers bucco-pharyngés ainsi que l'importance de l'inhalation de la fumée dans les cancers du poumon. Au décours de ses recherches, l'Institut national d'hygiène (INH) – qui devint plus tard l'Inserm – lui attribua, en 1960, la première unité de recherche de recherches

<sup>1</sup> Je remercie très sincèrement et chaleureusement Freddy Spira qui nous a accompagnés, Jean-Marie Gagliolo et moi-même, dans ce parcours initiatique.

<sup>2</sup> Lellouch J, Lazar P, Ducimetière P *et al.* Daniel Schwartz, un maître de la statistique médicale. *La Jaune et la Rouge*, 2010, 651 : 42-3.

<sup>3</sup> Spira A. Santé publique, statistique et renouveau de la médecine. *Savoir/Agir*, 2008, 5 : 89-105.

statistiques, l'unité 21, à l'hôpital Paul Brousse à Villejuif. Il la dirigea pendant 25 ans<sup>4</sup>. Daniel Schwartz fut à l'origine du développement d'une discipline, la statistique. Son objectif fut de progresser dans la réflexion sur la notion de causalité pour expliquer les maladies, proposer des traitements et avoir un modèle conceptuel complémentaire à ce qui existait. Son implication dans la recherche a été, entre autres, favorisée par la réforme hospitalo-universitaire proposée par Robert Debré (1958), créant les centres hospitaliers universitaires (CHU). En 1968, il fut nommé professeur à la faculté de médecine de Paris XI. Son héritage est immense et, sans pouvoir en faire ici un tour exhaustif, nous pouvons souligner le poids de ses travaux dans les essais thérapeutiques<sup>5</sup>. Autour de Daniel Schwartz et de l'unité 21, on parla de « l'école de Villejuif » et ses élèves rayonnèrent dans de nombreux domaines<sup>6</sup>. En outre, de nombreux étudiants furent formés à la biostatistique durant trente ans à l'Institut de Statistiques de l'Université de Paris, puis au Centre d'enseignement de la statistique appliquée à la médecine et à la biologie médicale (CESAM)<sup>7</sup>. En 1994, j'ai eu la chance d'assister à quelques-uns de ses cours, le CESAM permettant de valider les biostatistiques dans le DEA de biomathématiques de Paris 6, dirigé alors par Alain-Jacques Valleron.

### *L'école François Grémy*

En parallèle de Villejuif, un deuxième pôle d'épidémiologie et santé publique se développait à Paris, au sein de la Pitié Salpêtrière, avec François Grémy, médecin, physicien, biostatisticien, à la tête de l'U88, Informatique en milieu médical et hospitalier. Il fonda la première unité de recherche Inserm consacrée à l'informatique médicale à l'hôpital de la Pitié-Salpêtrière, dont il fut le directeur durant treize années (1968-1982)<sup>8</sup>. Ce fut à « l'école François Grémy » que furent formés Jean-François Boisvieux, Marcel Goldberg ou encore Roger Salamon.

De son côté, Marcel Goldberg contribua à la conception et à la réalisation d'un logiciel de gestion de base de données original (projet SARI), notamment appliqué à la banque d'information automatisée sur les médicaments (BIAM). À partir de 1971, il participa à la mise en place d'une base de données d'études épidémiologiques sur la population des salariés d'EDF-GDF (environ 150 000 travailleurs) ou encore de la cohorte Gazel (15 000 hommes et 5 500 femmes agents d'EDF-GDF). Aujourd'hui, la cohorte Constances, composée de personnes consultant des centres d'examen de santé de la Sécurité sociale et constituée d'un échantillon de 200 000 sujets représentatifs de la population française adulte affiliée au régime général de Sécurité sociale (plus de 80 % de la population) constitue une plateforme largement ouverte à la communauté des chercheurs et de la santé publique, afin de contribuer au développement de la recherche épidémiologique et de fournir des informations à visée de santé publique<sup>9</sup>.

<sup>4</sup> Site Histoire Inserm [[https://histoire.inserm.fr/les-femmes-et-les-hommes/pierre-ducimetiere/\(page\)/5](https://histoire.inserm.fr/les-femmes-et-les-hommes/pierre-ducimetiere/(page)/5)] (consulté le 29 juil. 2019).

<sup>5</sup> Éloge. Daniel Schwartz (1917-2009). *Revue d'épidémiologie et de Santé Publique*, 2009, 57 : 313-16.

<sup>6</sup> BenKimoun P. Daniel Schwartz, spécialiste de la statistique médicale. *Le Monde*, 8 sept. 2009, [[https://www.lemonde.fr/disparitions/article/2009/09/08/daniel-schwartz-specialiste-de-la-statistique-medicale\\_1237529\\_3382.html](https://www.lemonde.fr/disparitions/article/2009/09/08/daniel-schwartz-specialiste-de-la-statistique-medicale_1237529_3382.html)] (consulté le 29 juil. 2019).

<sup>7</sup> Encore aujourd'hui, au sein de Médecine Sorbonne Université, [<https://master-sante.sorbonne-universite.fr/diucsam/>] (consulté le 29 juil. 2019).

<sup>8</sup> [<https://histoire.inserm.fr/les-femmes-et-les-hommes/francois-gremy>] (consulté le 29 juil. 2019).

<sup>9</sup> Présentation de Constances, [<http://www.constances.fr/cohorte/presentation-constances.php>] (consulté le 29 juil. 2019).

### *La construction d'une communauté d'épidémiologistes*

Désireux de fédérer et donner une plus grande visibilité à la communauté francophone de l'épidémiologie et de la santé publique, Daniel Schwartz proposa de développer une revue scientifique indexée dans l'*Institute for Scientific Information*. Il se rapprocha alors de Raymond Sohier, rédacteur en chef de la *Revue d'hygiène et santé publique*. Ensemble, ils créèrent, en 1976, la *Revue d'Épidémiologie et de Santé Publique* (RESP), première revue d'épidémiologie française de niveau international. Dans un même élan, la même année, il fonda l'Association des épidémiologistes en langue française (Adelf), continuant ainsi à œuvrer pour structurer notre communauté<sup>10</sup>.

Entre 1970 et 1986, plusieurs épidémiologistes passés par les unités de Daniel Schwartz et de François Grémy créèrent d'autres équipes au sein des hôpitaux Paul-Brousse, Port-Royal, ou encore Kremlin-Bicêtre, comme par exemple, mais cela n'est, bien sûr, pas exhaustif :

- Philippe Lazar puis Denis Hémon, U170 « Épidémiologie et statistiques sur l'environnement et la santé » jusqu'en 2005, puis UMR 1018 « Épidémiologie et santé des populations »,
- Claude Rumeau-Rouquette puis Gérard Bréart, U149 « Épidémiologie de la santé des femmes et des enfants » à l'hôpital Baudelocque-Port-Royal à Paris de 1988 à 2005,
- Pierre Ducimetière, U258 « Épidémiologie cardiovasculaire » de 1982 à 1998, à l'hôpital Broussais à Paris, puis de 1999 à 2005 à l'hôpital Paul-Brousse, l'intitulé devenant « Épidémiologie cardiovasculaire et métabolique »,
- Alfred Spira puis Henri Leridon, U292 « Santé publique, épidémiologie, reproduction humaine » de 1985 à 2001.

En 1986, Eveline Eschwège succéda à Daniel Schwartz à la direction de l'unité 21 qui devint l'unité « Statistiques, Recherches cliniques et épidémiologiques, métabolisme, mode de vie ». Au-delà de Paris et de sa région, d'autres pôles se développèrent : par exemple, sous l'impulsion de Laure Papoz qui avait développé des enseignements d'épidémiologie aux côtés de Daniel Schwartz, à Montpellier (unité 500 « Épidémiologie des maladies chroniques et du vieillissement ») ; sous celle de Jean Martin à Nancy (unité 115 « Santé au travail et santé publique : méthodes et applications ») ; ou encore sous celle de Roger Salamon à Bordeaux (unité 330 « Épidémiologie, santé publique et développement »)<sup>11</sup>.

### **Panorama de la structuration entre épidémiologie et santé publique au sein de l'Inserm, 1964-2016**

En plus de la création d'unités, l'Inserm structurait sa communauté de chercheurs en épidémiologie et santé publique, notamment à l'initiative de la présidence et de la direction générale.

#### *La Direction de la Recherche Médico-sociale de l'Inserm*

Aux débuts de l'Inserm, la direction de la Recherche médico-sociale de l'Inserm (DRMS), sur le site du Vésinet, dans les Yvelines, était chargée des travaux d'épidémiologie descriptive et plus particulièrement de travaux portant sur l'organisation des soins en tentant d'intégrer des nouvelles approches comme les sciences sociales. Ses deux grandes missions étaient d'une part le recueil, la réunion, l'analyse des données d'épidémiologie descriptive afin de mieux apprécier la fréquence et

<sup>10</sup> Éloge. Daniel Schwartz (1917-2009). *Revue d'épidémiologie et de Santé Publique*, 2009, 57 : 313-316.

<sup>11</sup> [<https://histoire.inserm.fr/de-l-inh-a-l-inserm/les-directeurs-d-unite-de-recherche>] (consulté le 29 juil. 2019).

la distribution des maladies dans les domaines relevant de la santé publique, d'autre part l'établissement de plusieurs statistiques de routine, dont à partir de 1968, la statistique nationale des causes de décès, qui est demeurée jusqu'à ce jour une activité propre de l'Inserm au sein du CépiDc<sup>12</sup>. La DRMS permettait aussi d'effectuer l'interface avec les décideurs du monde de la santé.

Dès le milieu des années 1970, la DRMS ne correspondit plus aux besoins de développement de la biomédecine. Désormais, l'Inserm devait s'affirmer comme un organisme de recherche biomédicale puissant. Comme le rappelait Pierre Ducimetière, l'accent essentiel était mis sur la formation de cliniciens-biologistes d'une part, la création d'unités de recherche de médecine expérimentale d'autre part<sup>13</sup>.

### *La structuration de l'épidémiologie à l'Inserm depuis les années 1980*

Au sein de l'Inserm, la pluridisciplinarité des approches dans le domaine de la santé fut assurée par des réponses institutionnelles progressives. L'entrée des sciences humaines à l'Inserm se fit sous Philippe Lazar avec la création d'une intercommission des sciences humaines et sociales de la santé à laquelle participait Françoise Héritier, anthropologue au Collège de France. Dans cette dynamique, le Centre de recherche, médecine, sciences, santé et société (CERMES), fut fondé par Claudine Herzlich. Le Cermes réunit des chercheurs de plusieurs disciplines allant de l'histoire de la médecine à l'épidémiologie, en passant par la sociologie et l'économie<sup>14</sup>. À ce moment, furent aussi créés les centres d'investigation clinique (1992)<sup>15</sup> et les expertises collectives (1993)<sup>16</sup>. Par la suite, Claude Griscelli apporta un soutien financier et institutionnel fort à la santé publique en développant des actions concertées notamment en épidémiologie clinique. Il fut l'instigateur du Programme hospitalier de recherche clinique (PHRC) lancé par le ministère de la santé.

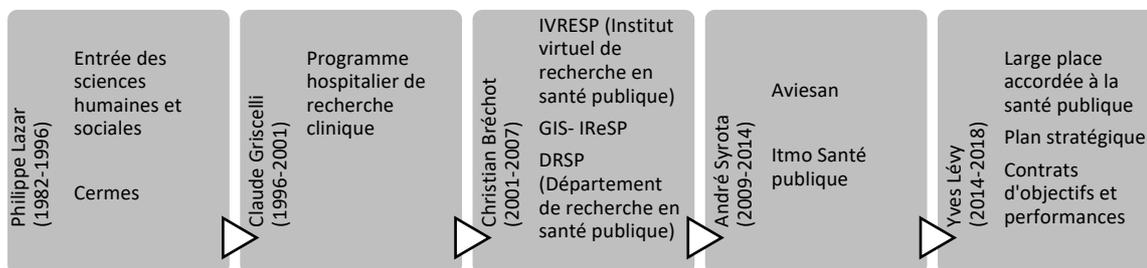


Tableau 1. Synthèse de la structuration de la recherche en santé publique depuis 1982, sous les différents directeurs généraux, puis Présidents Directeurs généraux de l'Inserm.

La structuration spécifique de la recherche en santé publique, mélange de recherche cognitive et de recherche opérationnelle, se développa en 2004 avec Christian Bréchet qui soutint la création d'un Institut virtuel de recherche en santé publique (IVRSP). Ce dernier regroupait les différents acteurs engagés dans la santé publique : des opérateurs de recherche aux caisses d'assurance maladie. En

<sup>12</sup> [<https://histoire.inserm.fr/les-lieux/centre-inserm-le-vesinet>] (consulté le 29 juil. 2019).

<sup>13</sup> Entretien avec Pierre Ducimetière, 19 juin 2002, par S Mouchet et J-F Picard, texte revu et amendé par Pierre Ducimetière, [[https://histoire.inserm.fr/les-femmes-et-les-hommes/pierre-ducimetiere/\(page\)/3](https://histoire.inserm.fr/les-femmes-et-les-hommes/pierre-ducimetiere/(page)/3)] (consulté le 29 juil. 2019).

<sup>14</sup> Site Cermes3, [<http://www.cermes3.cnrs.fr/fr/>], (consulté le 26 nov. 2019).

<sup>15</sup> Site Inserm, [<https://www.inserm.fr/recherche-inserm/recherche-clinique/centres-investigation-clinique-cic>] (consulté le 29 juil. 2019).

<sup>16</sup> Disponibles sur le site iPubli, [<http://www.ipubli.inserm.fr/handle/10608/1>] (consulté le 26 nov. 2019).

2007, à partir d'une initiative de l'Inserm, autour d'Alfred Spira, chercheur à l'Inserm et professeur de santé publique à l'Université Paris Sud 11, l'IRVSP se structura en Groupement d'Intérêt Scientifique « Institut de Recherche en Santé Publique » (GIS-IReSP). Portée par l'Inserm, cette structure développe et promeut aujourd'hui la recherche en santé publique en se fondant sur la communauté scientifique nationale. Parmi ses missions : il y a la définition d'une politique scientifique de recherche en santé publique tenant compte des enjeux de santé nationaux et internationaux, la question de la gestion d'outils collectifs, l'aide à la décision et l'expertise... Christian Bréchet permit, en outre, la création du département de Recherche en santé publique (DRSP), codirigé jusqu'en 2008 par Didier Guillemot et Christophe Tzourio.

En 2009, André Syrota, créa l'Alliance pour les sciences de la vie et de la santé (Aviesan) dont il devint le Président<sup>17</sup>. Il s'agissait de fédérer tous les acteurs de la recherche en sciences de la vie et de la santé. Ce fut au sein de cette alliance que furent créés les instituts thématiques multi-organismes (ITMO), dont l'un fut consacré à la santé publique. Cet institut a été dirigé par Gérard Bréart (directeur de recherche Inserm, Professeur de santé publique UPMC) entre 2008 et 2011, Jean-Paul Moatti (chercheur Inserm, professeur d'économie de la santé, Aix-Marseille) entre 2011 et 2014, Geneviève Chêne (professeur d'épidémiologie et santé publique, Université de Bordeaux) entre 2014 et 2017, et depuis, j'ai l'honneur d'en être la nouvelle directrice.

Sous l'impulsion d'Yves Lévy (2014-2018) qui suivit les recommandations du SAB (*Scientific Advisory Board*) international, une large place fut faite à la santé publique dans le plan stratégique de l'Inserm et dans le contrat d'objectifs et performances signés avec les ministères de la Santé et de la Recherche pour la période 2016-2020<sup>18</sup>.

Grâce aux efforts des communautés et de ses dirigeants successifs depuis plus d'une cinquantaine d'années, l'Inserm occupe aujourd'hui une place unique au niveau national, à l'interface de la recherche fondamentale, de la médecine et de la santé publique.

### *L'ITMO Santé publique et plan stratégique de l'Inserm : un plan de recherche 2016-2020*

La mise en place d'un plan de recherche en santé publique est l'objectif numéro 5 du plan stratégique de l'Inserm 2016-2020<sup>19</sup>. C'est dans ce cadre que les conseillers de l'ITMO santé publique provenant des équipes Inserm mais aussi plus largement d'autres institutions (IRD, INED, CNRS...) et d'autres disciplines (des sciences dites dures quantitatives aux sciences humaines et sociales) et l'IReSP ont travaillé pour définir les priorités stratégiques et les infrastructures, l'animation et faire des propositions sur les financements utiles et nécessaires au déploiement de ce programme.

Ce programme s'est développé avec l'ensemble des communautés mobilisées autour des défis de santé : une mortalité prématurée trop élevée, des inégalités sociales et territoriales de santé, la nécessaire transition organisationnelle qui nous fait passer d'un modèle centré sur la clinique (services hospitaliers, soins aigus, étiologie individuelle, innovations techniques, efficacité) à un modèle centré sur la santé (parcours ambulatoires, prévention et promotion de la santé, prise en charge globale, efficacité) et ainsi de la recherche clinique à la recherche en santé publique ou en population, le tout dans le contexte de la stratégie nationale de santé et du plan de santé publique qui donnent une large priorité à la prévention et l'organisation du système de santé. Quatre axes,

<sup>17</sup> Site Aviesan, [<https://aviesan.fr/>] (consulté le 26 nov. 2019).

<sup>18</sup> Inserm, *Plan stratégique 2016-2020* : 23-24, en ligne sur le site Inserm, [[https://www.inserm.fr/sites/default/files/2017-11/Inserm\\_PlanStrategique\\_2016-2020.pdf](https://www.inserm.fr/sites/default/files/2017-11/Inserm_PlanStrategique_2016-2020.pdf)] (consulté le 29 juil. 2019).

<sup>19</sup> *Ibid.*

cohérents, afin de renforcer la visibilité internationale et la réactivité de la recherche française en santé publique ont été travaillés dans le cadre de ce plan stratégique : déterminants de santé, prévention et recherche interventionnelle en santé des populations ; santé, travail, environnement ; politiques et services de santé ; concepts et méthodes dans les champs émergents ou innovants<sup>20</sup>.

Pour chaque axe, un certain nombre d'actions ont été mises en place. À titre d'exemple, et sans viser l'exhaustivité, notons la labélisation de nouvelles équipes de recherche notamment dans le domaine des sciences humaines et sociales ; des actions coordonnées de structuration des communautés avec l'IRESP ; la participation à la construction d'un plan de recherche en santé publique ; des actions européennes... Aujourd'hui, 72 équipes Inserm, au sein de 27 unités, sont affiliées à l'Institut thématique Santé publique et quatre chercheurs issus de ces équipes siègent au Conseil Scientifique de l'Inserm (mandature 2017-2022).

Mais la santé publique dépasse les frontières de l'Institut de santé publique et les interactions sont multiples avec les autres instituts avec qui nous partageons ces enjeux : l'Institut des technologies de la santé, l'Institut cancer, l'Institut génétique génomique et bio-informatique, l'Institut I3M immunologie, infectiologie et microbiologie, l'Institut physiopathologie, métabolisme et nutrition.

### *L'Inserm et le développement des outils en épidémiologie*

L'alerte et la surveillance de la santé de la population ont acquis, ces dernières années, un développement considérable. C'est en particulier parce que Pierre Ducimetière proposa à la direction générale de la Santé et à l'Inserm, au milieu des années 1980, de reconnaître l'activité des registres de morbidité qui, à la fois, réunissent les informations de surveillance utiles à la santé publique et participent à des travaux de recherche coopérative en santé des populations<sup>21</sup>. Près de 30 ans plus tard, l'Inserm, Santé publique France et l'Institut national du cancer (INCa) s'associent toujours activement aux registres et plus de 60 d'entre eux ont été évalués par le Comité d'Évaluation des Registres.

Avec Gilles Bloch, alors directeur général de la recherche et de la technologie (2006-2009) au ministère de la Recherche, Gérard Bréart et Alfred Spira, ont proposé de privilégier le développement des grandes cohortes en population en tant qu'infrastructures de recherche afin d'étudier certaines pathologies. C'était un choix difficile de cibler ainsi les moyens, et cela a mis des années à se formaliser, jusqu'au Grand emprunt 2009-2010. Aujourd'hui, les 13 cohortes (infrastructure, equipex, cohortes) financées par le Plan Investissement Avenir incluent plus de 600 000 personnes. Ces cohortes sont soit généralistes en population générale, soit sur des pathologies particulières.

### **Innovation et défis futurs de l'épidémiologie et de la santé publique**

Concernant les données de santé, que ce soit dans le cadre de la construction d'une plateforme nationale de données de santé ou dans le cadre de la pérennisation des cohortes mises en place ou du soutien des registres, il faut favoriser un accès large à ces données, ainsi que les appariements aux bases de données médico-administratives tout en respectant le cadre réglementaire mise en place et l'information aux sujets ayant accepté de participer. Les innovations en santé issues du développement de l'intelligence artificielle auront besoin d'être évaluées et suivies, notamment pour mesurer leur innocuité.

<sup>20</sup> *Ibid.*

<sup>21</sup> Discours de Pierre Ducimetière, lauréat du Prix d'honneur 2007 de l'Inserm, 4 déc. 2007, [[https://histoire.inserm.fr/les-femmes-et-les-hommes/pierre-ducimetiere/\(page\)/3](https://histoire.inserm.fr/les-femmes-et-les-hommes/pierre-ducimetiere/(page)/3)] (consulté le 26 nov. 2019).

Une réflexion nouvelle au sein des communautés est engagée : elle vise à organiser l'interface entre les équipes de recherche, les acteurs de santé publique et les décideurs publics, dans un maillage territorial et national, afin de conduire les études, les évaluations et la recherche accompagnant la décision publique. Elle peut être portée et accompagnée à la fois par les équipes de recherche et l'IReSP, qui mixent les partenariats utiles.

Au regard d'une animation structurante des communautés, le modèle des programmes transversaux de l'Inserm peut offrir des opportunités fortes d'innovation pouvant organiser les communautés dans une configuration nationale. Cela pourrait faire l'objet de co-construction avec les partenaires de l'IReSP – dont certains apportent des financements aux enjeux d'organisation du système de santé – et les communautés de recherche, le tout dans une perspective d'aide à la décision publique et aux politiques publiques. Le financement doit pour cela être à la hauteur des enjeux, pour les appels à projets en santé publique comme ils le furent pour structurer la recherche clinique.

Enfin, les enjeux portent aussi sur la formation au métier de chercheur en santé publique et à leur recrutement. Actuellement, les postes de chercheur en santé publique sont soit dans les établissements hospitalo-universitaires, soit à l'Inserm. Dans les hôpitaux universitaires, les possibilités de nomination dépendent des CNU (Conseil National des Universités) des professions médicales (médecine, pharmacie, odontologie), avec un recrutement préférentiel dans les professions médicales. La valence hospitalière de ces postes est à forte connotation soins, recherche clinique et système d'information hospitalier. Il n'existe pas, à ce jour, de CNU spécifique en santé publique pour des scientifiques non issus des professions médicales. Or, la recherche en santé publique a besoin de sortir de l'hôpital et du soin pour faire une large place à la recherche en promotion de la santé ou bien encore sur l'environnement... Elle se nourrit de la pluridisciplinarité, nécessitant d'attirer dans ce champ d'application, non seulement les médecins, mais aussi plus largement les chercheurs en sciences humaines et sociales (sociologue, anthropologue, économiste, démographe, historien...), lesquels pour être reconnus dans leurs propres disciplines doivent rester dans leur champ disciplinaire. Si nous voulons développer la recherche et l'enseignement en santé publique en décroissant les disciplines dans une perspective de coconstruction, une réflexion globale est ainsi à mener sur la structuration du champ.

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

L'auteure tient à remercier Céline Paillette (Comité pour l'histoire de l'Inserm) et Jean-Marie Gagliolo (Institut de Santé Publique, Inserm) pour leur apport à la construction et la rédaction de cet article.

