

2006

ACADÉMIE DES SCIENCES

L'ÉPIDÉMIOLOGIE HUMAINE

CONDITIONS DE SON
DÉVELOPPEMENT EN FRANCE,
ET RÔLE DES MATHÉMATIQUES

Sous la direction de
ALAIN-JACQUES VALLERON



INSTITUT DE FRANCE
Académie des sciences

L'épidémiologie humaine

Conditions de son développement en France, et rôle des mathématiques

RAPPORT SUR LA SCIENCE ET LA TECHNOLOGIE N° 23

Animateur : Alain-Jacques Valleron

ACADÉMIE DES SCIENCES



17, avenue du Hoggar
Parc d'activités de Courtabœuf, BP 112
91944 Les Ulis Cedex A, France

Rapports sur la science et la technologie

- *Sciences et pays en développement. Afrique subsaharienne francophone*
RST n° 21, 2006.
- *La recherche spatiale française*
RST n° 22, 2006.

© 2006, EDP Sciences, 17, avenue du Hoggar, BP 112, Parc d'activités de Courtabœuf,
91944 Les Ulis Cedex A

Tous droits de traduction, d'adaptation et de reproduction par tous procédés réservés pour tous pays. Toute reproduction ou représentation intégrale ou partielle, par quelque procédé que ce soit, des pages publiées dans le présent ouvrage, faite sans l'autorisation de l'éditeur est illicite et constitue une contrefaçon. Seules sont autorisées, d'une part, les reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective, et d'autre part, les courtes citations justifiées par le caractère scientifique ou d'information de l'œuvre dans laquelle elles sont incorporées (art. L. 122-4, L. 122-5 et L. 335-2 du Code de la propriété intellectuelle). Des photocopies payantes peuvent être réalisées avec l'accord de l'éditeur. S'adresser au : Centre français d'exploitation du droit de copie, 3, rue Hautefeuille, 75006 Paris. Tél. : 01 43 26 95 35.

ISBN 2-86883-796-4

Académie des sciences

Rapport Science et Technologie

Le Comité interministériel du 15 juillet 1998, à l'initiative du ministre de l'Éducation nationale, de la Recherche et de la Technologie, a confié à l'Académie des sciences l'établissement du rapport biennal sur l'état de la science et de la technologie.

Pour répondre à cette demande, l'Académie des sciences a mis en place en son sein le Comité « *Rapport Science et Technologie* » (RST), chargé de choisir les sujets d'étude et de suivre les travaux.

Chaque thème retenu est conduit par un groupe de travail animé par un membre ou un correspondant de l'Académie, entouré d'experts.

Chaque rapport est soumis au Comité RST, à un Groupe de lecture critique, et à l'Académie des sciences.

Entre 1999 et 2006, vingt-deux rapports ont ainsi été édités et remis au ministre délégué à la Recherche.

COMPOSITION DU COMITÉ RST

Alain ASPECT

Membre de l'Académie des sciences — Directeur de recherche au Centre national de la recherche scientifique, professeur à l'École polytechnique

Jean-François BACH

Secrétaire perpétuel de l'Académie des sciences — Professeur à l'université René-Descartes

Jean-Michel BONY

Membre de l'Académie des sciences — Professeur à l'École polytechnique

Christian BORDÉ

Correspondant de l'Académie des sciences — Directeur de recherche au Centre national de la recherche scientifique

Édouard BRÉZIN

Président de l'Académie des sciences — Professeur à l'université Pierre-et-Marie-Curie et à l'École polytechnique

Marie-Lise CHANIN

Correspondant de l'Académie des sciences — Directeur de recherche émérite au Centre national de la recherche scientifique

Geneviève COMTE-BELLOT

Correspondant de l'Académie des sciences — Professeur émérite de l'École centrale de Lyon

François CUZIN

Membre de l'Académie des sciences — Professeur à l'université de Nice-Sophia-Antipolis

Jean DERCOURT

Secrétaire perpétuel de l'Académie des sciences — Professeur émérite à l'université Pierre-et-Marie-Curie

Christian DUMAS

Membre de l'Académie des sciences — Professeur à l'École normale supérieure de Lyon

Michel FARDEAU

Correspondant de l'Académie des sciences — Professeur au Conservatoire national des arts et métiers, directeur médical et scientifique à l'Institut de myologie (Hôpital de la Pitié Salpêtrière)

Jules HOFFMANN

Vice-président de l'Académie des sciences — Directeur de l'Institut de biologie moléculaire et cellulaire de Strasbourg

Jean-Pierre KAHANE

Membre de l'Académie des sciences — Professeur émérite à l'université Paris-Sud Orsay

Daniel KAPLAN

Membre de l'Académie des sciences — Directeur de la société Fastlite

Henri KORN

Membre de l'Académie des sciences — Professeur honoraire à l'institut Pasteur et directeur de recherche honoraire à l'Institut national de la santé et de la recherche médicale

Nicole LE DOUARIN

Secrétaire perpétuelle de l'Académie des sciences — Professeur honoraire au Collège de France

Jean-Louis LE MOUËL

Membre de l'Académie des sciences — Physicien à l'Institut de physique du globe de Paris

François MATHEY

Membre de l'Académie des sciences — Directeur de recherche au Centre national de la recherche scientifique, professeur à l'École polytechnique

René MOREAU

Membre de l'Académie des sciences — Professeur à l'Institut national polytechnique de Grenoble

Olivier PIRONNEAU

Membre de l'Académie des sciences — Professeur à l'université Pierre-et-Marie-Curie

Jean-Pierre SAUVAGE

Membre de l'Académie des sciences — Directeur de recherche au Centre national de la recherche scientifique

Moshe YANIV

Membre de l'Académie des sciences — Professeur à l'Institut Pasteur et directeur de recherche au Centre national de la recherche scientifique

Coordination éditoriale :

Jean-Yves CHAPRON

Directeur du service des Publications de l'Académie des sciences, assisté de **Joëlle FANON**

AVANT-PROPOS

JEAN DER COURT

Secrétaire perpétuel de l'Académie des sciences

Le Comité interministériel de la recherche du 15 juillet 1998 a confié à l'Académie des sciences le soin d'établir un bilan biennal sur l'état des sciences et des techniques de la recherche en France. L'examen de l'état actuel de l'épidémiologie humaine en France, des conditions de son développement et du rôle renforcé que peuvent y jouer les mathématiques est présenté ici.

L'État consacre chaque année plus de 10 % du produit intérieur brut à la Santé. C'est dire l'importance que les citoyens, la représentation nationale et le gouvernement y attachent. Et pourtant, l'épidémiologie, l'une des disciplines fondant la politique de santé, n'est pas assez développée en France, en dépit des travaux de grande qualité mais peu nombreux que lui consacre un petit nombre de spécialistes au sein de structures émietées et peu coordonnées. Cet état de fait est analysé ici ; le bilan est dressé et des propositions de mesures susceptibles d'y remédier sont présentées.

Le rapport est constitué de quatre parties :

- une suite de 23 monographies regroupées en 9 chapitres présentant une analyse documentée établie par des spécialistes, adoptée collectivement par le groupe de travail, et terminée par une série de recommandations ;
- **une synthèse accompagnée de 14 orientations prioritaires constituant le cœur du rapport**, présentée dans un chapitre final établi par Alain-Jacques Valleron, animateur du groupe de travail, qui a été adoptée par l'Académie ;
- une série de remarques rédigées par 10 des acteurs majeurs de la discipline ou de son environnement, à l'issue de la réunion de « lecture critique » ;
- une présentation, à l'Académie, en séance plénière, par Jean-François Bach, Secrétaire perpétuel, accompagnée des remarques faites en séance.

L'épidémiologie est au cœur des sciences fondamentales et appliquées qui traitent de la santé publique. D'emblée, elle a envisagé les données médicales par des méthodes mathématiques, en particulier les statistiques. En 1760,

Bernoulli présente à l'Académie Royale des Sciences un mémoire concluant – grâce à une modélisation – au bon rapport bénéfice-risque comme méthode de prévention de la scarification de malades atteints de la variole. Puis, en France, P.C. Louis, en 1837, présente et développe la « *médecine numérique* » ; R. Villermet (1840) expose les conditions de travail dans les fabriques et les confronte statistiquement à la détérioration de la santé. Ce travail aboutit à une loi limitant le travail des enfants (1850). Les trois composantes de l'épidémiologie sont posées : travail statistique, connaissance médicale et importance sociétale.

Pendant près d'un siècle, l'épidémiologie ne se développe guère en France. L'essentiel de l'effort des scientifiques et des médecins suit la voie immense ouverte par Pasteur. L'épidémiologie ne se répand, en France, que vers 1960, mais dans le seul domaine médical pour l'enseignement et pour la recherche. L'épidémiologie s'est imposée par la découverte de facteurs de risque de plusieurs maladies très répandues (notamment le tabac pour le cancer des bronches et les maladies cardiovasculaires. . .), par son rôle dans l'identification et la description de maladies émergentes (sida, hépatite C, . . .), par la nécessité d'établir les liens entre pratiques industrielles et maladies (amiante, . . .), enfin, par l'évaluation de risques éventuels environnementaux (usines électronucléaires, OGM, téléphonie mobile, . . .).

La population des acteurs de la recherche et de l'enseignement en épidémiologie est difficile à cerner en France, comme l'est celle de bien d'autres domaines. La recherche épidémiologique se déroule dans des agences spécialisées récentes qui ne dépendent pas du ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche, dans deux établissements publics de recherche, agences de moyens et de programmes (Inserm, surtout, et IRD) associés à de nombreuses universités autonomes où œuvrent des enseignants chercheurs, des ingénieurs et des techniciens. Les équipes de recherche et d'enseignement sont intégrées aux seules facultés de médecine ; elles ne sont présentes, en tant qu'unités d'enseignement et de recherche, dans aucun des secteurs mathématiques, biologiques ou sciences humaines. En dépit de ces difficultés, le rapport tente une détermination numérique des acteurs et de leur production scientifique à partir du nombre des publications. Alors que la part de l'Europe est très voisine de celle des États-Unis, on constate que le Royaume-Uni se classe en première place, que l'Allemagne ne figure pas et que la France est en troisième position en Europe. Au fil des chapitres, il est établi que le petit nombre des équipes françaises occupe, dans les disciplines où elles travaillent, une position de qualité.

Dans le rapport, les différents types d'approche de l'épidémiologie sont exposés. L'épidémiologie descriptive cherche à établir des liens entre l'état de santé,

ou les caractéristiques cliniques et biologiques de la population, et leurs conditions de vie (géographiques, sociales, familiales, comportementales, ...). Cette méthode, très importante dans le monde, est délaissée en France, et le mode d'évaluation des carrières des chercheurs et enseignants chercheurs reconnaît mal ce type de recherche. La généralisation et l'accroissement de l'efficacité des concepts et des techniques informatiques ont considérablement accru les champs du possible en épidémiologie. Très utilisées pour appréhender la gestion de la santé publique, les informations regroupées en bases de données de grande taille ont vocation à être aussi des bases de données épidémiologiques. Cette utilisation, très développée en Europe, commence à l'être timidement en France. Le rapport discute ce verrou majeur et propose des voies d'amélioration.

L'approche de l'épidémiologie analytique est mieux appréciée, même si elle est plus coûteuse et difficile à mettre en œuvre. Elle cherche le lien entre une maladie donnée et des facteurs supposés intervenir. Elle exige souvent la constitution de cohortes suivies pendant des décennies.

L'épidémiologie expérimentale est utilisée pour l'évaluation statistique de médicaments ou de pratiques médicales. Ces méthodes, en raison de leur coût et de leur durée, mobilisent l'industrie pharmaceutique et des sites de recherche dédiés (notamment au VIH et aux hépatites, encore en petit nombre).

L'épidémiologie théorique, enfin, est en général utilisée dans des situations où le nombre de données est très faible et où la seule ressource est d'estimer, souvent grâce à des méthodes probabilistes, les conséquences prévisibles des rares connaissances disponibles. Il s'agit alors de travaux prédictifs d'un risque et susceptibles d'éclairer la décision relative à la nature des mesures préventives.

L'ouvrage expose ces différents aspects et préconise de lever les verrous qui, en France, freinent la recherche. Des propositions précises sont exposées.

Le rapport explicite les liens entre les trois composantes fondatrices de l'épidémiologie et montre leur apport mutuel.

1. Les sciences mathématiques, statistiques, probabilités.
2. La biologie humaine et les sciences médicales à partir de quatre exemples (maladies infectieuses, cardiovasculaires, vieillissement, cancers).
3. Les sciences humaines et sociales.

L'épidémiologie est en train de créer une révolution de la pratique médicale dans le monde. Sous la forme d'épidémiologie clinique, elle met en jeu les informations épidémiologiques recueillies auprès de groupe de malades sur la tolérance

et l'efficacité des traitements, et permet au praticien de prendre une décision thérapeutique individuelle face au malade qu'il a à traiter. Cela conduit à connaître la tolérance et l'efficacité des traitements, fondement d'une *médecine fondée sur les preuves*. Ce concept a apporté une révolution de la pratique médicale. Il en a résulté l'impérieuse nécessité de diffuser ces travaux de recherche de manière synthétique et critique afin d'informer les médecins praticiens. En particulier, tous les essais cliniques doivent être pris en compte, même ceux qui n'ont pas abouti à la mise sur le marché d'un médicament. La médecine praticienne, en France, s'est longtemps désintéressée de cette façon de faire nécessitant, d'une part, l'établissement de fiches (maintenant sous forme électronique) regroupées par disciplines, sous-disciplines, maladies et, d'autre part, l'effort et le temps de les consulter avant une prise de décision pour un malade. Seul un petit nombre de structures de recherche et d'unités d'enseignement s'en préoccupe ; ainsi la France a pris, dans le domaine de la valorisation des résultats de l'épidémiologie clinique auprès des praticiens, un retard considérable face aux États-Unis et à de nombreux pays européens. Une barrière linguistique est souvent invoquée.

Toute décision politique de santé publique, à laquelle contribue l'épidémiologie, nécessite une adhésion de la population. Pour ce faire, le rapport soumet des propositions pour mieux insérer, dans le système éducatif, une approche des notions de risque, de facteur de risque et des principes et des méthodes de l'épidémiologie. L'ouverture aux enseignements universitaires de biologie, de mathématiques, d'économie, de sociologie de l'épidémiologie est vivement recommandée et rendue possible par les nouvelles dispositions d'organisation des études supérieures qui ont introduit un fort degré de liberté inexistant auparavant. L'épidémiologie ne doit pas rester cantonnée au seul secteur médical. La création d'un nouveau grand établissement (École des Hautes Études de Sécurité Publique) est susceptible d'être la structure d'enseignement supérieur attendue, pour assurer ce développement en relation avec les agences, les universités et les établissements de recherche.

Avis du Groupe de lecture critique

La partie analytique du rapport aurait pu être plus exhaustive encore et des listes, des tableaux synthétiques auraient eu leur place.

L'aide à la décision et l'épidémiologie théorique deviennent, comme on le voit actuellement, de plus en plus nécessaires, mais pour être utilisée à bon escient, les décideurs doivent connaître et accepter les hypothèses qui sous-tendent la méthode et les incertitudes inhérentes ; des propositions sont faites à ce propos.

Enfin, la formation à l'épidémiologie, dans l'ensemble du système éducatif, est soulignée en notant que l'épidémiologie se trouve à la croisée de trois grands secteurs disciplinaires mais que si des spécialistes de chacun doivent exister, en outre, ils doivent collaborer au quotidien.

Les participants au groupe de lecture critique ont examiné le rapport dans sa partie analytique, puis en ont débattu. Dix d'entre eux ont bien voulu résumer par écrit leurs interventions. La nécessité de préciser la limitation du sujet à l'épidémiologie humaine s'impose, car l'épidémiologie animale est fort développée et enseignée au-delà des zoonoses. Les interventions enrichissent les données fournies dans les chapitres analytiques.

Les plus nombreuses interventions concernent les bases de données publiques, leur traitement statistique et de leur accessibilité aux épidémiologistes. Elles explicitent les possibilités et les contraintes, informent de l'existence de certaines bases non (ou peu) évoquées, en particulier, au sein des agences. Des propositions d'amélioration de la fluidité s'inscrivent dans le nouveau contexte institutionnel. Les conditions de traitement des cohortes restent difficiles, elles gagneraient en efficacité et en pérennité s'il existait des personnels dédiés. La part prépondérante de l'industrie du médicament dans l'épidémiologie associée aux études de mise sur le marché est explicitée.

L'Académie des sciences remercie les 15 participants aux nombreuses réunions du groupe de travail animé par Alain-Jacques Valleron, les 37 auteurs des monographies de la partie analytique, les participants au groupe de lecture critique et, en particulier, tous ceux qui ont bien voulu adresser leurs contributions écrites.

Jean-Yves Chapron, Directeur du service des Publications et Joëlle Fanon, chargée des rapports RST ont rendu possible et facilité la préparation, la rédaction et l'édition du rapport, qu'ils en soient aussi remerciés.

COMPOSITION DU GROUPE DE TRAVAIL

Animateur

Alain-Jacques VALLERON

Membre de l'Académie des sciences — Professeur à l'université Pierre-et-Marie-Curie — Praticien hospitalier — Unité de recherche « Épidémiologie, systèmes d'informations, modélisation » (Inserm UMR-S 707 et université Pierre-et-Marie-Curie)

Membres du groupe

Laurent ABEL

Directeur de recherche à l'Inserm — Codirecteur de l'unité de recherche « Génétique humaine des maladies infectieuses » (Inserm U 550 et université René-Descartes)

Pierre-Yves BOËLLE

Maître de conférences à l'université Pierre-et-Marie-Curie
Praticien hospitalier — Unité de recherche « Épidémiologie, systèmes d'information et modélisation » (Inserm UMR-S 707 et université Pierre-et-Marie-Curie)

Marcel GOLDBERG

Professeur à l'université de Versailles Saint-Quentin-en-Yvelines — Praticien hospitalier — Unité de recherche « Santé publique et épidémiologie des déterminants professionnels et sociaux de la santé » (Inserm U 687, université Versailles Saint-Quentin-en-Yvelines, université Paris-Sud et Cnamts)

Gilles KAHN†

Membre de l'Académie des sciences — Président-directeur général de l'Inria

Henri LÉRIDON

Correspondant de l'Académie des sciences — Directeur de recherche à l'Ined — Directeur de l'unité de recherche « Épidémiologie, démographie et sciences sociales » (Inserm U 569, Ined et université Paris-Sud)

- Roger SALAMON
Professeur à l'université Victor-Segalen –
Bordeaux II – Président de l'Association des
épidémiologistes de langue française (Adelf)
– Directeur de l'unité de recherche « Épidé-
miologie, santé publique et développement »
(Inserm U 593 et université Victor-Segalen)
- Philippe SANSONETTI
Membre de l'Académie des sciences –
Professeur à l'Institut Pasteur – Directeur
de l'unité de recherche « Pathogénie
microbienne moléculaire »
(Inserm U 786 et Institut Pasteur)
- Gilbert SAPORTA
Professeur, titulaire de la chaire de Statis-
tique appliquée au Conservatoire national
des arts et métiers – Président de l'Internat-
ional Association for Statistical Computing
- Daniel SCHWARTZ
Correspondant de l'Académie des sciences
– Professeur émérite à l'université Paris-Sud
- Guy THOMAS
Professeur à l'Université Pierre-et-Marie-
Curie – Praticien hospitalier – Directeur
de l'unité de recherche « Épidémiologie,
systèmes d'information, modélisation »
(Inserm UMR-S 707 et université
Pierre-et-Marie-Curie)
- Maurice TUBIANA
Membre de l'Académie des sciences –
Professeur émérite à l'université Paris-Sud –
Directeur honoraire de l'Institut
Gustave-Roussy
- Marc YOR
Membre de l'Académie des sciences –
Professeur à l'université
Pierre-et-Marie-Curie
- Coordonnateur**
- Jean DERCOURT
Secrétaire perpétuel de l'Académie
des sciences – Professeur émérite
à l'université Pierre-et-Marie-Curie

Chapitre 4.5

Frédéric Yves BOIS

Délégué scientifique — Direction des Risques chroniques de l'Ineris

Chapitre 5.1

Alexandre ALCAÏS

Chargé de recherche à l'Inserm — Unité de recherche « Génétique humaine des maladies infectieuses » (Inserm U 550 et université René-Descartes)

Jean-Laurent CASANOVA

Professeur à l'université René-Descartes — Praticien hospitalier — Codirecteur de l'unité de recherche « Génétique humaine des maladies infectieuses » (Inserm U 550 et université René-Descartes)

Chapitre 5.2

Philippe AMOUYEL

Directeur général de l'institut Pasteur de Lille — Directeur de l'unité de recherche « Épidémiologie et santé publique » (Inserm U 508 et Institut Pasteur Lille)

Chapitre 5.3

Annick ALPÉROVITCH

Directeur de recherche à l'Inserm — Unité de recherche « Neuro-épidémiologie » (Inserm U 708 et université Pierre-et-Marie-Curie)

Jean-François DARTIGUES

Professeur à l'université Victor Segalen – Bordeaux II — Praticien hospitalier — Unité de recherche « Épidémiologie, santé publique et développement » (Inserm U 593 et université Bordeaux II)

Chapitre 5.4

Catherine HILL

Chef du Service de Biostatistique et d'épidémiologie de l'Institut Gustave-Roussy

Simone BENHAMOU

Directeur de recherche à l'Inserm — Unité de recherche « Épidémiologie des cancers » (Inserm U 521 et université Paris-Sud)

Chapitre 7

Annie FAGOT-LARGEAULT

Membre de l'Académie des sciences —
Professeur au Collège de France, chaire
de Philosophie des sciences biologiques
et médicales

Chapitre 8.1

Philippe RAVAUD

Professeur à l'université Denis-Diderot —
Praticien hospitalier — Responsable du
département d'Épidémiologie, biostatistique
et recherche clinique de l'Hôpital Bichat

Chapitres 8.2 et 8.4

Michel SETBON

Directeur de recherche au CNRS —
Laboratoire d'Économie et de sociologie
du travail du CNRS (Aix-en-Provence)

Chapitre 8.3

Didier TORNÏ

Chargé de recherche à l'Inra —
Unité de recherche « Transformations
sociales et politiques liées au vivant »

Chapitre 9.2

Annie MAMECIER

Doyenne de l'Inspection générale
de l'éducation nationale —
Groupe Sciences de la vie et de la Terre

Chapitre 9.3

Antoine FLAHAULT

Professeur à l'université Pierre-et-Marie-
Curie — Praticien hospitalier —
Chef du département de Santé publique
de l'Hôpital Tenon

Chapitre 9.4

Jean-Christophe THALABARD

Professeur à l'université René-Descartes —
Praticien hospitalier — Responsable
de l'unité « Recherche clinique » du CHU
Hôpital Necker-Enfants malades

TABLE DES MATIÈRES

Rapport Science et Technologie	iii
Composition du Comité RST	v
Avant-propos	ix
Composition du groupe de travail	xv
Introduction	xxxi
CHAPITRE 1 Définition, historique et champs de l'épidémiologie humaine	1
1. Présentation générale de l'épidémiologie humaine	3
1.1. Le niveau « populationnel » incontournable en recherche médicale	3
1.2. Définition de l'épidémiologie	3
1.3. L'intégration de l'épidémiologie dans la recherche médicale . . .	5
2. Historique de l'épidémiologie : une science récente	5
3. Épidémiologie d'observation, épidémiologie expérimentale, épidémiologie théorique	10
3.1. L'observation en épidémiologie	11
3.2. Épidémiologie expérimentale	24
3.3. La poussée de l'épidémiologie théorique et de la modélisation .	27
3.4. Les « niveaux de preuve » en épidémiologie	29
3.5. Trois autres césures de l'épidémiologie	30

CHAPITRE 2 État des forces épidémiologiques en France 37

SOUS-CHAPITRE 2.1 L'épidémiologie humaine 39

1. La formation, les métiers, les moyens 39
 - 1.1. Qu'est-ce qu'un épidémiologiste ? 39
 - 1.2. Les autres métiers 40
 - 1.3. Les conditions de la réalisation d'études épidémiologiques . . . 40
2. Grands domaines et structures d'exercice de l'épidémiologie 43
 - 2.1. La recherche 43
 - 2.2. La surveillance épidémiologique 62
 - 2.3. Études diverses 65
 - 2.4. Aide à la décision et aux politiques de santé publique 65

SOUS-CHAPITRE 2.2 L'interface épidémiologie animale - épidémiologie humaine 71

1. Vétérinaires et épidémiologie 71
2. Danger animal 72
 - 2.1. Pour l'homme 72
 - 2.2. Pour les animaux domestiques 74
3. Unicité du vivant 75
4. Pistes de réflexion 76

CHAPITRE 3 Sources de données pour l'épidémiologie 79

SOUS-CHAPITRE 3.1 Les enquêtes de santé, les données démographiques, la mortalité... 81

1. L'organisation du système statistique en santé 82
 - 1.1. Le Cnis et la coordination des programmes des principaux
acteurs 82
 - 1.2. Un système néanmoins assez éclaté, comprenant des acteurs
dédiés à l'information 84
 - 1.3. Des questionnaires pour qui l'information statistique
et son analyse demeurent très secondes 87
 - 1.4. Un système demeuré longtemps isolé au plan international . . . 87
2. Quelques sources de données statistiques en santé publique 89
 - 2.1. Des enquêtes de santé « généralistes » 89
 - 2.2. La mortalité et la mortalité par causes :
quelques sources méconnues 93

SOUS-CHAPITRE 3.2	Le recueil de l'information dans les hôpitaux	97
1.	Le programme de médicalisation des systèmes d'information (PMSI) dans les établissements de soins publics et privés	98
2.	Intérêt de l'utilisation des données du PMSI à des fins d'épidémiologie descriptive pour les pathologies chroniques : l'exemple du cancer . . .	99
3.	Apport des techniques de chaînage des données du PMSI pour l'autoévaluation gérée par les médecins : l'exemple du Réseau périnatal de Bourgogne	103
4.	Apport des techniques de chaînage des données du PMSI et des registres pour les recherches sur l'utilisation des services de soins : l'exemple des États-Unis	105
CHAPITRE 4	Épidémiologie et sciences mathématiques	113
SOUS-CHAPITRE 4.1	Épidémiologie et statistique	115
1.	Statistique mathématique et épidémiologie	116
1.1.	Analyse des essais thérapeutiques	116
1.2.	Données manquantes	118
1.3.	Régression non paramétrique	118
1.4.	Méthodes géostatistiques	119
1.5.	Méthodes amplitude/fréquence	120
2.	Résultats de mathématique contemporaine appliqués à l'épidémiologie	122
2.1.	Méthodes de Monte-Carlo	122
2.2.	Critères de sélection de modèles	122
2.3.	Tests séquentiels	122
2.4.	Relations causales — Facteurs de confusion — Choix des covariables	124
3.	Les apports spécifiques de l'épidémiologie	125
3.1.	Apports méthodologiques à d'autres disciplines appliquées . . .	125
3.2.	Modèles de régression à effets aléatoires	126
3.3.	Travaux de recherche mathématique suscités par l'épidémiologie	126
SOUS-CHAPITRE 4.2	Épidémiologie et <i>data mining</i> ou fouille de données	129
1.	Définitions et historique	129
2.	Les outils	131
2.1.	Exploration « non supervisée »	131
2.2.	Prédiction ou apprentissage « supervisé »	131

3. Quelques applications en épidémiologie et santé publique	132
4. <i>Data mining</i> versus modélisation statistique	133
4.1. Le rôle des modèles	133
4.2. Problèmes spécifiques d'inférence et de validation dans les grandes bases de données	134
4.3. Penser la complexité	135

SOUS-CHAPITRE 4.3 Épidémiologie et informatique 137

1. Recueil et exploitation des données	138
2. Confidentialité : la (non-)identification des individus	139
3. Thésaurus, nomenclature, classification : des outils indispensables en perpétuelle évolution	140
4. Aide à l'analyse de données complexes par la modélisation informatique	141

SOUS-CHAPITRE 4.4 Épidémiologie théorique : modèles mathématiques et maladies transmissibles 143

1. L'approche « systèmes dynamiques » : équilibre, stabilité,	145
2. L'approche probabiliste et statistique : processus ponctuels, MCMC . .	148
3. L'approche « systèmes complexes » : apport des modèles structurés . .	149
4. L'approche « systèmes complexes » : les réseaux de contact	150
5. L'approche « contrôle » appliquée aux systèmes dynamiques : l'exemple de la viabilité	151
6. Autres développements récents	152

SOUS-CHAPITRE 4.5 La modélisation en épidémiologie du cancer 157

1. Historique et cadre général	158
1.1. Les mathématiques sont utiles à l'étude de toutes les étapes de la cancérogenèse	159
2. Évolution, besoins, perspectives	160

CHAPITRE 5 Épidémiologie et sciences biologiques : quatre exemples 165

SOUS-CHAPITRE 5.1 Épidémiologie génétique des maladies infectieuses 167

1. La variabilité de réponse à un agent infectieux	168
2. Les méthodes de l'épidémiologie génétique	171

2.1. L'analyse de liaison génétique	172
2.2. Les études d'association	173
3. Illustration : les infections mycobactériennes	176

SOUS-CHAPITRE 5.2 Du risque cardiovasculaire aux démences : les voies de l'épidémiologie moléculaire 185

1. De l'irruption de la biomédecine en épidémiologie cardiovasculaire	186
2. Entre épidémiologie et biologie, un rapprochement naturel	187
3. L'introduction des variables biologiques dans les études épidémiologiques : une nécessaire adaptation des outils biologiques	188
4. Des hypothèses sur les mécanismes d'apparition des maladies	189
5. Une approche moléculaire de la classification des maladies	190
6. De l'appropriation de l'épidémiologie par les biologistes et inversement	192

SOUS-CHAPITRE 5.3 Épidémiologie et vieillissement : des sciences sociales à la biologie moléculaire 197

1. Le vieillissement : un processus qui commence bien avant d'être âgé	198
2. Épidémiologie du vieillissement ou des maladies liées à l'âge ?	198
3. L'épidémiologie génétique de la longévité	200
4. L'âge : le meilleur marqueur épidémiologique du vieillissement ?	201
5. Le poids des facteurs sociaux et économiques	202

SOUS-CHAPITRE 5.4 Épidémiologie des cancers 207

1. Identification des facteurs de risque	208
1.1. Une méthode rigoureuse	208
1.2. Des résultats	209
1.3. La mesure des risques et l'importance de la dose	210
2. L'épidémiologie moléculaire des cancers	211

CHAPITRE 6 L'épidémiologie sociale 215

1. Pourquoi s'intéresser à la dimension sociale des problèmes de santé ?	217
2. Quelques règles de travail	220
3. L'apport d'autres disciplines	221
4. Les outils de l'épidémiologie sociale	225
4.1. Des variables d'une nature différente	225
4.2. Un champ particulier : travail et santé	226

4.3. Des références théoriques	226
4.4. L'analyse de biographies	227
4.5. Analyses multiniveaux	229
4.6. Les méthodes d'enquête	229
5. L'utilisation des résultats	230
6. Forces et faiblesses de l'épidémiologie sociale en France	231
CHAPITRE 7 Épidémiologie et causalité	237
CHAPITRE 8 L'épidémiologique : la base scientifique de la santé publique	247
SOUS-CHAPITRE 8.1 L'épidémiologie clinique, base scientifique pour la prise en charge rationnelle des malades	249
1. Quelques données historiques	249
2. Définition de l'épidémiologie clinique	250
3. Apport de l'épidémiologie clinique	251
4. Le développement et l'enseignement de l'épidémiologie clinique . . .	255
5. La situation française	255
SOUS-CHAPITRE 8.2 Épidémiologie et décision en santé publique	261
SOUS-CHAPITRE 8.3 « Mais pourquoi résistent-ils ? » Conditions de réalisation d'actions de santé publique sur une base épidémiologique	265
1. Une science en débat public	266
2. L'épidémiologie au risque de la politique	267
3. De l'épidémiologie de laboratoire à la pratique de santé publique . .	268
4. Quand l'épidémiologie réussit	269
SOUS-CHAPITRE 8.4 Force de conviction de la modélisation prévisionnelle épidémiologique dans l'action publique face au risque	273
1. Risque de transmission du vMCJ en transfusion sanguine	274
2. Évaluation du risque d'ESB aux États-Unis	276
3. La modélisation prévisionnelle épidémiologique entre besoin et pertinence	277

- 4. Le champ de la modélisation en relation avec sa force de conviction . 279
- 5. Vers un développement nécessaire et problématique
de la modélisation épidémiologique 281

CHAPITRE 9 Les conditions nécessaires au développement de l'épidémiologie 285

SOUS-CHAPITRE 9.1 La formation en épidémiologie humaine à l'université 287

- 1. Préalables 287
 - 1.1. Le positionnement académique de l'épidémiologie 287
 - 1.2. La faiblesse de la formation : un faux procès 289
 - 1.3. La formation : une grande variété 289
- 2. En France 290
 - 2.1. La formation à et par la recherche en épidémiologie 291
 - 2.2. Santé publique 291
 - 2.3. Médecine 294
 - 2.4. La formation continue 298
- 3. À l'étranger 299

SOUS-CHAPITRE 9.2 Épidémiologie, enseignement et éducation à la santé au lycée et au collège 301

- 1. L'épidémiologie humaine dans les programmes d'enseignement
des écoles et des collèges en France 301
 - 1.1. Une première sensibilisation à l'école primaire 301
 - 1.2. Une formation basée sur la liaison entre éducation à la santé
et éducation à l'environnement 302
 - 1.3. Une articulation entre instruction et éducation
aux responsabilités en classe de troisième 302
- 2. L'épidémiologie dans les enseignements de SVT au lycée 303
 - 2.1. Une imprégnation trop modeste dans les séries de culture
générale 303
 - 2.2. Un manque important de données épidémiologiques
pour la série scientifique 304
- 3. Quelques exemples d'ancrage de l'épidémiologie
dans les programmes étrangers 305
 - 3.1. Données des sites gouvernementaux de l'éducation
dans les pays francophones 305
 - 3.2. Données des sites de l'éducation dans un pays non francophone 308

4.	L'épidémiologie au service d'une éducation à la santé renouvelée	310
4.1.	Une ambition nationale non satisfaite	310
4.2.	Des recommandations utiles	311
4.3.	L'apport de l'épidémiologie aux innovations pédagogiques	311
4.4.	Épidémiologie et formation	311

SOUS-CHAPITRE 9.3 Les moyens nécessaires à la recherche épidémiologique 313

1.	Bonnes pratiques en épidémiologie : le coût de la qualité	314
2.	L'épidémiologie populationnelle est hautement consommatrice de ressources	315
3.	L'épidémiologie clinique repose sur des standards de qualité justifiant des coûts élevés	326
4.	Quel serait le coût d'un vaste projet épidémiologique mis en place en 2005 en France ?	327
5.	Les atouts et les faiblesses de la France comme lieu de réalisation des essais cliniques	329

SOUS-CHAPITRE 9.4 Études épidémiologiques en santé : aspects législatifs et réglementaires 333

1.	Les dispositifs protégeant la vie privée	334
1.1.	La loi relative à l'informatique, aux fichiers et aux libertés (loi 78-17 ; 06/01/1978)	334
1.2.	La charte des bonnes pratiques épidémiologiques	337
2.	La protection de l'intégrité physique des personnes	337
2.1.	La loi n° 88-1138 du 20/12/1988 dite « loi Huriet-Sérusclat »	337
2.2.	De la propriété des données et de l'obligation d'informer	339
2.3.	Les dispositions récentes : loi n° 2004-806 du 09/08/2004	340
3.	Éléments de comparaison internationale	342
3.1.	Les pays européens (Sénat, LC n° 89, 2001)	342
3.2.	Pays non européens	344
3.3.	Les IRB (International Review Board)	346
3.4.	Protection des personnes et Health Insurance Portability and Accountability Act — 1996 (HIPAA, 1996, Medical Privacy : National Standards to protect the privacy of personal health information)	347
4.	Le cas particulier des recherches gènes-environnement	347
5.	Protection des personnes et partage des données	349

Conclusion	353
1. Les constats	354
1.1. Une forte demande de la société en épidémiologie	354
1.2. Un positionnement difficile de l'épidémiologie par rapport aux autres disciplines scientifiques	354
1.3. Les moyens nécessaires à la recherche épidémiologique largement ignorés	355
1.4. La recherche épidémiologique intégrée à la recherche biomédicale	356
1.5. Le besoin en statisticiens, probabilistes, informaticiens, s'accroît	357
1.6. La recherche épidémiologique nécessite aussi des liens étroits avec les sciences humaines et sociales	357
1.7. Le déficit d'interaction entre « statistiques sanitaires » et recherche épidémiologique	358
1.8. Le sous-dimensionnement de la recherche et de l'enseignement épidémiologique en France	359
1.9. Faiblesse de la culture épidémiologique du public et des professionnels de santé	360
2. Les changements institutionnels récents en santé publique à prendre en compte	361
2.1. En préalable : distinguer santé publique et épidémiologie	361
2.2. Les agences sanitaires issues de la loi de 1998	362
2.3. L'École des hautes études en santé publique (EHESP)	363
2.4. Le projet d'Institut virtuel de recherche en santé publique	363
3. Orientations prioritaires du rapport (OPR)	364
Recommandations	375
Summary and recommendations	379
Groupe de lecture critique	383
Composition du Groupe de lecture critique	385
Commentaires de l'Agence française de sécurité sanitaire des aliments (Afssa)	389
Commentaire de la Caisse nationale de l'assurance maladie des travailleurs salariés (Cnamts)	391

Commentaire de la Commission nationale de l'informatique et des libertés	393
Commentaire de la Direction de la recherche, des études, de l'évaluation et des statistiques	395
Commentaire de l'École nationale vétérinaire d'Alfort (ENVA)	399
Commentaire de la Haute autorité de santé	403
Commentaires de l'Institut national de la santé et de la recherche médicale	405
Commentaire de l'Insee	407
Commentaire de l'Institut de veille sanitaire (InVS)	413
Commentaire du Leem	415
Présentation à l'Académie des sciences, par Jean-François Bach	417
Intervention de Maurice Tubiana, membre de l'Académie des Sciences	423

INTRODUCTION

Selon la définition donnée au chapitre 1, l'épidémiologie décrit les variations de fréquence des maladies dans les groupes humains, et recherche les déterminants de ces variations. Elle vise à la compréhension des causes des maladies, et à l'amélioration de leurs traitements et moyens de prévention.

L'épidémiologie humaine comprend trois volets :

- un volet *observationnel* divisé en épidémiologie descriptive et en épidémiologie analytique ; cette dernière analyse les relations entre facteurs de risques supposés et maladies ; son but ultime est d'identifier ceux de ces facteurs qui sont causaux ;
- un volet *expérimental*, domaine des essais thérapeutiques cliniques, des essais de prévention, et des essais d'intervention ;
- un volet *théorique* concernant principalement l'étude de la dynamique des maladies transmissibles.

Sur le plan des méthodes, l'épidémiologie moderne s'appuie depuis longtemps sur la statistique ; réciproquement, elle a été, et est, à la source de la découverte de nouvelles méthodes statistiques (cf. sous-chapitres 4.1 « Épidémiologie et statistique » et 4.2 « Épidémie et « *Data Mining* » ou fouille des données »). Depuis plusieurs années, on assiste à une forte implication de nouveaux champs des mathématiques en épidémiologie, notamment calcul des probabilités, théories des systèmes complexes, analyse numérique et modélisation en général (cf. sous-chapitres 4.3 « Épidémiologie et informatique », 4.4 « Épidémiologie théorique : modèles mathématiques et maladies transmissibles » et 4.5 « La modélisation en épidémiologie du cancer »).

Comme il sera rappelé dans le chapitre 1, l'origine historique de l'épidémiologie peut être trouvée en France, avec les noms de D. Bernoulli au XVIII^e siècle et — surtout — de P.A. Louis au XIX^e siècle, dont la méthode (la « méthode numérique » en médecine) et les élèves diffusèrent en Europe et aux États-Unis. Cependant, si la France peut se targuer d'avoir été un berceau de l'épidémiologie, l'analyse objective des effectifs des chercheurs et des laboratoires (cf. chapitre 2 « État des forces épidémiologiques en France ») montre qu'elle est actuellement sous-dimensionnée par rapport aux pays comparables. Pourtant, la demande en épidémiologie grandit sans cesse, aussi bien pour des motifs scientifiques

(parce que c'est une discipline intégrative en recherche biomédicale), que pour des motifs de demande sociale (parce que c'est la science amont de la santé publique). Comment répondre, qualitativement et quantitativement, à cette demande ? C'est pour répondre à cette question que l'Académie des sciences a suscité ce rapport sur l'épidémiologie humaine, les conditions de son développement (puisqu'elle est sous-développée en France) et le rôle des mathématiques (puisqu'il est en train de s'élargir).

Quatre constatations d'actualité soutiennent l'opportunité d'un tel rapport en 2006 :

- 1) **Nécessité de mieux faire connaître ce qu'est l'épidémiologie** : la forte demande d'épidémiologie, aussi bien de la part de la société que de celle des instances scientifiques, est en effet souvent accompagnée d'une méconnaissance de ce qu'est l'épidémiologie, de ses champs, de ses méthodes, et des coûts de la recherche dans son domaine. Il est donc utile de rassembler ces informations en un seul document.
- 2) **Explosion de l'information susceptible d'être analysée en épidémiologie** : le nombre de variables mesurables en pratique par individu a crû de façon exponentielle : sur le plan de la recherche, accès aux données génomiques ; sur le plan clinique, informatisation des dossiers médicaux, avec intégration des informations d'imagerie numérique, et de biologie grâce à la généralisation de l'automatisation ; existence de gigantesques bases de données médicales construites initialement dans un but de gestion économique ou d'organisation des soins, mais dont on peut espérer qu'elles seront de plus en plus valorisées en termes de recherche ; foisonnement, enfin, de bases de données dans les domaines de l'environnement, du travail, etc., susceptibles de fournir des informations à analyser en relation avec les données médicales. Ce changement de dimension des systèmes d'informations crée une nouvelle donne pour l'épidémiologie : il faut en conséquence adapter la formation des épidémiologistes, définir les nouvelles compétences qui devront être présentes dans les équipes, caractériser les nouveaux environnements techniques et législatifs adaptés à cette évolution.
- 3) **Demande sociétale accrue et modifiée vis-à-vis de l'épidémiologie** : la demande est *accrue* dans le domaine traditionnel de la mesure de la santé : plus qu'autrefois, la société veut connaître avec précision la situation sanitaire, et l'état des facteurs de risque de toute catégorie susceptible de faire l'objet d'actions de santé publique ; la société veut pouvoir fixer des objectifs quantitatifs de progrès en santé publique, comme

dans les autres domaines, mais elle se heurte, plus souvent qu'ailleurs, à l'absence de données de base. Cette demande forte, et consommatrice de moyens, d'épidémiologie très descriptive et opérationnelle, ne rentre pas directement dans la vision de la recherche épidémiologique qu'ont la majorité des chercheurs épidémiologistes. Avec l'évolution des outils d'observation que permet désormais l'informatique (cf. sous-chapitre 4.3 « Épidémiologie et informatique »), il y a sans doute nécessité de réfléchir à une redéfinition du rôle de l'observation de terrain dans l'activité de recherche épidémiologique. La demande sociétale vis-à-vis de l'épidémiologie est aussi *modifiée* par rapport à celle du passé, à cause de l'intérêt croissant porté aux risques sanitaires, notamment ceux concernant les risques faibles (exemple : risques cancérigènes dans l'alimentation, risque éventuel des téléphones portables, ...), les risques nouveaux émergents (exemple : pneumonie atypique, ...), voire ceux associés à des risques encore inconnus (exemple : risques liés à une attaque bioterroriste). Or, la méthode épidémiologique fondée sur le suivi rétrospectif, ou prospectif, de groupes de sujets convenablement choisis, n'est pas adaptée pour répondre à cette demande (un exemple simple d'inadaptation étant l'étude des risques épidémiologiques en cas de bioterrorisme). Il faut donc examiner quelles sont les nouvelles méthodologies à développer et quelle est la meilleure manière de faire pénétrer dans la société une réflexion clarifiant ce qui relève ou non de l'approche épidémiologique ou non et quelle doit être l'interprétation de ses résultats.

- 4) **L'épidémiologie, base de rationalisation en santé publique** : la quatrième raison d'actualité pour établir ce rapport est le changement profond survenu dans la culture médicale au cours des vingt dernières années : on a reconnu que c'était en s'appuyant sur des démarches scientifiques rigoureuses — et non plus sur l'« expérience » ou la « leçon des maîtres » — que l'on devait décider de la meilleure politique de prévention ou de l'attitude thérapeutique la plus efficace vis-à-vis d'une maladie, ou même de la prise en charge d'un malade donné. Cet appel à la science est en pratique un appel à l'épidémiologie, et en premier lieu (mais pas exclusivement) à l'épidémiologie expérimentale (essais thérapeutiques ou essais de prévention randomisés). Cette « médecine fondée sur la preuve scientifique » (ou, en anglais, *Evidence Based Medicine*) devient la base culturelle de la médecine clinique (cf. sous-chapitre 8.1 « L'épidémiologie clinique, base scientifique pour la prise en charge rationnelle des malades »), et est déjà l'instrument de travail des agences réglementaires de santé publique (Anaes, Afssaps, ...). On doit, alors, noter qu'en pratique les études épidémiologiques appropriées nécessitent souvent une approche multicentrique très lourde à organiser, très coûteuse, et une longue durée d'observation.

La recherche des conditions de développement de l'épidémiologie en France, et le rôle des mathématiques

Pour réaliser les enjeux que l'épidémiologie doit affronter, un certain nombre de conditions nécessaires sont à remplir, que ce rapport analysera et à propos desquelles il donnera des recommandations.

Ces conditions sont d'abord *matérielles* : l'ambition de la mise en place d'une plate-forme épidémiologique étudiant plusieurs dizaines ou centaines de milliers de personnes sur lesquelles les informations génomiques, comportementales, d'imagerie, etc., sont recueillies, implique évidemment l'identification d'une organisation — y compris en termes de constitution d'équipes de travail performantes — qui n'a rien à voir avec ce qui était nécessaire pour mener les grandes enquêtes épidémiologiques des années 1960 ou 1970. Des standards internationaux très exigeants — tels ceux de l'ICH (International Committee on Harmonisation) doivent être appliqués. Il est donc indispensable de bien identifier les conditions de la mise en œuvre, difficile et coûteuse, de telles études dans le contexte matériel et législatif français. Sans quoi, la base culturelle de la médecine française s'appuiera uniquement sur des études étrangères, donc menées dans un contexte de système de soins différent, posant de ce fait des problèmes d'extrapolation. Ces aspects sont traités dans les sous-chapitres 9.4 « Études épidémiologiques en santé : aspects législatifs et réglementaires » et 9.3 « Moyens nécessaires à la recherche épidémiologique ».

Il y a également des conditions *intellectuelles* à remplir pour construire les systèmes d'information les plus pertinents, et analyser de façon optimale les masses d'informations recueillies, que ce soit à l'occasion de protocoles de recherche, ou en utilisant de grandes bases de données médicales, environnementales ou comportementales constituées initialement dans un autre but que la recherche médicale. Il faut qu'une plus grande fraction du monde scientifique (et non pas les seuls épidémiologistes professionnels, en bien trop petit nombre) se mobilise autour de l'approche épidémiologique. Réciproquement, l'épidémiologie ne peut se développer sans une très forte interaction avec les autres disciplines des sciences de la vie (cf. chapitre 5 « Épidémiologie et sciences biologiques : 4 exemples ») afin, en particulier, d'intégrer dans ses analyses les données provenant des innovations technologiques, notamment en génomique. L'importance des facteurs sociaux, au sens large, dans la genèse, la prise en charge et le devenir des maladies, n'est plus à argumenter. Aussi, l'épidémiologie doit-elle aussi avoir une forte interaction avec certaines sciences humaines et sociales (cf. le chapitre 6, « L'épidémiologie sociale », faisant l'état des lieux de l'épidémiologie sociale). L'épidémiologie, par les connaissances qu'elle a fournies, a permis d'améliorer l'état de santé collectif et la prise en charge individuelle de malades. Les exemples abondent à travers le rapport. Elle a démontré l'existence de

risques inconnus, et trouvé les moyens de les prévenir. Mais la connaissance ne suffit pas pour toujours générer un comportement rationnel face à la santé, ni de la part des « décideurs », ni de la part des citoyens. L'épidémiologie a démontré que le tabac causait le cancer des bronches, et d'autres maladies. Les paquets de cigarettes l'affirment sur leur emballage. Pourtant on continue à fumer. C'est pourquoi il est essentiel qu'une recherche adaptée en sciences humaines permette de comprendre les conditions de la traduction de la connaissance épidémiologique en comportement de santé publique (cf. sous-chapitres 8.4 « Force de conviction de la modélisation épidémiologique au service de l'action publique face au risque » et 8.3 « Mais pourquoi résistent-ils ? Conditions de réalisation d'actions de santé publique sur une base épidémiologique »).

Le rapport analysera tout particulièrement en détail les relations entre mathématiques et épidémiologie.

En effet, en plus de la statistique, depuis une vingtaine d'années et de façon croissante, les mathématiques sont devenues indispensables en épidémiologie, surtout dans le domaine des maladies infectieuses. Les probabilités, la modélisation des systèmes complexes y trouvent notamment des champs d'application nouveaux, de grande importance pratique, stimulants sur le plan de la recherche et faisables car, désormais, les données sont là. L'implication des mathématiciens français en épidémiologie, est non seulement souhaitable, mais possible car les nouveaux problèmes de l'épidémiologie leur fournissent des sujets réels d'intérêt.

Finalement, parce que l'épidémiologie humaine est non seulement une science qui, à ce titre, mérite d'entrer dans la culture générale de chacun, mais aussi parce que c'est la science en amont de la santé publique, le rapport étudiera comment sa valorisation sociale peut être améliorée : d'abord, par le biais de l'enseignement des collégiens et lycéens qui ignorent tout — pour l'instant — des méthodes et résultats de l'épidémiologie, en particulier en ce qui concerne la détermination et la quantification des facteurs de risque biologiques, comportementaux, environnementaux (cf. sous-chapitre 9.2 « Épidémiologie, enseignement et éducation à la santé au lycée et au collège ») ; ensuite, par un enseignement plus large de l'épidémiologie à l'université, qui devrait s'étendre en dehors des seules facultés de médecine et mieux atteindre les étudiants de biologie, de mathématiques, et de sciences humaines et sociales (cf. sous-chapitre 9.1 « La formation en épidémiologie à l'université »).

CHAPITRE 1

Définition, historique et champs de l'épidémiologie humaine

ALAIN-JACQUES VALLERON ET DANIEL SCHWARTZ

1 | Présentation générale de l'épidémiologie humaine

1.1 Le niveau « populationnel » incontournable en recherche médicale

Nul n'imagine pouvoir découvrir un nouveau test diagnostique, un nouveau traitement, ou l'étiologie d'une maladie en étudiant un seul patient : l'observation courante et le bon sens nous rappellent que deux patients atteints de la même maladie ne présentent en général pas les mêmes symptômes, et deux patients atteints de la même maladie ne répondent pas de la même façon au même traitement. De même, l'exposition au même facteur de risque (le tabac, tel polluant, etc.) a des conséquences variables en termes de maladies.

L'existence de cette variabilité implique que les conséquences d'une exposition à un facteur de risque, le succès d'un traitement, ou le pronostic d'une maladie ne peuvent s'exprimer qu'en termes de probabilités. On évalue la probabilité (ou « risque ») de ces événements pour un **individu** en recherchant, dans des **groupes** convenablement choisis, la fréquence des événements auxquels on s'intéresse et dans quelle mesure celle-ci dépend de la présence de « cofacteurs ». Ceci caractérise la démarche épidémiologique.

L'épidémiologie est indispensable pour comprendre la physiopathologie d'une maladie chez l'homme : l'omniprésence de la variabilité ne peut être surmontée ni par le recours à des modèles animaux ou *in vitro*, par construction réductionniste, ni par l'accès à des informations jusque-là inconnues grâce à la découverte de nouvelles technologies, telles celles de la génomique : la cause d'une maladie, même simple, ne peut être trouvée en lisant quelques puces à ADN : on retrouva rapidement qu'en génomique aussi, il fallait employer la méthodologie de l'épidémiologie : mesurer les cofacteurs, travailler sur des groupes de taille suffisante constitués sans biais, etc., à tel point que des recommandations explicites furent données aux chercheurs pour qu'ils tiennent compte de cette variabilité incontournable (Tilstone, 2003).

Ainsi, même en considérant les avancées technologiques les plus récentes, la recherche médicale implique la prise en compte de la variabilité, donc une approche populationnelle, c'est-à-dire épidémiologique (étymologiquement, *δημος* est la population).

1.2 Définition de l'épidémiologie

Définition : l'épidémiologie étudie les variations de fréquence des maladies dans les groupes humains et recherche les déterminants de ces variations.

Elle vise en particulier à la recherche des causes des maladies et à l'amélioration de leurs traitements et moyens de prévention.

- L'épidémiologie concerne l'**ensemble** des maladies et situations pathologiques, et pas seulement les seules maladies transmissibles.
- La recherche épidémiologique de déterminants des maladies se fait en comparant les fréquences de ces maladies dans des groupes ou sous-groupes au cours du temps, en fonction du lieu, de la profession, des antécédents médicaux, de caractéristiques phénotypiques ou biologiques, etc.
- La recherche de ces « déterminants » vise en général à identifier des facteurs de risque, individuels ou collectifs, des maladies étudiées. L'épidémiologie permet alors de fournir des modèles prévisionnels reposant sur la connaissance de ces facteurs de risque.
- Dans le cas des maladies infectieuses, les « déterminants » étudiés peuvent concerner la nature des interactions entre sujets infectieux, sensibles, immunisés, par exemple la géométrie du réseau de contacts entre individus. De tels déterminants sont alors découverts en modélisant mathématiquement les interactions.
- La recherche des « déterminants » des variations de fréquence des maladies mène à étudier parallèlement les variations à travers les groupes humains des **facteurs** intervenant dans la survenue des maladies : on parle alors d'épidémiologie de la tension artérielle, de la pollution atmosphérique, et même d'épidémiologie du médicament.
- Les groupes humains sur lesquels travaille l'épidémiologiste peuvent être :
 - des populations « naturelles », définies par exemple par leur localisation ;
 - des groupes constitués spécifiquement pour les besoins de la recherche (exemple : les femmes en âge reproductif employées dans telle industrie et exposées à un toxique entre telle et telle année) ;
 - des groupes « théoriques » (exemple, pour les besoins de la modélisation d'une épidémie : une population imaginaire dans laquelle on suppose qu'il n'y a pas de structure d'âge. . .).

1.3 L'intégration de l'épidémiologie dans la recherche médicale

La démarche épidémiologique diffuse dans l'ensemble de la recherche médicale : pour documenter ceci, nous avons analysé¹ les 100 derniers articles originaux publiés avant le 31 juillet 2004 et dont le titre comprenait le mot *epidemiology* : 34 étaient signés par des chercheurs relevant d'équipes d'épidémiologie (au sens large), mais 41 étaient signés en premier par des cliniciens (avec dans 9 cas seulement un co-auteur épidémiologiste), 16 par un biologiste (avec dans 3 cas un co-auteur épidémiologiste) et 9 par des personnels d'agences gouvernementales ou d'ONG.

Une partie importante de la production des épidémiologistes s'exprime dans les revues de spécialités médicales, et non d'épidémiologie : on ne peut pas caractériser la recherche épidémiologique par l'analyse des articles scientifiques parus dans les seuls journaux spécialisés d'épidémiologie. En effet, les résultats majeurs des épidémiologistes (par exemple, récemment, ceux concernant les risques associés à la prise de traitements substitutifs hormonaux) sont publiés soit dans des journaux généralistes médicaux (*NEJM*, *Lancet*, etc.) ou non médicaux (*Nature*, *Science*, *PNAS*. . .), soit dans des journaux spécialisés du domaine concerné de la recherche médicale (*AIDS*, pour une recherche épidémiologique sur le sida, *Circulation* pour une recherche épidémiologique dans le domaine cardiovasculaire, etc.). Cette affirmation est confortée par l'analyse des publications 2001-2002 des chercheurs dirigeant des équipes ou unités Inserm relevant de l'épidémiologie : environ 30 % de leurs articles ont été publiés dans des revues spécialisées d'épidémiologie ou de biostatistique ; la majorité (70 %) ont été publiées dans des revues de spécialités médicales ou biologiques (60 %) ou généralistes (10 %). Parmi ceux-ci, les chercheurs épidémiologistes étudiés en signent environ la moitié en premier ou en dernier, ce qui indique leur rôle moteur dans la production de ces connaissances.

La remarque précédente est d'ailleurs essentielle au moment d'évaluer la production d'une équipe ou d'un chercheur épidémiologiste ; le succès de leur recherche se mesure en considérant les publications parues dans les meilleurs journaux « d'application », et pas seulement dans les journaux de leur discipline auxquels ils réservent plutôt leurs innovations méthodologiques.

2 | Historique de l'épidémiologie : une science récente

Le recueil systématique de données de mortalité commença à Londres en 1592, et leur publication hebdomadaire fut réalisée en 1603 ; la première

¹ Les études bibliométriques qui suivent ont été réalisées par Sophie Lesieur, Inserm U 707, que nous remercions vivement.

analyse détaillée de l'information fournie par les certificats de mortalité est due à John Graunt (1620-1674) qui publia en 1662 les *Natural and political observations (. . .) upon the bills of mortality* et les adressa à l'astronome Huyghens. La première table de mortalité est due à un autre astronome, Halley, qui en 1693 modélisa les données de mortalité de Breslau (maintenant Wroslaw). C'est cette table de vie qu'utilisera Bernoulli dans ses travaux sur la modélisation de la variole (voir encadré 2).

Les premiers travaux d'épidémiologie expérimentale habituellement cités sont ceux de James Lind (1716-1794) qui en 1747 étudia les épidémies de scorbut à bord des longs courriers et découvrit qu'un régime comportant de la vitamine C (en l'occurrence des citrons) permettait de contrôler ces épidémies, ce qui fut appliqué en pratique sur les bateaux dès la fin du XVIII^e siècle.

La littérature internationale identifie souvent comme « père fondateur de l'épidémiologie » un Français, Pierre Charles Louis (voir encadré 2) qui créa la « méthode numérique » en médecine. Un autre « père fondateur » est l'anglais William Farr (1807-1883), élève de Pierre C. Louis, qui précisa, dès 1838 (Farr, 1838), la notion de risque, et montra l'importance des analyses longitudinales (« cohortes ») pour évaluer les risques. Il reçut en 1839 la charge de collecter des données de morbidité de façon systématique et est considéré comme le fondateur de la surveillance épidémiologique. Les racines historiques de l'épidémiologie sociale sont également françaises, avec la mise en évidence par Louis René Villermé en 1826 que les plus grandes mortalités observées dans certains quartiers trouvaient plus leur explication dans la pauvreté de ses habitants que dans l'« insalubrité » de leurs logements qui était généralement la cause évoquée.

Les débuts de l'« épidémiologie analytique », dans laquelle on recherche les facteurs de risque des maladies à partir d'observations faites sur des populations, datent du milieu du XIX^e siècle, époque de la fondation de la Société anglaise d'épidémiologie (1850). Trois travaux fondateurs sont habituellement cités : celui de Panum, qui étudia en 1846 la dynamique de la rougeole aux îles Féroé, identifia le mode de transmission direct de personne à personne, et fournit une estimation de la durée d'incubation ; celui de Snow, qui, en 1854, mena une étude épidémiologique lors de l'épidémie de choléra à Londres, à l'issue de laquelle, grâce à la comparaison des fréquences de la maladie dans des quartiers desservis par des réseaux d'eau différents, il conclut qu'il y avait un agent transmissible à l'origine du choléra et que celui-ci était véhiculé par l'eau. À peu près en même temps, en 1847, Semmelweis à Vienne découvrit que la fièvre puerpérale était une maladie transportée, en l'absence d'hygiène appropriée, par les mains des soignants des femmes pendant leurs accouchements. Ces trois découvertes ont en commun d'avoir directement conduit à des applications pratiques en santé publique. De plus, elles ouvrent la série — encore

actuelle — des découvertes épidémiologiques effectuées avant la découverte des mécanismes biologiques correspondants ; ce n'est en effet que seulement 20 ans plus tard que Pasteur mettait en évidence l'existence d'agents infectieux explicatifs des maladies transmissibles.

La fin du XIX^e siècle et le début du XX^e siècle furent en revanche relativement silencieux peut-être parce que la démarche épidémiologique est fondée essentiellement sur l'analyse de la variabilité, alors que l'idéologie scientifique dominante de cette époque était le déterminisme. Pourtant, on assistait parallèlement à la naissance de la statistique. Deux dates clés sont 1922 (introduction en agronomie, par Fisher, du tirage au sort pour assurer la comparabilité des groupes traité et témoins) et 1928 (théorie des tests de Neyman et Pearson). On trouve ensuite les travaux annonciateurs de l'épidémiologie expérimentale moderne : les deux premiers essais thérapeutiques, avec tirage au sort et test statistique, concernèrent la tuberculose et furent réalisés par Amberson en 1931 sur 24 patients atteints de tuberculose pulmonaire, puis par Bradford Hill en 1948. On doit noter aussi les travaux étudiant l'association entre la présence de fluor dans l'eau et la fréquence des caries dentaires ; ces travaux, d'abord exploratoires en 1938 aboutirent à la conclusion en 1942 que le fluor avait bien un rôle étiologique dans la prévention des caries dentaires et se terminèrent en 1956 par une étude « quasi expérimentale » où il fut démontré par une étude comparative que lorsqu'on ajoutait du fluor dans l'eau de boisson, on diminuait de façon très importante la fréquence des caries dentaires.

Dans la période 1950-1980, la démarche épidémiologique s'impose comme méthode de découverte des facteurs de risque des maladies, principalement dans le domaine des maladies chroniques. Celles-ci, à l'exception de quelques maladies monogéniques rares, se caractérisent par leur caractère multifactoriel, avec intervention possible de multiples facteurs de risque, biologiques, environnementaux, comportementaux, etc. Ce fut la naissance de l'épidémiologie utilisant systématiquement la statistique : celle-ci s'appuya au début souvent sur des études écologiques, puis, de plus en plus, la recherche des facteurs de risque fut faite grâce à des études où l'on comparait systématiquement le passé de malades et de « témoins » (études cas-témoins), ou bien où on suivait comparativement des sujets exposés, ou non exposés, à un facteur de risque (études de cohortes). Ainsi, de 1948 à 1952, en comparant 1 465 malades atteints d'un cancer des bronches à des témoins, Doll et Hill incriminent le tabac comme cause de ce cancer. En France, des travaux menés dès 1954 sur le même sujet par D. Schwartz et P. Denoix (Schwartz *et al.*, 1961) permirent de renforcer cette hypothèse — encore discutée à l'époque — de la causalité de la relation tabac-cancer des bronches, ceci par la mise en évidence du rôle de l'inhalation. Deux grandes études de cohortes la confortèrent encore : dans la première, à nouveau de Doll et Hill, c'est l'ensemble du corps médical anglais qui fut suivi (Doll et Hill, 1954) ; dans la seconde, menée aux États-Unis par

Hammond, 60 000 enquêteurs bénévoles suivirent pendant 5 ans un million de sujets (Hammond, 1966). Une découverte épidémiologique remarquable de la même époque, à la fois parce qu'elle montre l'importance de l'épidémiologie descriptive (actuellement trop négligée : ce sera un « leitmotiv » de ce rapport), l'aspect multifactoriel des étiologies, et la capacité de l'épidémiologie à conduire à des découvertes essentielles quand elle est conduite de façon intégrée avec la biologie, est celle du lymphome de Burkitt (Burkitt, 1962). Ce chercheur, au cours de ce qu'il appela un « safari » (Burkitt, 1962) épidémiologique, voulut comprendre les variations d'incidence de cette tumeur, en particulier avec l'altitude. Il trouva une association avec la présence de moustiques, et finalement ces recherches identifièrent pour la première fois qu'un virus (le virus d'Epstein-Barr) pouvait causer un cancer humain, la malaria agissant comme cofacteur en stimulant la prolifération des lymphocytes B, et en y augmentant la probabilité de remaniements chromosomiques ayant pour conséquence l'activation des oncogènes ; grâce aux travaux de recherche qui ont suivi, en particulier ceux du Circ, cette maladie a maintenant quasiment disparu. L'épidémiologie expérimentale (« essais thérapeutiques randomisés ») eut une montée en puissance parallèle : dès 1979 un rapport du NIH aux États-Unis, analysant 985 études d'évaluation thérapeutique effectuées sur la période 1975-1979 en trouva 60 % où le tirage au sort était pratiqué. Certains de ces essais thérapeutiques furent de véritables coups de tonnerre dans le monde de la médecine. Par exemple, au cours des années 1970, l'essai de l'UGDP (University Group Diabetes Program) mit à bas la pratique du traitement par hypoglycémifiants oraux du diabète de type II en montrant que non seulement les sujets ainsi traités n'avaient pas moins de complications cardiovasculaires que les sujets recevant un simple régime, mais même qu'ils présentaient plutôt une légère... surmortalité (Prout *et al.*, 1972). Les années 1970 voient le développement de l'« épidémiologie sociale » qui étudie le rôle des facteurs comportementaux, environnementaux, et sociaux (pauvreté, travaux dangereux) dans les inégalités de morbidité et de mortalité, et qui, corrélativement, suppose que les modifications de ces facteurs doivent être considérées en priorité pour améliorer l'état de santé de la population.

Depuis 1980, on assiste à une nouvelle étape de l'histoire de l'épidémiologie à quatre égards :

- arrivée des maladies émergentes au premier plan des préoccupations de santé publique : découverte du sida en 1981, de l'hépatite C en 1990, du nvCJD en 1996, du Sras en 2003, crainte d'une pandémie grippale, etc. ;
- reconnaissance par les agences réglementaires du rôle clé de l'épidémiologie en tant que science devant fonder la pratique médicale (« médecine fondée sur la preuve » ou « *Evidence Based Medicine* ») ;
- intérêt sociétal pour les risques sanitaires « faibles », voire encore jamais observés (le voisinage des centrales nucléaires, le téléphone cellulaire, les OGM, ...);

- changement d'échelle qualitatif et quantitatif dans les capacités de recueil de données épidémiologiques, grâce aux nouvelles technologies biomédicales qui produisent à faible coût des masses de données nouvelles, directement informatisées ; explosion des capacités de gestion et traitement de l'observation.

Encadré 1.1

Les racines françaises de l'épidémiologie

Le travail présenté par Daniel Bernoulli à l'Académie royale des sciences en 1760 est considéré comme l'ancêtre de l'épidémiologie mathématique et sa méthodologie n'a pas vieilli (Valleron, 2000 ; Dietz et Heesterbeek, 2000). Bernoulli construisit un modèle mathématique de l'épidémiologie de la variole afin d'évaluer, dans ce qu'on appelle maintenant une démarche bénéfico-risque, l'intérêt de la variolisation (technique consistant à effectuer une scarification de l'enfant sain avec des croûtes prélevées sur un malade varioleux). Le résultat de son modèle était favorable à l'usage de cette technique et ses travaux déclenchèrent un grand intérêt, et une vive polémique de la part de d'Alembert qui discuta en détail le problème que représente la comparaison entre un risque immédiat (celui de la variolisation) et un bénéfico plus lointain (ne pas avoir la variole quelques années plus tard).

Au siècle suivant, on trouve en France plusieurs grands précurseurs de l'épidémiologie, reconnus comme tels dans la littérature internationale (Bollet, 1973). Ainsi le précurseur de l'épidémiologie clinique fut Pierre Charles Louis (1787-1872) qui fonda la « médecine numérique ». Son traité fut traduit aux États-Unis dès 1838. Il eut immédiatement un grand nombre de disciples dont William Farr en Angleterre, de nombreux épidémiologistes américains, et Joseph Skoda de Vienne, le maître et un des rares protecteurs de Semmelweis. Au lieu de se fier à la « tradition » ou à « l'autorité des maîtres », il avait décidé de collecter systématiquement les informations au lit du malade et d'analyser objectivement les résultats des traitements en fonction de ces informations. C'est ainsi que dans son travail le plus connu, il montra — à la fureur de Broussais — que l'usage de la saignée, pratiquée systématiquement grâce à des sangsues (environ 20 millions de sangsues importées en 1830...) pour traiter les pneumonies ne se justifiait pas. Il déclara devant l'Académie de médecine : « Sans l'aide de la statistique, rien qui ressemble à une véritable science médicale n'est possible ». Stimulé par l'approche de Pierre Charles Louis, dont l'utilité était débattue devant l'Académie des sciences (1835) et celle de médecine (1837) où on lui reprochait d'avoir « une démarche statistique mélangeant dans un même dénominateur des individualités différentes », Jules Gavarret (1809-1890),

polytechnicien et médecin, publia en 1840 ses *Principes généraux de statistiques médicales*. Il fut sans doute le premier à utiliser la démarche du test statistique afin d'évaluer si des proportions observées étaient significativement différentes. Ses premiers travaux furent appliqués au cas du choléra et il démontra que l'incidence n'était pas « significativement » différente chez les hommes et chez les femmes, alors qu'elle était différente dans certains quartiers par rapport à d'autres. Louis René Villermé (1782-1863) fut de son côté un précurseur mondialement reconnu de l'épidémiologie sociale : un de ses travaux importants concerna *Le tableau physique et moral des ouvriers dans les fabriques de coton, de laine et de soie* (1840) et eut comme conséquence pratique le vote d'une loi sur le travail des enfants (1850).

C'est aussi en France, à la fin du XIX^e siècle, que Bertillon créa la première classification des maladies permettant d'effectuer des relevés épidémiologiques systématiques et rigoureux d'entités cliniques bien définies (on en est actuellement à la dixième révision de cette classification internationale des maladies).

L'épidémiologie fut quasiment absente de France jusqu'à la fin des années 1950, où elle fut initiée grâce à la création simultanée par Daniel Schwartz de l'unité de recherches statistiques de l'Institut national d'hygiène (futur Inserm), et, sur le plan de l'enseignement, du Cesam (Centre d'enseignement de la statistique appliquée à la médecine) qui fournit, pour la première fois, une formation de très haute qualité, et accessible aux biologistes et médecins, y compris par correspondance. Le Cesam forma plus de 30 000 diplômés et fut l'outil principal de diffusion de la culture statistique et épidémiologique dans le milieu médical.

3 | Épidémiologie d'observation, épidémiologie expérimentale, épidémiologie théorique

L'épidémiologie comprend trois niveaux d'approche : l'observation, l'expérimentation et la théorisation. Ces niveaux sont d'importances très inégales :

- l'observation a été jusqu'ici le centre de gravité de l'activité de recherche épidémiologique ; elle a concerné majoritairement les travaux d'épidémiologie « analytique » visant à analyser les relations entre facteur(s) de risque et maladie(s). Les statistiques sanitaires qui sont, *a priori*, un outil naturel d'observation ont été jusqu'ici développées quasiment indépendamment de la recherche épidémiologique. Cependant, un intérêt croissant est manifesté à leur égard, d'autant plus que les nouvelles technologies

de l'information et de la communication leur donnent une envergure sans comparaison avec le passé. Ceci sera largement détaillé dans les chapitres suivants ;

- l'expérimentation, qui comprend principalement les essais randomisés thérapeutiques et de prévention, a pris une importance décisive en médecine et en santé publique. Ce domaine est en forte croissance et diffuse dans toutes les disciplines médicales. On ne peut plus le considérer comme relevant exclusivement de l'épidémiologie, sauf en ce qui concerne la recherche méthodologique ;
- l'épidémiologie théorique a représenté au cours du XX^e siècle, et jusqu'aux années 1980, une part faible, négligeable et hélas négligée, de la recherche épidémiologique. Cette situation évolue rapidement, en raison d'une part de la diffusion générale de la culture des méthodes de la modélisation et d'autre part, de la priorité que représentent depuis la fin du XX^e siècle les maladies infectieuses, qui sont un domaine prioritaire d'application de ces méthodes.

3.1 L'observation en épidémiologie

Les chercheurs en biologie, physique, sciences de l'Univers, chimie ont besoin de l'observation pour générer leurs hypothèses. Leur observation s'appuie sur des instruments et des méthodes sans cesse renouvelées par des innovations technologiques. Ils mettent au point des expériences permettant de tester ces hypothèses et de vérifier leur capacité prédictive. Finalement, ils émettent des théories, réalisent des modèles qu'ils confrontent autant que possible à l'observation. Observation, expérimentation et théorisation forment donc une boucle générant la découverte scientifique.

On pourrait s'attendre à ce que l'épidémiologie suive ce stéréotype et dispose donc d'outils d'observation puissants pour construire *in fine* ses théories. En réalité, les outils d'observation épidémiologique sont en général construits en aval des hypothèses de recherche, et non en amont pour les générer. Les Anglo-Saxons parlent à ce propos d'une démarche *hypothesis driven* (pilotée par l'hypothèse). En revanche, on doit noter la faiblesse des investissements en termes d'outils d'observation épidémiologique systématique à grande échelle destinée à faciliter la génération d'hypothèses (et non pas conséquences d'hypothèses déjà générées).

L'observation épidémiologique comporte deux volets : le premier (épidémiologie descriptive) vise à décrire les variations dans le temps, l'espace, etc., de l'état de santé, ou des facteurs de risque. Le second (épidémiologie analytique)

concerne les observations visant à analyser la relation possible entre un ou des facteurs de risque, et la survenue de une ou des maladies. De nombreux (la majorité ?) chercheurs épidémiologistes ne considèrent pas l'épidémiologie descriptive comme une réelle activité de recherche ; pour eux l'activité de recherche doit être guidée par une hypothèse de travail relative à la relation entre facteurs de risque et santé. Pour ce faire, l'épidémiologie analytique est le bon outil. Remarquons cependant que le clivage recherche/non recherche ne recouvre pas la séparation épidémiologie analytique/descriptive : bien évidemment, et l'histoire le démontre, on peut faire une découverte épidémiologique grâce à l'épidémiologie descriptive (un exemple remarquable est celui du lymphome de Burkitt déjà cité) ; réciproquement, une « étude » d'épidémiologie analytique peut procéder d'une demande sociale, et non pas d'une démarche de recherche scientifique.

3.1.1 L'épidémiologie descriptive : une trop faible priorité de la recherche épidémiologique

L'épidémiologie descriptive a pour objet de décrire les variations dans le temps, l'espace ou à travers les différents groupes sociaux, de l'état de santé et des facteurs de risque de la population à partir de données obtenues exhaustivement ou grâce à des échantillons représentatifs de la population.

