

## 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.