## EPIDEMIOLOGY DURING THE 19<sup>TH</sup> CENTURY

# Epidemiology, Statistics, and Public Health in the Victorian Britain, 1837-1902

**Anne HARDY**

Honorary Professor, Centre for History in Public Health  
London School of Hygiene and Tropical Medicine

### **Abstract**

*The modern history of statistics and epidemiology emerged in nineteenth-century England, following the introduction of demographic measures and concerns about public health in the years before 1850. This paper traces these developments and their consequences across the later nineteenth century, which were reflected in sharply falling death rates by 1900-1920.*

**Keywords :** *census, epidemiology, public health reform, Britain, history of the 19<sup>th</sup> century.*

### **Résumé**

#### **Épidémiologie, statistique et santé publique dans l'Angleterre Victorienne, 1837-1902**

L'histoire contemporaine de la statistique et de l'épidémiologie est apparue en Angleterre au XIX<sup>e</sup> siècle, à la suite de l'introduction de mesures démographiques et du poids croissants des enjeux de santé publique dans les années précédant 1850. Le présent article retrace ses développements et leurs conséquences à la fin du XIX<sup>e</sup> siècle, qui se traduisirent par une forte baisse des taux de mortalité entre 1900 et 1920.

**Mots-clés :** recensement, épidémiologie, réforme de la santé publique, Grande-Bretagne, histoire du XIX<sup>e</sup> siècle.

In 1973, the American academic Peter Stansky edited a book entitled *The Victorian Revolution: Government and Society in Victoria's Britain*, and since then, the concept of the Victorian revolution in government has been explored in numerous ways<sup>1</sup>. Stansky's volume covered political parties, education, high farming, the Civil Service and the intelligentsia, but made no mention of public health, local government, or statistics. Yet an important part of this profound revolution in Britain's style of government lay in the introduction of measures designed to control rapidly deteriorating urban and rural environments, and the declining quality of life and health of a rapidly increasing urban population. The tools for measuring these developments had been put in place before Victoria ascended the throne, and played a crucial part in stimulating the wide-reaching social reforms which became so characteristic of the Victorian era. The year 1801, saw the introduction of the first population census, to be repeated every ten years into the future, but it was not until 1837, the year Victoria came to the throne, that Civil Registration, the official recording of all births, marriages, deaths, and, importantly, causes of death, also became law.

## The Victorian Revolution in the government of public health

### *The importance of the census and civil registration*

These two tools for the measurement of population, the census and civil registration, made three things clear by 1851: that the size of the population of England and Wales had doubled since 1801, that population growth was most rapid in the industrial cities and in London, and that the principal causes of high death rates and premature deaths in these cities were infant mortality, and crowd diseases – smallpox, the childhood infections, fevers, diarrhoea, and tuberculosis. Already in 1839, the Health of the Towns Association had been established to try to combat some of these problems, and in 1850 the Epidemiological Society of London was created to study the problem of disease. English sanitarians were already studying European manifestations of cholera with interest, since the movement of the disease across the continent appeared to follow trade routes and lines of communication<sup>2</sup>. In 1832, 1848 and 1854, epidemic cholera at Hamburg was rapidly followed by the appearance of the disease in seaports on the east coast of England<sup>3</sup>. Meanwhile the accumulating evidence of the annually published cause-of-death statistics suggested that unsavoury environmental conditions were responsible for excess and premature deaths in British towns and cities. It was Edwin Chadwick's concern for the economic consequences of these premature deaths that led to the

<sup>1</sup> Stansky P. ed. *The Victorian Revolution: Government and Society in Victoria's Britain*. New York : NewViewpoints (sic), 1973.

<sup>2</sup> Hauser P. *Le choléra en Europe depuis son origine jusqu'à nos jours*. Paris: Société d'éditions scientifiques, 1897: 533.

<sup>3</sup> Cowper W. Precautionary advice to Local Boards of Health, with reference to Epidemic Cholera. *The Lancet*, 1857, II: 374-75.

publication of his landmark *Report on the Sanitary Condition of the Labouring Classes of Great Britain* in 1842<sup>4</sup>.

Chadwick's publication, although it is commonly agreed that it sparked the Victorian public health revolution, was not its only driver. The real spur to these highly important local government reforms was statistical cause-of-death information on cholera. Cholera visited England in 1832, 1848, 1854, and to a much lesser degree in 1866. The 1854 cholera outbreak was the occasion of a classic moment in Victorian public health history, when John Snow identified the source of a local cholera outbreak in Soho, as drinking water taken from the Broad Street pump. The local death records showed cholera deaths clustered in the streets about the pump, but the positive proof lay in a single solitary cholera death registered in the wealthy northern district of Hampstead on the furthest fringe of the city. The victim was a woman who liked the taste of the Broad Street water so much that she had a flask sent up to her house every day<sup>5</sup>. It was this statistical outlier that – in the age before bacteriology – proved the case against the Broad Street pump.

### *Local Government and the national organisation of public health*

It is not too extreme to say that the episode of the Broad Street pump began the British public health revolution. Although the City of London – London's world-famous financial district – had taken the innovative step of appointing a Medical Officer of Health in 1848, to supervise and improve sanitary arrangements and living conditions within its bounds with a view to reducing or controlling disease outbreaks, no other London authority had followed its example. But London was too important to be left at the mercy of epidemic disease, and in 1855 Parliament passed the Metropolis Local Management Act, effective from the beginning of 1856. Under this legislation, London was divided up into 25 districts, governed by rate-payer elected civil vestries, and each was required to appoint a Medical Officer of Health with sanitary inspectors in support, to improve sanitary conditions and to monitor, prevent or contain, outbreaks of infectious disease<sup>6</sup>. Cause of death information initially provided the knowledge base for these operations, but in 1889 the Infectious Diseases Notification Act considerably widened and improved that base by making notification of cases of the principal infectious diseases compulsory on doctors over the whole country. The system of public health management imposed on London in 1855 was extended to the whole of England and Wales in 1872. This national organisation consisted in urban and rural district councils, each of which was obliged to appoint a Medical Officer of Health with appropriate staff, as had been done in London. In 1889 a further organisational tier, independent of local government, was added at County level.

The 1872 Public Health Act did not merely establish a system of government responsible for managing local public health issues across England and Wales, it also created a Central Medical Department, under the newly established Local Government Board. This department was headed by the Chief Medical Officer, with a supporting staff of medical inspectors and a suitable secretariat. The Department's annual reports, including accounts of investigations undertaken by individual medical inspectors, were published annually among the Official (Parliamentary) Papers issued by successive governments. These investigations brought a new level of connectedness into epidemiological investigation in England and Wales. With information garnered country-wide from

<sup>4</sup> Chadwick E. *Report on the Sanitary Condition of the Labouring Population of Great Britain*. Flinn M W ed. Edinburgh: Edinburgh University Press, 1965.

<sup>5</sup> Snow J. *On the Mode of Communication of Cholera*. London: John Churchill, New Burlington Street, 1855, 2<sup>d</sup> ed.

<sup>6</sup> Hardy A. Public health and the expert: The London Medical Officers of Health, 1856-1900. In Mcleod R M ed. *Government and Expertise: Specialists, Administrators and Professionals, 1860-1919*. Cambridge: Cambridge University Press, 1998: 128-42.

registrars of deaths, from the weekly record of notifications of infectious disease (from 1889), and from local medical officers of health, the staff of the Central Medical Department travelled country-wide to investigate complex or puzzling disease incidents and occurrences, often involving more than one geographical area.

## Investigating diseases

### *Searching for the source of infection on the ground*

One such report, dating from 1890, provides an example of the quality and style of these investigations<sup>7</sup>. By the 1890s, cholera, typhoid and diarrhoea were known to be, mostly, caused by drinking water contaminated by human excrement containing the bacteria which caused these diseases. The 1890 report was concerned with the drinking water supply of County Durham, in the north-east of England, where the statistics indicated that an outbreak of typhoid had occurred over a wide area north and south of the River Tees in the six weeks from 7 September to 18 October. Since the river supplied the local drinking water, it was assumed that the river water was the source of infection, and in November Medical Inspector Frederick William Barry was sent from London to investigate. He started out from the market town of Darlington, to follow the course of the Tees towards its source. On his journey, during which he crossed and re-crossed the river, he explored numerous hamlets, villages, and farmsteads before reaching the pretty market town of Barnard Castle. Barnard Castle lies on the Durham side of the river, facing the villages of High and Low Stainsforth on the Yorkshire side. The area was already popular with holiday makers during the summer months, but Barry's eyes were not for the grandeur of the scenery. His mission was to discover the extent to which this natural environment was polluted by human sewage. As he tracked the drainage condition of houses, farms, and settlements, and the fall of drains, sewers, and open ditches into the river, his concern mounted. On reaching Barnard Castle, he was appalled to find that all sewage conduits emptied their contents into the river, and that layers of solid sewage lay deep on the foreshore. Never in his whole career, Barry noted, had he seen 'such a mass of stinking abominations' as lay alongside the Tees at Barnard Castle at the time of his visit.

### *Using statistics and census data*

Barry's rural excursion was not unusual for a Medical Inspector. Sanitary improvements had considerably reduced typhoid incidence in urban areas by the 1890s, and public health investigators were increasingly paying attention to rural conditions as the source of urban pollution and disease. The typhoid outbreak which Barry was investigating occurred over ten Durham registration districts, but before he finished work, the area was hit by a second outbreak, which lasted from 28 December 1890 until 7 February 1891. Some 3,000 cases were reported to have occurred in the combined outbreaks. As a result of the second outbreak, Barry's report on the earlier epidemic was delayed, and he was able to make use of the 1891 population census data in constructing his final analysis. His conclusions were stark: pollution of the Tees, aggravated by heavy rainfall and flooding, which had washed the filthy foreshores clean, and the consumption of the polluted Tees water by local residents, had generated both outbreaks. During those two epidemic periods, fifteen typhoid cases had occurred among Tees water drinkers for every one among those who did not drink Tees water.

<sup>7</sup> The following account of Barry's investigation is taken from Hardy A. *Salmonella Infections, Networks of Knowledge and Public Health in Britain, 1880-1975*. Oxford: Oxford University Press, 2015: 21-3.

### *The heyday of observational epidemiology at the turn of the century*

The evidence for Barry's indictment of the Tees water supply was circumstantial, but was none the less accepted by his contemporaries. The investigation was a classic of late Victorian observational epidemiology: a broad and detailed survey of the terrain in which the unusual typhoid outbreak had occurred. Barry gave a detailed statistical analysis of the outbreaks and of their incidence in the ten registration districts involved. The sanitary circumstances of the districts were described in detail (their geographical size, population size, occupational and class structure, number of houses with their rateable value and sanitary condition, the state of the roads, the nature of the local economy and the distribution of disease), and the local arrangements for the disposal of human excrement and refuse carefully recorded. Drinking water received particular attention: it was for this purpose that Barry undertook his up-river journey towards the source of the Tees. He gave the history of the local corporation water undertakings since 1870 in full, concluding that their deficiencies were responsible for the outbreaks.

Barry's 1890 excursion to Durham illustrates very neatly the use which late Victorian epidemiologists made of statistics. Registered deaths and causes of death informed local authorities that an epidemic was in progress. In a difficult case, or where, as in Durham, several local authority areas were involved, appeal would be made to the Medical Department at the Local Government Board for an experienced epidemiologist to investigate the problem. Taking into account the scale of the problem and the nature of the disease involved, an appropriate investigation would follow – with typhoid, as here, by tracking the water sources back from the site or location of the epidemic to origin, to identify the pollution access point and the source of that pollution. In all these investigations, whatever the disease involved, statistics recording cases and deaths from a particular disease in a particular local population provided the key outbreak identification data<sup>8</sup>. In the 1890s, this method of investigation was in its heyday. By the early 1900s, the new science of bacteriology was beginning to come between the statistics and the epidemiology<sup>9</sup>.

None the less, the purely statistical investigative epidemiology of the Victorian period proved extremely effective in resolving local public health problems as they related to infectious diseases. The statistics provided the momentum for compulsory infant vaccination against smallpox, begun in 1851, for the establishment from 1867 of isolation hospitals for cases of infectious disease, and in 1870 for the construction of the network of port sanitary authorities, to police the sanitary condition of shipping, and to prevent the import of dangerous infection diseases from abroad. All these measures were inspired by the information base, both statistical and geographical, which cause-of-death information provided<sup>10</sup>. From that database, the whole programme of preventive action was built up over the period between 1855 and 1900. By the end of Victoria's reign Britain's preventive structure encompassed administrative, environmental and legal provisions, and its effectiveness was demonstrated by falling death rates. Crude death rates for England and Wales stood at 22.4 per thousand living in the decade 1841-50, by 1891-1900 they stood at 18.2, and by 1911-20 at 14.0. If the real momentum of improvement in the people's health did not show itself until the first two

<sup>8</sup> These reports are contained in *British Parliamentary Papers: Reports of the Medical Officer to the Privy Council, 1859-1870* ; *Reports of the Medical Officer to the Local Government Board, 1872-1902*.

<sup>9</sup> Hardy A. On the Cusp: Epidemiology and Bacteriology at the Local Government Board, 1890-1905. *Medical History*, 1988, 42: 328-34.

<sup>10</sup> For this information, consult the annual and decennial reports of the Registrar-General for England and Wales, available in the published *British Parliamentary Papers series*.

decades of the twentieth century, it was the statistically and epidemiologically driven public health measures put in place during Victoria's reign that brought about that dramatic demographic achievement.

*The author declares that she has no conflicts of interest.*

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

## GRIPPE ESPAGNOLE

# Mort et résurrection du virus de la grippe espagnole

**Patrick BERCHE**

Professeur émérite de l'Université Paris Descartes

### Résumé

Décrypter le génome du virus de la grippe espagnole est un enjeu majeur pour comprendre la gravité de cette pandémie et prédire l'occurrence de nouvelles pandémies. On a pu identifier et ressusciter le virus léthal H1N1 de 1918. Ces remarquables progrès posent des problèmes éthiques concernant certaines manipulations génétiques d'agents infectieux avec « gain de fonction ».

**Mots-clés** : pandémie de grippe de 1918-1919, H1N1, génome viral, histoire du XX<sup>e</sup> siècle, Johan Hultin.

### Abstract

#### **Death and resurrection of the Spanish flu virus**

*Deciphering the genome of the Spanish flu virus is a major challenge to understand the severity of this pandemic and predicting the occurrence of new pandemics. The 1918 lethal virus H1N1 has been identified and resurrected. This remarkable progress raises ethical issues with respect to certain genetic manipulations of infectious agents with « gain of function ».*

**Keywords** : influenza pandemic flu 1918-1919, H1N1, viral genome, history, 20<sup>th</sup> century, Johan Hultin.

Les études historiques montrent que, depuis le XVI<sup>e</sup> siècle, on enregistre environ trois pandémies par siècle, dont certaines déciment les populations. Pourquoi ces pandémies et quels sont les mécanismes de leur émergence ? Les recherches sur le virus de la grippe espagnole ont fait progresser les connaissances sur ce fléau majeur, qui demeure une menace majeure pour l'avenir<sup>1</sup>.

### La grippe espagnole, un fléau énigmatique

La grippe espagnole de 1918 fut une terrible pandémie, clôturant la Première Guerre mondiale. Elle aurait fait entre 50 et 100 millions de victimes dans le monde et, en Europe, elle entraîna plus de morts que la Grande Guerre. Les premiers cas de grippe apparurent en mars 1918 dans le comté de Haskell dans le Kansas, aux États-Unis. L'Amérique qui venait d'entrer en guerre en 1917, aux côtés des Alliés, avait organisé un réseau de camps militaires d'entraînement pour accueillir des centaines de milliers de jeunes recrues volontaires, en partance vers la France. La grippe se répandit à travers ce réseau de bases militaires dans tous les États-Unis, puis en Europe et dans l'ensemble du monde, en quelques mois.

La pandémie évolua en deux vagues successives. La première dura de mars à août 1918, et fut relativement bénigne, associée à une mortalité de 0,1 %. La seconde beaucoup plus grave a sévi de septembre 1918 jusqu'au printemps 1919, entraînant 2 à 4 % de mortalité. La mort survenait parfois en deux à trois jours du fait d'une pneumonie virale massive, d'autres fois en dix à douze jours, à cause d'une surinfection pulmonaire bactérienne. Dans les années qui ont suivi, la grippe est devenue saisonnière avec une faible mortalité.

Pendant longtemps, la pandémie de la grippe espagnole qui s'abattit sur des populations épuisées par la guerre, demeura une énigme. Une première question était liée à son épidémiologie. Contrairement à la grippe saisonnière qui frappait les âges extrêmes de la vie, la grippe espagnole toucha avec prédilection les adultes jeunes entre 20 et 40 ans, qui en furent les principales victimes. Cela est probablement dû aux circonstances très particulières de la guerre de 14-18, qui engendrèrent des concentrations de jeunes soldats vivant en grande promiscuité dans les camps et les tranchées. Une autre question fut celle de l'agent responsable. Depuis les années 1890, on croyait – à tort – que la grippe était due au bacille de Pfeiffer, aujourd'hui appelé *Haemophilus influenzae*. Il fallut attendre 1933, pour que Christopher Andrewes, Wilson Smith et Patrick Laidlaw découvrent, à Londres, le virus responsable, en réussissant à l'inoculer au furet, puis en le cultivant sur œuf embryonné. Les deux principaux facteurs de virulence associés à l'enveloppe du virus, l'hémagglutinine et la neuraminidase, ont été découverts respectivement en 1941 et en 1949. On appela désormais ce virus H1N1. Son génome décrypté par séquençage est constitué de 13 500 nucléotides d'ARN. Il est fragmenté en huit segments portant chacun les différents gènes du virus, ce qui lui confère une importante flexibilité génétique. De nombreuses recherches sur le terrain et en

<sup>1</sup> Berche P. *Faut-il avoir peur de la grippe ? L'histoire des pandémies*. Paris : Odile Jacob, 2012.

laboratoire ont permis de montrer que les virus pandémiques apparaissent par recombinaison de ces segments provenant de différentes souches virales. Cela survient chez les hôtes domestiques intermédiaires infectées par plusieurs virus, tels que le canard et le porc. On s'est aussi aperçu que les oiseaux sauvages formaient le réservoir naturel des virus de la grippe.

Comprendre l'extrême virulence de certains virus pandémiques est un enjeu important pour prévenir les risques futurs associés aux virus de la grippe. Pour cela, il fallait retrouver le virus de la grippe espagnole de 1918.

## La recherche du virus de la grippe espagnole dans le permafrost

Au décours de la Seconde Guerre mondiale, l'idée fit son chemin que le virus de la grippe espagnole pourrait survivre dans le corps des victimes inhumées dans le permafrost du Grand Nord arctique. Le fléau avait décimé les populations inuites très sensibles au virus, avec des taux de mortalité avoisinant 90 %. Les restes des nombreuses victimes reposaient dans les cimetières des villages en Alaska. En 1951, Johan Hultin, jeune étudiant suédois travaillant à l'université de l'Iowa aux États-Unis, entreprit ainsi de consacrer sa thèse de recherche à l'isolement du virus disparu, à partir de prélèvements pulmonaires humains effectués sur des patients décédés en 1918 et inhumés en Alaska. Malgré toutes les précautions, il échoua à cultiver le virus. Il ne put soutenir sa thèse et abandonna sa carrière de chercheur.

Quelques décennies plus tard, entré en scène un médecin militaire, Jeffery Taubenberger, qui avait rejoint un nouveau laboratoire appartenant à l'*Armed Forces Institute of Pathology*, organisme dépendant du *Walter Reed Army Medical Campus*, au nord-est de Washington. Son premier travail avait porté sur l'identification de l'agent responsable d'une épidémie mystérieuse chez les dauphins et les phoques. En 1995, utilisant une technique puissante d'amplification de l'ADN, la PCR (*Polymerase chain reaction*<sup>2</sup>), il avait réussi à identifier, à partir de cadavres de dauphins en décomposition, deux morbillivirus inconnus, proches du virus de la rougeole. Encouragé par ses résultats et bien que ne connaissant rien du virus de la grippe, Jeffery Taubenberger décida de tenter d'amplifier le génome du virus de la grippe espagnole par amplification génique. Pour cela, il exploita les ressources du repositoire de Forest Glen (Maryland) où sont conservés, depuis la Guerre de Sécession, des millions de prélèvements d'origine humaine. On pouvait y trouver notamment des blocs de paraffine contenant des fragments de tissus pulmonaires, utilisés pour les préparations microscopiques des anatomopathologistes. Jeffery Taubenberger sélectionna plusieurs blocs provenant de soldats morts de la grippe espagnole en septembre et octobre 1918. Après un travail intense, il réussit, le 23 juillet 1996, à amplifier, par RT-PCR<sup>3</sup>, des fragments d'ARN viral à partir des prélèvements d'un jeune soldat, Roscoe Vaughan, tombé malade le 17 septembre 1918, et décédé le 26 septembre, à 21 ans. Environ 15 % du génome viral furent alors déterminés et les résultats furent publiés le 21 avril 1997 dans *Science*.

## Le génome viral décrypté

À cette époque, Johan Hultin avait 72 ans et passait une retraite paisible à San Francisco. Découvrant l'article de Jeffery Taubenberger, il fut à la fois enthousiaste et bouleversé, se remémorant le travail de la thèse qu'il n'avait jamais pu achever. Il lui écrivit pour lui faire part des

<sup>2</sup> La PCR (*Polymerase Chain Reaction*) est une méthode permettant l'amplification des millions de fois de quantités infimes (picogrammes) d'ADN *in vitro*, en utilisant des amorces spécifiques et une enzyme (Taq polymérase).

<sup>3</sup> La RT-PCR (*Retro-Transcriptase-polymerase Chain Reaction*) consiste à transformer l'ARN en ADN par une *rétrotranscriptase*, puis à réaliser une amplification par PCR.

recherches qu'il avait menées en 1951, se déclarant prêt à aller effectuer de nouveaux prélèvements, à ses propres frais, en Alaska. Début août 1997, Johan Hultin prit un vol commercial pour Anchorage, suivi d'un autre pour Nome, puis monta à bord d'un petit avion pour rejoindre Brevig, village d'environ 265 âmes. Là, il obtint la permission du Conseil du village pour réaliser des prélèvements au cimetière. Il exhuma la dépouille d'une jeune femme inuite de 18 ans, presque intacte, décédée en novembre 1918. Il la baptisa Lucy, « Lumière », nom du célèbre squelette de 3 millions d'années, découvert en Éthiopie. Les prélèvements furent envoyés à Jeffery Taubenberger qui réussit à amplifier les fragments du virus de 1918. Ainsi, la séquence complète du génome du virus de la grippe espagnole put être déterminée<sup>4</sup>. Finalement, Jeffery Taubenberger réussit à séquencer 5 isolats du virus de la grippe espagnole à partir de Lucy et de quatre patients décédés de septembre 1918 à février 1919 aux États-Unis et en Angleterre. Les cinq séquences étaient strictement identiques, à 3 nucléotides près sur 13 500 nucléotides. Cela confirme l'unicité du virus de la grippe espagnole au cours de la pandémie<sup>5</sup>.

### La résurrection du virus de la grippe espagnole

En 2005, plusieurs équipes du *Center for Diseases Control and Prevention* (CDC) et de l'*US Department of Agriculture* (USDA), sous l'égide du microbiologiste Terence Tumpey, ressuscitèrent le virus H1N1 disparu depuis 1918. Les chercheurs ont d'abord reconstitué l'ADN viral synthétisé *in vitro* à partir de la séquence du virus de la grippe espagnole, puis l'ont introduit dans des cellules préalablement infectées par un virus H1N1 inoffensif. L'ADN viral a été transcrit en ARN dans ces cellules, permettant ainsi de fabriquer quelques rares virus très virulents. Ces virus ainsi reconstitués ont été sélectionnés par passages chez la souris, recouvrant ainsi le virus H1N1 de la grippe espagnole<sup>6</sup>. Le virus fut ainsi reconstruit, réassemblé, puis testé dans un laboratoire de haute sécurité du *Southeast Poultry Laboratory Research* (USDA).

En dix ans, le redoutable virus H1N1 a été identifié et ressuscité. Il s'est avéré 40 000 fois plus virulent pour les singes que les virus H1N1 de la grippe saisonnière, confirmant son extrême dangerosité. Il s'agit en réalité de la fabrication d'une arme biologique redoutable, qui est aujourd'hui stockée quelque part dans un congélateur d'un laboratoire militaire de l'*Armed Forces Institute of Pathology*.