C'est cependant souvent à l'occasion de manques en « épidémiologie descriptive » que le public ou les décideurs soulignent le sous-développement de la recherche épidémiologique en France. Le besoin d'épidémiologie descriptive, et le manque de données épidémiologiques, apparaissent en effet clairement lorsqu'on consulte le rapport annexé au projet de loi sur la politique de santé publique voté en 2004. Ce rapport listait 100 objectifs quantifiables de santé publique. Parmi les 35 objectifs relatifs aux facteurs de risque (alcool, tabac, nutrition, etc.), 13 étaient considérés comme « ayant pour préalable la production d'informations épidémiologiques ». Parmi les 65 objectifs relatifs aux diverses pathologies, 22 étaient dans ce cas. Par exemple, un objectif affiché était de réduire de 30 % la dose de radiations reçue par la population au cours d'exams diagnostiques ; mais l'information épidémiologique sur ce qu'est cette dose n'existe pas ! Une telle méconnaissance est insupportable pour le public, et les « décideurs ».

L'épidémiologie descriptive repose actuellement majoritairement sur des données recueillies en dehors du milieu de la recherche académique, dans le cadre d'administrations centrales ou hospitalières et d'agences sanitaires qui produisent des « statistiques sanitaires ».

Des exceptions existent, et l'activité d'observation épidémiologique *stricto sensu* n'est pas absente de l'Inserm. Elle se déroule soit au sein même d'équipes

de recherche, soit par le biais de conventions avec des agences, etc. Le Comité national des registres lui-même est copiloté par l'Inserm et l'InVS. Parmi les exemples notables d'activités d'observation épidémiologique menées au sein d'équipes de recherche, on peut citer : les statistiques de causes de décès (cf. sous-chapitre 3.1 « Les enquêtes de santé, les données démographiques, la mortalité ») qui sont recueillies au sein de l'Inserm, dans une structure spécifique (le CEPIDC). Un certain nombre de groupes de recherche de l'Inserm ont une activité d'épidémiologie descriptive découlant directement de leur programme scientifique : ceci concerne divers registres de morbidité, par exemple relatifs aux malformations congénitales ou à la maladie de Creutzfeldt-Jakob. L'absence de données permettant de modéliser la dynamique des maladies infectieuses fréquentes et de développer de nouveaux algorithmes d'alerte a aussi motivé le développement à l'Inserm, il y a plus de 20 ans, d'un système d'observation télé-informatique épidémiologique en temps réel des épidémies de maladies infectieuses fréquentes (Sentinelles) considéré comme pionnier dans ce domaine (Graitcer et Thacker, 1986 ; McCormick, 1987). Dans un certain nombre de ces cas, ces recueils de données sont effectués dans le cadre de conventions entre l'Inserm et des agences, en général l'Institut national de veille sanitaire).

On peut aussi signaler des « observatoires » épidémiologiques tel Monica, observatoire européen dans le domaine cardiovasculaire, avec trois sites français s'appuyant sur des groupes de recherche épidémiologique performants du domaine. De même, alors que peu était connu de façon rigoureuse sur les comportements sexuels, une enquête de taille et de qualité fut effectuée à l'Inserm (l'ACSF : voir encadré 3.1.1). Ceci fut obtenu grâce au financement de l'ANRS, et ne l'aurait sûrement pas été sans la circonstance de l'épidémie de sida, malgré l'intérêt évident du sujet en épidémiologie et en santé publique.

Encadré 1.2

L' ACSF – Analyse des comportements sexuels en France

Malgré l'importance évidente du sujet, y compris en dehors du champ strict de l'épidémiologie, on fut obligé de constater, lors de l'épidémie de sida, qu'on ne disposait d'aucune donnée fiable sur la sexualité des Français depuis les travaux de Simon en 1970 alors que l'« évolution des mœurs » menait à l'évidence depuis à plus de promiscuité et donc à des risques accrus de contamination par voie sexuelle. D'autre part, les comportements vis-à-vis de la prévention des maladies sexuellement transmissibles étaient inconnus, alors que leur compréhension était nécessaire à la mise en œuvre de celle du sida. Une enquête nationale d'envergure portant sur 20 000 sujets de 16 à 69 ans, reposant sur des entretiens téléphoniques, fut menée en 1991-1992 avec le soutien de l'ANRS. L'enquête a fourni des données sur les

fréquences des rapports sexuels, les nombres de partenaires, les pratiques sexuelles, et les attitudes vis-à-vis de la prévention des maladies sexuellement transmissibles.

A. Spira, N. Bajos (1993). *ACSF : Les comportements sexuels en France : Rapport au ministre de la Recherche et de l'Espace*. Documentation française (La), Paris. 351 p. Voir aussi *Nature* (1992), **360** (6403) : 407-9.

Ces exemples d'implications d'équipes de recherche dans l'observation sont très positifs. Cependant, ils sont dispersés, parcellaires. Il faut noter que la France n'a pas encore investi dans la mise en place de systèmes d'observation épidémiologique puissants directement destinés à la recherche épidémiologique, associant mesures de la morbidité et évaluation des facteurs de risque. De tels exemples existent ailleurs et mèneront à un avantage positif (voir le cas du Nhanes — National Health and Nutrition Examination Survey — décrit dans le chapitre 9.3 « Moyens nécessaires à la recherche épidémiologique »). Aussi doit-on explorer de près ce que l'existant, obtenu dans d'autres finalités que celle de la recherche, peut apporter. C'est l'objet du paragraphe suivant.

3.1.2 Statistiques sanitaires, systèmes d'information médico-économiques et recherche épidémiologique

De nombreux systèmes d'information relatifs à la santé et aux soins dans la population et à l'hôpital sont disponibles en France (cf. chapitre 3 « Sources de données pour l'épidémiologie » et les actes de la journée d'étude organisée par le Cnis et la Drees en janvier 2000 (http://www.cnis.fr/doc/rapports/RAP_0025.HTM#document)). Ces systèmes, à de rares exceptions près, ont été développés en dehors de la recherche académique, car leur finalité est en général la gestion du système de santé ou l'amélioration de la prise en charge du patient. Ce sont donc les organismes ayant en charge ces missions qui ont la responsabilité de ces systèmes d'information. Malheureusement, on peut regretter que ces recueils considérables d'informations aient jusqu'ici très peu contribué au progrès des connaissances (ainsi qu'en atteste le très faible nombre de publications scientifiques françaises s'appuyant sur eux).

Un première raison en est que, souvent développés par un organisme particulier, naturellement pour répondre aux missions de cet organisme, ils sont sans liaison possible (et souvent, ni souhaitée) avec d'autres systèmes d'information développés par d'autres organismes qui apporteraient les informations complémentaires nécessaires sur le plan épidémiologique ; la mise en place de telles

liaisons entre systèmes d'information est souvent bloquée en évoquant la confidentialité des données médicales et les limitations (pourtant bien moins grandes qu'on ne le croit) de la réglementation en termes d'accès à des données sensibles (cf. sous-chapitre 9.4 « Études épidémiologiques en santé : aspects législatifs et réglementaires », ainsi que le rapport publié par la Drees en 2004²). En tout cas, la construction de dossiers « centrés sur le patient », grâce à la mise en liaison de bases de données différentes, aurait une grande rentabilité en recherche épidémiologique, sous réserve du respect nécessaire, et possible, de la confidentialité des données ; c'est ce qu'a montré l'exploitation des données des meilleures Health Maintenance Organizations (HMO) aux États-Unis : on trouve 422 références dans Pubmed pour la seule année 2003 dont le titre ou le résumé contiennent le sigle HMO.

Une seconde raison au manque de productivité de ces systèmes d'information en termes de recherche est le peu de communication entre les mondes de l'information sanitaire et celui de la recherche épidémiologique : les chercheurs épidémiologistes ne manifestent pas un grand intérêt pour les potentialités de tels systèmes, et ne se sont pas — ou n'ont pas été — associés à leur construction, à de rares exceptions près (voir l'encadré 3.1.3, relatif à l'infection à VIH). La qualité des données de ces systèmes d'information est souvent jugée insuffisante par les chercheurs pour leur travail. Ceci n'est pourtant pas systématiquement vrai et surtout oublie la dynamique de qualité des données qu'entraînerait leur meilleure valorisation en termes de recherche, dynamique qui profiterait d'ailleurs aussi, *in fine*, aux objectifs médico-administratifs initiaux de ces bases. En effet, les personnels de santé à la source du recueil de l'information dans ces systèmes font naturellement plus d'efforts pour fournir une information de qualité lorsqu'ils savent qu'elle sera utilisée par la recherche que lorsque la finalité est perçue comme uniquement « administrative ».

Pour conclure, il convient de souligner les progrès actuels : la possibilité d'accéder à des informations gérées par la Cnamts est désormais ouverte aux chercheurs (moyennant bien sûr le respect de règles précises) et ils peuvent ainsi compléter de façon très précieuse les recueils d'information dont ils disposent dans le cadre de leurs protocoles de recherche. Les bases de données disponibles sont maintenant de fait utilisées pour fournir des réponses rigoureuses à des questions épidémiologiques (ainsi, récemment, a-t-on pu facilement décrire avec précision les complications du diabète grâce à un échantillon représentatif de 10 000 diabétiques pris en charge par la Cnamts). De tels exemples sont encore trop exceptionnels et encore trop souvent limités à des sujets très finalisés. Notons, de plus, que des organismes relativement nouveaux (Drees au ministère de la Santé, Institut de veille sanitaire, ...) ont désormais la mission, les moyens

²« Élaboration d'un outil de mesure de la mortalité post-hospitalière », in vol. 40 de la série Études de la Drees, 2004.

et la volonté de dynamiser le recueil de l'information épidémiologique et travaillent déjà en collaboration avec des équipes de recherche. Ce mouvement doit absolument être poursuivi.

3.1.3 Les dossiers médicaux informatisés ou informatisables : une opportunité pour la recherche épidémiologique encore mal saisie

Qu'il faille surmonter le manque d'intégration entre la recherche épidémiologique et les outils d'information existants est particulièrement clair sur l'exemple des dossiers médicaux et carnets de santé divers : par exemple, depuis 1985, un système national d'information sur la santé de la mère et de l'enfant a été mis en place et tout enfant dispose d'un carnet de santé individuel (remanié en 1995) où sont notés tous les événements de santé importants concernant la période prénatale, au moment de la naissance (biométrie, examens cliniques, tests de dépistage), de l'enfance (examens du 8^e jour, du 9^e mois, du 24^e mois), vaccinations, etc. Les parents connaissent bien ce carnet de santé qui leur est réclamé périodiquement à différents moments de la vie de l'enfant et le conservent précieusement. Une partie de cette information (notamment les examens du 8^e jour, du 9^e et du 24^e mois font l'objet d'analyses statistiques locales (PMI) synthétisées par la Drees au niveau national). Il reste, hélas, que les millions de carnets de santé créés depuis de nombreuses années auraient pu être la base de futurs travaux épidémiologiques passionnants dans lesquels on corrèlerait les paramètres de l'enfant au cours de sa petite enfance avec des événements de l'âge adulte. Malheureusement, ils n'ont pas été informatisés et ne sont donc pas exploitables, pas même pour faire une étude simple d'épidémiologie des vaccins pourtant répertoriés de façon exhaustive par nécessité réglementaire. Ainsi, alors que 15 000 000 d'enfants ont été vaccinés contre l'hépatite B lors de la campagne de vaccination de 1995 (donc dans l'ère post-Internet) — dont on connaît les contestations qui en sont issues —, il est extraordinaire que, faute d'informatisation, aucune étude de suivi post-vaccinal systématique s'appuyant sur ces données ne soit possible.

La création d'un dossier électronique médical individuel a été décidée, dans le cadre de la loi du 9 août 2004 relative à l'assurance maladie, et devrait ouvrir de grands espoirs pour la recherche épidémiologique. Cependant, au cours de la préparation de cette loi, autant ont été mis en avant les bénéfices que ce dossier médical informatisé est susceptible d'apporter en termes de qualité de prise en charge du malade et de maîtrise des coûts, autant son potentiel pour la recherche épidémiologique n'a pas été souligné. Tel n'a pas été le cas, non plus d'ailleurs, dans le cahier des charges du projet anglais équivalent (qui bénéficiera d'un financement de 6 milliards de livres!).

Il faut affirmer qu'il est de l'intérêt général, aussi bien du point de vue de la recherche que du point de vue même du progrès de la prise en charge des malades, parce qu'il lui est intimement lié, que les nouveaux systèmes d'information sanitaires, médicaux, médico-économiques soient développés en y incluant une mission de recherche épidémiologique. Les problèmes de confidentialité des données existent bien sûr, mais la plupart sont surmontables, notamment grâce au travail en commun avec les associations de malades. L'exemple du sida (voir encadré 3.1.3 relatif au DMI2) a bien montré que l'on pouvait compter sur les patients pour être partie prenante, et non frein, de la recherche épidémiologique parce que c'est du progrès de la recherche que dépendront les progrès de la santé. Les défenseurs autoproclamés du « secret médical » devront réfléchir à la contradiction entre le souhait — passé dans la loi — que chaque malade ait accès à son dossier médical et la frilosité face à la mise en commun de ces données pour le progrès de la recherche médicale. Il est clair que certaines situations sont particulièrement difficiles de ce point de vue (maladies mentales, très grand âge), mais il est clair que le progrès de la prise en charge de ces personnes passe par la recherche et que la recherche passe par le recueil des données. Il faudra donc trouver les solutions technologiques et éthiques à ces problèmes, et non laisser ces populations à l'abandon en termes de recherche.

Encadré 1.3

Le DMI2, base de données Inserm-DHOS des séropositifs français au VIH (D. Costagliola, Inserm U720, Paris)

Ce projet associe la DHOS (Direction de l'hospitalisation) et l'Inserm et s'appuie sur la fusion des deux systèmes d'information : l'un, à visée médico-économique, créé par la DHOS en 1988 et l'autre, clinico-épidémiologique, créé à l'Inserm en 1989. Le recueil des données concerne tous les sujets suivis dans 62 hôpitaux répartis dans 29 des 30 centres d'information et de soins de l'immunodéficience humaine (Cisih), soit la très grande majorité des patients infectés par le VIH. Les données clinico-épidémiologiques du 30^e Cisih (Bordeaux) font l'objet, depuis le début, d'une base de données spécifique (GECSA).

Le DMI2 démontre que, moyennant les précautions nécessaires organisationnelles et informatiques, de tels systèmes peuvent parfaitement être développés en France y compris dans des cas aussi sensibles que celui de l'infection à VIH. En effet, le projet a reçu un avis favorable de la Cnil dès le 27 novembre 1991 et en pratique l'immense majorité des patients donne son consentement éclairé par écrit à sa présence dans cette base.

Cette base de données est considérable (96 099 personnes au 1^{er} décembre 2004) et montre que la France peut mener des projets de recherche épidémiologique de très grande ampleur en s'appuyant sur des systèmes

d'information existants, avec l'atout de la grande taille et du coût raisonnable (par rapport à des études spécifiques construites *ex nihilo* et nécessitant à chaque fois la mise en place d'une nouvelle enquête). De nombreux résultats importants ont pu être acquis grâce à cette base. Ainsi, elle a permis récemment pour la première fois de mettre en évidence une association entre la durée d'exposition aux inhibiteurs de protéase et le risque d'infarctus du myocarde (Mary-Krause *et al.*, 2003).

3.1.4 L'épidémiologie analytique : mesurer les risques, rechercher les causalités

L'objet de l'épidémiologie analytique est d'identifier les facteurs de risque des maladies et de quantifier leur importance. Cette identification est une condition nécessaire à la recherche de méthodes de prévention. Le chercheur épidémiologiste recherche également si les relations entre facteurs de risque et maladies sont causales. Une difficulté, qui a conduit au développement d'outils et méthodes complexes, vient de ce que, les maladies étant en général plurifactorielles, il faut être capable d'identifier les facteurs « réellement » causaux et ceux qui sont des *facteurs de confusion* (en anglais *confounding factors*), c'est-à-dire des facteurs dont la corrélation avec une maladie s'explique en réalité par leur corrélation aux « vrais » facteurs causaux.

Les facteurs de risque considérés peuvent être génétiques, biologiques (hypertension artérielle, glycémie, ...), comportementaux (régime alimentaire, tabagisme, ...), environnementaux (exposition professionnelle, exposition à des polluants, ...), psychologiques ou sociaux. Ils peuvent être **individuels** ou **collectifs**. La consommation de tabac, le taux de cholestérol sanguin, sont des facteurs de risque individuels ; l'exposition aux radiations naturelles dans le lieu de résidence est un facteur de risque collectif.

La recherche des relations causales entre un facteur de risque et une maladie implique la prise en compte d'une temporalité cohérente avec l'hypothèse testée qui est discutée dans le chapitre 7 « Épidémiologie et causalité » : par exemple, l'exposition à un cancérigène ne peut se traduire en cancer que des années, voire des dizaines d'années plus tard. L'information recueillie sur l'exposition aux facteurs de risque suspectés et sur la survenue de la maladie doit donc aussi s'étaler sur des dizaines d'années de distance. Les études *transversales* sont celles dans lesquelles les informations concernant le facteur de risque et la maladie sont collectées en même temps. Elles sont donc inadaptées à l'exemple précédent, mais peuvent être utilisées lorsque les facteurs de risque étudiés sont

attachés de façon permanente à l'individu, comme par exemple les facteurs génétiques. Les études *longitudinales* sont celles dans lesquelles on recueille au cours du temps les informations concernant les facteurs de risque et les informations concernant la maladie. On peut ainsi prendre en compte la temporalité de la séquence cause-effet.

Les études épidémiologiques « écologiques »

Dans les « études écologiques », on étudie la relation éventuelle entre un facteur de risque et une maladie, en étudiant leur corrélation à partir de données agrégées sur des unités géographiques et non pas à partir de données collectées au niveau de l'individu. On a analysé ainsi les variations de mortalité par cause cardiorespiratoire dans un certain nombre de villes américaines (Samet *et al.*, 2000) ou européennes (Katsouyanni *et al.*, 1997) en fonction des taux de pollution moyens relevés dans ces sites. Les résultats des « études écologiques » sont considérés avec prudence par les épidémiologistes du point de vue de leur capacité à découvrir des relations causales. En effet, deux paramètres peuvent être corrélés géographiquement sans l'être au niveau individuel. Les études écologiques permettent, en tout cas, de générer des hypothèses. Un exemple récent et intéressant d'hypothèse générée ainsi est celui de la relation observée entre la baisse de l'incidence de certaines pathologies infectieuses et, corrélativement, la hausse des maladies à forte composante immunitaire (asthme, diabète type-1, maladie de Crohn, sclérose en plaques) (Bach, 2002).

Un nouveau schéma « mixte » d'observation épidémiologique, notamment adapté à la prise en compte des facteurs de risque sociaux, émerge : certaines variables sont relevées au niveau de la communauté, d'autres au niveau individuel (en reprenant l'exemple précédent, la fonction respiratoire ou le niveau socioprofessionnel seront mesurés individuellement, mais la pollution atmosphérique sera mesurée au niveau d'une entité géographique). Des méthodes d'analyse statistique multiniveaux sont développées pour analyser ces données.

Études de cohorte (prospectives) et études cas-témoins (rétrospectives)

Les épidémiologistes distinguent deux grands schémas d'études dans la recherche des facteurs de risque :

- les *études de cohorte*, dans leur principe, comparent le devenir de deux groupes de sujets : l'un exposé aux facteurs de risque étudiés et l'autre non exposé. Ces études sont « prospectives » car elles suivent l'histoire

d'un sujet depuis son exposition aux facteurs de risque suspectés jusqu'au moment de sa maladie. On parle de « cohorte historique », quand le point de départ commence dans le passé (on parle aussi d'étude prospective « non concurrente ») ;

- le second schéma d'étude épidémiologique est l'*étude cas-témoins* dans laquelle on compare des cas d'une maladie à des témoins judicieusement choisis en ce qui concerne leurs expositions passées aux facteurs de risque d'intérêt. Ces études sont aussi appelées « rétrospectives » (car, sur le plan temporel, elles partent du présent (la maladie) et recherchent dans le passé l'exposition à des facteurs de risque).

Encadré 1.4

Trois exemples de cohortes françaises

Le GREA (Groupe de recherche sur l'épidémiologie de l'athérosclérose), créé à l'initiative de Daniel Schwartz, initia en 1967 la première grande cohorte française (Ducimetière *et al.*, 1981), toujours active (Empana *et al.*, 2004). Environ 7 000 sujets de la cinquantaine ont été ainsi suivis, certains pendant plus de 25 ans. Plus de 250 cardiopathies ischémiques ont été identifiées, permettant l'étude du rôle respectif des divers facteurs de risque.

PAQUID a été l'étude pionnière sur la démence chez la personne âgée (Alperovitch *et al.*, 2002) pour une présentation des nouvelles cohortes sur le sujet. Elle a été lancée en 1988 par l'unité Inserm « Épidémiologie, santé publique et développement » (Bordeaux). PAQUID a notamment montré que les « réserves » cognitives ou les stratégies cognitives liées au niveau d'éducation permettent de retarder son expression clinique.

Un exemple remarquable de cohorte hospitalière existe à l'Institut Gustave Roussy (Villejuif) : les malades irradiés thérapeutiquement depuis la création de cet institut sont suivis, ce qui permet d'évaluer le risque de second cancer en fonction de la dose reçue (évaluée dans les différents organes cibles par dosimétrie) et de fournir ainsi des données directes sur la cancérogénicité des radiations (voir, par exemple, le cas de l'irradiation après cancer du sein) (Rubino *et al.*, 2003).

Ces deux schémas de référence ont chacun leurs intérêts et inconvénients ; il serait faux de croire — comme trop souvent — que l'un (les cohortes) est forcément, toujours, supérieur à l'autre :

- les études de cohorte permettent de connaître les risques absolus de maladie des sujets soumis, ou non soumis, au facteur étudié. En revanche, les études cas-témoins permettent d'estimer seulement le risque relatif (rapport du risque de ceux qui sont soumis, au risque de ceux qui ne sont pas soumis au facteur étudié) ;
- les études de cohortes sont plus longues et plus coûteuses que les études cas-témoins : il est beaucoup plus coûteux de suivre une population de sujets exposés et non exposés pendant des dizaines d'années que de rechercher dans le passé de malades et de témoins les expositions aux facteurs de risque ;
- lorsque la maladie est rare, il est extrêmement difficile de faire une étude de cohorte (on n'observe pas assez de cas dans les deux groupes exposés et non exposés) et par conséquent l'outil de référence est l'étude cas-témoins ;
- l'étude de cohortes permet d'étudier toutes les pathologies liées à l'exposition à un facteur, alors que par définition l'étude cas-témoins, étant centrée autour d'une pathologie, permet d'étudier tous les facteurs de risque possible du passé, mesurables chez les cas et chez les témoins, et non pas seulement celui qui fonde l'hypothèse de travail, comme dans l'étude de cohorte. Cependant, le biais qui peut résulter d'une reconstitution différente du passé, selon qu'on investigate les malades ou les témoins, représente une grande limitation des études cas-témoins.

L'épidémiologie génétique

L'étude des facteurs de risque génétique est effectuée dans le cadre d'une sous-discipline identifiée de l'épidémiologie, avec laboratoires, journaux et sociétés savantes. Dans le chapitre 5 « Épidémiologie et sciences biologiques : 4 exemples », les diverses contributions, notamment celles de Alexandre Alcaïs, Jean-Laurent Casanova et Laurent Abel d'une part et de Philippe Amouyel, d'autre part, décrivent la complémentarité de l'épidémiologie génétique avec les autres approches en épidémiologie. L'épidémiologie génétique s'est appuyée longtemps sur des structures d'enquête spécifiques avec données familiales (*à minima*, à partir d'interrogatoires et de dossiers médicaux, de plus en plus à partir de prélèvements biologiques). Les mesures d'héritabilité, de pénétrance, ou la recherche de gènes impliqués, ont nécessité des méthodes statistiques spécifiques d'analyse de ces données (Ellsworth et Manolio, 1999).

L'épidémiologie génétique est actuellement renouvelée par les nouvelles possibilités expérimentales permettant la comparaison à grande échelle de l'expression de milliers de gènes à la fois entre des groupes de malades et de témoins.

Toutes les difficultés biostatistiques habituellement rencontrées sont présentes et amplifiées (par exemple l'interprétation de tests statistiques multiples). Mais, au-delà, le fait que de telles recherches s'adressent plus aux maladies polygéniques que monogéniques (identifiées facilement par les études familiales), et le fait que l'expression des gènes peut être modulée par des facteurs environnementaux, posent de nouveaux problèmes méthodologiques et fournissent notamment un nouveau champ d'opportunités pour la modélisation (par exemple, modélisation de réseaux de gènes).

L'avenir de la recherche épidémiologique, comme celui de l'ensemble de la recherche en sciences de la vie, ne peut être qu'associant les approches « génomiques ». Réciproquement, celles-ci nécessitent une approche épidémiologique, seule capable d'interpréter les masses de données recueillies. Les tailles des enquêtes nécessaires, leur nature (est-il préférable de travailler sur des populations relativement homogènes génétiquement, comme celle de l'Islande par exemple ? ou est-il préférable de travailler au contraire sur des populations hétérogènes ?) et les méthodologies d'analyse à développer sont de nouveaux défis. De très grandes enquêtes couplant épidémiologie et génomique sont en début de réalisation, ou en projet, et il faut espérer que la France saura être concurrentielle. Une des difficultés à surmonter concerne évidemment l'éthique, comme les difficultés vécues par le projet islandais précurseur (DeCode) (Abbott, 2004) le montrent.

L'épidémiologie analytique face aux nouveaux enjeux de sécurité sanitaire

La plupart des risques sanitaires ayant mobilisé l'attention depuis les années 1980 ont en commun de ne pouvoir être étudiés en pratique par les outils classiques de l'épidémiologie analytique : par exemple, on ne peut évaluer le risque éventuel des millions de Britanniques qui ont été exposés à l'agent de l'ESB en construisant une analyse de cohorte, avec des sujets exposés (difficiles à distinguer des autres) et non exposés qu'on suivrait pendant 60 ans avant de donner, peut-être, une réponse ; on ne peut non plus de façon fiable reconstituer comparativement des consommations alimentaires du lointain passé. On ne peut non plus évaluer l'efficacité d'une mesure de santé publique contre une attaque bioterroriste utilisant le virus de la variole en réalisant un essai randomisé d'intervention, etc.

Dans le cas de l'évaluation de la réalité, puis de la taille des risques liés aux très faibles doses d'agents physiques ou chimiques, on peut tenter de réaliser des études de cohorte ou cas-témoins. Mais le nombre de sujets nécessaires pour démontrer, grâce à des études de cohortes, l'existence d'effets éventuels est

en général inaccessible. Ainsi, pour démontrer l'existence d'un risque de cancer respiratoire chez les sujets exposés à 20 mSV lors d'examens diagnostiques de routine, il faudrait disposer de suivi de cohortes de 10 000 000 de sujets (Brenner *et al.*, 2003) !

Les études cas-témoins nécessitent des effectifs plus accessibles, mais leur principe nécessite de rechercher dans le (lointain) passé des expositions qui souvent ne furent pas mesurées et sont soumises au biais de souvenirs différentiels (les cas attachent plus d'importance que les témoins à une petite exposition passée).

Malgré ces difficultés, des équipes épidémiologiques tentent de réaliser des études cas-témoins ou de cohorte pour évaluer la réalité de ces risques, ou pour les quantifier. Le financement de telles études est relativement facile à obtenir, vu la demande sociale. Cependant, pour des raisons pratiques, ces études peuvent être de taille plus petite que celle que le calcul demande. On se trouve alors devant les risques statistiques classiques : risque de seconde espèce de ne pas voir un effet qui pourtant existe ; risque de première espèce de trouver un test significatif, alors qu'il n'y a pas d'effet. Ce dernier risque est de plus amplifié par le « biais de publication » selon lequel une étude « positive » (se terminant par un test statistique significatif d'association entre le facteur de risque et la maladie) a plus de chance d'être publiée (ou médiatisée) qu'une étude « négative ». Aussi, la littérature foisonne d'études épidémiologiques relatives aux faibles expositions dont les résultats souvent contradictoires provoquent un certain désarroi et/ou une certaine perplexité. Il est difficile à ce sujet de ne pas citer le déjà ancien éditorial de G. Taubes dans *Science* en 1995 « The epidemiology faces its limits » (Taubes, 1995), éditorial largement discuté dans les communautés épidémiologiques mais dont la force était... le bon sens : l'ambition de démontrer l'existence de risques relatifs très faibles est souvent déçue lorsqu'elle est mise en rapport avec le coût pharaonique des études nécessaires (vu le nombre de sujets requis) et la faible force de conviction des résultats obtenus.

De plus en plus, on a donc recours à la modélisation pour étudier les conséquences éventuelles de ces nouveaux risques sanitaires, en utilisant sa capacité pour traduire les hypothèses biologiques en conséquences épidémiologiques, ou en l'utilisant comme « planche à dessin » sur laquelle on étudie des scénarios d'action.

La recherche de la causalité

Le problème de la capacité de l'épidémiologie à démontrer des relations causales est souvent posé, non seulement par les scientifiques mais par le public en

général, à chaque fois qu'un résultat épidémiologique touche un problème sensible de société. Chez les biologistes, beaucoup pensent qu'une relation causale n'est démontrée que lorsque le mécanisme fondamental (si possible moléculaire) à sa base est démontré expérimentalement. Par réaction, face à ce qu'ils ressentent comme une dictature de la biologie, certains épidémiologistes défendent la primauté de la « plausibilité épidémiologique » sur la « plausibilité biologique » (Porta, 1999) avec l'argument historique des nombreux résultats démontrés par l'épidémiologie avant même que les outils expérimentaux biologiques nécessaires, voire les concepts, n'aient existé.

Rappelons ici seulement qu'en situation d'observation, les imputations causales ne sont pas déductives, mais inductives (Valleron, 2000). L'épidémiologie analytique ne peut donc que démontrer des associations entre facteurs de risque et maladies. Ce difficile problème de la recherche de la causalité en épidémiologie fait l'objet du chapitre 7.

3.2 Épidémiologie expérimentale

On parle d'épidémiologie expérimentale lorsque le niveau d'exposition des sujets au facteur étudié est fixé par le chercheur. Le modèle dominant de l'épidémiologie expérimentale est celui de l'essai randomisé³ (de l'anglais *random*, hasard, hérité lui-même du vieux français *randon*) dans lequel l'expérimentateur utilise le tirage au sort pour constituer les groupes à comparer. La majorité des essais randomisés sont destinés à évaluer l'effet (ou l'absence probable d'effet important) d'un nouveau traitement par rapport à un traitement de référence. On parle alors d'essais thérapeutiques randomisés. Les essais randomisés peuvent aussi concerner l'évolution de l'efficacité (ou de l'absence probable d'efficacité importante) d'une politique de prévention (un régime alimentaire à haute teneur en fibres et avec peu de graisse diminue-t-il le risque de cancer colorectal ?) ou d'un dépistage (une nouvelle technique d'imagerie appliquée au dépistage du cancer des bronches permettrait-elle d'abaisser la mortalité ?). Dans la majorité des essais randomisés, l'unité expérimentale est l'individu. Cependant, dans un nombre croissant de cas, l'unité expérimentale est le groupe de sujets. Par exemple, pour évaluer l'efficacité d'un système d'aide informatisé à la décision médicale, on tirera au sort les médecins disposant — ou non — de cet outil et on évaluera les résultats sur les groupes de patients que chacun a en charge.

On parle enfin d'expérience aussi à propos des « expériences naturelles » ou « quasi-expériences » (*quasi experimental designs*) : dans celles-ci, on juge que, bien que non attribué par l'expérimentateur, le niveau d'exposition subi par les

³Aussi appelé « essai contrôlé », par transfert de l'anglais *control*.

sujets résulte d'un processus indépendant de la relation éventuelle « exposition-maladie ». Par exemple, pour étudier l'étiologie du cancer du sein, on a comparé des Japonaises plus ou moins récemment immigrées aux États-Unis avec des Japonaises non immigrées (chez qui la fréquence du cancer du sein est beaucoup plus faible que celle des femmes nord-américaines). Ceci repose sur l'hypothèse que l'immigration est un processus indépendant des variables de mode de vie étudiées. Le risque des femmes japonaises immigrées dépendant de l'ancienneté de l'immigration (identique aux femmes nord-américaines pour les plus anciennes), ces recherches ont démontré l'importance de facteurs de style de vie dans le cancer du sein.

Majoritairement, l'épidémiologie expérimentale consiste donc en la réalisation d'« essais randomisés ». Ces essais ont pour caractéristique de comparer deux groupes (au moins) de sujets ne différant que par le facteur étudié (par exemple, le traitement). La volonté d'éviter tout biais mène en général — quand c'est techniquement possible — à réaliser ces essais en « double aveugle » (ni le sujet, ni le personnel soignant en contact avec ce sujet ne savent lequel des traitements le malade reçoit). L'organisation d'un essai thérapeutique humain nécessite un encadrement réglementaire extrêmement rigoureux, destiné d'une part à garantir l'éthique de la recherche et d'autre part à garantir la qualité technique du recueil et de l'analyse des données. Ces points sont analysés dans les sous-chapitres 9.3 « Moyens nécessaires à la recherche épidémiologique » et 9.4 « Études épidémiologiques en santé : aspects législatifs et réglementaires ». Le fait qu'il ne soit éthique de proposer un essai à un patient que si, *a priori*, ses chances sont équilibrées, quel que soit le « bras » de l'essai dans lequel il se trouve, mène nécessairement à entreprendre des essais qui ont *a priori* une faible probabilité de déboucher sur la démonstration d'une bien plus grande efficacité d'un des deux traitements que de l'autre. Cependant, la démonstration de l'absence de différence significative entre deux traitements est souvent intéressante et ne devrait d'ailleurs pas être qualifiée de « résultat négatif ». Par ailleurs, les essais démontrent aussi parfois des effets adverses imprévus qui sont de grand intérêt.

Encadré 1.5

Le traitement du risque de transmission maternofoetale du VIH

Au début de l'histoire du sida, la transmission maternofoetale du virus de l'immunodéficience humaine était très importante (de l'ordre de 25 à 30 % des enfants de mère infectée développaient, en l'absence de traitement existant, la maladie). Le seul traitement disponible était la zidovudine (AZT). Une équipe américaine (avec laquelle collaborèrent des équipes françaises) eut

le courage d'expérimenter l'efficacité du traitement de la mère par l'AZT sur le risque d'infection du nouveau-né. Cet essai thérapeutique, mené en double aveugle, montra une réduction énorme de ce risque (passant de 25 à 8 %) (Connor *et al.*, 1994). Depuis, les stratégies de traitement ont encore été améliorées et testées par de nouveaux essais thérapeutiques randomisés. Une femme séropositive a maintenant de très grandes chances de ne plus donner naissance à un enfant infecté par le virus. Notons aussi dans ce sujet le rôle de la modélisation (Rouzioux *et al.*, 1995) qui montra que l'enfant était infecté très tardivement *in utero*, guidant ainsi la thérapeutique.

Une équipe française montra les conditions dans lesquelles ce résultat pouvait être mis en pratique en Afrique, où ce mode de transmission est très important (Dabis *et al.*, 1999).

Les essais thérapeutiques représentent un examen de passage imposé par les agences réglementaires pour les nouvelles molécules. Leur organisation, très encadrée, qui doit suivre des normes internationales contraignantes, les rend extrêmement coûteux et hors de portée des équipes ne disposant pas d'une solide infrastructure biostatistique et surtout organisationnelle, spécialement dans le milieu « académique » : cependant, l'ANRS a eu la volonté et les moyens, en relation avec l'Inserm, de développer quelques centres d'essais thérapeutiques dans les domaines du VIH et des hépatites. De nombreux centres d'investigation clinique (CIC), initialement destinés aux études de type pharmacologique, essaient de développer une infrastructure minimale. Une réflexion approfondie sur la meilleure manière de conserver une initiative d'essais thérapeutiques dans le milieu de la recherche académique — compte tenu de contraintes organisationnelles et de coût — est absolument nécessaire. Ce problème est abordé dans le sous-chapitre 9.3 « Moyens nécessaires à la recherche épidémiologique ».

Le développement des essais thérapeutiques a induit une recherche biostatistique active. Ceci a concerné notamment la recherche de plans d'expérience puissants, économiques, permettant d'arriver au résultat le plus vite possible en minimisant tous les risques des hommes et femmes inclus dans ces essais, y compris le risque de constater rétrospectivement qu'ils étaient dans le mauvais « groupe ». L'omniprésence des considérations éthiques a mené par exemple au développement de plans séquentiels (Wald) et à l'utilisation de méthodes bayésiennes permettant de réduire les nombres de sujets (Couzin, 2004). Enfin, parce que les essais thérapeutiques durent plusieurs années, concernent des milliers de sujets, débouchent non seulement sur des résultats scientifiques et médicaux, mais sur des autorisations (ou refus) de mise sur le marché (donc sur des enjeux économiques majeurs), ils doivent bénéficier d'un environnement technique irréprochable et fonctionner selon des principes rigoureux d'assurance qualité.

Ceci a mené au développement d'outils informatiques avancés et très coûteux. On doit constater que si les Français développent en interne dans leurs équipes des outils informatiques de cette nature, ceux qui sont sur le marché sont tous étrangers.

3.3 La poussée de l'épidémiologie théorique et de la modélisation

Si l'on prend les très grandes revues généralistes (*Nature*, *PNAS*, *Science*) comme un témoin de l'intérêt de la communauté scientifique pour l'épidémiologie, on constate que très peu de grands travaux d'épidémiologie analytique sont publiés, mais que de nombreux travaux d'épidémiologie « mathématique » (c'est-à-dire utilisant la modélisation) y apparaissent régulièrement. Cette tendance s'explique du côté des épidémiologistes par le fait que de grands sujets d'actualité (telle l'étude de la dynamique des maladies infectieuses) nécessitent cette approche et par la culture mathématique des médecins et des biologistes, bien meilleure que dans le passé ; du côté des mathématiciens, la tendance est aidée par le fait que l'épidémiologie pose des problèmes intéressants, notamment en calcul des probabilités, et dans l'étude des systèmes complexes (notamment réseaux « petit monde » et réseaux « sans échelle ») et par l'intérêt nouveau général des mathématiciens pour la médecine et la biologie.

Les travaux d'épidémiologie théorique, comme ceux d'épidémiologie observationnelle, relèvent de deux catégories : ceux qui ont un but cognitif, et ceux qui sont destinés à aider la décision en santé publique.

Dans la première catégorie (travaux à visée cognitive), on trouve notamment ceux qui visent à identifier les lois gouvernant la dynamique des épidémies ou à estimer des paramètres de l'histoire naturelle des maladies cachées à l'observation directe (par exemple, la taille future d'une épidémie, la durée de l'incubation d'une maladie infectieuse). L'épidémie de sida a bien illustré dans la période récente pourquoi la modélisation mathématique était indispensable en épidémiologie. C'est la modélisation, et elle seule, qui, en interprétant les premières données recueillies à partir de malades contaminés au cours d'une transfusion datée, a fait comprendre que la caractéristique de cette maladie était une très longue durée d'incubation qui expliquait pourquoi seule une minorité des sujets séropositifs avaient développé la maladie clinique (détruisant ainsi l'idée de « porteur sain » alors correcte politiquement, mais dramatique par la perception du risque qu'elle construisait notamment chez le décideur politique).

Au cours des années récentes, la modélisation en épidémiologie s'est également développée : ainsi, grâce à de nouvelles catégories d'informations, il est

possible de tenir compte des structures de contact d'individus entre eux dans des petites ou grandes communautés (Eubank *et al.*, 2004). On doit noter également les travaux très prometteurs d'épidémiologie des maladies infectieuses dans lesquels les variations de la variabilité des séquences de virus observées dans des populations sont modélisées : par exemple pour dater le début d'une épidémie d'hépatite virale C (Pybus *et al.*, 2001) ou pour comprendre les variations annuelles du virus de la grippe (Ferguson *et al.*, 2003).

La seconde catégorie de travaux a pour objet principal d'aider à des décisions de santé publique en l'absence de données observables. Les deux catégories d'applications les plus importantes sont celles qui concernent la prévision de situations épidémiologiques futures et la planification de méthodes de lutte appropriée, et celles qui concernent l'évaluation des risques environnementaux très faibles, en deçà des capacités d'observation.

Dans le domaine des maladies infectieuses, on peut signaler :

- les travaux de modélisation destinés à apprécier la menace que représenterait dans une société moderne la diffusion malveillante d'agents du passé (variole, peste), et à comparer des stratégies d'action en cas d'attaque bioterroriste, par exemple avec le virus de la variole ;
- les travaux qui évaluent l'impact des mouvements de population (notamment par transport aérien), et corrélativement l'impact qu'aurait une restriction de ces mouvements, sur la diffusion d'une épidémie émergente ;
- les travaux de prévision de l'extension possible des maladies émergentes (Sras, maladies à prions, etc.) ;
- l'impact possible du réchauffement global sur l'extension de maladies transmissibles (fièvres hémorragiques, paludisme, choléra).

Dans tous les exemples, la modélisation permet de rassembler les connaissances disponibles dans un modèle même afin que les « décideurs » sachent se préparer à une épidémie d'ampleur, ou au contraire de taille limitée, et puissent comparer des stratégies de lutte sur une sorte de planche à dessin que leur procure la modélisation. La modélisation sert aussi à aider le décideur à évaluer l'efficacité d'interventions dans d'autres domaines que celui des maladies infectieuses : par exemple, on a ainsi montré que le recours aux diverses technologies d'aide à la reproduction n'était pas capable de compenser la baisse de fertilité des femmes après 35 ans (Léridon, 2004). C'est également la modélisation qui a chiffré les risques résiduels associés à la transfusion, permettant de piloter des choix de méthodes de réduction de ces risques.

Pour estimer les risques liés aux faibles doses, la modélisation est à nouveau le seul recours possible alors que l'observation épidémiologique ne permet pas de répondre. Par exemple, le risque éventuel de cancer aux très faibles doses, telles celles reçues au cours des examens radiologiques diagnostiques, a-t-il été estimé par recours à des modèles mathématiques. Encore faut-il que les hypothèses émises par les modélisateurs tiennent compte des progrès des connaissances de la biologie.

3.4 Les « niveaux de preuve » en épidémiologie

Quelle est la force de preuve de l'épidémiologie ? Bien évidemment, l'épidémiologie d'observation a la faiblesse d'être inductive, et non déductive (mais l'astronomie aussi, pour une large part). L'épidémiologie expérimentale apporte la démonstration de supériorité d'un traitement sur un autre, alors que l'épidémiologie d'observation ne l'apporte pas.

Parce que l'épidémiologie est devenue la science de référence de la santé publique, et le fondement de la « médecine fondée sur la preuve scientifique », les agences de santé publique, internationales et françaises, ont mis au point des « échelles de niveaux de preuve » dans lesquelles elles proposent une hiérarchie des études épidémiologiques au sommet desquelles on trouve l'approche expérimentale, puis les études de cohorte, puis les études cas-témoins, enfin les études épidémiologiques descriptives⁴ et les opinions d'experts. L'épidémiologie théorique n'est pas présente dans ces « échelles ».

De telles classifications sont utiles aux profanes pour interpréter l'information produite par les travaux scientifiques **visant à évaluer l'efficacité de traitements**. Malheureusement, elles sont utilisées trop souvent en dehors de cette indication première et appliquées à tous les domaines de recherche épidémiologique sans discernement. Au sein de l'épidémiologie analytique, les études de cohortes ne sont pas nécessairement « meilleures » que les études cas-témoins : de nombreux résultats majeurs ont été obtenus grâce à des études cas-témoins (l'exemple le plus frappant étant celui de la relation tabac-cancer), mais la recherche de facteurs étiologiques des maladies rares implique l'usage d'études cas-témoins et non de cohortes. Finalement, un inconvénient majeur de ces classifications est d'ancrer l'idée que l'épidémiologie descriptive, la narration des faits, est nécessairement en bas de l'échelle de qualité. C'est pourtant sur l'observation des faits et leur narration qu'est bâtie la science.

⁴Un exemple de telles classifications peut être trouvé sur le site de la Haute autorité de santé (<http://www.anaes.fr>).

Il faut donc veiller à ce que la notion de « niveau de preuve » soit prioritairement utilisée dans le domaine de l'évaluation des traitements et ne le soit pas aveuglément en milieu académique, menant à financer prioritairement seulement les études des niveaux de preuve jugés (faussement) élevés (l'épidémiologie expérimentale) alors que l'épidémiologie a aussi besoin de systèmes d'observation rigoureux et de qualité, d'études cas-témoins bien menées, notamment sur des maladies peu fréquentes pour lesquelles on ne dispose pas d'hypothèses étiologiques, et d'études transversales ou écologiques bien menées.

En bref, le meilleur classement qu'on puisse fournir des études épidémiologiques est celui qui considère l'originalité des hypothèses, la qualité des méthodes utilisées et la compétence des chercheurs et non pas — de façon aveugle — la place dans une classification arbitraire.

3.5 Trois autres césures de l'épidémiologie

En sus de la classification précédente de l'activité épidémiologique (observation, expérimentation, théorie), trois distinctions méritent d'être rappelées.

La distinction entre « épidémiologie explicative et épidémiologie pragmatique »

En 1967, apparut la distinction à faire entre les attitudes explicatives et pragmatiques selon la finalité des essais thérapeutiques (Schwartz et Lellouch, 1967). On montra que, de plus, cette distinction avait des conséquences sur la planification, l'analyse et l'interprétation des essais thérapeutiques. Elle est plus générale et s'applique à l'ensemble de l'épidémiologie. En effet la recherche épidémiologique poursuit deux buts différents et non contradictoires : le premier est un but de *recherche* : il s'agit de comprendre l'étiologie d'une maladie, son histoire naturelle, ses facteurs pronostiques. Le second est un but d'action : il s'agit de fournir des bases scientifiques aux choix de politique de santé publique, et même d'aider à déterminer la prise en charge optimale d'un patient donné. Cette distinction entre épidémiologies « explicative » et « pragmatique » (Schwartz, 1994) est bien éclairée lorsqu'on considère la recherche des facteurs de risque d'une maladie : dans une démarche « explicative », on ne peut se contenter de démontrer qu'un facteur donné est un facteur de risque (c'est-à-dire que les sujets présentant ce facteur ont un risque accru de maladie). On veut déterminer si le facteur est causal et comprendre le mécanisme sous-jacent. Dans une démarche « pragmatique », savoir qu'un facteur est un facteur de risque peut suffire à mettre en œuvre les actions de santé publique recherchées.

Épidémiologie populationnelle et épidémiologie clinique

L'épidémiologie a comme matériel d'étude des groupes humains. Dans le cas où le « groupe humain » considéré est un groupe d'individus de la population générale, et où la recherche vise à donner des résultats applicables en population générale, on parle d'*épidémiologie populationnelle* (en anglais : *Population Epidemiology*). Dans le cas où le « groupe humain » considéré est un groupe d'individus pris en charge pour le traitement d'une maladie, en général à l'hôpital, et où la recherche vise à optimiser la prise en charge individuelle des malades, on parle d'*épidémiologie clinique*. L'épidémiologie clinique est une discipline récente (début des années 1980) avec ses équipes de recherche, ses sociétés savantes, ses journaux spécialisés (exemple : *Journal of Clinical Epidemiology*) (cf. sous-chapitre 8.1 « L'épidémiologie clinique, base scientifique pour la prise en charge rationnelle des malades »).

Épidémiologie académique et épidémiologie d'intervention

L'épidémiologie « académique » est effectuée en milieu de recherche, en général avec des visées prioritairement cognitives. L'épidémiologie « d'intervention », un domaine en forte croissance, partage les méthodologies de l'épidémiologie académique, mais est plutôt effectuée en milieu non académique (agences sanitaires, par exemple). Elle a pour objet de résoudre des problèmes de santé publique survenant brusquement, grâce à l'utilisation de la méthode épidémiologique. Par exemple, lorsqu'un incident toucha la centrale nucléaire de Three Mile Island en 1979, le CDC (Center for Disease Control) d'Atlanta envoya immédiatement une équipe d'épidémiologistes d'intervention (les *Epidemic Intelligence Officers*, du nom du corps créé en 1950, à un moment où les armes biologiques étaient déjà redoutées, dans le contexte de la guerre de Corée). Cette équipe disposait de l'ensemble des moyens humains et matériels pour réaliser, sur le champ, une enquête épidémiologique avec dosimétrie et constitution d'un fichier prospectif de nature à renseigner ultérieurement de façon exhaustive et rigoureuse sur les éventuelles conséquences sanitaires de cet accident. Par contraste, la même organisation n'existait pas au moment de l'accident de Tchernobyl ; au contraire, avec retard, on a tenté de reconstituer des données qui, soit ne peuvent plus être connues (l'état des sujets au moment de l'accident), soit sont impossibles à recueillir suite au très important déplacement de populations. Une épidémiologie d'intervention performante, rénovée par l'utilisation de systèmes d'information en réseau attendit 1992 pour avoir une première existence institutionnalisée avec le Réseau national de santé publique ; elle a maintenant un cadre organisationnel fixé par la loi du 1^{er} juillet 1998, et des réalisations, avec l'Institut national de veille sanitaire (InVS). Le passage au niveau européen, inévitable pour la bonne raison que les épidémies

ne reconnaissent pas les frontières, est en cours. Bien sûr, toute classification est arbitraire et celle-ci n'échappe pas à la règle : des épidémiologistes « académiques » effectuent des recherches visant à développer des outils, et à obtenir des résultats, utiles pour l'épidémiologie d'intervention. Certaines des activités menées par les épidémiologistes d'intervention conduisent à des publications scientifiques. Cependant, la distinction entre les deux activités mérite d'être rappelée, à la fois pour rappeler la nécessité de développer le métier d'épidémiologiste d'intervention, lequel nécessite des qualités méthodologiques parfaites, et pour souligner que celui-ci ne doit pas être jugé à l'aune de la publication scientifique, mais à celle des résultats rendus à la santé publique.

Encadré 1.6

Épidémiologie d'intervention : deux exemples (J.C. Desenclos, InVS)

Fin 1999, le Centre national de référence des listerias qui centralise et type les souches humaines de *Listeria* détecta l'émergence d'un « sous-type » inhabituel réparti sur l'ensemble du territoire et alerta l'InVS. L'enquête épidémiologique menée sur les cinq premiers patients indiqua qu'ils avaient consommé des rillettes achetées dans différents magasins de la même enseigne. Un rappel des rillettes concernées et une information du public purent être organisés par le ministère de la Santé sur la base de ces faits. Des souches de *Listeria* génotypiquement identiques à celles des patients furent retrouvées dans ces rillettes, confirmant les résultats épidémiologiques (de Valk *et al.*, 2001). Une autre épidémie liée à un autre sous-type fut détectée début 2000. La consommation alimentaire des cas y étant moins démonstrative, une enquête cas-témoins fut effectuée en urgence et montra un excès de risque élevé chez les consommateurs de langue de porc en gelée (risque estimé plus de 75 fois supérieur (intervalle de confiance : 4,7 à 216,0) à celui de non-consommateurs semblables). Le public fut alors informé afin d'éviter la consommation de ce mets. La souche ne fut cependant pas retrouvée dans cette charcuterie qui fut à l'origine de 32 cas dont 9 décès (de Valk *et al.*, 2001).

Conclusion

L'épidémiologie a une longue histoire dans laquelle les Français, d'abord des mathématiciens au XVIII^e siècle, puis des médecins à l'avant-garde de leur

génération au XIX^e siècle ont eu une part importante, plus connue à l'étranger qu'en France.

La montée en puissance de l'épidémiologie au XX^e siècle, jusqu'à la période actuelle où elle est la science de référence pour les politiques de santé publique, s'est appuyée d'abord sur la statistique, qui a fourni les outils conceptuels (de la planification des études, avec la randomisation, jusqu'à l'analyse des résultats qui présente de nombreuses spécificités). Plus récemment, notamment suite à la résurgence de l'intérêt pour les maladies infectieuses, la modélisation a pris un rôle important.

Les résultats acquis par l'épidémiologie, depuis le début de son histoire jusqu'à maintenant, ont souvent ouvert de nouveaux champs de recherche biomédicale en identifiant des maladies jusque-là inconnues, ou des relations causales, ou probablement causales, entre des facteurs de risque et des maladies. L'épidémiologie et la biologie de base sont complémentaires. Ainsi, la compréhension des relations gènes-environnement, qui motive l'ensemble de la recherche biomédicale, ne peut se faire sans une approche populationnelle (donc épidémiologique).

L'épidémiologie, science globale des facteurs de risque des maladies, défie en rappelant sans cesse que pour comprendre les causes des maladies, pour trouver de nouveaux traitements efficaces, il y a nécessité d'approches intégrées, y compris incluant les sciences humaines et de l'environnement à côté de la biologie de base, quelles que soient les avancées spectaculaires de cette dernière. Mais l'épidémiologie est aussi la science en amont de la santé publique. Elle ne peut vivre indépendamment des priorités de santé publique, donc de santé du public. Celles-ci donnent une place de plus en plus grande à la sécurité sanitaire, en particulier dans les pays développés. À tort, ou à raison, le public est facilement préoccupé par l'impact possible des facteurs environnementaux sur la santé et est assez peu sensible, peut-être par manque d'information, aux progrès sanitaires évidents traduits par des indices aussi simples que l'augmentation de l'espérance de vie, avec et sans incapacités. Le public a besoin d'une information épidémiologique complète et rapide.

Au total, on peut penser que le rôle de l'épidémiologie va s'accroître car elle est indispensable dans la production des connaissances en recherche médicale, et dans l'aide à la décision en santé publique. La condition de son développement sera l'accroissement de son interaction avec les autres disciplines, notamment les mathématiques de plus en plus indispensables, compte tenu de l'augmentation de la taille des informations à traiter et du besoin croissant de théorisation, dont l'outil est la modélisation.

Bibliographie

- Abbott A. (2004). Icelandic database shelved as court judges privacy in peril. *Nature*, **429** (6988) : 118.
- Alpérovitch A. *et al.* (2002). Les études épidémiologiques sur le vieillissement en France : de l'étude Paquid à l'étude des Trois Cités. *C. R. Biologies*, **325** (6) : 665-72.
- Bach JF. (2002). The effect of infections on susceptibility to autoimmune and allergic diseases. *N Engl J Med*, **347** (12) : 911-20.
- Bollet AJ. (1973). Pierre Louis : the numerical method and the foundation of quantitative medicine. *Am J Med Sci*, **266** (2) : 92-101.
- Brenner DJ. *et al.* (2003). Cancer risks attributable to low doses of ionizing radiation : assessing what we really know. *Proc Natl Acad Sci, USA*, **100** (24) : 13761-6.
- Burkitt D. (1962). A children's cancer dependent on climatic factors. *Nature*, **194** : 232-4.
- Burkitt D. (1962). A « tumour safari » in East and Central Africa. *Br J Cancer*, **16** : 379-86.
- Connor EM. *et al.* (1994). Reduction of maternal-infant transmission of human immunodeficiency virus type 1 with zidovudine treatment. Pediatric AIDS Clinical Trials Group Protocol 076 Study Group. *N Engl J Med*, **331** (18) : 1173-80.
- Couzin J. (2004). The new math of clinical trials. *Science*, **303** (5659) : 784-6.
- Dabis F. *et al.* (1999). 6-month efficacy, tolerance, and acceptability of a short regimen of oral zidovudine to reduce vertical transmission of HIV in breast-fed children in Cote d'Ivoire and Burkina Faso : a double-blind placebo-controlled multicentre trial. DITRAME Study Group. Diminution de la TRANsmiSsion Mère-Enfant. *Lancet*, **353** (9155) : 786-92.
- Dietz K., Heesterbeek JA. (2000). Bernoulli was ahead of modern epidemiology. *Nature*, **408** (6812) : 513-4.
- Doll R., Hill AB. (1954). The mortality of doctors in relation to their smoking habits ; a preliminary report. *Br Med J*, **4877** : 1451-5.
- Ducimetière P. *et al.* (1981). *Les cardiopathies ischémiques : incidence et facteurs de risque : l'étude prospective parisienne*. Paris, Inserm.
- Ellsworth DL., Manolio TA. (1999). The Emerging Importance of Genetics in Epidemiologic Research III. Bioinformatics and statistical genetic methods. *Ann Epidemiol*, (4) : 207-24.

- Empana JP. *et al.* (2004). Sagittal Abdominal Diameter and Risk of Sudden Death in Asymptomatic Middle-Aged Men. The Paris Prospective Study I. *Circulation*, **110** (18) : 2781-5.
- Eubank S. *et al.* (2004). Modelling disease outbreaks in realistic urban social networks. *Nature*, **429** (6988) : 180-4.
- Farr W. (1838). On prognosis. *British Medical Almanach Suppl*, pp. 199-216.
- Ferguson NM., Galvani AP., Bush RM. (2003). Ecological and immunological determinants of influenza evolution. *Nature*, **422** (6930) : 428-33.
- Graitcer PL., Thacker SB. (1986). The French connection. *Am J Public Health*, **76** (11) : 1285-6.
- Hammond EC. (1966). Smoking in relation to the death rates of one million men and women. *Natl Cancer Inst Monogr*, **19** : 127-204.
- Katsouyanni K. *et al.* (1997). Short-term effects of ambient sulphur dioxide and particulate matter on mortality in 12 European cities : results from time series data from the APHEA project. Air Pollution and Health : a European Approach. *Bmj*, **314** (7095) : 1658-63.
- Léridon H. (2004). Can assisted reproduction technology compensate for the natural decline in fertility with age ? A model assessment. *Human Reproduction*, **19**.
- McCormick A. (1987). French lessons on surveillance of communicable diseases. *Br Med J (Clin Res Ed)*, **294** (6564) : 74-5.
- Mary-Krause M. *et al.* (2003). Increased risk of myocardial infarction with duration of protease inhibitor therapy in HIV-infected men. *AIDS*, **17** (17) : 2479-2486.
- Porta M. (1999). Biologic plausibility in causal inference : current method and practice. *Am J Epidemiol*, **150** (2) : 217-9.
- Prout TE. *et al.* (1972). The UGDP controversy. Clinical trials versus clinical impressions. *Diabetes*, **21** (10) : 1035-40.
- Pybus O.G. *et al.* (2001). The epidemic behavior of the hepatitis C virus. *Science*, **292** (5525) : 2323-5.
- Rouzioux C. *et al.* (1995). Estimated timing of mother-to-child human immunodeficiency virus type 1 (HIV-1) transmission by use of a Markov model. The HIV Infection in Newborns French Collaborative Study Group. *Am J Epidemiol*, **142** (12) : 1330-7.

- Rubino C. *et al.* (2003). Radiation dose, chemotherapy, hormonal treatment and risk of second cancer after breast cancer treatment. *Br J Cancer*, **89** (5) : 840-6.
- Samet JM. *et al.* (2000). Fine particulate air pollution and mortality in 20 U.S. cities, 1987-1994. *N Engl J Med*, **343** (24) : 1742-9.
- Schwartz D. *et al.* (1961). Results of a French survey on the role of tobacco, particularly inhalation, in different cancer sites. *J Natl Cancer Inst*, **26** : 1085-108.
- Schwartz D., Lellouch J. (1967). Explanatory and pragmatic attitudes in therapeutic trials. *J Chronic Dis*, **20** : 637-648.
- Schwartz D. (1994). Connaissance ou décision : attitude explicative ou attitude pragmatique. *Journal de la Société Statistique de Paris*, **135** (2) : 1-18.
- Taubes G. (1995). Epidemiology faces its limits. *Science*, **269** (5221) : 164-9.
- Tilstone C. (2003). Vital statistics. *Nature*, **424** : 610-612.
- de Valk H. *et al.* (2001). Two consecutive nationwide outbreaks of Listeriosis in France, October 1999-February 2000. *Am J Epidemiol*, **154** (10) : 944-50.
- Valleron AJ. *et al.* (1986). A computer network for the surveillance of communicable diseases : the French experiment. *Am J Public Health*, **76** (11) : 1289-92.
- Valleron AJ. (2000). Les rôles de la modélisation en épidémiologie. *C. R. Acad. Sci. III*, **323** (5) : 429-33.
- Valleron AJ. (2000). Mise en évidence des faits et recherche des causes en épidémiologie environnementale : enjeux méthodologiques. *C. R. Acad. Sci. III*, **323** (7) : 617-28.

CHAPITRE 2

État des forces épidémiologiques en France

SOUS-CHAPITRE 2.1

L'épidémiologie humaine

MARCEL GOLDBERG ET ROGER SALAMON

1 | La formation, les métiers, les moyens

1.1 Qu'est-ce qu'un épidémiologiste ?

Il est certainement utile de commencer ce chapitre en décrivant ce qu'est un épidémiologiste professionnel, comment il est formé, de quelles compétences et de quel environnement il doit disposer, tant cette discipline reste largement méconnue en France aujourd'hui, et tant reste répandue l'illusion qu'il suffit d'un peu de bonne volonté et d'une brève formation complémentaire à un médecin ou un scientifique d'une autre discipline pour se transformer en épidémiologiste.

L'épidémiologie est une spécialisation, intervenant après une première formation (médecine ou autres professions de santé, statistique, biologie, sciences sociales, ...). Elle nécessite un master 1 du LMD (ancienne maîtrise), obtenu au bout de deux ans, suivi d'un master 2 (ancien DEA), obtenu en un an, et le cas échéant une thèse d'université (deux ou trois ans).

1.2 Les autres métiers

Les épidémiologistes sont souvent amenés, selon leurs thèmes de recherche, à coopérer avec des spécialistes d'autres disciplines scientifiques, relevant de la biologie, de la clinique, de la modélisation, de l'informatique et de la statistique, ou des sciences humaines et sociales.

Mais, outre les collaborations interdisciplinaires entre groupes de chercheurs, souvent indispensables, la conduite des études épidémiologiques requiert qu'au sein même des équipes d'épidémiologie soient réunis des techniciens et ingénieurs dont les spécialités peuvent beaucoup varier : les compétences spécifiques dont l'épidémiologiste a besoin de s'entourer au quotidien sont en effet très différentes selon qu'il s'agit d'épidémiologie moléculaire, d'épidémiologie sociale, de modélisation d'épidémies, ou de surveillance de la mortalité.

On trouvera donc associés dans les équipes d'épidémiologie, selon leurs orientations scientifiques, des personnels de métiers aussi divers que des informaticiens, des techniciens d'enquête ou de recherche clinique (de formation paramédicale, en sociologie ou psychologie, etc.), des codeurs spécialisés (nosologues, nomenclaturistes), des hygiénistes industriels, etc. On se reportera au chapitre 9.1 « La formation en épidémiologie à l'université » pour une description plus complète des formations existant en France à l'Université.

1.3 Les conditions de la réalisation d'études épidémiologiques

Les conditions concrètes de la réalisation d'études épidémiologiques sont également largement méconnues et sont trop souvent assimilées à la production de statistiques courantes à partir de recueils de données de routine. Toute étude épidémiologique, si « petite » soit-elle, doit être réalisée en suivant un protocole rigoureux, qui doit être préparé dans tous les cas préalablement à sa mise en œuvre, de façon spécifique aux questions auxquelles il s'agit d'apporter une réponse¹ : les hypothèses doivent être clairement établies en s'appuyant sur une bibliographie complète, la population étudiée doit être définie de façon précise (critères d'inclusion et d'exclusion compris), les données à recueillir doivent être rigoureusement décrites de façon exhaustive et standardisée, les méthodes et conditions du recueil spécifiées dans les détails, des procédures de contrôle de qualité, de validation des données et de monitoring de l'étude doivent être mises en place, les analyses statistiques requièrent souvent la mise en œuvre de méthodes sophistiquées parfois spécifiques des analyses épidémiologiques (la

¹On peut se reporter à : Adelf, AEEMA, Aderest, Epiter. Recommandations — Déontologie et bonnes pratiques en épidémiologie. *Rev Épidém et Santé Publ*, 1999, 47 : 1S1-1S32.

« biostatistique » est devenue une véritable branche de la statistique) qu'il faut savoir choisir et utiliser à bon escient (cf. chapitre 4 « Épidémiologie et sciences mathématiques »). Bref, non seulement la réalisation d'études épidémiologiques est affaire de professionnels spécialisés, comme on l'a rappelé, mais elle nécessite beaucoup de temps de spécialistes divers (épidémiologistes bien sûr, mais aussi biostatisticiens, techniciens d'enquête, codeurs, etc.).

Sur le plan pratique, une « petite » recherche épidémiologique s'étale sur au moins deux ou trois ans, et peut mobiliser facilement une dizaine de personnes-années de travail. Des projets plus importants, comme par exemple des cohortes longitudinales de longue durée, ou de grandes études cas-témoins en population, peuvent mobiliser en permanence une (voire plusieurs) dizaine(s) de personnes, uniquement pour en assurer le bon déroulement au quotidien, avant même toute analyse épidémiologique proprement dite.

Le coût d'une étude épidémiologique peut être extrêmement variable, non seulement évidemment en fonction de l'effectif des sujets inclus et de la durée du suivi pour les études longitudinales, mais aussi selon les investigations menées auprès des sujets, ainsi que selon les données qui peuvent éventuellement être disponibles par ailleurs. Si on ne prend pas en compte le coût des examens et des analyses biologiques ou génétiques (IRM, dosage de métabolites sanguins, puces à ADN, etc.), qui sont spécifiques de certaines recherches et qu'on ne peut facilement généraliser, l'essentiel des coûts d'une étude épidémiologique concerne le recueil de données relatives aux sujets inclus. L'encadré 1.3 présente quelques exemples pour certaines études, et on trouvera une analyse plus détaillée des moyens nécessaires à la recherche épidémiologique au chapitre 9.3 « Les moyens nécessaires à la recherche en épidémiologie ».

Encadré 2.1-1

Coût des études épidémiologique

Études cas-témoins

Le coût d'une étude cas-témoins en population générale est estimé au minimum à 230 euros par sujet inclus. À titre d'exemple, dans l'étude Icare, qui porte sur les expositions aux facteurs de risques professionnels des cancers du poumon et des voies aérodigestives supérieures et qui inclut 9 000 sujets (6 000 cas et 3 000 témoins), la structure des coûts concerne essentiellement les salaires (épidémiologistes, enquêteurs, personnels réalisant le codage, la saisie et le contrôle de qualité des données, statisticiens, secrétariat) et les frais de fonctionnement (déplacements des enquêteurs, missions de coordination, réunions régulières d'enquêteurs, frais généraux). Ce sont les salaires qui représentent la part de loin la plus importante du coût des études (environ

90 %), notamment les salaires d'enquêteurs, car dans une étude en population générale, il faut compter environ une journée de travail d'enquêteur pour l'inclusion d'un sujet (recherche du sujet, contact pour rendez-vous, déplacement, interrogatoire, remplissage des questionnaires et autres recueils de données). Les **études cas-témoins en milieu hospitalier** sont habituellement moins coûteuses, car les patients peuvent être enquêtés pendant leur séjour à l'hôpital, ce qui réduit sensiblement certains frais.

Cohortes prospectives

On peut citer quelques exemples qui donnent des éléments de coût.

La cohorte Gazel permet le suivi de plus de 20 000 agents EDF-GDF depuis maintenant plus de 15 ans à partir de différentes sources déjà existantes (fichiers du personnel, mutuelles, services médicaux de l'entreprise qui ont construit des systèmes d'information épidémiologique, incluant un registre des cancers et un registre des pathologies cardiaques ischémiques) et par un auto-questionnaire postal annuel. **Le coût total annuel par sujet** (hors personnel permanent) **est de 25 euros environ**, constitué essentiellement des salaires des gestionnaires de données, des codeurs et du personnel de saisie, des dépenses d'impression, des frais postaux, de saisie, de codage, et d'informatique. Gazel a été conçue comme un laboratoire épidémiologique ouvert à la communauté scientifique et plus de 30 études sont actuellement en cours dans la cohorte ; elles concernent en moyenne 10 000 sujets par étude, soit l'équivalent d'au moins 300 000 sujets inclus, pour lesquels les coûts peuvent être très faibles, car de nombreuses données sont déjà recueillies et la logistique de suivi est prise en charge par le fonctionnement courant de la cohorte.

La cohorte Suvimax a assuré le suivi de 15 000 sujets en population générale pendant 8 ans. Le coût total (hors personnel permanent) a été estimé à 10 millions d'euros. **Le coût annuel par sujet inclus est de 83 euros environ**. Dans **l'Enquête santé et protection sociale** (Irdes-Cnamts), le recueil de données pour 10 000 sujets, sous-traité à une société de service, coûte 1,2 million d'euros par an (hors personnel permanent). **Le coût annuel par sujet inclus est de 120 euros environ**.

La cohorte Eden inclut 2 000 enfants suivis pendant 5 ans, soit 10 000 enfants/années. Le coût total (hors personnel permanent) est estimé à 2,5 millions d'euros, soit **250 euros/enfant/an**.

En Angleterre, **la cohorte Whitehall II** suit 10 000 fonctionnaires depuis 1985 ; son budget annuel (incluant les salaires de toute l'équipe, y compris celui de l'investigateur principal) est 1,8 millions de livres (soit environ 2,6 millions d'euros par an). **Le coût annuel par sujet inclus est de 260 euros environ**.

On voit que le coût du suivi d'une cohorte prospective peut varier d'un facteur de 1 à 10 selon le contexte. Ceci s'explique par diverses raisons : possibilité ou pas d'utiliser des dispositifs de recueil de données déjà en place (Gazel bénéficie de données provenant des services médicaux d'EDF-GDF, déjà validées par les épidémiologistes de l'entreprise, des fichiers du personnel incluant l'adresse postale à jour des sujets, des caisses de sécurité sociale, de l'infrastructure des centres d'examen de santé de la Sécurité sociale, etc.) ; type de données recueillies ; salaires des personnels déjà pris ou non en charge par les organismes de recherche.

Au total, si on intègre les salaires des membres permanents des équipes, **les coûts essentiels des études de cohorte prospective sont liés au recueil des données** : quand il est possible d'utiliser des sources de données validées préexistantes (cas de Gazel), le coût est très sensiblement inférieur à celui des suivis de cohortes qui doivent assurer intégralement l'ensemble des procédures de recueil et de suivi des sujets.

2 | Grands domaines et structures d'exercice de l'épidémiologie

À la fois discipline scientifique à visée cognitive et science appliquée destinée à aider à la décision clinique et de santé publique, l'épidémiologie s'exerce dans des contextes institutionnels très diversifiés. De façon schématique, on peut regrouper en trois grands domaines les contextes dans lesquels on rencontre des épidémiologistes « spécialistes » : la recherche, la surveillance épidémiologique, et les études diverses à visée essentiellement d'aide à la décision, publique ou privée.

2.1 La recherche

2.1.1 Structuration générale

Dans tous les pays où la recherche épidémiologique est développée, on observe une organisation des structures de recherche selon deux axes non exclusifs :

- un axe disciplinaire, où les laboratoires de recherche sont rattachés à des structures « généralistes » centrées sur l'épidémiologie, sous des dénominations diverses (écoles, instituts ou départements universitaires d'épidémiologie, de santé publique, . . .) ;

- un axe orienté vers les domaines d'application, cherchant à répondre à des besoins de connaissance qui concernent des objectifs descriptifs, étiologiques, d'évaluation ou de prévision dans les domaines concernés : maladies (cancer, maladies infectieuses, maladies cardiovasculaires, . . .), populations spécifiques (personnes âgées, mère et enfant, . . .), ensembles de facteurs de risque (environnement, risques professionnels, nutrition, . . .).

Ces axes sont complémentaires, et on rencontre à la fois des grandes structures d'enseignement et de recherche généralistes en épidémiologie, biostatistique et disciplines connexes, et des laboratoires spécialisés dans un domaine, qui sont soit insérés dans ces structures généralistes, soit rattachés à des organismes spécialisés (instituts du cancer, du vieillissement, . . .). Dans les grands pays scientifiques qui ont développé une recherche épidémiologique active, il n'existe pas d'organisme spécialisé dans un domaine de la santé qui ne possède sa division ou son département d'épidémiologie.

La recherche méthodologique couvre des aspects diversifiés : amélioration des protocoles d'étude, des méthodes de mesure, d'analyse statistique, de modélisation prévisionnelle. Elle est presque toujours intégrée au sein des structures d'épidémiologie généralistes ou spécialisées par problème : dans aucun pays, on ne rencontre de laboratoires qui soient situés au sein de facultés de sciences se consacrant spécifiquement au développement de recherches en méthodes épidémiologiques : c'est à l'intérieur de facultés de médecine, d'écoles de santé publique, d'instituts de recherche en épidémiologie ou consacrés à une pathologie (instituts du cancer, instituts de santé au travail ou environnementale, . . .) que se trouvent les équipes de chercheurs mathématiciens, statisticiens, ou informaticiens.

2.1.2 Organisation institutionnelle

En France, les structures de recherche épidémiologique ne sont qu'en partie organisées de façon similaire à ce qu'on observe dans les pays ayant une recherche épidémiologique forte. En effet, si une grande partie des laboratoires de recherche épidémiologique est orientée selon des problèmes de santé (une pathologie, un groupe de population, une famille de facteurs de risque pour la santé), ceux-ci ne sont habituellement intégrés, ni à de grandes structures d'enseignement et de recherche épidémiologique (qui n'existent pratiquement pas dans notre pays : cf. plus loin), ni à de grandes structures spécialisées dans un domaine (qui n'ont presque jamais, sauf dans le domaine du cancer, développé en leur sein de laboratoires d'épidémiologie²). Bien que pour la plupart

²Encore faut-il préciser que, même dans ce cas, il s'agit presque exclusivement de structures de gestion d'essais cliniques, et que l'essentiel de l'épidémiologie descriptive, étiologique ou évaluative se fait hors de ces organismes (unités de l'Inserm, registres des cancers).

hébergées physiquement dans des structures médicales (facultés de médecine, hôpitaux), elles n'ont généralement pas de lien direct avec l'activité scientifique de celles-ci. C'est par leur rattachement disciplinaire — commission scientifique spécialisée de l'Inserm ; sous-section spécialisée du CNU — qui permet de faire masse, malgré des effectifs restreints (*cf.* plus loin), et qui autorise une évaluation comparative nationale, que ces laboratoires ont pu développer une recherche de qualité.

Comme on le détaillera plus loin (*cf.* *Aspects quantitatifs*), la recherche épidémiologique est particulièrement concentrée au sein de l'Inserm, où historiquement elle a pris son développement. Ces toutes dernières années, un rapprochement s'est opéré avec les universités, dans le cadre notamment de la « mixité » qui s'est établie entre formations Inserm et établissements universitaires ; cette mixité aujourd'hui systématique (il n'existe pratiquement plus d'unités « propres » Inserm) a été voulue par les organismes concernés, mais elle n'a pas encore, pour l'essentiel, permis de développer des moyens conséquents en épidémiologie de la part des universités. Quelques-unes font cependant exception ; ainsi les universités Paris VI, Paris VII et Paris XI, Bordeaux II, Toulouse ou Nancy ont-elles consacré des moyens non négligeables à leurs unités mixtes sous des formes diverses et variables d'un établissement à l'autre : mise à disposition de locaux, affectation d'enseignants et de personnel IATOS. Cette situation s'explique sans doute par le fait que l'épidémiologie est considérée, sur le plan de l'organisation universitaire, comme une discipline « médicale » et s'inscrit donc dans le cadre des UFR médicales des universités. Il existe de ce fait un assez grand nombre de laboratoires hospitalo-universitaires d'épidémiologie, mais ceux-ci sont presque toujours de taille très réduite et ont à assumer de lourdes tâches hospitalières et d'enseignement, ce qui ne laisse que très peu de place à des activités de recherche. Ce n'est que dans les quelques établissements cités ci-dessus qu'une activité notable de recherche universitaire a pu se développer, grâce à l'appui majeur de l'Inserm relayé en partie par l'université dans le cadre d'unités mixtes. En dehors des UFR médicales, la recherche épidémiologique universitaire est pratiquement inexistante.

On a également vu se développer, plus tardivement qu'à l'Inserm, des équipes d'épidémiologie dans des structures diverses (Service de santé des armées, Registres de maladie), ainsi que dans d'autres établissements de recherche (à l'exception notable du CNRS), qui disposent d'équipes spécialisées dans certains domaines : Institut Pasteur (pathologie infectieuse), Inrets (accidents de circulation), INRS (risques professionnels), IPSN (rayonnements ionisants), IRD (nutrition, santé de la mère et de l'enfant, géographie de la santé, pathologie infectieuse), Inra et Afssa (épidémiologie vétérinaire). Ces structures hors Inserm sont cependant toutes de très petite taille et isolées au sein des organismes concernés.

L'épidémiologie clinique, présente depuis longtemps dans certains centres de lutte contre le cancer, s'est progressivement étendue dans de nombreux CHU. Elle reste cependant une activité marginale de la recherche clinique, et on observe même depuis quelques années un recul des moyens de ces structures là où elles étaient présentes depuis longtemps ; ainsi, il est significatif de noter qu'aucune des sept « Cancéropôles » dont la création a été proposée n'ait inclus un laboratoire d'épidémiologie.

Le secteur privé (industrie pharmaceutique et agroalimentaire, certaines entreprises, des sociétés d'étude) emploie en France encore très peu d'épidémiologistes, et ceux-ci n'ont habituellement pas d'activité de recherche relevant de cette discipline, mais se consacrent à des études diverses, à visée essentiellement commerciale.

2.1.3 Domaines scientifiques couverts

Les domaines scientifiques couverts en France par la recherche épidémiologique peuvent être décrits par pathologies, déterminants, ou populations cibles. On citera les principaux thèmes, ainsi que le nombre de laboratoires de recherche qui y consacrent tout ou partie de leur activité³.

Par pathologies : cancers (7 formations Inserm, un laboratoire IPSN, un laboratoire INRS) ; pathologies cardiovasculaires (5 formations Inserm) ; pathologies neurologiques et psychiatriques (7 formations Inserm) ; pathologies liées au vieillissement (6 formations Inserm) ; pathologies respiratoires non cancéreuses (4 formations Inserm) ; pathologies infectieuses (8 formations Inserm, un laboratoire Institut Pasteur, environ une douzaine de chercheurs IRD) ; pathologies de la grossesse, du nouveau-né et perturbations de la fécondité (4 formations Inserm) ; maladies métaboliques (2 formations Inserm) ; déficiences incapacités, handicaps (5 formations Inserm) ; pathologies ostéo-articulaires (1 formation Inserm) ; néphropathies (1 formation Inserm) ; accidents (1 formation Inserm, 1 laboratoire Inrets).

Par déterminants : environnement physicochimique (9 formations Inserm, un laboratoire INRS, un laboratoire IPSN) ; génétique (13 formations Inserm) ; environnement psychosocial, comportements (6 formations Inserm) ; alimentation et nutrition (5 formations Inserm) ; évaluation de la prévention, du dépistage et de la prise en charge thérapeutique (14 formations Inserm) ; risques liés aux pratiques thérapeutiques (6 formations Inserm).

³Les données utilisées ne sont pas parfaitement à jour, mais n'ont que très peu évolué depuis leur collecte ; d'autre part, la plupart des unités Inserm étant plurithématiques, elles se retrouvent dans plusieurs rubriques : le nombre total de formations Inserm est donc inférieur à la somme de celles qui sont citées.

Par « populations cibles » : santé de la femme et des enfants (1 formation Inserm) ; adolescents (1 formation Inserm) ; personnes âgées (3 formations Inserm) ; pays en voie de développement (4 formations Inserm, une vingtaine de chercheurs IRD répartis dans plusieurs formations : voir détail plus loin).

Une partie minoritaire, mais émergente, des chercheurs en épidémiologie a une activité de recherche quasi exclusive en **épidémiologie clinique**, une proportion importante d'entre eux étant des chercheurs hospitalo-universitaires ou hospitaliers. Ainsi, trois formations Inserm consacrent l'essentiel de leur activité de recherche à l'épidémiologie clinique, dans les domaines de l'infection à VIH, et sur des aspects méthodologiques (méthodes statistiques en recherche clinique, évaluation thérapeutique des maladies chroniques). Cette activité de recherche clinique, traditionnelle depuis longtemps dans certains centres de lutte contre le cancer, s'est progressivement répandue dans d'autres structures hospitalières. Par ailleurs, une proportion non négligeable des épidémiologistes qui consacrent l'essentiel de leurs travaux à l'étude descriptive ou étiologique, réalisent également une partie significative de leur activité de recherche dans divers aspects de l'épidémiologie clinique : évaluation des actions de prévention primaire des altérations de la santé, dépistage précoce des maladies et évaluation des actions de prévention secondaire, évaluation des méthodes diagnostiques, pronostiques et thérapeutiques. Ceci dans des domaines pathologiques très divers : cancers, maladies cardiovasculaires, maladies métaboliques, altérations de la fertilité, pathologies périnatale, infection à VIH et sida, pathologies neurologiques et liées au vieillissement. . . Il semble que l'épidémiologie clinique, qui était autrefois considérée comme négligée par les épidémiologistes et les cliniciens en France, bien qu'encore numériquement faible, ait connu un important développement ces dernières années.

Si cette énumération semble montrer que la plupart des grands domaines de la santé font l'objet de travaux de recherche épidémiologique, il faut cependant considérer que certains thèmes sont couverts par une seule équipe, voire parfois un seul chercheur ! D'autre part, si certains thèmes semblent être abordés par un nombre important de formations, le plus souvent il ne s'agit pour celles-ci que d'un aspect marginal de leurs recherches ; ainsi, tel laboratoire, consacré essentiellement à l'étude épidémiologique d'une pathologie spécifique, réalisera de temps à autres une étude incluant d'autres maladies, ou incluant des développements méthodologiques spécifiques, sans que cela représente un réel potentiel de recherche spécialisé.

Ce rapport n'a pas la prétention de présenter une réelle évaluation de la qualité de la recherche épidémiologique française. On peut cependant citer, sans être exhaustif, quelques domaines où des équipes françaises participent à une véritable compétition internationale. Il en est ainsi de l'épidémiologie génétique, de l'étude des risques pour la santé liés à l'environnement (notamment

l'environnement de travail), de la recherche méthodologique (biostatistique, modélisation), de recherches épidémiologiques sur certaines grandes pathologies (cancers, pathologies cardiovasculaires, pathologies neurologiques et vieillissement, pathologies respiratoires, infection par le VIH, pathologies de la reproduction).

Un certain nombre de domaines sont trop peu couverts en France par la recherche épidémiologique. Ainsi, en regard du poids considérable des affections concernées, la santé mentale est sans doute le domaine pathologique le moins développé actuellement en France ; qu'il s'agisse des toxicomanies, de la dépression, des suicides et plus généralement d'épidémiologie psychiatrique, le domaine des maladies psychiatriques et des conduites addictives est clairement trop peu développé. L'étude des maladies infectieuses et parasitaires en dehors du VIH reste confinée à très peu d'équipes, notamment en ce qui concerne les pays en développement : l'Institut Pasteur et les instituts Pasteur d'Outre-Mer ont une excellente compétence dans le secteur de la pathologie infectieuse, mais une certaine faiblesse en épidémiologie, bien qu'un laboratoire spécialisé ait été récemment créé ; l'IRD n'a pas non plus accordé de moyens importants à l'épidémiologie de la pathologie infectieuse, qui n'est couverte que par quelques chercheurs. Le potentiel de recherche concernant les maladies métaboliques et la nutrition est sous-dimensionné.

Encadré 2.1-2

Quelques résultats obtenus ces dix dernières années par des équipes françaises⁴

Épidémiologie génétique

De nombreux développements de méthodologie statistique adaptés aux analyses familiales des maladies multifactorielles ont été réalisés pour localiser sur le génome et caractériser les gènes impliqués dans ces maladies, identifier et modéliser les effets de ces gènes, mettre en évidence des interactions entre gènes et gène/environnement, estimer les risques associés aux facteurs génétiques, permettre l'étude de différents types de populations. Les principaux domaines d'application sont nombreux et les recherches, menées dans les équipes Inserm, ont conduit à mettre en évidence des régions du génome impliquées dans les pathologies multifactorielles (schizophrénie, autisme, asthme, lèpre, bilharziose, maladie coeliaque, sclérose en plaques,

⁴Cet encadré résume les principaux éléments d'un rapport établi par la Commission scientifique spécialisée de santé publique de l'Inserm : « La recherche en santé publique à l'Inserm : quels faits marquants au cours des dix dernières années ? », juin 2003.

diabète de type 2. . .). Des polymorphismes de gènes candidats ont été caractérisés comme étant associés à différentes pathologies (gènes du système rénine-angiotensine et maladies cardiovasculaires, gènes du métabolisme des xénobiotiques et cancers du poumon, complexe ICOS-CTLA4-CD28 et sclérose en plaques, gène de l'interféron gamma et taux d'immunoglobulines E associé à l'asthme, gène de l'enzyme de conversion de l'angiotensine et démence. . .). Le rôle de gènes connus et les risques associés ont été caractérisés (système HLA dans la polyarthrite rhumatoïde et la maladie coéliqua, gène APOE dans la maladie d'Alzheimer, gènes CDKN2A et MC1R dans le mélanome). Des interactions ont été mises en évidence entre gènes et autres facteurs de risque (cancer du sein) ou entre gènes et mesures de prévention (gène du récepteur bêta2-adrénergique et activité physique).

Cancer

La contribution à la connaissance des facteurs de risque des cancers est importante. Pour certains facteurs, cancérigènes connus, les recherches ont permis de préciser l'influence des modalités d'exposition, comme l'âge au début de l'exposition ou la dose (types de poussière de bois et cancers nasosinusiens), de l'association de différents facteurs (tabac et alcool dans le cancer de l'œsophage, facteurs familiaux et radiothérapie dans les cancers secondaires. . .), de gènes de prédisposition (tabac et polymorphismes des gènes du métabolisme dans les cancers du poumon). D'autres travaux ont suggéré de nouveaux facteurs de risque (fibres et graisses alimentaires et cancers digestifs, formaldéhyde et cancers nasosinusiens, rayons ultraviolets et mélanome oculaire, hydrocarbures polycycliques aromatiques et cancer de la vessie, produits agricoles et leucémie à tricholeucocytes, antécédents familiaux de cancers et leucémies de l'enfant, consommation de cannabis et cancer du poumon, traitement par tamoxifène après cancer du sein et cancer secondaire de l'endomètre). Des équipes ont démontré l'influence de facteurs génétiques dans la réponse aux infections par certains virus oncogènes (HTLV-1, cause de certaines leucémies, et HHV8 pour le sarcome de Kaposi) ou dans la prédisposition au mélanome associé aux ultraviolets.

Maladies cardiovasculaires

De nombreux résultats ont été obtenus dans l'étude des gènes candidats et ont permis d'approfondir la connaissance de facteurs biologiques (lipoprotéines, inflammation. . .) et environnementaux (activité physique de loisir. . .) de la maladie coronaire. On a montré l'intérêt de distinguer l'épaississement *intima-media* de l'évolution des plaques d'athérome dans la compréhension de l'histoire naturelle de l'athérosclérose carotidienne, et des facteurs de risque différentiels ont pu être mis en évidence. Les travaux menés sur l'épidémiologie de la mort subite de l'adulte ont été pionniers et reconnus comme tels au niveau international. Des facteurs de risque spécifiques comme la

mort subite parentale, le diabète, la fréquence cardiaque élevée au repos, le taux d'acides gras libres circulant ont été identifiés.

Maladies liées au vieillissement

Plusieurs études de cohorte françaises ont apporté des résultats importants : évaluation précise de la prévalence et de l'incidence de la démence en France, rôle du niveau d'études et de la profession sur le risque de démence, suggérant que les « réserves » cognitives permettent de retarder l'expression clinique, rôle possible des facteurs biologiques dans le déclin cognitif, détection précoce de la démence et description de l'histoire naturelle et des facteurs de risque de la maladie, mise en évidence de plusieurs facteurs de risque de fracture ostéoporotiques autres que la diminution de la densité minérale osseuse, en particulier les troubles du fonctionnement neuromusculaire et de la vision qui influencent le risque de chute et les capacités de protection lors de chute, influence du stress oxydant sur les pathologies oculaires liées à l'âge.

Essais cliniques

Dans le domaine des maladies infectieuses, deux essais initiés par des équipes françaises ont marqué l'évolution du traitement des patients infectés par le VIH. L'essai Delta a montré l'intérêt de bithérapies d'antirétroviraux pour ralentir l'évolution vers le stade sida et le décès ; l'interleukine-2, qui agit en stimulant l'activation et la prolifération des cellules T CD4 et CD8, a été évaluée avec succès. Par ailleurs, il a pu être montré qu'une augmentation des lymphocytes CD4+ au bout de 6 mois de trithérapie antirétrovirale constitue le marqueur pronostic le plus important pour évaluer la réponse aux multithérapies des patients ainsi traités et que certains facteurs génétiques de l'hôte sont à prendre en compte pour la connaissance du pronostic des patients. Des progrès ont été accomplis dans le domaine de la prévention de la transmission de la mère à l'enfant (efficacité des traitements courts par AZT de 36 semaines de gestation à l'accouchement), et on a pu monter l'efficacité du cotrimoxazole pour la prévention des infections opportunistes chez l'adulte.

Épidémiologie sociale

Des résultats importants ont été obtenus concernant le rôle des facteurs psychosociaux au travail sur la pathologie cardiovasculaire, la dépression et les troubles musculosquelettiques, ainsi que sur l'importance des relations sociales comme facteur protecteur. Dans le domaine du handicap, on a pu montrer les déterminants sociaux du passage de l'état d'incapacité à celui de dépendance.

Au total, si l'épidémiologie apparaît clairement comme une discipline solidement implantée à l'Inserm, qui constitue indéniablement la référence au niveau national, elle est très peu présente dans les autres organismes, qu'il s'agisse d'EPST ou des universités (malgré l'effort d'un petit nombre d'entre elles dans le cadre d'unités mixtes avec l'Inserm). De plus, elle n'est pas répartie de façon homogène dans tous les domaines d'intérêt potentiel, et à côté de points forts indiscutables, il existe des champs qui devraient être développés.

2.1.4 Aspects quantitatifs : formations de recherche et personnels

Le tableau 2.1-1 présente la répartition (en équivalents temps plein) des personnels appartenant aux formations rattachées fin 2004 à la Commission scientifique spécialisée concernée de l'Inserm (qui regroupe également la recherche en sciences sociales de la santé)⁵.

Sur les 30 formations concernées, on dénombrait 15 unités et 8 équipes autonomes ou incluses dans une unité majoritairement non épidémiologique, qui se consacrent à l'épidémiologie ; 3 services communs concourent également aux travaux épidémiologiques. Parmi ces laboratoires, un petit nombre a une orientation forte en méthodologie (biostatistique et modélisation) ou en sciences sociales. Trois IFR (à Paris-Sud, Bordeaux et Toulouse) sont orientés majoritairement vers l'épidémiologie. De plus, 7 unités rattachées à une autre commission scientifique spécialisée sont en partie engagées dans des recherches en épidémiologie. Il faut noter la grande stabilité du nombre des formations consacrées à l'épidémiologie, puisqu'en 2000, on comptait également 30 formations rattachées à la commission scientifique spécialisée concernée.

En termes d'évolution des personnels sur les cinq dernières années, on constate également que le potentiel de recherche à l'Inserm n'a pratiquement pas augmenté : on dénombrait 402 équivalents temps plein en 2000 et 408 en 2004 (toutes catégories de personnels confondues). Le nombre des chercheurs Inserm équivalents temps plein était de 114 en 2000, et de 115 en 2004. Il est important de noter la proportion très élevée de médecins parmi ces chercheurs (36 %), alors que dans l'ensemble des chercheurs de l'Inserm la proportion de médecins n'est que de 19 %. Quant aux chercheurs universitaires et hospitalo-universitaires, leur nombre, en équivalents temps plein, est passé de 54 en 2000 à 69 en 2004 ; celui des chercheurs d'autres EPST travaillant dans les laboratoires d'épidémiologie de l'Inserm de 23 en 2000 à une trentaine en 2004. Il est remarquable (et fortement dommageable) que le nombre d'ingénieurs et techniciens de l'Inserm travaillant dans ces laboratoires ait fortement diminué pendant

⁵Source : Inserm — Département Animation et partenariats scientifiques, mars 2005.

Nombre de laboratoires	Chercheurs Chercheurs		Chercheurs Chercheurs		ITA		ITA		Post-		Effectif total/		Effectif moyen/	
	Inserm	U/HU	autres EPST	autres EPST	Inserm	autres EPST	Inserm	autres EPST	doctorants	laboratoire	laboratoire	laboratoire	laboratoire	
30	115,8	69,1	30,7	157,1	19,3	16	408,1	13,6						

Tableau 2.1-1
 Personnel (équivalents temps plein) des formations Inserm travaillant en épidémiologie fin 2004. (U : universitaires ; HU : hospitalo-universitaires ; ITA : ingénieurs, techniciens et administratifs.)

cette période, passant de 175 équivalents temps plein en 2000, à 157 en 2004, cette chute n'étant pas compensée par les ITA d'autres EPST travaillant dans ces laboratoires (18 en 2000, 19 en 2004). On compte 16 post-doctorants équivalents temps plein dans les laboratoires d'épidémiologie en 2004, alors qu'ils étaient 17 en 2000. Globalement, en équivalents temps plein, l'effectif moyen par laboratoire est très faible et quasiment inchangé sur les cinq dernières années : 13,4 en 2000, et 13,6 en 2004. Rapporté à l'ensemble de l'Inserm, le domaine très vaste et transversal de la recherche en épidémiologie, ne représente que 8 % de l'ensemble des chercheurs.

Il faut souligner fortement que, malgré sa grande faiblesse numérique comparativement à d'autres pays de tradition scientifique (voir plus loin « *Quelques comparaisons internationales* »), le renforcement de l'effectif des chercheurs en épidémiologie à l'Inserm ne fait pas l'objet actuellement d'une attention excessive : ainsi, lors du concours de recrutement au titre de 2004 de cet organisme, un seul jeune chercheur épidémiologiste a pu être recruté en CR 2 malgré le très haut niveau de nombreux candidats, et aucun épidémiologiste n'a bénéficié d'un recrutement sur les postes en CDD ouverts pour la première fois à l'Inserm ! Comme de nouveau un seul poste de CR 2 a été ouvert au titre du concours 2005 pour l'ensemble de la commission scientifique spécialisée concernée (qui couvre l'ensemble des disciplines de la santé publique, c'est-à-dire l'épidémiologie, la biostatistique et les sciences sociales), on constate qu'il est prévu dans la période actuelle de n'augmenter le potentiel de jeunes chercheurs épidémiologistes de l'Inserm que d'une unité au maximum par an...

Quantitativement, la participation des universitaires à la recherche épidémiologique est sensiblement plus modeste. On dénombre au total 156 postes de titulaires (professeurs et maîtres de conférences) rattachés aux deux seules sous-sections du CNU qui sont concernées : *Épidémiologie, économie de la santé, prévention, et Biostatistique et sciences de l'information et de la communication* (données de mai 2004 ; cf. chapitre 9.1 « La formation en épidémiologie à l'université »). On peut considérer qu'environ le tiers de ces enseignants se consacre entièrement à l'épidémiologie, et qu'un autre tiers s'y consacre à mi-temps : on a donc au mieux environ 80 universitaires équivalents temps plein travaillant dans le domaine de l'épidémiologie ; il faut y ajouter un très petit nombre d'universitaires de facultés des sciences qui sont impliqués dans la recherche épidémiologique. Mais encore faut-il considérer que nombre des universitaires épidémiologistes sont rattachés à une formation Inserm et sont donc déjà comptabilisés ci-dessus.

Le Service de santé des armées compte quelques épidémiologistes regroupés essentiellement dans deux départements d'Épidémiologie et de Santé publique. Au total, on dénombre actuellement 19 épidémiologistes militaires : 3 professeurs agrégés, 8 spécialistes ou équivalents, 8 assistants ou équivalents.

Le département Sociétés et santé de l'IRD compte 5 unités de recherche et de service ayant l'épidémiologie comme discipline principale. Elles se consacrent essentiellement à la santé de la mère et de l'enfant en milieu tropical, à l'environnement et l'efficacité des interventions, au VIH/sida et maladies associées, à la nutrition. Deux autres développent des programmes de nature épidémiologique, sans que ce soit leur thématique principale, et emploient un ou des épidémiologistes. Au total, on recense 15 chercheurs et un ingénieur temps plein ayant suivi un cursus complet en épidémiologie, exerçant exclusivement dans ce domaine, et 11 chercheurs et deux ingénieurs ayant une compétence en épidémiologie, mais ne pratiquant cette discipline que pour une partie de leur temps. L'épidémiologie est en expansion à l'IRD depuis une dizaine d'années, et les associations avec les universités sont en progression (Montpellier I et II, Aix-Marseille, Paris I, Paris V et Paris VI). Il y a par ailleurs de très nombreuses associations avec des universités du Sud (Sénégal, Cameroun, Bénin, Bolivie, Thaïlande, etc.).

Le poids des autres structures citées plus haut (Institut Pasteur, INRS...) est très faible, comme on l'a déjà souligné. L'implantation sur le territoire français du Centre international de recherche sur le cancer (Circ), qui dispose d'un département d'épidémiologie, n'apporte qu'une contribution très marginale à la recherche épidémiologique nationale, même s'il joue un rôle très important au plan international. En effet, malgré la qualité des échanges entre chercheurs français et ceux de cet organisme dépendant de l'OMS, la nature même des activités épidémiologiques du Circ ne bénéficie pas plus à la France qu'à d'autres pays, puisque pour l'essentiel il a pour vocation de coordonner de grandes études internationales d'épidémiologie du cancer réalisées simultanément dans plusieurs pays, la France n'y étant pas toujours associée faute d'un nombre d'équipes suffisant pour participer à ces grands programmes épidémiologiques.

On ne dispose pas, en France comme dans les autres pays, d'un recensement véritable des épidémiologistes, qui serait difficile en tout état de cause, car, comme on l'a indiqué, l'épidémiologie est pratiquée de façon secondaire par certains cliniciens ou d'autres membres des professions de santé. Un indicateur possible des ordres de grandeurs, bien que très imparfait (des épidémiologistes professionnels très actifs n'y adhèrent pas, et bon nombre des adhérents n'ont qu'un rapport professionnel lointain avec l'épidémiologie), est le nombre d'adhérents aux deux principales sociétés savantes et professionnelles d'épidémiologie existant en France : l'Adelf (*Association des épidémiologistes de langue française*) la plus ancienne et la plus implantée parmi les chercheurs, et Epiter (*Association des épidémiologistes de terrain*) plus récente et qui concerne surtout les épidémiologistes travaillant dans les agences ou en coopération. L'Adelf compte moins de 500 adhérents français, dont 235 appartenant au milieu académique (enseignants et chercheurs), les autres provenant de l'administration de la santé, du secteur privé, des professions de santé et de milieux

divers. L'association Epiter a 400 adhérents, qui se répartissent différemment : environ 150 appartiennent à l'administration de la santé ou à la coopération, une soixantaine à des associations, bureaux d'étude ou à l'industrie privée, et un nombre équivalent à la médecine praticienne, le reste étant composé de membres du milieu académique (environ 50) et d'adhérents de statuts divers. Le nombre total de personnes adhérant à ces associations est d'environ 650, car elles ont de nombreux membres communs ; il faut par ailleurs souligner qu'une grande majorité des membres de ces associations qui sont hors du milieu académique n'a qu'une pratique ponctuelle de l'épidémiologie.

Un autre indicateur de la faiblesse quantitative de la recherche épidémiologique en France est le nombre de thèses d'université soutenues. Il n'existe que deux écoles doctorales qui ont le label « Santé publique » (Paris-Sud, cohabilitée avec Paris V, et Paris VI/VII) ; des thèses d'épidémiologie sont également préparées à Nancy et Bordeaux II, mais elles sont délivrées dans des écoles doctorales de sites regroupant toutes les disciplines. L'ensemble des écoles doctorales qui préparent à un doctorat d'épidémiologie n'a délivré au total que 35 diplômes de doctorat en 2003. La situation ne semble s'améliorer que lentement : parmi les 103 étudiants inscrits en 2005 dans les cinq DEA susceptibles de préparer à des thèses d'épidémiologie, les responsables estiment qu'environ 45 futurs thésards sont attendus dans cette promotion (cf. sous-chapitre 9.1 pour une présentation des enseignements universitaires).

Les comparaisons quantitatives que l'on peut faire avec la situation de l'épidémiologie dans d'autres pays de niveau de développement économique comparable (cf. encadré 2.1.4) montrent que la France est loin de se situer favorablement, et n'a pas su jusqu'à présent développer l'épidémiologie à un niveau satisfaisant sur le plan quantitatif. Ainsi, la taille moyenne des laboratoires d'épidémiologie de l'Inserm (qui représentent l'essentiel du potentiel de la recherche épidémiologique française) est modeste comparativement à certaines autres disciplines, ou à des structures de recherche épidémiologique dans d'autres pays : rappelons qu'ils comptent en moyenne 13,6 personnes équivalents temps plein (cf. tableau 2.1-1).

Un autre aspect particulièrement marquant est l'absence complète de grande structure d'enseignement et de recherche épidémiologique « généraliste » en France, qui pourrait se comparer à ce qui existe dans d'autres pays (cf. plus bas), malgré des projets (parfois anciens) qui ne se sont pas encore concrétisés. L'Institut de santé publique, épidémiologie et développement (Isped), rattaché à l'université Bordeaux II, qui est incontestablement la plus importante structure d'enseignement et de recherche épidémiologique en France, regroupant deux formations Inserm, un département universitaire qui délivre plusieurs diplômes en épidémiologie (cf. sous-chapitre 9.1 « La formation en épidémiologie à l'université ») et une structure d'épidémiologie hospitalière, comptabilisait en juillet

2004 un effectif total de 148 personnes. Dans un domaine qui ne relève pas de l'enseignement et de la recherche, mais de la surveillance épidémiologique (cf. plus loin), l'Institut de veille sanitaire (InVS) employait au total sur ses différents sites 251 personnes en juillet 2004.

Encadré 2.1-3

Quelques comparaisons internationales⁶

Structures d'enseignement supérieur et de recherche « généralistes ».

Aux États-Unis, on dénombre au sein des universités pas moins de 36 Graduate Schools of Public Health délivrant des PhD, et 57 Graduate Programs in Community Health ou Preventive Medicine délivrant des masters en dehors des écoles de santé publique. La taille de ces structures généralistes est élevée comparativement à la France : ainsi, le corps d'enseignants chercheurs de la John Hopkins School of Public Health de Baltimore est 1 133 personnes, dont 501 temps plein (effectif à peine inférieur à l'ensemble du potentiel français de recherche!), la Harvard School of Public Health emploie 918 personnes, dont 472 enseignants chercheurs, la Carolina School of Public Health 197 enseignants chercheurs temps plein, la UCLA School of Public Health a un effectif de 400 personnes, et la Berkeley School of Public Health de 253. Au Canada, en ne comptabilisant que le corps enseignant titulaire du département d'Épidémiologie et biostatistique de l'université McGill de Montréal, le total s'élève à 91 personnes ; si on y ajoute les autres personnels, l'effectif total est de 230 environ.

En Europe également, plusieurs pays possèdent d'importantes structures d'enseignement supérieur et de recherche généralistes. En Grande-Bretagne, la London School of Hygiene and Tropical Medicine compte 729 personnes (dont 405 enseignants et chercheurs), le seul département d'Épidémiologie et Santé publique de l'Imperial College 70 personnes, et même la section Santé publique de la petite université de Sheffield possède un corps enseignant de 32 personnes temps plein. Aux Pays-Bas, pays de 16 millions d'habitants, le département de Santé publique de l'université Erasmus de Rotterdam a un effectif de 125 personnes.

Ces exemples ont été choisis arbitrairement, car il existe de nombreuses autres structures d'enseignement et de recherche généralistes dans les pays cités. **Rappelons que l'Ispep de Bordeaux, de loin la plus importante**

⁶Les données citées proviennent des sites Web des institutions concernées, consultés en juillet et août 2004.

structure française dont le périmètre est comparable, emploie au total moins de 150 personnes, toutes catégories confondues.

Laboratoires spécialisés. Le département d'Épidémiologie et santé publique du London University College, qui se consacre exclusivement à l'épidémiologie sociale, compte 151 personnes, dont 113 enseignants et chercheurs ; la Division of Clinical Epidemiology du Royal Victoria Hospital de Montréal (qui est une des sept unités qui composent le département d'Épidémiologie et Biostatistique de l'université McGill) emploie 34 enseignants et chercheurs, sans compter les autres catégories de personnel.

Rappelons que le nombre moyen de chercheurs et enseignants chercheurs dans les unités Inserm d'épidémiologie est de 10,5.

Agences de santé publique chargées de la surveillance et de l'intervention épidémiologiques. Les Centers for Disease Control and Prevention (CDC) des États-Unis emploient plus de 8 500 personnes, alors que l'InVS en compte 251 ; les autres agences nationales françaises (Afsa, Afsaps, Affse) qu'il faut comptabiliser pour couvrir un périmètre équivalent aux CDC, emploient un nombre dérisoire d'épidémiologistes. Cette énorme disproportion se retrouve pour les structures spécialisées dans un domaine spécifique. Ainsi, le département Santé travail de l'InVS, en charge de la surveillance épidémiologique des risques professionnels en France, compte environ 20 personnes, alors que le Health & Safety Executive britannique en compte 4 500, le National Institute for Occupational Safety and Health (NIOSH) américain plus de 1 400, et même la Finlande (5,2 millions d'habitants) dispose du Finnish Institute of Occupational Health, qui emploie environ 800 personnes. Ces organismes couvrent un périmètre plus large que la stricte surveillance épidémiologique, mais les structures chargées de cette mission y sont très conséquentes.

Associations d'épidémiologistes. Alors que les associations françaises (Adelf et Epiter) comptent environ 650 membres au total, quelque 1 000 épidémiologistes sont inscrits à l'association d'épidémiologie des Pays-Bas, dont la population est presque quatre fois moins nombreuse ; environ 750 appartiennent au monde de la recherche académique : universités et autres organismes publics de recherche. Il en est de même pour environ 700 des 825 membres de l'association espagnole d'épidémiologie. La Société anglaise d'épidémiologie est la plus ancienne (elle a été créée en 1850), mais il n'a pas été possible d'obtenir de chiffres concernant le nombre de ses adhérents. Le nombre d'épidémiologistes professionnels au Canada, dont la population est de 32 millions d'habitants, est estimé à environ 900. Les États-Unis ne comptent pas moins de trois associations « généralistes » d'épidémiologie (American College of Epidemiology, American Epidemiology Society, Society for Epidemiologic Research), ainsi que des associations

d'épidémiologistes spécialisés (épidémiologie de l'environnement, pharmaco-épidémiologie, etc.). Il est bien entendu hasardeux de comparer ces chiffres, car les conditions d'adhésion à une association professionnelle peuvent différer beaucoup d'un pays à l'autre et avoir une signification différente ; il semble néanmoins que ces données confirment le nombre très restreint d'épidémiologistes français comparativement à d'autres pays.

2.1.5 La production scientifique : la place de la France dans la recherche internationale⁷

L'analyse bibliométrique est un exercice difficile, dont les résultats, notamment appliqués à l'évaluation individuelle, sont souvent discutables. Néanmoins, lorsqu'il s'agit comme ici de procéder à des comparaisons entre pays, elle permet de situer globalement la place d'un pays dans la hiérarchie de la recherche internationale, et fournit des indications intéressantes.

La base que nous avons choisie d'interroger en priorité est Medline (NLM), car elle est indubitablement l'outil bibliographique le plus répandu dans le domaine biomédical.

Une des difficultés est de définir ce qu'est une publication d'épidémiologie. En effet, en dehors des articles publiés dans des revues explicitement consacrées à l'épidémiologie, on peut rencontrer des articles d'épidémiologie dans des revues de santé publique ou de santé environnementale, ainsi que dans des revues médicales, généralistes ou spécialisées.

Sera présentée ici en détail l'une des analyses effectuées : elle repose sur l'analyse du nombre d'articles publiés dans la catégorie de journaux définis par les indexeurs de Medline comme appartenant à la catégorie « épidémiologie » (*JC = Epidemiology*). Les résultats sont présentés dans le tableau 2.1-2.

Deux ratios ont été calculés pour comparer la production de chaque pays à celle de la France : un ratio brut (pourcentage des publications du pays divisé par pourcentage des publications françaises), et un ratio pondéré par l'effectif de la population des pays comparativement à la France. Le total des pays analysés correspond à 94,5 % de la production mondiale.

En nombre absolu de publications, les États-Unis sont les premiers (1 008 publications sur la période), l'Europe des 15 est présente dans le tableau pour

⁷Cette partie a été rédigée à partir de l'analyse bibliométrique réalisée par Sophie Lesieur (Inserm Unité 707) ; nous remercions aussi Gilles Hejblum pour son aide.

Classement	Pays	Période	2003-2004	RF	RFp
		n	%		
1	Danemark	71	2,33 %	0,8	9,4
2	Norvège	49	1,61 %	0,6	7,6
3	Suède	83	2,72 %	1,0	6,6
4	Pays-Bas	129	4,23 %	1,5	5,7
5	Finlande	40	1,31 %	0,5	5,5
6	Australie	116	3,80 %	1,4	4,2
7	UK	297	9,73 %	3,5	3,6
8	Canada	152	4,98 %	1,8	3,4
9	Rép. Tchèque	38	1,25 %	0,5	2,6
10	Suisse	27	0,88 %	0,3	2,6
11	N. Zélande	14	0,46 %	0,2	2,5
12	États-Unis	1008	33,04 %	12,0	2,5
13	Israël	20	0,66 %	0,2	2,1
14	Irlande	10	0,33 %	0,1	1,8
15	Grèce	26	0,85 %	0,3	1,7
16	Italie	128	4,20 %	1,5	1,6
17	Pologne	68	2,23 %	0,8	1,3
18	France	84	2,75 %	—	—
19	Belgique	15	0,49 %	0,2	1,0
20	Espagne	56	1,84 %	0,7	1,0
21	Autriche	9	0,29 %	0,1	0,8
22	Russie	136	4,46 %	1,6	0,7
23	Chine	107	3,51 %	1,3	0,6
24	Allemagne	62	2,03 %	0,7	0,5
25	Japon	92	3,02 %	1,1	0,5
26	Turquie	18	0,59 %	0,2	0,2

Classement	Pays	Période	2003-2004	RF	RFp
27	Bésil	28	0,92 %	0,3	0,1
	Total		94,49 %		

n : nombre de références du pays. % : part du pays sur l'ensemble des publications de la période. RF : ratio des publications du pays par rapport aux publications de la France. RFp : ratio des publications pondéré par la taille population.

Tableau 2.1-2

Classement des productions d'articles par pays, rapportés à la taille de la population du pays.

La base Medline a été interrogée avec l'interface Silverplatter, sur une période couvrant l'année 2003 et l'année 2004 jusqu'en avril. Les articles d'épidémiologie ont été identifiés en recourant au champ « Journal Category » de la référence. Ce champ « JC » identifie de grands groupes de journaux correspondant aux grands domaines biomédicaux ; l'affectation au champ JC est effectuée par les indexeurs de la National Library of Medicine.

33,1 %, celle des 25 pour 36,6 %. La France est le 10^e pays producteur ; 4 pays européens la devancent (Grande-Bretagne, Pays-bas, Italie, Suède).

Le nombre relatif de publications (colonne RFp du tableau) permet de tenir compte de la taille très différente des pays. Plusieurs observations s'imposent : tout d'abord le dynamisme des pays du Nord qui, en nombre relatif, publient 5 à 9 fois plus que la France. Ensuite, celle-ci n'est classée que 18^e (elle est classée 10^e en nombre brut). Enfin, en Europe, la France est loin derrière le Royaume-Uni, le Danemark, la Norvège, la Suède. On note également la faible présence des pays asiatiques : Chine avec 3,5 %, Inde non présente.

Les autres analyses bibliographiques que nous avons effectuées ont toutes confirmé qualitativement ces classements : ces recherches ont été menées sur le sous-ensemble des sept journaux d'épidémiologie les plus cités et, dans ce fond, la production des pays européens a été recherchée. La France est largement devancée par les pays du Nord en nombre relatif, dans une proportion identique au tableau précédent.

Le constat global de ces comparaisons concernant la production scientifique en épidémiologie est une place médiocre de la France dans la compétition scientifique internationale. Il faut cependant nuancer l'interprétation de ce fait incontestable en tenant compte du nombre de chercheurs français dans la discipline, qui est particulièrement faible, comme on l'a montré plus haut. Sans qu'il soit possible de quantifier ce phénomène, faute de données fiables concernant le nombre de chercheurs en épidémiologie par pays, il semble en fait que

Organisme	Journaux d'épidémiologie	Tous journaux	
EPST	29	159	
<i>Inserm</i> ¹		23	119
CNRS		4	19
IRD		1	6
Inra		1	11
Cirad		0	4
Agences	3	21	
<i>Afssa</i>		0	8
<i>InVS</i>		3	13
Services hospitalo-universitaires ²	26	525	
Laboratoires universitaires ²	3	20	
Instituts privés	2	14	
Santé des armées	0	1	
Associations	0	2	
Observatoires régionaux	1	2	
Sociétés privées	0	10	
Total	64	754	

1 : presque toujours en mixité avec l'Université. 2 : affiliés ou non à un EPST.

Tableau 2.1-3

France : Publications par organismes (période 2003-2004).

La base Medline a été interrogée avec l'interface Silverplatter, sur une période couvrant l'année 2003 et l'année 2004 jusqu'en mars. Le tri sur l'appartenance institutionnelle du premier auteur a été fait après isolement des références en langue anglaise contenant le terme « France » dans l'adresse de l'auteur. Très souvent, les champs d'adresse sont mal renseignés, notamment lorsqu'ils proviennent d'institutions non universitaires. La répartition selon les organismes a donc été effectuée manuellement. Les chiffres reportés dans ce tableau sont sans doute sous-évalués par rapport à la production réelle de ces organismes.

Les nombres de publications sont présentés :

1^o colonne : nombre de publications parues dans les journaux de la catégorie JC = Epidémiologie (cf. légende du tableau 2.1-2).

2^o colonne : nombre de publications publiées dans l'ensemble des journaux référencés dans Medline et ayant été indexées à la NLM avec le qualificatif « Epidemiology ».

la productivité individuelle des chercheurs français est certainement très bonne, mais qu'ils sont trop peu nombreux pour permettre à notre pays une contribution plus importante à la recherche épidémiologique internationale.

En France même, on a cherché à quantifier la contribution des organismes selon leur nature (tableau 2.1-3). Deux analyses ont été faites. La première concerne les articles publiés dans les journaux classés sous la rubrique *Épidémiologie* par la NLM, dont l'adresse du premier auteur est en France ; la seconde analyse porte sur l'ensemble des articles comprenant un « *Qualificatif* »⁸ : *Epidemiology* dans la base Medline.

La production publiée dans des revues d'épidémiologie proprement dites se partage à parts à peu près égales entre EPST et services hospitalo-universitaires ; concernant ces derniers, il est difficile de savoir s'ils sont ou non affiliés à un EPST, ceci dépendant de la façon dont le premier auteur a indiqué son affiliation. Parmi les EPST, l'Inserm (dont la quasi-totalité des unités est mixte avec l'Université) exerce de fait un quasi-monopole (23 publications sur 29). Lorsque le dénombrement est fait dans l'ensemble des références indexées avec le *Qualificatif* : *Epidemiology*, l'Inserm reste très largement prédominant au sein des EPST et des agences, mais la part des services hospitalo-universitaires (affiliés ou non à un EPST) devient largement prépondérante (70 % de la production totale).

Ces dénombrements illustrent le fait que l'Inserm, associé le plus souvent à l'Université, est de très loin l'organisme qui en France a la plus importante activité de recherche en épidémiologie générale (publications dans des revues classées *Épidémiologie*), alors que l'épidémiologie clinique (publications avec *Qualificatif* : *Epidemiology* dans des revues médicales, qui correspondent le plus souvent à ce type de recherche) est l'apanage des hospitalo-universitaires, n'appartenant pas le plus souvent à la discipline universitaire épidémiologie.

Ils soulignent également que la contribution de l'Inserm et des hospitalo-universitaires à la recherche épidémiologique, générale et clinique, est tout à fait majeure (76 % du total des publications dans des revues classées *Épidémiologie*, 85 % des publications avec *Qualificatif* : *Epidemiology*), la part de chacun des autres organismes étant anecdotique.

2.2 La surveillance épidémiologique

On observe depuis quelques années en France une demande très forte et constamment croissante dans le domaine de la surveillance épidémiologique, institutionnalisée par la création des agences de santé et assimilés (InVS, Affsaps, Affsa, Afsse).

⁸Les qualificatifs, dans le thesaurus de la base Medline représentent des concepts généraux employés pour préciser le sens des descripteurs (mots-clés de l'article).

La surveillance épidémiologique peut être définie comme le suivi et l'analyse épidémiologiques systématiques et permanents d'un problème de santé et de ses déterminants à l'échelle d'une population, afin de les contrôler par des interventions au niveau individuel ou collectif, et d'identifier des phénomènes inconnus en termes d'effets ou de déterminants. Alors que la recherche épidémiologique a pour objectif l'acquisition de connaissances nouvelles et généralisables à des populations différentes de celle où l'étude a été faite, la surveillance a pour objectif l'acquisition de connaissances pour la population d'intérêt (« ici et maintenant ») mais *a priori* non généralisables à des populations différentes. À titre d'exemple, l'identification d'un agrégat de cas de cancer de la vessie dans une imprimerie, ou de mésothéliome dans un chantier naval sont des résultats importants pour un programme de surveillance épidémiologique, car ils signent la présence d'un problème impliquant une intervention (prévention, dépistage, réparation) dans les entreprises concernées (mais pas nécessairement dans d'autres entreprises de même type). En revanche, ils peuvent ne pas intéresser la recherche épidémiologique, car le rôle des colorants vis-à-vis des cancers de la vessie dans l'imprimerie, ou de l'amiante vis-à-vis du mésothéliome dans les chantiers navals sont aujourd'hui bien établis. Outre leur objectif de contrôle de la situation sanitaire de la population, les systèmes de surveillance épidémiologique ont également un rôle d'alerte vis-à-vis de phénomènes émergents, qu'ils cherchent à détecter le plus précocement possible (maladies émergentes, facteurs de risques inconnus ou dont les effets se modifient).

C'est actuellement presque exclusivement à l'InVS que s'est développée une activité de surveillance épidémiologique systématique d'un certain nombre de problèmes de santé humaine (maladies infectieuses et maladies chroniques) et de facteurs de risque (environnement général et professionnel).

L'Afssa a également développé des réseaux de surveillance épidémiologique. Au nombre d'une vingtaine, ils sont essentiellement consacrés au domaine de la pathologie vétérinaire, mais leur importance pour la santé humaine peut être considérable, car on sait que la plupart des grandes alertes sanitaires des vingt dernières années ont été dues à des maladies animales transmises à l'homme (voir sous-chapitre 2.2 « L'interface épidémiologie animale — épidémiologie humaine). Certains sont des réseaux « ciblés » sur une ou plusieurs maladies bien définies, d'autres des réseaux « globaux » concernant toutes les maladies d'une espèce animale. À titre d'exemple, on peut citer, dans le premier groupe, des réseaux ciblés sur la rage, l'ESB ou les salmonelloses ; dans le second des réseaux consacrés aux pathologies équinées, aviaires ou bovines, ainsi qu'aux alertes alimentaires.

L'InVS a connu un développement très rapide : créé en 1998 (il prenait la suite du RNSP créé en 1994 et composé alors d'une douzaine de personnes),

il emploie aujourd'hui environ 250 personnes (incluant le personnel déconcentré en régions) ; la comparaison des effectifs de l'InVS avec ceux de son homologue américain (*cf.* plus haut : *Aspects quantitatifs*) laisse prévoir que le développement de l'InVS n'en est qu'à ses débuts. Cette véritable explosion de la demande en professionnels de l'épidémiologie appliquée, par les agences de santé, n'est pas sans poser des problèmes concrets aujourd'hui et pour les années à venir. En effet, la formation des épidémiologistes à un niveau suffisant pour correspondre aux besoins des agences (thèse et quelques années d'expérience dans un domaine de spécialisation) ne peut être assurée que dans des laboratoires de recherche « académiques » (notamment à l'Inserm et dans quelques laboratoires hospitalo-universitaires, qui concentrent la quasi-totalité des moyens de la recherche épidémiologique française). Or, la faiblesse quantitative de la recherche (petit nombre de laboratoires d'épidémiologie, recrutements au compte-gouttes de chercheurs, d'enseignants et de personnel de soutien) ne permet pas actuellement d'assurer la formation d'épidémiologistes de niveau adéquat en nombre suffisant. Il est donc à craindre que le niveau de qualification des « épidémiologistes d'agence » soit insuffisant du fait d'un « effet de ciseau » (diminution de l'offre et augmentation de la demande).

Ceci est déjà largement vrai ; ainsi le département des Maladies chroniques et des traumatismes de l'InVS cherche depuis plus de deux ans le responsable de son unité Cancer sans trouver un spécialiste reconnu pour occuper ce poste ; plusieurs emplois d'épidémiologistes spécialistes des risques professionnels sont publiés depuis des mois, sans qu'aucun candidat correspondant au profil recherché ne se soit manifesté. On trouve à une moindre échelle le même type de difficultés pour d'autres structures de taille plus modeste, comme les ORS et registres de maladie, qui font également l'objet de demandes croissantes de connaissances épidémiologiques de la part des collectivités territoriales et des pouvoirs publics locaux et régionaux, phénomène destiné à s'amplifier lors de la mise en œuvre des dispositions de la loi de santé publique d'août 2004, qui prévoit notamment la création de « Groupements régionaux de santé publique » sous l'autorité du préfet de région.

Cette situation paradoxale, caractérisée par la création d'un nombre important d'emplois publics d'épidémiologistes dans les agences et par la diminution simultanée du nombre d'emplois dans le secteur académique, est la marque d'un défaut d'anticipation des pouvoirs publics, et de leur méconnaissance de ce que sont l'épidémiologie et les épidémiologistes, des nécessités et des contraintes de leur formation. Cette situation, qui ne pourra qu'empirer si un effort vigoureux et suffisamment long de formation et de développement des structures de recherche n'est pas fait, est porteuse de difficultés pour l'avenir, non seulement de la recherche, mais de la santé publique française.

2.3 Études diverses

Des organismes divers (collectivités locales, organismes de protection sociale, médecine du travail, médecine scolaire, départements d'information médicale (Dim) hospitaliers, laboratoires pharmaceutiques, entreprises industrielles, etc.) ont de plus en plus fréquemment besoin d'études épidémiologiques, essentiellement de type descriptif et évaluatif, dans une optique d'aide à la décision : description de l'état de santé de populations particulières, recours aux soins, pharmaco-économie, évaluation de besoins de santé, etc.

Ces études sont parfois menées en interne par les organismes concernés, ou elles sont confiées à des structures spécialisées. Cependant, leur réalisation se heurte à la disponibilité de spécialistes en épidémiologie, et là aussi, il est à craindre que la qualité des études soit souvent insuffisante. Le problème du trop petit nombre de laboratoires de recherche formateurs et la pénurie actuelle en épidémiologistes de niveau élevé qui en découle, évoquée plus haut, ne pourra que s'amplifier encore dans la période à venir, au fur et à mesure que ces nouveaux besoins, qui commencent à s'exprimer, se traduiront par une demande accrue.

2.4 Aide à la décision et aux politiques de santé publique

L'épidémiologie est une discipline essentielle en termes de connaissances scientifiques et de production des informations de base sur lesquelles se fondent en grande partie les décisions de santé publique. On peut citer quelques exemples d'utilisation directe de résultats de la recherche épidémiologique pour asseoir des décisions politiques en santé.

Ainsi les résultats de l'étude Paquid ont servi de référence pour la définition des modalités de prise en charge des démences. Les recherches sur la sexualité et la reproduction ont permis l'adaptation des stratégies de prévention. Les agences de sécurité sanitaire font régulièrement appel à l'expertise des épidémiologistes et biostatisticiens, qui ont, par exemple, apporté une contribution majeure à l'évaluation quantitative des risques liés aux infections à prions (notamment risque transfusionnel et risque alimentaire). Ce sont également des travaux de recherche épidémiologique sur les risques professionnels de cancer qui ont fondé des décisions importantes, comme des modifications de valeurs limites d'exposition, l'extension des critères de reconnaissance de maladies professionnelles, ou la constitution du Fonds d'indemnisation des victimes de l'amiante.

Une autre expression de la contribution de l'épidémiologie à la décision est celle des expertises scientifiques réalisées à la demande d'autorités et

d'organismes divers par l'Inserm et publiées dans sa collection « *Expertises scientifiques collectives* ». Celles-ci impliquent toujours une expertise épidémiologique, parfois de façon très majoritaire, comme le montre clairement la liste des titres figurant dans l'encadré 2.4. Une telle action est menée également de façon croissante dans les agences (Afssa, InVS, Afssaps). La majorité des problèmes aigus de santé publique exige l'expertise épidémiologique : on peut citer par exemple le récent rapport mené à l'Afssaps en septembre 2005 sur le traitement hormonal substitutif de la ménopause.

Encadré 2.1-4

Liste des expertises scientifiques collectives publiées par l'Inserm

- Psychothérapie : trois approches évaluées
- Tabac : comprendre la dépendance pour agir
- Santé des enfants et des adolescents : propositions pour la préserver
- Hépatite C : transmission nosocomiale — État de santé et devenir des personnes atteintes
- Tabagisme : prise en charge chez les étudiants
- Alcool : dommages sociaux — Abus et dépendance
- Troubles mentaux : dépistage et prévention chez l'enfant et l'adolescent
- Déficiences visuelles : dépistage et prise en charge chez le jeune enfant
- Asthme : dépistage chez l'enfant
- Cannabis : effets sur le comportement et la santé
- Alcool : effets sur la santé
- Éducation pour la santé des jeunes : démarches et méthodes
- Susceptibilités génétiques et expositions professionnelles
- Rythmes de l'enfant : de l'horloge biologique aux rythmes scolaires
- Dioxines dans l'environnement : quels risques pour la santé ?
- Lombalgies en milieu professionnel : quels facteurs de risque et quelle prévention ?
- Obésité et prévention : dépistage et prévention chez l'enfant
- Maladies parodontales : thérapeutiques et prévention
- Éthers de glycol : quels risques pour la santé ?

- Vaccinations : actualités et perspectives
- Carences nutritionnelles : étiologies et dépistage
- Plomb dans l'environnement : quels risques pour la santé ?
- Effets sur la santé des fibres de substitution à l'amiante
- Évaluer la mise à disposition du Subutex pour la prise en charge des usagers de drogue
- Étude épidémiologique générale sur le suivi des patients bénéficiant de prescription de Méthadone depuis 1993
- Insuffisance rénale chronique : étiologies, moyens de diagnostic précoce, prévention ?
- Risques héréditaires de cancers du sein et de l'ovaire : quelle prise en charge ?
- La migraine : connaissances descriptives, traitements et prévention
- Effets sur la santé des principaux types d'exposition à l'amiante
- Hépatites virales : dépistage, prévention, traitement
- Grande prématurité : dépistage et prévention du risque
- Imagerie médicale en France dans les hôpitaux publics
- Méningites bactériennes : stratégies de traitement et de prévention
- Ostéoporose : stratégies de prévention et de traitement
- Sida, maladies associées : pistes pour de nouveaux médicaments
- Rachialgies en milieu professionnel : quelles voies de prévention ?
- La grippe : stratégies de vaccination
- Artériopathie des membres inférieurs : dépistage et risque cardiovasculaire

Conclusions

En conclusion de ce chapitre, qui présente de façon succincte l'état des forces de la recherche épidémiologique française, on doit mettre l'accent sur les principales spécificités qui caractérisent notre pays dans le contexte de la compétition scientifique internationale.

En premier lieu, on ne peut qu'être frappé, lorsqu'on situe notre pays dans le contexte international, par **la très grande faiblesse numérique du potentiel de**

recherche français, quels que soient les indicateurs retenus : nombre de chercheurs et enseignants chercheurs, nombre de thèses soutenues, taille moyenne des laboratoires spécialisés et des structures généralistes, . . . Il faut souligner que sur le plan qualitatif, il apparaît cependant que le niveau de la formation des futurs chercheurs, ainsi que les modalités d'organisation et la qualité de la recherche épidémiologique sont très comparables à ce qu'on rencontre dans les pays les plus avancés, ce qui est attesté par le rang occupé par le nombre des publications issues des laboratoires français, non négligeable quand on le rapporte aux forces existantes.

Un second constat est celui de l'inadéquation des financements, et (peut-être surtout) des modalités de gestion des organismes de recherche, à la nature et l'ampleur des recherches épidémiologiques actuelles, qui menacent de rapidement marginaliser encore plus la France dans la compétition scientifique internationale. Ainsi, à titre d'exemple, on constate que les cohortes épidémiologiques françaises se caractérisent par leur taille relativement faible, aucune ne dépassant un petit nombre de dizaines de milliers de sujets (la plus grande cohorte française, l'étude E3N⁹ (Clavel-Chapelon *et al.*, 2002) a inclus 100 000 femmes), alors que certaines cohortes dans d'autres pays peuvent atteindre plusieurs centaines de milliers de sujets, voire plus. À titre d'illustration, on peut citer en Grande-Bretagne la *One Million Women Study* (Darling, 1998)¹⁰ le projet *UK Biobank*¹¹ qui prévoit le suivi prospectif de 500 000 personnes, ou la *General Practice Research Data Base* (Jick, 1991) qui gère les données de santé de 4 millions de personnes depuis plus de 15 ans. En Norvège, la *Norwegian Mother and Child Cohort Study* est une « cohorte de naissance », qui a inclus 100 000 femmes à la 18^e semaine de grossesse, puis leurs 100 000 nouveau-nés, ainsi que 70 000 pères, soit au total 270 000 personnes (Stoltenberg, 2003). La *Nurses'Health Study* a été mise en place aux États-Unis dès 1976 et assure le suivi prospectif de 122 000 infirmières ; une deuxième vague a été mise en place en 1989 et a inclus 117 000 femmes (Zhang, 2000). Par comparaison avec ces réalisations, plusieurs raisons expliquent la relative modestie des cohortes françaises : difficulté pour réunir les importants financements nécessaires et ceci sur de longues périodes ; impossibilité de disposer de personnel spécialisé stable et d'un niveau de qualification suffisant (même lorsque les financements nécessaires sont disponibles !), du fait de la rigidité des règles de gestion des personnels des établissements publics, alors que la durée des suivis de cohorte est incompatible avec un trop fort renouvellement des personnels techniques qui doivent assurer la continuité des procédures et des recueils de données ; extrême rigidité des règles de gestion

⁹Clavel-Chapelon F and the E3N-EPIC group. Secular trends of age at menarche and at onset of regular cycling in a large cohort of French women. *Human Repr* 2002 ; 17 : 228-232.

¹⁰Darling GM, Davis SR, Johns JA. Hormone replacement therapy compared with simvastatin for postmenopausal women with hypercholesterolemia. (1998) *N Eng J Med* ; 338 : 64.

¹¹<http://www.ukbiobank.ac.uk/status.htm>.

des fonds (même d'origine privée) des établissements publics, souvent incompatibles avec la mobilisation rapide de moyens adéquats (on se rapportera au sous-chapitre 8.2 « Épidémiologie et décision en santé publique » pour une analyse plus détaillée des moyens nécessaires à la recherche épidémiologique).

Recommandations

- Favoriser l'émergence de grandes structures généralistes en épidémiologie (écoles, instituts...), regroupant un nombre suffisant de laboratoires spécialisés.
- Inciter les grandes structures médicales spécialisées à développer en leur sein des laboratoires d'épidémiologie ; ceux-ci doivent être de taille suffisante, dotés d'une autonomie scientifique et rattachés à des instances scientifiques nationales d'épidémiologie, afin d'éviter que des épidémiologistes trop isolés ne soient réduits à la réalisation de travaux ancillaires).
- Augmenter sensiblement et durablement les emplois du secteur de la recherche et de l'enseignement.
- Assouplir les règles de recrutement de personnels techniques stables dans les établissements de recherche.
- Assouplir les règles de gestion des budgets.
- Veiller à un niveau de qualification suffisant des épidémiologistes employés dans les structures de santé publique et de surveillance épidémiologique.
- Favoriser la participation des cliniciens et autres professionnels de santé à la recherche d'épidémiologique, par des rattachements partiels plus aisés à des structures de recherche en épidémiologie.
- Favoriser les études multicentriques à l'échelle européenne, en mobilisant des budgets plus importants et en assouplissant les règles de gestion.

Bibliographie

- Clavel-Chapelon F. and the E3N-EPIC group (2002). Secular trends of age at menarche and at onset of regular cycling in a large cohort of French women. *Human Repr*, **17** : 228-232.
- Darling GM., Davis SR., Johns JA. (1998). Hormone replacement therapy compared with simvastatin for postmenopausal women with hypercholesterolemia. *N Eng. J. Med.*, **338** : 64.

- Jick H., Jick SS., Derby LE. (1991). Validation of information recorded on general practitioner based computerised data resource in the United Kingdom. *Br Med J*, **302** : 766-768.
- Stoltenberg C. (Jan. 2003). The Norwegian Network of Human Research Biobanks and Health Studies. Norwegian institute of public health, *Division of epidemiology*, Oslo.
- Zhang SM., Willett WC., Hernan MA., Olek MJ., Ascherio A. (2000). Dietary fat in relation to risk of multiple sclerosis among two large cohorts of women. *Am. J. Epidemiol*, **152** : 1056-64.

SOUS-CHAPITRE 2.2

L'interface épidémiologie animale – épidémiologie humaine

CHARLES PILET ET BERNARD TOMA

Au sein de cet inventaire des forces épidémiologiques en France, il est possible de situer l'apport des épidémiologistes de formation initiale vétérinaire et son évolution souhaitable.

1 | Vétérinaires et épidémiologie

En France, des vétérinaires travaillent en épidémiologie, pour une partie ou la totalité de leur temps, dans des structures consacrées soit à la santé animale, soit à la santé humaine.

Pour la santé animale, il en est ainsi de l'Inra, de l'Afssa, de la Direction générale de l'alimentation (DGAl) du ministère de l'Agriculture, du Cirad, des écoles nationales vétérinaires, pour ne citer que les principales de ces structures.

Leur formation en épidémiologie est assurée, en plus de la période initiale en écoles vétérinaires, grâce à différents moyens, spécifiques de l'épidémiologie animale (exemple : diplôme d'épidémiologie animale appliquée, CES d'épidémiologie animale. . .) ou communs à l'épidémiologie animale et à l'épidémiologie humaine (exemple : cours de l'Institut Pasteur, cours de Veyrier du lac (DEA) et doctorats d'université. . .).

Le nombre de vétérinaires travaillant en épidémiologie animale à temps plein dans ces structures est de l'ordre d'une vingtaine. Par ailleurs, au sein de l'Association pour l'étude de l'épidémiologie des maladies animales (AEEMA), un peu plus de 100 vétérinaires français sont directement intéressés par, et impliqués d'une manière ou d'une autre dans, l'épidémiologie animale.

En revanche, la présence de vétérinaires travaillant en épidémiologie dans des structures consacrées à la santé humaine est très réduite.

Sans prétendre à l'exhaustivité, on peut citer simplement :

- l'Institut Pasteur : six vétérinaires (*pro-parte*) ;
- l'Institut de veille sanitaire : deux vétérinaires ;
- les Cire (centres interrégionaux d'épidémiologie) : un vétérinaire ;
- la Direction générale de la santé : deux vétérinaires.

Le recours à des vétérinaires épidémiologistes (ou travaillant en partie en épidémiologie) dans des structures de santé humaine est donc très limité. Pourtant, le danger pour l'homme représenté par le monde animal est loin d'être négligeable. N'oublions pas en effet qu'au cours des vingt dernières années, la plupart des grandes alertes sanitaires ont eu l'animal pour origine.

2 | Danger animal

2.1 Pour l'homme

De tout temps, le danger du monde animal pour l'homme a été soupçonné ou reconnu. Des maladies majeures, d'origine exclusivement animale, ont pu être maîtrisées ou éradiquées, du moins en Europe occidentale. Il en est ainsi, par exemple, de la morve et de la rage. Quelques grandes maladies atteignant les animaux domestiques ont été pratiquement éradiquées en France (la tuberculose et la brucellose bovines).

Cependant, même pour de telles maladies, le risque de réintroduction accidentelle demeure, notamment à cause de la mondialisation des échanges, y compris d'animaux et de produits d'origine animale. La réintroduction périodique de la rage par importation illicite de chiens illustre bien ce risque pérenne.

D'autres maladies d'origine animale directe ou indirecte ont une incidence qui ne diminue pas, en particulier en raison de la fréquence du portage sain chez l'animal domestique ou sauvage : les salmonelloses, les infections par *Campylobacter jejuni*, par *Escherichia coli* O157H7, la fièvre Q, les hantaviroses, la pasteurellose et les autres bactérioses transmises par morsure, etc.

Enfin, diverses zoonoses sont apparues au cours des dernières décennies ou sont en cours d'émergence. Parmi elles, on peut citer, sans prétendre à l'exhaustivité :

- le nouveau variant de Creutzfeld-Jacob qui a été à l'origine de crises majeures de l'alimentation (viande bovine) dans de nombreux pays et a conduit à la mise en œuvre de mesures de lutte particulièrement coûteuses contre l'encéphalopathie spongiforme bovine ;
- les divers agents pathogènes (bactéries, virus, parasites. . .) hébergés par les nouveaux animaux de compagnie (NAC) en vogue dans les pays occidentaux et dont l'importation provoque l'introduction d'agents exotiques, comme récemment le virus de la variole du singe aux États-Unis ;
- le virus West-Nile, qui a envahi l'ensemble des États-Unis au cours des quatre dernières années et qui poursuit son expansion en Amérique du Nord en tuant des oiseaux, des chevaux et des humains. Ce virus s'est également signalé à plusieurs reprises depuis quelques années, dans le midi de la France ;
- le virus Hendra ;
- le virus Nipah, dont l'aire géographique de mise en évidence est en extension ;
- le virus grippal aviaire qui a été à l'origine, au cours de ces derniers mois, de la mort de plus d'une cinquantaine de personnes en Asie, mais qui constitue surtout une source potentielle de recombinants pouvant entraîner une pandémie. Quant on connaît la fréquence de la circulation de souches hautement ou faiblement pathogènes du virus grippal aviaire parmi l'avifaune, on mesure, là aussi, la pérennité du risque potentiel pour l'Homme. La situation dans plusieurs pays d'Asie demeure très préoccupante, avec l'isolement du virus grippal aviaire H5N1 chez le porc, espèce candidate pour faciliter la recombinaison entre souches d'origine humaine et d'origine aviaire ;

- le syndrome respiratoire aigu sévère (Sras) dont l'origine zoonotique est plus que probable.

2.2 Pour les animaux domestiques

Compte tenu de l'évolution de la situation épidémiologique des maladies infectieuses dans le monde, on peut affirmer que le risque représenté par la faune sauvage à travers les animaux domestiques, a augmenté de façon relative, et que, probablement, il continuera d'augmenter dans l'avenir.

Cette situation est due à au moins deux facteurs.

1. il est plus facile de maîtriser une maladie chez les animaux domestiques que chez les animaux sauvages.

Différents exemples, tant de zoonoses que de maladies spécifiquement animales, permettent d'illustrer cette affirmation. Ainsi, par exemple, en France, la tuberculose et la brucellose bovines sont en cours d'éradication grâce à des actions commencées il y a 30 à 40 ans. On peut penser que dans un petit nombre d'années, dans cette espèce animale, ces deux maladies auront complètement disparu. Mais, le risque du développement d'un réservoir sauvage des agents de ces deux maladies se précise, en particulier à cause du sanglier. En effet, les populations de sangliers ont beaucoup augmenté au cours des dernières années en France ; l'isolement du bacille tuberculeux a été réalisé sur ces animaux ; par ailleurs près de 30 % des sujets ont une réponse positive en sérologie de la brucellose.

La peste porcine classique a été éliminée chez le porc en France et dans différents pays européens. Elle persiste dans un réservoir sauvage (le sanglier) en France et en Allemagne notamment, menaçant à tout moment de resurgir chez le porc.

De même, la maladie d'Aujeszky est une maladie en voie de disparition en France chez le porc domestique. En revanche, elle se développe chez les sangliers sauvages.

Les maladies transmises par arthropodes, comme les arboviroses, notamment à partir de la faune sauvage, demeurent très difficiles à maîtriser ; l'exemple de l'infection par le virus West-Nile, sévissant périodiquement dans le sud de la France et qui, en quatre ans, a envahi l'ensemble des États-Unis, l'illustre bien.

En résumé, grâce à des actions de lutte, diverses maladies animales majeures sont maîtrisées ou en cours d'éradication chez les animaux domestiques, mais demeurent enzootiques dans la faune sauvage qui constitue un réservoir dangereux et permanent pour les animaux domestiques et, à travers eux, pour l'homme en ce qui concerne les zoonoses.

2. certaines espèces sauvages constituent un réservoir d'agents pathogènes non encore identifiés ou récemment reconnus.

Dans ce domaine, on peut citer les virus Hendra et Nipah, dont le réservoir semble être constitué par des chauves-souris, mais aussi le Sras dont le réservoir animal est en cours d'élucidation.

Il est certain que d'autres virus circulent de façon silencieuse au sein de diverses populations de la faune sauvage et qu'ils seront révélés accidentellement dans l'avenir par apparition de maladies graves chez l'homme ou chez des animaux domestiques.

En résumé, le danger animal pour l'homme est à considérer avec beaucoup d'attention, notamment celui de la faune sauvage car ce dernier est plus difficile à maîtriser.

3 | Unicité du vivant

Du point de vue biologique, l'homme n'est qu'une espèce animale parmi d'autres.

Nous avons rappelé précédemment que la plupart des grandes alertes sanitaires des vingt dernières années ont été dues à des maladies animales transmises directement à l'homme (encéphalite spongiforme bovine, grippe aviaire, virus West-Nile, fièvre hémorragique Ebola, Lassa, Marburg. . .) ou après une contamination initiale (sida).

Au plan du pouvoir pathogène des agents infectieux et parasitaires, les mêmes mécanismes gouvernent le spectre zoologique d'espèces réceptives d'un même agent chez les primates (comprenant l'homme) ou dans n'importe quel groupe d'espèces animales.

Au plan du diagnostic, du dépistage des maladies et de l'étude de leurs agents, les mêmes méthodes sont utilisées, que l'homme soit inclus ou non dans la liste des espèces concernées.

Au plan de l'épidémiologie, les modes de transmission des agents pathogènes sont semblables, indépendamment de l'espèce concernée, même si leur fréquence peut différer ; ainsi, les toxi-infections alimentaires collectives (TIAC) atteignent essentiellement l'homme, mais la transmission de l'ESB chez les bovins résulte d'un mécanisme épidémiologique tout à fait comparable.

Enfin, les outils épidémiologiques utilisés en épidémiologie animale sont les mêmes que ceux utilisés en épidémiologie humaine.

Les particularités de la pathologie animale résident essentiellement dans deux points :

- d'une part, la richesse et la diversité de la pathologie animale, rendue encore plus complexe par le grand nombre d'espèces pouvant être atteintes. Les vétérinaires connaissent bien cette pathologie ;
- d'autre part, les méthodes de lutte utilisables et l'opportunité de leur mise en place en fonction du rapport coût/bénéfice et de l'efficacité prévisible. Ici encore, les connaissances vétérinaires sont précieuses et indispensables pour toute prise de décision rationnelle dans la lutte contre les zoonoses.

4 Pistes de réflexion

- Le danger permanent de la **faune sauvage** pour l'homme, directement ou à travers des animaux domestiques ou des arthropodes, devrait conduire à un accroissement de l'attention, de la réflexion, de l'épidémiosurveillance et des actions de recherche orientées vers ce secteur du vivant. Ces actions en amont par rapport à la pathologie des animaux domestiques et de l'homme ont une importance capitale pour la protection de la santé animale et de la santé publique.
- Les actions d'épidémiosurveillance ou de recherche en épidémiologie de **maladies humaines à caractère zoonotique** devraient se faire en parfaite concertation avec des vétérinaires épidémiologistes ou en les intégrant dans les équipes médicales.
- Compte tenu de l'unicité du vivant, les actions de recherche en épidémiologie ou d'épidémiosurveillance des **maladies humaines sans caractère zoonotique** pourraient bénéficier de l'apport d'épidémiologistes à formation initiale fondée sur le vivant, par exemple, celle acquise par les vétérinaires.
- Par analogie avec l'existant dans des pays étrangers, comme la Grande-Bretagne par exemple, un **comité national des zoonoses** pourrait être

un lien de réflexion concertée, de rassemblement et de diffusion d'informations ainsi que de propositions d'actions de recherche, d'épidémiologie-surveillance ou de lutte contre les zoonoses. Une telle structure pourrait utilement être complétée par un groupe pluridisciplinaire par zoonose ou groupe de zoonoses voisines. À cet égard, à l'image des CDC américains (Centers for Disease Control and Prevention), il convient de se réjouir de la création prochaine d'un Centre européen de lutte contre les maladies transmissibles.

- La formation des médecins en facultés de médecine ne doit pas négliger les zoonoses, tant celles courantes que celles représentant des menaces sanitaires majeures.

En résumé, comme l'a rappelé récemment l'Académie nationale de médecine, la veille sanitaire, pour être efficace, doit s'exercer non seulement chez l'homme, mais en amont, sur l'animal, qu'il soit domestique ou sauvage (*Bull Acad natle Med*, 2003, **187** (6) : 1201).

Ainsi, il y aurait intérêt à associer systématiquement aux centres nationaux de référence impliqués dans la surveillance de maladies zoonotiques les systèmes et organismes impliqués dans la surveillance des maladies animales transmissibles, à l'exemple de ce qui est fait dans le cas de la listériose.

CHAPITRE **3**

Sources de données pour l'épidémiologie

SOUS-CHAPITRE 3.1

Les enquêtes de santé, les données démographiques, la mortalité. . .

PIERRE MORMICHE

Introduction

Ce chapitre présente, de façon nécessairement résumée, le système statistique de santé, son organisation et ses sources.

La réflexion est partie d'anecdotes dans lesquelles interviennent deux des principales sources statistiques en santé : les enquêtes de l'Insee¹ dites « Enquêtes décennales sur la santé et les soins médicaux » (enquêtes décennales) et « Enquête handicaps-incapacités-dépendance » (Hid) (*Courrier des Statistiques*, 1996 et 1998).

Au cours de la préparation d'Hid 1998-2001, une association de malvoyants alsacienne formula la question du nombre d'aveugles en France, ainsi que dans son département du Haut-Rhin. La réponse (après consultation de l'Inserm) fut

¹L'Institut national de la statistique et des études économiques, cf. ci-après, § 2.2.1.

un constat de... déficience : on ne savait pas, ou plus exactement on ne disposait ni d'un recensement ni d'une estimation statistique du nombre d'aveugles même au seul niveau national. Cela fait toucher du doigt la faiblesse de nos connaissances en épidémiologie descriptive. C'était certes en 1998, mais a-t-on beaucoup progressé depuis ?

Au cours de l'exploitation de l'enquête décennale de santé 1991-1992, on sortit un résultat inattendu : de 15 à 45 ans, les femmes étaient à chaque âge entre une fois et demie et deux fois plus nombreuses à porter des lunettes ou lentilles que les hommes. Des collègues britanniques établirent à notre demande des tableaux analogues pour leur pays : résultat semblable ! Les résultats furent publiés par l'Insee (*Insee-Première*, 1996), largement repris dans la presse quotidienne... C'était en 1996 ; nous n'avons trouvé trace depuis d'aucune recherche épidémiologique ayant cherché à réfuter, comprendre ou expliquer un tel écart entre genres — ni en France où cette publication parut, ni sur les sites anglo-saxons.

On trouvera donc ci-après une présentation résumée du système statistique de santé — son organisation puis les sources — avec des réflexions suscitées par les expériences précédentes sur ses relations médiocres avec la recherche épidémiologique.

1 | L'organisation du système statistique en santé

1.1 Le Cnis et la coordination des programmes des principaux acteurs

La statistique, née pour le prince afin de dénombrer ses sujets et les richesses de ses terres, est encore avant tout une affaire publique. Elle est depuis 1951 coordonnée par le Cnis (Conseil national de l'information statistique), organisme indépendant² placé « auprès de l'Insee », et « assurant la concertation entre les producteurs et les utilisateurs de statistiques publiques » (site Internet du Cnis : <http://www.cnis.fr>).

Il examine de façon très formalisée les projets d'enquête statistique publique :

- « Une des treize **formations spécialisées** du Cnis, émet un **avis d'opportunité**, attestant que l'enquête correspond à un besoin d'intérêt public et qu'elle ne fait pas double emploi avec d'autres sources statistiques ;

²Ses membres ne sont pas rémunérés, mais leurs éventuels frais de déplacement sont remboursés, le secrétariat assuré par un service de l'Insee, et les locaux de réunion fournis par le ministère des Finances.