### Enjeux de sécurité et d'éthique

Les origines du virus de la grippe espagnole demeurent cependant incertaines. Les données de séquence suggèrent qu'il n'est pas directement apparenté à une souche aviaire, mais qu'il aurait probablement transité par un hôte mammifère intermédiaire. En 2009, on a montré à partir des séquences d'un grand nombre de souches H1N1 de diverses origines, que le virus de la grippe espagnole aurait circulé chez le porc et chez l'homme depuis au moins 1911. La souche pandémique proviendrait d'un réassortiment entre une souche humaine circulante et des virus de mammifères (porcs, chevaux...).

<sup>4</sup> Taubenberger J *et al.* Initial genetic characterization of the 1918 "Spanish" influenza virus. *Science*, 1997, 275 :1793-96 ; Taubenberger J *et al.* Characterization of the 1918 influenza virus polymerase genes. *Nature*, 2005, 437 : 889-93.

<sup>5</sup> Taubenberger J, Reid A, Janczewski T *et al.* Integrating historical, clinical and molecular genetic data in order to explain the origin and virulence of the 1918 Spanish influenza virus. *Philosophical Transactions of the Royal Society of London, Series B*, 2001, 356 : 1829-39.

<sup>6</sup> Tumpey T *et al.* Characterization of the reconstructed 1918 Spanish influenza pandemic virus. *Science*, 2005, 310 : 77-80.

Reste à comprendre pourquoi le virus est devenu soudainement très virulent en septembre 1918 et cela à peu près partout dans le monde au même moment ? Il n'a pas été possible jusqu'ici d'identifier des séquences virales provenant de la première vague de la pandémie, du fait de sa faible mortalité. Il est hautement probable qu'on soit en présence du même virus lors des deux vagues : les soldats infectés lors de la première étaient totalement résistants à l'assaut mortel de la seconde. Il est plausible qu'il s'agisse d'une mutation ponctuelle du virus, comme on l'a observé lors d'épidémies survenues dans des élevages de volailles déclenchées par les virus de la grippe aviaire, au départ peu virulents, puis brutalement létaux du fait de mutations ponctuelles.

Parallèlement, une controverse est apparue récemment à propos de manipulations génétiques sur un virus H5N1 de la grippe aviaire. Ce virus est extrêmement virulent pour l'homme avec un taux de mortalité pouvant atteindre 50 %, mais il n'y a pas de contagiosité interhumaine. En créant certaines mutations du H5N1, il a été possible de le rendre contagieux chez un mammifère, le furet<sup>7</sup>. Malgré les progrès remarquables enregistrés ces dernières années, les travaux sur le virus de la grippe posent d'importants problèmes éthiques et de droit international sur le respect par les signataires des traités d'interdiction des armes biologiques<sup>8</sup>.

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

---

<sup>7</sup> Imai M, *et al.*, Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets, *Nature*, 2012 ; 486 : 420-428.

<sup>8</sup> Korn H, Berche P, Binder P, *Les menaces biologiques : Biosécurité et responsabilité des scientifiques*. Paris, PUF, 2008.



## 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.



## ÉPISTÉMOLOGIE DE L'ÉPIDÉMIOLOGIE

# La réflexion épistémologique en épidémiologie dans la seconde moitié du XX<sup>e</sup> siècle

**Joël COSTE**

Professeur de santé publique à l'université Paris Descartes

Directeur d'études à l'École Pratique des Hautes Études (Histoire de la médecine)

### Résumé

C'est dans les décennies 30-40 du XX<sup>e</sup> siècle que l'épidémiologie a pris le visage scientifique qu'on lui connaît actuellement, soutenue par une réflexion méthodologique mais aussi épistémologique qui s'est poursuivie jusqu'à aujourd'hui. L'article propose un aperçu de cette réflexion dans la seconde moitié du XX<sup>e</sup> siècle, notamment sur les questions fondamentales de la validité des études et de la causalité.

**Mots-clés** : épidémiologie, histoire, épistémologie, validité, causalité.

### Abstract

*Epistemological thinking in epidemiology in the second half of the 20<sup>th</sup> century*

*During the 1930s and 1940s, epidemiology took a scientific turn, supported by both a methodological and an epistemological reflection which continues to the present time. This article provides an overview of this reflection in the second half of the 20<sup>th</sup> century, particularly in relation to issues of study validity and causality.*

**Keywords** : *epidemiology, history, epistemology, validity, causality.*

Au XX<sup>e</sup> siècle, l'épidémiologie – l'étude de la distribution des maladies et de leurs déterminants chez l'homme<sup>1</sup> – a connu un essor remarquable et a acquis le statut de discipline scientifique grâce à une standardisation et une mathématisation accrues de ses méthodes. Elle a permis d'obtenir des connaissances d'une finesse inconnue jusque-là sur la fréquence de nombreuses maladies, dans les populations les plus variées. L'épidémiologie a aussi permis de mettre en évidence des variations courtes de la fréquence des maladies et de mettre en œuvre, avec une efficacité inédite, l'ancien projet de contrôle des maladies. La régression de nombreuses maladies infectieuses, puis, à la fin du XX<sup>e</sup> siècle, de plusieurs maladies chroniques (les maladies vasculaires, certains cancers) dans les populations occidentales est en partie redevable à des mesures préventives *mieux ciblées* grâce aux résultats d'investigations épidémiologiques fines : descriptives (identifiant des sous-populations à risque élevé), étiologiques (identifiant des facteurs de risque) ou évaluatives (confirmant l'efficacité des mesures préventives avant leur diffusion)<sup>2</sup>. Cet essor de l'épidémiologie s'est inscrit dans une continuité historique qui remonte à l'Antiquité grecque et dont on trouve des développements importants au Moyen Âge et à l'Époque Moderne. C'est toutefois au XIX<sup>e</sup> siècle que l'épidémiologie prit une orientation nettement quantitative, à la suite de la rencontre du mouvement statistique (initié par William Petty et amplifié au XVIII<sup>e</sup> siècle avec l'arithmétique politique<sup>3</sup>) avec le calcul des probabilités, développé notamment par Pierre-Simon de Laplace. Une seconde vague d'innovations, situable pour son début dans les décennies 1930-1940, a associé l'inventivité mathématique et l'inventivité méthodologique : des schémas d'enquête reposant sur les méthodes de comparaison contrôlée qui avaient été développées en biologie ont été introduits en recherche médicale et ont rapidement constitué des outils pertinents de l'étude tant des traitements (l'essai clinique randomisé) que de l'étiologie des maladies (les enquêtes étiologiques d'observation)<sup>4</sup>. Une réflexion épistémologique originale a été stimulée par cette introduction, en particulier sur les questions fondamentales de la validité des études et de la causalité.

## La validité des études

### *Austin Bradford Hill et la randomisation*

Parmi les pionniers de l'introduction en médecine des méthodes de comparaison contrôlée utilisées dans des disciplines biologiques comme l'agronomie et la génétique, certains développèrent une

<sup>1</sup> McMahon B, Pugh TF, Ipsen J. *Epidemiologic Methods*. Boston : Little, Brown, 1960.

<sup>2</sup> Coste J. Les maladies dominantes au XX<sup>e</sup> siècle. In Fantini B, Lambrichs LL dir. *Histoire de la pensée médicale contemporaine. Évolutions, découvertes, controverses*. Le Seuil : Paris, 2014 : 259-78.

<sup>3</sup> Martin T. dir. *Arithmétique politique dans la France du XVIII<sup>e</sup> siècle*. Paris : INED, 2003.

<sup>4</sup> Leplège A, Bizouarn P, Coste J. dir. *De Galton à Rothman. Les grands textes de l'épidémiologie au XX<sup>e</sup> siècle*. Paris : Hermann, 2011 ; Coste J, Leplège A. Pour l'épistémologie et l'histoire de l'épidémiologie. *Revue d'Épidémiologie et de Santé Publique*, 2009, 57, 5 : 317-8.

réflexion épistémologique sur les conditions de validité des études. Dans ses écrits des années 1950 sur l'essai clinique, le statisticien anglais Austin Bradford Hill (1897 – 1991) capitalisait l'expérience britannique en matière d'essais cliniques, illustrée quelques années plus tôt par l'éclatant succès de l'essai évaluant la streptomycine dans la tuberculose, publié en 1948<sup>5</sup>, qui contrastait avec l'échec de l'essai américain, abandonné suite aux réticences des médecins cliniciens à constituer un groupe contrôle<sup>6</sup>. Le succès de l'essai anglais tenait largement à la bonne réalisation de la randomisation, dont Augustin Bradford Hill avait eu la responsabilité. Ce dernier attendait de la randomisation tout autre chose que son aîné, « le fondateur de la statistique moderne »<sup>7</sup>, Ronald Aylmer Fisher (1890 – 1962) qui la promouvait comme la méthode assurant la validité des tests statistiques<sup>8</sup>. Il en attendait, quant à lui, la neutralisation des « biais personnels » dus aux interventions des investigateurs. Comme il l'exprime clairement dans « La philosophie de l'essai clinique » écrit en 1953<sup>9</sup>, la randomisation garantit trois choses : 1) que les préférences personnelles des médecins chargés d'inclure les malades dans l'essai n'interviennent pas dans la constitution des groupes de traitement ; 2) que leurs efforts pour surmonter ces préférences n'introduisent pas de nouveaux biais ; et 3) que des critiques extérieurs ne puissent mettre en doute le caractère complètement aléatoire des affectations de patients dans les groupes. Aucune autre procédure n'apporte en effet une garantie suffisante contre les interventions des cliniciens. Et c'est bien ce que les nombreux essais défectueux réalisés auparavant avaient montré : si l'on confie aux cliniciens la répartition des patients, ils interviennent, leurs préférences se manifestent et les groupes ne sont plus comparables. Seule une stricte randomisation permet de confier au hasard cette répartition, évitant ainsi les biais de sélection.

### *Les limites de l'essai randomisé*

Rapidement, l'essai randomisé devint la méthode de référence dans l'étude des traitements, non sans réticence ou résistance de larges secteurs de la médecine (chirurgie, médecine physique, psychothérapies...) et il fut même placé au sommet de la hiérarchie des preuves par les tenants de l'*Evidence Based-Medicine* dans les années 1990<sup>10</sup>. Très rapidement aussi, des auteurs questionnèrent la validité externe des essais réalisés, alors que les patients trop graves, trop âgés, présentant des pathologies associées étaient généralement exclus, et que les traitements étaient beaucoup mieux pris ou réalisés que dans la « vraie vie ». Il a été aussi montré que la randomisation n'est pas toujours nécessaire ni suffisante pour réaliser des tests d'hypothèse ou établir un lien de cause à effet, et qu'elle ne peut garantir la comparabilité stricte entre les groupes ni le contrôle des variables de confusion à moins que l'essai porte sur des effectifs considérables. Dans l'importante réflexion sur les limites de l'essai randomisé<sup>11</sup>, l'article publié par Daniel Schwartz et Joseph Lellouch dans le *Journal of Chronic Diseases* en 1967<sup>12</sup> se distingue particulièrement. Les auteurs y

<sup>5</sup> *Streptomycin Treatment of Pulmonary Tuberculosis. A Medical Research Council Investigation, BMJ, 1948, 2 : 769-82.*

<sup>6</sup> Marks HM. *The Progress of Experiment : Science and Therapeutic Reform in the United States, 1900-1990.* New York : Cambridge University Press, 1997 : 98-127.

<sup>7</sup> Rao CR. RA Fisher. The founder of modern statistics. *Statistical Science, 1992, 7: 34-48.*

<sup>8</sup> Armitage P. Fisher, Bradford Hill and randomization. *International Journal of Epidemiology, 2003, 32, 6: 925-8.*

<sup>9</sup> Hill AB. *Statistical Methods in Clinical and Preventive Medicine.* Washington DC: Institutes of Health Annual Lectures, 1953, chap. 1.

<sup>10</sup> Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA, 1992, 268, 17 : 2420-5.*

<sup>11</sup> Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. *Social Science in Medicine, 2018, 210 : 2-21.*

<sup>12</sup> Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutical trials. *Journal of Chronic Diseases, 1967, 20, 8 : 637-48.*

abordent une série de problèmes posés par ces essais, dans une perspective non seulement technique ou analytique, mais aussi conceptuelle et philosophique. Ils montrent qu'il faut distinguer deux attitudes pour « deux types de problèmes radicalement différents » : une « attitude explicative, dont l'objectif est la compréhension des phénomènes » et une « attitude pragmatique, dont l'objectif est la décision ». Dans la première, ils soulignent :

On cherche à découvrir s'il existe une différence entre deux traitements définis de manière stricte et habituellement simple ; leurs effets sont évalués, avec des critères ayant une signification biologique, dans une population de sujets définie de manière relativement arbitraire, mais qui est la plus propice à révéler une différence susceptible d'exister ; les méthodes statistiques utilisées pour déterminer le nombre de sujets nécessaire et pour analyser les résultats doivent minimiser les risques d'erreur de première et de deuxième espèce<sup>13</sup>.

Dans la seconde, ils indiquent :

On cherche à répondre à la question : lequel des deux traitements doit-on choisir ? Les traitements sont définis de façon souple et sont habituellement complexes : ils prennent en compte les traitements associés et les abandons. Leurs effets sont évalués avec des critères prenant en compte les intérêts des patients et les coûts au sens large. La population des patients étudiés est bien déterminée : ceux chez qui les conclusions de l'essai seront extrapolées. Les méthodes statistiques utilisées doivent limiter le risque de troisième espèce (celui de choisir le traitement le moins bon)<sup>14</sup>.

Il fallut toutefois attendre près de 30 ans et le milieu des années 1990 pour que les propositions de Daniel Schwartz et Joseph Lellouch dans leur article pionnier, sinon visionnaire, soient entendues et que le terme « pragmatique » soit utilisé pour une série d'essais randomisés portant sur des thérapeutiques ou des interventions complexes (éducatives, physiques, etc.) ou collectives ou réalisés en pratique courante ou de médecine générale : les « real world trials »<sup>15</sup>.

### *La validité des études étiologiques*

La question de la validité des études se posa avec plus de force encore pour les enquêtes étiologiques d'observation. Après une première phase où différentes sources d'erreurs (les *biais*) de ces études ont été décrites (ce fut la « période des catalogues » des biais en épidémiologie comme l'a appelée Paolo Vineis<sup>16</sup> avec, entre autres, les 8 sources d'erreur d'Edmond A. Murphy<sup>17</sup> en 1976, les 35 types de biais de David L. Sackett<sup>18</sup> en 1979), une première clarification conceptuelle a été opérée par Olli S. Miettinen en 1985<sup>19</sup>, qui distingua trois grands types de biais en miroir des trois caractéristiques principales de l'essai thérapeutique 1) le contrôle de l'exposition (par un placebo ou autre) et son absence à l'origine des biais de comparaison ou de confusion, 2) la randomisation et son absence à l'origine des biais de sélection, 3) l'évaluation en aveugle et son absence à l'origine des biais d'information.

<sup>13</sup> *Ibid.* : 647.

<sup>14</sup> *Ibid.*

<sup>15</sup> Dal-Ré R, Janiaud P, Ioannidis JPA. Real-world evidence: How pragmatic are randomized controlled trials labeled as pragmatic ? *BMC Medicine*, 2018, 16, 1 : 49.

<sup>16</sup> Vineis P. History of bias. In Morabia A. ed. *A History of Epidemiologic Methods and Concepts*. Basel : Birkhäuser, 2004 : 327-336.

<sup>17</sup> Murphy EA. *The logic of medicine*. Baltimore : Johns Hopkins University Press, 1976.

<sup>18</sup> Sackett DL. Bias in analytic research. *Journal of Chronic Diseases*, 1979, 32, 1-2 : 51-63.

<sup>19</sup> Miettinen OS. *Theoretical epidemiology, principles of occurrence research in medicine*. New York : Wiley, 1985 : 30.

Une autre étape importante dans cette clarification a été franchie grâce à Kenneth Rothman et Sander Greenland dans un chapitre de la deuxième édition de leur ouvrage *Modern Epidemiology* en 1996<sup>20</sup>. Dans ce texte qui a déjà fait date, les auteurs donnent non seulement une présentation complète, structurée et systématique des sources d'erreurs en épidémiologie, mais offrent des réflexions d'une portée plus large sur la question épineuse de la généralisation scientifique. Celle-ci est en effet considérée comme un processus d'élaboration d'hypothèses « universelles » abstraites, à bien séparer de la notion de « représentativité » et de « validité externe » d'une étude (qui concerne les inférences aux personnes extérieures à la population de l'étude). Une autre originalité de ce texte réside dans l'approche de l'étude épidémiologique « comme un exercice de mesure » ayant pour objectif d'estimer au mieux une « mesure de survenue de maladie ou d'effet ».

## La question de la causalité

### *Association ou causalité ?*

Austin Bradford Hill eut également un apport majeur dans la réflexion sur la causalité en épidémiologie, notamment avec son célèbre article « The Environment and disease, association or causation ? », publié en 1965<sup>21</sup>. La question que pose Hill dans cet article est en effet centrale : à quelles conditions peut-on passer de la constatation d'une association statistique entre des variables à un jugement de causalité ? Cette question n'est en effet pas de pure théorie en épidémiologie où il faut « déduire la causalité » pour « engager une action ». Austin Bradford Hill liste neuf aspects de l'association à considérer : la force, la consistance, la spécificité, la temporalité, le gradient, la plausibilité, la cohérence avec les connaissances médicales et biologiques, la possibilité d'une confirmation expérimentale et l'analogie avec d'autres causes connues. Cette liste d'aspects est souvent évoquée sous le nom de « critères de Hill » bien qu'il n'ait jamais employé ce terme. Austin Bradford Hill souligne qu'aucun des neuf aspects ne peut être considéré comme une preuve indiscutable ou une condition *sine qua non* et qu'il ne pense pas qu'il puisse y avoir des règles absolues en la matière, même si dit-il, « le meilleur soutien en faveur de l'hypothèse de la causalité<sup>22</sup> » est la confirmation expérimentale ou quasi expérimentale quand elle est possible. Austin Bradford Hill, comme David Hume et John Stuart Mill bien avant lui, ne s'est pas engagé dans une réflexion théorique sur la notion de causalité. C'est peut-être grâce à cette absence de théorisation que la liste d'Austin Bradford Hill connut un immense succès. D'autres « critères » furent proposés après Austin Bradford Hill, en particulier par Mervyn Susser<sup>23</sup>, qui distingua les propriétés essentielles d'une cause (association et ordre temporel), et les critères utiles pour décider de l'existence de ces propriétés essentielles dans un cas particulier : force, spécificité et consistance de l'association, « survivabilité », performance prédictive, cohérence (théorique, factuelle, biologique, statistique). À l'exception de la temporalité, condition *sine qua non* pour la notion de cause, aucun item de ces listes n'est nécessaire, et aucun n'est suffisant pour établir un jugement de causalité.

<sup>20</sup> Rothman KJ, Greenland S. Precision and validity in epidemiologic studies. In Rothman KJ, Greenland S. ed. *Modern Epidemiology*, Philadelphia : Lippincott, Williams & Wilkins, 1998 (2<sup>nd</sup> edition) : 115-134.

<sup>21</sup> Hill AB. The environment and disease: association or causation. *Proceedings of the Royal Society of Medicine*, 1965, 58 : 295-300.

<sup>22</sup> *Ibid.* : 299.

<sup>23</sup> Susser M. What is a cause and how do we know one ? A grammar for pragmatic epidemiology. *American Journal of Epidemiology* 1991, 133, 7 : 635-48.

### Approches déterministes ou probabilistes de la causalité

Dans les décennies qui ont suivi, la question complexe de la causalité, abordée par Bradford Hill de manière « pragmatique » pour la prise de décision en santé publique, a fait l'objet de réflexions soutenues de la part d'auteurs qui ont conceptualisé des modèles déterministes multicausaux de plus en plus complexes, dont les plus connus sont les « chaînes causales »<sup>24</sup>, les « toiles des causes » (*web of causation*)<sup>25</sup> dans lesquelles les maladies reconnaîtraient plusieurs ensembles de groupes de causes agissant de concert, les diagrammes causaux développés à partir des travaux de Judea Pearl<sup>26</sup>, ou encore le modèle des composantes causales suffisantes de Kenneth J. Rothman, publié dans *American Journal of Epidemiology* en 1976<sup>27</sup> qui reprenait les conditions INUS (pour *Insufficient but Necessary part of a condition which is itself Unnecessary but Sufficient for the result*) de John L. Mackie<sup>28</sup>. D'autres auteurs considérèrent les déterminismes des maladies agissant au niveau des populations, comme Abdel R. Omran qui proposa le modèle de la transition épidémiologique<sup>29</sup> et Mirko D. Grmek celui de pathocénose<sup>30</sup>. James S. Koopman quant à lui plaida pour des « systèmes » intégratifs couplant les modèles de transmission conçus pour les maladies infectieuses (au niveau populationnel) et les modèles multicausaux conçus pour les maladies chroniques (au niveau individuel)<sup>31</sup>. Parallèlement, des critiques plus radicales des modèles déterministes multicausaux émanèrent des tenants d'une causalité principalement ou exclusivement probabiliste, qui introduisirent dans le champ de l'épidémiologie les théories de Patrick Suppes ou de Wesley Salmon<sup>32</sup>.

### Conclusion

La longueur très limitée de cet article n'a pas permis d'entrer dans le détail des débats, parfois extrêmement techniques, sur la validité et la causalité, ni de faire réellement justice à la qualité de la réflexion des pionniers de l'introduction en médecine des méthodes de comparaison contrôlée. La visée pragmatique de ces auteurs ne souffre en revanche d'aucun doute : il s'agissait avant tout pour eux de convaincre les médecins de la validité des méthodologies utilisées pour justifier une intervention médicale clinique ou de santé publique. Dans les deux dernières décennies du XX<sup>e</sup> siècle, le développement de « spécialités » épidémiologiques, par l'ouverture ou plutôt par l'exploitation nouvelle d'anciens champs fréquentés par les statisticiens, comme la médecine clinique, l'usage des produits de santé, la génétique, la médecine sociale, pour ne citer que les plus étendus ou les plus fréquentés, s'est traduite par de nouveaux débats et de nouvelles réflexions épistémologiques dont la portée a pu dépasser les champs en question. Des réflexions ont aussi porté sur l'évaluation des conséquences sociétales et juridiques des expositions environnementales, ou

<sup>24</sup> Frost WH. ed. *Snow on cholera*. New York : The Commonwealth Fund, 1936 : IX-XXI.

<sup>25</sup> MacMahon B, Pugh TF, Ipsen J. *Epidemiologic...*, *op. cit.*

<sup>26</sup> Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology*, 1999, 10, 1 : 37-48.

<sup>27</sup> Rothman KJ, Causes. *American Journal of Epidemiology*, 1976, 104, 6 : 587-92.

<sup>28</sup> Mackie JL. Causes and conditions. *American Philosophical Quarterly*, 1963; 2, 4 : 245-55.

<sup>29</sup> Omran AR. The epidemiologic transition. A theory of the epidemiology of population change. *Milbank Memorial Fund Quarterly*, 1971, 49, 4 : 509-38.

<sup>30</sup> Grmek MD. Préliminaires d'une étude des maladies. *Annales. Économies, Sociétés, Civilisations*, 1969, 24, 6 : 1437-83.

<sup>31</sup> Koopman JS, Lynch JW. Individual causal models and population systems models in epidemiology. *American Journal of Public Health*, 1999, 89, 8 : 1170-4.

<sup>32</sup> Fagot-Largeault A. Épidémiologie et causalité. In Valleron A-J ed. *L'épidémiologie humaine : conditions de son développement en France, et rôle des mathématiques*. Les Ulis : EDP Sciences, 2006 : 239-45.

médicamenteuses, et sur la question de l'imputabilité individuelle de pathologies à ces expositions qui ne pourront seulement qu'être mentionnées ici<sup>33</sup>.

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

---

<sup>33</sup> Broadbent A. *Philosophy of epidemiology*. London : Palgrave & Macmillan, 2013 ; Coste J. Diverging approaches of pharmacovigilance and pharmacoepidemiology to assessing drug safety : epistemological and ethical implications. *Pharmacoepidemiol Drug Saf* , 2017, 26, 5 : 600-2 ; Gaille M dir. *Pathologies environnementales : identifier, comprendre, agir*. Paris : CNRS éditions, 2018.



## HISTOIRE DE L'ÉPIDÉMIOLOGIE

### L'étude de Framingham

Quel rôle dans l'histoire de l'épidémiologie des facteurs de risque ?