- le Comité du label du Cnis délivre un **avis de conformité** au regard des objectifs énoncés : il s'assure de la qualité de l'opération statistique, de son processus, des normes statistiques et comptables et des tests du questionnaire réalisés avant l'enquête. Il vérifie également qu'une concertation a été menée avec les partenaires concernés. » (*ibidem*).

« L'ensemble des enquêtes visées par le Cnis sont inscrites au **programme annuel** publié par un arrêté au Journal officiel. » (*ibidem*).

La compétence du Cnis s'étend à tous les thèmes de l'information statistique, et en particulier à celui de la santé et de la protection sociale, auquel une formation spécifique est dédiée. La Drees³ joue un rôle central dans son animation ; y participent les organismes publics et parapublics producteurs ou utilisateurs de statistiques de santé, les caisses d'assurance maladie, les syndicats et organisations professionnelles, des collectivités ou organismes locaux. On notera que la participation de l'Inserm est parfois importante mais assez irrégulière, celle du CNRS et des universités nulle, bien que les réunions soient peu fréquentes (une ou deux par an).

Par exemple, un groupe de travail mandaté par la formation sur « la connaissance statistique de l'état de santé de la population » a rendu son rapport en septembre 2002 (Cnis, 2002). Ce rapport fournit notamment une analyse thématique des informations recueillies par les diverses sources : enquêtes statistiques (43 enquêtes recensées), cohortes et enquêtes épidémiologiques de l'Inserm (48 sources recensées), registres qualifiés par le Comité national des registres (32 registres recensés). Il comporte une synthèse de l'état des lieux, ainsi qu'un ensemble de préconisations.

Dans la ligne de ce rapport, la formation s'est accordée dans sa réunion du 5 mai 2003 (*cf.* p. 26 de son compte rendu sur le site du Cnis), sur la constitution d'un groupe de travail spécifiquement consacré à la refonte des enquêtes relatives à la santé, qui devra « réfléchir à la construction d'un dispositif harmonisé s'appuyant, à court terme, sur l'enquête SPS du Credes⁴ et, dans un cycle quinquennal, sur les enquêtes en santé et la suite des enquêtes Hid. Nous souhaitons ici nous inscrire dans un schéma d'harmonisation européenne dont les premières enquêtes devraient avoir lieu en 2006-2016. (... dans ce cadre) nous ne pourrions plus limiter notre observation aux ménages mais nous devons, au contraire, l'étendre aux institutions. »

³Direction de la recherche, des études, de l'évaluation et des statistiques du ministère de la Santé et de la protection sociale (*cf.* paragraphe 2.2.2 ci-après).

⁴Centre de recherche, d'études et de documentation en économie de la santé (devenu en 2004 « Institut de recherches » : Irdes).

1.2 Un système néanmoins assez éclaté, comprenant des acteurs dédiés à l'information

1.2.1 L'Insee : un rôle important, une affectation de moyens aux statistiques de santé... symbolique

L'Insee, Institut national de la statistique et des études économiques, est le noyau du système statistique public en France. Parmi ses équivalents étrangers on citera l'Office for National Statistics anglais (ONS), Statistiques Canada, l'Instituto Nazionale di Statistica italien (IStat) ou encore l'Australian Bureau of Statistics (ABS).

Il compte quelques 6 500 agents, dont 1 800 cadres statisticiens ou gestionnaires, formés pour la plupart par les écoles de statistiques relevant de l'Institut; il faut ajouter 675 agents détachés dans des services statistiques de ministères ou autres organismes du secteur. Il s'appuie en outre sur un réseau de 800 enquêteurs professionnels; enquêteurs généralistes, ceux-ci ont pu assurer par exemple avec succès les interviews des enquêtes « Hid Prisons » auprès des détenus, ou de l'enquête « Sans domicile » auprès des SDF.

Très différente de celle pratiquée dans les EPST, la gestion des carrières est dominée par un souci de mobilité (3 ans en principe dans les postes de début de carrière et rarement plus de 5 ans ensuite). Elle vise à construire des profils de pluricompetences qui homogénéisent l'institution, mais ont certainement comme contrepartie l'absence de connaissance approfondie dans des champs (la santé par exemple) autres que la méthodologie statistique.

La politique de diffusion des données — y compris des bases de données individuelles anonymes — s'est récemment ouverte de façon rapide. À ce titre par exemple, les fichiers « détail » des enquêtes auprès des ménages ou individus, qui sont actuellement disponibles pour les chercheurs par l'intermédiaire du Centre Quételet, devraient être accessibles librement sur le site Internet de l'Institut avant la fin de l'année 2004.

L'Insee joue aussi un rôle méthodologique assez important en statistique de par ses écoles de formation. Actuellement, environ quatre sur cinq de leurs diplômés ne sont pas destinés à entrer dans l'Institut lui-même ou dans la fonction publique statistique en général.

Ce rôle comporte cependant un certain nombre de points faibles notamment dans le domaine de la santé. Ainsi l'Insee a conclu en 1998 une convention générale avec l'Inserm. Elle prévoyait entre autres un effort de développement des

échanges de personnels avec d'un côté plus de possibilités d'accueil de diplômés des écoles de l'Insee (on notera que l'Ensaï, École nationale de la statistique et de l'analyse de l'information, comporte une option « biostatistique » et une option « économie de la santé »), et aussi en retour des accueils temporaires de chercheurs Inserm associés à la conception et à la mise en œuvre d'enquêtes spécifiques ou d'ingénieurs de recherche comme enseignants ; les résultats sont demeurés limités⁵.

D'ailleurs, concernant les statistiques de santé, l'Insee s'écarte volontairement au profit de la Drees. Il n'intervient que de deux façons : soit en montant des enquêtes en population que la Drees n'a pas les moyens propres (en terme de réseau d'enquêtes ou de méthodologie) de prendre en charge — c'est classiquement le cas de l'enquête décennale sur la santé, et désormais de l'enquête Hid — ; soit, comme cela se fait de plus en plus, en intégrant à des enquêtes générales des modules sur tel ou tel champ de santé. Cela pose au moins deux types de problèmes :

- d'une part, bien que disposant de sources importantes, l'Insee n'a en temps ordinaire que de très peu d'agents travaillant sur le domaine de la santé (y compris la mortalité par cause). Compte tenu du principe de mobilité rappelé ci-dessus, il est assez clair qu'aucune accumulation de compétence spécifique au domaine n'est réalisable dans ces conditions ;
- d'autre part, les organes statistiques européens se font en principe une règle de traiter avec les instituts nationaux de statistiques. Par exemple, c'est avec l'Insee qu'ils ont discuté la mise sur pieds puis la révision de la série des enquêtes dites « panel européen », y compris leur module sur la santé ; il en sera sans doute de même pour le gros investissement qu'ils prévoient dans la conception et le démarrage d'une enquête européenne sur la santé. Il y a donc toujours une difficulté française à soutenir sur ces questions la discussion comme il le faudrait.

1.2.2 La Drees : une place récente mais centrale

La Drees, Direction de la recherche, des études, de l'évaluation et de la statistique, (Detape, Lacroix, 2001 ; Journal officiel, 1998) est sans doute depuis début 1999 la grande nouveauté du système statistique de santé publique. En regroupant l'ensemble des missions de recherche et d'observation statistique jusque-là assez éparpillées dans le ministère chargé de la Santé et de la Solidarité, elle a pu tout à la fois donner une cohérence à l'appareil de connaissance

⁵Ainsi le « Comité de coordination » constitué par la convention, et qui devait se réunir deux fois par an, n'a tenu à ce jour aucune session.

du ministère et le dégager en bonne part de l'emprise d'objectifs purement gestionnaires.

La concentration des moyens et l'accroissement des budgets lui ont permis de lancer, conduire ou appuyer de nombreuses enquêtes ; elle a ainsi pu doubler les sous-produits statistiques de la gestion du système (par exemple les données relatives à l'allocation personnalisée à l'autonomie — Apa) par des enquêtes autonomes s'intéressant directement à la santé des populations concernées (par exemple sa forte participation dans le financement et l'exploitation d'Had) ou à l'effet des aides sur leurs conditions de vie (enquête auprès des bénéficiaires de l'Apa).

La Drees intervient par deux autres voies dans la recherche en santé. D'une part, son regroupement avec la MiRe (Mission de Recherche) lui permet de financer d'assez nombreux appels d'offres ou conventions débouchant sur des travaux menés par des équipes de recherche. D'autre part, sa position institutionnelle l'associe à la tutelle de plusieurs des organismes de recherche du domaine : CTNERHI⁶, Ined, Inserm... et ses missions comportent explicitement une coordination avec l'Insee, l'InVS⁷ et les ORS⁸ (Journal officiel, 1999).

Forte de quelques 170 agents, dont quelques médecins et près d'un tiers de statisticiens détachés de l'Insee, la Drees connaît, comme ce dernier, une forte mobilité de la majeure partie de son personnel. Les inconvénients pour l'accumulation d'expérience en sont sans doute plus limités, compte tenu de la concentration de ses travaux sur les champs de la santé et la solidarité, mais ils peuvent néanmoins peser, par exemple sur la participation aux travaux et discussions internationaux, comme le suivi de la mise en place de statistiques européennes de santé ou le suivi de la révision de la classification internationale des handicaps et de son application dans les enquêtes.

1.2.3 L'Inserm et la mortalité par causes

Outre son rôle épidémiologique central, l'Inserm tient naturellement une place déjà décrite dans les statistiques de santé publique (cf. chapitre 2).

Il est notamment depuis 1968 le producteur des statistiques de mortalité par cause (établies par le Centre d'épidémiologie sur les causes médicales de décès : CEPIDC), sur lesquelles on reviendra en 3.2.1 ci-dessous.

⁶Centre technique national d'études et de recherches sur les handicaps et les inadaptations. Organisme sous la tutelle de la Das et de la Drees, très proche des associations de personnes handicapées.

⁷Institut de veille sanitaire.

⁸Observatoires régionaux de la santé.

1.3 Des questionnaires pour qui l'information statistique et son analyse demeurent très secondes

On trouvera dans une autre partie de cet ouvrage (cf. sous-chapitre 3.2), à propos du PMSI, une description de sous-produits statistiques de la gestion de l'appareil de santé et de leurs limites.

Ces limites relèvent d'un rapport plus général des questionnaires à la prise en compte des impératifs — et des bénéfices potentiels — de l'information et de la recherche, aujourd'hui en France. En témoigne dans le champ spécifique du handicap la situation que stigmatisait le rapport du Cnis de 1997 (Cnis, 1997) : il y estimait en effet à plus de 180 milliards de francs de l'époque le coût du handicap et de la dépendance, tout en constatant la quasi-nullité du budget public de recherche correspondant.

De son côté, la Cour des comptes évaluait plus récemment le montant de « l'effort public en faveur des personnes en situation de handicap » à 26 milliards d'euros en 2001 (Cour des comptes, 2003, page 50) — ceci ne comprenant pas les aides aux personnes âgées dépendantes. Simultanément, une évaluation, qualifiée de très approximative, des financements publics de la recherche sur le handicap s'établit dans le même rapport à 120 millions d'euros en 2001. Soit un taux d'investissement de l'ordre de 4,5 pour mille.

Ce manque d'attention aux bénéfices potentiels des investissements de recherche reste assez dominant. Un aspect fréquent est l'absence de continuité dans l'appui aux chercheurs et dans leur consultation. Deux exemples : on a vu la Cnamts investir assez largement dans le financement de la grosse enquête Insee sur le handicap (Hid : 3 millions de francs de l'époque), puis s'abstenir de toute exploitation des fichiers, que ce soit par elle-même ou par des commandes passées à des unités de recherche spécialisées — travaux qui lui auraient sans doute coûté moins du dixième de la somme investie dans la constitution de la base de données. Autre exemple, celui de conseils généraux ayant financé des extensions locales d'échantillons, puis ne s'étant plus préoccupés d'obtenir le moindre tableau de résultats... alors que d'autres, qui n'avaient pas jugé utile de le faire, se sont quatre ou cinq ans plus tard lancés dans des travaux beaucoup plus incertains pour tirer de l'enquête nationale des estimations départementales.

1.4 Un système demeuré longtemps isolé au plan international

La plupart des organismes cités entretiennent de nombreux rapports internationaux avec les chercheurs ou organismes de leur spécialité. C'est trop

manifeste pour avoir besoin d'exemples dans les cas de l'Ined ou de l'Inserm ; le CTNERHI a, malgré ses faibles moyens, joué un rôle essentiel dans le suivi de la réforme de la Classification internationale des handicaps (CIH) ; le Credes et la Drees réalisent de nombreuses études comparatives internationales, notamment sur l'organisation des systèmes de santé ou d'aide sociale ; l'Insee participe aux instances de concertation statistique européennes, aux fédérations internationales de statisticiens. En outre, des chercheurs de l'Ined et de l'Inserm ont participé activement — à titre « individuel » — à de nombreuses actions statistiques européennes (par exemple à la définition et l'harmonisation des calculs d'espérances de vie en santé dans le cadre « d'Euro-Rêves »). . . Et pourtant la situation est peu satisfaisante. Pourquoi ?

On a proposé ci-dessus une explication : à savoir que les principales enquêtes statistiques en santé publique étaient réalisées par l'Insee qui, trop peu investi jusqu'à présent dans la réflexion sur la santé, n'avait ni les moyens ni la compétence pour participer dans ce domaine aux groupes de travail et commissions internationales.

Le résultat a été que la participation française à l'élaboration des indicateurs ou des opérations statistiques européens en santé a souvent été insuffisante. Sur des points où l'expérience ou la compétence globales de nos acteurs nous auraient permis de rectifier des questionnements ou des normes inadéquates, faute d'avoir été présents lors de leur discussion, nous avons dû les intégrer et les appliquer.

Jusqu'à présent, les remontées de données nationales au niveau européen consistaient principalement à fournir à Eurostat des « indicateurs » d'état de santé ou de divers facteurs de risque (tabagisme par exemple) ; la définition des indicateurs devait être conforme aux définitions d'Eurostat, mais les méthodes d'élaboration, les sources, voire la rédaction des questions et des items de réponse étaient laissées à l'appréciation des pays membres.

Ceci a commencé à changer et évoluera probablement vers un système d'enquêtes européennes sur la santé. On passera ainsi d'une étape d'harmonisation à des outils relativement standardisés. Dès 2006-2007, ces enquêtes ou dans un premier temps ces modules d'enquêtes nouveaux entreront en concurrence avec nos propres outils.

Si l'on ne peut qu'approuver ces objectifs, le mode d'élaboration des outils laisse craindre de fortes pertes de qualité. Choix des objectifs, méthodes et des questionnaires sont en effet déterminés par des consensus d'experts et très rarement soumis à une validation expérimentale. Cette stratégie de consensus n'est pas précisément la caractéristique habituelle de la démarche scientifique ; mais à n'y intervenir que trop peu — pour ce qui concerne le domaine des statistiques

de santé — les instituts français ne peuvent espérer développer ni faire entendre de propositions alternatives.

Or s'agissant d'enquêtes qui mobilisent des centaines d'enquêteurs et des millions d'euros, il est peu probable qu'il y ait place pour deux systèmes concurrents.

2 | Quelques sources de données statistiques en santé publique

Comme indiqué plus haut, on trouvera une vue d'ensemble des sources statistiques sur l'état de santé de la population dans le rapport du Cnis (Cnis, 2002).

Ce paragraphe présente les principales enquêtes de santé et quelques-unes des statistiques de mortalité.

2.1 Des enquêtes de santé « généralistes »

La France dispose de trois grandes enquêtes en population, représentatives de la population vivant en métropole, au moins en domicile ordinaire. Elles couvrent l'ensemble du champ de la santé tel qu'il est actuellement défini par l'OMS : l'accès aux soins médicaux, les pathologies et les différentes dimensions du handicap — y compris la participation sociale. Elles fondent dans tous ces domaines de nombreuses estimations de prévalence relatives à la santé.

2.1.1 Les enquêtes décennales sur la santé et les soins médicaux de l'Insee

C'est l'une des plus anciennes sources disponibles. Après une première version datant de 1959, trois éditions ont été menées avec une méthodologie quasiment inchangée en 1970-71, 1980-81 et 1991-92. L'édition 2002-2003 a apporté de nombreuses innovations, tout en préservant la comparabilité avec les précédentes.

Cette source a longtemps été délaissée par les épidémiologistes, pour deux raisons concourantes. Un manque de communication, tout d'abord (dont la responsabilité est partagée entre les divers acteurs), entraînant l'ignorance pure et simple de l'existence de certaines sources — ou la conviction qu'elles n'étaient pas accessibles. Une réserve assez partagée ensuite à s'engager dans l'exploitation d'une morbidité déclarative, dont on sait les biais qu'elle peut entraîner comme source d'évaluation de la morbidité « objective » ; réserve suffisamment

forte pour qu'un directeur général de l'Inserm intervenant ès qualité dénonce dans les discussions du Cnis « le gaspillage que constituerait la réalisation de telles enquêtes ». Il nous semble que cette réticence devrait pouvoir être aujourd'hui dépassée, pour trois raisons convergentes :

- la méthode de base de l'enquête, très attentive à laisser le moins d'espace possible aux oublis. Ceci en utilisant d'abord classiquement un questionnaire sur une liste complète de types de pathologies ; puis en multipliant les allers-retours entre les recours aux soins et leurs motifs, tout au long d'un suivi des personnes s'étalant sur 3 mois (période ramenée à 2 mois en 2003) ; enfin en appuyant les interviews par la relecture des réponses par une équipe médicale — qui réalise après l'enquête le codage exhaustif des pathologies ;
- la forte médicalisation de notre société depuis une trentaine d'années, qui en multipliant les recours a certainement réduit la part des pathologies ignorées et rapproché la morbidité déclarée de la morbidité diagnostiquée, et celle-ci de la morbidité « diagnosticable » ;
- les évolutions intervenues dans la dernière édition de l'enquête, en particulier celle qui a développé des modules spécifiques consacrés à un certain nombre de grandes pathologies, et celle qui a proposé à une partie des personnes interrogées des bilans de santé dans les centres d'examens de la Sécurité sociale.

On notera qu'un questionnaire relatif au handicap conçu par des chercheurs de l'Inserm avait été intégré dès 1980. Repris en 1991 (et 2002), il a fourni la seule base nationale disponible en France à ce jour pour estimer l'évolution des espérances de vie en santé, plus précisément des espérances de vie sans incapacités (EVSI) ; ses résultats ont eu un fort retentissement, en établissant avec une bonne certitude — et contrairement aux thèses internationales dominantes au début des années 1990 — que les années de vie gagnées (rappelez-vous que l'espérance de vie globale croît en France depuis plusieurs décennies au rythme de 2,5 années tous les dix ans) étaient des années sans incapacités (Robine, Mormiche, 1993). Ces résultats sont notamment utilisés aujourd'hui comme base de référence dans les travaux de projection des nombres de personnes âgées dépendantes.

2.1.2 Les enquêtes santé-protection sociale du CreDES

Le CreDES a mis en place depuis 1988 l'enquête Santé et protection sociale (SPS), à la demande et en collaboration avec la Cnamts, rejointe depuis par la Canam, la MSA, la FNMF, la Drees plus récemment et l'InVS.

Elle est réalisée à partir de l'échantillon permanent d'assurés sociaux de France métropolitaine (Epas). Son principal objectif est de décrire et d'analyser les relations entre consommation de soins, état de santé, caractéristiques socio-économiques et couverture sociale. Menée d'abord tous les ans, puis tous les deux ans, elle associe la méthode de l'interview par téléphone ou par déplacement d'enquêteurs à celle des questionnaires auto-administrés. Depuis 1988, sur dix ménages contactés, six ont accepté de participer à l'enquête, ce qui représente environ 7 000 ménages et 20 000 individus. Parmi ceux-ci, certains ont été interrogés quatre fois en quatorze ans.

Il s'agit en effet de la seule opération, menée actuellement en France, qui permette d'associer de façon individuelle, pour une majorité d'enquêtés, les informations déclarées aux consommations présentes dans les fichiers de l'assurance-maladie. Sa courte périodicité lui confère une réactivité permettant d'aborder des questions conjoncturelles, soit dans le but de contribuer à l'évaluation de politiques de santé, soit pour apporter des éléments de connaissance sur les comportements de soins des Français et leurs opinions concernant le système de santé. Ainsi, SPS a par exemple contribué à informer les pouvoirs publics sur la proportion de personnes renonçant à des soins pour raisons financières. Elle alimente, avec d'autres, la réflexion sur la Couverture maladie universelle (CMU) et participe actuellement à son évaluation.

Comme les enquêtes décennales, SPS ne couvre que les personnes habitant en France métropolitaine en domicile ordinaire.

2.1.3 L'opération Handicaps-incapacités-dépendance de l'Insee

Les réflexions menées dans le champ du handicap au cours de la décennie 1990 ont toutes conduit, de façon récurrente, au même double constat : celui du retard de la France dans le domaine des statistiques nationales sur le handicap, l'incapacité et la dépendance, et celui de l'hétérogénéité et de l'insuffisance de l'appareil d'information sur ces sujets. « L'amélioration de l'information statistique sur le handicap » a été soulignée comme indispensable par le rapport du groupe de travail *ad hoc* du Cnis en 1997 (Cnis, 1997).

Ceci a abouti en 1998 à l'enquête « Handicap, incapacités, dépendance » (Hid) qui a été réalisée pour toutes les personnes vivant à domicile ou en institution, en prenant appui pour la partie Ménages sur le « Recensement général de la population française de mars 1999 ».

L'architecture, la méthode et les questionnaires de l'enquête ont été élaborés par un Groupe de projet, en charge d'épauler l'Insee dans la préparation et comprenant des chercheurs de l'Ined, de l'Inserm (IFRH et IFRV⁹), du Credes, de la Drees et du CTNERHI.

Hid recouvre un ensemble d'enquêtes de trois types : auprès des personnes handicapées en institution (deux passages fin 1998 et fin 2000) ; auprès des ménages (deux passages fin 1999 et fin 2001), précédée d'une phase de filtrage réalisée en mars 1999 (questionnaire « Vie quotidienne et Santé » rempli pour 360 000 personnes) ; auprès de la population retenue en milieu carcéral, réalisée par l'Ined en mai 2001 (1 284 répondants).

Elle a donné lieu à plusieurs colloques de recherche et un colloque de « restitution » à l'intention des acteurs du domaine du handicap (RFAS, 2003). Une centaine d'articles, études ou rapports ont été publiés à ce jour. On trouvera une documentation abondante la concernant sur le site Web de l'IFR « Handicap » (site Hid).

L'Inserm, en tant qu'institution, a apporté son soutien à cette initiative et des chercheurs des IFRH et IFRV ont joué, comme on l'a indiqué, un rôle primordial dans la conception des questionnaires et de la méthodologie.

Cependant, comme le rapportait la Cour des comptes citée plus haut, on ne peut s'empêcher de s'interroger sur la prise en compte par la recherche épidémiologique du tournant pris par l'OMS dès le début des années 1980 dans sa conception de la santé. Depuis cette date en effet, au-delà des pathologies, les aspects lésionnels, fonctionnels et sociaux, dont l'importance dans la vie des populations ne saurait être trop soulignée, ont été inscrits en tant que tels dans le champ de la santé. L'adoption successive de la Classification internationale du handicap (CIH) puis du fonctionnement (CIF) en font foi.

La statistique française a certes été longue à prendre en compte cette évolution, puisque l'enquête Hid n'est intervenue que dix-huit ans après l'adoption de la CIH. Cependant, la statistique n'est en cette matière au moins que descriptive ; on attend donc avec une certaine impatience que la recherche épidémiologique investisse ce domaine « nouvellement » reconnu de ses capacités analytiques et explicatives. Or, on peut craindre que son engagement en ce sens ne soit encore ni entamé, ni même explicitement envisagé. Une discussion approfondie de la pertinence d'un tel élargissement et des conditions de son développement est aujourd'hui nécessaire.

⁹ Respectivement Institut fédératif de recherches sur le handicap, et sur le vieillissement.

2.2 La mortalité et la mortalité par causes : quelques sources méconnues

2.2.1 Les délais

L'état civil permet de disposer de statistiques *a priori* exhaustives puisque chaque naissance, mariage et décès en France fait l'objet d'un acte établi par l'officier d'état civil de la commune. Le nombre élevé de ces dernières et leur taille parfois très réduite expliquent la forme de la transmission à l'Insee, qui n'est encore que pour partie télématique, d'où des délais de transmission-saisie-contrôles assez élevés : des statistiques provisoires, par sexe, âge et regroupements de communes, sont disponibles à la fin de l'année $n + 1$, les statistiques définitives (par commune détaillée) comportent un délai de 1 an et demi.

Grâce à l'enquête « villes » décrite plus loin, un bilan démographique annuel national est dressé chaque année en janvier $n + 1$, qui présente le volume et une première analyse des mouvements démographiques (naissances, mariages, décès) de l'année écoulée (Doisneau, 2003).

Les causes de décès figurent sur un bulletin anonyme rempli par le médecin constatant le décès. Il est transmis au CEPIDC de l'Inserm via la direction départementale de l'Action sanitaire et sociale (Ddass) ; le CEPIDC entreprend d'une part le chiffrage des causes de décès selon la Cim 10, et d'autre part un appariement avec le fichier anonymisé reçu de l'Insee au cours de l'année $n + 1$, afin de contrôler la complétude et l'identité des deux procédures de transmission.

Cette comparaison permet à l'Insee de compléter le fichier des décès, et de produire un fichier anonyme définitif sans les causes, au second semestre de l'année $n + 2$. Quant au chiffrage des causes de décès, il aboutit à la production d'un fichier anonyme avec causes de décès actuellement seulement disponible au second semestre de l'année $n + 3$.

Ce sont naturellement les fichiers de décès par causes qui intéressent en premier lieu l'épidémiologie. Pour le long terme, ils ont par exemple permis aux chercheurs de l'Ined d'établir des séries longues (depuis 1925) détaillées et cohérentes par-delà les nombreuses révisions de la Cim (Vallin, Meslé, 1988 ; Meslé, Vallin, 1995).

Au total, les études françaises sur ce thème semblent néanmoins peu nombreuses : moins d'une vingtaine de recherches originales, dont la moitié concernant des comparaisons interrégionales (Salem *et al.*, 2000).

2.2.2 Court terme et enquête « Villes »

L'enquête « Villes » est une enquête mensuelle de comptage des naissances, mariages et décès auprès d'un échantillon de communes. Elle permet ainsi de produire chaque mois une estimation provisoire des mouvements enregistrés en France métropolitaine au cours du mois précédent. Ces données sont publiées dans le *Bulletin mensuel de statistiques*. Elles sont également accessibles sur le site Internet de l'Insee (www.indices.insee.fr).

Elle est réalisée par l'Insee depuis 1946. Elle couvre environ les deux tiers des naissances, 40 % des décès et un quart des mariages. L'extrapolation à la France métropolitaine est réalisée sur la base des résultats de l'état civil des années précédentes.

Suite à la canicule de l'été 2003, un système de transmission des décès d'un échantillon de 140 grandes communes (informatisées) est en cours de mise au point, qui permettra de transmettre en quelques jours à l'InVS un volume quotidien des décès.

2.2.3 Coup de projecteur sur les échantillons de mortalité et l'échantillon démographique permanent (EDP)

Des sources initiées par l'Insee permettent d'étudier la mortalité selon des critères variés, et les caractéristiques et trajectoires sociodémographiques. Il s'agit d'échantillons spécifiques, tirés du recensement de la population, auxquels on adjoint les événements ultérieurement enregistrés par l'état civil. Ils cumulent les renseignements sociaux recueillis au recensement (niveau d'études, activité, profession, migrations récentes, immigration. . .), beaucoup plus complets et de qualité beaucoup plus assurée que ceux figurant dans les bulletins de décès, avec les renseignements démographiques.

On citera d'abord le plus ancien : l'échantillon tiré du Recensement de 1954, constitué de 450 000 hommes de 30 à 69 ans et de leurs épouses éventuelles, a permis d'établir la hiérarchie sociale des taux de mortalité en France sur la période 1955-1979. Sa confrontation à l'échantillon tiré du recensement de 1975 a permis d'établir l'évolution des inégalités sociales de 1954 à 1980.

De même ont été constitués à partir du recensement de 1999 deux échantillons (« Grands âges », 150 000 personnes de plus de 85 ans et « Très grands âges », 80 000 personnes de 93 ans et plus) dédiés à l'étude de la mortalité aux grands âges, qui reste encore aujourd'hui très mal modélisée. L'étude des causes de décès, celle des écarts sociaux à ces âges devrait être fécondes.

Enfin, comme l'ONS britannique, l'Insee a construit et entretient depuis 1968 un échantillon démographique dit « permanent », qui met en série les réponses individuelles aux recensements successifs pour une sous-population d'environ 600 000 personnes, et les enrichit des événements d'état civil. Source unique, mais d'un accès très encadré (pour le respect des conditions d'anonymat). . . il est peu exploité (près de 100 fois moins de travaux que sur son équivalent anglais). Est-ce l'effet d'un manque d'information ou de moyens, du cloisonnement entre organismes, des conditions d'accès à la source. . . ?

Bibliographie

- Beaumel C., Doisneau L., Vatan M. (2003). La situation démographique en 2001 — Mouvement de la population, *Insee-Résultats, série Société*, n° 18, septembre 2003. *Disponible sur le Web à l'adresse suivante : <http://www.insee.fr/fr/ppp/collections/>*
- Cnis (1997). Handicap et dépendance : l'amélioration nécessaire du système statistique, *Rapport du Cnis*, n° 35, juillet 1997.
- Cnis (2000). Actes de la journée d'études organisée par le Cnis et la Drees sur l'observation statistique en matière de santé, *Rapport du Cnis*, n° 60, juin 2000. *Disponible sur le Web à l'adresse ci-dessus.*
- Cnis (2002). Connaissance statistique de l'état de santé de la population, *Rapport du Cnis*, n° 76, septembre 2002.
- Cour des comptes (2003). *La vie avec un handicap*, Rapport public particulier, juin 2003.
- Courrier des Statistiques* (1996). L'enquête décennale sur la santé 1991-1992 : bilan de la réalisation et des apports. n°s 79-80, décembre 1996.
- Courrier des Statistiques* (1998). L'enquête HID de l'Insee : objectifs et schéma organisationnel, n°s 87-88, décembre 1998. *Un site web fournit en ligne la documentation de l'enquête (questionnaires, articles méthodologiques, dictionnaires des fichiers) et la liste des travaux et études menés à partir de HID : <http://ifr-handicap.inserm.fr>*
- Detape Y., Lacroix J. (2001). Le système français de statistique publique : les services statistiques ministériels. *Courrier des statistiques*, n°s 98-99, septembre 2001, pages 46-49. *Disponible sur le Web à l'adresse suivante : <http://www.insee.fr/fr/ppp/collections/>*
- Doisneau L. (2003). Bilan démographique 2002 : légère diminution des naissances. *Insee-Première*, n° 882, janvier 2003. *Disponible sur le Web à l'adresse suivante : <http://www.insee.fr/fr/ppp/collections/>*

- Insee-Première* (1996). Les lunettes à la loupe : les femmes portent plus souvent des lunettes que les hommes. n° 475, juillet 1996. Disponible sur le Web à l'adresse précédente.
- Journal officiel (1998). Décret n°s 98-1079 du 30 novembre 1998 portant création de la Drees. *Journal officiel*, n° 278, 1^{er} décembre 1998, page 18095. Disponible sur le Web à l'adresse suivante : <http://www.legifrance.gouv.fr/WAspad/RechercheSimpleJorf.jsp>
- Journal officiel (1999). Arrêté du 22 février 1999 portant organisation de la Drees. *Journal officiel*, n° 46, 24 février 1999, page 2835. Disponible sur le Web à l'adresse ci-dessus.
- Meslé F., Vallin J. (1995). *Reconstructing Long Trend Series of Causes of Death : the Case of France*, European Association for Population Studies. Seminar at Urgüp, Turkey, octobre 1995.
- Paris V., Renaud T., Sermet C. (2002). Des comptes de la santé par pathologie : un prototype sur l'année 1998. *Questions d'Économie de la santé*, n° 56, septembre 2002 et *Rapports du Credes*, n° 1480, mai 2003.
- RFAS (2003). Handicaps, Incapacités, Dépendance : actes du colloque d'octobre 2002. *Revue française des affaires sociales*, vol. 57, n°s 1-2, janvier-juin 2003.
- Robine J-M., Mormiche P. (1993). L'espérance de vie sans incapacité augmente. *Insee Première*, n° 281, octobre 1993. Disponible sur le Web à l'adresse suivante : <http://www.insee.fr/fr/ppp/collections/>
- Salem G., Rican S., Jouglu E. (2000). *Atlas de la santé en France*, Volume 1 — Les causes de décès. Drees-MiRe, John Libbey, février 2000.
- Site Cnis / Site web présentant les fiches d'enquêtes approuvées par le Comité du Label de 1999 à 2004 : http://www.cnis.fr/ind_doc.htm
- Site HID / Site web présentant la documentation sur l'enquête HID (questionnaires, méthodes, ordre du jour des colloques et liste des publications) : <http://ifr-handicap.inserm.fr>
- Vallin J., Meslé F. (1988). Les causes de décès en France de 1925 à 1978. Avec une étude des variations géographiques, de G. Caselli et V. Egidi. Ined, *Travaux et documents*, Cahier n° 115, 1988.

SOUS-CHAPITRE 3.2

Le recueil de l'information dans les hôpitaux

CATHERINE QUANTIN ET JEAN-MARIE RODRIGUES

Le recueil et la gestion de l'information médicale liée au patient ont pour premier objectif les soins et la continuité des soins. Les informations qui sont rassemblées dans le dossier médical ou le dossier du patient sont de plus en plus utilisées à des fins collectives, telles que la recherche clinique, épidémiologique, économique et évaluative. Toutefois, les informations contenues dans le programme de médicalisation du système d'information (PMSI) et recueillies dans les établissements de soins publics ou privés semblent sous-utilisées au regard de leurs potentialités. Après avoir rappelé les fondements du recueil des informations du PMSI, nous évoquerons les tentatives d'utilisations de ces données en matière d'épidémiologie descriptive clinique en exposant les freins à ces utilisations, puis nous exposerons les perspectives offertes par l'apport des techniques de chaînage qui permettent de raisonner non seulement en termes d'hospitalisation mais aussi en termes de patient, et ainsi d'autoriser le suivi des populations.

1 | Le programme de médicalisation des systèmes d'information (PMSI) dans les établissements de soins publics et privés

Depuis plus d'une vingtaine d'années, les pays industrialisés rencontrent des difficultés financières qui finissent par menacer leur système de santé (Grémy, 2004). Dans cet objectif, il était nécessaire d'estimer les dépenses de santé à partir de classifications spécifiques. À la fin des années 1960, Fetter (Fetter, 1980), a développé le système de classification américain permettant d'obtenir des DRG (Diagnosis Related Group) pour classer les produits hospitaliers et apprécier la formation des coûts de production en les rapportant à la pathologie traitée.

Autrement dit, dans la logique du système de Fetter, l'hôpital est considéré comme une entreprise de production multiple. Hospitalisés pour des affections diverses, les malades nécessitent des soins qui diffèrent selon leurs maladies mais aussi selon leur âge et les moyens diagnostiques et thérapeutiques disponibles. Si chaque malade constitue un cas particulier et implique un séjour dont le coût ne peut être exactement prévu, on peut malgré tout établir des catégories de patients qui présentent des caractères communs, en termes de pathologie, d'âge et de sexe, nécessitent des traitements analogues et entraînent des dépenses du même ordre de grandeur, afin de suivre l'évolution des coûts de l'activité des établissements.

La classification en DRG est utilisée aux États-Unis dans le cadre du système de paiement prospectif du programme d'assurance *Medicare* destiné aux personnes âgées et handicapées ainsi que du programme *Medicaid* (Bright, 1989) pour les personnes démunies. Révisée annuellement depuis 1983, cette classification connaît de nouvelles applications dans d'autres secteurs du système de soins (ambulatoire, réseau de soins, ...). En Europe, ce type de classification a été introduit dans les années 1980 et a été adapté aux spécificités des systèmes de santé européens (Rodrigues, 2001). Entre 1982 et 1994, 14 pays européens ont développé leur propre système d'information afin de collecter et d'analyser les données administratives et médicales d'un établissement de soins. Certains pays comme le Portugal, la Belgique, la Norvège, l'Irlande et l'Italie, utilisent ce système pour financer leurs hôpitaux, en régulant les budgets hospitaliers en fonction de l'activité médicale produite. En France, la « Tarification à l'activité » s'appuie sur ce type de classification pour le financement de ses hôpitaux d'une part et des cliniques privées d'autre part.

Ainsi en France, à partir de résumés de sortie standardisés (RSS), actuellement recueillis pour tous les patients hospitalisés en court séjour, l'algorithme de classification, appelé « fonction groupage », classe les séjours des patients

dans les groupes homogènes de malades (GHM) issus des DRG, en réunissant des cas comparables en termes de pathologie et de mobilisation de ressources hospitalières (Dusserre, 1999). Ce recueil s'appuie sur des variables du résumé d'unité médicale (RUM) renseigné pour chaque passage dans une unité médicale (les RUM correspondant à un séjour hospitalier étant regroupés en RSS) : le diagnostic principal, c'est-à-dire le motif de prise en charge qui a mobilisé l'essentiel de l'effort médical et soignant au cours de l'hospitalisation, le ou les diagnostics associés, l'âge, le sexe et les actes diagnostiques et thérapeutiques.

Disposant d'une classification nationale d'environ 650 GHM, chaque hôpital, compte tenu de la distribution des RSS recueillis, peut décrire son activité médicale en classant ces résumés dans les différents GHM, et déterminer son « *case mix* » ou distribution des cas traités. Jusqu'en 2003 ce *case mix* permettait aux Agences régionales d'hospitalisation (ARH) à partir des RSS rendus anonymes (RSA) de définir l'indice statistique d'activité de chaque établissement afin de distinguer, parmi les établissements dont le coût moyen par cas est très élevé, ceux qui prennent en charge des cas particulièrement complexes de ceux qui traitent les cas de manière très coûteuse. Ces derniers établissements, considérés comme surdotés, faisaient alors l'objet d'une proposition d'écrêtage dans le cadre de la négociation du budget global.

Le projet de tarification à l'activité (T2A) (arrêté 31 décembre 2003, Fessler JM, 2003), mis en place en 2004, a pour ambition d'aller plus loin que la simple modulation des dotations globales de financement des hôpitaux publics et privés participant au service public hospitalier. Il introduit une harmonisation des modes de financement de ce secteur avec celui des cliniques privées, soumises jusqu'alors à un système de tarification à la journée et à un forfait lié aux actes réalisés. Dans son rapport annuel au Parlement sur la Sécurité sociale en septembre 2002, la Cour des comptes exprimait également ces grandes attentes vis-à-vis de ce projet, « car les établissements dont les coûts sont supérieurs au tarif national sont incités à les réduire alors que ceux dont les coûts sont inférieurs en reçoivent les bénéfices ».

2 | Intérêt de l'utilisation des données du PMSI à des fins d'épidémiologie descriptive pour les pathologies chroniques : l'exemple du cancer

Les données de morbidité fournies par les registres et les données de mortalité sont très utilisées pour fournir des estimations d'incidence du cancer au niveau national. Après avoir rappelé les principes et les limites de ces deux modes de recueil, nous essaierons de dégager quel pourrait être l'apport complémentaire des données du PMSI.

Les statistiques de causes médicales de décès font partie des informations de base dans tous les pays du monde pour évaluer le type et l'importance des maladies ou des accidents mortels (Bouyer, 1995). En France, la statistique des causes de décès est actuellement établie à partir des informations recueillies dans deux documents : le bulletin de décès (établi par l'officier d'état civil à la mairie) et le certificat de décès, établi par le médecin (Michel, 2000) selon le modèle proposé par l'OMS (Jouglan, 1998) qui permet de différencier la cause immédiate du décès, la cause initiale, les causes sous-jacentes ainsi que les comorbidités. Les parties anonymes et confidentielles de ces deux documents sont transmises par la mairie au médecin de santé publique de la DDASS puis au Centre d'épidémiologie sur les causes médicales de décès (CEPIDC). La statistique nationale de mortalité est alors établie annuellement (Bouvier-Colle, 1998 ; Jouglan, 2000).

Toutefois, même si celle-ci est nationale et exhaustive, l'information apportée est limitée par plusieurs facteurs (Bouyer, 1995 ; Tubiana, 1991). En effet, l'identification de la cause de décès n'est pas toujours évidente pour les personnes atteintes de plusieurs maladies, notamment les personnes âgées. De plus, le cancer ne sera pas enregistré sur le certificat de décès des patients cancéreux qui décèdent suite à une autre maladie ou suite à un accident (Powell, 1991). Enfin, les taux de mortalité n'englobent pas les patients ayant survécu et de ce fait les personnes jeunes ou les cancers de bon pronostic, comme par exemple ceux de la thyroïde, sont sous-représentés.

Les registres en population ont pour objectif principal le recensement systématique, continu et exhaustif des nouveaux cas de cancer et de leurs caractéristiques dans une population définie sur une base géographique. Ainsi les registres de population fournissent, entre autres, des informations sur l'incidence du cancer et sur ces variations temporelles avec la mise à jour régulière de statistiques internationales (Parkin, 1997 ; Ménégoz, 1995). En 1965, l'OMS a créé à Lyon le Centre international de recherche sur le cancer (CIRC), et, depuis 1966, l'International Association of Cancer Registry (IACR) permet le fonctionnement en réseau des registres du cancer au niveau international (Wagner, 1991). Aujourd'hui, on dénombre environ 200 registres de population au monde, dont la plupart sont situés en Europe et en Amérique du Nord et sont consacrés à l'enregistrement des cancers. La population couverte par ces registres peut être nationale comme par exemple au Danemark, en Norvège, en Finlande ou en Islande ou porter sur des régions spécifiques d'un pays comme c'est le cas en France, en Italie ou aux États-Unis.

En France, le développement des registres de cancer s'est accéléré depuis 1975, donnant lieu à la création en 1986 du Comité national des registres (CNR) qui est placé sous la double tutelle des ministères chargés de la Santé (Institut de veille sanitaire — InVS) et de la Recherche (Inserm). Une vingtaine

de registres du cancer qualifiés peut ainsi être dénombrée. Depuis 1991, ces registres sont regroupés en un réseau Francim qui participe également à plusieurs actions concertées européennes notamment par le biais du Réseau européen du registre du Cancer du programme Eurocare (Sant, 2003), ainsi que du réseau Eurocim, ce qui laisse supposer une bonne homogénéité des données. La mise en place d'une procédure de qualification des registres français depuis la création du CNR est également un argument en faveur de la qualité des registres, qui s'appuient sur des procédures standardisées pour le recueil de l'information et sa gestion. Toutefois, l'extrapolation des résultats obtenus au niveau national est parfois limitée par l'absence de couverture de l'ensemble de la population française car moins de la moitié des régions de France métropolitaine disposent d'un registre reconnu par le CNR, et près de la moitié des registres sont spécialisés (limités à un organe particulier).

L'intérêt des bases de données PMSI apparaît donc clairement (comme nous le montrerons au paragraphe 3) dans la mesure où il fournit des bases de données structurées et codées de manière standardisée et exploitables au niveau national, régional (Gross, Cancéropôles) et local. L'utilisation de ces données pour l'allocation budgétaire des établissements de santé constitue un incitatif important en faveur d'une bonne exhaustivité et d'une bonne qualité des données, éléments qui sont par ailleurs réglementés. Ainsi, les résultats de l'analyse de gestion hospitalière nationale (AGHN) publiés sur le site Internet officiel du PMSI (www.atih.sante.fr) indiquent une exhaustivité de l'ordre de 100 % pour la base nationale des établissements publics et privés depuis 1998. Sur le plan de la qualité, le taux d'erreur global diminue au cours du temps d'après la littérature (Minvielle, 1991 ; Colin, 1994 ; Cauvin, 1997 ; Le Minous, 2000), avec des estimations inférieures à 10 % pour les données administratives et à 20 % pour les données médicales. Outre les programmes d'assurance de qualité développés dans certains établissements (D'alche-Gautier, 1999 ; Briand, 2000), les procédures de contrôle de qualité *a posteriori* sur échantillons effectués en interne par les DIM et en externe par les médecins inspecteurs et les médecins conseils territoriaux sont renforcées par des procédures de gestion de la qualité permettant la détection automatique d'erreurs. On peut citer par exemple, outre les algorithmes de détection d'erreurs inclus dans la fonction de groupage en GHM, le logiciel Nestor, développé par le Comité technique régional d'information médicale (Cotrim) d'Ile-de-France, qui permet un contrôle lexical, syntaxique et sémantique du RSS.

Par ailleurs, la population française des patients atteints de cancer est bien représentée dans les bases de données PMSI. En effet, la majorité des moyens thérapeutiques et une partie des moyens de diagnostics utilisés contre le cancer nécessitent une hospitalisation ou une venue dans le secteur court séjour des établissements de santé. Les patients traités chirurgicalement par chimiothérapie ou par radiothérapie sont donc enregistrés de façon systématique dans les bases

de données PMSI. Toutefois, certaines personnes sont susceptibles d'échapper au PMSI, soit parce qu'elles ne sont jamais prises en charge pour cancer (par exemple certaines personnes âgées) ou parce qu'elles sont suivies uniquement en consultation, comme certaines tumeurs de la peau et hémopathies chroniques. On est toutefois en droit de penser que ces situations restent rares et qu'elles seront couvertes en grande partie à l'avenir par l'extension du PMSI au recueil de l'activité externe des établissements.

Les bases de données PMSI peuvent donc apporter une contribution notable à l'épidémiologie descriptive des cancers selon deux niveaux. Tout d'abord, elles sont déjà utilisées par les registres existants, à titre de source complémentaire, notamment pour la recherche des patients qui sont pris en charge pour leur cancer dans une région différente de celle de leur lieu de résidence. La collaboration entre les DIM et les registres existe dans plusieurs régions, notamment en Normandie. En Bourgogne, la collaboration entre le DIM du CHU de Dijon et le registre bourguignon des cancers digestifs (Faivre, 2001) ainsi que le registre des hémopathies malignes de Côte-d'Or (Carli, 1999), a été bénéfique aux deux types de structures, sous forme de reconnaissance de cas pour les registres et de faux diagnostics pour le DIM. D'autre part, les informations recueillies par le PMSI pourraient être utilisées pour le calcul des estimations d'incidence du cancer pour les régions non couvertes par un registre en population (Trombert-Paviot, 2003). En effet, l'application de la procédure Foin (Trouessin, 1997) permet depuis peu de réaliser le chaînage des séjours d'un même patient dans l'ensemble des établissements de soins publics et privés au niveau régional comme au niveau national. Dans ce cadre, les données PMSI pourraient aider à l'estimation et l'analyse des variations d'incidence à différents échelons géographiques sur le territoire national, tout en tenant compte des variations des pratiques de prise en charge (Trombert-Paviot, 2003).

Ceci supposerait de définir une méthode d'identification des cas incidents à partir du PMSI qui pourrait reposer soit sur la première identification d'un code diagnostic de cancer, pour l'ensemble des séjours d'un même patient, soit sur des algorithmes plus complexes alliant les codes diagnostics aux codes actes spécifiques de ce cancer (McBean, 1993 ; McClish, 1997), l'exclusion des cancers prévalents reposant sur la recherche du même cancer dans les bases PMSI des années antérieures (Warren, 1999). Cette méthode d'identification des cas incidents se heurte à la problématique des erreurs de recueil des données médicales du PMSI. Toutefois, celles-ci sont essentiellement le fait d'erreurs dans la hiérarchisation des diagnostics, c'est-à-dire dans le choix du diagnostic principal. L'information est alors bien présente dans le résumé mais à une autre place. Pour éviter ces erreurs, la méthode la plus simple serait alors de rechercher la première occurrence du diagnostic de cancer pour un patient, qu'il s'agisse d'un diagnostic principal, relié ou associé, selon les règles du codage du PMSI, diffusées par l'Agence technique de l'information sur l'hospitalisation (ATIH).

De plus, l'intérêt des établissements, dans le cadre de la T2A, étant de coder ce qui est le mieux valorisé et non pas ce qui donne le meilleur reflet de la réalité épidémiologique, l'utilisation épidémiologique des données PMSI supposerait la définition d'indicateurs et de règles de codage spécifiques. Cela est prévu et le rapport technique a été introduit avec le format du RUM en 2005 (possibilité de fichier complémentaire à visée épidémiologique, associé au RSS).

3 | Apport des techniques de chaînage des données du PMSI pour l'autoévaluation gérée par les médecins : l'exemple du Réseau périnatal de Bourgogne

La région Bourgogne compte 1 800 000 habitants et environ 18 000 naissances annuelles. Un réseau périnatal s'est mis en place progressivement depuis 1992 (Cornet, 2001). Il est actuellement composé des 16 établissements publics et des 4 établissements privés de la région qui prennent en charge les femmes enceintes, les mères et les nouveau-nés. Un comité médical constitué de représentants des 20 établissements (20 obstétriciens et 20 pédiatres) a la responsabilité de l'évaluation et de l'organisation des soins au sein du réseau périnatal. Le développement d'un recueil continu d'indicateurs périnatals est apparu indispensable à la dynamique du réseau et a été mis en place dès 1998. Le choix et la définition des items ont été assurés par un groupe de travail multidisciplinaire qui a retenu 51 items. Chacun d'eux a reçu une définition médicale précise et consensuelle.

Le PMSI a été choisi comme support du recueil, bien que son objectif premier soit médico-économique. La base de données du PMSI présentait l'avantage d'être déjà existante et quasi-exhaustive, puisque obligatoire dans tous les établissements, qu'ils soient publics ou privés. Ainsi, le recueil systématique des items préalablement définis a été intégré au recueil obligatoire du PMSI, conduisant à un PMSI « élargi ». Quarante des 51 items choisis ont pu être codés de façon standardisée selon la Classification internationale des maladies 10^e révision (Cim 10) et le catalogue des actes imposés pour le codage du PMSI. Les items absents de la Cim 10 concernaient essentiellement les facteurs de risques psychosociaux maternels, les antécédents maternels, l'âge gestationnel et les données nominatives supplémentaires nécessaires au chaînage de fichiers anonymes. Pour ces variables, deux solutions ont été appliquées. Dans la plupart des cas, il a été possible d'obtenir un consensus régional sur une utilisation particulière de codes de la Cim 10 : par exemple, le code Z 768 « sujet ayant recours aux services de santé dans d'autres circonstances précisées » est utilisé par tous les établissements pour coder l'item « grossesse non connue de l'équipe obstétricale le jour de l'accouchement ». Dans d'autres cas, beaucoup plus rares, il n'a

pas été possible d'utiliser la Cim 10, notamment pour le codage de l'âge gestationnel qui se fait en semaines d'aménorrhée (de 22 à 43 SA). Pour la saisie de ces items, il a été nécessaire d'élargir les possibilités de saisie des logiciels PMSI (adjonction d'une fiche à chaque résumé d'unité médicale (Rum), permettant la saisie de données supplémentaires, et constituant ainsi un « Rum élargi »).

La qualité des données est indispensable si l'objectif du recueil est l'évaluation. Elle est contrôlée en premier lieu sur les fichiers nominatifs dans chaque établissement, puis sur le fichier global anonyme de façon centralisée. Des procédures de validation spécifiques au chaînage de fichiers et à la périnatalité sont appliquées par le pédiatre et l'informaticien de la cellule d'évaluation du Réseau périnatal (exhaustivité du recueil de tous les « Rum élargis » d'un même patient et des données utilisées pour le chaînage cohérence des données administratives. Les données saisies en sus du Rum font l'objet de contrôles particuliers. À titre d'exemple, en 2002, après validation, 78 âges gestationnels manquent sur 18 108 nouveau-nés.

Parallèlement, pour chaque couple mère-enfant, la cohérence des données médicales recueillies est vérifiée. Les incohérences sont détectées par des procédures informatiques et vérifiées manuellement par le pédiatre. La plupart des erreurs sont facilement identifiables par l'analyse des Rum élargis, et corrigées directement par le pédiatre sans nécessité de retour au dossier obstétrico-pédiatrique. Le dossier du patient est parfois ressorti mais cette activité ne représente pour les établissements que quelques heures de travail par an et par service.

Les fichiers extraits de chaque établissement sont centralisés pour traitement au CHU de Dijon (Cellule d'évaluation du réseau périnatal). Toutefois, conformément à la législation (directive 95/46/CE du 24 octobre 1995 ; loi n° 78-17 du 6 janvier 1978), ces fichiers ne peuvent sortir des établissements pour exploitation qu'après avoir été rendus anonymes. Le chaînage de données anonymes est rendu possible par l'utilisation du logiciel Anonymat (Quantin, 1998). Ce logiciel a reçu l'accord du Service central de la sécurité des systèmes d'information, le SCSSI. Son utilisation dans le cadre des fichiers médicaux bourguignons est autorisée par la Cnil (n° 98003718). Ce programme assure une transformation irréversible (méthode de hachage) des données d'identification en un repère utilisable pour le chaînage. Les fichiers sont donc rendus anonymes dans chaque établissement avant transmission au CHU pour chaînage et traitement.

Les données obtenues sont utilisées à des fins d'évaluation de la qualité des soins périnataux : analyse des pratiques médicales (conformité à la charte du réseau périnatal, aux recommandations de bonne pratique clinique), analyse de l'organisation des soins, données épidémiologiques descriptives. Les résultats font l'objet d'une analyse annuelle collégiale et non anonyme par les médecins

du réseau. Ceux-ci restent propriétaires de ces informations, condition *sine qua non* de l'efficacité de ce système d'audit interne (Sagot, 2003). Le système de recueil des données d'évaluation périnatale adopté en Bourgogne montre ainsi la possibilité de développer un recueil continu et exhaustif de données PMSI élargies et chaînées pour évaluer les pratiques périnatales à l'échelle d'une région. D'autres travaux réalisés à l'échelle de la région Rhône-Alpes (Rabilloud, 1998) soulignent également le potentiel des données PMSI pour cette évaluation.

4 | Apport des techniques de chaînage des données du PMSI et des registres pour les recherches sur l'utilisation des services de soins : l'exemple des États-Unis

Le National Cancer Institute et les Centers for Medicare and Medicaid Services (CMMS) (Potosky, 1993) ont organisé une base de données partagée et permanente, par le chaînage d'une part des données des registres de cancers participant au programme Surveillance, Epidemiology and End Results (SEER) pour les années 1973-1999, et d'autre part, des données de prises en charge et de facturation des assurances *Medicare* et *Medicaid* jusqu'en 2002. Cette base de données s'appelle SEER-Medicare (<http://healthservices.cancer.gov/seermedicare/>).

Le chaînage se fait sur les données suivantes : numéro de sécurité sociale, prénom, nom, initiale intermédiaire, date de naissance et date de décès. L'évaluation a montré que le chaînage de ces données était possible pour environ 93 % des personnes de 65 ans et plus.

Les avantages et inconvénients de cette base de données ont été évalués pour le dépistage (Freeman, 2002), la chirurgie des cancers (Cooper, 2002 ; Du, 2000), la radiothérapie (Virnig, 2002), la chimiothérapie (Warren, 2002) et les complications des traitements (Potosky, 2002).

Parmi les études portant sur l'ensemble de la trajectoire de soin, on peut citer celles sur les taux de survie après chirurgie conservatrice pour le cancer du sein, les protocoles de chimiothérapie utilisés et leurs résultats (Du, 2003), l'utilité des données DRG pour les recherches (Freeman, 2000), l'effet des états dépressifs sur le diagnostic, le traitement et la survie des femmes âgées (Goodwin, 2004), l'utilisation des soins à domicile (Lackan, 2004) et l'évaluation de l'efficacité du dépistage et du traitement sur le résultat (Randolph, 2002).

Cet exemple illustre bien les possibilités offertes par le chaînage des données épidémiologiques issues de bases de données ayant des objectifs différents ou

semblables pour partager entre chercheurs de diverses disciplines les informations recueillies souvent au prix d'efforts importants.

Conclusion et recommandations

Les données issues des bases médico-administratives du PMSI constituent une source d'informations précieuses pour l'épidémiologie descriptive et pour la planification, à l'échelon régional et national. En effet, il s'agit de la seule source d'information assurant la couverture de la totalité du territoire national, bénéficiant de procédures de contrôle d'exhaustivité et de qualité à la fois internes aux établissements et externes par les tutelles hospitalières. Pourtant, ces données sont à l'heure actuelle sous-exploitées à des fins d'épidémiologie descriptive probablement, pour les raisons suivantes :

- une finalité économique, sans implication claire des acteurs concernés par l'utilisation épidémiologique ;
- un enregistrement par séjour et non pas par patient.

Ces deux limites pourraient tout à fait être évitées à condition de mettre en place les préconisations suivantes :

- la définition claire d'objectifs épidémiologiques à partir de la proposition d'indicateurs définis conjointement par les sociétés savantes représentant les différentes spécialités médicales et chirurgicales au niveau national, l'InVS et l'Inserm ainsi que l'Atih afin d'intégrer et d'harmoniser les règles de codage de ces indicateurs dans le cadre des règles du PMSI, tout en veillant à ne pas nuire aux objectifs médico-économiques princeps de ce recueil ;
- le développement du chaînage des résumés d'un même patient permettant de rassembler les informations des différentes hospitalisations de ce patient pour tous les établissements publics et privés qui l'ont pris en charge, ainsi qu'avec les données de mortalité collectées par l'Insee et le CEPIDC. Ceci suppose d'introduire dans le dispositif de chaînage actuel (logiciel Foin de la Cnamts) des informations permettant de relier les données d'hospitalisation avec les données de mortalité et notamment la commune de naissance du patient. En effet, les informations contenues dans le logiciel Foin, bien qu'elles s'appuient sur le numéro de sécurité sociale de l'assuré et la date de naissance du patient, ne permettent pas de faire le lien avec les données de mortalité individuelles. Il est bien entendu que le dispositif de hachage, assurant l'anonymat des informations, selon les

recommandations de la Cnil (Vuillet-Tavernier, 2000), devra être maintenu de façon à assurer l'anonymat d'informations traitées au niveau national et la confidentialité due au patient ;

- le renforcement des procédures d'assurance de qualité au niveau des établissements et de contrôle de qualité des données au niveau régional pour s'assurer non seulement du respect des règles de codage du PMSI à finalité économique mais également de celles à finalité épidémiologique.

Il semble en effet indispensable d'adapter l'outil d'évaluation médico-économique aux besoins de l'épidémiologie descriptive et de la planification pour augmenter son efficacité et permettre son appropriation par les professionnels de santé et les décideurs, prérequis indispensable à l'installation d'une démarche d'assurance-qualité efficace.

Remerciements

Je remercie tout particulièrement pour leur aide précieuse dans la relecture de ce manuscrit : A. Bernard, M.J. Bismuth, J.P. Bodin, P.M. Carli, J. Faivre, B. Gouyon, F. Gremy, F. Kohler, P. Metral, A. Musat, A. Tursz et F. Van Roekeghem.

Bibliographie

Arrêté 31 décembre 2003 fixant les prélèvements provisionnels à opérer sur le produit au titre de l'année 2004 de la constitution sociale de solidarité.
<http://www.sante.gouv.fr/htm/dossiers/t2a/accueil.htm>

Bouyer J., Hémon D., Cordier S. *et al.* (1995). *Épidémiologie : Principes et méthodes quantitatives*. Paris, Éditions Inserm.

Bouvier-Colle M., Vallin J., Hatton F. (1998). *Mortalité et causes de décès en France*. Paris, Éditions Inserm ; 1990 – Ministère de l'Emploi et de la solidarité, Haut comité de la santé publique. *La santé en France 1994-1998*. Paris, La Documentation française.

Briand S., Chalé J., Fabre D. (2000). Qualité des résumés de sortie standardisés. Comment mobiliser les médecins ? *Gestions Hospitalières*, 188-193.

Bright RA., Avorn J., Everitt DE. (1989). Medicaid data as a resource for epidemiologic studies : strengths and limitations. *Journal Clinical Epidemiological*, **42** (10) : 937-45.

- Carli PM., Coebergh JWW., Verdecchia A. and the Eurocare Working Group (1999). Variation in survival of adult patients with haematological malignancies in Europe since 1978. *European Journal Cancer*, **34** : 2253-63.
- Cauvin J., Hardy B., Jehan P. *et al.* (1997). Qualité du codage et conséquences en ISA et journées théoriques : à partir des fichiers de 100 dossiers recodés par établissements. *Journal d'Économie Médicale*, **15** (3-4) : 195-206.
- Colin C., Ecochard R., Delahaye F. *et al.* (1994). Data quality in a DRG-based information system. *International Journal Quality Health Care*, **6** (3) : 275-80.
- Cooper GS., Virnig B., Klabunde CN., Schussler N., Freeman J., Warren JL. (2002). Use of SEER-Medicare data for measuring cancer surgery. *Medical Care*, **40** (8 Suppl) : 43-48.
- Cornet B., Gouyon JB., Binquet C. *et al.* (2001). Évaluation régionale en périnatalité : mise en place d'un recueil continu d'indicateurs. *Revue d'Épidémiologie et de Santé Publique*, **49** : 583-593.
- D'alche-Gautier M., Hamon-Poupinel V., Henriot L. *et al.* (1999). La qualité des informations médicales du PMSI : comparaison de deux enquêtes à deux ans d'intervalle. *Journal d'Économie Médicale*, **17**(1) : 29-36.
- Dusserre L., Ducrot H., Allaërt FA. (1999). *L'information médicale : l'ordinateur et la loi*, 2^e édition, EM inter.
- Du X., Freeman JL., Goodwin JS. (1999). Information on radiation therapy treatment in patients with breast cancer : the advantages of the linked Medicare and SEER data. *Journal Clinical Epidemiology*, **52** : 463-470.
- Du X., Freeman JL., Warren JL., Nattinger AB., Zhang D., Goodwin JS. (2000). Accuracy and completeness of Medicare claims data for surgical treatment of breast cancer. *Medical Care*, **38** (7) : 719-727.
- Faivre J., EUROCARE study group (2001). Variation in survival of patients with digestive tract cancers in Europe, 1978-1989. *European Journal Cancer Prevention*, **10** (2) : 173.
- Fessler JM., Frutiger P. (2003). *La tarification hospitalière à l'activité : éléments d'un débat nécessaire et propositions*. Gestion des Ressources. Fonction cadre de santé. Éditions Lamarre.
- Fetter RB., Shin Y., Freeman JL. *et al.* (1980). Case Mix definition by Diagnosis Related Groups. *Medical Care*, **18** (suppl 2) : 1-53.
- Freeman JL., Zhang D., Freeman DH., Goodwin JS. (2000). An approach to identifying incident breast cancer cases using Medicare claims data. *Journal Clinical Epidemiology*, **53** : 605-614.

- Freeman JL., Klabunde CN., Schussler N., Warren JL., Virnig BA., Cooper GS. (2002). Measuring breast, colorectal and prostate cancer screening with Medicare claims data. *Medical Care*, **40** (suppl 8) : 36-42.
- Goodwin JS., Zhang DD., Ostir GV. (2004). Effect of depression on diagnosis, treatment, and survival of older women with breast cancer. *Journal American Geriatrics Society*, **52** (1) : 106-111.
- Grémy F., Priollaud N. (2004). *On a encore oublié la santé ! Propositions pour une médecine de qualité et de solidarité*. Collection « les propos d'un homme libre ». Éditions Frison-Roche.
- Jougla E., Pavillon G., Rossollin F., de Smedt M., Bonte J. (1998). Improvement of the quality and comparability of causes of death statistics inside the European Community. Eurostat Task Force on « causes of death statistics ». *Revue d'Épidémiologie et de Santé Publique*, **46** (6) : 447-56.
- Jougla E., Rican S., Pequignot F. et al. (2000). Inégalités sociales de mortalité. In : *Les inégalités sociales de santé*, Inserm La Découverte, 147-62, Paris.
- Lackan NA., Ostir GV., Freeman JL., Mahnken JD., Goodwin JS. (2004). Decreasing variation in the use of hospice among older adults with breast, colorectal, lung and prostate cancer. *Medical Care*, **42** (2) : 116-122.
- Le Minous F., Jarno P., Ledoux B. et al. (2000). Évaluation de la qualité du codage des séjours de cancérologie groupés dans le GHM 675. *Journal d'Économie Médicale*, **18** (1) : 51-62.
- Ménégoz F., Chérié-Challine L. (1998). *Le cancer en France : incidence et mortalité, situation entre 1975 et 1995*. La Documentation française.
- McBean AM., Babish JD., Warren JL. (1993). Determination of lung cancer incidence in the elderly using medicare claims data. *American Journal Epidemiology*, **137** (2) : 226-34.
- McClish DK., Penberthy L., Whittemore M. et al. (1997). Ability of medicare claims data and cancer registries to identify cancer cases and treatment. *American Journal Epidemiology*, **145** (3) : 227-33.
- Michel E., Jougla E. (2000). Main indicators used for the descriptive analysis of medical causes of death. *Bulletin Cancer*, **87** (10) : 755-9.
- Minvielle E., de Pourvoirville G., Jeunemaître A. (1991). PMSI contrôle de la qualité du codage. *Gestions Hospitalières*, **302** : 17-22.
- Organisation mondiale de la santé (1993). *Classification Internationale des Maladies*, 10^e révision. Genève, OMS.

- Parkin DM., Whelan SL., Ferlay J., Raymond L., Young J. (1997). Cancer incidence in five continents volume VII. *IARC Scientific Publications*.
- Potosky AL., Riley GF., Lubitz JD., Mentnech RM., Kessler LG. (1993). Potential for cancer related health services research using a linked Medicare-tumor registry database. *Medical Care*, **31** (8) : 2-748.
- Potosky AL., Warren JL., Riedel ER., Klabunde CN., Earle CC., Begg CB. (2002). Measuring complications of cancer treatment using the SEER-Medicare data. *Medical Care*, **40** (suppl 8) : 62-68.
- Powell J. (1991). Data sources and reporting. In : *Cancer registration : principles and methods*. Jensen O., Parkin D., Mac Lennan R., Muir C., Skeet R. (eds). IARC Scientific Publications, 29-42.
- Quantin C., Bouzelat H., Allaert FA. *et al.* (1998). Automatic Record Hash Coding and Linkage for Epidemiological Follow-up Data Confidentiality. *Methods Information Medicine*, **37** : 271-7.
- Rabilloud M., Ecochard R., Guilhot J. *et al.* (1998). Study of the variations of the caesarean sections rate in Rhône-Alpes region (France) : Effect of women and maternity service characteristics. *European Journal Obstetrics Gynecology Reproductive Biology*, **78** : 11-7.
- Randolph WM., Mahnken JD., Goodwin JS., Freeman JL. (2002). Using Medicare data to estimate the prevalence of breast cancer screening in older women : comparison of different methods to identify screening mammograms. *Health Services Research Decision*, **37** (6) : 1643-1657.
- Rodrigues JM., Trombert Paviot B., Martin C. *et al.* (2001). Les systèmes DRG en Europe et dans le monde. *Competence. Hospital Management Forum*, 9-12.
- Sagot P., Gouyon-Cornet B., Gouyon JB. (2003). Le PMSI élargi : système d'évaluation du réseau de soins périnataux de Bourgogne. *Gynécologie Obstétrique Fertilité*, **31** : 162-177.
- Sant M., Aareleid T., Berrino F. *et al.* (2003). Eurcare – 3 : survival of cancer patients diagnosed 1990-94, results and commentary. *Annals Oncology*, **14** (suppl 5) : 61-118.
- Trombert-Paviot B., Martin C., Vercherin P., Rodrigues JM. (2003). From case mix data bases to health geography. In : *Proceedings of the 19th International PCS/E Working Conference*, 295-301, Washington (USA).
- Trouessin G., Allaert FA. (1997). FOIN : a nominative information occultation function. *MIE*, **43** : 196-200.
- Tubiana M. (1991). Trends in cancer mortality. *Bulletin Cancer*, **78** (5) : 401-3.

- Virnig BA., Warren JL., Cooper GS., Klabunde CN., Schussler N., Freeman J. (2002). Studying radiation therapy using SEER-Medicare linked data. *Medical Care*, **40** (suppl 8) : 49-54.
- Vuillet-Tavernier S. (2000). Réflexions autour de l'anonymat dans le traitement des données de santé. *Médecine et Droit*, **40** : 1-4.
- Wagner G. (1991). History of cancer registration. In : *Cancer registration : principles and methods*. Jensen O. Parkin D., Mac Lenan R., Muir C., Skeet R. (eds.). IARC Scientific publications, 3-6.
- Warren JL., Feuer E., Potosky AL. et al. (1999). Use of medicare hospital and physician data to assess breast cancer incidence. *Medical Care*, **37** (5) : 445-56.
- Warren JL., Harlan LC., Fahey A., Virnig BA., Freeman JL., Klabunde CN., Cooper GS., Knopf KB. (2002). Utility of the SEER-Medicare data to identify chemotherapy use. *Medical Care*, **40** (suppl 8) : 55-61.

CHAPITRE 4

Épidémiologie et sciences mathématiques

SOUS-CHAPITRE 4.1

Épidémiologie et statistique

GUY THOMAS, MARC YOR ET PIERRE-YVES BOËLLE

Dans le domaine du vivant, les sources de variabilité sont très nombreuses, et pour la plupart inconnues. Même si l'on imaginait pouvoir connaître tous les paramètres intervenant dans le phénomène étudié, un modèle prenant en compte tous ces degrés de liberté serait beaucoup trop complexe pour être utile. Dans ce contexte, les épidémiologistes ont été amenés :

- à développer des protocoles spécifiques de recueil de données ;
- à modéliser les phénomènes étudiés comme des phénomènes aléatoires ;
- à analyser les données recueillies dans le cadre théorique de la statistique mathématique.

Les interactions entre épidémiologie et statistique mathématique sont riches et bilatérales. Certains problèmes rencontrés en épidémiologie sont à l'origine de méthodes statistiques originales et spécifiques, dont l'étude théorique est menée par des mathématiciens. Réciproquement, des résultats de mathématique contemporaine trouvent un domaine d'application en épidémiologie. Cette dynamique interactive peut seule garantir la qualité de la recherche en épidémiologie, alors que la disponibilité croissante de logiciels « clés en main » permet désormais à des utilisateurs naïfs de mener des analyses complexes sans maîtrise des outils méthodologiques.

L'objet de ce chapitre est d'illustrer les interactions entre épidémiologie et statistique mathématique par des exemples appartenant à différents domaines de l'épidémiologie, mais sans aucune prétention à l'exhaustivité.

1 | Statistique mathématique et épidémiologie

Les fondements de l'approche statistique sont évidemment mathématiques, même si cette caractéristique peut être masquée par la possibilité d'appliquer certains résultats sans avoir à maîtriser la théorie sous-jacente. La disponibilité croissante de logiciels mettant en œuvre simplement et rapidement les traitements statistiques les plus complexes ne peut que renforcer l'écart entre compréhension et utilisation d'une méthode.

1.1 Analyse des essais thérapeutiques

Dans un protocole d'essai clinique (on considérera ici un protocole de phase III, de recherche de supériorité ou d'équivalence), on retrouve généralement les caractéristiques suivantes : un nombre de patients à inclure est fixé dès le début du protocole, grâce à un calcul basé sur la différence entre traitements que l'on souhaite déceler avec des risques acceptés ; la conclusion n'est connue qu'une fois l'ensemble des patients observés, à l'issue d'une analyse finale ; l'analyse est conduite dans un cadre statistique « fréquentiste ».

L'analyse de tels essais peut ne pas nécessiter de compétences mathématiques particulières, par exemple si l'on effectue une simple comparaison de moyennes. Néanmoins, la théorie mathématique qui sous-tend la validité de la méthodologie statistique fait appel à des concepts mathématiques comme la convergence en loi d'une suite de variables aléatoires, qui ne sont enseignés que dans le 2^e cycle des études supérieures.

Le protocole type de l'essai contrôlé reste régulièrement critiqué, en raison de problèmes éthiques, de l'écart entre la recherche d'un résultat « statistiquement significatif » et de sa pertinence clinique, et également en raison du temps nécessaire pour obtenir une conclusion (souvent plusieurs années). Des développements récents de statistique mathématique ont été réalisés qui ciblent certains de ces problèmes, au prix toutefois d'une plus grande complexité de l'analyse. Un exemple important est le développement des méthodes de tests répétés ou séquentiels. La propriété recherchée est alors de pouvoir conclure à l'efficacité d'une intervention en s'appuyant sur le nombre le plus réduit possible d'unités

expérimentales. La répétition des tests, ainsi que la poursuite de l'essai conditionnellement aux résultats antérieurs, posent cependant des problèmes mathématiques ardues pour déterminer quelle procédure garantit un risque d'erreur global spécifié à l'avance.

Par exemple, la recherche d'une fonction de coût optimale pour répartir cette erreur au cours des analyses successives est un terrain de recherche encore actif. Bien qu'*a priori* d'essence différente, la recherche de méthodes de randomisation adaptative, où l'intervention attribuée aux patients nouvellement recrutés dans un essai est biaisée vers celle qui semble la plus efficace au vu des données antérieures (stratégie *Play the Winner*) possède des caractéristiques équivalentes à celle des essais séquentiels. Les développements mathématiques nécessaires pour établir les propriétés de ces méthodes sont très techniques, notamment lorsque l'on veut garantir une taille finie à ces essais : les démonstrations sont réalisées grâce aux propriétés de fonctionnelles du mouvement brownien, dont le niveau mathématique est celui d'un troisième cycle. L'implémentation en pratique de ces méthodes reste limitée, d'une part à cause de problèmes logistiques (l'analyse des données doit être contemporaine du déroulement de l'essai), mais également en raison de la plus grande complexité de la méthode. Cependant au moins deux logiciels spécialisés sont actuellement commercialisés (PESTTM, EASTTM) qui permettent de simplifier tant la planification que la conduite de tels essais.

Un changement beaucoup plus radical dans la conduite des essais cliniques est celui qui est annoncé dans un récent éditorial de la revue *Science* : l'auteur constate que « si la majeure partie des sciences ont travaillé pour que la connaissance *a priori* soit largement intégrée dans les analyses ultérieures, il n'en est pas de même pour les essais cliniques ». Bien qu'une grande quantité d'information *a priori* soit néanmoins utilisée dans la construction d'un essai, il est vrai que l'analyse finale de l'essai repose sur les seules données recueillies au cours de l'essai, privilégiant une forme d'objectivité de l'analyse en dehors de tout *a priori*. Dans les essais de phase I/II, c'est-à-dire de recherche de dose et toxicité, ce paradigme tend de plus en plus à être remplacé par des approches « bayésiennes » où l'information connue avant l'essai et obtenue en cours d'essai est utilisée dans la conduite du protocole. Les essais de phase III sont encore peu réalisés selon l'approche bayésienne, bien qu'il ait été suggéré que cela puisse s'avérer très fructueux, notamment dans le cas des maladies rares. Ces approches « bayésiennes » sont généralement plus complexes dans leur mise en œuvre, puisqu'elles reposent sur des calculs (essentiellement) numériques de densités conditionnelles. Ici, bien qu'il n'existe pas encore de logiciels « clés en main » permettant la planification et l'analyse de ces essais, des plates-formes non spécifiques sont déjà disponibles (par exemple BUGSTM).

1.2 Données manquantes

Il n'est généralement pas acceptable, et toujours préjudiciable, d'exclure de l'analyse les données qui sont incomplètes. C'est cependant la méthode qui est généralement appliquée dans les analyses épidémiologiques, notamment parce que les logiciels d'analyse procèdent à l'élimination des données incomplètes avant analyse.

Éviter les données manquantes est donc une priorité qui doit se gérer principalement lors du recueil des données par le contrôle de qualité. Cependant, l'utilisation de méthodes statistiques récentes permet, le cas échéant, de rendre compte de ces données manquantes lors de l'analyse. Par exemple, des techniques de type EM (« expectation, maximisation ») sont quelquefois possibles. Il s'agit généralement, par une méthode itérative où les données manquantes sont remplacées par leur moyenne prédite, de parvenir à une estimation des paramètres. Le développement de la méthode est cependant *ad hoc* et doit être adapté à chaque jeu de données.

Une autre voie de recherche à visée plus globale est de procéder à une imputation des données manquantes. À l'aide d'un modèle prédictif ajusté sur les données observées (généralement un modèle non paramétrique, ou par ré-échantillonnage), les valeurs manquantes sont « imputées », avec une variabilité qui reflète l'incertitude du modèle prédictif. On peut donc obtenir un jeu de données complet, qui a l'avantage d'être analysable avec des outils statistiques classiques. Il faut généralement procéder à des imputations multiples pour garantir la validité statistique de la procédure.

Bien évidemment, des hypothèses sont nécessaires à la validité de ces approches : ainsi, elle a été établie pour des données manquantes « totalement au hasard » ou « au hasard », mais la meilleure prise en compte des données manquantes « non au hasard », ou des données « à manquantes informatives » reste un champ de recherche ouvert. Cependant, on ne sait pas caractériser, pour un jeu de données, la nature des données manquantes : « au hasard » ou « informatives ». La disponibilité des méthodes d'imputation multiple dans les logiciels statistiques standard peut alors poser des problèmes d'utilisation indiscriminée.

1.3 Régression non paramétrique

Dans les études où les facteurs de confusions sont nombreux et/ou de nature quantitative, l'estimation non biaisée des relations statistiques d'intérêt repose sur l'utilisation de modèles de régression, où les facteurs de confusion figurent en tant que covariables. Dans ce contexte, les modèles de régression dits « non

paramétriques » ont l'intérêt majeur de ne pas imposer d'hypothèse *a priori* sur la forme mathématique de la relation entre la variable à expliquer et les covariables. Ces modèles sont mathématiquement complexes. Néanmoins, il existe aujourd'hui des logiciels qui permettent d'effectuer des régressions non paramétriques en épargnant à l'utilisateur tout aspect technique. Cette apparente simplicité peut conduire à des conclusions erronées.

Exemple : mortalité et pollution atmosphérique

Les relations entre pollution atmosphérique et mortalité font l'objet de nombreuses recherches. Dans ce contexte, il est devenu habituel d'utiliser une approche non paramétrique pour décrire les covariables (jour de la semaine, saisonnalité, tendances, facteurs météorologiques) dans le modèle de la variable à expliquer, et de garder un modèle paramétrique simple pour la relation entre l'indicateur de pollution et la mortalité. Un article récent (Dominici *et al.*, 2002) illustre les risques inhérents à l'utilisation « naïve » des logiciels de régression non paramétrique dans ce contexte. En prenant pour exemple le problème de l'estimation de la relation entre la mortalité chez les plus de 75 ans et la concentration en particules de l'atmosphère, les auteurs montrent que les procédures automatiques de logiciels « clés en main » peuvent conduire à des biais importants, et que les options par défaut des programmes doivent être adaptées en fonction du problème à traiter.

Ainsi, plus les outils méthodologiques sont puissants et flexibles, plus les mathématiques sous-jacentes sont complexes, plus leur utilisation doit être encadrée par des épidémiologistes qui en maîtrisent les aspects théoriques.

1.4 Méthodes géostatistiques

La cartographie des maladies transmissibles a pour but immédiat de représenter géographiquement l'incidence, mais également d'analyser la dissémination spatiotemporelle de la maladie, ouvrant ainsi la possibilité de prédire ou de contrôler celle-ci.

En géologie, l'analyse de gisements doit s'appuyer sur une vue partielle des sous-sols, obtenue généralement par carottage. Extrapoler la composition du sous-sol à partir de quelques sondages a donc été l'objet d'une recherche spécifique qui a débouché sur la méthode du krigeage (Matheron, 1970). À partir de données pareillement prélevées par « sondage » auprès d'un échantillon de médecins répartis sur un territoire, il est donc possible d'appliquer le même type

de méthodes en épidémiologie (Carrat et Valleron, 1992). La figure 4.1-1 permet de juger l'intérêt de la méthode dans la représentation graphique de la grippe en France.

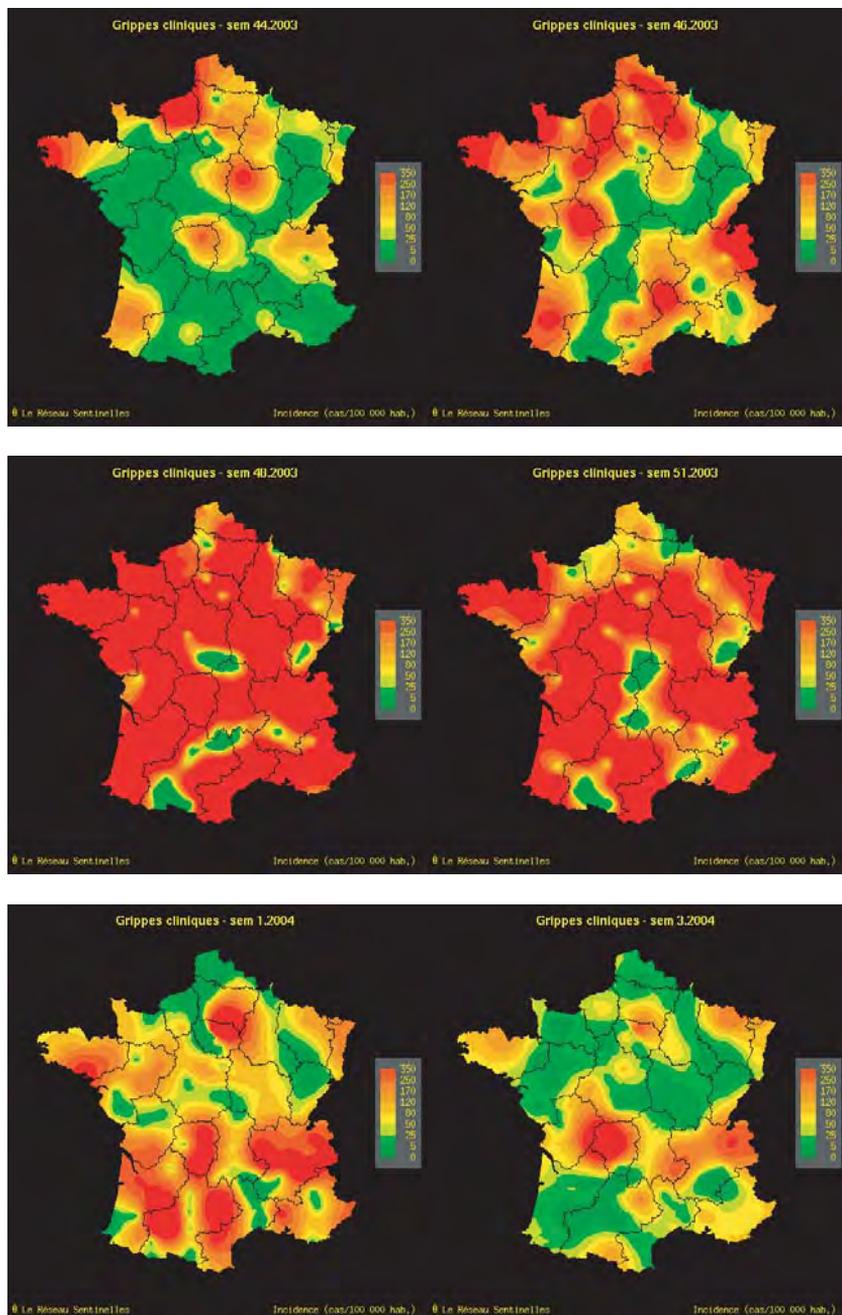
Plusieurs problèmes conceptuels et statistiques se posent :

- comment prendre en compte la dynamique temporelle de l'épidémie ? il s'agit d'un problème lié à la dynamique des épidémies, et que l'on n'avait donc pas à traiter en géostatistique ;
- la proximité géographique est-elle pertinente dans la diffusion des épidémies ? Ou faut-il plutôt prendre en compte une géographie liée aux transports ?

L'utilisation de l'approche bayésienne hiérarchique permet aujourd'hui de répondre à certains de ces aspects, notamment dans la prise en compte de la dynamique temporelle.

1.5 Méthodes amplitude/fréquence

Dans l'étude des maladies transmissibles, l'existence de fronts de propagation devrait être observée si la dimension spatiale géographique est prépondérante. On peut également s'attendre à ce que cette propagation soit plus apparente pour des maladies impliquant un vecteur (moustiques par exemple). De fait, on a observé de tels fronts pour la propagation de la dengue en Thaïlande. L'analyse permettant cette mise en évidence emprunte à la théorie du signal, avec une décomposition en modes empiriques. Cette méthode permet de décomposer sur une famille de fonctions empiriques (et non pas une base de fonctions choisie par avance, comme une base d'ondelettes) la variance des données observées. Dans le cas de la dengue, on a ainsi pu mettre en évidence l'existence d'un mode de période 3 ans, qui correspond au balayage du pays par des fronts d'infection partant de Bangkok. De telles analyses ont également été conduites avec la rougeole en Grande-Bretagne, en s'appuyant par exemple sur une décomposition par ondelettes, avec la mise en évidence de sources de cas, correspondant aux grandes communautés, et de puits, correspondant à des communautés trop petites pour entretenir la maladie mais susceptibles de constituer des réserves de personnes à risque. L'application de ces méthodes permet de mettre en évidence des effets ténus, indétectables dans les données brutes. Il est cependant clair que, pour beaucoup d'épidémiologistes, ces résultats restent très abstraits, car correspondant à des méthodes encore peu connues. Une formation / accompagnement dans la mise en place et l'interprétation est donc nécessaire.

**Figure 4.1-1**

Représentation de la dynamique de l'épidémiologie de syndromes grippaux en France en 2003-2004 : cartes hebdomadaires. (Données : Réseau sentinelles, cf. <http://www.u707.upmc.fr/sentiweb>).

2 | Résultats de mathématique contemporaine appliqués à l'épidémiologie

2.1 Méthodes de Monte-Carlo

Récemment, les approches de type Monte-Carlo ont connu un développement intense. Leur utilisation en épidémiologie permet l'estimation des paramètres dans le cadre de modèles complexes (modèles bayésiens notamment). D'un point de vue général, les techniques Monte-Carlo permettent de calculer efficacement des intégrales inaccessibles aux méthodes de calcul numérique déterministes. Pour les mettre en œuvre, il faut pouvoir simuler des lois de probabilité, domaine qui a donné naissance à de très nombreux travaux fondamentaux. On peut citer dans ce champ l'apparition récente de méthodes de simulation exacte en temps fini (Propp et Wilson, 1996), fondées sur des méthodes de couplage utilisées pour l'étude théorique des chaînes de Markov. Ces méthodes n'étaient à l'origine applicables qu'à des modèles ne comportant qu'un nombre fini d'états, mais les travaux les plus récents étendent le domaine d'application à des modèles plus généraux.

2.2 Critères de sélection de modèles

L'épidémiologiste est souvent confronté au problème de sélectionner le « meilleur » modèle au sein d'une famille de modèles rivaux. Historiquement, la première solution a consisté à traiter ce problème comme un problème statistique de test d'hypothèse, mais cette approche était limitée à certaines familles de modèles (familles de modèles emboîtés) et conduisait à des incohérences logiques. Le deuxième type de solution est apparu au cours des années 1970, avec l'utilisation de critères de sélection construits sur des arguments heuristiques. Ce n'est qu'au cours des dix dernières années qu'a été élaborée une théorie satisfaisante des critères de sélection, couvrant le domaine de l'estimation non paramétrique. Par exemple, le problème du choix optimal de la largeur de bande d'un histogramme (problème d'apparence élémentaire) est l'objet d'un article publié en 2000 (Castellan, 2000). Le critère permettant de déterminer la largeur de bande d'un histogramme peut être utilisé sans en comprendre l'origine, mais les mathématiques sous-jacentes correspondent aux travaux de recherche les plus récents de mathématiciens professionnels.

2.3 Tests séquentiels

Lors de la mise en place d'un essai clinique, le calcul du nombre de sujets nécessaires est l'étape déterminante qui assure que la quantité d'information recueillie (information statistique, au sens de l'information statistique de Fisher

par exemple) permettra de détecter une différence d'intérêt clinique avec une grande probabilité. L'information de Fisher croît linéairement avec la taille de l'échantillon, et en pratique celle-ci devient disponible au cours de l'essai avec l'observation ou la mesure des événements d'intérêt chez les patients, et non pas lorsque tous les patients ont été observés.

La construction de tests séquentiels a pour objet d'exploiter cette augmentation graduelle de l'information afin de permettre d'obtenir une conclusion plus rapidement qu'un essai classique de taille fixe. Les premières approches séquentielles (du type test du rapport de vraisemblance de Wald) souffraient cependant d'une limitation pratique : en théorie, on était certain de pouvoir prendre une décision, et en moyenne cette décision serait prise avant un test classique de mêmes caractéristiques, mais aucune garantie n'était donnée sur la taille maximale nécessaire. Deux innovations ont permis de rendre les essais séquentiels plus attrayants. Il s'agit d'une part des tests « triangulaires », et d'autre part de l'étude des propriétés statistiques des tests répétés.

Le test triangulaire renvoie à la représentation graphique de l'information recueillie au cours d'un essai. La trajectoire de celle-ci est suivie dans le plan (moyenne, variance) de la statistique de test, et des frontières triangulaires sont tracées dans ce plan qui assurent un risque de première espèce égal à 5 %. Contrairement au test séquentiel de Wald, les frontières sont concourantes et non parallèles, ce qui garantit que l'on conclut en temps fini, avec un effectif maximal excédant de peu celui du test classique.

La deuxième méthode correspond à une extension plus intuitive de l'approche classique : plutôt que de faire un test unique, on réalisera plusieurs tests, en nombre fixé *a priori*, en s'arrangeant pour que le risque global soit de 5 %. Des approches pragmatiques peuvent être employées : faire les premiers tests avec un risque très faible (par exemple 1 ‰) pour n'arrêter l'essai précocement que si une différence massive existe, et pratiquer le dernier avec un risque presque semblable à celui que l'on aurait eu en taille fixée (approche de Peto). Des approches optimales sont également possibles, dans lesquelles on essaie de limiter au maximum la taille de l'échantillon nécessaire.

Bien que simples d'énoncé, ces deux approches donnent lieu à des problèmes mathématiques très techniques. Par exemple, l'étude des trajectoires d'une statistique de test dans le plan moyenne/variance peut être ramenée en première approximation à l'étude des fonctionnelles du mouvement brownien, comme décrit par Siegmund (1985). Avec cette formulation, l'étude de certaines propriétés devient possible. On peut déterminer les propriétés d'un test où l'on examine les événements à des temps fixés d'avance (avec un nombre aléatoire de mesures entre deux temps), ou lorsque l'on réalise un test à une cadence fixée par un nombre d'événements donné. En particulier, ces derniers développements

donnent lieu à des calculs mathématiques de corrections liées au dépassement des frontières avant que le test soit effectué. Il en résulte des tests dits « en arbre de Noël » où les frontières sont corrigées pour prendre en compte l'augmentation discrète de l'information.

Enfin, l'estimation de quantités d'intérêt suite à la pratique de ces tests demande des précautions particulières, étant donné que la quantité d'information recueillie est conditionnelle à la différence observée.

En pratique, ces tests demeurent peu utilisés. On peut donner deux éléments d'explication : le premier est que l'économie potentielle en taille d'échantillon peut s'avérer insuffisante au regard de l'addition en complexité tant logistique que technique. Le second est que ce type d'approche demande une compétence statistique qui ne s'acquiert généralement pas avant le 3^e cycle et demandent un travail technique conséquent de la part du statisticien.

2.4 Relations causales – Facteurs de confusion – Choix des covariables

L'existence de facteurs de confusion constitue l'un des obstacles essentiels à l'élucidation de relations causales à partir de données non expérimentales. Dans la mesure où le phénomène de confusion est un concept de nature causale, les tentatives d'élucidation et de définition dont il a été l'objet dans un cadre strictement statistique sont restées infructueuses et ont abouti à une littérature abondante mais obscure.

Le paradoxe de Simpson est une illustration spectaculaire du problème des facteurs de confusion dans les études non expérimentales. Il correspond à la situation où un événement C augmente la probabilité de E dans une population p , tout en diminuant la probabilité de E dans toute sous-population de p . Par exemple, on observe que la proportion des patients qui guérissent est plus élevée dans le groupe traité que dans le groupe témoin. Cependant, si l'on considère l'effet du traitement séparément chez les hommes et chez les femmes, on observe que la proportion de patients qui guérissent est plus faible dans les groupes traités que dans les groupes témoins. Cette constatation est évidemment paradoxale et contraire au bon sens, si l'on interprète la relation statistique traitement/guérison comme causale (plus généralement, toute relation statistique entre deux variables peut être inversée par ajustement sur d'autres variables).

La solution de ce paradoxe a été donnée par Pearl (2000) dans le cadre de la théorie mathématique de la causalité. Pearl a montré que les concepts de facteurs de confusion et d'ajustement étaient de nature causale, et ne pouvait donc

être définis et manipulés rigoureusement qu'à l'aide de modèles de causalité. Il a notamment établi des critères mathématiques qui permettent de décider quelles sont les variables sur lesquelles on doit effectuer un ajustement (« *back-door* » *criterion*).

3 | Les apports spécifiques de l'épidémiologie

3.1 Apports méthodologiques à d'autres disciplines appliquées

Dans un certain nombre de cas, les épidémiologistes ont été à l'origine du développement de méthodes spécifiques, qui ont par la suite trouvé d'autres domaines d'application.

3.1.1 Modèle *logit*

Historiquement développé pour décrire des relations dose/réponse en épidémiologie, le modèle *logit* est aujourd'hui très utilisé en économétrie, où il concurrence largement les modèles *probit* traditionnellement utilisés.

3.1.2 Analyse des données de survie

L'exemple le plus spectaculaire de la diffusion d'une approche épidémiologique est peut-être celui de l'analyse des données de survie. En pratique, il s'agit d'observer le temps jusqu'à la survenue d'un événement d'intérêt (décès, rechute, guérison, ...). Dans de telles études les données recueillies sont caractérisées par le fait qu'elles sont incomplètes, au moins pour des raisons pratiques liées à la durée limitée de l'étude. En effet, il est clair que l'on ne peut connaître la durée de survie que pour les personnes qui seront décédées. Si l'on choisit d'arrêter l'étude avant d'avoir observé le décès de toutes les personnes suivies, certaines personnes seront encore en vie, et on saura seulement que leur délai de survie est au moins supérieur à leur durée de suivi dans l'étude. Cette caractéristique est appelée censure (ici « à droite »), et exige des méthodes statistiques spécifiques.

C'est le statisticien anglais David Cox qui a proposé en 1972 le modèle, connu aujourd'hui comme le « modèle de Cox », afin d'analyser les données de survie tout en tenant compte de caractéristiques individuelles qui pouvaient influencer la durée de survie. Le modèle de Cox repose sur une modélisation

semi-paramétrique de la fonction de risque instantané associée à une technique d'estimation par vraisemblance partielle. Le succès du modèle de Cox dans le domaine de l'épidémiologie s'est plus récemment étendu à des domaines comme la fiabilité ou les mathématiques financières. Dans cet exemple, ce sont donc en quelque sorte les épidémiologistes qui ont joué le rôle de précurseurs, en donnant naissance à une technique qui est à présent adoptée par d'autres champs disciplinaires.

3.2 Modèles de régression à effets aléatoires

Ces modèles ont été imaginés par les épidémiologistes pour analyser les données issues d'enquêtes longitudinales, ou les ensembles de données structurées de manière hiérarchique. Dans ce type d'étude, un même sujet est observé à plusieurs reprises au cours d'une certaine période. Ou bien, on recueille des données sur des individus qui appartiennent à des sous-groupes homogènes (par exemple : fréquentant la même école, ou habitant la même ville...). De ce fait, les données issues d'un même sujet, ou d'un sous-groupe homogène reflètent l'influence de facteurs communs (facteurs individuels, géographiques, environnementaux, sociaux). Comme il n'est généralement pas possible de faire l'inventaire exhaustif, et encore moins de mesurer ces facteurs de similarité, l'approche statistique consiste à prévoir dans le modèle d'analyse l'existence d'une dépendance stochastique entre les observations d'un même individu ou d'un même sous-groupe. Si cette précaution n'est pas prise, les conclusions des tests statistiques sont incorrectes (car l'estimation de la variance est inexacte). Les modèles à effets aléatoires permettent précisément d'introduire et de modéliser une dépendance stochastique : ils consistent à considérer que la valeur observée chez un sujet donné est la somme d'un effet fixe, commun à tous les sujets, et d'un effet aléatoire reflétant les caractéristiques inobservées du sujet ou du sous-groupe auquel il appartient.

3.3 Travaux de recherche mathématique suscités par l'épidémiologie

Il arrive que des épidémiologistes imaginent une solution heuristique pour répondre à leurs besoins spécifiques, et que l'étude théorique de la solution (et donc de ses conditions de validité) devienne un thème de recherche mathématique au cours des années suivantes.

3.3.1 Méthodes de vraisemblance

Les méthodes de vraisemblance ont été introduites par Fisher dans les années 1920 à 1925 sur des arguments heuristiques. L'étude théorique des propriétés

des estimateurs du maximum de vraisemblance a constitué de 1930 à nos jours un domaine de recherche pour de nombreux mathématiciens parmi lesquels on peut citer Hotelling, Doob, Cramer, Wald, Neyman et LeCam.

La méthode de vraisemblance partielle imaginée par Cox (1971, 1975) a été formalisée et validée sur le plan théorique au cours des 15 années suivantes (Andersen et Gill, 1982 ; Wong, 1986).

3.3.2 Le *bootstrap*

La méthode du *bootstrap* a été décrite en 1979 par Efron (qui se définit lui-même comme un biostatisticien). L'étude des propriétés théoriques de cette méthode s'est faite au cours des années suivantes (1980-1990), engendrant ainsi de nombreux travaux par des mathématiciens comme Bickel, Freedman, Beran et Hall. Giné et Zinn (1990) ont étendu la validité du *bootstrap* aux processus empiriques, dans des espaces fonctionnels très généraux.

3.3.3 Formalisation de l'inférence causale

Comme l'a dit Pearl lui-même, les problèmes et paradoxes de l'épidémiologie (de l'économétrie et des sciences sociales) ont été à l'origine de sa recherche sur la théorie mathématique de la causalité (Pearl, 2000). Réciproquement, les résultats de Pearl apportent aujourd'hui la solution à la question de l'inférence causale à partir de données non expérimentales.

Conclusion

Comme les autres sciences appliquées, les méthodes de l'épidémiologie comportent des méthodes spécifiques — qui concernent à la fois des protocoles de recueil de données et les méthodes d'analyse de ces données — et des méthodes générales de formalisation mathématique. Les interactions entre épidémiologistes et mathématiciens constituent une condition essentielle du développement de la recherche en épidémiologie.

Recommandations

- « Contacts » mathématiciens/épidémiologistes : au sein de quelles structures ?
- Encadrement des utilisateurs de logiciels.

Bibliographie

- Andersen PK., Gill RD. (1982). Cox's regression model for counting processes : a large sample study. *Ann Statist*, **10** : 1100-1120.
- Carrat F., Valleron AJ. (1992). Epidemiologic mapping using the « kriging » method : application to an influenza-like illness epidemic in France. *Am J Epidemiol*, **135** : 1293-1300.
- Castellan G. (2000). Sélection d'histogrammes à l'aide d'un critère de type Akaike. *C. R. Acad. Sci. Paris Sér. I Math*, **330** : 729-732.
- Cox DR. (1972). Regression models and life tables (with discussion). *J Roy Statist Soc B*, **34** : 187-220.
- Cox DR. (1975). Partial likelihood. *Biometrika*, **62** : 269-276.
- Dominici F. et al. (2002). On the use of generalized additive models in time-series studies of air pollution and health. *Am J Epidemiol*, **156** : 193-203.
- Efron B. (1979). Bootstrap methods : another look at the jackknife. *Ann Statist*, **7** : 1-26.
- Giné E., Zinn J. (1990). Bootstrapping general empirical measures. *Ann Prob*, **18** : 851-869.
- Matheron G. (1970). *La théorie des variables régionalisées et ses applications*. Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau.
- Pearl J. (2000). *Causality. Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, UK.
- Propp JG., Wilson DB. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Rand Struc Alg*, **9** : 223-252.
- Siegmund D. (1985). *Sequential analysis. Tests and confidence intervals*. Springer, New-York.
- Wong Hu. (1986). Theory of partial likelihood. *Ann Statist*, **14** : 88-123.

SOUS-CHAPITRE 4.2

Épidémiologie et *data mining* ou fouille de données

GILBERT SAPORTA

1 | Définitions et historique

Le *data mining* que l'on peut traduire par « fouille de données » apparaît au milieu des années 1990 aux États-Unis comme une nouvelle discipline à l'interface de la statistique et des technologies de l'information : bases de données, intelligence artificielle, apprentissage automatique (*machine learning*).

David Hand (1998) en donne la définition suivante : « *Data Mining consists in the discovery of interesting, unexpected, or valuable structures in large data sets* ». Voir également Fayyad *et al.* (1996) et Friedman (1997).

La métaphore qui consiste à considérer les grandes bases de données comme des gisements d'où l'on peut extraire des pépites à l'aide d'outils spécifiques n'est certes pas nouvelle. Dès les années 1970 Jean-Paul Benzécri n'assignait-il pas le même objectif à l'analyse des données ? : « *L'analyse des données est un outil pour dégager de la gangue des données le pur diamant de la véridique nature* ».

On a pu donc considérer que bien des praticiens faisaient du *data mining* sans le savoir.

On confondra ici le *data mining*, au sens étroit qui désigne la phase d'extraction des connaissances, avec la découverte de connaissances dans les bases de données (KDD ou *Knowledge Discovery in Databases*) (cf. Hébrail et Lechevallier, 2003).

Comme l'écrivent ces derniers auteurs :

« *La naissance du data mining est essentiellement due à la conjonction des deux facteurs suivants :*

- *l'accroissement exponentiel dans les entreprises, de données liées à leur activité (données sur la clientèle, les stocks, la fabrication, la comptabilité...) qu'il serait dommage de jeter car elles contiennent des informations-clés sur leur fonctionnement (...) stratégiques pour la prise de décision ;*
- *les progrès très rapides des matériels et des logiciels (...).*

L'objectif poursuivi par le data mining est donc celui de la valorisation des données contenues dans les systèmes d'information des entreprises. »

Les premières applications se sont faites dans le domaine de la gestion de la relation client qui consiste à analyser le comportement de la clientèle pour mieux la fidéliser et lui proposer des produits adaptés. Ce qui caractérise la fouille de données (et choque souvent certains statisticiens) est qu'il s'agit d'une analyse dite « *secondaire* » de données recueillies à d'autres fins (souvent de gestion) sans qu'un protocole expérimental ou une méthode de sondage ait été mis en œuvre.

Quand elle est bien menée, la fouille de données a apporté des succès certains, à tel point que l'engouement qu'elle suscite a pu entraîner la transformation (au moins nominale) de services statistiques de grandes entreprises en services de *data mining*.

La recherche d'information dans les grandes bases de données médicales ou de santé (enquêtes, données hospitalières, etc.) par des techniques de *data mining* est encore relativement peu développée, mais devrait se développer très vite à partir du moment où les outils existent. Quels sont les outils du *data mining* et que peut-on trouver et prouver ?

2 | Les outils

On y retrouve des méthodes statistiques bien établies, mais aussi des développements récents issus directement de l'informatique. Sans prétendre à l'exhaustivité, on distinguera les méthodes exploratoires, où il s'agit de découvrir des structures ou des comportements inattendus, de la recherche de modèles prédictifs où une « réponse » est à prédire, mais on verra plus loin que l'acception du terme « modèle » diffère fondamentalement de son sens habituel.

2.1 Exploration « non supervisée »

2.1.1 Analyse des données : visualisation, classification

Les techniques de projection orthogonale sur des sous-espaces : analyse en composantes principales, analyse des correspondances, permettent de réduire efficacement la dimension, du point de vue du nombre de variables. Les méthodes de classification visent à former des groupes homogènes d'unités en maximisant des critères liés à la dispersion (*k-means*). Des extensions non linéaires (splines, noyaux, etc.) étendent le champ de ces méthodes classiques.

2.1.2 Recherche de règles d'association

Cette méthode est une des innovations du *data mining* : introduite en 1993 par des chercheurs en base de données d'IBM, elle a pour but de rechercher des conjonctions significatives d'événements. Typiquement une règle de décision s'exprime sous la forme : si (A et B) alors C mais il s'agit d'une règle probabiliste et non déterministe. On définit le support de la règle comme la probabilité d'observer à la fois la prémisse X et la conclusion Y : $P(X \cap Y)$ et la confiance comme $P(Y/X)$. Parmi les règles ayant un support et une confiance minimale, on s'intéressera à celles où $P(Y/X)$ est très supérieur à $P(Y)$. Les premières applications ont concerné les achats dans les grandes surfaces : parmi les milliers de références disponibles et les millions de croisements, identifier les achats concomitants qui correspondent à des fréquences importantes. Cette méthode s'étend bien au-delà de ce type d'application. L'originalité tient essentiellement à la complexité algorithmique du problème.

2.2 Prédiction ou apprentissage « supervisé »

Inutile d'évoquer ici les techniques de régression bien connues. La méthode la plus typique du *data mining* est certainement celle des arbres de décision : pour prédire une réponse Y , qu'elle soit numérique ou qualitative, on cherche tout

d'abord la meilleure partition de l'ensemble des données (en général en deux sous-ensembles) issue d'une partition effectuée sur les prédicteurs et on itère dans chacun des sous-ensembles : la croissance exponentielle de l'arbre est contrôlée par des critères d'arrêt de type coût-complexité ainsi que par l'usage de données de validation qui permettent d'éliminer les branches non pertinentes.

Cette technique conduit à des règles de décision très lisibles, d'où son succès, et hiérarchise les facteurs explicatifs. À l'opposé en termes de lisibilité, les logiciels de *data mining* proposent souvent des méthodes hautement non linéaires comme les réseaux de neurones, les machines à vecteurs de support (SVM). Même si les règles de décision ont une forme mathématique explicite, celle-ci est en général très complexe et ces méthodes sont utilisées comme des boîtes noires.

Une autre approche consiste à complexifier des méthodes simples : les arbres de décision étant souvent instables, on va en utiliser plusieurs obtenus sur des données rééchantillonnées par *bootstrap* : la décision finale s'obtient par une procédure de vote s'il s'agit d'un problème de classification, ou de moyenne pour un problème de régression : c'est le *bagging*. Citons également le *boosting*, qui consiste à améliorer des procédures de décision en surpondérant les unités mal classées, et en itérant le processus.

3 | Quelques applications en épidémiologie et santé publique

L'utilisation des méthodes de *data mining* en épidémiologie et santé publique est en forte croissance. Comme dans d'autres domaines, c'est la disponibilité de vastes bases de données historiques (on parle maintenant d'entrepôts de données) qui incite à les valoriser, alors qu'au dire de beaucoup de spécialistes elles sont actuellement sous-utilisées.

Pour n'en citer que deux, la revue *Artificial Intelligence in Medicine* et le *Journal of the American Medical Informatics Association* y consacrent de plus en plus d'articles.

La plupart des publications portent sur les arbres de décision et les règles d'association (Lavrac, 1999). Parmi les domaines traités, mentionnons la recherche de facteurs de risque pour les accidents domestiques, le diabète (Figure 4.2-1), les suicides, les infections nosocomiales (Brossette *et al.*, 1998), la détection de la fraude (*Medicare*, Australie). Ces publications mentionnent souvent la découverte de règles inattendues et efficaces.

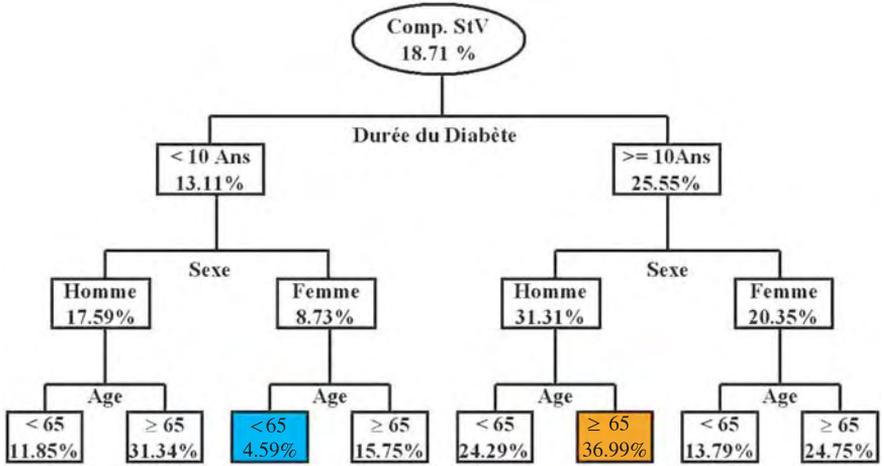


Figure 4.2-1

Prévision de la complication de Saint-Vincent (Projet Data Diab, A. Duhamel, Lille 2).

La recherche en génomique et la protéomique font également de plus en plus appel à des techniques de *data mining*.

4 | *Data mining* versus modélisation statistique

4.1 Le rôle des modèles

La notion de modèle en fouille de données prend un sens particulier : un modèle est une relation entre des variables, exprimable sous une forme analytique ou algorithmique qui ne provient pas d'une théorie mais réalise un bon ajustement aux données. Ainsi, il est courant d'explorer différents modèles (linéaires, non linéaires) en faisant varier les paramètres (nombre de couches dans un réseau de neurones, noyau pour des SVM, etc.) jusqu'à obtenir les meilleures prédictions. On est très loin de la démarche usuelle de modélisation, mais plutôt dans une optique pragmatique où il ne s'agit pas forcément de comprendre mais simplement de prévoir. Rappelons quand même qu'un modèle, au sens classique, n'est qu'une simplification de la réalité et comme le disait George Box : « Tous les modèles sont faux, certains sont utiles ».

Cette démarche n'est pas pour autant du pur empirisme et se fonde sur une théorie solide, celle de l'apprentissage statistique : un modèle réalise un compromis entre sa capacité à rendre compte des données d'apprentissage et sa capacité de généralisation à de nouvelles données. Plutôt que des indices statistiques

de type vraisemblance pénalisée (critères d'Akaike ou de Schwarz) reposant sur des hypothèses distributionnelles, le choix d'un modèle en *data mining* se fait en fonction de ses performances sur d'autres données que celles qui ont servi à le choisir et le caler, d'où l'emploi de méthodes de validation croisée (les données sont divisées en plusieurs parties, chacune étant prédite à l'aide du reste des données) ou de mesures de capacité de type dimension de Vapnik-Cervonenkis.

4.2 Problèmes spécifiques d'inférence et de validation dans les grandes bases de données

L'inférence statistique classique a été développée pour traiter des « petits » échantillons. En présence de très grandes bases de données, le paradoxe est que tout devient significatif : par exemple, pour un million d'individus, l'hypothèse d'indépendance entre deux variables sera rejetée au risque 5 % si le coefficient de corrélation linéaire est supérieur en valeur absolue à 0,002, ce qui est sans intérêt pratique. L'inférence classique ne fonctionne plus et la fouille des grandes bases de données amène à repenser la notion de test et conduit ainsi à des recherches nouvelles.

L'échantillonnage ne perd cependant pas ses droits car il est souvent préférable de travailler sur une partie de la base que sur la totalité. L'exhaustivité des traitements n'est souvent qu'un argument commercial des éditeurs de logiciels. Un problème demeure cependant, celui de la représentativité de la base : même très grande, on ne peut garantir que les futures observations se comporteront comme les passées, d'autant plus que la base n'a pas été constituée à des fins de traitement statistique. Des recherches originales portent sur ce point.

L'alimentation continue des bases de données de façon quasi automatique pose également des problèmes nouveaux connus sous le nom de flots de données (*data streams*) qu'il faut traiter « à la volée » sans devoir reprendre à chaque fois l'ensemble des données disponibles (Domingos et Hulten, 2000).

Quand, en fouille de données, on a exhibé une structure ou une association intéressante et inattendue, on n'est pas certain de sa validité. En effet, avec une exploration combinatoire, il est inévitable de trouver toujours quelque chose ! Ce problème est proche de celui bien connu des comparaisons multiples, mais à une toute autre échelle. On trouve ce genre de situations dans la recherche de règles d'associations ou dans l'analyse des puces à ADN où on réalise des milliers de tests simultanément. La théorie pertinente est celle du contrôle du taux de fausses découvertes de Benjamini et Hochberg (1995) qui fait l'objet de recherches en plein essor, voir l'article de Ge *et al.* (2003).

La « découverte » de règles intéressantes par la fouille de données doit donc être considérée comme une phase exploratoire, nécessitant une validation ultérieure, mais avec des outils différents. Même en cas de validation, le problème de la causalité reste posé.

4.3 Penser la complexité

Il est illusoire de croire que des modèles simples peuvent toujours convenir à des situations complexes. Les modèles de régression (linéaire ou logistique) ont l'avantage de l'interprétabilité, mais cela ne suffit plus en présence de phénomènes fortement non linéaires. Il faut alors souvent plonger les données dans des espaces de grande dimension en utilisant des opérateurs de régularisation.

Le traitement d'images médicales ou de puces à ADN en est une illustration frappante : il pose un défi dû à la fois à la complexité des données et au rapport inhabituel entre le nombre de variables et le nombre d'observations. Le nombre de variables est souvent considérablement plus grand que celui des observations : une image d'un megapixel en couleur correspond à trois millions de variables. . . La théorie de l'apprentissage déjà évoquée (Hastie *et al.*, 2001) fournit le cadre théorique adapté tout en faisant le lien avec des aspects bien connus des statisticiens : estimation fonctionnelle (splines de lissage, estimateurs à noyaux) et régression non paramétrique.

Conclusion

La disponibilité accrue de bases de données médicales de plus en plus vastes (carte Vitale, données hospitalières, grandes enquêtes, etc.) sera un domaine de prédilection pour les méthodes de fouille de données, et des découvertes pourront certainement en être tirées. Ces découvertes doivent être validées par des techniques différentes des tests de la statistique classique.

Une nouvelle forme d'inférence pour les grands ensembles de données est en train d'émerger et le *data mining* est aussi une source de recherches théoriques et pas seulement un ensemble de techniques empiriques. Le *data mining* n'est certainement pas une mode éphémère, mais une démarche et des outils appropriés à l'analyse des très grandes bases de données. Il serait dommageable de laisser ce champ aux seuls informaticiens, car, de par leur formation à l'aléatoire et à la compréhension de la variabilité, statisticiens et épidémiologistes sont les plus à même d'en tirer profit et d'en déjouer les pièges.

Recommandations

- Associer dans les formations universitaires épidémiologie et bio-informatique.
- Enseigner aux épidémiologistes les techniques et les outils de la fouille de données, ainsi que les bases de données.
- Organiser des groupes de travail interdisciplinaires à l'instar du « Dimacs-Working Group on Data Mining and Epidemiology » de Rutgers.

Bibliographie

- Benjamini Y., Hochberg Y. (1995). Controlling the false discovery rate : a practical and powerful approach to multiple testing. *JRSSB*, **57** : 289-300.
- Brossette SE. *et al.* (1998). Association rules and data mining in hospital infection control and public health surveillance. *J Am Med Inform Assoc*, **5** (4) : 373-81.
- Domingos P., Hulten G. (2000). Mining high-speed data streams. ACM SIGKDD, Fayyad U.M., Piatetsky-Shapiro G., Smyth P., Uthurusamy R. (eds.) (1996). *Advances in Knowledge Discovery and Data Mining*. Menlo Park, California, AAAI Press.
- Friedman JH. (1997). Data mining and statistics : what's the connection ? <http://www-stat.stanford.edu/~jhf/ftp/dm-stat.ps>
- Ge Y., Dudoit S., Speed TP. (2003). Resampling-based multiple testing for DNA microarray data analysis. *Test*, **12** : 1-77.
- Hand DJ. (1998). Data mining : statistics and more ? *The American Statistician*, **52** : 112-118.
- Hastie T., Tibshirani R., Friedman J. (2001). *The Elements of Statistical Learning*, Springer.
- Hébrail G., Lechevallier Y. (2003). Data Mining et Analyse des données. In : *Analyse des données*, G. Govaert (ed.). Hermes, 323-355.
- Lavrac N. (1999) Selected techniques for data mining in medicine. *Artificial Intelligence in Medicine*, **16** : 3-33.
- Saporta G. (2000). *Data Mining and Official Statistics*, Quinta Conferenza Nazionale di Statistica, ISTAT, Rome.
- Journal de la SFdS* **142**, 1, (2001) : numéro spécial sur le *data mining*.

SOUS-CHAPITRE 4.3

Épidémiologie et informatique

PIERRE-YVES BOËLLE

Plusieurs pays ont annoncé, voire débuté, le développement d'un dossier médical informatisé individuel « universel », parmi lesquels : **le Royaume-Uni**, dans le cadre du « National IT program for the NHS » (doté de 5 milliards de livres) qui vise 50 millions de dossiers accessibles par 30 000 professionnels ; **les États-Unis** avec le « National Health Information Infrastructure » ; **la France** avec le dossier médical informatisé annoncé en 2004.

Ces projets créent pour la santé publique, et particulièrement l'épidémiologie, l'opportunité de réaliser des études, en population ou à l'hôpital, dans lesquelles des problèmes pratiques couramment rencontrés dans les enquêtes pourraient être réduits en grande partie : description rétrospective de l'histoire d'un sujet, parfois sur plusieurs dizaines d'années ; mise en place une surveillance épidémiologique à « très large spectre » ; coûts engendrés par le recueil, la vérification et la validation des données ; diffusion des résultats auprès de la population.

Les sciences de l'information sont au centre de cette nouvelle évolution de l'épidémiologie. Nous décrivons dans la suite certaines interactions qui se mettent en place aujourd'hui entre les deux disciplines, ainsi que des pistes de développement suscitées par l'évolution des questions et des données épidémiologiques.

1 | Recueil et exploitation des données

Actuellement, un recueil de données spécifique est généralement mis en place pour les besoins d'une enquête épidémiologique. Par exemple, pour connaître la prise en charge des patients infectés par le VIH en France, le dossier médico-épidémiologique de l'immunodéficience humaine, a été développé dès la fin des années 1980, avec une participation forte de l'Inserm et de la Direction des hôpitaux. Ce dossier informatisé, élaboré dans un cadre de recherche, a permis de suivre environ 100 000 patients infectés au cours du temps et sert de base à la connaissance de l'épidémiologie du sida en France.

Depuis les années 1990, des progiciels organisés autour de systèmes de gestion de bases de données sont couramment utilisés, et des études sont réalisées aujourd'hui directement sur Internet sans support papier (McAlindon *et al.*, 2003). Les grands éditeurs de logiciels de systèmes de gestion de base de donnée et d'analyse statistique commercialisent de tels produits, et des solutions ont également été développées localement dans la recherche publique. Quelques petites entreprises (*start-up*) sont présentes, en proposant des solutions logicielles adaptées à des pathologies ou organisations de travail spécifiques. La formalisation du recueil des données, par exemple en définition d'une étude, d'un patient, d'une visite, l'intégration de contraintes de cohérence dans le recueil, le retour d'information vers les investigateurs ont donné lieu à une recherche qui a pu être intégrée dans ces progiciels. Cependant, ceux-ci restent avant tout des « cahiers d'observation électroniques », dédiés à une étude spécifique, plutôt que des véritables dossiers informatisés.

Parallèlement, il existe également des recueils de données en routine qui ont un intérêt épidémiologique : par exemple les décès (Centre d'épidémiologie sur les causes médicales de décès, Inserm – CepiDc, sous-chapitre 3.1), les séjours hospitaliers (Programme de médicalisation du système d'information, PMSI – voir sous-chapitre 3.2), les prescriptions/remboursements (organismes d'Assurance maladie), les données de pharmacovigilance. . .

L'exploitation de ces données, déjà informatisées, a pu servir de base à plusieurs études épidémiologiques : notamment dans l'étude de la mortalité spécifique du cancer et d'autres maladies ayant une morbidité et mortalité conséquentes. Des problèmes pratiques subsistent dans l'exploitation de ces données non recueillies spécifiquement : le temps nécessaire à l'obtention de données consolidées, incompatible avec une réactivité rapide (quelques années pour les décès, un semestre pour l'hôpital), la qualité des données recueillies et de leur codage, souvent auprès de professionnels mal informés des conséquences d'un codage erroné, et la confidentialité.

Le recueil de données de routine permet toutefois la mise en place d'études épidémiologiques rapides. Ainsi, en Angleterre, le « GPRD — General Practitioner Research Database » (Walley et Mantgani, 1997), qui recueille les événements de santé de 5 % de la population anglaise en médecine générale, a donné lieu à des études dans de nombreuses pathologies (plus de 270 à ce jour). Aux États-Unis, certaines HMO (« Health Management Organizations », organisations rassemblant organismes de remboursement et de délivrance des soins) se sont dotées de systèmes de gestion de dossiers médicaux électroniques qui, en exploitant l'évolution longitudinale du patient, permettent de déterminer si les patients sont traités conformément à la meilleure connaissance médicale du moment, et dans certains cas d'appliquer certaines recommandations de santé, au moyen de « pense-bêtes » édités lors des consultations (Garg *et al.*, 2005).

Bien que la France ait été pionnière dans le domaine de la surveillance épidémiologique appuyée sur les technologies de l'information, notamment avec le Réseau national téléinformatique de surveillance et d'information sur les maladies transmissibles (Valleron *et al.*, 1986) (RNTMT, l'ancêtre du réseau « Sentinelles »), cet avantage ne s'est pas maintenu. Il n'y a pas aujourd'hui de système de collection de données de santé utile pour la surveillance épidémiologique doté d'une couverture comparable à celle trouvée dans les pays anglo-saxons (voir par exemple les études Nhanes décrites dans le sous-chapitre 9.3).

L'extension des recueils systématiques à l'ensemble des événements de santé, visée par les dossiers médicaux universels, est évidemment prometteuse. Ces développements posent des problèmes pratiques, par exemple dans le stockage de l'information, sa préservation, ainsi que celui de la puissance de calcul nécessaire à son traitement. Dans les deux projets de dossier électronique de santé anglo-saxons, pour lesquels un cahier des charges détaillé existe déjà, il faut noter que la possibilité d'une recherche épidémiologique de bonne qualité n'est pas mise en avant.

Il est aujourd'hui essentiel pour les épidémiologistes de s'associer plus activement aux spécialistes des sciences de l'information qui détermineront les plans de collection de données de santé, et de favoriser l'accès à ces bases de données ainsi qu'à des moyens de traitement informatique performants.

2 | Confidentialité : la (non-)identification des individus

Le croisement d'informations recueillies pour un même individu dans des bases de données informatisées issues de divers domaines (par exemple : décès,

registres, hôpitaux, prescriptions, génomique, emploi, environnement, administration, météorologie, assurances, etc.) permettrait la mise en place d'enquêtes épidémiologiques aujourd'hui inédites ou impossibles à réaliser. Afin d'augmenter la puissance des études, et de détecter de faibles différences de risques, il est impératif que ces recoupements se fassent sur une base individuelle, c'est-à-dire que l'on puisse relier les observations pour un même individu dans des bases de données différentes.

Un problème informatique central est alors que les règles de confidentialité, souvent différentes pour chaque source, et généralement restrictives lorsqu'il s'agit de données liées à la santé, pourraient être mises à mal lors de ces recoupements. Il est établi par exemple que la connaissance de 80 SNP (*single nucleotide polymorphism*, l'unité de base de la variabilité génétique) suffirait à identifier de manière totalement unique tout individu dans le monde (Lin *et al.*, 2004) !

D'autre part, les pays se sont dotés de lois régissant l'exploitation des données de santé, qui reposent principalement sur la notion de consentement donné pour une recherche particulière. Quelles devront être les aménagements pour permettre l'exploitation des données issues des dossiers médicaux universels ? La gestion globale de la confidentialité dans les bases de données informatisées utilisées ou potentiellement utilisables en épidémiologie peut ainsi être l'origine d'une recherche originale en informatique.

3 | Thésaurus, nomenclature, classification : des outils indispensables en perpétuelle évolution

L'utilisation de bases de données informatisées pour la recherche épidémiologique est subordonnée à l'existence d'une structuration et d'une indexation des informations collectées. Grâce à la recherche en informatique médicale, des solutions satisfaisantes ont déjà émergé. Par exemple, l'ensemble des publications biomédicales sont indexées grâce au vocabulaire contrôlé « MeSH : Medical Subject Headings » (Lowe et Barnett, 1994) (l'analogue du Pacs des médecins).

Les systèmes de classification des maladies (par exemple la Classification internationale des maladies, Cim10), de nomenclature médicale (par exemple Systematic Nomenclature of Medicine, Snomed) (Cote et Robboy, 1980), sont aujourd'hui disponibles, et la recherche en informatique médicale a permis de rendre opérationnels ces systèmes à l'aide par exemple de gestionnaires d'ontologies. Leur intégration dans des applications informatiques médicales

est possible et a largement dépassé le stade de la recherche (dossier patient électronique, système d'aide à la prescription, etc.). Le transcodage entre termes de différentes classifications est possible (Unified Medical Language System, UMLS), bien que non exempt de problèmes. Par ailleurs, des protocoles d'échange assurant l'interopérabilité entre logiciels ont été définis (par exemple HL7 : Health Level 7) et un regain d'intérêt existe pour le développement de ces fonctionnalités en utilisant par exemple XML. Ces composantes font d'ailleurs partie du cahier des charges des projets anglais « National IT Program for the NHS » et américain « National Health Information Infrastructure » pour le dossier médical universel.

Le développement de thésaurus, nomenclatures et classifications demeure l'objet d'une recherche active, mais ils sont d'ores et déjà utilisés en routine. L'apport des technologies de l'information doit être de faciliter leur utilisation pratique, notamment dans les projets à grande échelle et avec une multitude d'intervenants ; ainsi que de garantir l'exhaustivité et la fiabilité des informations recueillies.

4 | Aide à l'analyse de données complexes par la modélisation informatique

Afin d'exploiter pleinement l'information épidémiologique contenue dans les bases de données médicales actuelles et à venir, on doit s'orienter vers l'analyse automatique. Dès à présent, il serait possible d'appliquer des « règles » simples à des données informatisées, à des fins de connaissance et de surveillance (Lindberg *et al.*, 1993) : par exemple pour surveiller l'évolution de l'obésité, ou de l'hypertension artérielle et sa prise en charge.

Cependant, l'analyse informatique, comme l'interprétation, de certaines données médicales ou biologiques pose des problèmes de recherche à l'heure actuelle non résolus. Prenons par exemple la détection du cancer du sein : on s'appuie aujourd'hui sur les mammographies, sur lesquelles on recherche des « microcalcifications ». Il s'avère aujourd'hui impossible de proposer une méthode automatique qui ait des caractéristiques de sensibilité et spécificité compatibles avec une utilisation en population générale. De même, il n'existe pas d'interprétation automatique des données recueillies dans les scanners, ou lors d'échographies, bien que des progrès notables aient été réalisés par exemple dans le recalage de données (par exemple pour réorienter la direction d'une radiographie) ainsi que pour la fusion de données acquises de plusieurs sources différentes.

Un enjeu des sciences et technologies de l'information est donc de développer modèles et techniques qui permettront cette automatisation. Une première étape vise à améliorer l'analyse de données. Ceci repose par exemple sur la composition de données intégrant l'information brute et les résultats de simulations informatiques basées sur des modèles mathématiques de processus physiologiques, pharmacodynamiques et épidémiologiques pertinents. Par ailleurs, les méthodes d'apprentissage basées sur le *data mining* pourraient être mise à profit pour des données très complexes pour lesquelles il n'existe pas de modèle théorique satisfaisant.

Bibliographie

- Cote RA., Robboy S. (1980). Progress in medical information management. Systematized nomenclature of medicine (SNOMED). *JAMA*, **243** : 756-62.
- Garg AX., Adhikari NK., McDonald H. *et al.* (2005). Effects of computerized clinical decision support systems on practitioner performance and patient outcomes : a systematic review. *JAMA*, **293** : 1223-38.
- Lin Z., Owen AB., Altman RB. (2004). Genetics. Genomic research and human subject privacy. *Science*, **305** : 183.
- Lindberg DA., Humphreys BL., McCray AT. (1993). The Unified Medical Language System. *Methods Inf Med*, **32** : 281-91.
- Lowe HJ., Barnett GO. (1994). Understanding and using the medical subject headings (MeSH) vocabulary to perform literature searches. *JAMA*, **271** : 1103-8.
- McAlindon T., Formica M., Kabbara K., LaValley M., Lehmer M. (2003). Conducting clinical trials over the internet : feasibility study. *BMJ*, **327** : 484-7.
- Valleron AJ., Bouvet E., Garnerin P. *et al.* (1986). A computer network for the surveillance of communicable diseases : the French experiment. *Am J Public Health*, **76** : 1289-92.
- Walley T., Mantgani A. (1997). The UK General Practice Research Database. *Lancet*, **350** : 1097-9.

SOUS-CHAPITRE 4.4

Épidémiologie théorique : modèles mathématiques et maladies transmissibles

PIERRE-YVES BOËLLE, NOËL BONNEUIL, JEAN-PIERRE FRANÇOISE,
MICHEL LANGLAIS ET KHASHAYAR PAKDAMAN

L'épidémiologie théorique des maladies transmissibles est devenue une discipline à part entière, distincte de la démographie théorique et de l'écologie théorique, et offre un terrain fertile en applications et en problèmes pour les mathématiques (Hethcote, 2000). Nous illustrons ceci dans ce chapitre, en montrant comment les modèles ont pu stimuler la réflexion sur des problèmes épidémiologiques et devenir des outils d'aide à la décision en santé publique.

Introduction

La modélisation mathématique dynamique d'une épidémie liée à une maladie transmissible permet d'analyser les conditions de son démarrage, de mettre au point et de comparer des stratégies de contrôle ou de prévention. L'importance et l'acceptation croissante de cette approche est manifeste si l'on considère la liste des articles parus dans des revues telles que *Nature*, *Science* et *The New*

England Journal of Medicine au cours des dernières années : évaluation de l'impact des maladies émergentes comme le Sras (Lipsitch *et al.*, 2003 ; Riley *et al.*, 2003), stratégies de vaccination antivariolique en cas d'attaque bioterroriste (Bozzette *et al.*, 2003), surveillance du démarrage d'une possible pandémie grippale (Ferguson *et al.*, 2004), détermination du rapport coût/efficacité des traitements pour l'infection à VIH (Freedberg *et al.*, 2001).

Parallèlement à ces exemples visant à une prévision quantitative, il existe une modélisation plus qualitative dont le but est de formaliser certains mécanismes pour mieux les comprendre. Les maladies à prions fournissent un remarquable exemple d'une telle approche : dès 1967, quinze ans avant que Stanley Prusiner (1982) introduise la dénomination « prion » (Prusiner, 1982), J.S. Griffith décrivait plusieurs modèles biochimiques qui établissaient la plausibilité d'un agent infectieux purement protéique, donc dépourvu de matériel génétique propre (Griffith, 1967). Griffith proposait également que l'agent infectieux de la tremblante du mouton puisse être de cette nature ; ceci a été établi plus tard par Prusiner. Les modèles ont donc également un intérêt « spéculatif », dont le but est d'inciter les expérimentateurs à ne pas négliger ou abandonner une hypothèse qui peut apparaître incongrue de prime abord. Profitons de cet exemple pour insister sur l'importance des données expérimentales comme base à la modélisation, même lorsque celle-ci est spéculative. Des travaux expérimentaux avaient montré que l'agent de la tremblante du mouton résistait aux traitements qui dénaturaient le matériel génétique, mais pas les protéines : cette observation avait motivé Griffith dans sa démonstration de la plausibilité d'un agent infectieux protéique.

Les deux approches décrites ci-dessus illustrent deux utilisations de la modélisation épidémiologique : l'une quantitative pour aider la décision en santé publique, l'autre spéculative pour montrer la plausibilité d'un mécanisme. En pratique, les mêmes techniques peuvent être mises en œuvre dans ces deux démarches, en combinant spéculation et prévision. Ce point est illustré dans la section suivante qui présente un bref historique de la modélisation dynamique des épidémies en faisant ressortir les principaux concepts et les approches. Nous concluons alors sur les nouvelles directions de recherche qui se mettent en place aujourd'hui dans le domaine.

Soulignons dès l'introduction que les problèmes issus de la modélisation des épidémies ont motivé des travaux mathématiques originaux, qui ont acquis une vie indépendante de leur motivation initiale. La théorie de la percolation en est un exemple fameux : celle-ci émerge notamment dans l'analyse de la propagation d'un agent infectieux dans un verger (Broadbent *et al.*, 1957 ; Hammersley, 1957) ; mais les phénomènes de percolation font aujourd'hui partie des domaines classiques d'application de la théorie des probabilités, sans que les motivations initiales soient nécessairement rappelées.

Modèles mathématiques des épidémies

En épidémiologie, on utilise la statistique plus que toute autre théorie mathématique. Une exception notable concerne l'étude des épidémies où des travaux théoriques et pratiques réalisés depuis le début du XX^e siècle ont fait appel à une très large gamme de compétences mathématiques. Bien que les principales revues d'épidémiologie (*American Journal of Epidemiology*, *Epidemiology*, *Epidemiologic Reviews*) publient peu de tels travaux, ce sont ceux qui reçoivent la plus grande exposition dans la communauté scientifique générale : dans l'infection à VIH dans les années 1980-90 (May *et al.*, 1987), l'encéphalopathie spongiforme bovine et le nouveau variant de la maladie de Creutzfeldt-Jacob (Ghani *et al.*, 1998 ; Valleron *et al.*, 2001), la fièvre aphteuse (Keeling *et al.*, 2001), et plus récemment le syndrome respiratoire aigu et sévère (Lipsitch *et al.*, 2003 ; Riley *et al.*, 2003).

Pourquoi la modélisation mathématique a-t-elle pris une telle importance dans l'épidémiologie des maladies transmissibles ? On peut invoquer, *a minima*, la faillite des comparaisons basées sur des extrapolations simples dans ce domaine. En effet, si l'incidence d'un cancer est divisée par deux dans une population de taille pareillement réduite, il n'en irait pas de même pour la taille d'une épidémie, à cause d'effets de seuils décrits plus bas.

Les premiers modèles théoriques modernes sont l'œuvre de En'ko, Hamer et Ross et datent du début du XX^e siècle. Ceux-ci proposent la description « compartimentale » de l'histoire naturelle d'une infection en trois phases : « Susceptibles » avant l'infection, « Infectieux » pour une personne infectée et contagieuse, « Retirés » pour les personnes guéries, immunes ou décédées. La deuxième étape importante est due à Kermack et McKendrick (1927). Enfin, depuis le début des années 1980, dans le sillage de la collaboration prolifique d'un théoricien, Sir Robert May, et d'un écologiste et biologiste, Roy Anderson, l'intérêt des grands journaux généralistes s'est renouvelé (Anderson *et al.*, 1979 ; May *et al.*, 1979).

1 | L'approche « systèmes dynamiques » : équilibre, stabilité, ...

La motivation des travaux de Kermack et McKendrick est leur constat que « La caractéristique la plus notable dans l'étude des épidémies est la difficulté de trouver un facteur causal qui expliquerait la diversité des magnitudes des épidémies ».

Dans les articles correspondant à cette étude (Kermack *et al.*, 1927 ; Kermack *et al.*, 1932 ; Kermack *et al.*, 1933), les auteurs décrivent ce qui est maintenant connu sous le nom de modèle « SIR ». Sous la forme la plus simple, ce modèle correspond au système d'équation suivant :

$$\begin{cases} \frac{dS(t)}{dt} = \lambda - \beta S(t) I(t) - \mu S(t) \\ \frac{dI(t)}{dt} = \beta S(t) I(t) - \gamma I(t) - \mu I(t) \\ \frac{dR}{dt} = \gamma I(t) - \mu R(t) \end{cases}$$

où $S(t)$, $I(t)$, $R(t)$ correspondent au nombre de personnes dans chaque compartiment à un instant t donné, et plusieurs paramètres sont utilisés pour rendre compte de la dynamique de cette population : λ correspond aux naissances dans la population, supposées toutes susceptibles, $1/\mu$ est la durée de vie moyenne, $1/\gamma$ la durée infectieuse moyenne et β correspond au taux de contact menant à une transmission effective de la maladie. En pratique, ce modèle restitue de manière saisissante la forme des épidémies observées et rapportées dans la littérature (figure 4.4-1).

L'analyse des propriétés d'équilibre de ce système d'équations fait apparaître un résultat d'une portée pratique inattendue, le « théorème du seuil » : *pour qu'une épidémie puisse avoir lieu, il faut une communauté susceptible de taille supérieure à un seuil déterminé par deux termes : la contagiosité de l'infection et la durée de la phase infectieuse*. Ce seuil serait par exemple de l'ordre de 200 000 à 300 000 pour la rougeole dans les pays occidentaux (Bailey, 1975), en l'absence de vaccination. D'autre part, la taille de l'épidémie peut facilement s'expliquer en fonction de la fraction de susceptibles de la population avant l'épidémie.

Plusieurs faits surprenants furent également tirés de l'analyse mathématique du modèle :

- la dissémination d'une maladie repose avant tout sur le nombre de personnes qui ne l'ont pas eue, beaucoup plus que sur le nombre initial de personnes infectées ;
- il est normal d'échapper « naturellement » à une épidémie, puisque celle-ci s'arrête avant d'utiliser tous les susceptibles ;
- une situation endémique stable peut prévaloir dans une population, lorsque l'incidence peut être exactement compensée par les retraits de personnes infectieuses.

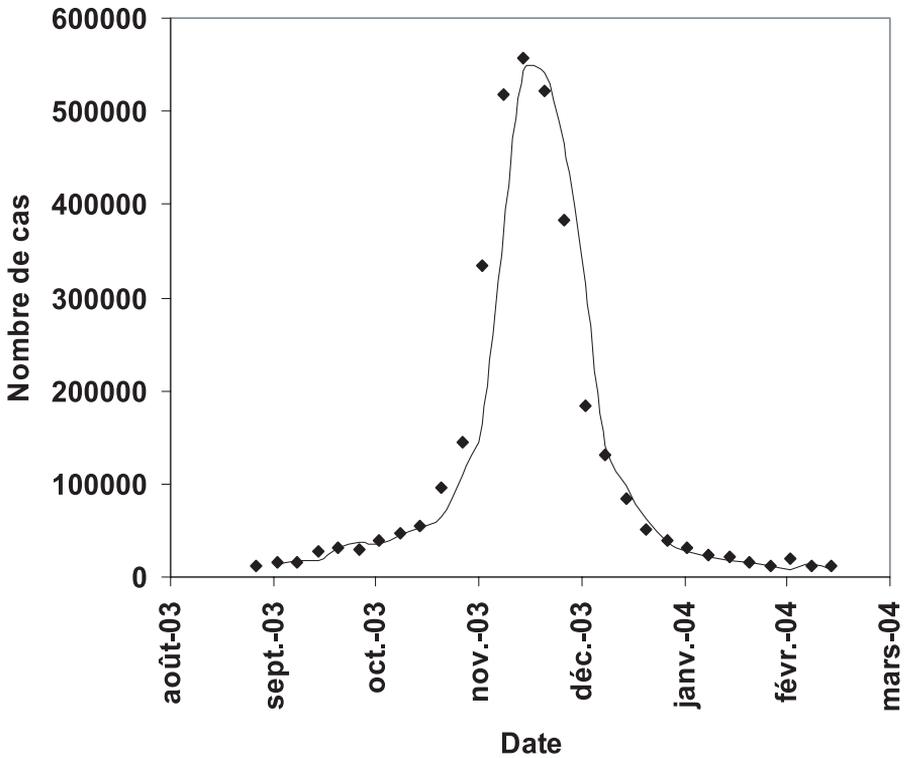


Figure 4.4-1

Courbe d'incidence épidémique. Nombre de cas de syndromes grippaux observés en France en 2003/2004 (données du réseau Sentinelles de l'Inserm) et ajustement d'un modèle mathématique du type SIR (en trait continu).

Le théorème du seuil a été reformulé par la notion du ratio de reproduction de l'épidémie, noté R_0 , dont l'interprétation usuelle est le nombre de cas directement infectés par un unique sujet infecté dans une population entièrement susceptible. L'intuition, et les mathématiques, montrent que lorsque R_0 est plus grand que 1, on aura une épidémie ; et pas lorsque R_0 sera inférieur à 1.

Suivant les travaux de Kermack et McKendrik, de très nombreuses extensions ont été proposées : prendre en compte les effets aléatoires (Bailey, 1975 ; Andersson *et al.*, 2000), structurer les compartiments (voir § 3), estimer des paramètres à partir de données (Dietz, 1993 ; Mollison *et al.*, 1994).

Bernoulli avait montré dès le XVII^e siècle l'intérêt de la modélisation mathématique dans l'étude de la variolisation, pratique précurseur de la vaccination antivariolique (Valleron, 2000). L'étude des conditions d'une vaccination optimale est toujours à l'origine de nombreux développements en épidémiologie

théorique. En effet, du point de vue d'un individu, il semble que la prévention de la maladie passe toujours par la vaccination. Mais l'existence d'un seuil critique de population susceptible en deçà duquel une épidémie est impossible montre qu'il n'en est rien : un individu pourrait échapper à la maladie sans être vacciné dans une telle population : ceci est l'immunité grégaire. Avec cette propriété, il apparaît également possible d'éradiquer une maladie transmissible : si la proportion de susceptibles est toujours inférieure au seuil critique, alors la maladie doit disparaître. Ceci a été le cas avec la variole, et on espère aujourd'hui l'éradication de la poliomyélite et de la rougeole.

La vaccination pose encore aujourd'hui des problèmes nouveaux, pour lesquels la modélisation mathématique demeure indispensable. Notamment, la disparition programmée de maladies amène à réfléchir sur la nécessité de renforcer la vaccination ou de l'arrêter ; la diminution rapide de l'immunité vaccinale est également à l'origine de questions de santé publique que l'on peut étudier par modélisation.

2 | L'approche probabiliste et statistique : processus ponctuels, MCMC

Pour passer de modèles « spéculatifs » à des modèles plus « quantitatifs », la principale difficulté rencontrée est de donner une valeur aux paramètres du modèle. La confrontation des modèles mathématiques d'épidémie avec des données réelles est nécessaire, avec la difficulté que les paramètres des modèles épidémiques ne sont pas tous directement observables : s'il est possible de relier, par exemple, la durée d'excrétion virale ou de colonisation avec la durée de la phase infectieuse, il n'existe pas de moyen simple d'observer la valeur des taux de contact efficaces au cours de la vie infectieuse.

On est donc confronté au problème statistique de la construction d'estimateurs de paramètres inobservables, à partir de données de terrain le plus souvent partielles et/ou agrégées. De nombreux développements ont donné un cadre probabiliste à l'étude des épidémies, et aujourd'hui ceci se fait généralement dans le cadre de la théorie des processus ponctuels (Neyman a d'ailleurs été un précurseur de ce type de description dans les systèmes biologiques), avec des techniques statistiques « de Monte-Carlo ». Des estimateurs aux propriétés de martingales ont été développés (Picard, 1980). L'utilisation de techniques modernes, combinant augmentation de données et chaînes de Markov de Monte-Carlo est également prometteuse (O'Neill *et al.*, 2000).

3 | L'approche « systèmes complexes » : apport des modèles structurés

Les individus d'une population ne sont généralement pas égaux devant l'exposition à une épidémie ; cependant les caractéristiques qui expliquent cette variabilité sont généralement délicates à identifier à partir des données de terrain. Si l'on peut structurer la population selon ces facteurs, on pourra proposer un modèle qui reflète mieux les données de terrain. Une liste de tels facteurs peut comprendre :

l'âge chronologique d'un individu. Ce paramètre intervient dans l'étude des maladies infantiles, lorsque la fertilité est affectée ou que la transmission verticale n'est pas négligeable. Des modèles intégrant l'évolution démographique d'une population sont nécessaires dans ce cas, mais d'analyse mathématique et numérique complexe. Les modèles de Leslie, de Lotka, Sharpe *et al.* intègrent ces aspects ;

la structuration spatiale. La localisation géographique d'un individu peut avoir un fort impact sur la démographie et sur les densités locales d'individus, comme sur les taux de contacts entre individus et les structurations sociales locales. Dans ce cas, on pourra choisir de maintenir le taux de contact constant avec la densité, ou au contraire proportionnel à celle-ci. La structuration spatiale peut être formulée dans des modèles mathématiques de type réaction-diffusion incluant des termes de transport. L'analyse de ceux-ci pose des problèmes d'analyse mathématique très complexes. Cependant, l'exemple récent de la propagation fulgurante du Sras à l'échelle de la planète à partir d'un foyer unique montre toutefois qu'un parasite peut franchir de grandes distances quasi instantanément, invalidant ainsi l'approche de réaction-diffusion. Une représentation sous formes de *patches* du domaine spatial reliés les uns aux autres par des connexions, avec un modèle de type SEIRS dans chaque *patch*, sera alors plus opérationnelle. La complexité viendra alors de la taille des systèmes différentiels à traiter ;

la structuration sociale. Les différents modes de propagation d'un parasite — amical/agressif, aérienne, transmission sexuelle, indirecte, partage de seringues contaminées — n'ont pas le même impact dans une population donnée selon le groupe d'individus considéré. La propagation est essentiellement liée au nombre et au type de contacts efficaces entre individus à l'intérieur d'un groupe ou entre individus appartenant à des groupes distincts. Une fois la population à étudier structurée en groupes socialement distincts, la propagation d'un parasite entre les groupes peut être modélisée par un modèle du type *patches* précédent ;

l'âge de l'infection. La contagiosité d'une maladie transmissible peut changer au cours du temps chez un individu infecté, une surinfection peut survenir ou au contraire l'immunisation se renforcer après une nouvelle exposition. La gestion des échelles de temps entre la durée de vie des individus et la durée de vie d'un infectieux ou la durée du stade infectieux rend alors plus délicate l'étude théorique et numérique du modèle.

4 | L'approche « systèmes complexes » : les réseaux de contact

On a précisé plus haut comment le ratio de reproduction de la maladie (R_0) pouvait être utilisé pour quantifier si les conditions sont propices à la propagation d'un agent infectieux et comment il était devenu une mesure commune du potentiel épidémique d'une maladie. Il est toutefois facile d'oublier que l'importance de ce paramètre, et l'estimation de sa valeur à partir des données, sont indissociables du formalisme du modèle SIR. En d'autres termes, l'existence d'un seuil épidémique est tributaire de la validité du modèle SIR pour cette épidémie.

La théorie du seuil doit-elle être remise en cause ? La structuration des compartiments dans le modèle SIR ne remet pas fondamentalement en cause l'existence même d'un seuil, même si son calcul devient plus compliqué (il est alors défini comme la plus grande valeur propre d'une matrice jacobienne en un point d'équilibre). May et Anderson (1988) ont montré que la propriété de seuil demeurerait, tant que la variance des taux de contacts entre compartiments était finie.

Que se passe-t-il dans le cas contraire ? Il est d'abord nécessaire de développer un nouveau formalisme pour décrire les contacts entre individus. Prenons l'exemple des maladies sexuellement transmissibles (MST) pour illustrer comment les données de terrain sont centrales dans ce problème. Dans ces maladies, on peut obtenir des informations directes et détaillées sur les contacts entre individus (par exemple l'étude ACSF — Analyse des comportements sexuels des Français) et reproduire les caractéristiques des réseaux de partenaires. En pratique, des limitations apparaissent : la simultanéité temporelle des partenariats est généralement mal mesurée, des incohérences notables apparaissent dans les données : par exemple, les hommes déclarent en moyenne plus de partenaires que les femmes. Il est donc impossible de nourrir le modèle à partir des seules données brutes.

Face à ces difficultés, des travaux plus « spéculatifs » sont menés pour identifier les caractéristiques nécessaires et suffisantes des contacts dans la propagation des épidémies. Pastor-Satorras et Vespignani (Pastor-Satorras *et al.*, 2001) ont examiné par exemple la dissémination de virus informatiques sur Internet, et

montrent que le théorème du seuil n'est pas forcément satisfait dans des réseaux où le nombre de connections des sites suit une loi de puissance. D'autres travaux qualitatifs et spéculatifs de même nature ont suivi cet exemple, notamment pour étudier les stratégies qui permettent d'enrayer la propagation d'épidémies dans ces réseaux hautement hétérogènes. En plus de susciter de nombreuses études numériques et exploratoires, l'intérêt de ces résultats a été d'attirer l'attention sur l'importance de bonnes observations pour descriptions et modélisations des mécanismes de contacts.

Ces travaux soulignent la sensibilité des résultats au choix de la structure des contacts entre individus. La conclusion que l'on peut en tirer est donc qu'il est important de ne pas se limiter aux modèles de contact « empruntés » à d'autres disciplines, jadis la chimie et aujourd'hui les réseaux informatiques. Les modèles épidémiologiques doivent s'orienter vers une description des contacts à partir de données et d'observations appropriées.