**Élodie GIROUX**

Maître de conférences en philosophie des sciences, Université Jean Moulin Lyon 3  
Institut de Recherches Philosophiques de Lyon, EA4187

#### Résumé

L'« épidémiologie des facteurs de risque » constitue un tournant dans l'histoire de l'épidémiologie au milieu du XX<sup>e</sup> siècle. L'identification de « facteurs de risque » de maladie est le cœur de cette épidémiologie étiologique. Or, ce concept apparaît comme chargé d'ambiguïtés et l'épidémiologie des facteurs de risque fait aujourd'hui l'objet de critiques comme étant trop centrée sur le niveau individuel au détriment du niveau populationnel et trop focalisée sur le calcul de risques relatifs. Cet article propose un détour par l'histoire de l'enquête prospective cardiovasculaire de Framingham qui débute en 1947, souvent considérée comme emblématique de cette épidémiologie, pour tenter d'apporter un éclairage sur ces ambiguïtés et ces critiques.

**Mots-clés :** enquête de Framingham, enquête de cohorte, épidémiologie moderne, facteur de risque, analyse multivariée, histoire des sciences.

#### Abstract

***The Framingham Heart Study. Its role in the history of risk factor epidemiology***

*At the middle of the 20<sup>th</sup> century, the “risk factor epidemiology” is a turning point in the history of epidemiology. The concept of risk factor is at the core of a new form of causal epidemiology. Yet this concept is ambiguous and the risk factor epidemiology is today criticized as too individualistic and as inducing risk relativism. Through a history of Framingham Heart Study, a cardiovascular cohort study that began in 1947 and that is often seen as emblematic of the “risk factor epidemiology”, this paper aims at lightening these ambiguities and criticisms.*

**Keywords :** *Framingham Heart Study, cohort study, modern epidemiology, risk factor epidemiology, multivariate analysis, history of science.*

Au milieu du XX<sup>e</sup> siècle<sup>1</sup>, dans un contexte de « transition épidémiologique<sup>2</sup> », l'épidémiologie élargit son objet d'investigation aux pathologies chroniques, en particulier les cancers et les maladies cardiovasculaires, devenues prévalentes dans les pays industrialisés. Ce qu'on appelle désormais « l'épidémiologie moderne<sup>3</sup> » ou « l'épidémiologie des facteurs de risque<sup>4</sup> » est associée à des méthodes d'enquête observationnelle dédiée à la recherche causale et à l'identification de « facteurs de risque » de ces maladies dont l'étiologie est complexe. Des enquêtes prospectives de cohorte et des enquêtes cas-témoins<sup>5</sup> sur les maladies cardiovasculaires et le cancer du poumon débutent dans les années 1940-1950 aux États-Unis et en Grande-Bretagne. Dans cette épidémiologie causale, les statisticiens mais aussi les cliniciens jouent un rôle déterminant. La nouvelle statistique inférentielle – qui s'est développée avec les travaux de Ronald Fisher, mais aussi de Jerzy Neyman et Egon Pearson entre autres, dans le premier tiers du XX<sup>e</sup> siècle – encourage le déploiement d'enquêtes de terrain : il devient possible de faire des inférences avec un degré mesuré d'incertitude, malgré les conditions hors laboratoire. Le test de signification et les analyses multivariées sont alors mis au service de l'analyse causale.

L'enquête prospective cardiovasculaire de Framingham, commencée en 1947 – et qui se poursuit encore aujourd'hui – occupe une place particulière dans l'histoire de l'épidémiologie des facteurs de risque. Elle est souvent considérée comme un modèle paradigmatique pour le protocole de l'enquête de cohorte<sup>6</sup>. De plus, il est admis qu'elle a joué un rôle fondamental dans le développement et la

<sup>1</sup> Cet article est issu d'un travail de thèse : Giroux É. Épidémiologie des facteurs de risque : genèse d'une nouvelle approche de la maladie, thèse de philosophie, Université Paris 1 Panthéon-Sorbonne, 2006.

<sup>2</sup> Omran AR. The epidemiologic transition. A theory of the epidemiology of population change. *Milbank Memorial Fund Quarterly*, 1971, 49, 4 : 509-38.

<sup>3</sup> Rothman KJ. *Modern Epidemiology*. Boston : Little, Brown and Company, 1986.

<sup>4</sup> Nous utilisons cette expression sans connotation critique contrairement au débat initié par Petr Skrabanek qui introduisit le terme de « risk-factorology ». Voir Skrabanek P. Risk-factor epidemiology : science or non-science ? In Anderson D ed. *Health, Lifestyle and Environment*. London : Social Affairs Unit, 1991 : 47-56. Un débat s'en est suivi dans le *Lancet* : Perry I.J. Risk factor epidemiology. *Lancet*, 1997, 350, 9086 : 1256 ; Susser M. Does risk factor epidemiology put epidemiology at risk? Peering into the future. *Journal of Epidemiologic Community Health*, 1998, 52 : 608-11.

<sup>5</sup> L'enquête cas-témoins consiste à comparer un groupe de personnes atteintes d'une maladie (cas) à un groupe de sujets qui n'ont pas la maladie étudiée (témoins). Il s'agit de rechercher des facteurs d'exposition antérieurs à la maladie et susceptible de pouvoir l'expliquer. L'étude de cohorte est quant à elle souvent longitudinale et consiste à suivre dans le temps des expositions suspectées d'être en cause dans la maladie étudiée et à relever les événements pathologiques.

<sup>6</sup> La plupart des manuels d'épidémiologie cite cette étude comme exemple pour définir le protocole de l'étude de cohorte : celui de Brian MacMahon, un grand classique dans l'épidémiologie anglo-saxonne, est un des premiers à expliciter la différence entre ces deux protocoles d'enquête : MacMahon B, Pugh TF, Ipsen J. *Epidemiologic Methods*. London : J&A Churchill, 1960 : 212. Voir aussi Susser M. Epidemiology in the United States after the World War II : the evolution of technique. *Epidemiologic Reviews*, 1985, 7 : 147-77 ; Doll R. Cohort studies : history of the method I. prospective cohort studies. In Morabia A ed. *A History of Epidemiologic Methods and Concepts*. Basel : Birkhäuser, 2004 : 243-74.

diffusion de « l'approche facteurs de risque<sup>7</sup> » de la maladie<sup>8</sup>. L'expression même de « facteur de risque » a été utilisée et diffusée dans le contexte biomédical à partir des résultats de cette enquête, dès le début des années 1960<sup>9</sup>. Enfin, cette enquête a contribué au développement d'une approche prédictive, fondée sur la constitution de « profil de risque » cardiovasculaire, à partir de la combinaison de certains facteurs de risque pour un même individu.

Mais pourquoi cette enquête est plus emblématique que les autres enquêtes prospectives de cohorte qui lui sont contemporaines ? Certes, les trois aspects que nous venons d'évoquer ont contribué à sa renommée. Mais si l'on regarde de plus près, l'enquête est loin de correspondre aux standards méthodologiques de l'enquête prospective de cohorte aujourd'hui en vigueur. Les choix méthodologiques ont été le fruit d'une négociation plutôt tâtonnante entre professionnels de santé publique, cliniciens et statisticiens, et le projet d'étude étiologique n'était au départ que secondaire. Si ces trois aspects expliquent la place qu'elle a prise dans l'histoire de l'épidémiologie, il convient cependant de nuancer. Ce faisant, nous souhaitons montrer que la critique actuelle de l'épidémiologie des facteurs de risque<sup>10</sup> est en partie un héritage de ces tâtonnements constitutifs de l'enquête de Framingham.

## L'étude de Framingham : un modèle de l'étude prospective de cohorte ?

### *Une population générale et mixte et un suivi prospectif*

Framingham est une ville qui se situe à trente kilomètres environ de Boston, capitale du Massachusetts et haut lieu de la cardiologie, avec Paul White<sup>11</sup> qui contribue à la conception de l'étude et d'autres éminents cardiologues comme Samuel Levine. La ville compte environ 28 000 habitants quand l'enquête débute en 1947. La cohorte est d'une taille importante, 5 209 individus, soit un peu moins d'un tiers des habitants qui ont entre 30 et 60 ans. Deux éléments contribuent au caractère précieux des données et à sa supériorité par rapport aux autres cohortes qui débutent alors aux États-Unis et en Grande-Bretagne<sup>12</sup>. D'une part, elle est constituée à partir d'une population générale, une population géographique dite « normale » et « non sélectionnée » alors que les autres enquêtes portent principalement sur des « populations artificielles<sup>13</sup> », c'est-à-dire des groupes professionnels ou issus d'un registre comme celui des assurés. D'autre part, elle inclut des femmes (2873). Aujourd'hui, cette inclusion fait la richesse des données de Framingham qui permettent d'obtenir de précieuses informations sur le rôle de la différence des sexes quant à l'étiologie de ces maladies. Ces choix sont assez étonnants en 1947 pour une étude prospective. À cette époque, la

<sup>7</sup> Aronowitz R. *Les maladies ont-elles un sens ?* Paris : Synthélabo, 1999 : 225.

<sup>8</sup> On a aussi souvent donné un rôle central à l'épidémiologie du cancer. Voir Berlivet L. "Association or causation ?" The debate on the scientific status of risk factor epidemiology, 1947-c.1965. *Clio Medica*, 2005, 75 : 39-74.

<sup>9</sup> Le premier usage de l'expression dans une publication médicale serait issu de l'analyse des données de l'enquête. Kannel WB, Dawber T, Kagan A *et al.* Factors of risk in the development of coronary heart disease, six-year follow-up experience, the Framingham study. *Annals of Internal Medicine*, 1961, 55 : 33-48.

<sup>10</sup> Voir note 4.

<sup>11</sup> Médecin personnel du président Eisenhower, P. White, exerçait dans le prestigieux hôpital de Boston : le *Massachusetts General Hospital*. Son manuel, *Heart Disease*, publié en 1931, fut réédité de nombreuses fois et traduit dans plusieurs langues. Paul White était présent dans les principales institutions représentatives de la cardiologie américaine, membre fondateur de l'*American Heart Association*, directeur du *National Advisory Heart Council*. Il fut très influent dans le développement de la cardiologie et dans l'importance que prit l'épidémiologie au sein de la médecine cardiovasculaire.

<sup>12</sup> Notamment à Los Angeles (1949), Albany (1953) et à Chicago (1957).

<sup>13</sup> NHLBI, Framingham Archives : G Meadors, Proposed study of the epidemiology of cardiovascular disease : rough draft, 1, 1948.

population principalement concernée par les maladies cardiovasculaires est celle des hommes. Or, pour avoir des résultats assez rapides sur la force de l'association statistique mesurée par le calcul du risque relatif (rapport des incidences), il vaut mieux cibler un groupe d'individus pour lesquels les événements cardiovasculaires ont des chances d'être les plus fréquents.

Une autre source d'étonnement concerne le fait que le type d'enquête retenu fut d'emblée celui, long et coûteux, de l'enquête prospective. Pour le cancer du poumon, les études initiées à la même époque sont d'abord des études cas-témoins. L'approche prospective n'est adoptée que dans un second temps, dans le cadre d'un débat sur la solidité des inférences causales que l'on peut tirer d'études cas-témoins. Le statisticien Ronald Fisher, notamment, critiquait l'hypothèse de la consommation de tabac, positivement associée au cancer du poumon, arguant de la faiblesse et des biais inhérents à l'étude cas-témoins. C'est alors que l'approche prospective est adoptée par Bradford Hill afin de donner de nouveaux éléments de preuve en faveur de l'inférence causale. Cette approche a un avantage épistémologique et méthodologique. Elle permet un recueil plus rigoureux des données d'exposition, en temps réel, et non rétrospectivement, en s'appuyant sur le souvenir des patients. Surtout, le suivi permet de garantir que l'exposition précède l'effet observé, élément essentiel pour le respect de notre intuition de la causalité. La temporalité fait partie de la liste des éléments aidant la décision de causalité dans ce qu'il est convenu d'appeler les « critères de causalité de Hill »<sup>14</sup> qui servent toujours aujourd'hui à l'inférence causale en épidémiologie.

### *Un projet à plusieurs strates*

Toutefois, à Framingham, ce qui semble avoir présidé au double choix d'une population générale et mixte et d'un suivi prospectif, ne relève pas explicitement de motifs liés à la méthodologie de la recherche étiologique. Premièrement, le projet de départ, en 1947-1948, dans le cadre de l'*US Public Health Service* (US PHS), est avant tout centré sur un objectif d'évaluation de l'incidence. Il s'agit alors de mener une enquête dans le style des enquêtes prospectives de morbidité de ce Service pour obtenir une évaluation précise de l'évolution du nombre de ces maladies dans la population nord-américaine<sup>15</sup>. Ce n'est que dans un second temps, en 1949, dans le cadre de son transfert au *National Heart Institute* (NHI), tout juste créé, que l'étude devient plus clairement un projet de recherche étiologique. De fait, les *National Institutes of Health* (NIH) constituent alors une sorte d'excroissance de l'*Hygienic Laboratory*, la division de l'*US PHS* consacrée à la recherche, et l'enquête fut considérée comme ayant finalement davantage un intérêt de recherche que d'action de santé publique. Deuxièmement, pour les maladies cardiovasculaires, l'enquête cas-témoins est en réalité difficile à réaliser car le nombre d'hypothèses étiologiques à tester est beaucoup trop important par rapport au cancer du poumon. Il fut ainsi difficile de ne retenir que 28 hypothèses étiologiques parmi les 300 initialement suggérées par le comité de conseil technique, en 1948. La forme prospective est plus adéquate dans cette configuration.

### *Le problème de la représentativité*

Un élément déterminant de la qualité de ce type d'étude concerne la modalité de sélection de l'échantillon et sa représentativité. Mais, la cohorte de Framingham est d'abord constituée de personnes volontaires. Elle débute dans un esprit proche d'un programme communautaire de prévention et de dépistage sur la tuberculose qui avait eu lieu entre 1917 et 1923 et qui avait reposé

<sup>14</sup> Hill A.B. Environment and disease : association or causation ? *Proceedings of the Royal Society of Medicine*, 1965, 58 : 295-300.

<sup>15</sup> Giroux É. Origines de l'étude prospective de cohorte : Épidémiologie cardiovasculaire américaine et étude de Framingham. *Revue d'Histoire des Sciences*, 2011, 64-2 : 297-318.

sur du volontariat<sup>16</sup>. Quand le projet prend une forme plus explicite de recherche étiologique, Felix Moore, statisticien au NHI, en charge des aspects statistiques et méthodologiques de l'enquête au moment de son transfert, introduit de l'aléatoire dans la sélection, en s'appuyant sur la liste alphabétique des résidents de la ville. En effet, depuis les travaux de Ronald Fisher, en particulier, la logique de la statistique inférentielle défend l'échantillonnage aléatoire pour garantir la possibilité du traitement probabiliste des résultats. 14% des volontaires de la population initiale sont malgré tout conservés dans la cohorte finale, pour ne pas risquer de mettre en péril les liens avec la population de cette ville.

Par ailleurs, la population de Framingham est peu représentative de la population nord-américaine. Le médecin Thomas Dawber, directeur de l'étude de 1950 à 1965, le concédera : « la population de cette ville du New England n'était pas elle-même totalement représentative des États-Unis. Il n'y avait pratiquement aucun Noir, aucun Oriental et la composition de la population blanche n'était pas nécessairement celle des populations blanches d'ailleurs<sup>17</sup> ». Néanmoins, ce n'est un problème à ses yeux que pour une étude d'incidence mais pas pour une étude étiologique centrée sur la comparaison inter-individuelle<sup>18</sup>.

Ainsi, aussi bien concernant la constitution de la cohorte, sa représentativité et la mise en place du protocole, l'étude de Framingham est loin de correspondre aux canons actuels de l'enquête prospective. On observe plutôt des tâtonnements et des ajustements successifs.

## Facteur de risque et modélisation de la multifactorialité

### *Du tableau de contingence à l'analyse multivariée*

Une autre raison explique la renommée de l'enquête de Framingham : elle a conduit à l'adaptation d'outils statistiques pour l'analyse des données épidémiologiques, en particulier l'analyse multivariée devenue si centrale dans l'analyse étiologique.

Une collaboration se met en place entre cliniciens à Framingham et statisticiens sur le site des *NIH* à Bethesda (Washington). Après son transfert au *NHI*, l'étude est successivement dirigée par des cliniciens d'ailleurs non formés en épidémiologie (ils se formeront par la suite) : d'abord Thomas Dawber, puis William Kannel, de 1966 à 1979, mais présent bien avant. Sur place, dans une clinique dédiée à l'enquête, les examens cliniques sont réalisés sur une durée de trente minutes pour chaque individu de la cohorte tous les deux ans. Les données collectées sont ensuite envoyées pour analyse aux statisticiens<sup>19</sup>. De leur côté, les statisticiens traitent ces données et sont aussi amenés à examiner des problèmes statistiques théoriques et méthodologiques soulevés par leur grand nombre et leur nature longitudinale. Ces travaux donnent d'ailleurs lieu à des publications dans des revues de statistique<sup>20</sup>.

Rapidement, des modèles multifactoriels se révèlent nécessaires pour démêler les associations entre les nombreux facteurs étudiés. Si dès 1957, les enquêteurs sont en mesure de publier des

<sup>16</sup> National Tuberculosis Association. *Framingham Monograph*, n° 10, Final Summary Report : 1917-1923 inclusive, Mass, 1918-1924.

<sup>17</sup> Notre traduction. Dawber TR. *The Framingham Study. The Epidemiology of Atherosclerotic Disease*. Cambridge, Mass : Harvard University Press, 1980 : 21-22.

<sup>18</sup> Dawber TR, Meadors GF, Moore FEJ. Epidemiological approaches to heart disease : the Framingham Study. *American Journal of Public Health*, 1951, 41 : 281.

<sup>19</sup> É. Giroux, *Épidémiologie ...*, *op. cit.* : 276.

<sup>20</sup> Kahn HA. A method for analyzing longitudinal observations on individuals in the Framingham Heart Study. In Goldfield ED ed. *American Statistical Association, Proceedings of the Social Statistics Section*. Washington : American Statistical Association, 1961: 156-160.

résultats conduisant à identifier sept hypothèses (hypertension, hypercholestérolémie, surpoids, âge, sexe, hypertrophie ventriculaire gauche, tabagisme) associées à une différence significative de risque<sup>21</sup>, la méthode classificatoire traditionnelle de l'épidémiologie est ici confrontée à ses limites. On ne peut pas assurer que chacun de ces facteurs contribue de manière indépendante au risque, ni évaluer s'il y a des interactions entre ces facteurs. En outre, cette approche fait perdre des informations avec la catégorisation plus ou moins arbitraire des variables d'exposition qui sont quantitatives et continues, comme les niveaux de pression artérielle et de cholestérolémie, en « élevé », « medium », « limite » et « bas ». Aussi les statisticiens ont-ils tendance à freiner les cliniciens pressés de conclure à l'existence de « facteurs de risque », afin d'identifier des variables sur lesquels il est possible d'intervenir. Ainsi, dès 1961, le clinicien William Kannel a recours au concept de « facteurs de risque » dans une publication<sup>22</sup> tandis que le statisticien Tavia Gordon explique sa réticence face à une terminologie plus pragmatique que rigoureusement étiologique<sup>23</sup>. Les statisticiens sont plus prudents dans leurs inférences à partir de simples associations statistiques, certes significatives, mais qui sont identifiées dans le cadre de la méthode classificatoire.

Dans une publication de 1962, Jerome Cornfield, statisticien des NIH, propose d'utiliser une fonction discriminante linéaire, intégrant les variables de cholestérolémie et de pression artérielle et permettant d'étudier leur contribution indépendante dans l'effet<sup>24</sup>. En 1967, il publie avec Jeanne Truett et William Kannel un article défendant l'importance de recourir à des équations multivariées pour conserver la nature continue des variables d'exposition étudiées et pour obtenir des informations quantitatives sur la manière dont les effets se combinent dans le risque. Le grand avantage de ces équations, par rapport aux tableaux de contingence, c'est qu'il est alors possible d'intégrer un grand nombre de variables :

La méthode analytique traditionnelle de l'épidémiologiste, la classification croisée multiple, devient rapidement impraticable à mesure que le nombre de variables à étudier augmente. Donc, si 10 variables sont observées et que chaque variable doit être étudiée à seulement trois niveaux, par exemple les niveaux de cholestérol à moins de 225 mg/100 ml, 225-274, et 275 et plus, il y aurait 59 049 cellules dans la classification croisée multiple<sup>25</sup>.

Le recours à ce type d'équations, aujourd'hui centrales pour l'identification de facteurs de risque en épidémiologie, contribua à consolider l'analyse et à démêler l'écheveau étiologique. Il devenait déjà un peu plus pertinent, en tout cas aux yeux des statisticiens, de parler de « facteur de risque » de maladie.

### *Une focalisation sur des variables individuelles et biologiques*

Une des critiques contemporaines de l'« épidémiologie des facteurs de risque » est qu'elle se centre trop sur des facteurs individuels et biologiques, au détriment de facteurs populationnels et sociaux. C'est indéniablement ce qui s'est passé à Framingham<sup>26</sup>. Les hypothèses qui sont

<sup>21</sup> Dawber TR, Moore FEJ, Mann GV. Coronary heart disease in the Framingham Study. *American Journal of Public Health*, 1957, 47, 2 : 4-24.

<sup>22</sup> Kannel W, Dawber T, Kagan A *et al.* Factors of risk in the development of coronary heart disease, six-year...., *op. cit.*

<sup>23</sup> Tavia Gordon évoque « une terminologie teintée par l'élan vers l'action », courrier à J Truett et J Cornfield, 1<sup>er</sup> nov. 1966, NHLBI, Framingham Archives.

<sup>24</sup> Cornfield J. Joint dependence of risk of coronary heart disease on serum cholesterol and systolic blood pressure : a discriminant function analysis. *Federation Proceedings*, 1962, 21, 2 : 58-61.

<sup>25</sup> Notre traduction. Truett J, Cornfield J, Kannel W. A multivariate analysis of the risk of coronary heart disease in Framingham. *Journal of Chronic Diseases*, 1967, 20 : 511.

<sup>26</sup> Giroux É. The Framingham study and the constitution of a restrictive concept of risk factor. *Social History of Medicine*, 2013, 26, 1 : 94-112.

confirmées comme « facteurs de risque », mais aussi celles qui sont retenues pour l'étude, sont essentiellement de nature biologique et clinique. Au départ, des hypothèses psychiques, environnementales et sociales sont proposées. Mais la mesurabilité, la reproductibilité et la facilité de prélèvement, dans le cadre d'un examen clinique qui ne doit pas durer trop longtemps, ni être trop invasif, ont été déterminantes. Le statut économique et les traits psychiques sont rapidement abandonnés. Après une étude de faisabilité, le régime alimentaire, le tempérament nerveux ou la constitution physique, le sont à leur tour.

On peut aussi expliquer cette focalisation sur des facteurs individuels, cliniques et biologiques, par le fait que l'étude est davantage passée sous l'influence de cardiologues et de cliniciens, après des débuts ancrés dans la santé publique. Les cardiologues jouèrent un rôle central dans le choix des 28 hypothèses quand l'enquête fut remaniée pour devenir une recherche étiologique. L'enquête est d'ailleurs souvent décrite par ses promoteurs comme une sorte d'« enquête clinique » à grande échelle plus qu'une étude épidémiologique à proprement parler<sup>27</sup>.

D'autres éléments ont contribué à ce que les facteurs de risque propulsés sur le devant de la scène de la prévention cardiovasculaire (hypertension, hypercholestérolémie) soient plus particulièrement de nature biologique : la possibilité de disposer de traitements biochimiques permettant de réduire le risque, ou plus généralement, la manipulabilité de ces facteurs. C'est alors l'industrie pharmaceutique qui a joué un rôle important, pour faire de l'hypertension et de l'hypercholestérolémie, deux des principales cibles de la prévention cardiovasculaire<sup>28</sup>.

## Le profil de risque et ses ambiguïtés

L'enquête de Framingham est aujourd'hui aussi connue pour être à l'origine des profils de risque permettant de calculer rapidement, en pratique de médecine générale, un risque cardiovasculaire global<sup>29</sup>. Ces équations ont, en effet, été établies à partir des données de la cohorte. Les cliniciens de l'enquête, mais aussi l'*American Heart Association*, ont rapidement souhaité utiliser les équations multivariées, introduites initialement par Jerome Cornfield dans un but analytique, pour établir des prédictions individualisées et un profil de risque cardiovasculaire. L'équation prédictive – ou profil de risque – permet de calculer le risque qu'un individu a, sur plusieurs années, de contracter tel événement cardiovasculaire, en fonction de la présence et du niveau des variables identifiées comme des facteurs de risque.