5 | L'approche « contrôle » appliquée aux systèmes dynamiques : l'exemple de la viabilité

Comment contrôler un système dynamique de sorte qu'une propriété qualitative soit conservée au cours du temps ? Comment manœuvrer un système dynamique de sorte qu'une propriété qualitative soit obtenue avant un temps donné ? Au lieu de considérer le comportement asymptotique du système, la question du maintien dans un ensemble donné se pose à chaque instant, en phase transitoire, avec ou sans attracteurs, sans besoin d'optimiser un critère intertemporel souvent rendu irréaliste par l'incertitude inhérente au comportement futur de l'agent.

L'ensemble de tous les états initiaux à partir desquels il existe une trajectoire restant dans un ensemble fixé correspondant à la propriété qu'on souhaite voir se perpétuer joue un rôle important. On l'appelle le *noyau de viabilité*. L'ensemble des états initiaux à partir desquels il existe au moins une trajectoire atteignant une cible donnée est *le domaine de capture*. En un état donné, l'ensemble des contrôles pour lesquels il existe une trajectoire restant dans l'ensemble fixé ou bien atteignant la cible est *la loi de régulation*.

En épidémiologie des pistes sont ouvertes :

- quelles politiques de vaccination, à quel âge, en quelles proportions ? Cette question a été envisagée par plusieurs auteurs. Cependant, il s'agit souvent d'examiner les conséquences d'une politique donnée. En viabilité,

nous posons le problème à l'envers : quelles mesures de vaccination au cours du temps de sorte que l'épidémie soit toujours maintenue en dessous d'un certain seuil, ou bien soit ramenée en dessous d'un seuil donné ? Les contrôles ne sont pas fixés au départ, mais sont adaptatifs aux changements de l'environnement ;

- sans avoir besoin de spécifier un critère d'optimisation intertemporel, sans avoir besoin de spéculer sur le comportement futur du système (comme des oscillations), l'analyse de viabilité offre des procédures de sélection de devenirs viables, en respectant les contraintes d'état ;
- l'existence, l'unicité et la viabilité des solutions des systèmes dynamiques complexes utilisés en épidémiologie reposent parfois sur des hypothèses de constance des risques instantanés, et mettent en jeu des fonctions supposées intégrables au sens L1, ou dérivables en un sens faible. Le concept d'enveloppe d'invariance issu de la théorie de la viabilité permet d'explorer ces questions dans l'espace beaucoup plus grand des fonctions à graphe fermé, en utilisant le concept judicieux de dérivée contingente. En dynamique des populations, Aubin, Bonneuil et Maurin (Aubin *et al.*, 2000) ont étudié le système de Lotka-McKendrick. Il reste des choses à faire en épidémiologie, où il vaut mieux savoir si les solutions existent, sont uniques, ou sous quelles conditions elles sont viables ;
- équations mutationnelles : lorsque la variable d'état est elle-même un ensemble, l'évolution du système s'écrit sous forme d'équation mutationnelle. Par exemple, quelle politique de contrôle sanitaire mettre en œuvre afin de maintenir un territoire infecté dans des limites prescrites ? Ce travail est en cours et mené par Karine Pichard (université de Pau).

6 | Autres développements récents

Un travail récent, paru dans la revue *Nature*, combine de manière intéressante les nouvelles approches de la modélisation en épidémiologie (Eubank *et al.*, 2004). L'objectif de ce travail est d'étudier l'impact d'une attaque bio-terroriste en milieu urbain et de comparer les interventions possibles. Bien que classique dans ses objectifs et ses conclusions, la méthodologie utilisée rend ce travail remarquable. En effet, plutôt que de faire des hypothèses *a priori* sur les contacts entre individus, les interactions entre individus dans leur modèle intègrent des données réelles de circulation urbaine et de démographie de la ville de Portland aux États-Unis. L'architecture des réseaux de contact n'est donc pas posée *a priori*, comme c'est le cas dans beaucoup de modèles, et n'est pas empruntée aux réactions chimiques ou à Internet : il s'agit d'une approche qui

exploite directement les facteurs qui régissent réellement les contacts entre les personnes.

Cette approche est également liée à de nouvelles techniques de modélisation, plus détaillée, dite « par agents » ou « individu-centrée ». Dans celle-ci, on privilégie une description « locale » des interactions à la description macroscopique. De telles approches sont par exemple essentielles lorsque l'on souhaite pouvoir aborder des aspects « intra-hôte », par exemple l'évolution génétique des parasites, et des aspects populationnels comme les caractéristiques de la transmission entre hôtes. Reposant aujourd'hui avant tout sur des simulations, ces approches demandent le développement d'outils analytiques spécifiques.

Enfin, avec l'explosion des données génomiques, à la fois sur les virus, bactéries ou leurs hôtes, il faut aujourd'hui imaginer des méthodes qui permettront d'intégrer ces données dans la modélisation. Des approches séduisantes ont été proposées, qui exploitent les propriétés d'échantillonnage du nombre de variants génétiques dans des populations d'effectif non stationnaire. Il est alors possible, en utilisant comme seule donnée un échantillon de la population virale à un moment donné, et en connaissant la vitesse de l'horloge moléculaire associée aux mutations, d'estimer la vitesse de croissance de la population porteuse de l'organisme considéré. Des résultats extrêmement convaincants ont été obtenus concernant par exemple l'histoire de l'épidémie d'hépatite C en Égypte (Pybus *et al.*, 2001) ou de l'évolution du virus de la grippe (Ferguson *et al.*, 2003). L'unification des modèles classiques et de ces approches génomiques est un terrain qui semble offrir des possibilités immenses en épidémiologie théorique (Grenfell *et al.*, 2004).

Conclusion

L'épidémiologie théorique des maladies transmissibles est devenue une discipline à part entière, distincte de la démographie théorique et de l'écologie mathématique (Hethcote, 2000). Celle-ci se caractérise par une combinaison de techniques issues de la modélisation des populations, de la virologie et la microbiologie, et offre un terrain d'application immense pour les mathématiques et les statistiques. Au-delà des résultats théoriques, cette discipline vise à fournir des bases quantitatives pour la réflexion de santé publique autour des maladies infectieuses. Dans cette direction, on notera que, si l'utilisation de modèles basés sur des arbres de décision, bien que sans aucune complexité mathématique, devient un outil commun dans la communauté épidémiologique et médicale (Freedberg *et al.*, 2001), l'utilisation de modèles plus vraisemblables (et plus complexes mathématiquement) serait cependant souhaitable.

Le développement de ce domaine doit s'appuyer sur les points suivants :

Renforcer les interactions. Les interactions entre épidémiologistes et mathématiciens/probabilistes sont informelles. On pourrait favoriser l'émergence de projets du type ACI Mathématiques pour l'épidémiologie.

Recueillir des données. On est aujourd'hui arrivé à une bonne connaissance théorique des propriétés des modèles mathématiques simulant des épidémies. Cette étape « spéculative » doit laisser place à une étape de validation, nourrie des données de terrain. Pour cela, celles-ci doivent être recueillies prospectivement et à grande échelle. Pour avoir de bons modèles, utiles en santé publique, il faut avoir avant tout de bonnes données.

Faire des modèles des outils de santé publique. Les effets d'une décision de santé publique peuvent parfois être anticipés en s'appuyant sur des résultats de modélisation. De la part des modélisateurs, cela nécessite de franchir une étape entre aspects « spéculatifs » du modèle et aspects « quantitatifs » utiles pour baser une décision.

Explorer les limites des modèles. Comme indiqué plus haut, si la modélisation peut aider à interpréter des données, elle n'a pas pour finalité de remplacer l'observation et l'expérimentation réelle. Les conditions qui font qu'un modèle est utilisable, que l'on aura confiance dans ses résultats et que ceux-ci pourront convaincre doivent être mieux étudiées.

Bibliographie

- Anderson RM., May RM. (1979). Population biology of infectious diseases : Part I. *Nature*, **280** (5721) : 361-7.
- Andersson H., Britton T. (2000). *Stochastic epidemics and their statistical analysis*. Springer-Verlag, New York.
- Aubin J., Bonneuil N. *et al.* (2000). Non-linear Structured Population Dynamics with Co-Variates. *Mathematical Population Studies*, **9** (1) : 1-31.
- Bailey N. (1975). *The Mathematical Theory of Infectious Diseases*. Griffin, London.
- Bozzette SA., Boer R. *et al.* (2003). A model for a smallpox-vaccination policy. *N Engl J Med*, **348** (5) : 416-25.

- Broadbent S., Hammersley J. (1957). Percolation processes I. *Proc. Cambridge Philos Soc*, **53** : 629-641.
- Dietz K. (1993). The estimation of the basic reproduction number for infectious diseases. *Stat Methods Med Res*, **2** (1) : 23-41.
- Eubank S., Guclu H. *et al.* (2004). Modelling disease outbreaks in realistic urban social networks. *Nature*, **429** (6988) : 180-4.
- Ferguson NM., Fraser C. *et al.* (2004). Public health. Public health risk from the avian H5N1 influenza epidemic. *Science*, **304** (5673) : 968-9.
- Ferguson NM., Galvani AP. *et al.* (2003). Ecological and immunological determinants of influenza evolution. *Nature*, **422** (6930) : 428-33.
- Freedberg KA., Losina E. *et al.* (2001). The cost effectiveness of combination antiretroviral therapy for HIV disease. *N Engl J Med*, **344** (11) : 824-31.
- Ghani AC., Ferguson NM. *et al.* (1998). Estimation of the number of people incubating variant CJD. *Lancet*, **352** (9137) : 1353-4.
- Grenfell BT., Pybus OG. *et al.* (2004). Unifying the epidemiological and evolutionary dynamics of pathogens. *Science*, **303** (5656) : 327-32.
- Griffith JS. (1967). Self-replication and scrapie. *Nature*, **215** (105) : 1043-4.
- Hammersley J. (1957). Percolation processes II. *Proc Cambridge Philos Soc*, **53** : 642-645.
- Hethcote H. (2000). The mathematics of infectious diseases. *SIAM Review*, **42** (4) : 599-653.
- Keeling MJ., Woolhouse ME. *et al.* (2001). Dynamics of the 2001 UK foot and mouth epidemic : stochastic dispersal in a heterogeneous landscape. *Science*, **294** (5543) : 813-7.
- Kermack WO., McKendrick AG. (1927). Contributions to the mathematical theory of epidemics-I. *Proc Roy Soc Lond A*, **115** : 700-721.
- Kermack WO., McKendrick AG. (1932). Contributions to the mathematical theory of epidemics-II. The problem of endemicity. *Proc Roy Soc Lond A*, **138** : 55-83.
- Kermack WO., McKendrick AG. (1933). Contributions to the mathematical theory of epidemics-III. Further studies of the problem of endemicity. *Proc Roy Soc Lond A*, **141** : 94-122.
- Lipsitch M., Cohen T. *et al.* (2003). Transmission dynamics and control of severe acute respiratory syndrome. *Science*, **300** (5627) : 1966-70.

- May RM., Anderson RM. (1979). Population biology of infectious diseases : Part II. *Nature*, **280** (5722) : 455-61.
- May RM., Anderson RM. (1987). Transmission dynamics of HIV infection. *Nature*, **326** (6109) : 137-42.
- Mollison D., Isham V. *et al.* (1994). Epidemics : models and data. *J R Stat Soc Ser A Stat Soc*, **157** (1) : 115-49.
- O'Neill P., Balding D. *et al.* (2000). Analysis of infectious disease data from household outbreaks by Markov chain Monte Carlo methods. *Applied Statistics*, **49** : 517-542.
- Pastor-Satorras R., Vespignani A. (2001). Epidemic spreading in scale free networks. *Phys Rev Lett.*, **86** : 3200-3203.
- Picard P. (1980). Applications of martingale theory to some epidemic models. *Journal of Applied Probability*, **17** : 583-599.
- Prusiner SB. (1982). Novel proteinaceous infectious particles cause scrapie. *Science*, **216** (4542) : 136-44.
- Pybus OG., Charleston MA. *et al.* (2001). The epidemic behavior of the hepatitis C virus. *Science*, **292** (5525) : 2323-5.
- Riley S., Fraser C. *et al.* (2003). Transmission dynamics of the etiological agent of SARS in Hong Kong : impact of public health interventions. *Science*, **300** (5627) : 1961-6.
- Valleron AJ. (2000). The roles for modeling in epidemiology. *C. R. Acad. Sci. III*, **323** (5) : 429-33.
- Valleron AJ., Boëlle PY. *et al.* (2001). Estimation of epidemic size and incubation time based on age characteristics of vCJD in the United Kingdom. *Science*, **294** (5547) : 1726-8.

SOUS-CHAPITRE 4.5

La modélisation en épidémiologie du cancer

FRÉDÉRIC YVES BOIS

L'apport des mathématiques à l'épidémiologie des cancers peut se faire à trois niveaux. Le premier correspond à celui de l'épidémiologie analytique, qui étudie le lien entre des expositions environnementales d'origines diverses (alimentaires, physiques, chimiques), des facteurs constitutionnels (notamment génétiques) et la survenue ou l'évolution de cancers. À ce niveau, la modélisation concerne à la fois le lien cancer/exposition mais aussi la reconstruction des expositions dans le temps. Le second niveau concerne les définitions de seuil dans le cadre des actions de prévention et de contrôle des instances réglementaires, où des techniques de modélisation stochastique des processus de cancérogenèse sont de plus en plus sollicitées. Le dernier niveau concerne l'épidémiologie expérimentale, où de nombreux modèles sont conçus pour l'analyse de données cliniques, en vue du développement de thérapies ou d'outils de pronostic. Si tous ces derniers travaux ont un intérêt appliqué certain, ils ont aussi une portée scientifique et peuvent être l'occasion de développements mathématiques originaux. La figure 4.5-1 résume l'articulation des différentes disciplines de recherche autour de la cancérogenèse et de ses facteurs.

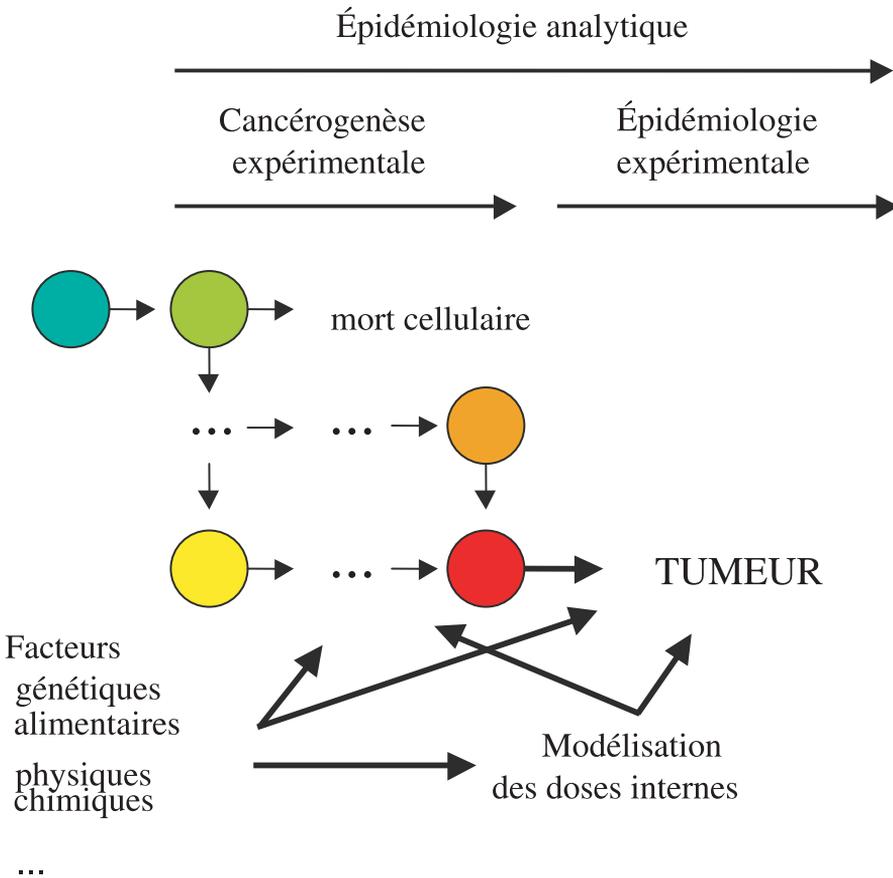


Figure 4.5-1
 Articulation des différentes disciplines de recherche épidémiologique ou toxicologique autour de la cancérogénèse et de ses facteurs étiologiques.
 La multiplicité apparente des étapes de la cancérisation et celles des facteurs la régulant rendent difficile l'identification des causes « mineures » de cancer. Les mathématiques sont utilisées par chacune des disciplines pour modéliser tout ou partie du processus.

1 | Historique et cadre général

C'est à partir de 1950 que les premières théories mathématiques de la cancérogénèse furent développées (Iversen et Arley, 1950; Nordling, 1953). Elles s'appuyaient sur des travaux concernant les processus stochastiques (Kendall, 1952). Leurs motivations et applications visaient à expliquer des résultats expérimentaux chez l'animal ou l'évolution de l'incidence des cas de cancer humains en fonction de l'âge. Naturellement, des applications de prévention ou de pronostic étaient également recherchées (Armitage et Doll, 1954). Les étroites

fenêtres thérapeutiques des chimiothérapies et des thérapies par irradiation poussèrent ensuite au développement de modèles mathématiques quantitatifs pour l'optimisation des doses de traitement atteignant les tumeurs. Ces applications ont ensuite eut une vie propre que bouleversent maintenant les données de la génétique.

1.1 Les mathématiques sont utiles à l'étude de toutes les étapes de la cancérogenèse

1.1.1 Modélisation du processus de cancérogenèse

C'est l'observation épidémiologique d'une évolution non linéaire de l'incidence des cancers en fonction de l'âge qui motiva les premiers travaux de modélisation mathématique de la cancérogenèse (Armitage et Doll, 1954). Ce fut, pour les statisticiens collaborant avec des épidémiologistes, l'occasion de proposer des modèles stochastiques markoviens offrant de bonnes descriptions de l'évolution des taux de cancer avec l'âge (Kendall, 1960 ; Whittemore et Keller, 1978 ; Moolgavkar et Venzon, 1979). L'application des modèles les plus simples à l'étude de la cancérogenèse expérimentale a été nettement moins fructueuse, car ils prennent mal en compte les mécanismes de réparation de l'ADN, les interactions cancer/hôte, ou les phénomènes d'apoptose. Les travaux les plus récents tentent de remédier à cette situation (Bois et Compton-Quintana, 1992 ; Kopp-Schneider, 1997 ; Dunson, 2000 ; van Leeuwen et Zonneveld, 2001 ; Little et Wright, 2003, <http://www.maths.dundee.ac.uk/mbg/>).

1.1.2 Modélisation des étapes cliniques et des traitements thérapeutiques du cancer

Indépendamment du traitement de données d'épidémiologie analytique ou de toxicologie animale, la modélisation de l'évolution de tumeurs avérées a une utilité indéniable pour le dépistage et la mise au point de thérapies anticancéreuses (Tubiana et Koscielny, 1999). Les modèles mathématiques utilisés en épidémiologie expérimentale font appel au calcul différentiel classique (Barbolosi et Iliadis, 2001 ; Simeoni *et al.*, 2004), au calcul stochastique (Ciampi *et al.*, 1986 ; Smolle et Stettner, 1993) mais aussi aux simulations stochastiques (Periti *et al.*, 1979 ; Düchting et Vogelsaenger, 1987 ; Michelson, 1990) ou aux automates cellulaires (Qi *et al.*, 1993). Des développements récents font appel à la théorie des graphes (Radmacher *et al.*, 2001).

1.1.3 Modélisation de la dose interne

La relation entre la dose (de radiation, médicament, cancérigène chimique) administrée et la dose interne, active, n'est pas forcément linéaire. La complexité des mécanismes d'absorption, diffusion, transport, métabolisme, élimination et les cascades d'événements secondaires menant à la dose interne font que l'on s'est très tôt orienté vers des modèles basés sur l'anatomophysiologie (Teorell, 1937). C'est dans les années 1970 que les progrès du calcul ont permis l'utilisation courante des modèles pharmacocinétiques « physiologiques » (Dedrick *et al.*, 1970) ou des fantômes numériques (Kerr *et al.*, 1976). Les modèles actuels font le plus souvent appel à la résolution numérique de systèmes différentiels ordinaires ou aux dérivées partielles. Ils sont utilisés en thérapeutique, en clinique, ou en évaluation des risques pour prendre en compte des différences physiologiques entre individus, affiner le calcul des relations dose-réponse, et reconstruire les expositions passées.

1.1.4 Les liens avec la statistique

Les modèles les plus complexes s'accommodent mal de méthodes fréquentistes classiques d'estimation des paramètres. Des développements intéressants font appel aux modèles statistiques hiérarchiques (Greenland, 2000), aux statistiques bayésiennes (Begg *et al.*, 1998) à l'analyse non linéaire de séries temporelles (De Gunst et Luebeck, 1994 ; Davidian et Giltinan, 1995).

2 | Évolution, besoins, perspectives

L'épidémiologie expérimentale ou clinique dispose de « bons » modèles, validés, prédictifs, opérationnels. La recherche dans ce domaine doit se poursuivre pour intégrer les progrès des outils de dépistage, de diagnostic, et de traitement. Une grande partie de cette recherche est directement financée par l'industrie de la santé.

L'épidémiologie analytique est plus mal lotie. Les modèles existants intègrent mal la multiplicité des causes de cancers ; les modèles les plus solides nécessitent le recours à des techniques statistiques complexes. Cependant, des modèles simplifiés, voire simplistes, comme le modèle multi-étapes dans l'évaluation du risque ou les approximations par des modèles linéaires ou logistiques en épidémiologie descriptive, restent encore largement utilisés, car considérés bien souvent comme suffisants pour décrire correctement l'évolution temporelle des

cancers. On peut s'accommoder de ces modèles simplifiés s'il s'agit d'interpoler des données, mais ils n'ont guère de justification scientifique quand il s'agit d'évaluer les risques aux faibles doses ou pour de faibles durées d'exposition.

Comment améliorer les modèles utilisés pour l'évaluation quantitative de la relation dose-risques ou celle de l'impact sur les populations ? Il est nécessaire de mieux comprendre les régulations fines des premières étapes de la cancérogenèse. L'analyse quantitative des régulations génétiques de la différenciation cellulaire, discipline faisant fortement appel à la statistique et aux mathématiques, devrait faire progresser la question. La solution ne sera probablement pas simple puisque plus de 1 % des gènes sont susceptibles de mutations contribuant au développement de cancers (Futreal *et al.*, 2004). Pour s'en approcher, il est nécessaire de développer et financer des recherches intégratives alliant biologie moléculaire, toxicologie, bio-informatique et épidémiologie.

Bibliographie

- Armitage P., Doll R. (1954). The age distribution of cancer and a multi-stage theory of carcinogenesis. *British Journal of Cancer*, **8** : 1-12.
- Barbolosi D., Iliadis A. (2001). Optimizing drug regimens in cancer chemotherapy : a simulation study using a PK-PD model. *Computers in Biology and Medicine*, **31** : 157-172.
- Begg CB. *et al.* (1998). A new strategy for evaluating the impact of epidemiologic risk factors for cancer with application to melanoma. *Journal of the American Statistical Association*, **93** : 415-426.
- Bois FY., Compton-Quintana PGE. (1992). Sensitivity analysis of a new model of carcinogenesis. *Journal of Theoretical Biology*, **159** : 361-375.
- Ciampi A. *et al.* (1986). Multi-type Galton-Watson process as a model for proliferating human tumour cell populations derived from stem cells : estimation of stem cell self-renewal probabilities in human ovarian carcinomas. *Cell and Tissue Kinetics*, **19** : 129-140.
- Davidian M., Giltinan DM. (1995). *Nonlinear Models for Repeated Measurement Data*. London, Chapman & Hall.
- De Gunst MC., Luebeck EG. (1994). Quantitative analysis of two-dimensional observations of premalignant clones in the presence or absence of malignant tumors. *Mathematical Biosciences*, **119** : 5-34.
- Dedrick RL. *et al.* (1970). Interspecies correlation of plasma concentration history of methotrexate (NSC-740). *Cancer Chemotherapy Reports*, **54** : 95-101.

- Düchting W., Vogelsaenger T. (1987). An approach of modelling and simulating the spread of heterogeneous tumor cells in a three-dimensional tissue segment. *Computers and Mathematics with Applications*, **14** : 783-792.
- Dunson DB. (2000). Models for papilloma multiplicity and regression : applications to transgenic mouse studies (with erratum). *Applied Statistics*, **49** : 19-30.
- Futreal PA. et al. (2004). A census of human cancer genes. *Nature Reviews Cancer*, **4** : 177-183.
- Greenland S. (2000). Principles of multilevel modelling. *International Journal of Epidemiology*, **29** : 158-167.
- Iversen S., Arley N. (1950). On the mechanism of experimental carcinogenesis. *Acta Pathologica Microbiologica Scandinavica*, **27** : 773-803.
- Kendall DG. (1952). Les processus stochastiques de croissance en biologie. *Annales de l'Institut Henri Poincaré*, **13** : 43-108.
- Kendall DG. (1960). Birth and death processes and the theory of carcinogenesis. *Biometrika*, **47** : 13-21.
- Kerr GD. et al. (1976). A mathematical model of a phantom developed for use in calculations of radiation dose to the body and major internal organs of a Japanese adult. *Journal of Radiation Research*, **17** : 211-229.
- Kopp-Schneider A. (1997). Carcinogenesis models for risk assessment. *Statistical Methods in Medical Research*, **6** : 317-340.
- Little MP., Wright EG. (2003). A stochastic carcinogenesis model incorporating genomic instability fitted to colon cancer data. *Mathematical Biosciences*, **183** : 111-134.
- Michelson S. (1990). A system for Monte-Carlo simulation of heterogeneous tumor cell populations. *Computers and Mathematics with Applications*, **20** : 139-148.
- Moolgavkar SH., Venzon J. (1979). Two-events models for carcinogenesis : incidence curves for childhood and adult tumors. *Mathematical Biosciences*, **47** : 55-77.
- Nordling CO. (1953). A new theory on the cancer-inducer mechanism. *British Journal of Cancer*, **7** : 68-72.
- Periti P. et al. (1979). Mathematical models of the neoplastic growth control. *Medical Biology and Environment*, **7** : 185-214.

- Qi, A-S. *et al.* (1993). A cellular automaton model of cancerous growth. *Journal of Theoretical Biology*, **161** : 1-12.
- Radmacher MD. *et al.* (2001). Graph models of oncogenesis with an application to melanoma. *Journal of Theoretical Biology*, **212** : 535-548.
- Simeoni M. *et al.* (2004). Predictive pharmacokinetic-pharmacodynamic modeling of tumor growth kinetics in xenograft models after administration of anticancer agents. *Cancer Research*, **64** : 1094-1101.
- Smolle J., Stettner H. (1993). Computer simulation of tumour cell invasion by a stochastic growth model. *Journal of Theoretical Biology*, **160** : 63-72.
- Teorell T. (1937). Kinetics of distribution of substances administered to the body. *Archives Internationales de Pharmacodynamie et de Thérapie*, **57** : 205-240.
- Tubiana M., Koscielny S. (1999). The rationale for early diagnosis of cancer — The example of breast cancer. *Acta Oncologica*, **38** : 295-303.
- van Leeuwen IMM., Zonneveld C. (2001). From exposure to effect : a comparison of modeling approaches to chemical carcinogenesis. *Mutation Research-Reviews in Mutation Research*, **489** : 17-45.
- Whittemore AS., Keller JB. (1978). Quantitative theories of carcinogenesis. *SIAM Reviews*, **20** : 1-30.

CHAPITRE 5

Épidémiologie et sciences biologiques : quatre exemples

SOUS-CHAPITRE 5.1

Épidémiologie génétique des maladies infectieuses

LAURENT ABEL, ALEXANDRE ALCAÏS ET JEAN-LAURENT CASANOVA

L'épidémiologie, dans sa dimension analytique, s'attache à identifier les facteurs de risque des pathologies. Historiquement, les premiers facteurs de risque considérés étaient extérieurs à l'individu (exogènes), par exemple de nature chimique, physique ou biologique, afin de rendre compte des distributions spatio-temporelles des maladies étudiées. Rapidement, des facteurs de risque propres à l'individu (endogènes), comme l'âge ou le sexe, ont été pris en compte dans les analyses réalisées. Puis, ces facteurs endogènes ont été raffinés avec l'identification de tout un ensemble de paramètres biologiques, en particulier mesurés dans le sang, pouvant être parfois traités comme de véritables phénotypes intermédiaires de la pathologie finale. Enfin, il est devenu possible de considérer des marqueurs génétiques permettant ainsi de tester le rôle de facteurs de risque réellement constitutionnels (et transmissibles/héritables) des individus. L'étude de ces facteurs génétiques nécessite des méthodes d'analyse spécifiques qui ont donné naissance à l'épidémiologie génétique.

Parmi les facteurs de risque biologique, les agents infectieux tiennent une place à part. Tout d'abord, parce que l'étude des épidémies est à l'origine même de l'épidémiologie. Ensuite, dès le milieu du XIX^e siècle, la théorie microbienne de Pasteur démontrait que les agents infectieux sont la cause de nombreuses

pathologies, ce qui, *a priori*, mettait un terme au débat sur l'origine des maladies infectieuses. Néanmoins, si l'agent infectieux est nécessaire, il est, en général, loin d'être suffisant au développement de la pathologie dont il est responsable. Cette variabilité de réponse à un agent infectieux constitue donc une opportunité unique d'étudier le rôle de facteurs de risque « modificateurs », en particulier constitutionnels de l'individu, lorsque celui-ci est exposé au facteur de risque causal indispensable à la survenue de la maladie. L'épidémiologie génétique des maladies infectieuses cherche à atteindre cet objectif et représente une illustration remarquable de l'interaction entre épidémiologie et biologie.

1 | La variabilité de réponse à un agent infectieux

Une observation épidémiologique fondamentale dans le domaine des maladies infectieuses est la grande variabilité individuelle de réponse aux agents infectieux. Globalement, elle se traduit par le fait qu'une proportion des sujets exposés à un agent infectieux ne développe pas de maladie clinique. Il s'agit bien sûr d'une observation très ancienne, en particulier dans des contextes épidémiques, dont, par exemple, on peut retrouver la notion dans ce vers de La Fontaine tiré des *Animaux malades de la peste* : « Ils ne mourraient pas tous, mais tous étaient frappés ». Cette variabilité qui dépend de l'agent infectieux considéré et du contexte géographique et historique dans lequel on se situe, n'est pas toujours simple à évaluer. Par exemple, dans la tuberculose, on estime qu'environ 10 % des sujets infectés par le bacille développeront une tuberculose clinique au cours de leur vie (Casanova et Abel, 2002). Cette seule constatation est néanmoins fondamentale puisqu'elle implique que la majorité des sujets sont naturellement résistants au bacille. Cette variabilité a également été retrouvée dans des contextes non naturels d'infection. Chez l'homme, on peut citer l'exemple du dramatique épisode provoqué par l'injection accidentelle d'une souche virulente de *M. tuberculosis* à Lübeck en 1926. Sur 251 enfants qui reçurent la même quantité de mycobactérie, 77 décédèrent, 127 eurent des signes radiologiques de la maladie, tandis que 47 restèrent asymptomatiques. Bien sûr, la très grande majorité des résultats dans un contexte d'infections expérimentales proviennent des modèles animaux qui ont largement confirmé le rôle majeur des facteurs génétiques de l'hôte dans le contrôle des infections (Buer et Balling, 2003).

Une autre observation majeure, également assez ancienne, est que cette variabilité de réponse n'est pas distribuée au hasard dans la population, mais qu'il existe souvent une forte corrélation familiale. Si les modes de transmission de l'agent infectieux peuvent souvent rendre compte d'une certaine dépendance familiale, il est vite apparu qu'ils ne pouvaient pas à eux seuls expliquer l'ensemble du phénomène. Plusieurs études épidémiologiques ont cherché à quantifier ce

risque familial indépendamment du risque inhérent à la contagion. On peut citer ici un remarquable travail réalisé au Danemark sur des sujets ayant été placés très tôt dans des familles adoptives, et qui a estimé les risques de décès précoces de ces sujets liés à certaines grandes pathologies (cancers, maladies vasculaires, infections) selon les causes de décès précoces de leurs parents adoptifs et de leurs parents biologiques (Sorensen *et al.*, 1988). Lorsque le parent biologique était considéré, le risque relatif le plus fort (5,8) de décès était observé pour les causes infectieuses. À l'inverse, le risque relatif lié aux infections n'était pas différent de 1 quand le parent adoptif était considéré, ce qui suggérait fortement une importante composante génétique dans la prédisposition aux décès précoces de cause infectieuse. On peut aussi en rapprocher les travaux montrant de fortes différences entre différents groupes ethniques dans les prévalences de nombreuses maladies infectieuses, comme la tuberculose (Stead *et al.*, 1990). D'autres approches, plus spécifiques de l'épidémiologie génétique, ont confirmé et affiné ces résultats. Parmi celles-ci, les études de jumeaux, qui ont permis de retrouver un taux de concordance pour la maladie plus grand chez les jumeaux monozygotes que chez les dizygotes, en particulier dans la lèpre et la tuberculose (Casanova et Abel, 2002). Plus récemment, l'analyse de ségrégation, qui teste spécifiquement si les distributions familiales observées d'un certain phénotype, par exemple une maladie, sont compatibles avec la transmission d'un gène influençant de façon majeure ce phénotype, est une méthode qui a permis de détecter le rôle d'un gène majeur dans des pathologies mycobactériennes, parasitaires et virales (Abel et Dessein, 1998 ; Alcais et Abel, 2004).

L'interprétation, en particulier biologique, et les différentes hypothèses que l'on peut formuler à partir de ces observations, est indiquée dans la figure 5.1-1 qui résume les différents facteurs pouvant jouer un rôle dans les deux grandes étapes du processus conduisant de l'exposition à un agent infectieux au développement d'une maladie clinique. Les facteurs influençant directement l'exposition au microbe ont bien sûr un effet majeur sur la première étape qu'est l'acquisition de l'infection elle-même. Les facteurs propres à l'agent infectieux (par exemple des facteurs de virulence) peuvent potentiellement intervenir à tous les stades du processus. Enfin, les facteurs de l'hôte, génétiques et non génétiques (par exemple l'âge ou l'état nutritionnel), jouent un rôle majeur dans toutes les étapes du processus, essentiellement par leur influence sur la réponse immunitaire à l'agent infectieux. La part respective de ces différents facteurs est souvent difficile à apprécier du fait de l'absence d'études les évaluant dans leur globalité, et dépend beaucoup de l'échelle spatiotemporelle à laquelle on se place. Les facteurs d'exposition sont bien sûr au premier rang pour expliquer les différences de prévalence observées dans beaucoup de maladies infectieuses à l'échelle mondiale, alors que des facteurs microbiens (par exemple la survenue de mutations augmentant la virulence d'une souche) peuvent rendre compte du déclenchement d'une épidémie. En revanche, il est maintenant clair que les facteurs de l'hôte, en particulier génétiques, sont les plus importants pour expliquer

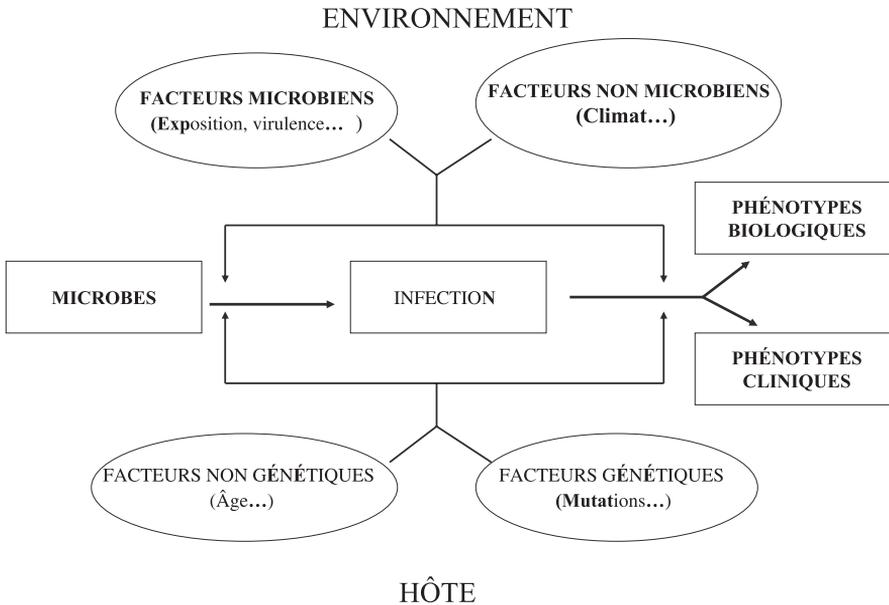


Figure 5.1-1

Facteurs influençant la réponse à un agent infectieux.

Le développement d'une infection après une exposition à un microbe dépend de la réponse immunitaire innée et/ou adaptative. Une fois l'infection établie, des phénotypes biologiques peuvent être mesurés et des manifestations cliniques peuvent éventuellement survenir. Chacune de ces étapes est sous le contrôle de facteurs de l'environnement et de l'hôte. Les facteurs environnementaux peuvent être microbiens (mode d'exposition, facteurs de virulence...) ou non microbiens (conditions climatiques...). Les facteurs de l'hôte peuvent être génétique (variants de l'ADN) ou non génétiques (âge, état général, lésions cutanéomuqueuses).

la variabilité observée dans des populations vivant dans une zone limitée (par exemple le même village) sur une période de quelques années. Ainsi, dans une étude portant sur une population brésilienne vivant dans un village endémique pour le parasite *Schistosoma mansoni*, responsable de la bilharziose hépatointestinale, la variabilité des niveaux d'infection, mesurés par le nombre d'œufs du parasite dans les selles, était expliquée à environ 25 % par des facteurs d'exposition (importance du contact avec l'eau infestée par les larves parasitaires) et l'âge, et à plus de 40 % par l'effet d'un gène majeur (Abel *et al.*, 1991). Au total, toute étude globale des facteurs influençant la réponse à un agent infectieux devra intégrer à la fois une composante épidémiologique « classique » et une composante biologique « au sens large ». C'est l'objectif des approches d'épidémiologie génétique.

2 | Les méthodes de l'épidémiologie génétique

Les méthodes de l'épidémiologie génétique associent des informations de nature épidémiologique, telles que la mesure de facteurs de risque connus pour influencer l'expression de la maladie étudiée (en particulier les facteurs influençant le risque de contamination par l'agent infectieux et/ou modulant les capacités immunitaires de l'hôte), et de nature génétique, comme les liens familiaux entre les individus de l'échantillon ou encore le typage de marqueurs. L'établissement de la carte du génome humain sur la base de marqueurs hautement polymorphes (Dib *et al.*, 1996) et la disponibilité croissante de marqueurs dialéliques (ou SNPs pour *Single Nucleotide Polymorphisms*) (Wang *et al.*, 1998) sont maintenant des outils fondamentaux pour ces études d'épidémiologie génétique. Ainsi, du fait de l'évolution de l'information disponible, l'épidémiologie génétique est maintenant beaucoup « plus génétique » qu'il y a seulement 7 ou 8 ans ! De nombreuses méthodes ont été (et sont) développées, qui ont chacune leurs avantages et leurs inconvénients détaillés dans les paragraphes suivants. De fait, il n'existe pas de stratégie optimale unique pour identifier les gènes impliqués dans les maladies infectieuses humaines, ou plus généralement dans les maladies dites « multifactorielles », et le choix du meilleur protocole pour une étude donnée dépend de nombreux facteurs relatifs au phénotype (nature, fréquence, distribution familiale...), à la population, aux moyens disponibles et aux connaissances préalables (Alcaïs et Abel, 2004).

Schématiquement, deux situations polaires peuvent être rencontrées en génétique humaine des maladies infectieuses. Dans des infections rares et souvent sévères, la structure familiale (par exemple des parents consanguins) ou les relations familiales entre les sujets atteints, sont évocateurs d'une hérédité simple monogénique. C'est le cas du syndrome de prédisposition mendélienne aux infections par des mycobactéries peu virulentes détaillé plus bas. Le plus souvent, la prédisposition génétique apparaît plus complexe et polygénique. Toutefois, cette distinction, quoique classique, est un peu artificielle puisque d'autres gènes peuvent influencer substantiellement l'expression clinique d'une prédisposition dite « mendélienne » et que dans un contexte polygénique, un gène, alors distingué comme majeur, peut avoir un effet clairement prédominant. En conséquence, si certains aspects des différentes stratégies visant à identifier les facteurs génétiques dans ces situations extrêmes peuvent être spécifiques, la plupart seront similaires, comme l'illustre la figure 5.1-1, en proposant une démarche commune depuis le phénotype jusqu'à la caractérisation fonctionnelle des mutations. Les paragraphes suivants présentent une vue d'ensemble des deux outils méthodologiques les plus utilisés en épidémiologie génétique, l'analyse de liaison génétique et les études d'association.

2.1 L'analyse de liaison génétique

Dans l'étude des traits complexes comme les maladies infectieuses, l'analyse de liaison (*linkage*) a pour objectif de localiser des régions chromosomiques contenant un ou des gène(s) d'intérêt, soit en s'intéressant à quelques régions candidates, soit en effectuant un criblage du génome. L'intérêt majeur de cette dernière approche, qui est maintenant la plus commune, est de s'assurer que l'ensemble des gènes majeurs impliqués dans le contrôle du phénotype soient identifiés. En particulier, le criblage du génome permet de découvrir de nouveaux gènes, et donc de nouveaux mécanismes physiopathologiques qui n'étaient pas identifiés *a priori*. Néanmoins, même lorsqu'elles sont fructueuses, les analyses de liaison aboutissent à l'identification de régions chromosomiques s'étendant le plus souvent sur 10 à 20 centimorgans (environ 10 à 20 mégabases) et pouvant contenir plusieurs centaines de gènes. On est donc ici dans une stratégie de génération d'hypothèses par des méthodes de nature statistique, hypothèses qui devront être validées par d'autres approches épidémiologiques (études d'association) puis biologiques (études fonctionnelles).

Le principe général de l'analyse de liaison est de rechercher des marqueurs génétiques, et donc des régions chromosomiques, qui ségrègent de façon non aléatoire avec le phénotype d'intérêt au sein des familles (Ott, 1999). En fonction des connaissances disponibles ou des hypothèses que l'on est prêt à faire sur le mode de transmission génétique du phénotype étudié, on distingue les méthodes modèle-dépendantes (*model-based*) des approches modèle-indépendantes (*model-free*). L'analyse modèle-dépendante par la méthode classique dite « des *lod-scores* » (Morton, 1955) nécessite de définir le modèle génétique, c'est-à-dire de spécifier la relation entre le phénotype et les facteurs qui peuvent influencer son expression, en particulier la présence d'un gène diallélique (d, D). Dans le contexte d'un phénotype binaire (malade/sain), ce modèle génétique devra spécifier, en plus de la fréquence de l'allèle de susceptibilité D, le vecteur des pénétrances c'est-à-dire la probabilité qu'un individu soit atteint, sachant son génotype (dd, Dd, ou DD) et la valeur de certains facteurs de risque pertinents (âge, niveau d'exposition à l'agent pathogène...). Ce modèle génétique est généralement estimé par l'analyse de ségrégation qui est la première étape pour déterminer le mode de transmission d'un phénotype donné à partir d'un échantillon familial. Le but de l'analyse de ségrégation est donc de discriminer entre différents facteurs (environnementaux ou génétiques) pouvant expliquer une ressemblance familiale, avec comme objectif principal de tester l'existence d'un gène appelé « gène majeur ». Le terme « majeur » ne signifie pas qu'il s'agit du *seul* gène impliqué dans le contrôle du phénotype mais que, parmi les gènes impliqués, il en est un qui a un effet suffisamment important pour être distingué des autres.

La méthode « des *lod-scores* » est certainement la plus puissante méthode d'analyse de liaison génétique lorsque le modèle génétique spécifié est correct. C'est le cas dans des situations d'hérédité monogénique où un modèle génétique simple peut être supposé. Une illustration élégante dans le cadre du syndrome de susceptibilité aux infections mycobactériennes fut l'identification de *IFNGR1*, un gène codant une chaîne du récepteur de l'interféron gamma (Jouanguy *et al.*, 1996 ; Newport *et al.*, 1996), en utilisant une méthode d'analyse de liaison modèle-dépendante particulière, la cartographie par homozygotie (*homozygosity mapping*) qui fait l'hypothèse d'un allèle de susceptibilité récessif rare dans des familles consanguines (Lander et Botstein, 1987). Cependant, une spécification incorrecte du modèle peut conduire à une perte sévère de puissance pour détecter la liaison. Dans le cadre des maladies multifactorielles complexes, où le modèle génétique est par essence difficilement identifiable, il est souvent préférable d'utiliser des méthodes d'analyse de liaison, dites « modèles-indépendantes », qui ne nécessitent pas d'explicitier la relation phénotype-génotype (Ott, 1999). La méthode la plus populaire est la méthode des paires de germains (frères ou sœurs) qui recherche si deux germains ayant une certaine ressemblance phénotypique (par exemple deux germains atteints dans le cas d'un phénotype malade/sain) ont une ressemblance, au niveau du marqueur génétique, plus grande que celle attendue du fait de leur seul lien de parenté. Cette ressemblance au marqueur est mesurée par la proportion d'allèles hérités d'un même ancêtre, c'est-à-dire les allèles identiques par descendance ou IBD (en espérance, cette proportion est égale à 50 % pour des germains lorsque la ségrégation est aléatoire). Le test de liaison génétique consiste alors à estimer la proportion d'allèles IBD au sein d'un échantillon de paires de germains atteints et à la comparer à la valeur attendue sous l'hypothèse nulle d'absence de liaison génétique, c'est-à-dire 50 %. De très nombreuses extensions de ces méthodes modèle-indépendantes ont été développées, par exemple pour l'analyse de phénotypes quantitatifs et d'apparentés plus éloignés que des germains (Alcaïs et Abel, 2004).

2.2 Les études d'association

Les études d'association sont utilisées pour tester le rôle de polymorphismes (définis classiquement par une fréquence d'au moins 1 % dans la population) au sein de gènes définis comme candidats soit du fait de leur fonction (gène candidat par hypothèse), soit du fait de leur localisation (gène candidat par expérience) (Figure 5.1-2). Les études d'association classiques sont des études en populations de type cas/témoins où l'on compare la fréquence d'un allèle particulier, ou plus généralement la distribution des génotypes, pour un marqueur génétique, en général un SNP, entre un échantillon de sujets non apparentés atteints et un échantillon de sujets non apparentés sains (Sasieni, 1997). On se retrouve ainsi dans une situation très proche d'une étude épidémiologique

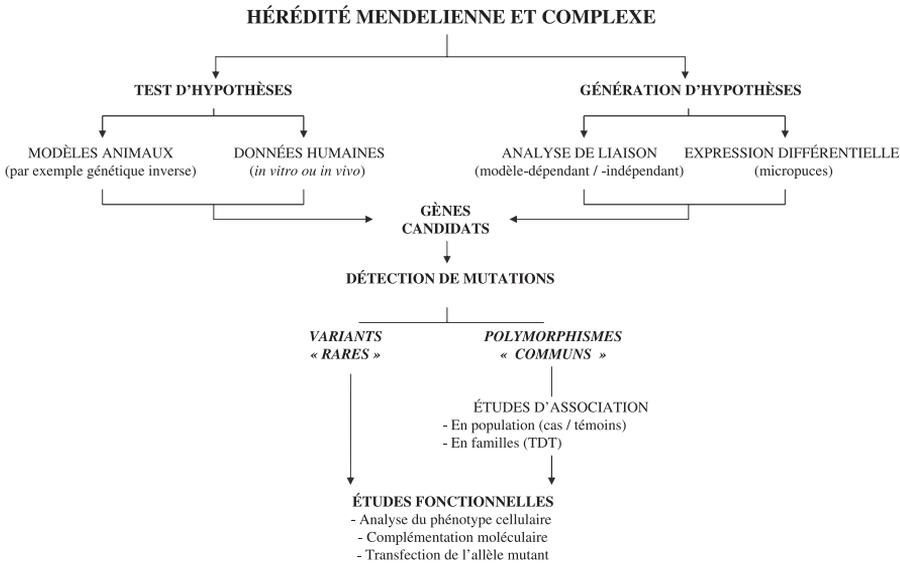


Figure 5.1-2

Identification des gènes de prédisposition aux maladies infectieuses chez l'homme.

Deux stratégies peuvent être envisagées. L'approche « test d'hypothèses » nécessite de spécifier des gènes candidats sur la base des connaissances déjà acquises sur le phénotype étudié ou un phénotype similaire (modèles expérimentaux, données humaines *in vitro*. . .). Pour les traits monogéniques, où des mutations rares ont un fort impact fonctionnel, le lien entre le génotype et le phénotype cellulaire sera en général relativement aisé à établir par les études fonctionnelles. Dans le cas des traits complexes, le rôle de polymorphismes dans un gène candidat sera testé par des études d'association en population ou en famille. La validation fonctionnelle des polymorphismes associés sera souvent difficile puisque leur effet sera a priori subtil et relativement modeste. En l'absence de gène candidat, ou lorsque l'approche gène candidat s'est révélée infructueuse, on pourra générer des hypothèses par un criblage du génome. Actuellement, ces criblages sont le plus souvent réalisés par des études de liaison génétique qui peuvent être modèle-dépendantes (model-based, c'est-à-dire la méthode des lod scores) ou modèle-indépendantes (model-free, par exemple les méthodes de paires de germains). Cependant, il est probable que les criblages fonctionnels (à l'aide des biopuces) seront de plus en plus utilisés dans les années à venir.

classique où le facteur de risque testé est de nature génétique. Une association statistiquement significative entre un polymorphisme donné et un phénotype peut avoir plusieurs interprétations :

- la variation allélique est causale de la variation phénotypique, c'est-à-dire le polymorphisme testé est le polymorphisme fonctionnel lui-même ;
- le polymorphisme testé n'a aucun rôle direct sur le phénotype mais il est en déséquilibre de liaison avec le polymorphisme fonctionnel ;
- l'association est due à une stratification de population ;
- elle est survenue du seul fait du hasard, c'est-à-dire, il s'agit de l'erreur de type I.

Dans l'objectif d'identifier le gène impliqué, seules nous intéressent les deux premières interprétations. Afin de mieux comprendre ce point, considérons deux locus L1 et L2. L1 présente un polymorphisme avec deux allèles D et d, et D est l'allèle fonctionnel augmentant le risque de développer la maladie. L2 présente également deux allèles, A et T, et correspond au locus qui a été génotypé et testé. L'explication la plus simple d'une association positive entre l'allèle A de L1 et la maladie étudiée est que A est lui-même le variant fonctionnel (L1 et L2 sont confondus et A correspond à D). La seconde explication est que A n'ait pas de rôle direct sur le phénotype mais soit en déséquilibre de liaison avec D. Déséquilibre de liaison signifie liaison génétique entre L1 et L2, généralement assez forte (par exemple L1 et L2 peuvent être dans le même gène), ET l'allèle A est préférentiellement associé à D, c'est-à-dire l'haplotype A-D est plus fréquent que le simple produit des fréquences de A et de D (par exemple, la plupart des cas présents sont dus au même allèle ancestral D et l'ancêtre qui a transmis cet allèle était porteur de l'haplotype A-D). Il est d'ailleurs à noter que la détermination de ces haplotypes (l'ensemble des allèles de différents polymorphismes situés sur un même chromosome) est souvent difficile et fournit un exemple instructif de la dialectique incessante entre épidémiologie et biologie moléculaire. Tandis que l'épidémiologie a proposé des solutions *in silico* le plus souvent structurées autour des méthodes de maximum de vraisemblance (Tregouet *et al.*, 2004), la biologie moléculaire a mis au point une technique permettant de déterminer directement *in vitro* un haplotype entier qui pourrait soulager l'épidémiologie génétique de cette sérieuse difficulté (Michalatos-Beloin *et al.*, 1996).

La troisième interprétation, la stratification de population, est une situation potentiellement fréquente qui peut conduire à une association « non informative » entre un marqueur génétique et un phénotype. Ainsi, une étude cas-témoins conduite dans une population qui serait un mélange de deux sous-populations dont l'une présente à la fois une fréquence de la maladie et de l'allèle A plus élevée que la seconde conclurait à une association entre A et la maladie, sans que l'allèle A (ou un autre allèle proche) ait le moindre rôle dans la maladie en question. Les deux approches qui ont été proposées pour surmonter ce problème sont une parfaite illustration du va-et-vient permanent entre épidémiologie et biologie moléculaire. La biologie moléculaire a proposé une solution dite « des contrôles génomiques » (Devlin et Roeder, 1999). L'idée consiste à génotyper en plus du marqueur d'intérêt un certain nombre de marqueurs sélectionnés aléatoirement sur le génome. On compare la distribution de ces marqueurs entre les deux échantillons de cas et de témoins permettant ainsi de détecter un certain niveau de stratification cryptique de populations, d'où on déduit un facteur correctif pour le test du marqueur principal (plus la différence sera grande sur des marqueurs aléatoires, plus le facteur correctif sera fort). L'épidémiologie a proposé d'éliminer les biais liés au choix de témoins « indépendants » en utilisant des témoins familiaux. C'est le fondement du développement récent des études d'association familiales comme le Test du Déséquilibre de Transmission

(TDT) (Spielman *et al.*, 1993). Dans cette approche, l'unité d'échantillon est totalement différente puisqu'il s'agit d'un trio constitué d'un cas et de ses deux parents. Le principe général de la méthode est ensuite d'utiliser les allèles parentaux non transmis aux cas comme allèles témoins de ceux qui sont effectivement présents chez le cas. On se retrouve en quelque sorte dans une situation d'étude appariée (chaque allèle du cas a son propre allèle contrôle correspondant à l'allèle parental non transmis) et ces données peuvent d'ailleurs s'analyser par des méthodes de régression logistique conditionnelle (Schaid et Rowland, 1998) comme dans les études cas/témoins appariées classiques.

Enfin la dernière interprétation, l'erreur de type I, mérite une place à part entière en raison de l'explosion des capacités de la génétique moléculaire, avec comme conséquence l'inflation importante du nombre d'hypothèses proposées pour validation à l'épidémiologie génétique. Du point de vue méthodologique, il a fallu trouver des solutions adaptées à cette situation de tests multiples le plus souvent corrélés. Cette problématique se pose d'ailleurs de la même façon dans le contexte précédemment évoqué des criblages du génome par analyse de liaison génétique où plusieurs centaines de marqueurs génétiques sont testés. Une première approche conservatrice s'est appuyée sur la correction de Bonferroni, c'est-à-dire directement multiplier la *p-value* observée par le nombre de tests réalisés. Cependant, des alternatives plus adaptées à ce contexte de tests multiples où la structure de corrélation peut être spécifiée ont été développées. Cela a été le cas dans les contextes de criblage complet où la modélisation d'un événement rare (un test significatif) sur le génome entier s'est appuyée sur la distribution de Poisson (Lander et Kruglyak, 1995), puis plus récemment dans la situation des études d'association en se fondant sur un calcul du nombre de tests efficaces à partir de la variance des valeurs propres dérivées de la matrice de corrélation entre les marqueurs testés (Nyholt, 2004). Une fois encore, la biologie moléculaire a été un moteur important de la recherche en épidémiologie.

3 | Illustration : les infections mycobactériennes

La lèpre, causée par *Mycobacterium (M.) leprae*, est l'infection mycobactérienne la plus fréquente après la tuberculose, avec environ 600 000 nouveaux cas diagnostiqués annuellement dans le monde (Jacobson et Krahenbuhl, 1996). L'expression de la maladie résulte de l'interaction entre l'agent pathogène et le système immunitaire du sujet infecté. Tandis que la plupart des individus développent une immunité efficace contre l'infection, les autres vont présenter des manifestations cliniques très variées en fonction de l'intensité de la réponse immunitaire du sujet. Ce spectre clinique s'étend entre deux formes dites « polaires » (pôle tuberculoïde et pôle lépromateux) *via* de nombreuses formes interpolaires encore appelées « *borderline* » (Jacobson et Krahenbuhl, 1996).

Chez la souris, il a été démontré que la réponse à l'infection par le bacille de la lèpre murine, *M. lepraemurium*, était en grande partie sous le contrôle d'un gène situé sur le chromosome 1 et appelé *Nramp1* pour la résistance naturelle (Vidal *et al.*, 1993), et sous la dépendance de gènes liés au complexe majeur d'histocompatibilité pour la résistance acquise (Lam-Yuk-Tseung et Gros, 2003). Chez l'homme, l'existence d'une composante génétique est maintenant admise sur la base de très nombreuses études, en particulier épidémiologiques (Casanova et Abel, 2002).

Dans un premier temps, les biologies expérimentale et humaine ont proposé de nombreux gènes candidats par hypothèse pour être testés par l'épidémiologie génétique. Si certains parallèles se sont avérés infructueux, d'autres ont permis une certaine dissection des mécanismes génétiques humains mis en jeu dans la réponse à *M. leprae*. Ainsi, la première étape du processus conduisant de l'exposition à l'infection par le bacille est vraisemblablement en partie sous l'influence de l'homologue humain du gène murin *Nramp1*, dénommé *NRAMP1* (Alcais *et al.*, 2000). Au sein de ces individus infectés, une minorité va présenter des manifestations cliniques qui vont ensuite se polariser vers une forme tuberculoïde ou lépromateuse. De nombreuses études ont confirmé l'implication de la région HLA (en particulier les gènes de classe II) (Casanova et Abel, 2002), et suggéré le rôle d'autres gènes comme celui codant le Toll Like Receptor 2 (Krutzik *et al.*, 2003), dans le processus final de polarisation. Toutefois, la composante génétique responsable des manifestations cliniques *per se* (c'est-à-dire, indépendantes car en amont des formes cliniques polarisées) n'avait pas pu être déterminée à partir des hypothèses de la génétique moléculaire.

Pour identifier le contrôle génétique de la lèpre *per se*, un criblage complet du génome a été entrepris sur un échantillon de familles vietnamiennes qui a montré une liaison génétique très significative avec une région chromosomique située en 6q25 (Mira *et al.*, 2003). Cette région a ensuite été saturée par des marqueurs dialléliques (SNPs), et une association a été retrouvée avec des SNPs localisés dans un bloc de 80 kilobases chevauchant la région régulatrice commune au gène *PARK2*, dont des mutations sont responsables de certaines formes juvéniles de maladie de Parkinson, et au gène co-régulé *PACRG* (Mira *et al.*, 2004). Des études haplotypiques ont montré que la combinaison de deux allèles à risque sur deux SNPs significatifs était associée à un risque cinq fois plus élevé de développer la maladie. Ce résultat a été retrouvé dans une population brésilienne, identifiant ainsi ces variants dans la région régulatrice partagée par *PARK2* et *PACRG* comme des facteurs de risque universels pour la lèpre *per se*. Fonctionnellement, les deux gènes sont impliqués dans l'une des cascades majeures de dégradation des protéines cellulaires, la cascade d'ubiquitination-protéolyse, une voie qui n'avait jamais été évoquée dans ce contexte de l'infection par *M. leprae*. En miroir du paragraphe précédent, cette étude d'épidémiologie

génétiq ue a fourni une nouvelle hypothèse physiopathologique que la génétique moléculaire humaine et expérimentale devra valider et préciser.

L'autre maladie mycobactérienne commune, la tuberculose, causée par *M. tuberculosis*, connaît actuellement une résurgence inquiétante. Elle pose des problèmes de santé publique bien supérieurs à la lèpre puisque environ un tiers de la population mondiale est infectée par *M. tuberculosis*, et un rapport de l'OMS estimait entre 8 et 9 millions le nombre de nouveaux cas cliniques en 2000 avec 1,9 millions de décès dus à la maladie (WHO, 2001). Comme pour la lèpre, il existe de nombreuses observations épidémiologiques en faveur du rôle majeur des facteurs génétiques de l'hôte dans la réponse au bacille : grande variabilité individuelle de réponse (environ 90 % des sujets infectés ne développent pas la maladie clinique, dramatique accident de Lübeck précédemment mentionné), fortes différences interethniques, ressemblance plus grande entre jumeaux monozygotes qu'entre jumeaux dizygotes (Casanova et Abel, 2002). . . Néanmoins, contrairement à la lèpre, il y a eu jusqu'à présent peu d'études familiales d'épidémiologie génétique dans la tuberculose. Un seul criblage du génome a été publié à ce jour sur un échantillon assez hétérogène avec des résultats plutôt décevants (Bellamy *et al.*, 2000). La majorité des travaux réalisés ont été des études d'association explorant le rôle de certains gènes candidats dont nous avons discuté les limites, avec la nécessité de reproduire les résultats et, idéalement, de les valider par des études fonctionnelles. Il est clair que cette validation fonctionnelle est loin d'être simple car l'effet de ces polymorphismes au niveau biologique est probablement modéré et donc difficile à mettre en évidence de façon concluante. Jusqu'à présent, les polymorphismes de certains antigènes HLA de classe II et du gène *NRAMP1* sont ceux pour lesquels les résultats les plus convaincants ont été observés (Casanova et Abel, 2002), même si aucune base moléculaire de leur effet n'a été établie. Ceci contraste avec l'identification récente de mutations causales de certaines infections mycobactériennes qu'il est particulièrement intéressant de décrire dans ce contexte.

Une relation causale a été démontrée, par des méthodes de biologie moléculaire, entre certaines mutations rares de cinq gènes intervenant dans l'immunité médiée par l'interféron (IFN) γ et le syndrome de prédisposition sélective à des infections par des mycobactéries peu virulentes comme le BCG et les mycobactéries environnementales dites « non tuberculeuses » (MNT) (Casanova et Abel, 2002 ; Casanova et Abel, 2004). L'IFN γ est une des principales cytokines activatrices des macrophages et agit grâce à un récepteur qui se compose de deux chaînes (IFN γ R1 et IFN γ R2) et induit par l'intermédiaire de la molécule STAT1 (*signal transducer and activator of transcription*), la transcription de gènes cibles. Différentes mutations des gènes codant pour ces trois molécules (*IFNGR1*, *IFNGR2*, *STAT1*) détériorent ainsi les réponses cellulaires à l'IFN γ . La sécrétion d'IFN γ est induite après la fixation de l'interleukine (IL)12 et de l'IL23, deux hétérodimères qui partagent la sous-unité p40, sur leurs récepteurs

respectifs qui comportent chacun deux chaînes dont une est commune, IL12R β 1. Le défaut de la sous-unité p40 ou de la chaîne IL12R β 1 causé par des mutations dans les gènes *IL12B* ou *IL12RB1*, entraîne une production diminuée d'IFN γ , et est également responsable d'une prédisposition aux infections mycobactériennes.

Il est très intéressant de noter que, dans ce syndrome, il existe une forte corrélation entre le génotype et le phénotype clinique. Les défauts complets de réponse à l'IFN γ sur les gènes *IFNGR1*, *IFNGR2*, et *STAT1* sont responsables d'infections disséminées et en général fatales dans la petite enfance, alors que les défauts partiels de ces mêmes gènes ou les défauts portant sur les gènes *IL12RB1* et *IL12B* sont responsables de phénotypes plus atténués (infections locales et récurrentes) se révélant parfois à l'âge adulte. De plus, la pénétrance pour les défauts complets en IL12R β 1 est incomplète, et a été estimée autour de 40 % dans une large étude dans les familles des patients (Fieschi *et al.*, 2003). Enfin, plusieurs sujets porteurs de ces déficits génétiques ont présenté une tuberculose clinique, ce qui représente les premiers cas de prédisposition mendélienne confirmée à *M. tuberculosis* (Casanova et Abel, 2004). Un enfant avec un défaut complet d'IFN γ R1, et un autre avec un défaut d'IL12/23p40 ont développé une tuberculose clinique, après avoir été atteints d'infections par BCG/MNT. De plus, plusieurs patients présentant soit un défaut partiel d'IFN γ R1 (un cas) soit un défaut complet d'IL12R β 1 (quatre cas) ont présenté une tuberculose clinique comme seul phénotype.

Il est clair que ces observations de tuberculose dues à ces mutations mendéliennes sont rares et n'expliqueront certainement pas les millions de cas annuels. Elles sont néanmoins fondamentales car elles démontrent formellement l'existence d'une prédisposition génétique à la tuberculose. Une hypothèse séduisante consiste alors à considérer cette susceptibilité causale à la tuberculose chez de rares individus comme l'extrémité d'un spectre continu de prédisposition dont l'autre extrémité correspondrait à une prédisposition beaucoup plus modérée et polygénique conférée par des allèles plus fréquents (ou une combinaison de certains de ces allèles) dans la population générale (Abel et Casanova, 2000). Cette hypothèse est également confortée par les résultats d'une analyse de liaison génétique sur une grande généalogie d'Amérindiens canadiens (Greenwood *et al.*, 2000) qui montrait un effet génétique majeur (risque relatif de 10), et donc une situation intermédiaire entre une hérédité purement mendélienne et une prédisposition polygénique. Cette étude est une avancée importante car elle indique que, dans certains contextes épidémiologiques, la prédisposition génétique à la tuberculose peut résulter de l'effet d'un gène majeur, qui est relativement plus facile à identifier et à valider sur le plan fonctionnel que les polymorphismes à effet modéré.

Conclusion

L'ensemble de ces résultats sur les maladies mycobactériennes offre une remarquable illustration de l'interaction indispensable entre épidémiologie et biologie dans la recherche des facteurs génétiques de l'hôte. Les explorations approfondies sur le plan immunologique et génétique moléculaire de certains patients ont ainsi permis de valider, au moins dans certains contextes, l'hypothèse concernant le rôle des facteurs génétiques de l'hôte dans la variabilité de réponse à un agent infectieux, qui avait été formulée à partir des observations épidémiologiques. Ainsi, le fait que seules certaines personnes développent des infections graves par des mycobactéries peu virulentes, alors que la quasi-totalité de la population y est exposée, est expliqué par des facteurs génétiques de l'hôte. La démonstration formelle reste à établir pour les infections beaucoup plus communes par les mycobactéries plus virulentes, mais les premières observations de tuberculose « mendélienne » sont déjà une première étape. De plus, ces études de génétique moléculaire ont généré de nouvelles hypothèses sur les gènes candidats à tester dans les infections communes. Ainsi, le gène *IL12RB1* a fait l'objet d'exploration dans la tuberculose pulmonaire, et une première association avec des polymorphismes du promoteur de ce gène a été décrite au Maroc (Remus *et al.*, 2004). Bien sûr, le rôle fonctionnel de ces polymorphismes reste à établir.

De façon symétrique, les études d'épidémiologie génétique menées sur la lèpre ont conduit à formuler des hypothèses biologiques totalement nouvelles sur les voies de réponse à *M. leprae* (Mira *et al.*, 2004). Fonctionnellement, les deux gènes identifiés, *PARK2* et *PACRG*, sont impliqués dans la cascade d'ubiquitination-protéolyse, dont la finalité est d'identifier au niveau cellulaire les protéines devenues inutiles afin d'initier leur dégradation par le protéasome. Grâce à son activité E3 ubiquitine-ligase, *PARK2* reconnaît des substrats spécifiques et induit leur polyubiquitination qui est un signal de ciblage vers la voie de dégradation par le protéasome. Ainsi, des mutations dans *PARK2* sont responsables de formes juvéniles de maladie de Parkinson, probablement par accumulation de substrats de *PARK2* conduisant à la dégénérescence des neurones dopaminergiques de la voie nigro-striatale (Corti et Brice, 2003). L'implication de cette cascade d'ubiquitination-protéolyse dans la réponse à *M. leprae* devra maintenant être validée, et ses mécanismes fins disséqués par les techniques expérimentales de la biologie moléculaire. De façon plus large encore, cette découverte inattendue de l'épidémiologie génétique soulève ainsi la possibilité fascinante que des anomalies dans la façon dont les cellules gèrent les protéines superflues (c'est-à-dire, les protéines devenues inutiles ou reconnues comme étrangères) pourraient être à l'origine de pathologies assez communes et très différentes dans leur présentation, comme des maladies infectieuses ou des maladies neurodégénératives.

L'épidémiologie génétique, telle que nous l'avons présentée tout au long des paragraphes précédents, s'intéresse aux relations entre les variations observées directement au niveau de l'ADN et les phénotypes appréciés au niveau de l'individu et/ou de la population. L'explosion technologique récente en biologie a introduit une étape intermédiaire entre l'ADN et l'individu en générant un grand nombre de nouveaux phénotypes *au niveau cellulaire*, comme ceux relatifs à l'expression des gènes (transcriptome) ou des protéines (protéome). Dans un premier sens, il s'agit d'une évolution. Ces mesures cellulaires pourront être considérées comme des phénotypes à part entière sur lesquels les concepts et les outils de l'épidémiologie génétique classique, telle que nous l'avons définie plus haut, seront appliqués. Dans un second sens il s'agit d'une révolution. En effet, il est maintenant permis d'étudier le lien entre les variations au niveau de l'ARN (le transcriptome) ou de la protéine (le protéome) et les phénotypes mesurés au niveau de l'individu et/ou de la population. En miroir de la carte génétique définie sur des critères de distance génétique (basée sur la recombinaison chromosomique) ou physique (basée sur la séquence d'ADN), on peut imaginer une carte fonctionnelle des gènes reposant sur des « distances fonctionnelles » définies à partir d'analyses de corrélation, d'analyses de cluster, d'analyses multidimensionnelles ou encore d'arbres de classification (Horvath et Baur, 2000). Il s'agit d'une telle modification conceptuelle qu'une nouvelle discipline pourrait bien voir le jour : « l'épidémiologie post-génomique ». En effet, on peut concevoir une cartographie fonctionnelle où les distances fonctionnelles pourraient être comparées à des distances attendues (semblable à l'analyse de liaison génétique) et différents éléments/variants d'une même cascade physiologique seraient plus ou moins associés au phénotype d'intérêt (semblable aux études d'association). Ces approches nécessitent des méthodes d'analyse résolument nouvelles, dont certaines sont déjà en cours de développement, et représenteront probablement un champ d'investigation majeur dans les années à venir. L'épidémiologie génétique symbolise donc un exemple unique où l'interaction entre deux disciplines, l'épidémiologie et la biologie, trouve un accomplissement dans la genèse d'une troisième, hybride et pourtant féconde !

Bibliographie

- Abel L., Casanova JL. (2000). Genetic predisposition to clinical tuberculosis : bridging the gap between simple and complex inheritance. *Am J Hum Genet*, **67** : 274-7.
- Abel L., Demenais F., Prata A., Souza AE., Dessein A. (1991). Evidence for the segregation of a major gene in human susceptibility/resistance to infection by *Schistosoma mansoni*. *Am J Hum Genet*, **48** : 959-70.
- Abel L., Dessein AJ. (1998). Genetic epidemiology of infectious diseases in humans : design of population-based studies. *Emerg Infect Dis*, **4** : 593-603.

- Alcaïs A., Abel L. (2004). Application of genetic epidemiology to dissecting host susceptibility/resistance to infection illustrated with the study of common mycobacterial infections. *In : Susceptibility to infectious diseases : the importance of host genetics*, Bellamy R. (ed). Cambridge University press, Cambridge, 7-44.
- Alcaïs A., Sanchez FO., Thuc NV., Lap VD., Oberti J., Lagrange PH., Schurr E., Abel L. (2000). Granulomatous Reaction to Intradermal Injection of Lepromin (Mitsuda Reaction) Is Linked to the Human NRAMP1 Gene in Vietnamese Leprosy Sibships. *J Infect Dis*, **181** : 302-308.
- Bellamy R., Beyers N., McAdam KP., Ruwende C., Gie R., Samaai P., Bester D., Meyer M., Corrah T., Collin M., Camidge DR., Wilkinson D., Hoal-Van Helden E., Whittle HC., Amos W., van Helden P., Hill AV. (2000). Genetic susceptibility to tuberculosis in Africans : a genome-wide scan. *Proc Natl Acad Sci, USA*, **97** : 8005-9.
- Buer J., Balling R. (2003). Mice, microbes and models of infection. *Nat Rev Genet*, **4** : 195-205.
- Casanova JL., Abel L. (2002). Genetic dissection of immunity to mycobacteria : the human model. *Annu Rev Immunol*, **20** : 581-620.
- Casanova JL., Abel L. (2004). The human model : a genetic dissection of immunity to infection in natural conditions. *Nat Rev Immunol*, **4** : 55-66.
- Corti O., Brice A. (2003). Maladie de Parkinson : que nous apprennent les gènes des formes familiales ? *Med Sci (Paris)*, **19** : 613-9.
- Devlin B., Roeder K. (1999). Genomic control for association studies. *Biometrics*, **55** : 997-1004.
- Dib C., Faure S., Fizames C., Samson D., Drouot N., Vignal A., Millasseau P., Marc S., Hazan J., Seboun E., Lathrop M., Gyapay G., Morissette J., Weissenbach J. (1996). A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature*, **380** : 152-4.
- Fieschi C., Dupuis S., Catherinot E., Feinberg J., Bustamante J., Breiman A., Altare F. *et al.* (2003). Low Penetrance, Broad Resistance, and Favorable Outcome of Interleukin 12 Receptor beta1 Deficiency : Medical and Immunological Implications. *J Exp Med*, **197** : 527-535.
- Greenwood CM., Fujiwara TM., Boothroyd LJ., Miller MA., Frappier D., Fanning EA., Schurr E., Morgan K. (2000). Linkage of tuberculosis to chromosome 2q35 loci, including NRAMP1, in a large aboriginal Canadian family. *Am J Hum Genet*, **67** : 405-16.