Des tensions sont là aussi notables dans les échanges entre les cliniciens à Framingham et les statisticiens aux NIH. Les statisticiens, beaucoup plus prudents, soulignent l'importance de bien distinguer usage analytique et usage prédictif des équations. Pour Tavia Gordon, il est nécessaire préalablement de tester « l'aptitude prédictive » de l'équation, c'est-à-dire de s'assurer que les équations effectuées à partir de la cohorte de Framingham valent pour les données issues d'autres études.

Par ailleurs, les facteurs retenus pour les équations sont différents en fonction de l'usage. La valeur prédictive d'un facteur n'est pas toujours identique à son importance étiologique. C'est le cas du surpoids qui est étiologiquement important, mais prédictivement faible : il constitue un facteur de

<sup>27</sup> Giroux É. De l'épidémiologie de santé publique à l'épidémiologie clinique : quelques réflexions sur les rapprochements entre épidémiologie et clinique (1920-1980). *Bulletin d'Histoire et d'Épistémologie des Sciences de la Vie*, 2012, 19, 1 : 21-43.

<sup>28</sup> Greene JA. *Prescribing by numbers: drugs and the definition of disease*. Baltimore : The Johns Hopkins University Press, 2007.

<sup>29</sup> Giroux É. Les modèles de risque en médecine : quelles conséquences pour la définition des normes et pour le jugement clinique ? Exemple du risque cardiovasculaire. In Fagot-Largeault A dir. *Émergence de la médecine scientifique*, Paris : Éditions matériologiques, 2012 : 199-216.

risque causal de maladies cardiovasculaires, mais il n'est pas pour autant un bon prédicteur quand il est associé aux autres facteurs retenus pour leur valeur prédictive dans l'équation de prédiction. À cela, il faut ajouter que l'équation prédictive privilégie des facteurs plus facilement mesurables en clinique. C'est, en effet, l'objectif de prévention qui conduit à introduire la notion de « profil de risque ». En 1976, la première équation prédictive fait l'objet d'une publication dans l'*American Journal of Cardiology*<sup>30</sup>. D'autres suivront, avec des adaptations à d'autres populations, pour ajuster et affiner la capacité prédictive de l'équation.

## Conclusion

L'histoire de l'enquête de Framingham relativise la place qu'elle a souvent prise dans les manuels comme paradigme de l'étude prospective de cohorte. Elle permet de comprendre que certaines critiques aujourd'hui adressées à l'épidémiologie des facteurs de risque ont leur source dans cette enquête, dans ses débuts tâtonnants ainsi que dans l'articulation complexe de la nouvelle statistique inférentielle et de la clinique, composantes clés de la nouvelle forme que prend l'épidémiologie à partir du milieu du xx<sup>e</sup> siècle. Les méthodes d'enquête et d'analyse des données se mettent progressivement en place, dans une négociation entre cliniciens et biostatisticiens. La statistique inférentielle est intégrée à l'épidémiologie, mais adaptée : c'est une permanente recherche de compromis entre exigence d'objectivité et faisabilité de l'enquête. Cette négociation se fait aussi en fonction des pathologies considérées et de la complexité de leur étiologie. Le concept de facteur de risque et les modèles prédictifs de risque viennent avant tout des cliniciens dans une optique plus pragmatique qu'analytique, leur objectif étant de développer pour les maladies cardiovasculaires une prévention individualisée. Ainsi, l'histoire de l'enquête de Framingham apporte un éclairage sur la manière dont la nouvelle statistique inférentielle a pénétré l'épidémiologie traditionnelle pour l'amener à jouer ce rôle dans l'identification de facteurs de risque de maladie et dans le développement consécutif d'une prévention individualisée du risque.

Si l'on en revient aux critiques récentes de l'« épidémiologie des facteurs de risque », il est possible de conclure que l'enquête de Framingham a eu un rôle déterminant pour infléchir la recherche épidémiologique du côté de ce qui constitue les principaux pôles d'accusation de ces critiques. Tout d'abord, la restriction des facteurs de risque à des facteurs biologiques et individuels que l'on observe dans la recherche à Framingham participe de *l'individualisation du risque* et d'un privilège accordé aux facteurs jugés plus aisément manipulables que des facteurs sociaux ou environnementaux. Ensuite, cette épidémiologie est critiquée pour se centrer sur des *mesures relatives de risque* – il est question de « relativisme du risque » – au détriment de mesures absolues<sup>31</sup>. Les analyses des données de Framingham sont, de fait, centrées sur des comparaisons inter-individuelles pour calculer des risques relatifs dont on considère qu'ils donnent une bonne évaluation de la force de l'association en vue de l'inférence causale. Dans cette logique, la variance entre populations et les mesures absolues de risque tendent à être négligées alors qu'elles sont tout aussi importantes pour l'évaluation des risques. Les enquêteurs insistent en effet sur le primat de la variance à l'intérieur de la population par rapport à la variance entre populations dans une telle recherche étiologique et semblent avoir l'idée que ces mesures relatives sont transférables à d'autres populations. Enfin, et c'est lié, l'approche facteurs de risque est aussi critiquée pour manquer de rigueur dans l'analyse causale : il s'agirait d'une approche « boîte noire » qui met en évidence des corrélations, sans apporter de connaissances sur le mécanisme qui relie l'exposition et l'effet. Or, s'il

<sup>30</sup> Kannel WB, MacGee DL, Gordon T. A general cardiovascular risk profile : the Framingham Study. *The American Journal of Cardiology*, 1976, 38 : 46-51. Oppenheimer GM. Profiling risk : the emergence of coronary heart disease epidemiology in the United States (1947–70). *International Journal of Epidemiology*, 2006, 35, 3 : 720-30.

<sup>31</sup> Voir Broadbent A. *Philosophy of epidemiology*, Palgrave Macmillan, 2013 : 129-44.

est clair que l'analyse multivariée contribue à renforcer la solidité de l'inférence causale en permettant d'évaluer la part indépendante de chaque facteur dans le risque, on observe, à Framingham, des divergences entre cliniciens et statisticiens sur l'analyse de la causalité et une certaine confusion, de la part des cliniciens en tout cas, entre inférence causale et inférence prédictive.

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



## ÉPIDÉMIOLOGIE ET MALADIES INFECTIEUSES

### Un monde moderne

L'épidémiologie des maladies infectieuses avant sa réinvention des années 1980-1990

#### Gaëtan THOMAS

Docteur de l'EHESS, enseignant à Sciences Po Paris  
membre du Centre de recherches historiques EHESS

#### Résumé

Le Centre international de l'enfance (CIE, 1949-1999) s'inscrivait doublement dans une modernité épidémiologique, via ses méthodes statistiques et via le parrainage du père de l'épidémiologie moderne en France, Daniel Schwartz. Dès lors, si le CIE présentait tous les gages de rigueur méthodologique, comment expliquer son absence des récits sur la réinvention de l'épidémiologie française dans les années 1980-1990 ?

**Mots-clés** : vaccination, essais cliniques, Centre international de l'enfance, épidémiologie, agences sanitaires.

#### Abstract

**A modern world : the epidemiology of infectious diseases before its reinvention in the 1980s-1990s**

*This article unpacks the International Children's Center (ICC, 1949-1999) relationship to an epidemiological modernity, through both its actual statistical methods and its advisor Daniel Schwartz, considered the father of modern French epidemiology. Given its centrality in this regard, how can we explain the absence of the ICC from narratives of the reinvention of French epidemiology in the 1980s and 90s ?*

**Keywords** : immunization, clinical trials, International Children's Center, epidemiology, health agencies.

Le début des années 1990 est décrit d'ordinaire comme une période de transformation radicale de la santé publique française. Selon un récit largement admis, la reconnaissance d'une faillite de la gestion des maladies infectieuses aurait conduit à la création d'institutions spécialisées, à un renouvellement des doctrines autour de la notion de sécurité sanitaire et, enfin, à l'arrivée d'une nouvelle génération d'acteurs. Les experts du Conseil supérieur d'hygiène publique de France (CSHPF), une instance consultative affaiblie par des soupçons de conflits d'intérêts dans les années 1980<sup>1</sup>, auraient alors perdu le monopole des savoirs sanitaires au profit de nouvelles structures administratives – les « agences sanitaires<sup>2</sup> ». Pourvues d'une double mission d'expertise et de police, ces dernières faisaient valoir un lien direct avec le ministère de la santé, un gage d'efficacité selon leurs promoteurs.

Les chercheurs de sciences sociales ont confirmé ce récit propagé par les responsables en poste dans les années 1990, sans toutefois manquer de pointer les ressorts politiques des déclarations sur les carences ou les faillites d'une « culture française » de santé publique<sup>3</sup>. Un des chapitres les plus discutés renvoie à l'institutionnalisation d'une épidémiologie des maladies infectieuses. Dénommée tantôt épidémiologie d'intervention, tantôt épidémiologie de terrain, cette forme de quantification était fortement influencée par les concepts et les pratiques de l'agence fédérale américaine des *Centers for Disease Control* d'Atlanta<sup>4</sup>. Les historiens ont expliqué que l'institutionnalisation de cette épidémiologie fut amorcée dans les années 1980 au moyen d'une double mobilisation, orchestrée depuis l'État et la Fondation Mérieux. En 1992, la création du Réseau national de santé publique (RNSP) marquait une étape importante. L'agence se consolida au long des années 1990 et développa de nouveaux domaines d'expertise, en matière de vaccination par exemple.

Malgré des déclarations contraires, ce nouveau paysage ne fut pas créé ex nihilo. Il existait une santé publique institutionnalisée en dehors de l'arène du CSHPF et avant l'essor des agences sanitaires. De ce point de vue, le tournant des années 1990 s'assimile davantage à un cycle wébérien de création-destruction qu'à l'apparition soudaine d'un réseau d'institutions créées dans l'urgence, en réponse à la multiplication des controverses. Or, si on connaît bien le versant créateur de ce cycle,

<sup>1</sup> Buton F. Une agence tous risques ? La veille sanitaire comme savoir de gouvernement. Mémoire pour l'habilitation à diriger des recherches, Université Lyon 2, 2015, vol. 2.

<sup>2</sup> Benamouzig D, Bensaçon J. Administrer un monde incertain : les nouvelles bureaucraties techniques. Le cas des agences sanitaires en France. *Sociologie du travail*, 2005, 47 : 301-22.

<sup>3</sup> Fassin D. Au cœur de la cité salubre. La santé publique entre les mots et les choses. In Fassin D et Dozon J-P dir. *Critique de la santé publique. Une approche anthropologique*. Paris : Balland, 2001 : 47-72.

<sup>4</sup> Buton F. Une agence...., *op. cit* : 63-140 ; Buton F. De l'expertise scientifique à l'intelligence épidémiologique : l'activité de veille sanitaire. *Genèses*, 2006, 65, 4 : 71-91 ; Berlivet L. Exigence scientifique et isolement institutionnel : l'essor contrarié de l'épidémiologie française dans la seconde moitié du XX<sup>e</sup> siècle. In Jorland G, Opinel A, Weisz G ed. *Body Counts, Medical Quantification in Historical & Sociological Perspectives*. Montréal : McGill-Queen's University Press, 2005 : 335-58 ; Berlivet L. Une santé à risques. L'action publique de lutte contre l'alcoolisme et le tabagisme en France, 1954-1999. Thèse de doctorat, Université de Rennes, 2000 ; Drucker J. Des CDC d'Atlanta à l'Institut de veille sanitaire en passant par le Réseau national de santé publique : l'essor de l'épidémiologie d'intervention en France. *Revue française des affaires sociales*, 1997, 3-4 : 63-70.

c'est moins le cas de ce qui a été détruit. Une institution aussi importante pour la vaccination que le Centre international de l'enfance (CIE, 1949-1999) s'avère absente des récits sur la mise en place des agences sanitaires. Pourtant, son département spécialiste de l'épidémiologie des vaccins, la Station pilote, s'intégrait pleinement à une modernité épidémiologique, via ses méthodes statistiques et via son association avec le rénovateur de la discipline, Daniel Schwartz.

### **L'épidémiologie du CIE, à la croisée de la médecine coloniale, postcoloniale, et de la santé publique métropolitaine**

Jusqu'à l'essor du RNSP dans les années 1990, le domaine français de la vaccination s'est organisé autour du CIE (1949-1999). Robert Debré, son fondateur, était un mandarin multimensionné<sup>5</sup>. Il donna à « son » centre tous les atouts d'une institution internationale, à commencer par le nom, alors que le CIE relevait du droit français et avait un personnel essentiellement métropolitain. L'institution avait pour mission de diffuser la discipline fondée par Robert Debré, la pédiatrie sociale, au moyen d'activités menées en priorité dans des territoires africains, sous domination française jusqu'au tournant des années 1960. Ces activités étaient de trois ordres : un enseignement postuniversitaire destiné à des boursiers, une recherche pluridisciplinaire censée incarner l'approche holistique de la santé de l'enfant promue par Debré – mais ne questionnant jamais les effets sanitaires de la domination coloniale – et la diffusion de connaissances au moyen d'une bibliothèque et de nombreuses publications. Le CIE était pris en tenaille entre l'impératif d'assurer la vie d'une discipline, l'attachement au pré carré africain de la France et le désir de rayonner sur une scène internationale, en partie incarnée par l'Organisation mondiale de la santé (OMS), critique envers les puissances colonisatrices et allergique à la défense des intérêts nationaux<sup>6</sup>.

Un niveau de difficulté s'ajoute si l'on considère les activités du CIE en matière de vaccination. La Station pilote fait figure d'anomalie puisqu'elle a d'abord travaillé en métropole, à la différence des autres services du CIE tournés vers l'Afrique coloniale et l'international. Durant sa première décennie d'existence, elle s'est consacrée presque exclusivement à la standardisation du vaccin contre la tuberculose, le BCG, au moyen d'essais cliniques réalisés sur des écoliers français. À partir du milieu des années 1960, elle a étendu son expertise à tous les vaccins de l'enfance et développé des collaborations décisives avec l'Institut Mérieux. Ce n'est qu'en 1970 qu'elle a entamé une série d'essais vaccinaux en Afrique subsaharienne, tout en réduisant, sans y renoncer, les activités menées en région parisienne<sup>7</sup>. Très impliquée dans l'internationalisation des politiques vaccinales, elle a importé en France des outils et des catégories épidémiologiques en usage à l'OMS à partir du début des années 1980. Signe de la reconnaissance acquise sur la scène nationale, elle a ensuite obtenu le statut de Centre national de référence pour les vaccinations de l'enfance en 1984. Deux ans plus tard, la directrice de la Station pilote siégeait au tout nouveau Comité technique des vaccinations (CTV), l'instance dédiée à l'édiction des politiques vaccinales<sup>8</sup>. Il faut donc souligner le paradoxe selon lequel le département d'une institution tournée vers l'Afrique est devenu le cœur battant de l'expertise vaccinale métropolitaine.

<sup>5</sup> Berlivet L. The Ubiquitous Mandarin. Notes on the Social Organization of Elite Medicine in the Twentieth Century, *Medicina Nei Secoli*, 2008, 20 : 847-69.

<sup>6</sup> Thomas G. La routine vaccinale. Enquête sur un programme français de rationalisation par les nombres. Thèse de doctorat EHESS, 2018 ; Pearson J L. *The Colonial Politics of Global Health. France and the United Nations in Postwar Africa*. Cambridge : Harvard University Press, 2018.

<sup>7</sup> Thomas G. La France et l'Afrique, en parallèle. In Marcilloux P, Dénéchère Y dir. *Le Centre international de l'enfance, 1949-1997. Des archives à l'histoire*. Rennes : Presses universitaires de Rennes, 2016 : 153-68.

<sup>8</sup> Guérin N. Les activités relatives au BCG et aux vaccinations au Centre international de l'enfance. In Marcilloux P et Dénéchère Y dir. *Le Centre .....*, op. cit. : 124.

Cette situation a façonné les programmes de recherche. Intégrant des problématiques propres à l'Afrique de l'Ouest, la Station pilote s'est par exemple interrogée sur les facteurs nutritionnels dans la réponse immunologique aux vaccins. Elle a aussi intégré le monde de l'épidémiologie du développement. Dans les années 1990, alors que son statut de Centre national de référence l'ancrait en France, elle rejoignit un projet centré sur l'Afrique de l'Ouest, EPIGEPS – pour épidémiologie, informatique appliquée et gestion des programmes de santé<sup>9</sup>.

Il existait donc une institution reconnue qui se consacrait à l'étude épidémiologique des maladies infectieuses et de la vaccination avant l'apparition des agences sanitaires. Il faudrait d'ailleurs compléter ce tableau en adjoignant au CIE des sections de l'Institut national d'hygiène (INH) puis de l'INSERM qui travaillèrent sur la vaccination jusqu'à la fin des années 1970. Mais à la différence des membres des unités INH-INSERM dont les pratiques statistiques étaient jugées dépassées<sup>10</sup>, ceux du CIE n'exprimèrent pas le sentiment d'être méprisés par les tenants du renouveau de l'épidémiologie française. De fait, la Station pilote avait trouvé en Daniel Schwartz un allié de premier plan sur la scène scientifique nationale.

### Parrain et neveu : Daniel Schwartz au CIE

Peu après la création du CIE en 1949, Robert Debré plaça les services de recherche sous le patronage de « conseillers scientifiques ». René Dubos, un chercheur en poste *au Rockefeller Institute* de New York, fut sollicité pour le laboratoire de biologie. Pour la Station pilote, Robert Debré choisit son disciple, le pédiatre Raymond Mandé, une personnalité aussi impliquée dans la création du CIE. Ce dernier exerçait une direction officieuse, ses prérogatives lui donnaient le dernier mot dans les réunions de service. À partir de 1955, un parrain plus iconoclaste s'associa à la Station pilote, le conseiller en statistiques Daniel Schwartz, dont l'expertise dédoublait celle de la statisticienne salariée par le CIE<sup>11</sup>.

Au-delà d'une dimension genrée frappante – des hommes salariés par d'autres institutions supervisaient le travail d'équipes plutôt féminines – ces parrainages introduisaient une nouvelle modalité de l'école-famille chère à Robert Debré. Ce dernier utilisait souvent une métaphore familiale pour se référer à la constitution d'une école de pédiatrie ou aux équipes plus resserrées qu'il dirigeait<sup>12</sup>. La métaphore trouvait ici une incarnation parfaitement littérale, puisque Daniel Schwartz était son neveu. Le compagnonnage de Daniel Schwartz avec le CIE débuta deux ans avant que celui-ci ne terminât une enquête sur l'étiologie du cancer qui l'installa dans les habits de renovateur de l'épidémiologie et lui ouvrit les portes de l'INH, où il se vit offrir la direction d'une unité.

L'implication de Daniel Schwartz dans les activités quotidiennes du CIE était bien moindre que celle de Raymond Mandé. Si on s'en tient aux traces visibles dans les archives, Daniel Schwartz participa à plusieurs réunions et coécrivit au moins un article avec la statisticienne du centre<sup>13</sup>. Ceci dit, son influence dépassait ces interventions épisodiques. On la retrouve dans la place dévolue à la statistique dans le travail expérimental du CIE. Avant de publier son premier manuel, *Méthodes*

<sup>9</sup> Archives du ministère de la Santé, 2010/081-19, « CIE Epigepts », notes de la réunion du 16 juillet 1993.

<sup>10</sup> Entretien avec France Lert mené par l'auteur, Paris, 7 avril 2014. Sur la santé publique à l'INSERM, voir Berlivet L. *Between Expertise and Biomedicine: Public Health Research in France after the Second World War. Medical History*, 52, 4 : 471-492.

<sup>11</sup> Archives CIE, (université d'Angers), ci-après ACIE, 1CIDEF13, rapport d'activité, 1956.

<sup>12</sup> Debré R. *L'honneur de vivre*. Paris, Stock Hermann, 1974 : 142.

<sup>13</sup> Orssaud E, Schwartz D. Corrélations entre les résultats des diverses techniques. *Bulletin de l'Union internationale contre la tuberculose*, 1960, 30, 2 : 251-3.

*statistiques à l'usage des médecins et des biologistes* (1963<sup>14</sup>), Daniel Schwartz fit paraître en 1956, dans la revue du CIE, un article où il plaidait pour l'intervention d'un ou d'une statisticienne à deux moments clés de l'expérimentation, dans la préparation d'un essai, puis lors de l'analyse des données :

L'importance du schéma expérimental [...] montre que le rôle du statisticien est double : il doit assurer d'abord l'organisation de l'expérience, et ensuite l'interprétation des résultats ; on connaît bien cette deuxième tâche, mais on ignore trop souvent la première : c'est une fois l'expérience terminée qu'on va voir un statisticien non consulté au préalable dans l'établissement du schéma, on lui apporte triomphalement quelques milliers de chiffres, et on lui demande ce qu'il faut en faire. Si le statisticien propose de les mettre à la corbeille, on verra de l'orgueil là où n'y a que position de rigueur : l'exploitation des résultats est absolument conditionnée par l'organisation de l'expérience<sup>15</sup>.

Il ressort de ce texte que la vaccination constitue un champ d'expérimentation rêvé pour le statisticien. Daniel Schwartz y remarque qu'un essai vaccinal pose moins de problèmes éthiques qu'un essai clinique évaluant un produit thérapeutique puisque le premier n'expose pas le malade à un placebo. Le CIE comparait des vaccins aux qualités similaires, la distribution des sujets dans les différents bras de l'essai n'avait pas de conséquence significative sur la santé des individus. Cependant, la faible variabilité des produits testés n'évacuait pas l'ensemble des problèmes éthiques, relatifs au consentement par exemple.

Plus tard, la présence de Daniel Schwartz dans l'organigramme du CIE sécurisa le prêt de calculateurs par son unité Inserm, l'unité 21. Les employés du CIE se contentaient d'instruments rudimentaires dans les locaux du bois de Boulogne. En 1975, la statisticienne Evelyne Orssaud ne disposait que d'un calculateur Olivetti Programma 101, une grosse machine à calculer commercialisée à partir de 1965, souvent présentée comme le premier ordinateur personnel. Les calculs complexes des essais vaccinaux furent effectués sur les machines autrement plus puissantes de l'unité 21 à Villejuif<sup>16</sup>.

L'unité 21 ne réalisa pas de travaux relatifs aux maladies infectieuses et à leur contrôle. L'intérêt de Schwartz et son équipe allait à l'étude des maladies chroniques. Comme l'a montré Luc Berlivet, ils tendirent à disjoindre leur définition de la statistique médicale des phénomènes infectieux, adoptant l'expression de « statistique médicale » pour souligner leurs différences avec l'épidémiologie de dénombrement et la santé publique<sup>17</sup>. La formule de statistique médicale exprimait un tropisme pour les questions d'étiologie dont la résolution passait par des innovations de méthode. Le mot d'épidémiologie ne fut revendiqué que lors de la parution d'un manuel de *Méthodes en épidémiologie* en 1970 – qui ne tenait toujours pas compte des maladies infectieuses<sup>18</sup>. Malgré leur distance vis-à-vis des maladies infectieuses comme objet d'étude, Daniel Schwartz et son équipe ont soutenu des pratiques éloignées de leur conception étroite de la statistique médicale défendue à l'intérieur de l'INH-INSERM. Comme le montre l'exemple du CIE, ils apportèrent un appui méthodologique et instrumental à des équipes moins portées sur l'innovation disciplinaire<sup>19</sup>.

<sup>14</sup> Schwartz D. *Méthodes statistiques à l'usage des médecins et des biologistes*. Paris : Flammarion, 1969.

<sup>15</sup> Schwartz D. La méthode statistique dans la recherche médicale. *Courrier. Centre international de l'enfance*, 1956, 2 : 75-84.

<sup>16</sup> ACIE, ICIDEF483, besoins en informatique de la Station pilote, 19 fév. 1975.

<sup>17</sup> Berlivet L. Une santé à risques..., *op. cit.*

<sup>18</sup> Rumeau-Rouquette C, Schwartz D. *Méthodes en épidémiologie*. Paris, Flammarion, 1970, 273 p.