- Horvath S., Baur M. (2000). Future directions of research in statistical genetics. *Stat Med*, **19** : 3337-3343.
- Jacobson RR., Krahenbuhl JL. (1999). Leprosy. *Lancet*, **353** : 655-60.
- Jouanguy E., Altare F., Lamhamedi S., Revy P., Emile JF., Newport M., Levin M., Blanche S., Seboun E., Fischer A., Casanova JL. (1996). Interferon-gamma-receptor deficiency in an infant with fatal bacille Calmette-Guerin infection. *N Engl J Med*, **335** : 1956-61.
- Krutzik SR., Ochoa MT., Sieling PA., Uematsu S., Ng YW., Legaspi A., Liu PT., Cole ST., Godowski PJ., Maeda Y., Sarno EN., Norgard MV., Brennan PJ., Akira S., Rea TH., Modlin RL. (2003). Activation and regulation of Toll-like receptors 2 and 1 in human leprosy. *Nat Med*, **9** : 525-32.
- Lam-Yuk-Tseung S., Gros P. (2003). Genetic control of susceptibility to bacterial infections in mouse models. *Cell Microbiol*, **5** : 299-313.
- Lander E., Kruglyak L. (1995). Genetic dissection of complex traits : guidelines for interpreting and reporting linkage results. *Nat Genet*, **11** : 241-7.
- Lander ES., Botstein D. (1987). Homozygosity mapping : a way to map human recessive traits with the DNA of inbred children. *Science*, **236** : 1567-70.
- Michalatos-Beloin S., Tishkoff S., Bentley K., Kidd K., Ruano G. (1996). Molecular haplotyping of genetic markers 10 kb apart by allele-specific long-range PCR. *Nucleic Acids Res*, **24** : 4841-4843.
- Mira MT., Alcais A., Van Thuc N., Moraes MO., Di Flumeri C., Hong Thai V., Chi Phuong M., Thu Huong N., Ngoc Ba N., Xuan Khoa P., Sarno EN., Alter A., Montpetit A., Moraes ME., Moraes JR., Dore C., Gallant CJ., Lepage P., Verner A., Van De Vosse E., Hudson TJ., Abel L., Schurr E. (2004). Susceptibility to leprosy is associated with PARK2 and PACRG. *Nature*, **427** : 636-640.
- Mira MT., Alcais A., Van Thuc N., Thai VH., Huong NT., Ba NN., Verner A., Hudson TJ., Abel L., Schurr E. (2003). Chromosome 6q25 is linked to susceptibility to leprosy in a Vietnamese population. *Nat Genet*, **33** : 412-5.
- Morton NE. (1955). Sequential tests for the detection of linkage. *Am J Hum Genet*, **7** : 277-318.
- Newport MJ., Huxley CM., Huston S., Hawrylowicz CM., Oostra BA., Williamson R., Levin M. (1996). A mutation in the interferon-gamma-receptor gene and susceptibility to mycobacterial infection. *N Engl J Med*, **335** : 1941-9.

- Nyholt DR. (2004). A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet*, **74** : 765-9.
- Ott J. (1999). *Analysis of human genetic linkage*. Johns Hopkins University Press, Baltimore.
- Remus N., El Baghdadi J., Fieschi C., Feinberg J., Quintin T., Chentoufi M., Schurr E., Benslimane A., Casanova JL., Abel L. (2004). Association of *IL12RB1* polymorphisms with adult pulmonary tuberculosis in Morocco. *J Infect Dis*, **190** (3) : 980-7.
- Sasieni PD. (1997). From genotypes to genes : doubling the sample size. *Biometrics*, **53** : 1253-61.
- Schaid DJ., Rowland C. (1998). Use of parents, sibs, and unrelated controls for detection of associations between genetic markers and disease. *Am J Hum Genet*, **63** : 1492-506.
- Sorensen TI., Nielsen GG., Andersen PK., Teasdale TW. (1988). Genetic and environmental influences on premature death in adult adoptees. *N Engl J Med*, **318** : 727-32.
- Spielman RS., McGinnis RE., Ewens WJ. (1993). Transmission test for linkage disequilibrium : the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet*, **52** : 06-16.
- Stead WW., Senner JW., Reddick WT., Lofgren JP. (1990). Racial differences in susceptibility to infection by *Mycobacterium tuberculosis*. *N Engl J Med*, **322** : 422-7.
- Tregouet DA., Escolano S., Tiret L., Mallet A., Golmard JL. (2004). A new algorithm for haplotype-based association analysis : the Stochastic-EM algorithm. *Ann Hum Genet*, **68** : 165-77.
- Vidal SM., Malo D., Vogan K., Skamene E., Gros P. (1993). Natural resistance to infection with intracellular parasites : isolation of a candidate for Bcg. *Cell*, **73** : 469-85.
- Wang DG., Fan JB., Siao CJ., Berno A., Young P., Sapolsky R., Ghandour G. *et al.* (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science*, **280** : 1077-82.
- WHO (2001). *Global tuberculosis control*. Geneva, Switzerland. WHO.

SOUS-CHAPITRE 5.2

Du risque cardiovasculaire aux démences : les voies de l'épidémiologie moléculaire

PHILIPPE AMOUYEL

L'épidémiologie est une discipline qui, dans la diversité qui la caractérise, fournit des outils, des modèles et des concepts permettant de répondre à des questions fondamentales ou de conforter, chez l'homme, des hypothèses issues de la recherche biomédicale fondamentale ou finalisée.

Les travaux de recherche sur les facteurs de risque cardiovasculaire fournissent un exemple des approches qu'il est possible de développer et des apports de l'épidémiologie à la connaissance scientifique et médicale. Appliquées à la biomédecine, ces approches permettent de dépasser les concepts nosologiques qui cloisonnent aujourd'hui la médecine et la science. Ainsi, par exemple, la convergence de ces approches vers une épidémiologie de plus en plus moléculaire a permis de relier deux grands domaines : le risque cardiovasculaire et le risque neurodégénératif, ouvrant le champ à de nouvelles hypothèses préventives et thérapeutiques.

Dans les évolutions récentes d'une recherche en biologie qui commence à développer des analyses moléculaires à haut débit génératrices d'énormes volumes

d'information, l'épidémiologie, par sa maîtrise de la biostatistique, de la stochastique et de l'informatique, et par sa réflexion holistique constitue une discipline de choix pour progresser dans la compréhension de la complexité du vivant.

Ces approches d'épidémiologie moléculaire sont généralisables, mais les engagements demandés doivent faire l'objet de choix stratégiques et politiques visant à investir massivement dans ces domaines.

1 | De l'irruption de la biomédecine en épidémiologie cardiovasculaire

Les maladies cardiovasculaires, au-delà de leur impact majeur sur la santé des populations humaines, constituent un des modèles les plus travaillés de maladie multifactorielle. Dès les premiers travaux de la vaste étude épidémiologique prospective de Framingham (Truett *et al.*, 1967), la multiplicité des déterminants de ces maladies est apparue et leur intégration au sein de modèles multivariés d'estimation du risque cardiovasculaire s'est rapidement imposée, ouvrant la voie aux approches de prévention et de prise en charge de ces maladies. À ce jour, plus de 250 facteurs de risque ont été décrits et parmi eux de nombreux marqueurs biologiques correspondant pour certains à des phénotypes biologiques intermédiaires et pour d'autres à des facteurs de susceptibilité génétique.

L'épidémiologie des maladies cardiovasculaires est née, il y a plus de 50 ans, avec les réflexions sur la mesure du risque de maladie coronaire chez l'adulte (*Measuring the risk of coronary heart disease in adult population groups*, 1957) et la constitution de cohortes épidémiologiques. Au cours de ce demi-siècle, trois étapes majeures ont permis de développer les connaissances et de faire progresser la prévention et la prise en charge de l'épidémiologie cardiovasculaire : l'observation épidémiologique, les progrès physiopathologiques et l'expérimentation préventive.

L'observation épidémiologique, telle qu'elle a été développée dans les études prospectives (Amouyel et Deverly, 2000), a permis de mettre en évidence un certain nombre de déterminants, d'en quantifier leurs effets conjugués et d'en apprécier leurs interrelations. Ces effets étaient estimés à partir du calcul de la probabilité de développer une maladie cardio-vasculaire dans un délai donné, c'est-à-dire plus simplement du risque cardiovasculaire.

Parallèlement à ces travaux épidémiologiques, de nombreuses recherches physiopathologiques accumulaient des données montrant la multiplicité, la complexité et l'hétérogénéité des mécanismes biologiques à l'origine des maladies

cardiovasculaires. Les épidémiologistes se sont appropriés ces travaux, en ont extrait de nouvelles variables, de nature « biologique », et les ont intégrées dans leurs modèles.

Enfin, l'arrivée de nouveaux moyens de prise en charge de ces multiples déterminants a été l'occasion de réaliser des expérimentations préventives (The Lipid Research Clinics Coronary Primary Prevention Trial results, 1984a, 1984b) dans le champ des maladies cardiovasculaires. L'objectif de ces essais, dont la méthodologie était guidée par l'épidémiologie et la statistique, consistait essentiellement à valider la réduction du niveau de risque en prenant en charge un déterminant, selon les protocoles standardisés et rigoureux des essais cliniques.

C'est ainsi que les concepts de risque cardiovasculaire et d'approche préventive visant à réduire ce risque ont été validés au début des années 1990, grâce aux essais de prévention secondaire (The Scandinavian Simvastatin Survival Study Group, 1994) et de prévention primaire (Shepherd *et al.*, 1995) du risque cardiovasculaire, faisant entrer la prévention cardiovasculaire dans le champ de la médecine fondée sur les preuves. Dans ces essais de prévention, c'est un facteur de risque biologique, le cholestérol LDL, qui était la cible des traitements et dont l'importance avait pu être mise en évidence par la conjonction de ces trois approches.

2 | Entre épidémiologie et biologie, un rapprochement naturel

La plupart du temps, la mise en évidence de nouveaux déterminants est issue de travaux relevant de la biologie fondamentale ou de la physiologie animale. Ces résultats proviennent de travaux développés *in vitro* ou *in vivo* chez l'animal, dont la pertinence et l'importance chez l'homme sont souvent difficiles à établir. S'appuyant sur ce contexte physiopathologique, l'épidémiologie peut en partie répondre à cette question de la pertinence chez l'homme, ou, à tout le moins, fournir des évidences quant à l'importance de ces facteurs dans tel ou tel domaine.

Un autre élément majeur de l'étroit rapprochement qui existe entre l'épidémiologie et la biologie est l'explosion des techniques de biologie moléculaire, et en particulier le développement de l'amplification de l'ADN par polymérase thermorésistante. Cette technique, susceptible de détecter d'infimes variations du génome humain, a permis aux épidémiologistes d'apporter un substrat biologique et scientifique à la mesure de la variabilité du génome. En effet, si l'on considère de façon schématique les sources de variabilité d'un événement, trois sont à prendre en compte : la variabilité environnementale, la variabilité individuelle, liée à des fondements spécifiques de nature génétique, et l'interaction

de ces deux sources de variabilité. Le développement de l'analyse des polymorphismes de l'ADN et des techniques d'analyse statistiques en génétique au cours des quinze dernières années a ouvert aux épidémiologistes de nouvelles perspectives d'exploration et de compréhension des pathologies.

Les maladies cardiovasculaires, par leur fréquence et les connaissances physiopathologiques accumulées depuis plusieurs dizaines d'années, sont un des sujets les plus étudiés par les épidémiologistes et les plus susceptibles d'interagir avec la biomédecine et de bénéficier de l'intrusion de la biologie en épidémiologie, notamment de la biologie moléculaire. Parmi les travaux récents qui ont pu avancer grâce à ces apports et ont eu un retentissement scientifique important, ceux relatifs à l'implication de l'enzyme de conversion de l'angiotensine et de son gène dans la maladie coronaire constituent un exemple intéressant.

Au début des années 1990, l'analyse des protéines du système rénine angiotensine et le clonage des gènes correspondants avaient permis de mettre en évidence une enzyme clé de cette voie métabolique impliquée à différents niveaux dans la régulation du système cardiovasculaire, l'enzyme de conversion de l'angiotensine I en angiotensine II (Corvol *et al.*, 1995). Cette protéine possède, *in vitro* et *in vivo*, des activités biologiques nombreuses, parmi lesquelles une influence sur la tonicité des vaisseaux. L'analyse du gène de l'enzyme de conversion avait mis en évidence un polymorphisme sous la forme d'une délétion dont la présence ou l'absence était associée à une variation significative de la concentration et de l'activité de l'enzyme elle-même (Rigat *et al.*, 1990). Cependant, le lien entre ce polymorphisme et un trait physiopathologique chez l'homme se révélait difficile. En utilisant une étude épidémiologique cas-témoins spécialement conçue pour prendre en compte des variables biologiques et génétiques, une association entre cette délétion et la survenue d'un infarctus du myocarde a pu être mise en évidence (Cambien *et al.*, 1992). Ce facteur de susceptibilité génétique a fait, par la suite, l'objet de nombreuses études, confirmant ou infirmant cette observation (Agerholm-Larsen *et al.*, 2000), rapportant de nouvelles associations (Amant *et al.*, 1997) et faisant évoluer les liens entre biologie et épidémiologie.

3 | L'introduction des variables biologiques dans les études épidémiologiques : une nécessaire adaptation des outils biologiques

Les travaux de biologie permettent la plupart du temps aux épidémiologistes d'identifier un marqueur biologique intermédiaire directement ou indirectement

associé, pouvant être analysé ensuite dans des populations humaines. La multiplication de ces marqueurs, la nécessité d'avoir accès à des prélèvements biologiques, le recours à des méthodes relevant de l'association statistique sous toutes ses formes ont amené les épidémiologistes à s'engager eux-mêmes dans ce type d'approches. Ils ont développé un savoir-faire dans la conception et la constitution de larges échantillons d'individus utiles à la mesure de la distribution des phénomènes morbides dans les populations humaines. L'adjonction de bases d'information biologique à ces études s'est donc imposée dès lors que la recherche biologique offrait des marqueurs susceptibles d'être mesurés de manière fiable et reproductible dans les tissus biologiques ainsi conservés.

Cependant, la réalisation d'analyses biologiques de toute nature sur de larges échantillons rebute souvent les biologistes tant par la répétitivité de la tâche que par le coût et les contraintes de contrôles de qualité inhérentes à ces travaux sur de grands effectifs. Aussi, les épidémiologistes ont-ils dû souvent développer leurs propres outils d'analyse biologique, jusqu'à monter des laboratoires de biologie spécialisés dans ce type de travaux.

Parmi les questions majeures liées à la biologie en épidémiologie, se pose celle de la conservation des échantillons. Compte tenu de la durée des études épidémiologiques, proportionnelle aux durées d'évolution des pathologies humaines, il est nécessaire de prévoir une conservation des échantillons biologiques garantissant leur intégrité plusieurs dizaines d'années durant. La question de l'organisation du stockage des échantillons et de leur pérennité est donc cruciale. Là encore une adaptation des techniques et des procédures est indispensable afin d'assurer cette conservation, l'altération inhomogène des échantillons biologiques au cours du temps pouvant générer de nombreux biais. C'est cette préoccupation, entre autres, qui a amené l'Inserm et le ministère de la Recherche (direction de la Technologie) à lancer l'appel d'offre « Cohorte et collection » afin de recenser et professionnaliser la constitution et l'entretien des collections d'échantillons biologiques. En effet, étant donné le temps et l'investissement humain et financier nécessaires à la constitution et à l'exploitation des études épidémiologiques, il est indispensable de garantir la qualité de ces collections qui constituent alors des outils de recherche permettant, pour certains d'entre eux, de vérifier nombre d'hypothèses biologiques récentes sans avoir recours à la mise en place d'une nouvelle étude longue et coûteuse.

4 | Des hypothèses sur les mécanismes d'apparition des maladies

Depuis près de 15 ans, l'intrusion des variables biologiques dans les études et les modèles d'analyse épidémiologique a permis d'aborder certaines questions

de la biologie sous un angle différent, plus ouvert, et d'apporter des fondements biologiques à des hypothèses issues de corrélations statistiques.

Un premier exemple d'apport relève du développement d'hypothèses sur les mécanismes d'apparition des maladies chroniques. Le concept de susceptibilité individuelle est défini essentiellement en réponse à un facteur d'environnement. Étant donné la rapidité d'évolution de l'environnement dans nos sociétés occidentales, il est possible de concevoir l'apparition de certaines maladies chroniques comme liée aux effets délétères de gènes initialement sélectionnés pour des effets bénéfiques dans des conditions environnementales très différentes. Cette théorie, développée initialement entre autres par Neel (1982), suggère en effet que certains gènes, qu'il qualifie de « gènes d'épargne », favorables à la survie dans des conditions difficiles (malnutrition, grandes migrations) deviendraient défavorables dans les conditions environnementales actuelles (suralimentation, sédentarité). Des illustrations de ces hypothèses existent pour des maladies comme le diabète ou l'hypertension artérielle (Zimmet et Thomas, 2003). En effet, nos ancêtres avaient été sélectionnés pour garantir la meilleure survie possible dans un quotidien fondé sur la chasse et la cueillette. Au cours des dernières décennies, l'environnement de l'*Homo sapiens* a changé de manière majeure, totalement inadapté au profil chasseur-cueilleur. L'adoption d'un style de vie sédentaire occidental et l'accès à des aliments hautement caloriques plaqués sur un profil génétique sélectionné pour favoriser les génotypes d'épargne a conduit à des modifications majeures du métabolisme augmentant la prévalence du diabète et de l'obésité et élevant ainsi le niveau de risque cardiovasculaire global des populations. Les études développées dans de larges échantillons de populations issues de régions variées du globe et présentant une grande variabilité quant aux conditions environnementales, permettent aux épidémiologistes de détecter des interactions majeures entre facteurs de susceptibilité génétique et d'environnement, impliquées dans la genèse des maladies, et d'en déduire des mécanismes potentiels de leur prévention.

5 | Une approche moléculaire de la classification des maladies

Un second exemple d'apport de cette intrusion massive de la biologie dans les approches épidémiologiques est de proposer des classifications différentes des maladies. Ces classifications sont fondées sur des mécanismes physiopathologiques moléculaires dont les liens ont pu être établis par le rapprochement de pathologies touchant des systèmes différents au travers de déterminants moléculaires communs. La caractérisation de la réponse à un déterminant selon la susceptibilité d'un individu tend à modifier ainsi la nosologie de pathologies hétérogènes, la clinique et l'évolution constituant jusqu'alors les principaux

éléments de classification. Un exemple est fourni par les liens qui ont été mis en évidence entre maladies cardiovasculaires et démences, au travers de déterminants génétiques communs.

Depuis plusieurs années, les épidémiologistes avaient pu noter des associations entre certains facteurs de risque cardiovasculaire, comme l'hypertension artérielle, et la survenue d'une démence (Skoog *et al.*, 1996). L'existence possible de déterminants communs entre maladies vasculaires et maladie d'Alzheimer est à rapprocher d'autres observations épidémiologiques, cliniques et anatomopathologiques suggérant que la maladie d'Alzheimer et les accidents vasculaires cérébraux sont associés plus fréquemment que ne le voudrait le hasard (Hofman *et al.*, 1997). Ainsi les cerveaux de patients non déments décédés de maladies coronaires, contiennent-ils plus de plaques séniles que ceux des sujets décédés d'autres causes (Sparks *et al.*, 1990). Ces observations sont également renforcées par les travaux de ces dix dernières années sur les facteurs de susceptibilité génétique de la maladie d'Alzheimer (Helbecque et Amouyel, 2004). Ainsi, un facteur de susceptibilité génétique comme le polymorphisme de l'apolipoprotéine E constitue-t-il un déterminant du risque cardio-vasculaire et neurodégénératif. Cette protéine, essentiellement connue pour son rôle de transport des lipides dans la circulation sanguine et pour ses interactions avec des récepteurs cellulaires, possède un polymorphisme à la fois structurel et génétique (Dallongeville *et al.*, 1992). L'apolipoprotéine E existe sous trois isoformes protéiques fréquentes : l'APOE3, E4 et E2. Chacune de ces isoformes est codée par une forme différente du gène de l'APOE appelés respectivement $\epsilon 3$, $\epsilon 4$ et $\epsilon 2$. L'isoforme APOE4 codée par l'allèle $\epsilon 4$ est associée à une élévation du LDL cholestérol athérogène et à une augmentation du risque d'infarctus du myocarde par rapport à l'APOE3, alors que l'APOE2 est associée à une diminution du risque de maladie cardiovasculaire (Luc *et al.*, 1994). La présence d'au moins un allèle $\epsilon 4$ de l'APOE dans le génotype d'un patient s'est révélée être le facteur de risque présentant la plus grande force d'association avec la maladie d'Alzheimer (Farrer *et al.*, 1997). Les individus porteurs de l'allèle $\epsilon 4$ auraient un risque de développer la maladie d'Alzheimer deux à seize fois plus important que les non-porteurs selon les études. Cet effet semble être identique chez l'homme et chez la femme, mais semble moins marqué dans les classes d'âges inférieures à 60 ans et supérieures à 80 ans. Cette relation, quoique d'intensité variable, est observée dans la plus grande majorité des populations. Enfin, il existe une troisième situation pathologique dans laquelle l'allèle $\epsilon 4$ de l'APOE ne joue plus un rôle délétère, mais un rôle protecteur : c'est la dégénérescence maculaire liée à l'âge (Souied *et al.*, 1998).

On peut donc constater qu'un même déterminant biologique peut être à l'origine de trois processus pathologiques différents selon le type de tissu où il s'exprime et les conditions physiopathologiques qui l'entourent. C'est grâce à des études épidémiologiques associant divers événements à un même déterminant

biologique que ce type d'hypothèse a pu être émis, ouvrant la voie à une épidémiologie moléculaire.

6 | De l'appropriation de l'épidémiologie par les biologistes et inversement

Devant les apports réguliers et originaux de l'épidémiologie à la biologie, les biologistes ont commencé à trouver un certain intérêt à ces approches impliquant des échantillons de populations humaines et exploitant de vastes banques et bases de données biologiques pour y tester certaines hypothèses. Les potentialités ainsi révélées les ont conduits à se référer souvent à ces approches pour tenter de répondre à des questions spécifiques ou plus souvent pour renforcer l'intérêt de telle ou telle découverte de biologie fondamentale en clinique. Cet engouement pour l'utilisation de l'épidémiologie a amené la publication de nombreux travaux, certains laissant à désirer quant à la qualité des méthodologies utilisées et favorisant la multiplication des biais : problème de puissance statistique, échantillonnages biaisés, biais d'indication, biais de sélection. . .

Cette banalisation des bases de données biologiques associées aux études épidémiologiques a fait oublier, pour ceux qui ne les avaient pas constituées, que de tels travaux sont initialement construits pour répondre à une question scientifique précise. Aussi, bien que pouvant être utilisées à d'autres fins dans certaines circonstances bien définies, elles ne peuvent constituer en soit des collections pouvant répondre à n'importe quelle hypothèse. Enfin, lorsque ces bases de données biologiques sont exploitées, il est important de tenir compte de l'implication des chercheurs épidémiologistes et des investissements financiers et humains qui ont contribué à leur mise en place, souvent lourde, longue et difficile.

Continuant dans cette logique de développement de bases de données biologiques, un projet a vu le jour au Royaume-Uni, ayant pour objectif de mettre en place au niveau national une vaste étude épidémiologique prospective suivant 500 000 personnes de 45 à 69 ans sur 10 ans, la « UK biobank » (Barbour, 2003). Son objectif affiché est de pouvoir corrélérer des facteurs d'exposition génétique et environnementale avec les affections détectées à la fin de la période de surveillance, parmi lesquelles les décès par maladies coronaires, par accidents vasculaires cérébraux, et par bronchopneumopathies obstructives chroniques, les cas de cancers du poumon et de la prostate, les cas de diabète et les fractures de hanche. Cette étude a été diversement accueillie par les scientifiques selon leur spécialité, les épidémiologistes en percevant les limites, malgré la taille (Clayton et Mckeigue, 2001), et les biologistes s'enthousiasmant beaucoup plus, fascinés par l'effectif de 500 000. L'investissement initial de l'étude a été fixé à 45 millions livres mais les estimations de coût sur l'ensemble du programme

sont comprises entre 70 et 500 millions livres. Pour le moment, l'efficacité d'un tel projet, au-delà de sa visibilité politique, reste difficilement évaluable. Il est probable que cette étude permettra d'obtenir certains résultats, mais que les objectifs de mise en évidence des déterminants des différentes pathologies ciblées dépendront du nombre de cas effectivement recensés et de leur définition, comme dans toute étude épidémiologique.

Parallèlement, les épidémiologistes se sont mis à développer systématiquement des bases de données biologiques dans la plupart de leurs études. Ces travaux, entraînant un acte médical invasif chez la personne interrogée, imposent d'obtenir un certain nombre d'autorisations préalables auprès des comités *ad hoc*, alourdissant d'autant les études. Certains épidémiologistes ont développé leur propre laboratoire de préparation, de manipulation et de stockage des échantillons afin de mieux maîtriser les conditions de gestion et de conservation de ces vastes volumes d'échantillons biologiques ; en particulier dans le domaine de la biologie moléculaire. Les épidémiologistes se sont lancés dans la réalisation de leurs propres caractérisations de polymorphismes génétiques, grâce à l'automatisation de ces techniques. Ces conditions les ont amenés à émettre des hypothèses biologiques ou à vérifier, indépendamment des biologistes, l'intérêt clinique de résultats biologiques développés *in vitro* ou *in vivo* en laboratoire, en oubliant parfois que les résultats ainsi obtenus par associations statistiques nécessitaient de solides présomptions de causalité pour être retenus.

En fait, cette appropriation commune de champs d'expérience aussi divers que l'épidémiologie et la biologie témoigne de la nécessité de la cohabitation de ces deux spécialités dans le cadre d'une interdisciplinarité forte. Des épidémiologistes doivent être intégrés dans certains laboratoires de biologie afin d'assurer une meilleure adéquation entre méthodologie et hypothèses biologiques. Inversement, des biologistes doivent être intégrés dans certains laboratoires d'épidémiologie afin d'utiliser au mieux l'outil biologique et ses hypothèses dans le contexte des études épidémiologiques.

Le développement massif et la diffusion des techniques de biologie à haut débit dans les domaines de la génomique et de la post-génomique sont venus apporter des outils à la mesure des volumes de données manipulés habituellement par les épidémiologistes, renforçant encore la nécessité de cette interdisciplinarité. Néanmoins, cette révolution silencieuse commence à se heurter aujourd'hui à deux écueils majeurs : les coûts inhérents à ces nouvelles technologies et la masse des données à manipuler.

Enfin, cette orientation vers une forte interdisciplinarité est soutenue par les éditeurs de revues scientifiques à facteurs d'impact élevés (Editors, Freely associating, 1999), qui aujourd'hui conditionnent la publication de résultats d'associations à des critères à la fois épidémiologiques et biologiques solides :

échantillons de grande taille (au moins 1 000 individus), probabilité critique faible, plausabilité biologique forte, analyse de la fonctionnalité des gènes incriminés, multiples études indépendantes, forces d'association élevées.

Conclusion et perspectives

Au cours des vingt dernières années, l'épidémiologie a commencé à s'imposer au sein de la recherche biomédicale comme un outil incontournable. La complexité des marqueurs biologiques pouvant être analysés a évolué de pair avec les progrès technologiques de la biologie. Ainsi la finalisation du programme génome humain, la miniaturisation des technologies, l'exécution massivement parallèle des analyses selon une biologie à grande échelle ont naturellement convergé vers une épidémiologie capable de son côté de concevoir et de manipuler de vastes échantillons de populations précisément décrits.

Cette convergence suscite un engouement et des attentes des biologistes qui dépassent parfois ce que peuvent pratiquement apporter les épidémiologistes. Comme dans toute recherche, les études épidémiologiques ne peuvent répondre de manière correcte qu'à une question correctement posée. Les études ou plus généralement « les cohortes », comme les biologistes aiment à les nommer, ne peuvent être considérées comme des outils universels permettant de répondre à toutes les questions. Cette confusion a été entretenue par la possibilité de conserver des banques de tissus biologiques et de pouvoir les utiliser bien longtemps après que l'étude soit terminée pour vérifier des hypothèses qui n'auraient même pas pu être envisagées 20 ans avant, au démarrage de l'étude. Cette capacité à constituer des modèles apportant des réponses originales à des questions de biologie fondamentale, les confusions entre grands effectifs, puissance et force d'association ont amené certains pays à s'engager dans la voie utopique de la construction de « la cohorte absolue » permettant de répondre à tout.

Il est donc important que cette convergence entre épidémiologie et biomédecine puisse être favorisée avec raison et équilibre. Les épidémiologistes français ont une certaine avance dans ce domaine. Cependant, les choix stratégiques et politiques favorisant les investissements dans cette interdisciplinarité sont indispensables pour maintenir cette avance. Certains pays nord-américains et européens se sont déjà résolument engagés dans cette direction.

Bibliographie

Agerholm-Larsen B. *et al.* (2000). ACE gene polymorphism in cardiovascular disease : meta-analyses of small and large studies in whites. *Arterioscler Thromb Vasc Biol*, **20** : 484-492.

- Amant C. *et al.* (1997). D allele of the angiotensin I-converting enzyme is a major risk factor for restenosis after coronary stenting. *Circulation*, **96** : 56-60.
- Amouyel P., Deverly A. (2000). Risque cardiovasculaire global : définition, stratégies d'évaluation et de prise en charge. *Thérapies*, **55** : 533-539.
- Barbour V. (2003). UK Biobank : a project in search of a protocol ? *Lancet*, **361** : 1734-1738.
- Cambien F. *et al.* (1992). Deletion polymorphism in the gene for angiotensin-converting enzyme is a potent risk factor for myocardial infarction. *Nature*, **359** : 641-644.
- Clayton D., McKeigue PM. (2001). Epidemiological methods for studying genes and environmental factors in complex diseases. *Lancet*, **358** : 1356-1360.
- Corvol P., Williams TA., Soubrier F. (1995). Peptidyl dipeptidase A : angiotensin I-converting enzyme. *Methods Enzymol*, **248** : 283-305.
- Dallongeville J. *et al.* (1992). Modulation of plasma triglyceride levels by apoE phenotype : a meta-analysis. *J Lipid Res*, **33** : 447-454.
- Editors. Freely associating (1999). *Nature Genetics*, **22** : 1-2.
- Farrer LA. *et al.* (1997). Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. *JAMA*, **278** : 1349-1356.
- Helbecque N., Amouyel P. (2004). Commonalities between genetics of cardiovascular disease and neurodegenerative disorders. *Curr Opin Lipidol*, **15** : 121-127.
- Hofman A. *et al.* (1997). Atherosclerosis, apolipoprotein E, and prevalence of dementia and Alzheimer's disease in the Rotterdam Study. *Lancet*, **349** : 151-4.
- Luc G. *et al.* (1994). Impact of apolipoprotein E polymorphism on lipoproteins and risk of myocardial infarction. The ECTIM Study. *Arterioscler Thromb*, **14** : 1412-1419.
- Measuring the risk of coronary heart disease in adult population groups (1957). A Symposium. *Am J Pub. Health*, **47** : 1-63.
- Neel JV. (1982). In *The genetics of diabetes mellitus*. Kobberling J., Tattersall R. (Eds), Academic Press, London, pp. 283-296.
- Rigat B. *et al.* (1990). An insertion/deletion polymorphism in the angiotensin I-converting enzyme gene accounting for half the variance of serum enzyme levels. *J Clin Invest*, **86** : 1343-1346.

- Shepherd J. *et al.* (1995). Prevention of coronary heart disease with pravastatin in men with hypercholesterolemia. *N Engl J Med*, **333** : 1301-1307.
- Skoog I. *et al.* (1996). 15-year longitudinal study of blood pressure and dementia. *Lancet*, **347** : 1141-1145.
- Souied EH. *et al.* (1998). The epsilon 4 allele of the apolipoprotein E gene as a potential protective factor for exudative age-related macular degeneration. *Am J Ophthalmol*, **125** : 353-359.
- Sparks DL. *et al.* (1990). Cortical senile plaques in coronary artery disease, aging and Alzheimer's disease. *Neurobiol Aging*, **11** : 601-607.
- Truett J. *et al.* (1967). A multivariate analysis of the risk of coronary heart disease in Framingham. *J. chron. Dis*, **20** : 511-524.
- The Lipid Research Clinics Coronary Primary Prevention Trial results (1984a). I. Reduction in incidence of coronary heart disease. *JAMA*, **251** : 351-364.
- The Lipid Research Clinics Coronary Primary Prevention Trial results (1984b). II. The relationship of reduction in incidence of coronary heart disease to cholesterol lowering. *JAMA*, **251** : 365-374.
- The Scandinavian Simvastatin Survival Study Group (1994). Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease : the Scandinavian Simvastatin Survival Study (4S). *Lancet*, **344** : 1383-1389.
- Zimmet P., Thomas CR. (2003). Genotype, obesity and cardiovascular disease—has technical and social advancement outstripped evolution? *J Intern Med*, **254** : 114-125.

SOUS-CHAPITRE 5.3

Épidémiologie et vieillissement : des sciences sociales à la biologie moléculaire

ANNICK ALPÉROVITCH ET JEAN-FRANÇOIS DARTIGUES

L'épidémiologie du vieillissement s'est développée plus tardivement que la recherche épidémiologique sur les maladies cardiovasculaires ou les cancers. Son essor s'est amorcé au début des années 1980 et a coïncidé avec la prise de conscience que le vieillissement des populations représentait un des défis majeurs auxquels nos sociétés devraient faire face dès le début du XXI^e siècle. L'épidémiologie joue déjà et jouera dans l'avenir un rôle majeur pour décrire, comprendre les déterminants et trouver des solutions aux problèmes associés au vieillissement. Aucune des difficultés méthodologiques de l'épidémiologie du vieillissement ne lui est spécifique mais toutes sont majorées par le fait que dans le vieillissement, à la différence des autres domaines de la santé, le phénomène étudié affecte l'ensemble de l'individu et l'ensemble des individus.

1 | Le vieillissement : un processus qui commence bien avant d'être âgé

Les pays nord-américains et le Royaume-Uni sont pratiquement les seuls à disposer de données épidémiologiques permettant d'estimer les effets à long terme d'expositions ou événements survenus à différents âges de la vie. Ces données sont à l'origine de résultats importants, notamment sur le lien entre les caractéristiques staturopondérales du nouveau-né et du jeune enfant et le risque de maladie vasculaire ou métabolique à l'âge adulte (Berenson, 2002 ; Lawlor *et al.*, 2004). Des facteurs mesurés à l'âge de 40-50 ans peuvent être prédictifs d'événements (démence, maladies vasculaires) survenant 20 à 30 ans plus tard. Des données suggèrent même que certains des facteurs de risque des pathologies cognitives et vasculaires associées au vieillissement sont repérables dès l'enfance (Snowdon *et al.*, 1996). Pour comprendre les phénomènes liés au vieillissement, pour concevoir des actions préventives efficaces, il faut donc des études prospectives de longue durée. L'âge minimal d'inclusion dans ces études varie en fonction des objectifs. Il peut être de 70 ou 75 ans dans certains travaux sur la démence dont plus de 80 % des cas surviennent après 70 ans (Letenneur *et al.*, 1999). Mais à ces âges, il peut être trop tard pour agir sur des facteurs de risque dans un objectif de prévention. Étudier le vieillissement, ce n'est pas étudier seulement des populations âgées. Par exemple, la Rotterdam Study, une des plus productives des études sur le vieillissement, porte sur des personnes de 55 ans et plus. Certaines pathologies, la démence notamment, étant très rares avant 65 ans, il est nécessaire de suivre de telles cohortes pendant de nombreuses années pour que l'incidence des événements (c'est-à-dire maladies) étudiés soit suffisante pour en identifier les facteurs de risque. Le nombre d'événements dépendant à la fois de la taille de la cohorte et de la durée du suivi, les épidémiologistes mesurent la puissance d'une étude et estiment le risque de maladie en termes de « *personnes-années* ». Par exemple, la taille optimale pour une étude prospective sur les facteurs de risque des accidents vasculaires cérébraux est 100 000 *personnes-années*. Il faut donc suivre 10 000 personnes pendant 10 ans si l'on veut se placer suffisamment en amont de la survenue de la maladie. Ces chiffres donnent une idée des efforts humains et matériels que représente une étude de cohorte. Il faut aussi souligner que ces études ne peuvent se faire sans la participation active de la population concernée.

2 | Épidémiologie du vieillissement ou des maladies liées à l'âge ?

L'avancée en âge s'accompagne d'un déclin de toutes les fonctions physiologiques. Du système immunitaire à la cognition, la diminution des performances

n'a de conséquences cliniquement perceptibles qu'à partir d'un certain seuil. L'âge auquel ce seuil est atteint varie d'une personne à l'autre pour une fonction donnée et d'une fonction à l'autre pour une personne donnée. Les méthodes d'exploration préclinique et les marqueurs biologiques permettant de mesurer le déclin fonctionnel associé au vieillissement bien avant qu'apparaissent ses conséquences cliniques sont largement utilisés en épidémiologie. De grandes études en population générale ont fourni une description précise des altérations des fonctions physiologiques en fonction de l'âge et de leur variabilité d'un individu à l'autre. Quelle que soit la fonction considérée, la variabilité des performances entre individus est très importante dans la population âgée. L'étude des facteurs de cette variabilité conduit à des questions de fond sur les frontières entre vieillissement normal, vieillissement accéléré et vieillissement pathologique. Pour répondre à ces questions — dont les enjeux scientifiques et de santé publique sont majeurs — il faut confronter les données fournies par des études longitudinales populationnelles à des modèles mathématiques de vieillissement. Ainsi, les résultats de la modélisation épidémiologique des relations entre vieillissement cognitif normal et maladie d'Alzheimer ont fait l'objet d'un riche débat scientifique contradictoire (Brayne *et al.*, 1998 ; Ritchie *et al.*, 1995). Observation et modélisation permettent aujourd'hui de répondre négativement à la question posée en 1994 par Drachman : « *If we lived long enough, will we all be demented?* ». Mais tout n'est pas clair pour autant. Depuis que des centres de la mémoire ont été mis en place dans différents pays, dont récemment la France, neurologues et neuropsychologues s'intéressent à un tableau clinique d'altération des fonctions cognitives, désigné internationalement sous le nom de Mild Cognitive Impairment (MCI) (Petersen *et al.*, 2001). Il n'y a consensus ni sur les critères définissant ce syndrome, ni sur les relations entre MCI et maladie d'Alzheimer (MA), ni sur la nature même de l'entité (maladie d'Alzheimer pré-démontielle ou sujet à risque d'évolution vers la MA). Est-il vrai, comme l'affirment certains, que la grande majorité des MCI évolue vers une maladie d'Alzheimer ? Si ce n'est pas le cas, existe-t-il des facteurs prédictifs de l'évolution bénigne ou maligne (vers la démence) d'un MCI ? On ne peut aujourd'hui proposer aux personnes atteintes de la maladie d'Alzheimer que des traitements symptomatiques dont l'efficacité est modérée et transitoire. La recherche de molécules susceptibles de prévenir les lésions cérébrales à l'origine de cette maladie ou d'en ralentir le développement est très active, en raison des énormes enjeux médico-économiques. La cible prioritaire de ces traitements pourrait être les personnes présentant un MCI, ou au moins une partie d'entre elles. Mieux connaître l'évolution des MCI afin d'être à même d'identifier les personnes les plus à risque de démence est l'un des objectifs majeurs actuels de la recherche épidémiologique sur le vieillissement cérébral.

Les problématiques épidémiologiques concernant le vieillissement d'autres fonctions telles que la vision, l'audition ou la motricité sont très proches de celles qui viennent d'être évoquées pour les fonctions cognitives. En revanche,

l'approche des maladies dont l'incidence augmente fortement avec l'âge et la majorité des cas survient après 65 ans (maladie de Parkinson, des accidents vasculaires cérébraux, etc.) ne se distingue pas, globalement, de celle de pathologies touchant des populations plus jeunes. Néanmoins, plus l'âge avance, plus la mesure des facteurs de risque devient difficile. C'est plusieurs décennies avant le début de la maladie qu'il faut rechercher les expositions professionnelles, comportements et autres facteurs étiologiques. L'erreur de mesure sur les facteurs de risque est d'autant plus grande que les personnes concernées présentent un déclin cognitif. Cette erreur de mesure n'est jamais sans conséquence. Au mieux, elle diminue la puissance de l'étude, qui pourrait alors passer à côté de réels facteurs de risque. Les épidémiologistes sont très attentifs aux éventuels biais, tant dans la conception du protocole d'une étude que dans l'analyse critique de ses résultats. Au passage, on peut regretter que, quel que soit le domaine, la médiatisation de certains résultats vers le grand public ne reflète que rarement la prudence des épidémiologistes.

3 | L'épidémiologie génétique de la longévité

Les progrès de la biologie moléculaire ont permis d'introduire l'étude du gène dans les recherches épidémiologiques sur le vieillissement. À nouveau, il est nécessaire de distinguer dans les études des polymorphismes de l'ADN celles sur la longévité et celles sur des maladies associées au vieillissement. La découverte d'une relation entre le polymorphisme de l'apolipoprotéine E et la maladie d'Alzheimer est l'un des faits scientifiques majeurs des dix dernières années. Cette relation a été initialement décrite dans des formes familiales de la maladie (Corder *et al.*, 1993). Les études épidémiologiques ont établi que l'isoforme APOE4 codée par l'allèle $\epsilon 4$ du gène était aussi un facteur de risque de la maladie d'Alzheimer sporadique, qui représente 95 % des cas de cette démence. Elles ont aussi permis d'estimer le risque de maladie en fonction du nombre d'allèles $\epsilon 4$ (0, 1 ou 2). Dans le prolongement de cette découverte, certains ont proposé d'utiliser le polymorphisme de l'apolipoprotéine E dans le diagnostic de la maladie d'Alzheimer. Cette proposition a soulevé de vives critiques de la part des épidémiologistes dont les données démontraient la très faible valeur prédictive d'un tel test. En revanche, si pour évaluer des stratégies de prévention, il devenait légitime d'identifier des groupes à haut risque de maladie d'Alzheimer, le polymorphisme de l'apolipoprotéine E serait l'un des éléments du profil de risque. Sur la base des études conduites dans les années 1990, les épidémiologistes pourraient bientôt être en mesure d'établir un profil de risque de maladie d'Alzheimer en fonction de caractéristiques sociodémographiques (âge, sexe, niveau d'études), médicales, biologiques et génétique et d'estimer le risque de survenue de la maladie pour un profil donné. Pourrait-on un jour utiliser ce profil pour définir la stratégie de prise en charge des patients,

comme les cliniciens le font dans les maladies cardiovasculaires ? Cela dépendra en premier lieu des possibilités de prévention de la maladie d'Alzheimer, prévention qui est la principale finalité de grandes études épidémiologiques en cours actuellement. Mais l'utilisation d'une caractéristique génétique dans un profil de risque pose une question éthique plus générale, à laquelle la société devrait auparavant répondre.

Par ailleurs, des biologistes ont entrepris des travaux visant à identifier des gènes de la longévité. Des études ont, dans cet objectif, comparé des centenaires à des sujets plus jeunes. Leurs résultats ont été assez décevants, en raison de biais, certains complexes et d'autres triviaux, qui n'existent pas dans l'étude des gènes de la longévité de certaines espèces animales, mais ne peuvent être ignorés lorsque l'on étudie des populations humaines. Nul besoin d'être un fin épidémiologiste pour concevoir que, pour être centenaire, il faut avoir échappé à de nombreuses maladies mortelles (cancer, maladies cardiovasculaires, maladies respiratoires, pour ne citer que quelques causes de décès importantes). On peut donc s'attendre à ce que certains polymorphismes liés à la longévité correspondent à des gènes impliqués dans des maladies fréquentes et mortelles : si l'allèle d'un gène est facteur de risque de la maladie, et donc de mortalité, les autres allèles seront liés à la longévité. Ainsi par exemple, on observe que la fréquence de l'allèle $\epsilon 4$, facteur de risque des maladies vasculaires et de la maladie d'Alzheimer, diminue avec l'âge et, par voie de conséquence, celle des autres allèles de l'apolipoprotéine E augmente (Ewbank, 2002). Les interactions entre des facteurs génétiques et environnementaux liés à la mortalité constituent, par ailleurs, l'une des difficultés de l'étude des facteurs de risque des maladies associées au vieillissement. Les biais induits par ces interactions peuvent être estimés par des modèles mathématiques (Elbaz et Alperovitch, 2002).

L'étude de gènes susceptibles d'intervenir dans les mécanismes généraux du vieillissement, comme ceux codant les enzymes intervenant dans le métabolisme des radicaux libres, a été décevante. Le projet d'une vaste étude prospective sur l'épidémiologie génétique de la longévité et des maladies associées au vieillissement, défendu par certains chercheurs, n'a pas suscité un grand enthousiasme de la part des épidémiologistes. Un tel projet, une « UK biobank » (Barbour, 2003) pour la population âgée de plus de 65 ans, nécessiterait des investissements considérables pour des retombées scientifiques et en santé publique très incertaines.

4 | L'âge : le meilleur marqueur épidémiologique du vieillissement ?

La variabilité, à âge égal, de certains signes du vieillissement est observable par tous. Existe-t-il un marqueur biologique qui mesurerait le vieillissement

mieux que l'âge ? Quelles doivent être les qualités d'un tel marqueur ? Ces questions ont été et restent un sujet de discussion entre biologistes, épidémiologistes et statisticiens. Un très grand nombre de paramètres biologiques changent avec l'âge. Le taux de sulfate de déhydroépiandostérone (DHEA), par exemple, diminue avec l'âge et des données épidémiologiques suggèrent qu'il est corrélé à quelques troubles associés au vieillissement et à la mortalité (Mazat *et al.*, 2001). Mais cela ne suffit pas pour en faire un meilleur indicateur épidémiologique du vieillissement que l'âge lui-même. Si l'on considère une fonction physiologique donnée, est-il concevable de trouver un marqueur lié à l'âge, qui prédit mieux que l'âge le déclin de cette fonction, sans être — évidemment — un marqueur de la fonction elle-même ? Si l'on considère maintenant le vieillissement global, la question se complique car il est alors nécessaire de définir, autrement que par l'âge, ce qu'est « être vieux ». Daniel Schwartz en propose une définition. Le vieillissement d'un individu est mesuré par le temps qui le sépare de sa mort : on est d'autant plus vieux que ce temps est court. Quel marqueur pourrait, à l'échelle populationnelle, être un meilleur prédicteur de l'espérance de vie que l'âge ? La biologie et l'épidémiologie ne semblent pas à même de répondre à cette question dans un avenir proche.

5 | Le poids des facteurs sociaux et économiques

À l'heure où la biologie moléculaire devient l'un des outils de la recherche épidémiologique sur le vieillissement, des études françaises et étrangères récentes viennent rappeler que les facteurs sociaux et économiques ont un rôle primordial dans le vieillissement et ses conséquences. Les inégalités sociales dans le domaine de la santé sont très fortes dans la population âgée. En France, la catégorie socioprofessionnelle reste l'un des déterminants majeurs de la longévité. Le niveau d'études et le niveau économique expliquent une très large part de la variabilité de certains phénomènes associés au vieillissement. Notamment, comme l'ont montré de nombreuses études épidémiologiques et particulièrement l'étude française Paquid (Letenneur, 1999), les performances cognitives et le risque de démence sont très liés au niveau d'études ou à la catégorie socioprofessionnelle. L'interprétation de cette association n'est pas univoque et le débat entre épidémiologistes sur cette question est loin d'être clos.

Les caractéristiques socio-économiques sont associées à une multitude de facteurs : mode de vie et comportements, expositions professionnelles, attitude vis-à-vis des problèmes de santé. On conçoit donc qu'elles puissent constituer, dans les études populationnelles, un marqueur à la fois grossier et puissant de l'exposition à un ensemble de facteurs de risque susceptibles d'influencer l'apparition de maladies ou leur pronostic. Les facteurs sociaux influencent aussi la qualité

de la prise en charge des patients. La consommation de médicaments potentiellement inappropriés pour les personnes âgées de 65 ans et plus a été analysée dans une étude épidémiologique prospective française, l'étude des 3 Cités (The 3C Study Group). Sur la base de critères internationalement reconnus (Beers, 1997), il a été montré que 40 % des personnes consommaient au moins un médicament inapproprié. Cette proportion était plus élevée chez les femmes que les hommes et était inversement liée au niveau d'études. Chez les personnes âgées les plus socialement défavorisées, cette proportion était de 50 %. Sachant que cette consommation peut, selon le médicament, entraîner des hospitalisations pour effets secondaires sévères, des troubles cognitifs, des chutes, etc., on a un exemple de la complexité des relations entre les facteurs sociaux et la santé. Le travail des épidémiologistes est de décrire cette complexité, d'en identifier les facteurs et de modéliser leurs interrelations.

Les relations entre niveau d'études et risque de démence peuvent aussi s'interpréter en s'appuyant sur des concepts de la neuropsychologie et l'utilisation de techniques non invasives d'imagerie cérébrale dans des études populationnelles. Les résultats de plusieurs études épidémiologiques indiquent que, à lésions cérébrales égales, la sévérité des troubles cognitifs dépend du niveau d'éducation. Comme toutes les maladies neurodégénératives, les démences sont la résultante de deux processus : l'un détruisant le cerveau, l'autre compensant ces destructions grâce à la plasticité cérébrale. Un des éléments favorisant la plasticité cérébrale est l'existence de capacités de réserve constituée essentiellement de réseaux neuronaux peu ou non utilisés dans les tâches cognitives de la vie quotidienne et qui pourraient être mobilisés en cas de besoin, notamment de lésion cérébrale (Stern, 2002). Ces réserves cognitives pourraient dépendre du niveau d'études et d'autres facteurs sociaux.

Conclusion et perspectives

Le champ de l'épidémiologie du vieillissement est immense. Sur des questions de santé publique majeures, les besoins de recherche sont considérables. Quelques exemples : les stratégies de prévention des maladies vasculaires ont été définies, pour la plupart, à partir d'études épidémiologiques dans des populations d'âge moyen inférieur à 65 ans et majoritairement masculines. Elles n'ont pas été suffisamment évaluées dans des populations âgées. Si l'étude Systeur (Forette *et al.*, 2002) a démontré l'efficacité du traitement de l'hypertension artérielle après 70 (75 ?) ans, la question se pose encore pour les hypolipidémiants. Par ailleurs, certains traitements des maladies chroniques de l'adulte pourraient avoir des effets que l'on ne soupçonnait pas sur certaines pathologies liées à l'âge. Des études épidémiologiques suggèrent que les hypolipidémiants, pour revenir à cet exemple, pourraient avoir un effet protecteur vis-à-vis de la

démence. Une génération de femmes dont une large proportion aura utilisé des traitements hormonaux (contraception, traitement hormonal substitutif de la ménopause) pendant de très nombreuses années atteint aujourd'hui la soixantaine. L'impact — positif ou négatif — de cette exposition sur les problèmes de santé associés au vieillissement doit être évalué. Devant la part croissante des plus de 75 ans dans la population, il est important de se poser la question de l'existence de facteurs de risque de cancer spécifiques à cet âge ou, plus généralement de l'épidémiologie des cancers dans la population âgée. Dans le domaine du vieillissement aussi, les équipes françaises sont très bien placées dans la compétition internationale et sont à même d'avoir un rôle de tout premier plan dans des enjeux de recherche et de santé publique majeurs.

Bibliographie

- Barbour VUK. (2003). Biobank : a project in search of a protocol. *Lancet*, **361** : 1734-1738.
- Beers MH. (1997). Explicit criteria for determining potentially inappropriate medication use by the elderly. An update. *Arch Intern Med*, **157** : 1531-6.
- Berenson GS. (2002). Childhood risk factors predict adult risk associated with subclinical cardiovascular disease – The Bogalusa Heart Study. *Am J Cardiol*, **90** (10C) : 3L-7L.
- Brayne C., Calloway P. (1988). Normal ageing, impaired cognitive function, and senile dementia of the Alzheimer's type : a continuum? *Lancet*, **1** (8597) : 1265-1267.
- Corder EH., Saunders AM., Strittmatter WJ., Schmechel DE., Gaskell PC., Small GW., Roses AD., Haines JL., Pericak-Vance MA. (1993). Gene dose of apolipoprotein E type A allele and the risk of Alzheimer's disease in late onset families. *Science*, **261** : 921-923.
- Drachman DA. (1994). If we live long enough, will we all be demented? *Neurology*, **44** : 1563-1565.
- Elbaz A., Alperovitch A. (2002). Bias in association studies resulting from gene-environment interactions and competing risks. *Am J Epidemiol*, **155** : 265-73.
- Ewbank DC. (2002). Mortality differences by APOE genotype estimated from demographic synthesis. *Genetic Epidemiol*, **22** : 146-155.
- Forette F., Seux ML., Staessen JA. *et al.* (2002). The prevention of dementia with antihypertensive treatment. *Arch Intern Med*, **162** (18) : 2046-2052.

- Lawlor DA., Davey-Smith DG., Ebrahim S. (2004). Birth weight is inversely associated with coronary heart disease in post-menopausal women : Findings from the British women's heart and health study. *J Epidemiol Community Health*, **58**(2) : 120-125.
- Letenneur L., Gilleron V., Commenges D., Helmer C., Orgogozo J., Dartigues JF. (1999). Are sex and educational level independent predictors of dementia and Alzheimer's disease ? Incidence data from the Paquid project. *J Neurol Neurosurg Psychiatr*, **66** : 177-183.
- Mazat L., Lafont S., Berr C., Debuire B., Tessier JF., Dartigues JF., Baulieu EE. (2001). Prospective measurements of dehydroepiandrosterone sulphate in a cohort of elderly subjects : Relation ship to gender, subjective health, smoking habits, and 10-year mortality. *PNAS*, **98** : 8145-8150.
- Petersen RC., Stevens JC., Ganguli M., Tangalos EG., Cummings JL., DeKosky ST. (2001). Practice parameters : early detection of dementia : mild cognitive impairment (an evidence-based review). Report of the quality standards subcommittee of the American Academy of Neurology. *Neurology*, **56** : 1133-1142.
- Ritchie K., Kildea D. (1995). Is senile dementia "age-related" or "ageing-related" ? – Evidence from meta-analysis of dementia prevalence in the oldest old. *Lancet*, **346** : 931-934.
- Snowdon DA., Kemper SJ., Mortimer JA. et al. (1996). Ability in early life and cognitive function and Alzheimer's disease in late life. Findings from the Nun Study. *JAMA*, **275** : 528-531.
- Stern Y. (2002). What is cognitive reserve ? Theory and research application of the reserve concept. *J Int Neuropsychol Soc*, **8** : 448-60.
- The 3C study group (2003). Vascular factors and risk of dementia : design of the Three-City study and baseline characteristics of the study population. *Neuroepidemiol*, **22** : 316-325.

SOUS-CHAPITRE 5.4

Épidémiologie des cancers

CATHERINE HILL ET SIMONE BENHAMOU

Introduction

Le cancer est une maladie très fréquente, et la recherche de ses causes est un domaine d'interaction pour les biologistes et les épidémiologistes. En effet, faire la preuve qu'une exposition est cancérigène pour l'homme nécessite de rassembler des informations expérimentales et des observations sur l'homme. Cette preuve sert de base pour argumenter les décisions de limiter les expositions, elle est donc essentielle pour la santé publique.

Les avancées considérables en biologie moléculaire durant les deux dernières décennies ont modifié le champ et les méthodes de l'épidémiologie du cancer. Les études actuelles intègrent ces avancées pour explorer les mécanismes de la carcinogenèse, identifier des sous-groupes d'individus susceptibles, et augmenter la précision de l'estimation de l'exposition et la prédiction de l'issue de la maladie.

1 | Identification des facteurs de risque

1.1 Une méthode rigoureuse

L'Organisation mondiale de la santé, dans son Centre international de recherche sur le cancer (Circ, connu aussi sous son nom anglais : International Agency for Research on Cancer IARC) basé à Lyon, évalue depuis 1969 les expositions humaines pour lesquelles il existe une suspicion de cancérogénicité. La méthodologie de ces évaluations est standardisée. Les résultats sont publiés dans des monographies dont l'index est cumulatif, et chaque monographie commence par un préambule qui décrit la méthode (www.iarc.fr). La décision d'évaluer ou de réévaluer une exposition est prise par un groupe d'experts en fonction des besoins et de la disponibilité de nouvelles informations, les évaluations en projet sont listées sur le site Internet de l'agence.

Une fois la décision prise d'évaluer une exposition, un groupe de travail réunissant des spécialistes est constitué et une revue de la littérature est réalisée. Ce groupe international doit :

- vérifier que les données nécessaires ont été rassemblées ;
- sélectionner les données les plus pertinentes en fonction de leur qualité scientifique ;
- résumer ces données de façon à ce que le lecteur puisse suivre le raisonnement du groupe de travail ;
- évaluer les résultats des études épidémiologiques et expérimentales ;
- évaluer les données permettant de comprendre les mécanismes d'action ;
- faire une évaluation globale de la cancérogénicité pour l'homme.

À la fin de cette évaluation, chaque exposition est classée dans une des catégories suivantes :

- groupe 1 : l'exposition est cancérogène pour l'homme ;
- groupe 2 : l'exposition est probablement ou possiblement cancérogène pour l'homme. Sont classées dans cette catégorie des expositions pour lesquelles les données vont depuis « presque suffisantes pour conclure à la cancérogénicité pour l'homme » jusqu'à « aucune donnée sur l'homme, mais cancérogénicité expérimentale chez l'animal ». On distingue le groupe 2A des cancérogènes humains probables (données limitées chez

l'homme et suffisantes chez l'animal, ou pas de données sur l'homme, des données suffisantes chez l'animal et un mécanisme d'action compris et extrapolable à l'homme) et le groupe 2B des cancérogènes possibles (données limitées sur l'homme et insuffisantes chez l'animal, ou pas de données sur l'homme et des données suffisantes chez l'animal) ;

- groupe 3 : l'exposition n'est pas classable quant à sa cancérogénicité pour l'homme. En général cette catégorie correspond à des expositions pour lesquelles les données sur la cancérogénicité pour l'homme sont inadéquates, et les données sur l'animal sont inadéquates ou limitées ;
- groupe 4 : l'exposition n'est probablement pas cancérogène pour l'homme. Cette catégorie correspond aux expositions pour lesquelles les données suggèrent une absence de cancérogénicité chez l'homme et dans les expériences sur l'animal.

Les poids respectifs des données biologiques et épidémiologiques dans le classement d'une exposition dépendent de la compréhension du mécanisme de carcinogenèse qu'apportent les données biologiques. Si les données expérimentales sont très convaincantes, on aura besoin de très peu de données sur l'homme pour conclure ; si les données expérimentales manquent ou ne sont pas concluantes, les données épidémiologiques joueront un rôle essentiel.

1.2 Des résultats

La liste des expositions déjà évaluées figure sur le site du Circ.

Parmi les 95 expositions certainement cancérogènes pour l'homme, le tabac est de loin la plus meurtrière en France. Environ un quart de la mortalité par cancer lui est attribuable (Hill et Laplanche, 2003). Les méfaits du tabac ont été démontrés essentiellement sur la base d'arguments épidémiologiques, corroborés par des expériences sur l'animal. Les premiers éléments de preuve ont été apportés dans les années 1950 (Doll et Hill, 1950 ; Wynder et Graham, 1950 ; Schwartz et Denoix, 1957), mais il a fallu attendre 1976 pour avoir une loi anti-tabac en France et pour voir le début de la réduction de consommation dans la population masculine. Cette loi, très en avance sur son temps, a été seulement partiellement respectée. Elle a été renforcée par une nouvelle loi en 1991 qui, associée à une substantielle augmentation du prix, a conduit à l'arrêt de l'augmentation de la consommation féminine. Il faut au moins 50 ans pour mesurer complètement les effets du tabagisme sur une génération. En 1951, deux épidémiologistes anglais Bradford Hill et Richard Doll ont eu l'idée géniale de prendre les médecins comme sujets d'étude. Ils ont écrit à tous les médecins enregistrés en Grande-Bretagne en leur demandant de répondre à quelques questions, et

en particulier à des questions sur leur consommation de tabac. Ils les ont ensuite suivis en notant simplement s'ils étaient vivants ou non, et quand ils mourraient, en enregistrant la cause de leur décès. L'objectif était, entre autres, d'étudier les risques liés au tabac. Les résultats ont été publiés exactement 50 ans après la première publication, le professeur Sir Richard Doll étant signataire des deux articles (Doll *et al.*, 2004). Cette étude montre l'importance des effets néfastes du tabac. Environ la moitié des personnes qui continuent à fumer des cigarettes en meurent, un quart dans la force de l'âge c'est-à-dire entre 35 et 69 ans, et un quart ensuite. Les hommes nés vers 1920 en Angleterre semblent être la première génération au monde dans laquelle on peut évaluer directement l'ensemble des risques liés à la cigarette et les effets bénéfiques de l'arrêt du tabac. Dans cette génération, la cigarette triple le risque de décès à chaque âge. En moyenne, les fumeurs de cigarettes meurent environ 10 ans avant les non-fumeurs. Arrêter à 60, 50, 40, ou 30 ans améliore l'espérance de vie de 3, 6, 9 ou 10 ans.

Ces données épidémiologiques sont très importantes, mais les données biologiques jouent aussi un rôle dans la démonstration. La mesure du niveau des adduits de l'ADN formés après exposition à la fumée du tabac et l'étude des polymorphismes génétiques qui déterminent le métabolisme des carcinogènes du tabac ou la réparation de l'ADN sont des voies de recherche active (Wiencke, 2002 ; Vineis *et al.*, 2004).

Le cancérigène le plus important pour la population française après le tabac est l'alcool : en effet, on buvait traditionnellement beaucoup d'alcool en France. Ainsi, en 1950, la consommation répartie sur l'ensemble de la population de 15 ans et plus était de 5 verres par jour contre 1,4 aux États-Unis et en Grande-Bretagne. La situation s'est améliorée nettement puisque aujourd'hui on ne boit plus que 2,8 verres par jour en France. La démonstration de la cancérigénicité de l'alcool est entièrement fondée sur des arguments épidémiologiques. Les mécanismes restent assez mystérieux.

1.3 La mesure des risques et l'importance de la dose

Une chose est de conclure qu'une substance est cancérigène, une autre est de mesurer les risques encourus. Oui le tabac et l'amiante sont cancérigènes (et pas l'aspartam), mais les risques encourus par un fumeur de 18 ans qui va rester fumeur toute sa vie et ceux liés à l'amiante encourus par une personne qui n'a jamais été exposé professionnellement à cette substance sont extrêmement différents, le premier a deux chances sur trois que ce soit son tabagisme qui le tue, le second n'encourt aucun risque particulier.

Dans la méthodologie du Circ, l'observation d'une relation entre la dose d'exposition et l'effet cancérigène est un argument de preuve de la causalité de l'association entre l'exposition et le risque de cancer. Cependant tout est dans la dose (phrase attribuée aussi bien à Paracelse qu'à Claude Bernard, voire à Louis Pasteur!).

Quand on s'interroge sur les risques de cancer liés à une exposition humaine, on commence par étudier les risques dans les populations les plus exposées. Ainsi, si on veut savoir si les champs électromagnétiques sont cancérigènes, va-t-on s'intéresser en premier lieu aux personnes travaillant sur des radars ou intervenant sur les lignes à haute tension. Si on veut mesurer le risque de cancer en fonction de l'exposition au radon, on va s'intéresser aux mineurs et particulièrement aux mineurs d'uranium, car dans ces mines les taux de radon sont et surtout étaient élevés. Se pose ensuite la question de l'extrapolation aux faibles doses des effets détectés à des doses élevés. Cette extrapolation est l'objet de nombreux débats dans le domaine des radiations ionisantes.

2 | L'épidémiologie moléculaire des cancers

Le rôle de l'hérédité dans l'étiologie des cancers est devenu une source d'intérêt majeur dans la communauté scientifique. Cet intérêt résulte de connaissances qui ont émergé durant les deux dernières décennies concernant les mécanismes moléculaires de la carcinogenèse.

Le cancer est une maladie complexe, dépendant de nombreux facteurs, facteurs de susceptibilité génétique et autres facteurs de risque (âge, sexe, mode de vie, etc.). La compréhension des mécanismes physiopathologiques à l'origine des cancers constitue donc un enjeu majeur en termes de santé publique. La mise en évidence des gènes et des facteurs modulant l'expression de ces gènes dans le déterminisme des pathologies cancéreuses représente un véritable défi et nécessite la mise en œuvre d'études multidisciplinaires associant cliniciens, biologistes, statisticiens et généticiens épidémiologistes.

Selon les types de cancer, la composante génétique est plus ou moins importante. Des gènes prédisposant à des formes familiales de cancers ont été caractérisés (exemple, BRCA1 et BRCA2 dans les cancers du sein et de l'ovaire, CDKN2A dans le mélanome malin cutané et APC dans le cancer colorectal). Le nombre d'individus porteurs d'une mutation constitutionnelle prédisposant à ces formes de cancer, et partant, le risque attribuable dans la population, sont relativement faibles. Il est maintenant bien établi que les gènes à forte pénétrance, c'est-à-dire dont les mutations sont associées à un risque élevé de

cancer, expliquent seulement une faible proportion de l'ensemble des cancers (IARC Workshop Report, 2004).

Cependant, la très grande majorité des cancers sont sporadiques et ont certainement une étiologie multifactorielle dans laquelle des gènes de susceptibilité agissent conjointement avec des facteurs environnementaux. L'augmentation de risque associée à ces gènes est généralement modérée. Cependant, une proportion non négligeable de cancers pourrait leur être attribuée si les variants augmentant le risque étaient fréquents.

La caractérisation du génome humain a généré une masse considérable d'informations sur la variabilité génétique. Un enjeu majeur pour la communauté scientifique est de savoir comment relier ces variants à des critères fonctionnels et importants cliniquement. La recherche fondamentale se situe au centre de cette problématique en cherchant à caractériser les effets des variants des gènes sur leur expression, leur régulation et leur fonction sur un plan moléculaire. Une approche complémentaire est celle de l'épidémiologie moléculaire, qui, par des études d'association, permet de mettre en évidence des relations entre des variants constitutionnels et des maladies au niveau de la population.

L'identification de nombreux polymorphismes à l'ADN (*Single Nucleotide Polymorphisms*, SNP) a stimulé le développement d'hypothèses pour expliquer la grande variabilité individuelle dans la susceptibilité au cancer observée par les épidémiologistes. Un exemple bien documenté à l'heure actuelle est celui du cancer du poumon. Bien que ce cancer soit essentiellement attribuable au tabac, tous les fumeurs ne développent pas un cancer. Peto *et al.* (2000) ont estimé que le risque cumulé de décès par cancer du poumon à 75 ans (en absence d'autres causes de décès) était de 16 % chez les hommes fumeurs et de 10 % chez les femmes fumeuses. Aussi, de nombreuses études épidémiogénétiques ont été entreprises durant la dernière décennie afin de rechercher d'éventuels facteurs génétiques pouvant expliquer que certains fumeurs soient plus susceptibles que d'autres de développer un cancer lié au tabac (Wu *et al.*, 2004). Cette susceptibilité individuelle aux carcinogènes du tabac chez l'homme peut refléter des différences dans le métabolisme des substances carcinogènes du tabac ou dans la réparation des lésions de l'ADN. Des progrès considérables ont été faits dans notre connaissance des facteurs de susceptibilité génétique aux cancers liés au tabac.

Ces études d'épidémiologie moléculaire prennent une ampleur sans précédent à l'heure où des milliers de SNP vont être identifiés par de nouvelles technologies de génotypage. Leur succès repose sur le recrutement d'un nombre important de sujets, sur la qualité des données cliniques et épidémiologiques, et sur le développement de méthodologies statistiques et d'outils informatiques adaptés à la complexité des données (Vineis *et al.*, 2004). De telles approches

devraient permettre d'identifier les gènes intervenant dans le déterminisme des cancers, leurs interactions avec les facteurs environnementaux, et à terme des sous-groupes d'individus à haut risque de cancer.

L'épidémiologie moléculaire des cancers consiste également à étudier d'autres bio-marqueurs que ceux relatifs à la susceptibilité individuelle, en particulier des marqueurs d'exposition aux carcinogènes tels que les adduits à l'ADN, ou les altérations génétiques acquises dans les tissus. Un certain nombre d'études ont été réalisées en intégrant différents types de biomarqueurs. Par exemple, le niveau d'adduits à l'ADN en hydrocarbures aromatiques polycycliques liés à une exposition au tabac, le polymorphisme du gène codant la glutathion S-transférase M1 (GSTM1) et les échanges entre chromatides sœurs dans les leucocytes ont été analysés dans une même étude sur le cancer du poumon chez les femmes (Tang *et al.*, 1998). Les résultats de cette étude mettent en évidence une interaction entre le niveau d'adduits et la délétion du gène GSTM1, mais pas entre chacun de ces deux biomarqueurs et la présence d'échanges entre chromatides sœurs, suggérant ainsi qu'un niveau élevé d'adduits à l'ADN pourrait contribuer à l'identification des sous-groupes d'individus plus susceptibles à développer un cancer.

Conclusion

L'épidémiologie des cancers a et aura de plus en plus de liens avec la biologie. En effet, nous disposons aujourd'hui de technologies très performantes permettant d'étudier simultanément et rapidement des dizaines de milliers de gènes. Les domaines d'application sont nombreux, tant en épidémiologie qu'en recherche clinique, rendant ainsi possible l'identification de variants génétiques prédisposant au cancer, ou la caractérisation des stratégies thérapeutiques adaptées aux profils génétiques des tumeurs.

Bibliographie

- Doll R., Hill AB. (1950). Smoking and carcinoma of the lung ; preliminary report. *Br Med J*, **2** (4682) : 739-48.
- Doll R., Peto R., Boreham J., Sutherland I. (2004). Mortality in relation to smoking : 50 years' observations on male British doctors. *BMJ*, **328** (7455) : 1519.
- Hill C., Laplanche A. (2003). Tabagisme et mortalité : aspects épidémiologiques. *BEH*, **22-23** : 98-100.

- IARC Workshop Report (2004). *In : Mechanisms of carcinogens – Contributions of molecular epidemiology. IARC Scientific Publications*, n° 157.
- Peto R., Darby S., Deo H., Silcocks P., Whitley E., Doll R. (2000). Smoking, smoking cessation, and lung cancer in the UK since 1950 : combination of national statistics with two case-control studies. *BMJ*, **321** : 323-329.
- Schwartz D., Denoix PF. (1957). (French investigations on the etiology of bronchopulmonary cancer ; role of tobacco.) *Sem Hop*, **33** (62/7) : 3630-43.
- Tang DL., Rundle A., Warburton D., Santella RM., Tsai WY., Chiamprasert S., Hsu YZ., Perera FP. (1998). Associations between both genetic and environmental biomarkers and lung cancer : evidence of a greater risk of lung cancer in women smokers. *Carcinogenesis*, **19** : 1949-53.
- Vineis P., Schulte PA., Carreon T., Bailer AJ., Medvedovic M. (2004). Issues of design and analysis in studies of gene-environment interactions. *In : Mechanisms of carcinogens – Contributions of molecular epidemiology. IARC Scientific Publications*, n° 157.
- Wiencke JK. (2002). DNA adduct burden and tobacco carcinogenesis. *Oncogene*, **21** (48) : 7376-91.
- Wu X., Zhao H., Suk R., Christiani DC. (2004). Genetic susceptibility to tobacco-related cancer. *Oncogene*, **23** : 6500-23.
- Wynder EL., Graham EA. (1950). Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma ; a study of 684 proved cases. *J Am Med Assoc*, **143** (4) : 329-36.

CHAPITRE 6

L'épidémiologie sociale

HENRI LÉRIDON ET MARCEL GOLDBERG

1 | Pourquoi s'intéresser à la dimension sociale des problèmes de santé ?

On observe que, dans la plupart des populations, les problèmes de santé n'affectent pas tous les groupes sociaux de la même façon. Les écarts entre groupes constitués selon des critères de sexe, d'âge, d'appartenance sociale, de niveau d'instruction... pour la fréquence d'une maladie (ou des décès qui en résultent) sont souvent de même ampleur que ceux résultant d'autres facteurs usuellement étudiés en épidémiologie : ces facteurs de risque « classiques » (comme le tabac, l'alcool, l'alimentation ou les conditions de travail) ne peuvent donc pas expliquer à eux seuls les inégalités de santé. Comme le constat s'applique, de façon plus ou moins marquée, à tous les pays, le phénomène ne peut pas être imputé non plus seulement à un système politico-social donné. De plus, les inégalités observées sont parfois identiques (dans leur niveau et leur structure) pour des maladies aux mécanismes physiopathologiques très différents (par exemple tuberculose et schizophrénie), ce qui peut faire douter d'une origine biologique commune.

Il faut donc tenter d'expliquer ces résultats, qui obligent à se poser des questions allant bien au-delà des interprétations purement biologiques ou génétiques. On pouvait d'abord songer aux différences de comportements et de conditions de vie ou de travail, qui pourraient expliquer pourquoi certains individus sont plus particulièrement exposés à certains risques, ou développent des formes plus graves de maladies faute de soins appropriés : mais on a montré que ces différences ne suffisent pas à expliquer entièrement les écarts entre groupes sociaux.

Dans l'ensemble, l'épidémiologie sociale est d'abord une épidémiologie « d'observation », s'appuyant sur des sources statistiques existantes ou, comme on le verra, sur des données de cohortes dont elle constitue un outil d'analyse important. Elle peut être particulièrement importante dans l'analyse des suivis de traitements au long cours, dans lesquels les caractéristiques sociales du patient conditionnent en partie son observance et d'autres comportements pouvant modifier l'efficacité du traitement. Symétriquement, l'existence de la maladie et la nature des traitements peuvent affecter la situation sociale de l'individu. L'encadré 6.1 le montre pour le cas de l'infection à VIH.

Encadré 6.1

Les inégalités sociales dans la maladie VIH chronique

Le cadre d'étude constitué par la maladie chronique permet une approche de la dynamique des inégalités sociales de santé complémentaire de celle

apportée par les études en population générale. En effet, ce cadre permet d'observer de façon fine les différents processus intriqués qui sont à l'origine des inégalités sociales de santé : rôle propre des caractéristiques socio-économiques sur l'évolution de l'état de santé des personnes malades ; rôle de la maladie chronique sur l'évolution de la situation sociale selon les caractéristiques initiales des personnes ; rôle de ces conséquences sociales de la maladie sur l'état de santé (Dray-Spira et Lert, 2003).

Dans le contexte actuel de large accès aux traitements antirétroviraux hautement actifs, l'infection par le VIH en France, du fait de ses spécificités (maladie chronique touchant des sujets jeunes aux caractéristiques sociales contrastées), constitue un cadre d'étude des inégalités sociales de santé particulièrement intéressant. Les données publiées suggèrent l'existence de nombreuses interrelations entre l'état de santé des personnes infectées par le VIH, la maladie VIH elle-même et les conditions de vie. Ces relations ont pu être masquées au début de l'épidémie du fait de l'histoire naturelle de l'infection VIH qui écrasait les différences sociales, mais la diffusion des traitements efficaces et la chronicisation de la maladie ont donné lieu à la mise en évidence de l'existence d'inégalités sociales dans les conséquences de l'infection à VIH.

Influence de la situation sociale sur l'état de santé dans le contexte du VIH

Dans différents pays, de nombreuses études prospectives destinées principalement à suivre le devenir des patients infectés par le VIH ont montré l'influence de certains facteurs d'ordre socio-économique — sexe, origine ethnique, niveau d'éducation, niveau de ressources, statut vis-à-vis de l'emploi, niveau de soutien social, mode de contamination par le VIH — sur la progression clinique et/ou biologique de la maladie (Egger *et al.*, 2002 ; Lee et Rotheram-Borus, 2001 ; Rapiti *et al.*, 2000). Plusieurs types de différences sont susceptibles d'expliquer les inégalités sociales rapportées dans ces études.