<sup>19</sup> Sur le soutien de l'équipe Schwartz apporté à d'autres membres de l'INH-INSERM, Gaudillière J-P. *Inventer la biomédecine. La France, l'Amérique et la production des savoirs du vivant (1945-1965)*. Paris, La découverte, 2002 : 242.

## La statistique, un gage de modernité expérimentale

Le rapport du CIE à la modernité épidémiologique dépassait le seul parrainage de Daniel Schwartz. La Station pilote – dont le nom rappelle les stations agricoles si importantes dans l'histoire des relations entre statistiques et expérimentation<sup>20</sup> – revendiquait une adhésion aux normes des essais cliniques modernes<sup>21</sup>. Quels en étaient les principes fondamentaux ? La randomisation d'abord, soit la répartition au hasard des produits et sujets testés<sup>22</sup>, et le double aveugle, une expression désignant la méconnaissance, par l'expérimentateur et le sujet, du bras de l'essai dans lequel celui-ci se trouve<sup>23</sup>. Le développement de ces règles, visant à contrôler différents biais, incarne l'aboutissement d'une « médecine scientifique » dont la légitimité provient de l'application de la statistique probabiliste à l'expérimentation clinique.

Les essais cliniques mêlant ces deux critères restèrent l'exception jusqu'aux années 1960 aux États-Unis, et les années 1970 en Europe, mais la Station pilote mit un point d'honneur à randomiser dès ses essais sur la standardisation du BCG lancés au tournant des années 1950. Entre 1948 et 1958, elle vaccina plus de 40 400 enfants au vaccin contre la tuberculose pour évaluer l'influence sur l'allergie de l'âge des vaccins, de la température de conservation, de la lumière, de la lyophilisation ou de la concentration des bacilles. Reconnaissaient les limites et le caractère répétitif des essais BCG, les membres de la Station pilote élargirent leurs recherches à tous les vaccins de l'enfance à partir du milieu des années 1960.

Les essais sur « l'immunisation multiple du nourrisson » démarrés en 1964 visaient à établir des calendriers vaccinaux. Ils testaient non pas un produit mais des dates d'injections dans leurs rapports mutuels. Le principe de double aveugle ne pouvait s'appliquer – comment dissimuler la date à laquelle une injection est réalisée, puisque c'est bien la date qui est testée ? – mais la randomisation restait de vigueur. Le recours à d'innombrables tests sérologiques, avant et après les injections, inscrivait ces travaux dans une mouvance dynamique au tournant des années 1960 – l'épidémiologie sérologique<sup>24</sup>.

La statistique permit au CIE de clamer une indépendance scientifique lorsque les laboratoires pharmaceutiques accrurent leur participation aux essais, via la fourniture des vaccins d'abord, puis des analyses sérologiques. Néanmoins l'usage systématique de la statistique traduisait moins une posture morale de défiance à l'égard des industriels – la Station pilote rechercha activement leur soutien financier – qu'une adhésion aux normes scientifiques de l'époque.

Enfin, la Station pilote resta attentive aux innovations de méthode relatives à d'autres pratiques que les essais cliniques. En 1977, Nicole Guérin, alors directrice adjointe de la Station pilote, fut l'une des premières Françaises à suivre une partie de la formation Epidemic Intelligence Service des CDC d'Atlanta, cinq ans avant que l'industriel Charles Mérieux n'y envoie des médecins stagiaires qui inaugurèrent, selon le récit admis, le mouvement d'institutionnalisation français de l'épidémiologie de terrain. Au début des années 1980, le CIE importa en France des

<sup>20</sup> Parolini G. In pursuit of a science of agriculture : the role of statistics in field experiments. *History and Philosophy of the Life Sciences*, 2015, 37, 3 : 261-281.

<sup>21</sup> Chalmers I. Joseph Asbury Bell and the birth of randomized trials. *Journal of the Royal Society of Medicine*, 2007, 100, 6: 287-293.

<sup>22</sup> Hacking I. Telepathy : origins of randomization in experimental design. *Isis*, 1988, 79 : 427-451 ; Marks H. *La médecine des preuves. Histoire et anthropologie des essais cliniques (1900-1990)*. Le Plessis-Robinson, Les empêcheurs de penser en rond, 1999 : 208.

<sup>23</sup> Kaptchuk T J. *Intentional Ignorance : A History of Blind Assessment and Placebo Controls in Medicine*. Bulletin of the History of Medicine, 1998, 72, 3 : 389-433.

<sup>24</sup> Ipsen J, Bowen H. Effects of Routine Immunization of Children with Triple Vaccine (Diphtheria-Tetanus-Pertussis). *Amer. J. Pub. Health Nat. Health*, 1955, 45 : 312-8.

techniques de sondage promues par l'OMS de calcul de couverture vaccinale<sup>25</sup>. Parce que ces sondages contournaient l'absence de bureaucratie sanitaire en produisant des données de terrain instantanées, ils s'inscrivaient dans la philosophie de la *shoe-leather epidemiology* défendue par les CDC. Ils avaient d'ailleurs été mis au point par un de leurs épidémiologistes<sup>26</sup>.

## Conclusion

À l'issue de ce parcours, il ne fait plus doute qu'une généalogie de la réinvention de l'épidémiologie doit intégrer le CIE. Cependant, ce constat vise moins à revendiquer une antériorité qu'à interroger une absence, double au demeurant, qui renvoie aussi bien au démantèlement du CIE en 1999 qu'à son évacuation des discours sur l'émergence, quelques années plus tôt, d'une épidémiologie moderne des maladies infectieuses. On pourrait voir dans cette omission des enjeux de présentation personnelle d'acteurs ayant pris part à la création des agences sanitaires. Il semble plutôt que CIE présentait des caractéristiques inassimilables à la nouvelle modernité sanitaire. Au tournant des années 1990, cette modernité était moins méthodologique et disciplinaire, comme celle des décennies d'après-guerre, que politique, liée à la situation sociale de l'expertise.

Alors que les pratiques de veille et d'alerte mises en place dans les années 1990 tendaient à s'inscrire dans un cadre national, tout en recourant au registre de la menace épidémique mondialisée, l'épidémiologie du CIE prolongeait la tradition de la médecine coloniale. Dans la deuxième partie du vingtième siècle, nombre d'essais vaccinaux furent menés dans les anciennes possessions françaises, si bien que le savoir sur la vaccination et les maladies infectieuses s'élabora sur différents terrains, en métropole et en Afrique de l'Ouest, parfois simultanément<sup>27</sup>. Or, l'Afrique francophone se détacha brusquement de l'horizon des épidémiologistes lorsqu'ils intégrèrent le nouvel environnement institutionnel des années 1990, alors même que cette région avait constitué un lieu de formation et d'intervention pendant des décennies. La seconde caractéristique problématique du CIE a trait à ses relations avec l'industrie pharmaceutique. Les controverses sanitaires des années 1990 mirent en évidence le coût politique des conflits d'intérêts. Les essais du CIE s'étaient développés dans le cadre d'une relation serrée, non régulée et non publique, entre industriels et épidémiologistes. Cette proximité présentait de nombreux avantages pour le CIE, dont des financements et l'accès aux dernières innovations. Mais dans un contexte de défiance à l'égard de l'autorité médicale et de l'État, ces liens représentaient une réelle faiblesse institutionnelle.

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

<sup>25</sup> Guérin N. Méthodes d'évaluation de la couverture vaccinale. *Revue des professions de santé*, 1982, 28 : 62.

<sup>26</sup> Henderson R, Sundaresan T. Cluster sampling to assess immunization coverage : a review of experience with a simplified sampling method. *Bulletin of the World Health Organization*, 1982, 60, 2 : 253-260.

<sup>27</sup> Thomas G. La France et l'Afrique, en parallèle..., *op. cit.* : 155-170.



## APRÈS LE DÉLUGE DES DONNÉES DE SANTÉ

# Données massives et médico-administratives Nouveaux enjeux pour l'épidémiologie

## Grégoire Rey

Directeur du CépiDc (Centre d'épidémiologie sur les causes médicales de décès)  
Inserm, Le Kremlin-Bicêtre

### Résumé

La production de données massives, le développement de méthodes adaptées à leur traitement et le mouvement d'ouverture et de partage des données publiques viennent modifier les métiers et les pratiques de l'épidémiologie. Le Système National des Données de Santé (SNDS) créé en 2016 est emblématique de ces enjeux. Il peut être réutilisé avec un coût limité, sans perdu de vue, sans biais de mémoire, et pour des pathologies et des expositions rares. Cependant, sa complexité peut fréquemment amener à de nombreuses erreurs d'exploitation et son chaînage avec d'autres bases de données accroît sensiblement son intérêt. Seul le développement d'outils mutualisés de documentation, d'exploitation statistique et de chaînage permettrait d'en tirer le meilleur parti. Ce sont les ambitions du Health Data Hub, créé par la loi en 2019, et dont la mission s'étend à la quasi-totalité des données personnelles de santé produites par le service public.

**Mots-clés :** données massives, méthodes épidémiologiques, SNDS, données administratives des demandes de remboursement des soins de santé.

### Abstract

#### **Massive and administrative health data : new challenges for epidemiology**

*The production of massive data, the development of methods to process them and the open data and data sharing movement are changing the professions and practices of epidemiology. The National System of Health Data (SNDS) created in 2016 is emblematic of these issues. It can be reused with a limited cost, without lost to follow up, memory bias, and for rare pathologies and exposures. However, its complexity can frequently lead to many operating errors and its linkage with other databases significantly increases its interest. Only the development of shared tools for documentation, statistical exploitation and chaining would make the best of it. These are the ambitions of the Health Data Hub. The hub was created by law in 2019, and its mission extends to almost all personal health data produced by the public service.*

**Keywords :** Massive data, Epidemiologic methods, SNDS, Healthcare data

## Production, exploitation et partage des données de santé

Au cours des vingt dernières années, le volume et la diversité des données disponibles pour des traitements de recherche dans le domaine des sciences de la vie et de la santé, et donc pour l'épidémiologie, a connu un accroissement sans précédent. Celui-ci a été et est toujours alimenté par :

- le développement des systèmes d'information en santé et la dématérialisation des documents administratifs en général, et dans le domaine de la santé en particulier,
- la génération de données « omiques », dont font partie les données génomiques (issues du décodage du génome), transcriptomiques (issues de la transcription du génome en ARN) et protéomiques (issues des protéines codées par le génome),
- la production de données biologiques, physiologiques et environnementales (exposome), en particulier avec l'utilisation de plus en plus généralisée de GPS et d'objets connectés.

### *Données massives et développement des outils informatiques et statistiques*

Les progrès dans la production de données se sont accompagnés de développement des moyens de calculs et des méthodes statistiques tout aussi spectaculaires. Par exemple, l'évolution des méthodes d'analyse de l'effet des vagues de chaleur sur la mortalité en France, et plus particulièrement la comparaison de l'analyse des vagues de 1976 et de 2006, est à cet égard édifiante. Des données équivalentes étaient produites sur les deux périodes, la totalité des enregistrements provenant des certificats de décès et traitée depuis 1968 par l'Inserm, et les données de températures enregistrées par Météo France dans des stations couvrant l'ensemble du territoire. Pour autant, malgré l'observation d'une hausse de 22 % de la mortalité dans des communiqués de presse de l'Inserm en octobre 1976, dont on sait aujourd'hui qu'elle correspondait à plus de 5 000 décès en excès<sup>1</sup>, le lien avec la chaleur n'était alors que suggéré. Trente ans plus tard, pour analyser la vague de chaleur sensiblement plus intense survenue entre le 11 et le 28 juillet 2006, il a été possible de construire et d'utiliser un modèle prédictif basé sur les températures observées pour calculer non seulement le nombre de décès en excès (environ 2200), mais également le nombre de décès en excès que l'on aurait attendus si la relation entre chaleur et mortalité était restée identique à celle d'avant le traumatisme de la canicule de 2003 (environ 6500 décès). Cette différence de 4300 décès a pu ainsi être interprétée comme une baisse de la vulnérabilité de la population française à la chaleur due notamment à la mise en place d'un système de prévention et d'alerte en 2004<sup>2</sup>. Ainsi, les outils

<sup>1</sup> Rey G, Jouglé E, Fouillet A, *et al.* The impact of major heat waves on all-cause and cause-specific mortality in France from 1971 to 2003. *Int Arch Occup Environ Health*, 2007, 80, 7 : 615-26.

<sup>2</sup> Fouillet A, Rey G, Wagner V, *et al.* Has the impact of heat waves on mortality changed in France since the European heat wave of summer 2003 ? A study of the 2006 heat wave. *Int J Epidemiol*, 2008 ; 37, 2 : 309-17.

statistiques permettant d'accroître les connaissances pour améliorer l'aide à la décision et évaluer les interventions ont considérablement progressé.

### *L'Open Data dans le champ de la santé*

Enfin, sur les dix dernières années, le mouvement dit *Open Data* a, peu à peu, gagné le domaine de la santé. Ce courant idéologique s'inscrit dans une tendance qui considère l'information publique comme un bien commun dont la diffusion est d'intérêt public et général. Suivant cette logique dans le domaine de la santé, la transparence des informations produites doit nécessairement, in fine, donner lieu à une amélioration des décisions et de l'état de santé général. Il est largement admis que l'*Open Data* est un moyen nécessaire pour parvenir à un haut niveau de reproductibilité des résultats d'analyse, et donc d'intégrité scientifique<sup>3</sup>. En revanche, l'*Open Data* n'est pas suffisante pour garantir le bon usage des données, si on considère que certaines d'entre elles peuvent être biaisées ou induire des comportements contre productifs. Les Anglo-Saxons le signalent régulièrement à propos des indicateurs de qualité des soins hospitaliers<sup>4</sup>. Le principe vertueux de l'*Open Data* repose donc sur le présupposé que tout indicateur couramment utilisé fait nécessairement l'objet d'un débat public raisonnable et intelligible, guidant spontanément à une amélioration des outils et méthodes. Sur un plan plus économique, l'*Open Data* est un excellent moyen de limiter le gâchis des données, bien souvent très coûteuses à produire et insuffisamment utilisées par les producteurs eux-mêmes. L'accélération de l'innovation dans les usages des données est de plus vue comme un facteur de croissance économique important.

En pratique, en France, dans le champ de la santé, ce mouvement a donné lieu à la création d'une commission *Open Data* en santé, dont les conclusions dans son rapport de 2014<sup>5</sup> ont été à l'origine de l'article 193 de la loi de modernisation de notre système de santé. Cette loi crée le système national des données de santé (SNDS), regroupement et chaînages de bases de données de santé fondées sur des procédures administratives, dans le but d'ouvrir l'accès à ces données pour le grand public, les administrations et la recherche, les études et les évaluations. Allant, encore plus loin, l'article 41 de la loi relative à l'organisation et à la transformation du système de santé de 2019 étend le périmètre du SNDS notamment à toutes les données recueillies dans le cadre de soins remboursés par l'assurance maladie et aux données d'enquête dans le domaine de la santé chaînées avec les données administratives de santé.

### *Le partage de données et la recherche académique*

Dans le champ de la recherche académique, les financements accordés aux projets et l'évaluation des équipes de recherche ont fait intervenir de façon croissante le critère du partage des données. C'est notamment le cas des investissements d'avenir accordés pour les grandes cohortes, ou de l'évaluation des registres, deux cadres qui demandent aux porteurs la définition de critères ouverts et transparents pour permettre à des équipes tierces d'accéder aux données produites. Parallèlement les institutions de recherche internationales comme l'Inserm ont pris des engagements publics en ce sens<sup>6</sup>. Les revues scientifiques exigent de plus en plus souvent un engagement à pouvoir transmettre

<sup>3</sup> Research Integrity, Open Science, and Health Policy, en ligne [<https://www.bmj.com/content/363/bmj.k4309/rr-0>] (consulté le 11 mai 2019).

<sup>4</sup> Shahian DM, Iezzoni LI, Meyer GS, *et al.* Hospital-wide mortality as a quality metric: conceptual and methodological challenges. *Am J Med Qual Off J Am Coll Med Qual*, avr. 2012, 27, 2 : 112-23.

<sup>5</sup> Commission open data en santé : rapport, en ligne, [<http://www.ladocumentationfrancaise.fr/rapports-publics/144000397/index.shtml>] (consulté en juillet 2014).

<sup>6</sup> Walport M, Brest P. Sharing research data to improve public health. *Lancet Lond Engl*, 12 fév. 2011 ; 377 (9765) : 537-9.

les données pour reproduire les analyses, et demandent parfois directement une transmission des données. Les principes FAIR (pour *Findable, Accessible, Interoperable* et *Reusable*) guident, autant que possible, la démarche d'ouverture et de partage de ces données<sup>7</sup>.

Ces évolutions changent les métiers de l'épidémiologie, qui historiquement consistait souvent pour une étude à ce qu'une même équipe formalise un questionnement scientifique plus ou moins vaste, élabore et mette en œuvre un protocole de recueil de données dédié et assure son analyse. L'équipe avait alors en son sein la connaissance fine de la donnée, de ces limites potentielles, et des procédures permettant sa mise en qualité. La réutilisation scientifique de données complexes et imparfaites produites par des tiers suppose la mise en œuvre de démarches méthodologiques nouvelles.

Nous donnons dans la suite de cet article une illustration de ces enjeux appliqués à l'utilisation du SNDS, dans son acception médico-administrative définie par la loi de modernisation de notre système de santé du 26 janvier 2016. Des exemples sont donnés de problématiques associées à la production des données des causes médicales de décès.

## Le SNDS pour la recherche

Le Système National des Données de Santé (SNDS) regroupe et chaîne les données suivantes :

- le Système National d'Information Inter-Régimes de l'Assurance Maladie (SNIIRAM) et en son sein le DCIR (Datamart Consommation Inter-Régime). Celui-ci est constitué essentiellement des données de remboursement de soins issues des consultations médicales, incluant les honoraires pratiqués, et des feuilles de soins numérisées ou dématérialisées, permettant notamment de connaître la liste des médicaments achetés et remboursés. Le système, géré par la Caisse Nationale d'Assurance Maladie (Cnam), permet également de connaître le régime d'appartenance et la présence d'une Affection de Longue Durée (ALD),
- le Programme de Médicalisation des Systèmes d'Information (PMSI) est une base de données collectée par l'Agence Technique de l'Information Hospitalière (ATIH), elle rassemble les informations concernant les diagnostics des patients, les actes prodigués lors des séjours, et leur origine et leur destination, à des fins de calcul des enveloppes financières annuelles attribuées aux établissements de santé publics et privés,
- la Base des Causes Médicales de Décès (BCMD), qui contient les diagnostics et entités nosologiques mentionnés sur les certificats de décès, collectée et traitée par l'Institut National de la Santé et de la Recherche Médicale (Inserm),
- les données médico-sociales relatives aux personnes handicapées, issues des Maisons Départementales des Personnes Handicapées (MDPH) et rassemblées par la Caisse Nationale de Solidarité pour l'Autonomie (CNSA), ces données n'étant pas encore intégralement standardisées et centralisées au moment de la rédaction de cet article.

Ces données, dès lors qu'elles sont produites, sont pseudonymisées (elles ne contiennent pas d'identifiant direct en clair) et transmises à la Cnam, organisme responsable du traitement SNDS.

L'accès à ces données est réglementé de façon précise<sup>8</sup>. Dans tous les cas, les accès, les extractions, les traitements et les appariements aux données du SNDS doivent se faire dans un cadre conforme à un référentiel de sécurité défini par un arrêté. Ce haut niveau de sécurité résulte de la

<sup>7</sup> Wilkinson MD, Dumontier M, Aalbersberg IJ, *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*, 15 mars 2016 ; 3 : 160018.

<sup>8</sup> Lesaulnier F. Recherche en santé et protection des données personnelles à l'heure du Règlement général relatif à la protection des données. *Médecine Droit*, 2019, 158 : 103-11.  
[<http://www.sciencedirect.com/science/article/pii/S124673911830112X>] (consulté le 15 nov 2018).

sensibilité de données de santé recueillies en routine, couvrant la quasi-totalité de la population Française.

Outre le mouvement *Open Data*, la mise en qualité progressive de ces données<sup>9</sup> et le nombre croissant d'études et de recherches menées ont peu à peu mis en évidence leur fort potentiel pour la recherche en santé publique, notamment en épidémiologie<sup>10</sup>. En effet, contrairement aux données d'enquêtes épidémiologiques nécessitant un recueil auprès des personnes, elles ne sont soumises ni à des biais de mémoire, ni à des risques de perdus de vue. C'est pourquoi le nombre de publications portant sur le SNIIRAM a crû à un rythme très soutenu au cours des années 2010. Entre autres publications, des travaux retentissants ont été menés sur le nombre de décès associés à la prise de Médiator hors autorisation de mise sur le marché<sup>11</sup>, sur le lien entre prise de contraceptif oraux et le risque d'embolie pulmonaire, d'AVC et d'infarctus<sup>12</sup>, ou sur le risque d'hospitalisation et de décès associé à la prise de Baclofène pour traiter l'alcoololo-dépendance<sup>13</sup>. Comme les exemples donnés ci-dessus le suggèrent, de nombreuses études sur le SNIIRAM ont été menées par ou en collaboration avec la Cnam. D'autres équipes ont, peu à peu, acquis une bonne connaissance de ces données, mais malgré une documentation et des outils de communication renforcés (forums, formations et réunions régulières), le risque d'obtenir des résultats biaisés par des artefacts de production demeure élevé. Cela est d'autant plus vrai lorsque l'on considère les études portant sur les multiples composantes du SNDS.

## Enjeux méthodologiques liés aux données du SNDS

Le SNDS est un système de données médico-administratives. À l'exception des données sur les causes médicales de décès, la finalité de leur production n'est pas l'exploitation statistique à des fins d'épidémiologie.

### *Qualité et comparabilité*

Pourtant, même les données des causes médicales de décès doivent être traitées avec précaution. En effet, les données finalement en base sont d'abord tributaires de la qualité et de la comparabilité des déclarations faites par les médecins certificateurs. Or l'information déclarée dépend de la connaissance que le médecin a du cas, du niveau de précision qu'il peut déclarer, et éventuellement de la transmission de cette information à l'Inserm. Pour exemple des conséquences d'une comparabilité limitée à l'étape de la transmission de l'information, la distribution spatiale du nombre de décès par suicide, tels qu'ils sont enregistrés à l'Inserm, est biaisée par rapport à la distribution réelle. Les zones géographiques dans lesquelles la proportion de suicides est la plus élevée

<sup>9</sup> Tuppin P, Rudant J, Constantinou P, *et al.* Value of a national administrative database to guide public decisions : From the système national d'information interrégimes de l'Assurance Maladie (SNIIRAM) to the système national des données de santé (SNDS) in France. *Rev Epidemiol Sante Publique*, oct. 2017 ; 65 Suppl 4 : S149-67.

<sup>10</sup> Weill A, Rudant J, Coste J. Utilisation des données de l'assurance maladie française pour étudier l'usage et les effets des médicaments en vie réelle : revue de 216 articles publiés entre 2007 et 2016. *Rev Epidemiol Sante Publique*, 1<sup>er</sup> juin 2017 ; 65 : S127.

<sup>11</sup> Weill A, Païta M, Tuppin P, *et al.* Benfluorex and valvular heart disease : a cohort study of a million people with diabetes mellitus. *Pharmacoepidemiol Drug Saf*, déc. 2010 ; 19, 12 : 1256-62.

<sup>12</sup> Weill A, Dalichampt M, Raguideau F, *et al.* Low dose oestrogen combined oral contraception and risk of pulmonary embolism, stroke, and myocardial infarction in five million French women: cohort study. *BMJ*, 10 mai 2016 ; 353: i2002.

<sup>13</sup> Chaignot C, Zureik M, Rey G, Dray-Spira R, *et al.* Risk of hospitalisation and death related to baclofen for alcohol use disorders: Comparison with nalmefene, acamprosate, and naltrexone in a cohort study of 165 334 patients between 2009 and 2015 in France. *Pharmacoepidemiol Drug Saf*, nov. 2018 ; 27, 11 : 1239-48.

(Bretagne, Franche-Comté) sont les régions dans lesquelles les causes de décès sont les mieux renseignées, tandis que les zones dans lesquelles cette proportion est la plus faible (Île de France, Rhône-Alpes) sont celles dans lesquelles les Instituts Médico-Légaux renvoient le moins d'information, générant une forte proportion de données manquantes<sup>14</sup>. Dans ce cas, une analyse des données brutes induirait des erreurs d'appréciation, voire des prises de décision sur le ciblage d'une politique de prévention, erronée. Plus généralement, l'identification et la correction de ce type de biais nécessitent des études spécifiques guidées par la connaissance fine du processus de production de la donnée au plus près du terrain<sup>15</sup>. Ce questionnement typique d'une approche épidémiologique rigoureuse doit être posé régulièrement, et généralisé à l'ensemble des données du SNDS.

### *Documentation*

La donnée sur les causes de décès est codée par l'Inserm en suivant les règles de la Classification Internationale des Maladies (CIM). Cette classification évolue avec la connaissance médicale, et il est fréquent que ces évolutions impactent la statistique finale en France et à l'étranger<sup>16</sup>. On constate par exemple que le passage de la neuvième à la dixième révision de la CIM est à l'origine d'une baisse de 30 % du nombre de décès ayant la pneumonie comme cause initiale. En tant qu'acteur de la mise en œuvre des règles de la CIM, le producteur des données identifie aisément l'origine de cette évolution. En revanche, la diffusion de ces données à une équipe tierce pose la question de la formation des équipes et de la documentation de ces évolutions. Cela est d'autant plus difficile pour l'Inserm s'il s'agit de faire accéder l'utilisateur à toutes les évolutions de codage, et pour la Cnam et l'ATIH s'il s'agit de documenter toutes les raisons possibles d'évolutions de remboursement ou d'applications des multitudes de règles. Les enjeux de la constitution d'une documentation structurée et d'outils pour faciliter sa consultation sont donc essentiels.

### *Identification de phénotype*

Dans le cadre d'études épidémiologiques, le chercheur souhaite souvent pouvoir identifier la date d'incidence d'une pathologie, son stade ou sa gravité. Dans le SNDS, ces informations sont recueillies suivant différents objectifs et sont rarement identifiables de façon claire et univoque. Il est alors nécessaire de combiner les informations de différentes sources (médecine de ville et hospitalisation) pour constituer l'approximation d'une variable d'incidence avec une fiabilité acceptable. Ces combinaisons d'information sont également appelées algorithmes d'identification de phénotype. Ces derniers sont aujourd'hui définis le plus souvent à dire d'experts. Différentes initiatives visant à généraliser ou centraliser la constitution d'algorithmes ont été menées par la Cnam<sup>17</sup> ou le réseau Redsiam<sup>18</sup>. La validation et/ou l'élaboration de ces algorithmes en population générale sont idéalement obtenues en présence d'un *gold standard* obtenu par le chaînage des

<sup>14</sup> Richaud-Eyraud E, Rondet C, Rey G. Transmission of death certificates to CépIdc-Inserm related to suspicious deaths, in France, since 2000. *Rev Epidemiol Sante Publique*, mars 2018 ; 66 (2) : 125-133

<sup>15</sup> Richaud-Eyraud E, Gigonzac V, Rondet C, *et al.* État des lieux des pratiques et de la rédaction des certificats de décès par les instituts médico-légaux en France, en 2016, dans la perspective de la mise en place d'un volet complémentaire du certificat de décès. *Rev Médecine Légale*, 1<sup>er</sup> févr. 2018 ; 9, 1 : 1-9.

<sup>16</sup> Rey G, Aouba A, Pavillon G, *et al.* Cause-specific mortality time series analysis : a general method to detect and correct for abrupt data production changes. *Popul Health Metr*, 2011 ; 9 : 52.

<sup>17</sup> Cartographie des pathologies et des dépenses, [<https://www.ameli.fr/l-assurance-maladie/statistiques-et-publications/etudes-en-sante-publique/cartographie-des-pathologies-et-des-depenses/index.php>] (consulté le 21 nov. 2017).

<sup>18</sup> Goldberg M, Carton M, Doussin A, *et al.* Le réseau REDSIAM (Réseau données Sniiram) -Spécial REDSIAM. *Rev Epidemiol Sante Publique*, 1<sup>er</sup> oct. 2017 ; 65 : S144-8.

données des patients identifiés par un registre de morbidité<sup>19</sup> avec les données de ces patients dans le SNDS, et possiblement par l'application de méthodes d'apprentissage statistique<sup>20</sup>.

### *Incomplétude des champs couverts*

Malgré leur volume, certaines données essentielles à l'épidémiologie sont absentes ou très insuffisamment renseignées dans le SNDS. C'est notamment le cas de données relatives au mode de vie, au niveau socio-économique, et aux diverses expositions environnementales. Une approximation du niveau socio-économique calculé à l'échelle de la commune de domicile est régulièrement utilisée<sup>21</sup>. Bien qu'intéressant pour de nombreux traitements portant par exemple sur les inégalités sociospatiales de santé, cette approximation est très insuffisante pour des analyses fines. Des chaînages avec des données sociales individuelles, comme les données du recensement ou les données administratives de l'assurance vieillesse sont absolument nécessaires<sup>22</sup>.

### *Dépendance des biais au protocole de l'étude*

Pour toutes les raisons évoquées plus haut, il existe des écarts aussi bien sur les variables à expliquer (survenue de maladie ou état de santé) que sur les variables explicatives (soin, expositions diverses) entre les informations que l'on souhaite mesurer et les informations réellement mesurées. Ces écarts sont susceptibles de créer des biais lorsque la mesure d'une association se fait dans une finalité d'interprétation causale, comme c'est souvent le cas en épidémiologie<sup>23</sup>. Sans viser une qualité parfaite de l'information, inexistante dans l'absolue, il est nécessaire d'interroger ce type de biais pour les différents usages des données et en fonction des objectifs poursuivis par les études. À défaut de pouvoir éliminer ces biais, l'incertitude que ces écarts génèrent sur les estimations doit être prise en compte.

### *Méthodes statistiques associées*

À l'international, les données médico-administratives suscitent tout autant d'intérêt. En raison de la grande quantité d'informations qu'elles peuvent chainer pour chaque individu et de la taille des populations couvertes, elles permettent de construire des populations exposées et non exposées, par exemple à la prise d'un médicament, présentant de nombreuses caractéristiques communes et considérées comme comparables, pour émuler des pseudo-interventions contrôlées<sup>24</sup>. Les méthodes statistiques développées à cette fin ne permettent pas de s'affranchir de toutes les précautions évoquées dans les paragraphes précédents, mais elles sont innovantes, souples et avec un minimum d'hypothèses non vérifiées, visant à tirer le meilleur parti de la grande dimension des données<sup>25</sup>.

<sup>19</sup> Défini comme un recueil continu et exhaustif de données nominatives, intéressant un ou plusieurs événements de santé, dans une population géographiquement définie.

<sup>20</sup> S'appuyant sur des corpus de données, les plus souvent volumineux, pour optimiser avec un minimum d'hypothèses sur la structure des données la capacité prédictive d'un algorithme.

<sup>21</sup> Rey G, Jouglu E, Fouillet A, Hemon D. Ecological association between a deprivation index and mortality in France over the period 1997-2001: variations with spatial scale, degree of urbanicity, age, gender and cause of death. *BMC Public Health*, 2009 ; 9 : 33.

<sup>22</sup> Rey G. Mesures des inégalités socio-spatiales de santé. Séminaire inégalités sociales de santé, 3 déc. 2015, DREES, Paris.

<sup>23</sup> Hernán MA, Robins J. Causal Inference. Boca Raton: Chapman & Hall/CRC, à paraître.

<sup>24</sup> Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. *Am J Epidemiol*, 15 avr. 2016 ; 183, 8 : 758-64.

<sup>25</sup> Van der Laan MJ, Rose S. *Targeted learning: causal inference for observational and experimental data*. New York, Springer, 2011.

Leur mise en œuvre fait appel à des techniques d'apprentissage statistique nécessitant des infrastructures informatiques de calcul hautes performances. Il est important de préciser ici que le résultat optimal obtenu par ces nouvelles méthodes peut être très insatisfaisant si les biais évoqués ci-dessus, ou d'autres que nous n'avons pas la place de discuter ici, sont trop importants. C'est toujours le rôle de l'épidémiologiste d'en évaluer la pertinence.

### Enjeux organisationnels

Pour résumer les éléments permettant de capitaliser l'ouverture des données de santé, en particulier celles du SNDS, il est indispensable de disposer :

- d'une expertise très fine sur les données produites, et une capacité à documenter cette connaissance et à la faire partager en structurant l'information pour la rendre interrogeable, autonomisant ainsi les chercheurs dans leur exploitation sur une finalité donnée,
- d'outils informatiques et de gestion des accès aux données permettant de respecter la confidentialité des traitements en suivant la réglementation et les référentiels de sécurité applicables, tout en accédant à des ressources de calculs hautes performances permettant de traiter des données volumineuses avec des méthodes innovantes,
- d'un cadre de gouvernance et de réglementation simplifiée des données pour permettre notamment de réaliser des chaînages entre bases de données, et ainsi obtenir tout autant des données couvrant des champs complémentaires (ex. données sociales) que des données d'incidence faisant office de *gold standards* (ex. données de registres).

Ces trois éléments sont en étroite interaction, typiquement parce que les règles de gouvernance et la réglementation des données sont souvent associées à des principes de sécurité des traitements et d'expertise des producteurs de données. Aucune équipe ne dispose aujourd'hui à la fois de l'expertise épidémiologique et de compétences expertes sur les données produites, sur la réglementation, et en système d'information, et l'obtention pour chaque équipe d'épidémiologie de l'ensemble de ces compétences demanderait des moyens considérables. Ces constats incitent naturellement à identifier le besoin d'un cadre structurant, multidisciplinaire et mutualisé.

C'est dans ce contexte que les annonces présidentielles de mars 2018 sur la stratégie nationale en intelligence artificielle ont permis la mise en place d'un *Health Data Hub*, appelé Plateforme des Données de Santé dans la loi de 2019 relative à l'organisation et à la transformation du système de santé. Cette plateforme prendra la forme d'un Groupement d'Intérêt Public (GIP), aura pour ambition de permettre le partage et l'appariement beaucoup plus généralisé de données de santé, et proposera des outils facilitant leur exploitation dans des conditions sécurisées.

Si elles sont mises au service de la recherche en épidémiologie, s'appuyant notamment sur le savoir accumulé et les initiatives existantes, ces évolutions pourraient largement faciliter la mise en place de projets ambitieux dont les résultats seront très utiles à la santé publique.

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

## SANTÉ GLOBALE

### Santé Globale et mesure

Un nouveau paradigme pour de nouveaux acteurs

**Grégoire LURTON**

Health Metrics Specialist, Bluesquare, Doctorant, University of Washington

#### Résumé

Le champ de l'action sanitaire global a été bouleversé dans les années 2000 par l'apparition de nouveaux acteurs. Ces acteurs ont renouvelé la gouvernance de la santé globale en introduisant de nouvelles pratiques axées sur la quantification. Dans le même temps, de nouveaux outils de collecte de données et de nouvelles pratiques computationnelles se sont développés pour l'analyse des données sur la santé des populations. Ces évolutions se sont conjuguées pour donner naissance à un nouveau paradigme de mesure en santé globale, moins déterminé par la définition de normes de collecte de données et plus axé sur l'analyse et la construction d'indicateurs composites. Ce nouveau paradigme a renforcé la redistribution des pouvoirs dans la santé globale, avec une légitimité accrue des nouveaux acteurs tenant un discours scientifique et au détriment d'acteurs politiques traditionnels.

**Mots-clés** : santé globale, mesure, recherche sur les systèmes de santé publique, gouvernance, science des données.

#### Abstract

##### **Measurement in Global Health : a new paradigm for new actors**

*Global Public Health has been radically changed in the early 2000s when new actors entered the field. By introducing new methods based on quantification, these actors renewed Global Health governance. In the meantime, new tools were developed for data analysis and population health modeling. These evolutions have interacted with each other to create a new paradigm for measurement in Global Health, geared towards the modeling of composite metrics more than towards the definition of measurement norms. This new paradigm strengthened a shift in power in Global Health, where actors able to hold a scientific discourse accrued their legitimacy, while traditional political actors place shrank.*

**Keywords** : global health, measurement, health systems research, governance, data science.

Si l'émergence du terme « santé globale » est associée à l'apparition de nouveaux acteurs et de nouveaux moyens de financement dans le domaine de la santé internationale, elle l'est aussi à l'émergence d'un champ scientifique marqué par la création de diplômes universitaires et de revues scientifiques spécialisés<sup>1</sup>. On peut sommairement décrire ce champ comme la combinaison d'un champ politique de l'action sanitaire et d'un projet de connaissance envisageant les questions de santé à un niveau global. L'histoire de l'émergence de ce champ s'inscrit par ailleurs dans celle de luttes d'influences autour de la maîtrise de la gouvernance sanitaire mondiale<sup>2</sup>.

La quantification est elle-même à la croisée de dimensions politique et scientifique et Alain Desrosières a décrit la dualité de l'activité du statisticien, entre administrateur et savant<sup>3</sup>. Dans le cadre de la santé globale, cette articulation prend des formes spécifiques liées à l'identité des acteurs impliqués et à un moment technique et politique spécial. Dans cet article, en tant que *Data Scientist* spécialisé dans les données sanitaires des pays en développement et doctorant auprès de l'Institute for Health Metrics and Evaluation (IHME), j'analyse comment le développement de la santé globale a contribué à redéfinir les conditions de la quantification des questions sanitaires.

## Construction d'une infrastructure de mesure pour la santé globale

### *Les nouveaux acteurs de la santé globale*

Dans un livre consacré à la mesure du climat, Michael Edward distingue deux grands projets des infrastructures globales de mesure<sup>4</sup>. Par *Making Global Data*, il désigne les efforts pour créer des outils de collecte de données standardisées à travers le monde, comme la définition par l'Organisation mondiale de la santé (OMS) de la classification des maladies et des causes de décès ou les normes de collecte de tout type de données, du suivi des patients séropositifs aux indicateurs de qualité des laboratoires. Par *Making Data Global*, Michael Edwards désigne les activités dont l'objectif est de fournir des résultats uniformes au niveau mondial, à partir de données disparates. Le

<sup>1</sup> La plupart des grandes revues médicales internationales ont aujourd'hui une déclinaison pour la *Global Health*, comme *The Lancet Global Health* ou *The BMJ Global Health*.

<sup>2</sup> Brown T, Cueto M, Fee E. The World Health Organization and the Transition From "International" to "Global" Public Health. *American Journal of Public Health*, 2006, 96, 1 : 62-72.

<sup>3</sup> Desrosières A. L'administrateur et Le Savant. *Courrier des Statistiques*, 1998, 87-88 : 71-79.

<sup>4</sup> Edwards P. *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*. Cambridge : The MIT Press, 2010.

*Global Burden of Disease (GBD)*<sup>5</sup>, qui modélise le fardeau des maladies dans le monde en compilant des milliers de sources de données est un exemple type de cette approche<sup>6</sup>.

L'articulation de ces deux projets dans la santé globale reflète les évolutions politiques et techniques des dernières décennies. Au tournant des années 2000, la gouvernance sanitaire mondiale a été bouleversée par l'arrivée de nouveaux acteurs, que ce soient des acteurs privés comme la *Bill and Melinda Gates Foundation (BMGF)*, ou de nouvelles institutions internationales comme le Fonds mondial de lutte contre le sida, la tuberculose et le paludisme (*The Global Fund to fight Aids, Tuberculosis and Malaria, GFATM*)<sup>7</sup>. Ces organisations sont devenues des opérateurs essentiels de la santé globale, la BMGF et le GFATM dépensant chaque année plus de trois milliards de dollars chacun dans des programmes de santé globale. Ces organisations s'inspirant des pratiques du *new public management*, la culture de l'audit, la mesure des résultats et l'évaluation de l'impact des politiques et programmes de santé deviennent des enjeux essentiels pour les acteurs de la santé globale<sup>8</sup>, et généralisent un prisme analytique axé sur le suivi de batteries d'indicateurs censés mesurer l'avancement des projets et leur impact sur la santé des populations<sup>9</sup>.

### *Morcellement des systèmes statistiques*

La construction de l'infrastructure de mesure de la santé globale se fait dans une interaction entre des capacités techniques et des enjeux de pouvoir sur la production du chiffre. Les évaluations d'impact des projets sanitaires sont financées par les acteurs de la santé globale pour justifier leur action et promouvoir le succès de projets qui se situent sur un marché global<sup>10</sup>. David Reubi a décrit comment cette rencontre entre la culture de l'audit et l'expertise épidémiologique a donné naissance à une « redevabilité épidémiologique » visant à faire le lien direct entre des activités de programmes de santé et des modifications de la santé des populations<sup>11</sup>. Les méthodes pour établir cette redevabilité nécessitent des capacités techniques et des infrastructures de calcul importantes dont seuls quelques acteurs disposent et sont capables de produire et de faire publier un discours quantitatif répondant aux canons de la connaissance en santé globale qu'ils contribuent eux-mêmes à définir.

Les systèmes statistiques nationaux étant souvent considérés comme insuffisants pour répondre aux exigences de « suivi-évaluation », on a assisté à l'apparition de systèmes parallèles créant au nom de la spécificité des infrastructures de collecte de données différentes pour chaque acteur, et parfois pour chaque projet. Cette privatisation des systèmes de mesure mène à une confiscation de la production quantitative par des acteurs souvent étrangers, affichant les codes de la scientificité de

<sup>5</sup> Murray C, Lopez A. Measuring the Global Burden of Disease. *New England Journal of Medicine*, 2013, 369 : 448-457.

<sup>6</sup> Ce projet, débuté dans les années 1990, à la Banque Mondiale, est hébergé depuis 2007 à l'IHME, où il bénéficie d'investissements conséquents de la Fondation Bill et Melinda Gates. Les résultats du GBD constituent aujourd'hui une référence incontournable dans la compréhension de la santé mondiale et informent les stratégies de la Fondation Gates et de nombreux acteurs de la santé globale.

<sup>7</sup> Buissonnière M. La nouvelle donne de la santé globale : dynamiques et écueils. *Revue internationale de politique de développement*, 2012, 3, 3 : 77-92.

<sup>8</sup> Gaudillière JP. Un Nouvel Ordre Sanitaire International ? *Ecologie & Politique*, 2016, 1 : 107-24. ; Szlezák N, Bloom B, Jamison D *et al.* The Global Health System: Actors, Norms, and Expectations. *Transition, PLoS Medicine*, 2010, 7, 1 : e1000183.

<sup>9</sup> Engle Merry S. Measuring the World: Indicators, Human Rights, and Global Governance. *Current Anthropology*, 2011, 52, S3 : S83-95.

<sup>10</sup> Rajkotia Y. Beware of the Success Cartel : A Plea for Rational Progress. *BMJ Global Health*, 2018, 3, 6 : e001197.

<sup>11</sup> Reubi D. Epidemiological Accountability : Philanthropists, Global Health and the Audit of Saving Lives. *Economy and Society*, 2018, 47, 1 : 83-110.

leur travail, et jetant le discrédit sur la capacité des acteurs nationaux à fournir des données et des analyses de qualité. Ainsi, des ONG internationales intervenant dans des hôpitaux publics se dispensent de transmettre les rapports standards des autorités nationales – arguant que ces rapports peuvent être mal définis ou ne leur paraissent pas utiles – et privilégient leurs propres systèmes de rapportage validés par des comités scientifiques localisés dans les pays du Nord d'où elles sont issues.

## **Making Global Data : la collecte de données en santé globale**

La collecte de données dans les pays en développement a été transformée par des innovations récentes. Citons deux exemples illustrant l'ampleur de cette transformation.

### *Le District Health Information System : un outil clé*

La collecte des données des services de santé dans les pays en développement est souvent faite par transmission et compilation successives de données à différents niveaux de la pyramide sanitaire. Par exemple, le nombre de patients vus en consultation est compilé par chaque formation sanitaire, puis agrégé au niveau du district sanitaire, puis aux niveaux provincial et national. L'informatisation de ces systèmes a longtemps reposé sur la saisie de données dans des fichiers transmis par email ou sur clé USB. Chaque système de santé définissant une solution technique *ad hoc* en utilisant des outils parfois peu adaptés à la saisie et au suivi de données<sup>12</sup>, la capacité d'analyse de ces données se limitait souvent au suivi d'indicateurs prédéfinis<sup>13</sup>. Depuis les années 2000, un système d'entrepôt de données destiné spécifiquement à la gestion de données sanitaires, a été développé par des chercheurs de l'Université d'Oslo<sup>14</sup>. Le *District Health Information System* (DHIS) est aujourd'hui utilisé dans 67 pays pour le suivi de l'activité des systèmes sanitaires. Il est devenu un outil standard permettant la collecte, la transmission et le partage de données sanitaires pour deux raisons principales. D'une part, il a construit une base de données reproduisant la structure typique d'un système d'information sanitaire. D'autre part, il promeut une nouvelle économie de la collecte de données basée sur la réutilisation « d'éléments de données » et sur la mise en place de « standards flexibles »<sup>15</sup>. Cette professionnalisation de la collecte de données par l'utilisation d'un outil spécialisé conduit à une meilleure disponibilité des données des systèmes de santé et à une meilleure transférabilité de ces données pour des usages secondaires.

### *Les outils mobiles de collecte de données : un changement fondamental*

Un second exemple des changements dans les capacités de collecte de données est le développement d'outils mobiles. Avant l'apparition de ces outils, la mise en œuvre d'une étude sur un territoire comme celui République Démocratique du Congo (RDC) nécessitait le transport de volumes de questionnaires papier considérables. Les questionnaires remplis devaient être acheminés vers les centres d'étude pour être saisis. À chaque étape de cette chaîne, des données pouvaient disparaître par perte ou destruction du matériel d'enquête. De plus, la longue chaîne de production

<sup>12</sup> Ces fichiers sont la plupart du temps issus des logiciels bureautiques Microsoft : Excel, Access, parfois Word ou d'autres logiciels peu adaptés à la gestion de données.

<sup>13</sup> Braa J, Hedberg C. The Struggle for District-Based Health Information Systems in South Africa. *The Information Society*, 2002, 18, 2 : 113-27.

<sup>14</sup> Braa J. A Data Warehouse Approach Can Manage Multiple Data Sets. *Bulletin of the World Health Organization*, 2005, 83, 8 : 638-39.

<sup>15</sup> Braa J, Hanseth O, Heywood A. *et al.* Developing Health Information Systems in Developing Countries : The Flexible Standards Strategy. *Management Information Systems Quarterly*, 2007, 31, Suppl 1 : 381-402.

d'une donnée comportait de multiples occasions d'erreur et peu d'opportunités de suivi et de correction.

Ce type de travail a été fondamentalement modifié par l'utilisation d'outils mobiles de collecte de données. Les données directement saisies sur tablettes sont rapidement transmises par les équipes de collecte, à travers les réseaux 3G. Le faible délai entre la collecte et la transmission des données vers des serveurs en assurant la sécurité limite les risques de perte et la richesse des métadonnées disponibles permet par ailleurs un suivi précis des campagnes de collecte de données et la traçabilité des résultats. En outre, la simplicité et le faible coût de ces outils, diminuent le coût technique et humain de la mise en place d'enquêtes. On assiste donc au développement de sources de données multiples, facilement accessibles et transférables entre acteurs.

## **Making Data Global : nouvelles pratiques de modélisation de la santé globale**

### *L'évolution des conditions d'utilisation des données*

L'accès aux données de la santé globale a été modifié par l'apparition de nouvelles pratiques dans la recherche en épidémiologie. Les données brutes des enquêtes démographiques et de santé<sup>16</sup> (EDS) financées par la *United States Agency for International Development* (USAID) sont cataloguées et mises à disposition des chercheurs en demandant l'accès. De même, l'IHME fournit un accès à son catalogue de données, le *Global Health Data Exchange* qui répertorie des milliers de sources et donne des liens vers ces données ou les contacts pour les obtenir<sup>17</sup>. Les chercheurs en santé des populations ont donc aujourd'hui, sans sortir de leur bureau, accès à une richesse de données sans précédent.

Toutefois, la disponibilité de ces données ne signifie pas qu'elles soient directement utilisables. Un travail est nécessaire pour leur compréhension et leur mise en cohérence. Un exemple : l'alignement des données de différentes sources pour un même système de santé est un problème récurrent et le développement de cartes sanitaires de référence par la consolidation de différentes sources de données disponibles est un travail à part entière. Même après la réalisation d'une telle carte en RDC<sup>18</sup>, il est en permanence nécessaire de refaire les correspondances entre ces données de référence et les nomenclatures utilisées par les différents acteurs du système de santé congolais.

Les compétences et les outils mis en œuvre pour la mesure de la santé mondiale évoluent donc. S'il est plus facile d'avoir accès à des données brutes, l'apparition de besoins nouveaux pour le catalogage et la standardisation renouvelle les problématiques.

### *L'entrée tardive de la santé globale à l'âge computationnel*

Devant la diversification des données utilisables, il est nécessaire de dégager la cohérence de résultats issus de différentes sources. Cette problématique est un genre identifié de la littérature épidémiologique. La méta-analyse consiste, par exemple, à comparer les résultats de certains traitements mesurés dans des études cliniques pour offrir une estimation robuste des effets de ces traitements. Dans le cas de données descriptives à l'échelle mondiale, il faut réestimer la valeur et la

<sup>16</sup> Les EDS sont de larges enquêtes sur les caractéristiques démographiques et l'état de santé des populations. Dans les pays dont les états civils sont défaillants, ce sont souvent les seules sources d'informations représentatives sur les populations. Le programme EDS est financé par USAID, et mis en œuvre par le cabinet américain ICF Macro. Voir le site des EDS, [<https://www.dhsprogram.com/>] (consulté le 4 novembre 2019).

<sup>17</sup> Site du Global Health Data Exchange, [<http://ghdx.healthdata.org>] (consulté le 5 juillet 2019).

<sup>18</sup> Site Carte sanitaire de la République démocratique du Congo, [<http://cartesanitairercd.org>] (consulté le 5 juillet 2019).

variation dans le temps et dans l'espace de paramètres multiples et corrélés. La dimensionnalité du problème nécessite la mise en œuvre de modèles de méta-régression tels que le modèle DisMOD<sup>19</sup>. La complexité de ces modèles fait qu'ils nécessitent des ressources de calcul importantes et font entrer la santé globale dans un âge computationnel.

La santé globale est entrée dans cet âge computationnel relativement plus tardivement que d'autres domaines. Les disciplines fournissant la grammaire quantitative de la santé publique que sont l'épidémiologie et la biostatistique sont plus préoccupées par les questions d'inférence causale et d'attribution des risques que par celles de grands modèles d'équilibres généraux. Si la climatologie avait commencé à utiliser ces modèles d'équilibre généraux dès les années 1940, une telle approche dans le domaine de la santé des populations date du début des années 2000, avec les premiers travaux sur le GBD. De même, si la simulation des agents ou les méthodes bayésiennes<sup>20</sup> ont été très tôt adoptées par des chercheurs en sciences humaines, elles l'ont été plus tardivement dans des domaines comme la démographie<sup>21</sup>.

L'apparition et la diffusion des méthodes de computation dans la mesure de la santé globale sont concomitantes à l'émergence de besoins d'éléments de décision à une échelle globale, pour des acteurs en quête de légitimité symbolique. Que ce soit le GFATM promouvant un modèle de financement *basé sur les résultats* ou la BMGF poursuivant des investissements informés par des preuves scientifiques, l'utilisation d'éléments chiffrés pour déterminer les priorités de financement d'actions sanitaires sert aussi à justifier la place de ces organisations en démontrant leur objectivité et la scientificité de leur approche.

### *Nouvelles modalités dans la gouvernance de la santé globale*

L'évolution de la gouvernance de la santé globale joue donc un rôle clé dans l'innovation des méthodes et des techniques. L'apparition de la redevabilité épidémiologique appelle en effet un plus grand recours à l'utilisation de modèles censés permettre l'analyse théorique de problèmes complexes. En 2009, Reuben Granich et collègues publient le résultat de « modèles mathématiques » pour démontrer l'efficacité de la stratégie *Test and Treat*<sup>22</sup>. En 2014, au-delà d'un bilan humain lourd, les projections épidémiologiques alarmistes effectuées par le *Center for Disease Control and Prevention (CDC)* ont contribué à mobiliser la communauté internationale autour de l'épidémie d'Ébola en Afrique de l'Ouest<sup>23</sup>. Finalement, des outils tels que *The Lives Saved Tool*<sup>24</sup>, proposant de quantifier l'impact d'interventions sanitaires en termes de nombre de vies sauvées sont courants dans les pratiques d'évaluation et de communication des acteurs de la santé globale.

<sup>19</sup> Flaxman A, Vos T, Murray C. *An Integrative Metaregression Framework for Descriptive Epidemiology*. Seattle : University of Washington Press, 2015.

<sup>20</sup> La simulation d'agent est une classe de modèles utilisant la simulation d'individus et de leurs interactions pour analyser des phénomènes complexes. Les méthodes bayésiennes sont des modèles statistiques reposant sur une représentation des quantités analysées sous formes de distributions de probabilité, permettant une description complexe de l'incertitude des résultats.

<sup>21</sup> Bryant J, Graham P. Bayesian Demographic Accounts: Subnational Population Estimation Using Multiple Data Sources. *Bayesian Analysis*, 2013, 8, 3 : 591-622 ; Billari F, Ongaro F, Prskawetz A. Agent-Based Computational Demography. In Billari F, Prskawetz A ed. *Agent-Based Computational Demography*. Heidelberg : Physica-Verlag HD, 2003 : 1-17.

<sup>22</sup> Stratégie visant à l'élimination du VIH par un dépistage systématique des personnes à risque et une mise sous traitement dès le diagnostic positif. Granich R, Gilks C, Dye C. *et al.* Universal Voluntary HIV Testing with Immediate Antiretroviral Therapy as a Strategy for Elimination of HIV Transmission : A Mathematical Model. *The Lancet*, 2009, 373, 9657 : 48-57.

<sup>23</sup> Meltzer M, Atkins C, Santibanez S *et al.* Estimating the Future Number of Cases in the Ebola Epidemic – Liberia and Sierra Leone, 2014–2015. *Morbidity and Mortality Weekly Report*, 2014, 63, 3 : S1-14.

<sup>24</sup> Walker N, Tam Y, Friberg IK. Overview of the Lives Saved Tool (LiST). *BMC Public Health*, 2013, 13 Suppl 3 : S1.

Il faut souligner que la valeur de ces approches est souvent essentiellement performative. Leurs résultats nourrissent des discours se voulant rigoureux pour défendre la nécessité ou le succès d'un investissement. La valeur et le crédit apportés aux différents exercices et outils cités plus haut sont cependant rarement justifiés par une validation rigoureuse. Les projections faites par le CDC d'Atlanta sur l'épidémie Ebola de 2014 ont en effet surestimé l'ordre de grandeur réel de l'épidémie<sup>25</sup>. Des critiques récurrentes des résultats du GBD portent sur l'extrapolation dans des zones où très peu de résultats directement mesurés sont disponibles, et où la validation ou l'infirmerie des résultats est quasiment impossible.

L'augmentation de la technicité de la production de connaissance mène donc à l'apparition de ce que Bruno Latour appelle des « centres de calcul »<sup>26</sup>. Ces centres concentrent d'énormes moyens et sont le lieu de production de la plupart des données de la santé globale avant qu'elles deviennent une connaissance légitime. On a évoqué l'IHME pour le GBD ou ICF Macro pour les EDS, on peut aussi parler de *WorldPop*, le consortium financé par la BMGF et basé à l'université de Southampton qui modélise la répartition spatiale des populations à des échelles fines<sup>27</sup>. La complexité du travail effectué dans ces centres, plus que l'exactitude vérifiée de leurs résultats, en fait des lieux de pouvoir, seuls à même de produire la connaissance qui guidera les décisions politiques et financières.

## Le nouveau paradigme de mesure de la santé globale

En décrivant les évolutions de l'activité de mesure dans la santé globale, on voit donc deux évolutions concomitantes. D'une part, l'arrivée de nouveaux acteurs dans la gouvernance sanitaire mondiale a fait émerger une nouvelle rationalité quantitative, privilégiant le suivi d'indicateurs spécifiques sur les résultats de programmes de santé particuliers. Pour mesurer ces indicateurs, on a privilégié la création de sous-systèmes spécialisés de collecte de données, répondant aux besoins de bailleurs de fonds ou d'intervenants techniques, et au détriment de la cohérence des systèmes d'information sanitaire nationaux. Cette nouvelle rationalité quantitative s'est donc faite au prix d'une fragmentation et d'une privatisation des sources de données et d'une complexification des outils quantitatifs. Dans le même temps, la facilitation de la collecte des données et le développement de méthodes de calcul complexes ont redéfini la façon de penser la santé globale en concentrant la production de connaissance dans quelques grands centres de calcul capables de modéliser la multitude de sources de données aujourd'hui disponibles.

En reprenant la dualité administrative et scientifique de l'activité statistique, on voit une redistribution des pouvoirs entre le savant et l'administrateur. Les progrès dans les méthodes de collecte et de partage de données ont abouti à une déqualification et dé-problématisation des activités essentielles de nomenclature et de définition. Il ne s'agit plus de définir les grandes catégories qui permettent de nommer et de mesurer le réel de la santé globale<sup>28</sup>, mais plutôt de les reconstruire *a posteriori*, à partir d'éléments épars, à l'aide d'infrastructures de calcul imposantes. Institutionnellement, cette évolution est symbolisée par un partage et un déplacement de l'activité de mesure de la santé globale de l'OMS vers de grandes institutions scientifiques recevant des financements privés. Alain Desrosières notait par ailleurs que la faible institutionnalisation des

<sup>25</sup> Chowell G, Viboud C, Simonsen *et al.* Perspectives on model forecasts of the 2014-2015 Ebola epidemic in West Africa : lessons and the way forward. *BMC Medicine*, 2017, 15 : 42.

<sup>26</sup> Latour B. *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, Mass : Harvard University Press, 1987.

<sup>27</sup> Site de WorldPop [<https://www.worldpop.org/>] (consulté le 1<sup>er</sup> novembre 2019).

<sup>28</sup> Bowker G, Leigh Star S. *Sorting Things out : Classification and Its Consequences*, Cambridge, Mass : MIT Press, 1999.

systèmes statistiques des pays en développement était une l'opportunité pour des innovations méthodologiques<sup>29</sup>. Il faut souhaiter que le nouveau paradigme de la mesure en santé globale puisse participer à cette créativité méthodologique, plus qu'au renforcement de structures de pouvoir scientifique et politique dont la légitimité doit aujourd'hui être questionnée.

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

---

<sup>29</sup> Desrosières A, La mesure du développement : un domaine propice à l'innovation méthodologique. *Revue Tiers Monde*, 2013, 213 : 23.

## Quelques considérations personnelles

Sur l'évolution de l'épidémiologie ces dernières décennies en France

**Marcel GOLDBERG**

Professeur émérite d'épidémiologie et de santé publique, Université Paris Descartes

### Résumé

Marcel Goldberg, Professeur émérite d'épidémiologie et de santé publique, Université Paris Descartes, est docteur en médecine, en biologie humaine et en mathématique appliquée. Il a dirigé de 1983 à 2003 une unité de recherche en épidémiologie de l'Inserm. Il a aussi présidé la Commission scientifique spécialisée en épidémiologie de l'Inserm et l'Association des épidémiologistes de langue française. Il est actuellement membre de l'Unité mixte Cohortes épidémiologiques en population, UMS 011. Les différentes fonctions qu'il a occupées lui ont permis d'avoir une vision « longitudinale » de l'épidémiologie et de son évolution.

**Mots-clés** : épidémiologie, cohorte, données massives.

### Abstract

#### ***Some personal considerations on the evolution of epidemiology in recent decades in France***

*Marcel Goldberg, Emeritus Professor of Epidemiology and Public Health, Paris Descartes University, holds a doctorate in medicine, human biology and applied mathematics. From 1983 to 2003, he led an epidemiology research unit at Inserm. He also chaired Inserm's Epidemiology Scientific Commission and the Association of French Epidemiologists. He is currently a member of the Population-based Epidemiological Cohorts Unit, UMS 011. By the various positions he held, he has earned a "longitudinal" vision of epidemiology and its evolution.*

**Keywords** : *epidemiology, cohort, big data.*

Beaucoup a été dit concernant divers aspects de l'épidémiologie d'hier et d'aujourd'hui durant ce colloque consacré à *l'Histoire de l'épidémiologie - Enjeux passés, présents et futurs*. Je voudrais simplement y apporter un éclairage complémentaire. Le texte qui suit n'a aucune prétention historique, ni épistémologique, ni scientifique d'une façon générale. Il s'agit simplement du témoignage d'un observateur et acteur de la scène épidémiologique dans notre pays, qui a exercé la discipline pendant plus de 40 ans dans différents contextes institutionnels, académiques ou pas : Inserm, université, agence de santé publique (Institut de veille sanitaire), services médicaux d'une grande entreprise publique. J'ai également été largement impliqué dans l'animation du milieu de l'épidémiologie, en ayant siégé pendant 12 ans à la Commission scientifique spécialisée de l'Inserm (dont un mandat comme président) ; j'ai eu aussi l'honneur de présider pendant plusieurs années l'Association des épidémiologistes de langue française. Ces fonctions m'ont permis d'avoir une vision « longitudinale », pour employer un terme épidémiologique, du milieu et de son évolution pendant plusieurs décennies.

### **Des méthodes de plus en plus sophistiquées, des données de plus en plus diversifiées et nombreuses**

Sur le plan méthodologique et technique, on a assisté à diverses évolutions de l'épidémiologie pendant les décennies écoulées. Grâce à l'énorme développement des capacités de calcul et de stockage des outils informatiques, deux phénomènes se sont développés en parallèle : d'une part l'accroissement quasi vertigineux de la quantité des données de différentes natures disponibles et provenant de sources nouvelles, d'autre part la sophistication des méthodes d'analyse.

Les données mobilisées par les études épidémiologiques aujourd'hui proviennent en grande partie de sources qui n'étaient pas disponibles voici quelques décennies comme les données génétiques de séquençage, de nouvelles méthodes d'imagerie ou d'images de satellites et, plus récemment encore, les données tirées des objets connectés ou des réseaux sociaux. Les données médico-administratives issues de l'activité hospitalière ou liées au remboursement des soins sont utilisées depuis longtemps, mais elles ont pris une ampleur gigantesque avec la mise en place d'entrepôts de données hospitalières et du Système national des données de santé qui couvrent la totalité de la population vivant en France. La taille des jeux de données utilisables par les épidémiologistes peut parfois se compter en téra- voire en pétaoctets et ils constituent ce qu'on appelle aujourd'hui les *Big Data*.

En parallèle, grâce à l'augmentation des capacités de calcul des ordinateurs, et du fait du volume des données à traiter, des méthodes statistiques de plus en plus sophistiquées ont été développées. Les années récentes ont vu le retour en force de méthodes d'intelligence artificielle particulièrement performantes.

Il est clair que la disponibilité de masses gigantesques de données et de méthodes permettant de les analyser va considérablement élargir les possibilités d'investigation, modifier les conditions d'exercice de l'épidémiologie et faire apparaître de nouveaux métiers au sein de la discipline. Mais,

ces possibilités nouvelles ne sont pas sans danger... Comme le rappelait Rodolfo Saracci dans un article récent<sup>1</sup>, la validité des données doit toujours prévaloir sur leur nombre et « big data » n'est pas synonyme de « good data ». Bien souvent, on ne sait rien (ou trop peu) sur les conditions du recueil des données provenant de ces sources, sur les personnes auprès desquelles elles ont été collectées ou sur la population dont elles sont issues. Et leur quantité, si elle peut compenser partiellement certaines imprécisions des mesures, ne pourra évidemment pas corriger les nombreux biais de toute nature susceptibles d'être rencontrés dans un contexte le plus souvent non contrôlé. Bien au contraire, des résultats issus de l'analyse de masses énormes de données peuvent donner à tort l'illusion de résultats solides, ceci d'autant plus qu'évidemment toute différence, si minime soit-elle, sera statistiquement significative.

Quel que soit le caractère innovant des méthodes d'analyse de données massives et leur apport indéniable dans certains domaines, elles ne remplaceront jamais ni la recherche sur les mécanismes biologiques, ni la recherche clinique, ni l'expertise épidémiologique. À cet égard, les résultats issus d'algorithmes complexes, dont les concepteurs eux-mêmes ne sont pas capables d'expliquer le fonctionnement, peuvent générer des résultats trompeurs s'ils ne sont pas « encadrés » par les fondamentaux d'une démarche épidémiologique rigoureuse.

## Les cohortes en population

Alors que, dans le monde anglo-saxon, les années qui ont suivi la guerre ont vu se mettre en place des cohortes en population qui ont largement contribué au développement de l'épidémiologie « moderne », comme la cohorte Framingham<sup>2</sup>, ou la *British Doctors Study*<sup>3</sup>, en France, il a fallu attendre 1967 pour voir l'Étude prospective parisienne, première cohorte en population<sup>4</sup>. Ce travail pionnier est resté pendant longtemps isolé. La mise en place d'importantes cohortes en population comme outil épidémiologique a émergé voici environ trois décennies. Au tournant des années 1980-1990, plusieurs cohortes importantes ont été mises en place<sup>5</sup>, suivies par d'autres cohortes de personnes âgées dans les années suivantes et celles des Investissements d'avenir, après 2010. Je voudrais insister sur certains aspects de l'évolution des cohortes en France durant les décennies récentes.

Elles se caractérisent tout d'abord par leur taille de plus en plus grande. Alors que l'Étude prospective parisienne comptait moins de 8 000 participants, la cohorte E3N avait, dès 1990, recruté 100 000 femmes. Gazel qui, avec 20 000 participants lors de sa mise en place voici exactement trente ans, semblait une cohorte de grande taille, est suivie aujourd'hui de Constances, dix fois plus importante.

Elles ont également un caractère de plus en plus « généraliste ». Les premières générations de cohortes étaient centrées sur un domaine spécifique de la santé, comme la pathologie

<sup>1</sup> Saracci R. Epidemiology in Wonderland : Big Data and precision medicine. *Eur J Epidemiol*, 2018, 33 (3) : 245-257.

<sup>2</sup> Oppenheimer GM. Becoming the Framingham Study. *Am J Pub Health*, 2005, 95 :602-10.

<sup>3</sup> Doll R, Peto R, Wheatley K, Gray R, Sutherland I. Mortality in relation to smoking: 40 years' observations on male British doctors, *Br Med J*, 1994, 309 : 901-11.

<sup>4</sup> Richard JL, Ducimetière P, Bonnaud G, Claude JR *et al.* Incidence et évaluation du risque des maladies coronaires. L'étude Prospective Parisienne. *Arch. Mal. Cœur Vaiss*, 1977, 70 : 531-540.

<sup>5</sup> Dartigues JF, Gagnon M, Michel P *et al.* Le programme de recherche Paquid sur l'épidémiologie de la démence. Méthodes et résultats initiaux. *Rev Neurol*, 1991, 147 : 225-30 ; Goldberg M, Leclerc A, Bonenfant S, Chastang JF *et al.* Cohort profile: the GAZEL Cohort Study. *Int J Epid*, 2007, 36 : 32-39 ; Clavel-Chapelon F, Van Liere MJ, Giubout C, Niravong MY *et al.* E3N, a French cohort study on cancer risk factors. E3N Group. Étude épidémiologique auprès de femmes de l'Éducation Nationale. *Eur J Cancer Prev*, 1997, 6 (5) : 473-8.

cardiovasculaire, les cancers ou les démences. Les objectifs scientifiques se sont considérablement élargis à des pathologies diverses et certaines cohortes prenant en compte les possibilités d'accumuler de nombreux cas de maladies diverses offertes par leurs importants effectifs, annoncent explicitement ne pas avoir d'orientation spécifique en termes de pathologies étudiées.

Ces premières générations de cohorte avaient été conçues comme des outils pour la poursuite des objectifs scientifiques propres des équipes qui les mettaient en place dans un domaine pathologique spécifique, ce qui était cohérent avec la spécialisation de celles-ci. Leur utilisation était donc essentiellement le fait de ces équipes. Avec l'augmentation de la taille des cohortes, réunissant des effectifs permettant l'étude de pathologies diverses, et l'augmentation parallèle de leur coût, est apparue la nécessité d'ouvrir l'accès de leurs données à la communauté scientifique, leur donnant un rôle d'infrastructure de recherche dépassant les intérêts scientifiques propres des équipes qui les conduisent.

Ces tendances se retrouvent à l'échelle internationale avec une obligation de partage des données imposée par la plupart des organismes qui financent des cohortes. On assiste également à la mise en place de « mega-cohortes » de plusieurs centaines de milliers de participants, dépassant le million pour les plus importantes. Mais, même ces cohortes géantes ne permettent pas toujours de réunir des effectifs suffisants pour l'étude de phénomènes rares. Cela a suscité, particulièrement dans la dernière décennie, la constitution de consortiums de cohortes, destinés à permettre la mise en commun de données. Les consortiums les plus récents associent plusieurs dizaines de cohortes dont les effectifs totaux se comptent en millions de sujets. Il devient de plus en plus fréquent de lire des publications portant sur le suivi individuel de plus d'un million de personnes.

## **De nouveaux domaines d'investigation**

Un autre aspect qui a été peu abordé lors du colloque est celui du renforcement, voire de l'émergence de nouveaux domaines scientifiques dans le champ de l'épidémiologie. Voici quelques décennies, l'essentiel de la recherche épidémiologique était une « épidémiologie d'organe » centrée sur des pathologies spécifiques. Progressivement, on a vu un important essor de thèmes comme l'étude des effets de l'environnement sur divers aspects de la santé. Certains domaines, pourtant importants, voire fondateurs, dans l'histoire de l'épidémiologie, comme l'étude des risques professionnels ou celle des déterminants sociaux de la santé, étaient un peu tombés en désuétude dans le paysage de la recherche épidémiologique française. Dans un contexte national et international de reconnaissance de plus en plus aiguë des inégalités sociales et territoriales de santé, ces thèmes ont connu une forte progression depuis les années 1990, et on a vu apparaître de nouvelles équipes très actives dans ces domaines.

## **La place grandissante de l'épidémiologie dans l'élaboration et la conduite des politiques de santé, et l'apparition de nouveaux acteurs**

Voici trois ou quatre décennies, l'épidémiologie était le domaine quasi exclusif du monde académique, et tout particulièrement de l'Inserm qui exerçait pratiquement un monopole sur la discipline. Et même si l'essentiel des travaux de ses équipes était consacré à la recherche étiologique, c'était aussi l'Inserm qui réalisait les études à visée de santé publique.

Progressivement les pouvoirs publics ont été de plus en plus demandeurs d'informations de nature épidémiologique pour l'élaboration et la conduite des politiques de santé, et de nouveaux organismes sont apparus. Le mouvement a été entamé avec les Observatoires régionaux de la santé, puis des agences de santé ont été créées dans divers domaines : surveillance sanitaire, médicaments, environnement... Des équipes d'épidémiologie se sont constituées dans certaines de ces agences, de

même qu'à la Caisse nationale d'assurance maladie ; d'autres, sans réunir des équipes en interne, sont devenues des financeurs importants de la recherche épidémiologique. Aujourd'hui, l'épidémiologie constitue une des disciplines majeures sur laquelle s'appuient ces agences de santé, qui sont plus proches des décideurs de santé publique que les chercheurs du monde académique, ce qui permet à l'épidémiologie d'intervenir plus directement dans les décisions de santé publique.

### **Vers la reconnaissance de l'épidémiologie comme discipline scientifique à part entière**

Pour conclure, je voudrais rappeler que pendant longtemps l'épidémiologie a été considérée dans les milieux de la recherche biomédicale comme une discipline mineure, voire non véritablement scientifique, tout juste bonne à donner des chiffres et des corrélations statistiques. Je me réjouis de constater que son statut de science à part entière commence à être reconnu, comme le montre le fait que notre discipline a aujourd'hui deux représentants à l'Académie des sciences et un au Collège de France. Et je suis persuadé que ce n'est qu'un début...

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



**Le Comité pour l'histoire de l'Inserm**

Bureau 1446

Inserm, 101, rue de Tolbiac, 75654, Paris Cedex 13

Plus d'information sur [www.inserm.fr](http://www.inserm.fr)

Contactez-nous en écrivant au secrétariat scientifique : [celine.paillette@ext.inserm.fr](mailto:celine.paillette@ext.inserm.fr)