– *Différences dans l'accès au système de soins.* L'utilisation des traitements antirétroviraux hautement actifs a permis de mettre en évidence de façon flagrante des différences d'accès aux soins parmi les personnes séropositives. Ces différences dans l'accès aux soins peuvent se produire à chacune des étapes de la trajectoire médicale : délais entre la contamination et le diagnostic puis entre le diagnostic et le début de la prise en charge ; lieu de prise en charge ; caractéristiques de la structure de prise en charge ; modalités du suivi médical ; accès aux traitements antirétroviraux ; type de traitement antirétroviral prescrit.

– *Différences dans l'observance thérapeutique.* Les multithérapies anti-rétrovirales, traitements lourds et contraignants, posent des problèmes

d'observance thérapeutique. L'adhésion au traitement dépend de phénomènes dynamiques : durée d'évolution de la maladie VIH, événements de vie négatifs ou interactions avec le système de soins. Les conditions de vie et caractéristiques sociodémographiques et psychocomportementales des malades affectent aussi l'observance thérapeutique.

– *Différences dans la morbidité associée à l'infection par le VIH.* En Europe et aux États-Unis, depuis 1998, les complications hépatiques de l'infection à VHC (virus de l'hépatite C) constituent la première cause de décès parmi les personnes atteintes. D'autres comorbidités fréquentes (maladies mentales, tuberculose, pneumopathies) peuvent également avoir un impact non seulement sur le pronostic de la maladie VIH elle-même par le biais d'une interaction directe avec l'infection par le VIH et/ou d'une prise en charge médicale rendue plus complexe, mais aussi sur la mortalité non directement liée à ce virus. En France, parmi les patients de la cohorte APROCO décédés entre 1997 et 2000, la cause du décès était indépendante du VIH dans un tiers des cas.

– *Différences dans les niveaux d'exposition au risque d'autres pathologies.* Les comportements à risque tels que la consommation de tabac, d'alcool, de drogue ou d'autres toxiques, tout comme la précarité des conditions de vie, peuvent constituer par eux-mêmes des facteurs de risque pour d'autres pathologies ou pour des traumatismes, notamment les cancers, la tuberculose, les suicides, les overdoses mortelles ou les accidents. Ces comportements et conditions de vie ont également vraisemblablement une influence sur les comportements relatifs à la santé, qu'il s'agisse des habitudes alimentaires ou du niveau d'activité physique qui pourraient favoriser voire compliquer certains effets indésirables des traitements antirétroviraux, ou du suivi médical systématique en dehors de la maladie VIH elle-même (suivis dentaire, gynécologique, etc.) qui, dans le cadre d'une maladie chronique, prend toute son importance en termes de prévention.

Conséquences sociales de la maladie VIH

– *Sur le statut vis-à-vis de l'emploi.* Le statut vis-à-vis de l'emploi joue un rôle majeur sur les conditions de vie des personnes atteintes de maladie chronique. Le statut vis-à-vis de l'emploi est lié non seulement à la sévérité de l'infection VIH, mais aussi aux niveaux d'éducation et de qualification professionnelle et à la place des sujets sur le marché du travail : à état d'avancement de la maladie comparable, le risque de perte d'emploi est plus élevé chez les patients ayant un niveau d'éducation faible, un emploi précaire, un travail physique, un faible contrôle sur le rythme et l'organisation de leur travail, et chez les patients issus de minorités ethniques. Dans le contexte d'une maladie de longue durée, l'accès, le maintien et le retour à l'emploi sont plus difficiles pour les personnes ayant un autre handicap sur le marché du travail (personnes les moins éduquées, non qualifiées, ayant une expérience

professionnelle faible ou ancienne) ou pour certains étrangers n'ayant pas de statut légal permettant de travailler. Le maintien, l'accès ou le retour à l'emploi sont également liés aux attitudes des employeurs et au climat de tolérance ou de discrimination vis-à-vis de l'infection VIH dans le milieu du travail.

– *Sur le niveau de ressources.* En Australie, 45 % des personnes séropositives interrogées en 1996 déclaraient avoir des revenus inférieurs au seuil de pauvreté, et 73 % rapportaient une diminution de leurs ressources consécutive à la maladie. La source principale de revenus était constituée par une aide sociale pour 58 % des patients. La survenue de difficultés économiques était associée de façon indépendante avec le statut vis-à-vis de l'emploi, l'état de santé déclaré et le sexe.

– *Sur les relations avec l'entourage.* Les phénomènes de discrimination liée à la maladie, différents selon les groupes de patients, peuvent provoquer ou renforcer un isolement social. Parmi un échantillon de patients suivis en France, la majorité (60 %) vivaient seuls en 2001 ; plus du tiers rapportent avoir été l'objet d'attitudes de rejet ou de discriminations liées au VIH/sida dans les différents domaines de leur vie sociale.

En révélant les inégalités sociales face à la maladie ou plus généralement dans les comportements de santé, l'épidémiologie sociale peut aider à un meilleur ciblage des politiques publiques. C'est évident dans le cas de maladies professionnelles, mais c'est vrai aussi de maladies non professionnelles : on ne peut, certes, pas modifier l'appartenance sociale d'un individu (encore que sortir une personne de la marginalité sociale soit un objectif légitime), mais on peut tenter d'agir sur un certain nombre de comportements qui sont propres à son groupe, quitte à renvoyer à des processus d'éducation s'inscrivant dans le long terme.

2 | Quelques règles de travail

Posons d'abord quelques préalables importants :

- il ne s'agira pas, ici, de discuter des approches exclusivement sociologiques de l'analyse des expositions au risque ou des comportements de santé, mais de s'intéresser aux approches épidémiologiques du problème, qui s'appuient donc sur des outils classiques de la discipline ou sur des outils plus spécifiques adaptés d'autres champs disciplinaires ;

- on restera centré sur les questions de santé proprement dites ; les différences dans l'accès aux soins, par exemple, constitueront — dans ce contexte — une variable explicative plutôt qu'à expliquer ;
- on se situera principalement au plan observationnel : description, analyse de facteurs de risques (de maladies et de leur issue fatale éventuelle), mais aussi recherche d'hypothèses permettant de comprendre les inégalités et efforts pour les tester. Indiquons cependant que, dans divers pays, on a réalisé des évaluations, selon des méthodes épidémiologiques, de tentatives d'interventions. Nous évoquerons aussi, brièvement, certains cadres théoriques qui ont été proposés, notamment par des sociologues et des anthropologues ;
- point essentiel, enfin : compte tenu de la nature des variables en cause, des approches pluridisciplinaires seront nécessaires, impliquant notamment la sociologie, la psychosociologie, l'anthropologie, l'économie ou la démographie. Il est donc important de préciser d'abord les rôles plus spécifiques de ces diverses disciplines.

3 | L'apport d'autres disciplines

Nous en aborderons quatre : la démographie, les sciences sociologiques, les sciences économiques et la géographie.

La démographie s'intéresse principalement à *la mortalité et aux causes de décès*, parce que son matériau privilégié est l'état civil. Les données permettant d'interpréter la cause de décès (telle que décrite dans les nomenclatures internationales de type Cim de l'OMS, par exemple) concernent habituellement : le sexe, l'âge, le statut matrimonial, le lieu de résidence, le statut socioprofessionnel. Les analyses démographiques permettent donc surtout d'identifier les groupes les plus à risque (de décès, pour une maladie donnée), selon les caractéristiques énumérées ci-dessus, et leurs évolutions temporelles ; l'inférence causale est limitée (cf. encadré 6.2). Autre domaine d'intérêt : *la reproduction*. On s'intéressera moins, ici, aux déterminants des niveaux actuels de la fécondité, celle-ci étant maintenant largement contrôlée par les couples, qu'aux variables « intermédiaires » qui permettent de comprendre comment ces niveaux sont obtenus : comportements sexuels, recours à la contraception et à l'avortement volontaire, effet des infections sexuellement transmissibles... De plus, l'exigence croissante de la part des couples d'obtenir le nombre exact d'enfants souhaité (et dans les délais souhaités) donne encore plus d'importance aux éventuels facteurs biologiques limitatifs et aux remèdes possibles (aide médicale à la procréation).

Encadré 6.2

Les différentielles sociales de mortalité

On constate de très importantes différences sociales en terme de mortalité dans tous les pays, mais la situation de la France, si elle reste mal connue, semble particulièrement préoccupante de ce point de vue par rapport à d'autres pays de niveau économique comparable (Kunst *et al.*, BMJ 1998).

Ainsi, le tableau 6.1, qui considère la mortalité toutes causes, montre qu'il existe en France pour les hommes adultes dans la période 1990-1995, un rapport entre catégories extrêmes de presque 2,5 pour le risque de décéder, et que l'augmentation du risque suit un gradient monotone selon la hiérarchie sociale. Quand on détaille les causes à l'origine des décès en comparant les ouvriers et employés aux cadres supérieurs et professions libérales, on constate que les différences sociales de mortalité concernent la plupart des maladies et des accidents, l'écart pouvant atteindre un rapport de 1 à 10 pour les maladies liées à l'alcool (Tableau 6.2).

Qui plus est, il semble bien que les inégalités sociales de mortalité n'ont pas diminué ces dernières décennies, malgré l'augmentation continue de l'espérance de vie de la population française, comme on le voit au tableau 6.3 qui montre l'évolution de la mortalité pour les principales catégories socio-professionnelles entre 1975 et 1995, comparativement à celle des cadres supérieurs et professions libérales.

Sources : Kunst *et al.* (1998) ; Mesrine (1999).

La sociologie, la psychosociologie et l'anthropologie fournissent des cadres de référence théorique permettant à la fois d'interpréter des comportements pouvant augmenter ou diminuer les risques de contracter une maladie, et d'analyser les autres sources d'inégalités tenant aux modes d'organisation de la société ou aux inégalités de « capital social » entre individus. Les observations anthropologiques, par exemple, peuvent être très précieuses pour signaler un problème jusque-là méconnu, ou pour proposer une interprétation de phénomènes observés (surmortalité, par exemple) en fonction de comportements spécifiques. On peut citer le cas de la maladie de Kourou qui a mis sur la piste de prions (la tradition était, dans une tribu de Nouvelle-Guinée, de manger les cerveaux des défunts) (Mathews *et al.*, 1968) ou celui du cancer du nasopharynx, dans lequel un facteur alimentaire a été clairement mis en évidence dans certaines populations (Poirier *et al.*, 1987). Sur un plan plus sociologique, deux courants de pensée principaux peuvent être cités.

PCS	Risque de mortalité par rapport aux cadres
Cadres, professions libérales	1,0
Agriculteurs exploitants	1,3
Artisans, commerçants	1,3
Professions intermédiaires	1,3
Employés	2,2
Ouvriers qualifiés	2,0
Ouvriers non qualifiés	2,4

Tableau 6.1

Mortalité des hommes selon la catégorie socioprofessionnelle (PCS). France, 1990-1995.

Hommes âgés de 30 à 64 ans et en activité au début de la période. Groupe professionnel au début de la période considérée.

L'approche « objectiviste » est la plus proche de l'épidémiologie classique. Elle suppose que les individus basent leurs comportements, de façon assez rationnelle, sur l'information (bonne ou mauvaise) dont ils disposent, en cherchant à minimiser les risques de conséquences néfastes. Dans le *Health Belief Model*, par exemple, la décision d'adopter un comportement résulte de la perception qu'a l'individu de sa vulnérabilité à un risque (contracter le VIH, par exemple), de la gravité des conséquences qui en résultent, des coûts et des bénéfices qu'il attache à l'adoption de moyens de prévention. Dans cette perspective, la *logique de préservation de la santé* est pensée et posée comme étant la plus structurante des comportements des individus.

L'approche « constructiviste » met davantage l'accent sur les processus de définition sociale du risque, et sur l'interaction entre un ensemble de risques et de logiques parfois contradictoires auxquels un même individu est soumis ; elle ne préjuge pas d'une place prépondérante d'une de ces logiques dans les conduites de préservation de la santé. L'importance du capital social et culturel, notamment tel que Bourdieu l'incorpore dans la notion d'*habitus*, est souvent évoquée ici. Dans ce contexte, le comportement d'un acteur peut sembler à première vue « incohérent », si l'on ne tient pas compte de l'ensemble des enjeux pour l'environnement social et relationnel de la personne.

L'économie s'intéresse notamment (pour le domaine qui nous occupe ici) aux inégalités dans l'accès aux soins pouvant résulter des inégalités de revenus ou de prise en charge par le système de protection sociale, à l'évaluation

	Ouvriers-Employés	Cadres supérieurs Professions libérales	Surmortalité ouvriers/employés
	(a)	(c)	(a/c)
Tumeurs	119,2	35,1	3,4
– poumon	30,8	8,9	3,5
– VADS (1)	32,3	3,0	10,8
– intestin	5,4	2,4	2,3
– estomac	3,8	1,2	3,1
– pancréas	3,7	1,7	2,2
– encéphale	3,3	2,3	1,4
– autres	39,9	15,7	2,5
Cardiovasculaire	52,4	18,4	2,8
– infarctus du myocarde	23,8	9,7	2,5
– accidents vasc.-cérébraux	10,5	3,0	3,5
– autres	18,1	5,7	3,2
Morts violentes	104,7	37,6	2,8
– suicides	39,2	13,8	2,8
– accidents circulation	31,2	11,3	2,8
– autres accidents	25,7	9,7	2,7
– autres	8,6	2,7	3,1
Alcoolisme (2)	25,0	2,5	10,0
Maladies respiratoires	7,2	1,6	4,5
Sida	14,9	16,6	0,9
Diabète	1,7	0,3	5,8
Causes inconnues	15,4	7,9	1,9
Autres causes	25,3	7,2	3,5
Ensemble	365,9	127,3	2,9

Source : SC8-Inserm ; taux standardisés par âge pour 100 000 (population de référence : 1990).
(1) Voies aérodigestives supérieures et œsophage. (2) Cirrhoses et psychoses alcooliques.

Tableau 6.2

Taux de décès par cause – France : hommes actifs de 25-54 ans, 1987-1993.

de l'efficacité des actions de santé selon les populations visées, à la « rationalité » des décisions individuelles en matière de santé et à leurs variations sociales. Le champ étant beaucoup plus investi par des économistes que par

Groupe professionnel	Période 1975-1980		Période 1982-1987		Période 1990-1995	
	Risque relatif	% de la pop.	Risque relatif	% de la pop.	Risque relatif	% de la pop.
Agriculteurs exploitants	1,3	11,3	1,2	9,3	1,3	6,1
Artisans, commerçants	1,4	11,0	1,3	10,8	1,3	10,9
Cadres, prof. libérales	1	10,7	1	13,4	1	16,7
Prof. intermédiaires	1,2	17,2	1,2	20,4	1,3	21,3
Employés	2,0	10,1	1,9	11,1	2,2	9,5
Ouvriers qualifiés	1,9	24,3	1,8	24,0	2,0	25,7
Ouvriers non qualifiés	2,6	15,4	2,5	11,0	2,4	9,8

Tableau 6.3

Évolution de la mortalité des hommes selon les groupes sociaux de 1975 à 1995.

Hommes âgés de 30 à 64 ans et en activité au début de la période considérée. Groupe professionnel au début de la période considérée.

des épidémiologistes, nous ne considérerons pas qu'il relève de l'épidémiologie sociale au sens strict et n'en traiterons pas ici.

La géographie propose une façon de catégoriser les observations sur une base spatiale. Elle permet donc de mettre en évidence des différences régionales (une opposition Nord-Sud en termes de mortalité en France, par exemple) et de les relier à d'autres observations géographiques portant sur des variables potentiellement explicatives (Salem *et al.*, 1999). La visualisation de ces similitudes est souvent très éclairante, mais les corrélations ainsi proposées doivent être interprétées avec prudence : il faut, notamment, se méfier de l'*ecological fallacy* consistant à transposer au niveau des individus les corrélations portant sur des agrégats géographiques (Kleinbaum, 2002).

4 | Les outils de l'épidémiologie sociale

4.1 Des variables d'une nature différente

L'épidémiologie sociale s'appuie d'abord sur les outils classiques de l'épidémiologie, qui traite de « facteurs de risque » pouvant se situer à divers niveaux. Mais ici *la nature des variables en cause impose une certaine prudence* : on ne peut pas mettre exactement sur le même plan le fait d'avoir tel profil professionnel et la présence de tel gène ou le résultat de tel dosage biologique. En effet, le fait d'appartenir à une classe sociale donnée ne peut pas être considéré comme

inhérent à l'individu (comme le fait d'avoir hérité de tel ou tel gène), mais bien plutôt comme le résultat d'une construction sociale de la société — et même de ceux qui observent celle-ci et qui y constituent des catégories —, et d'un itinéraire personnel depuis la famille d'origine jusqu'à la situation actuelle. Une caractéristique sociale a donc toutes les chances d'être multidimensionnelle par construction. Il existe divers marqueurs de la situation sociale de l'individu qui sont utilisés en épidémiologie sociale, comme le diplôme le plus élevé ayant été obtenu, les revenus, la PCS (Profession et catégorie sociale résumant à la fois le type de travail effectué et la situation vis-à-vis de l'emploi) ou la profession du conjoint, chacune de ces variables correspondant à une facette particulière de la position socio-économique. Les mécanismes par lesquels chaque variable est associée à la santé peuvent être spécifiques ; ainsi le diplôme indique notamment la capacité à s'approprier des informations relatives à la santé et permet de repérer la situation sociale au début de la vie adulte, les revenus déterminent largement des conditions de vie matérielles, tandis que la profession reflète entre autres l'exposition à des facteurs de risque professionnels et des contraintes de travail. Ces différents indicateurs socio-économiques sont complémentaires. Par ailleurs, il apparaît de plus en plus que la trajectoire sociale des individus tout au long de la vie influe sur l'état de santé à l'âge adulte.

4.2 Un champ particulier : travail et santé

La santé au travail constitue un champ privilégié pour l'épidémiologie sociale. Au-delà des expositions physiques et chimiques, dont la prise en compte s'impose à l'évidence, on s'intéresse de plus en plus au stress au travail, possible générateur de troubles mentaux, d'altération de la qualité de vie, mais aussi de problèmes cardiovasculaires et de santé mentale, par exemple (Marmot *et al.*, 1998 ; Niedhammer *et al.*, 1998 ; Cheng *et al.*, 2000). Caractériser un individu par son activité professionnelle, c'est en même temps définir une partie (peut-être la plus déterminante) de son « environnement » immédiat et des expositions auxquelles il est soumis, en se donnant les moyens d'objectiver celles-ci.

4.3 Des références théoriques

L'épidémiologie sociale fait aussi référence, comme on l'a dit, à des modèles théoriques globaux. On a notamment proposé deux modèles importants. Le modèle « matérialiste » renvoie aux conditions matérielles de vie et de travail, au manque de ressources matérielles personnelles pour faire face à des conditions défavorables, aux limitations de ces mêmes ressources dans l'environnement social ; dans cette approche, on admet que l'organisation et le fonctionnement du

système économicosocial sont producteurs d'inégalités entre individus et groupes sociaux (Lynch *et al.*, 2000). Le modèle « psychosocial », de son côté, part du constat que l'augmentation des ressources individuelles ne garantit pas automatiquement ou indéfiniment un meilleur accès aux soins, une meilleure santé ou une moindre mortalité. Ce serait la distribution inégalitaire des ressources matérielles et culturelles qui jouerait le rôle majeur, au-delà des niveaux absolus, lesquels peuvent s'élever sans que l'on constate aucun progrès sanitaire (Marmot, 2001). Les approches théoriques de Durkheim et de Bourdieu sont fréquemment utilisées ici.

4.4 L'analyse de biographies

Le fait que les facteurs auxquels on s'intéresse appartiennent souvent à l'histoire personnelle de l'individu, et que certaines des variables utilisées peuvent changer au cours de la vie de l'individu, conduisent à recourir aux *méthodes d'analyse des biographies (life-course analysis)* qui mettent en jeu des « cohortes vie entière » : il s'agit de l'adaptation de modèles statistiques multivariés à l'étude de l'incidence d'événements, successifs et/ou concurrents. L'interprétation des résultats fait appel à diverses théories : le *latency model* postule que certains événements ne produisent leurs effets qu'après des années de latence, tout comme certaines expositions à des nuisances environnementales n'ont d'effets cancérigènes qu'à long terme ; le *pathway model* met l'accent sur l'effet cumulatif de ces facteurs précoces et surtout sur ses interactions avec les conditions et événements vécus tout au long de la vie (Kuh et Ben-Schlomo, 1997). Ces analyses supposent évidemment de suivre des individus pendant de longues périodes, de préférence depuis la naissance (ou même pendant la grossesse), puisque l'on sait maintenant que la grossesse, la naissance et les premières années de la vie conditionnent beaucoup d'événements ultérieurs (Poulton *et al.*, 2002) (*cf.* encadré 6.3).

Encadré 6.3

L'intérêt des suivis sur grandes cohortes

Les enquêtes de cohorte portant sur des échantillons constitués en population générale ou sur des sous-groupes particuliers (salariés du secteur public ou d'une entreprise publique, habitants de grandes villes. . .) mais non sélectionnés sur des critères de maladie ou d'exposition à des risques spécifiques, sont d'intérêt tout spécial en épidémiologie sociale. Elles permettent, en effet, de recueillir des informations sur un grand nombre d'indicateurs d'appartenance sociale et de comportements, de repérer l'évolution de ces variables dans le temps, et de mieux appréhender les interactions entre variables.

Elles sont particulièrement efficaces quand leurs données peuvent être couplées à des informations externes solides en matière de santé, comme des registres de maladies ou de décès.

Un exemple d'étude de cohorte à fort contenu social est la Whitehall Study anglaise (Marmot *et al.*, 1987). Elle porte sur des fonctionnaires britanniques suivis pendant dix ans pour analyser les facteurs de risque de décès par cancer et maladies cardiovasculaires. La mortalité des fonctionnaires occupant les fonctions les moins qualifiées est trois fois supérieure à celle des administrateurs. Fait plus surprenant, lorsque l'on contrôle pour une série de facteurs dont l'association avec le risque de décès par cancer ou maladie cardiovasculaire est établie (obésité, habitude de fumer, peu d'activité physique, hypertension. . .), le rapport de risque est encore supérieur à 2.

Les Britanniques sont aussi les spécialistes de cohortes permettant le suivi d'enfants depuis la naissance et jusqu'à l'âge adulte, voire sans limite de durée. La plus ancienne de ces cohortes a été constituée en 1946 (NSHD) : commencée par une étude de périnatalité auprès de 17 000 nouveau-nés, elle s'est poursuivie sur un sous-échantillon de 5 400 enfants qui ont fait l'objet de 20 enquêtes depuis 1946. Un second échantillon de même taille a été constitué en 1958 (NCDS), et un autre en 1970 (BCS70), cette fois suivis intégralement. Toutes ces cohortes sont encore « actives » aujourd'hui : plus de 70 % de la population d'origine est toujours en observation. Tout récemment, les Anglais ont lancé en 2000 la « Millenium Cohort Study » (MCS) : plus de 21 000 enfants nés en 2000-2001, dont une partie des enfants nés de la cohorte 1970. Bien que centrées sur les questions de développement de l'enfant dans son contexte familial, social et économique, ces études ont aussi apporté des résultats significatifs en santé publique. Dans un récent bilan de l'expérience de ces cohortes britanniques, John Bynner (2002) écrivait :

« La première évidence d'un risque associé au fait de fumer pendant la grossesse pour le développement ultérieur de l'enfant a été tirée des données NCDS. On a de la même façon mis en évidence l'intérêt du lait maternel par rapport au lait artificiel. La valeur comparée des tests de vision et d'audition réalisés aux différents âges a pu être démontrée. Plus récemment, on a pu montrer le rôle de la pollution atmosphérique sur l'asthme et les difficultés respiratoires. Les effets positifs et négatifs des vaccinations ont été étudiés dans ces cohortes, ainsi que l'étiologie de maladies graves comme le cancer. L'intérêt croissant pour les recherches sur les inégalités de santé a grandement bénéficié des données rassemblées dans les cohortes d'enfants, comme en témoigne par exemple le récent rapport Acheson au gouvernement (1998) ».

Un intérêt spécifique se développe aujourd'hui pour la mise en relation de la santé des enfants avec leur environnement, physique et social. Aux États-Unis, le projet de National Children Study, piloté par le NICHD, ambitionne

ainsi de suivre 100 000 enfants depuis la grossesse jusqu'à l'âge adulte. Un projet intégrant les aspects britanniques et américains est en cours d'élaboration en France (20 000 enfants).

Sources : D.S. Acheson (1998). J. Bynner (2002). M.G. Marmot *et al.* (1987).

4.5 Analyses multiniveaux

L'importance attachée à la description de l'environnement de l'individu conduit à considérer que les variables utiles peuvent se situer à plusieurs niveaux : caractéristiques personnelles d'abord (comme le sexe, l'âge, le niveau d'études...), mais aussi modes d'insertion sociale (données sur le type d'habitat environnant, sur l'accessibilité au système de soins, sur les relations sociales, etc.) ; on peut analyser ceux-ci en retenant plusieurs niveaux superposés, par exemple en premier lieu le quartier de résidence (qui détermine largement le cadre de la vie quotidienne : Diez Roux *et al.*, 2001) et en second lieu une unité géographique plus vaste qui soit pertinente en termes d'équipement médical ou de réglementation sociale. Les *modèles d'analyse multiniveaux* (Goldstein, 1995) permettent de traiter de façon statistiquement correcte ces diverses catégories de variables, en évitant un double écueil : soit de considérer que les caractéristiques personnelles d'un individu suffisent à déterminer complètement ses comportements, soit de considérer que ceux-ci sont entièrement déterminés par l'appartenance à telle ou telle entité géographique ou autre. Dans cette approche, la description de « l'environnement communautaire » tient une place essentielle : elle peut reposer, par exemple, sur des « indicateurs de quartier » dont les dimensions sont potentiellement innombrables. L'un des problèmes est de les réduire à des variables synthétiques pertinentes, sans les dénaturer ou en forcer le sens.

4.6 Les méthodes d'enquête

Lorsqu'il s'agit de décrire des différences sociales de santé et les variables qui y sont associées, l'épidémiologie sociale s'appuie souvent sur des enquêtes en population générale, qui posent des problèmes différents des études cas-témoins ou des suivis cliniques le plus souvent utilisées en épidémiologie. Constituer un échantillon en population générale, ou dans un large sous-ensemble (tout un secteur professionnel, par exemple), suppose de disposer d'une bonne base de sondage et d'avoir les moyens d'accéder aux individus inclus dans l'échantillon (qui peuvent être dispersés sur tout le territoire). Il faut ensuite évaluer le degré de représentativité de l'échantillon des répondants, estimer les biais potentiels liés aux valeurs manquantes, procéder éventuellement à un « redressement »

de cet échantillon, et finalement construire des estimateurs des variances des résultats obtenus, tenant compte des modalités souvent complexes du plan de sondage. Ces problèmes ne sont pas spécifiques à l'épidémiologie sociale (les enquêtes sociodémographiques y sont depuis longtemps confrontées), mais ils fournissent l'occasion de collaborations fructueuses avec les statisticiens spécialistes des sondages.

De très intéressants travaux d'épidémiologie sociale ont aussi été réalisés sur des groupes sociaux relativement homogènes, comme les agents de la fonction publique ou ceux d'une grande entreprise publique. Ces études, qui sont plus centrées sur la compréhension des mécanismes de détermination sociale de la santé, prennent alors souvent la forme d'études de cohortes, permettant de mettre en œuvre les méthodes biographiques évoquées plus haut (Führer *et al.*, 2002). En « neutralisant » l'une des caractéristiques de la diversité sociale, elles permettent — néanmoins — de mieux analyser le rôle de certains facteurs d'origine sociale qu'une trop grande hétérogénéité sociale risquerait de masquer.

5 | L'utilisation des résultats

Les conclusions des études épidémiologiques classiques peuvent souvent être traduites en recommandations pour des actions de santé publique. Si l'on identifie, par exemple, le tabac comme cause probable de nombreux cancers, il faudra s'efforcer de réduire le nombre de fumeurs dans la population ; si l'on constate la présence d'une substance nuisible dans l'air ou dans l'eau de boisson, il faudra s'efforcer de réduire les émissions pour améliorer l'environnement et/ou demander aux personnes exposées d'adopter certains comportements de protection. En revanche, si l'on observe que les ouvriers ou les personnes peu diplômées sont plus exposés à certains risques que d'autres, on ne peut pas leur demander de changer de catégorie sociale ou de passer leur baccalauréat. Des interventions sont cependant possibles à deux niveaux.

Si les populations concernées constituent un groupe spécifique et repérable dans la population, on peut alors cibler des campagnes d'intervention sur ces groupes précis, soit en intervenant sur le lieu de travail (s'il s'agit d'une exposition professionnelle), soit en s'appuyant sur les réseaux sanitaires ou les systèmes d'aide sociale. Si ces populations ne sont pas directement repérables, parce que les expositions ou comportements en cause ne sont pas le fait de groupes sociaux spécifiques, il faut alors penser en termes de programmes en population générale, s'appuyant sur les bases théoriques mentionnées plus haut. On pourra, par exemple, chercher à modifier un comportement « à risque » en expliquant le lien entre le comportement et le risque encouru, lien qui n'est pas toujours évident à comprendre. L'insistance mise sur l'utilisation du préservatif dans les campagnes anti-sida en est un bon exemple : en une dizaine d'années,

par exemple, le taux d'utilisation du préservatif au premier rapport (rapport qui peut être particulièrement à risque, le partenaire n'étant pas forcément connu depuis longtemps) est passé de 10-15 % à plus de 80 %, sans que les personnes concernées aient pu être approchées personnellement, dans l'immense majorité des cas, et sans qu'elles aient forcément tout compris des mécanismes de transmission précis du VIH.

Dans d'autres circonstances, l'amélioration de la santé passera par des actions se situant plus « en amont », c'est-à-dire des actions d'éducation en matière de nutrition, d'hygiène personnelle, de mode de vie, et d'information sur les possibilités offertes par le système de santé du pays.

Enfin, les résultats des travaux d'épidémiologie sociale peuvent parfois servir de guide pour définir des politiques publiques dans des domaines divers : politique de l'emploi, politique de la protection sociale, politique de la ville, politique de l'éducation, etc. En effet la mise en évidence des implications sur la santé de trajectoires socioprofessionnelles trop parcellisées, de certaines circonstances sociofamiliales, de certains environnements résidentiels dans les quartiers de vie donne des indications essentielles pour le choix des politiques publiques et leur évaluation.

6 | Forces et faiblesses de l'épidémiologie sociale en France

Ce domaine de recherche épidémiologique est encore peu présent en France : on ne compte aucune unité de recherche consacrant toute son activité à l'épidémiologie sociale, et sans doute pas plus de quatre ou cinq qui abordent ce thème à l'occasion de travaux centrés sur des problèmes de santé spécifiques. Dans un article récent de présentation de ce domaine de recherche (dans une revue française : Goldberg *et al.*, 2003), les auteurs listaient 167 références : 23 % concernaient des auteurs français, mais pour plus de la moitié il s'agissait de contributions à 3 ouvrages parus en français, et seulement 7 % étaient des publications dans des revues de bon niveau international.

À l'inverse, le domaine est bien présent à l'étranger : dans la liste bibliographique mentionnée ci-dessus, on relève 14 % d'articles parus dans des revues généralistes en médecine de très bon niveau (*New England Journal of Medicine*, *British Medical Journal*, *Lancet* et même *Science*) et 16 % dans les meilleures revues d'épidémiologie (*International Journal of Epidemiology*, *American Journal of Epidemiology*, *American Journal of Public Health*). On trouve, notamment aux États-Unis et en Grande-Bretagne, d'importants laboratoires se consacrant exclusivement à l'épidémiologie sociale : ainsi, le *Department of Epidemiology and Public Health* du *University College* de Londres compte environ 150 personnes.

L'une des raisons de la faiblesse de la situation française pourrait tenir au fait que les données disponibles dans le système statistique sont peu nombreuses. Certes, pour l'étude des *causes de décès*, la France est l'un des rares pays à coder à la fois la cause de décès et le statut social dans les bulletins de décès, ce qui — malgré les limites des informations recueillies dans ces bulletins — est un avantage certain. Mais pour les études de morbidité, on doit constater que — jusqu'ici — le système hospitalier et le système de protection sociale n'ont guère pu être mobilisés. Il faut souhaiter que les nécessités d'une meilleure gestion de ces systèmes conduisent à l'élaboration de données utiles non seulement à la gestion quotidienne des acteurs, mais aussi à la recherche épidémiologique. Une concertation entre gestionnaires et chercheurs sera nécessaire, tant pour définir la nature des variables à prendre en compte que pour organiser les manières de les rendre disponibles aux chercheurs (voir sous-chapitres 3.2 « Le recueil de l'information dans les hôpitaux et 9.3 « Moyens nécessaires à la recherche épidémiologique »).

Au-delà des sources statistiques résultant de fichiers construits spécifiquement ou non, on doit recourir à des enquêtes. Celles-ci doivent reposer sur de vastes échantillons, les processus d'interactions entre variables étant nombreux. Or, s'il existe de nombreux exemples de telles réalisations à l'étranger, ils sont trop rares en France. Ainsi, un récent rapport de l'Insee constate les insuffisances françaises dans le domaine des suivis de longue durée des personnes, contrairement à ce qui a été développé, souvent depuis très longtemps, par les organismes chargés de la statistique publique dans d'autres pays (Chaleix et Lollivier, 2004).

Il faut aussi signaler le faible intérêt porté par la communauté scientifique française des sciences sociales à l'étude de la santé et particulièrement de ses déterminants sociaux. Alors que bon nombre des épidémiologistes étrangers spécialisés sont en fait des sociologues ou des psychosociologues de formation et d'insertion institutionnelle, on ne rencontre pratiquement pas ce type d'épidémiologistes en France. Or, la collaboration interdisciplinaire est particulièrement importante dans un champ de recherche qui se situe à la rencontre de plusieurs domaines. De plus, pratiquement par définition, la recherche en épidémiologie sociale doit être intimement liée aux terrains dans lesquels elle s'effectue, et les modèles importés d'autres sociétés doivent, plus que dans les autres domaines de l'épidémiologie, être mis à l'épreuve de la société française.

Autre source de difficulté : comme il arrive souvent pour une discipline « d'interface », les spécialistes en épidémiologie sociale ont du mal à se faire reconnaître par leurs pairs « purs épidémiologistes » comme « purs sociologues ». La dimension réduite de la discipline ne lui permet évidemment pas de revendiquer des instances spécifiques d'évaluation, mais lorsque pas un seul des membres de l'instance concernée n'est familier avec le champ de recherche, le handicap est certain.

Une conséquence inévitable : la discipline ne peut pas contribuer suffisamment à aider les pouvoirs publics dans la définition de la politique de santé du pays. Parfois, même des informations purement descriptives, mais très importantes pour la définition et la mise en œuvre d'une politique, font défaut. Cela est apparu de façon particulièrement claire lors de l'examen de la loi de santé publique adoptée en juillet 2004 par le Parlement. Cette loi s'appuie sur de nombreux objectifs quantifiés de santé à atteindre : or, pour une proportion très élevée de ces objectifs, les données de base permettant de construire des indicateurs de suivi de l'application des mesures prévues par la loi n'existent simplement pas.

Recommandations

- Intégrer plus systématiquement des variables de situation sociale dans les grandes études épidémiologiques.
- Encourager des projets ambitieux, de grande dimension et de longue durée, à l'instar des meilleurs exemples étrangers.
- Encourager les sociologues français à « s'emparer » de l'épidémiologie sociale comme le font leurs collègues étrangers (l'épidémiologie sociale existe au sein de départements de sociologie, et pas seulement dans ceux de santé publique).
- Inciter les systèmes d'évaluation à mieux prendre en compte les spécificités de la discipline.

Bibliographie

- Acheson DS. (1998). *Independent inquiry into inequalities in Health Report*. The Stationery Office, London.
- Bartley M., Owen C. (1996). Relation between socio-economic status, employment, and health during economic change, 1973-1993. *BMJ*, **313** : 445-453.
- Ben-Schlomo Y., White I., Marmot MG. (1996). Does the variation in the socio-economic characteristics of an area affect mortality? *BMJ*, **312** : 1013-1014.
- Berkman LF., Kawachi I. (eds.) (2000). *Social Epidemiology*. Oxford University Press, New York.
- Bresleau N., Chilcoat H., Susser E., Matte T., Liang K., Peterson E. (2001). Stability and change in children's intelligence quotient scores : a comparison of two socioeconomically disparate communities. *Am J Epidemiol*, **154** : 711-717.

- Bynner J. (2002). British Birth Cohorts Studies : Research Prospects. Paper for the *LASMAS Meeting on Longitudinal Studies*, Paris, 7-8 January 2002.
- Cassel J. (1976). The contribution of the social environment to host resistance. *Am J Epidemiol*, **104** : 107-123.
- Chaleix M., Lollivier S. (2004). *Outils de suivi des trajectoires des personnes en matière sociale et d'emploi*. Insee, n° 98/B010, Paris.
- Cheng Y., Kawachi I., Coakley E., Schwartz J., Colditz GA. (2000). Association between psychosocial work characteristics and health functioning in American women : prospective study. *BMJ*, **320** : 1432-1436.
- Courgeau D., Baccaïni B. (1997). Analyse multiniveaux en sciences sociales. *Population*, **52** (4) : 831-863.
- Cox DR. (1972). Regression models and life tables (with discussion). *J Royal Statist Soc B*, **34** : 187-220.
- Dahl E. (1996). Social mobility and health : cause or effect? *BMJ*, **313** : 435-436.
- Diez Roux AV. (2002). Places, people and health. *Am J Epidemiol*, **155** : 516-519.
- Diez Roux AV. *et al.* (2001). Neighborhood of residence and incidence of coronary heart disease. *New Engl J Med*, **345** : 99-106
- Dray-Spira R., Lert F. (2003). Social health inequalities during the course of chronic HIV disease in the era of highly active antiretroviral therapy. *AIDS*, **17** : 283-290.
- Egger M., May M., Chene G., Phillips AN., Ledergerber B., Dabis F. *et al.* (2002). Prognosis of HIV-1-infected patients starting highly active antiretroviral therapy : a collaborative analysis of prospective studies. *Lancet*, **360** (9327) : 119-129.
- Fuhrer R., Shipley MJ., Chastang JF., Schmaus A., Niedhammer I., Stansfeld SA., Goldberg M., Marmot MG. (2002). Socioeconomic position, health and possible explanations : a tale of two cohorts. *Am J Public Health*, **92** : 1290-1294.
- Goldberg M., Melchior M., Leclerc A., Lert F. (2003). Épidémiologie et déterminants sociaux des inégalités de santé. *Rev Epidemiol Santé Publique*, **51** : 381-401.
- Goldstein H. (1995). *Multilevel Statistical Models*. Edward Arnold, London.

- House J., Landes K., Umberson D. (1988). Social relations and health. *Science*, **241** : 540-544.
- Karasek R., Theorell T. (1990). *Healthy Work : stress, productivity and the reconstruction of working life*. NY : Basic Books, New York.
- Kleinbaum DG. (2002). Epidemiologic methods : the "art" in the state of the art. *J Clin Epidemiol*, **55** : 1196-1200.
- Kogevinas M., Pearce N., Susser M., Boffetta P. (1997). *Social Inequalities and Cancer*. Lyon, International Agency for Research on Cancer. IARC Scientific Publications.
- Kuh D., Ben-Schlomo Y. (eds.) (1997). *A lifecourse approach to chronic disease epidemiology*. Oxford University Press, Oxford.
- Kunst AE, Groenhouf F., Mackenbach JP., Health EW. (1998). Occupational class and cause specific mortality in middle age men in 11 European countries : comparison of population based studies. EU Working Group on Socioeconomic Inequalities in Health. (Comment). *BMJ*, **316** : 1636-1642.
- Leclerc A., Fassin D., Grandjean H., Kaminski M., Lang T. (eds.) (2000). *Les inégalités sociales en santé*. La Découverte, Paris.
- Lee M., Rotheram-Borus MJ. (2001). Challenges associated with increased survival among parents living with HIV. *Am J Public Health*, **91** (8) : 1303-1309.
- Leon D., Walt G. (eds.) (2000). *Poverty, inequality and health*. University Press, Oxford.
- Link BG., Phelan J. (1996). Understanding sociodemographic differences in health. The role of fundamental social causes. *Am J Public Health*, **86** (4) : 471-473.
- Lynch JW., Davey Smith G., Kaplan GA., House JS. (2000). Income inequality and mortality : importance to health of individual income, psychosocial environment, or material conditions. *BMJ*, **320** : 1200-1204.
- Mackenbach JP., Bakker MJ. (2002). Reducing inequalities in health. A *European perspective*. Routledge, London, New York.
- Marmot MG. (2001). Inequalities in health. *New Engl J Med*, **345** : 134-136.
- Marmot MG., Bosma H., Hemingway H., Brunner E., Sa S. (1997). Contribution of job control and other risk factors to social variations in coronary heart disease incidence. *Lancet*, **350** : 235-239.
- Marmot MG., Fuhrer R., Ettner SL., Marks NF., Bumpass LL., Ryff CD. (1998). Contribution of psychosocial factors to socio-economic differences in health. *The Milbank Quarterly*, **76** (3) : 403-448.

- Marmot MG., Kogevinas M., Elston MA. (1987). Social/economic status and disease. *Annu Rev Public Health*, **8** : 111-135.
- Marmot MG., Shipley MJ., Rose G. (1984). Inequalities in death : specific explanations of a general pattern ? *Lancet*, **1** (8384) : 1003-1006.
- Mathews JD., Glasse R., Lindenbaum S. (1968). Kuru and cannibalism. *Lancet*, **2** (7565) : 449-452.
- Mesrine A. (1999). Les différences de mortalité par milieu social restent fortes. In : *Données sociales 1999*, 228-235. Insee, Paris.
- Niedhammer I., Goldberg M., Leclerc A., Bugel I., David S. (1998). Psychosocial factors at work and subsequent depressive symptoms in the Gazel cohort. *Scand J Work Environ Health*, **24** : 197-205.
- Poirier S., Ohshima H., de-The G., Hubert A., Bourgade MC., Bartsch H. (1987). Volatile nitrosamine levels in common foods from Tunisia, south China and Greenland, high-risk areas for nasopharyngeal carcinoma (NPC). *Int J Cancer*, **39** : 293-296.
- Poulton R., Caspi A., Milne BJ., Thomson WM., Taylor A., Sears M., Moffit TE. (2002). Association between children's experience of socioeconomic disadvantage and adult health : a life-course study. *Lancet*, **360** : 1640-1645.
- Rapiti E., Porta D., Forastiere F., Fusco D., Perucci CA. (2000). Socioeconomic status and survival of persons with AIDS before and after the introduction of highly active antiretroviral therapy. Lazio AIDS Surveillance Collaborative Group. *Epidemiology*, **11** (5) : 496-501.
- Ravaud JF., Mormiche P. (2000). Handicaps et accidents. In : *Les inégalités sociales de santé*, A. Leclerc, D. Fassin, H. Grandjean, M. Kaminski, T. Lang (eds.) La Découverte, Paris.
- Rose G. (1985). Sick individuals and sick populations. *Int J Epidemiol*, **30** (4) : 809-817.
- Salem G., Rican S., Jouglu E. (1999). *Atlas de la santé en France. Volume 1 : Les causes de décès*. John Libbey Eurotext, Paris.
- Von Korff M., Koepsell T., Curry S., Diehr P. (1992). Multilevel analysis in epidemiologic research on health behaviors and outcomes. *Am J Epidemiol*, **135** : 1077-1082.
- Wilkinson RG. (1996). *Unhealthy societies. The afflictions of inequality*. Routledge, London.

CHAPITRE 7

Épidémiologie et causalité

ANNE FAGOT-LARGEAULT

La notion de causalité en épidémiologie est probabiliste. Elle tient à la notion de dépendance/indépendance des événements, définie par Laplace au début de son *Essai philosophique sur les probabilités* : « La probabilité . . . est le rapport du nombre des cas favorables à celui de tous les cas possibles. [. . .]. Si les événements sont indépendants les uns des autres, la probabilité de l'existence de leur ensemble est le produit de leurs probabilités particulières. [. . .]. Quand deux événements **dépendent l'un de l'autre**, la probabilité de l'événement composé est le produit de la probabilité du premier événement, par la probabilité que cet événement étant arrivé l'autre arrivera » (Laplace, 1825).

L'idée d'une « **recherche des causes par les probabilités** » remonte au XVIII^e siècle, elle fut systématisée par Laplace (1774, 1778). Elle ne signifie pas qu'il existe une définition purement mathématique de la relation causale, mais seulement qu'il existe des critères du lien causal formulables en termes de théorie des probabilités. Ces critères formels ne sont pas des critères suffisants de causalité, mais ce sont des critères puissants lorsqu'ils peuvent être appliqués. Leur application ne peut se faire qu'en admettant des présupposés d'arrière-plan. D'une part, lorsque Laplace parle d'événements, il ne s'agit pas de faits singuliers, mais de classes de faits (événements répétables, génériques). L'analyse probabiliste ne s'applique pas à des faits historiquement avérés, mais non répétables. D'autre part, pour détecter des dépendances entre événements, il faut admettre que dans l'Univers **tout n'est pas lié à tout** : il n'y a de dépendances causales intéressantes que s'il y a des événements indépendants les uns des autres. Celui qui, tel le brachmane de Voltaire, professe que le moindre petit fait est « lié à tous les événements de l'Univers, passés, présents et futurs », s'interdit l'investigation de liens causals significatifs.

Dans la perspective probabiliste, il y a **présomption de causalité** chaque fois que, de deux événements, l'occurrence de l'un modifie la probabilité d'occurrence de l'autre. Cette notion a été formalisée par Suppes (1970) :

« L'événement C_t est une cause présomptive de l'événement E_t si, et seulement si :

$$(1) \quad t' < t, \quad (2) \quad P(C_{t'}) > 0 \quad \text{et} \quad (3) \quad P(E_t|C_{t'}) > P(E_t). \quad \text{»}$$

Familièrement parlant, concevoir un facteur causal comme un facteur qui augmente la probabilité d'occurrence d'un effet, c'est rendre compte de l'intuition qu'une cause est efficace. Dire qu'elle est efficace n'est pas dire qu'elle produit l'effet infailliblement. C'est dire qu'elle favorise l'apparition de l'effet, ou qu'elle en augmente les chances. Les causes, au sens des épidémiologistes, sont des « facteurs de risque ». Cette terminologie est récente. À la suite de Laplace, les mathématiciens français du XIX^e siècle employaient le mot « influence » : un

facteur est causal s'il a une **influence** sur l'apparition de l'effet. L'influence se mesure à l'**écart** entre la probabilité (conditionnelle) de l'événement en présence de la cause, et la probabilité (simple) de l'événement (toutes causes confondues – Laplace appelait « possibilité » de l'événement cette probabilité simple).

D'où viennent les **hypothèses causales**? Le travail de l'épidémiologiste de terrain a été comparé à celui du **détective** qui identifie un suspect et le suit à la trace. Edward Jenner rapporte comment la vérole des vaches « progresse du cheval au pis de la vache et de la vache au sujet humain ». John Snow observe que la propagation du choléra suit la route des navires, qu'il se transmet par voie alimentaire. Ignaz Semmelweis soupçonne que l'agent de la fièvre puerpérale passe de la salle d'autopsie à la maternité par les mains des internes. Carlos Chagas s'intéresse à un insecte hématophage vecteur d'un trypanosome qu'il injecte par les déjections qui souillent sa piquûre (Buck *et al.*, 1988). La différence entre épidémiologiste et détective est que le policier-détective tente de débrouiller une **histoire singulière**, tandis que l'épidémiologiste vise à reconstituer les étapes d'une **histoire générique**. Mirko Grmek (1995) a loué ces « détectives médicaux » qui ont dessiné les « filières du sida », et brossé le tableau de l'épidémie, avec ses modalités typiques de transmission (par la sexualité, par les seringues, de la mère à l'enfant) et ses facteurs aggravants, comme la « brèche transfusionnelle » qui transforma « un chemin étroit en voie royale ». Les scénarios caractéristiques du sida ne se confondent, ni avec les histoires individuelles des personnes atteintes, ni avec l'histoire globale de la pandémie. Ces dernières sont uniques. Les scénarios sont généraux. C'est ce qui permet l'analyse statistique. Notons aussi que l'analyse statistique va introduire de la **discontinuité** dans le scénario, sélectionnant des « événements » (c'est-à-dire des **classes** d'événements, exemple : séropositivité de la mère/séropositivité de l'enfant) entre lesquels un lien de probabilité sera établi.

Dans les sciences sociales, on a le plus souvent des scénarios causals une approche **théorique** (*top-down*). On modélise la situation, on trace un diagramme avec des flèches indiquant le sens des liens de causalité, puis on regarde jusqu'à quel point le diagramme « colle » avec les faits connus. Dans les sciences biomédicales, l'approche est plus souvent empirique, **inductive** (*bottom-up*). Le père de l'épidémiologie clinique, Pierre C.A. Louis, enseignait à ses étudiants à rédiger leurs observations, de manière à pouvoir les comparer, afin d'en abstraire le schéma évolutif : « marche de la maladie », ou « histoire naturelle de la maladie ». Ces schémas suggèrent des liens de causalité. La médecine, à vrai dire, surabonde en hypothèses étiologiques, et le rôle de l'épidémiologiste-statisticien sera plus souvent de soumettre des conjectures causales existantes à l'examen critique que d'inventer des hypothèses causales neuves. Certes, Cournot (1843) envisageait la possibilité pour un statisticien de faire sur un ensemble de données des « coupes » systématiques dans l'espoir de détecter une « anomalie » (différence significative) restée inaperçue jusque-là, c'est-à-dire, liée à une partition

à laquelle on n'avait pas songé. Mais alors la « théorie des hasards » prévoit qu'il finira par tomber sur une différence fortuite. En pratique, les hypothèses causales intéressantes sont le plus souvent suggérées par l'observation biologique et clinique. La méthodologie statistique est moins utile pour **découvrir** des liens causals inédits que pour **tester** la solidité de liens causals présumés.

Conscient de la **variabilité** des phénomènes biologiques, Laplace se méfiait de l'impression qu'une influence s'exerce. On savait que dans l'espèce humaine le rapport des sexes à la naissance est d'environ 18 garçons pour 17 filles. Laplace observa qu'à Londres le rapport était durablement plus élevé qu'à Paris, cette « plus grande facilité » des naissances masculines suggérant qu'il pouvait y avoir un facteur londonien (climatique ?) favorisant les naissances mâles. Fallait-il chercher une « cause spéciale » de l'écart observé ? Sur une plus longue période d'observation l'écart s'effaça. « C'est ici surtout qu'il est nécessaire d'avoir une méthode rigoureuse pour distinguer, parmi les phénomènes observés, ceux qui peuvent dépendre du **hasard**, de ceux qui dépendent de causes particulières » (Laplace, 1778). Doll et Peto (1981), qui ont soigneusement examiné les pièces du dossier tabac, estiment que, jusqu'en 1950, on pouvait douter du lien entre la fumée du tabac et le cancer bronchopulmonaire (CBP). Les habitudes tabagiques étaient si répandues que l'impression de certains pneumologues ou chirurgiens (que le CBP était plus fréquent chez les gros fumeurs que dans la population générale) demandait à être contrôlée. Et l'incidence apparemment croissante du CBP, dont on avait tenté de montrer qu'elle suivait la courbe des ventes de cigarettes, pouvait être un artefact tenant à l'amélioration des diagnostics. Mais après 1950 les données, de fiabilité croissante, de la certification des causes de décès, sont corroborées par celles des registres de cancer : l'épidémie de CBP est réelle. Les premières enquêtes rétrospectives établissent que les malades ont, en moyenne, beaucoup plus fumé que les témoins. Et les enquêtes prospectives confirment que le risque de CBP augmente avec la quantité fumée, la durée d'exposition au tabac, le fait d'inhaler la fumée, etc. Il y a un lien. Ce lien est-il causal ?

Présomption n'est pas preuve. On sait que des statisticiens éminents, comme Ronald Fisher et Jacob Yerushalmy, résistèrent à l'idée que la fumée du tabac cause le cancer. Ils concédaient qu'un lien statistique était établi. Un lien statistique signale un lien causal, il n'est pas un lien causal. La fumée du tabac, plutôt que la cause du cancer (le facteur qui induit l'augmentation du risque), pouvait être un simple marqueur (un témoin de l'augmentation du risque). Les fumeurs pouvaient posséder un trait X (génétique, par exemple) les distinguant des non-fumeurs, et rendant compte à la fois de leur penchant tabagique et de leur susceptibilité au CBP. Quels moyens l'épidémiologiste a-t-il de distinguer entre **facteurs** et **marqueurs** de l'augmentation du risque ?

Observons d'abord que les outils mathématiques que nous utilisons pour l'analyse du monde réel induisent habituellement de la réversibilité. Piaget disait que la réversibilité est une conquête de la pensée. Les associations statistiques sont symétriques (si A est corrélé à B, B est corrélé à A). Laplace (1774) a montré que le raisonnement probabiliste qui passe par l'équation de Bayes marche aussi bien pour remonter de l'effet à la cause (diagnostic étiologique) que pour pronostiquer l'effet à partir de la cause. Or, les relations causales sont **asymétriques**. On peut « remonter » les chemins causals par la pensée, mais les processus causals « descendent » le temps. Ce présupposé de la recherche causale est souvent implicite, mais il est fondamental : **l'effet ne précède pas la cause** (*after cannot cause before*). Laplace le suggère discrètement en parlant d'un « premier » événement dont dépend un autre qui « arrivera » (cf. ci-dessus). Suppes le marque clairement dans sa définition : si C ne précède pas E dans le temps, il ne peut pas en être une cause présomptive. Ce critère chronologique est étranger à la théorie des probabilités. Il tient à ce que les scénarios causals, tout génériques qu'ils soient, gardent la non-réversibilité des processus historiques dont ils sont extraits. On ne peut pas imputer à un médicament C un effet nocif E s'il s'avère que l'effet E s'est déclaré avant la prise du médicament. Certes, le critère chronologique ne saurait servir à prouver positivement la causalité (sous peine de commettre la faute de raisonnement : *post hoc, ergo propter hoc*). Le fait que E a suivi la prise du médicament ne suffit pas à établir la responsabilité (causale) du médicament. Mais parmi les critères d'imputabilité, le **critère chronologique** sert à éliminer les liens absurdes. C'est pourquoi certains ont vu dans la chronologie une « logique » causale (Davis, 1985). C'est même la seule logique qu'on puisse trouver à la causalité. Les relations causales n'ont pas d'autre « bonne » propriété logique ; en particulier, elles ne sont pas transitives. De ce que A cause B, et B cause C, il ne suit pas que A cause C.

Lorsqu'un lien causal présomptif est « compatible avec la chronologie », s'assurer de sa réalité est rarement une tâche simple. Le travail de Doll et Peto sur les causes du cancer donne un exemple remarquable de la méthodologie qui, dans le cas du CBP, a conduit de la présomption à la preuve — sinon à la preuve absolue, du moins à la conviction que la **convergence des arguments** rassemblés fait qu'il est devenu déraisonnable de continuer à douter de l'action causale du tabac : « *The conclusion that cigarette smoke is a direct cause of cancer derives from many different types of epidemiological evidence, combined with the fact that the smoke can cause cancer in experimental animals* ». Parmi les arguments avancés :

- les **données** factuelles disponibles à partir de 1950 sont nombreuses et fiables (on ne peut pas attribuer le lien statistique à des sources biaisées) ;
- le « tiers facteur » X qui eût pu être la « cause commune » du CBP et de la propension à fumer n'a pas été identifié, son identification est devenue

improbable (même s'il est impossible de prouver *a priori* sa non-existence), et **aucune autre hypothèse** n'explique les faits observés mieux que celle de l'action causale du tabac ;

- l'efficacité du tabac à induire le CBP a été évaluée par des **mesures** de la **force du lien** causal, en termes de « risque relatif » (RR : nombre par lequel la présence de C multiplie la probabilité d'occurrence de E), ou en termes de proportion des cas de E qui seraient évités en l'absence de C (Doll et Peto estiment que 9 CBP sur 10 seraient évités si personne ne fumait « fraction évitable du risque ») ;
- certes, l'essai contrôlé randomisé qui eût fourni le plus haut niveau de preuve n'a pas été réalisé chez l'homme, pour d'évidentes raisons éthiques, mais on dispose de données **expérimentales** sur l'animal, et sur l'homme de données d'observation simples (mortalité), ou comparatives (enquêtes cas-témoins, suivi de cohortes).

L'histoire de l'épidémiologie montre que, sur les **critères de causalité**, on a beaucoup tâtonné (cf. Hill, Evans, Lilienfeld). Les manuels donnent des listes de critères (exemple : Jenicek et Cléroux). La méthodologie adaptée à chaque situation, et le niveau de preuve à atteindre pour décider qu'un lien est causal, sont affaire de jugement. La maladie coronarienne est multifactorielle. Nombreux sont les facteurs qui ont été soupçonnés d'y jouer un rôle : hypertension artérielle, tabagisme, obésité, cholestérol, diabète, sédentarité, caractère colérique, etc. À première vue, ces facteurs ne sont pas indépendants les uns des autres, la fraction du risque attribuable à chacun est petite, la preuve de leur influence nécessite des études épidémiologiques d'une envergure formidable, longues et coûteuses. En 1984 un groupe de travail américain conclut qu'il existait désormais assez d'**indices concordants** (génétiques, épidémiologiques, cliniques, expérimentaux) pour qu'on puisse affirmer « *beyond reasonable doubt* » que l'hypercholestérolémie n'est pas seulement un marqueur du risque coronarien, mais une cause de maladie coronarienne (*JAMA-France*, 1985, **10**, 104 : 819). C'était une décision importante : elle menait à considérer la réduction des taux de cholestérol sanguin comme un objectif de santé publique. Sur quels arguments reposait-elle ? D'une part, des essais d'intervention bien contrôlés avaient montré que la réduction des taux de cholestérol sanguin réduisait efficacement à elle seule la mortalité coronarienne, et on avait calculé de combien. D'autre part, la recherche physiologique commençait à élucider le mécanisme par lequel une surcharge en cholestérol lèse les artères et finit par les boucher. Plausibilité biologique et mesure de l'efficacité convergeaient pour emporter la décision.

Qu'ajoute à l'évidence statistique la connaissance du **mécanisme** par lequel la cause produit l'effet ? Wesley Salmon, qui a beaucoup médité sur le concept statistique de causalité, juge ce concept insuffisant, parce qu'il fait voir les liens

causals comme des relations entre événements discrets, négligeant les « processus continus qui transmettent l'influence causale d'une région à l'autre de l'espace-temps ». Découvrir le mécanisme de l'action causale permettrait de rétablir la continuité perdue, en interpolant entre des événements C et E qu'on avait arbitrairement disjoints le fil d'une histoire. Les chercheurs de laboratoire tendent parfois à penser que les épidémiologistes ont besoin du calcul parce qu'ils raisonnent sur des données grossières, prélevées à un niveau superficiel, tandis que celui qui se livre à une observation fine n'a pas besoin de compter. Mais la mise en évidence sous le microscope de fibres d'amiante fichées dans une plèvre atteinte de mésothéliome, si elle suggère un mécanisme, ne suffit aucunement à prouver que l'amiante cause ce cancer. À tous les niveaux de la réalité il existe de multiples phénomènes contigus et causalement indépendants. Et la « nécessité » du lien de C à E ne sort pas renforcée de ce qu'on a individualisé les étapes d'une chaîne causale intermédiaire (ou de plusieurs, s'il y a plusieurs routes possibles) : vu la variabilité biologique, à chaque maillon de la chaîne le processus causal peut bifurquer, et l'impression de contingence s'en trouver accrue. Savoir raconter comment s'opère le passage de C à E satisfait à une exigence touchant la continuité de la propagation des influences causales, mais cela ne dispense pas d'établir par le calcul la réalité de l'influence.

Bibliographie

- Buck C., Llopis A., Najera E., Terris M. (eds). (1988). *The Challenge of Epidemiology. Issues and Selected Readings*, Washington D.C. : PAHO-WHO, Scientific Publication n° 505.
- Cournot MAA. (1843). *Exposition de la théorie des chances et des probabilités*, Hachette, Paris.
- Davis JA. (1985). *The Logic of Causal Order*, Sage, London.
- Doll R., Peto R. (1981). *The Causes of Cancer. Quantitative Estimates of Avoidable Risks of Cancer in the United States Today*, Oxford University Press.
- Evans AS. (1978). Causation and disease. A chronological journey. *Am J Epidemiol*, **108** : 249-258.
- Fagot-Largeault A. (1986), Approche médicale de la causalité dans les systèmes complexes, *Archives internationales de physiologie et de biochimie* (Liège), **94** (4) : C85-C94. (1989), *Les causes de la mort. Histoire naturelle et facteurs de risque*, Vrin, Paris. (2003), Preuve et niveau de preuve dans les sciences biomédicales, In : *La vérité dans les sciences*, Changeux, dir., Odile Jacob, Paris, 215-236.
- Grmek M. (1989). *Histoire du sida. Début et origine d'une pandémie actuelle*, Payot, Paris ; 2^e éd. revue et augt., 1995.

- Hill AB. (1965). The environment and disease : association or causation ? *Proc Roy Soc Med*, **58** : 295-300.
- Jenicek M., Cléroux R. (1982), *Épidémiologie : principes, techniques, applications*, (chap. 7). (1985), *Épidémiologie clinique* (chap. 2). Edisem, St Hyacinthe et Maloine, Paris.
- Laplace PS. de, *Œuvres complètes*, publiées sous les auspices de l'Académie des sciences, Gauthiers-Villars, Paris, 14 vol., 1878-1912 (posth.). Includ : *Mémoire sur la probabilité des causes par les événements* (1774), tome VIII, 27-65 ; *Mémoire sur les probabilités* (1778), tome IX, 383-485 ; *Théorie analytique des probabilités* (1812) et *Essai philosophique sur les probabilités* (1814-1825), tome VII.
- Lilienfeld AM., Lilienfeld DE. (1980). *Foundations of Epidemiology*, Oxford University Press (chap. 12).
- Pearl J. (2000). *Causality. Models, Reasoning and Inference*, Cambridge University Press.
- Salmon WC. (1980). Probabilistic causality, *Pacific Philosophical Quarterly*, **61** : 50-74. (1984), *Scientific Explanation and the Causal Structure of the World*, Princeton University Press.
- Suppes P. (1970). A probabilistic theory of causality. *Acta Philosophica Fennica*, Fasc. XXIV, North Holland, Amsterdam.

CHAPITRE 8

L'épidémiologique : la base
scientifique de la santé
publique

SOUS-CHAPITRE 8.1

L'épidémiologie clinique, base scientifique pour la prise en charge rationnelle des malades

PHILIPPE RAVAUD

1 | Quelques données historiques

Pierre Charles Alexandre Louis (1787-1872), considéré comme le fondateur de l'épidémiologie clinique, développa vers 1830 une méthode standardisée de recueil et d'analyse des données révolutionnaire pour l'époque, appelée méthode numérique. Son apport clinique le plus connu concerne la remise en cause de la sacro-sainte saignée dans les maladies inflammatoires. Il fit école en France, mais également en Suisse (Semmelweis), en Angleterre (Farr), en Amérique où ses élèves (Holmes, Shattuck, Bowditch) ont créé l'école américaine de médecine clinique (Morabia, 1996).

Si le terme épidémiologie clinique a été utilisé pour la première fois par Paul en 1938 (Paul, 1938), les débuts de l'épidémiologie clinique moderne datent

du milieu des années 1960. Alvan Feinstein et David Sackett, confrontant leurs conceptions, développèrent cette nouvelle discipline (Fletcher, 2001). La première unité de recherche en épidémiologie clinique fut créée en 1966 à New York (State University of New York, Buffalo, États-Unis), suivie par le département d'Épidémiologie clinique et de biostatistique de la Mac Master University au Canada en 1967. En 1967 et 1968, Sackett et Feinstein publièrent séparément des articles précisant la définition et le champ d'application de l'épidémiologie clinique (Feinstein, 1968a, b, c ; Sackett, 1969).

Des enseignements d'épidémiologie clinique débutèrent à Yale et à Mac Master en 1970. Un programme d'enseignement conjoint de la médecine clinique et de l'épidémiologie clinique, créé en 1974 (Robert Woods Johnson Clinical Scholars Programs), donna un essor important à l'épidémiologie clinique aux États-Unis. Plusieurs ouvrages majeurs en épidémiologie clinique furent publiés dans les années 1980, successivement par Fletcher (University of North Carolina) en 1982, par Sackett (University of Mac Master ; Sackett *et al.*, 1985) et Feinstein (Yale University) en 1985 (Feinstein, 1985).

En 1988, le *Journal of Chronic Diseases* fut renommé *Journal of Clinical Epidemiology*.

2 | Définition de l'épidémiologie clinique

Contrairement à l'épidémiologie classique, qui étudie et interprète la maladie en tant que phénomène de groupe sur la base de multiples données individuelles, l'épidémiologie clinique utilise les informations épidémiologiques recueillies auprès des groupes de malades pour une meilleure prise de décision clinique face à un malade donné. Ainsi selon Feinstein, « l'épidémiologie clinique est une extension de l'épidémiologie classique (traditionnellement orientée vers les stratégies en santé publique et communautaire), c'est un champ d'activités qui recouvre l'ensemble des décisions cliniques telles qu'elles se présentent lors de la prise en charge de l'individu (patient) tant en ce qui concerne le diagnostic, le pronostic et les mesures thérapeutiques que d'autres jugements médicaux » (Feinstein, 1985).

Sackett définit l'épidémiologie clinique comme « l'application par un médecin directement en charge de malades, de méthodes épidémiologiques et biométriques à l'étude des processus diagnostiques et thérapeutiques dans le but d'améliorer la santé ».

Le champ d'application de l'épidémiologie clinique concerne donc particulièrement l'évaluation de la tolérance et de l'efficacité des traitements, de la valeur diagnostique des tests, l'établissement du pronostic chez un individu. On utilise pour cela les méthodes de l'épidémiologie classique mais également des méthodes spécifiques.

Certains considèrent que l'épidémiologie clinique constitue la science de base de la recherche clinique (Gail, 2000) et de la médecine clinique (Straus, 2004).

3 | Apport de l'épidémiologie clinique

Comme l'a souligné Sackett, le développement majeur et la force de l'épidémiologie clinique sont essentiellement liés à son rôle central et prépondérant dans cinq évolutions récentes et parallèles considérées pour certaines d'entre elles comme des révolutions dans le domaine de la santé : la production des données probantes (*evidence generation*), l'évaluation de ces données (*evidence appraisal*), la recherche de ces données (*evidence retrieval*), la synthèse de ces données (*evidence synthesis*) et enfin l'utilisation de ces données (*evidence application*). De nombreuses autres disciplines (médecine clinique, biostatistique, épidémiologie et sciences sociales, en particulier) ont aussi contribué à ces progrès (Sackett, 2002).

Ces évolutions récentes ont été favorisées et/ou rendues indispensables par l'explosion de la littérature médicale (plus de 10 000 revues biomédicales et 12 millions d'articles indexés dans la base de données Medline). Cette masse d'informations sans cesse renouvelée explique les difficultés constatées de transposition des résultats de la recherche à la pratique clinique et le fossé existant parfois entre les pratiques des médecins et les données scientifiques disponibles.

Parmi les multiples apports de l'épidémiologie clinique, le concept d'*Evidence-Based Medicine*, ou médecine fondée sur les faits prouvés, et la collaboration Cochrane sont parmi les plus marquants.

Le développement de l'analyse critique de la littérature médicale, sous l'impulsion de Haynes et Sackett (Mac Master University), a constitué une étape importante dans les années précédant l'apparition du concept d'*Evidence-Based Medicine*. Ce terme a été introduit par Guyatt, qui, avec l'aide d'un groupe de médecins et d'épidémiologistes cliniciens, a proposé en 1992 une nouvelle forme d'enseignement de la médecine basée sur un auto-enseignement actif sous forme de résolution de problèmes cliniques (*Evidence-based medicine working group*, 1992). L'EBM était définie comme une formation à l'utilisation

consciente et judicieuse des meilleures données actuelles de la recherche clinique, disponibles à partir d'une recherche systématique dans la littérature pour soigner un malade donné. L'EBM est maintenant devenue une partie intégrante de l'enseignement initial de la médecine et de la formation médicale continue (réf 1-4, Edito Straus BMJ 30/10/2004). Progressivement, ce qui n'était initialement qu'une forme nouvelle d'enseignement est devenu également une forme nouvelle d'exercice de la médecine correspondant si étroitement aux aspirations conjointes des malades, des médecins et des décideurs qu'elle s'est imposée en moins de 10 ans comme la seule voie réaliste de pratiquer la médecine en assurant aux malades des soins de qualité et en aidant les médecins à se préserver des pressions multiples (administration, industrie pharmaceutique...) dont ils sont l'objet. Dans cette approche, on préconise que le médecin utilise un modèle de décision clinique à trois composants prenant en compte son expérience clinique, les préférences du malade et les données actuelles de la recherche. Il semble illusoire d'espérer que les médecins praticiens réalisent une bibliographie et analysent de manière critique les articles originaux; *a contrario*, l'utilisation d'informations EBM déjà triées (par exemple à partir de journaux de synthèse tel l'*EBM Journal* (Davidoff, 1995) ou l'*ACP Journal Club*, d'ouvrages ou de sites spécialisés) semble plus réaliste. Ainsi, un ouvrage tel *Clinical Evidence*, conçu à la demande du National Health Service anglais comme un recueil des preuves disponibles, régulièrement remis à jour, est distribué gratuitement à tous les médecins anglais et italiens (financement gouvernemental) et aux médecins américains (financement par une fondation : United Health Foundation). Ces documents sont disponibles sous forme papier (mise à jour tous les six mois), Internet (mise à jour tous les mois) et chargeable sur un assistant digital personnel. Des traductions japonaise, russe, allemande, française existent pour une version courte de ce document (Clinical Evidence, 2003). Des données *evidence-based* sont également accessibles aux malades. Des sites tel www.besttreatments.org, développé par le *British Medical Journal*, ont pour but de les aider avec leur médecin à décider quel sera le traitement qui leur sera le mieux adapté (Nash, 2003). Le concept de l'EBM s'est étendu également à la décision en santé publique avec l'*Evidence-Based Public Health* (Jenicek, 1997 ; Woolf, 1999).

Les mêmes besoins entraînent la création de la Cochrane collaboration en Angleterre. Archie Cochrane, épidémiologiste écossais, a argumenté dès 1972 que les essais cliniques randomisés pouvaient seuls séparer les interventions efficaces de celles qui étaient dangereuses ou inutiles. Quelques années plus tard, il écrivit « qu'une grande faiblesse de notre profession était de ne pas avoir organisé une synthèse critique, par spécialité ou sous-spécialité, revue périodiquement, de tous les essais contrôlés randomisés ». Inspiré par les idées de Cochrane, Iain Chalmers créa en 1974 le premier registre d'essais randomisés en médecine périnatale regroupant initialement plus de 3 500 essais publiés, puis rapidement des revues systématiques dans ce domaine (Grant, 1981 ;

Chalmers, 1986). En 1992, le NHS finança la création d'un « centre Cochrane » à Oxford pour faciliter la préparation des revues systématiques d'essais randomisés dans le domaine de la santé. En 1993, lors d'un colloque de la New York Academy of Sciences, Iain Chalmers et un groupe de 76 autres personnes co-fondèrent la Cochrane collaboration (Chalmers, 1993). Rapidement l'idée se développa et se diffusa dans le monde avec la création avant la fin de l'année 1994 de six autres centres Cochrane en Europe, Amérique du Nord et Australie. L'objectif annoncé de la Cochrane collaboration était d'aider médecins, malades et décideurs à prendre des décisions bien informées dans le domaine des soins, en préparant, en mettant à jour, et en favorisant l'accessibilité de revues systématiques évaluant l'effet des interventions dans le domaine de la santé. Ceci est bien sûr d'un intérêt évident pour les médecins praticiens mais également pour les chercheurs qui, identifiant plus facilement les domaines où les connaissances sont incomplètes, peuvent orienter leurs recherches (Straus, 2004). L'importance de ce travail qui fut comparée à celle du projet d'identification du génome humain (Naylor, 1995), constitue selon Sackett le principal apport de l'épidémiologie clinique (Sackett, 2002).

Dans la dernière version disponible de la Cochrane Library (www.cochrane.org) (2004, Issue 3), on retrouve 2074 revues systématiques couvrant de multiples champs de la médecine pour les pays développés ou en voie de développement. Les revues systématiques produites par la Cochrane collaboration, réalisées par plus de 7000 volontaires, sont régulièrement mises à jour. Selon des estimations basées sur le rythme de publications des essais randomisés et de réalisation des revues systématiques, c'est seulement en 2015 que les 10000 revues Cochrane nécessaires pour synthétiser les données disponibles dans les différents domaines de la médecine seront réalisées (Mallett, 2003). La Cochrane Library comprend également un registre des essais publiés (Cochrane Central Register of Controlled Trials) regroupant plus de 415000 essais randomisés, soit 25 % de plus que dans la base de données Medline.

La Cochrane library est accessible gratuitement pour tous les médecins (financement gouvernemental) dans de nombreux pays : Australie, Danemark, Royaume-Uni, Finlande, Irlande, Norvège, Afrique du Sud, Espagne ainsi que par l'intermédiaire du réseau HINARI (Health International Access to Research Initiative) pour les pays classés par la Banque mondiale comme de revenus faibles ou moyens (Grimshaw, 2004). Par ailleurs, il existe plus de 5000 abonnements institutionnels dans le monde et plus de 2,5 millions d'utilisateurs individuels.

Aujourd'hui, la Cochrane collaboration (organisme à but non lucratif) est confrontée à un problème majeur de financement. Bien que les gouvernements australien, danois, anglais, irlandais, norvégien, sud-africain et belge financent

des centres Cochrane, de nombreux centres, y compris parmi les plus actifs, tel le centre canadien, ne disposent pas de financement pérenne.

Les épidémiologistes cliniciens ont eu ainsi un rôle important, non seulement dans la production de connaissances mais aussi dans les développements méthodologiques rendus nécessaires par cette production. Ils ont été ainsi largement impliqués dans la création de standards de publication, devenus incontournables dans les domaines des essais thérapeutiques (CONSORT statement ; Moher, 2001), des études diagnostiques (initiative STARD ; Bossuyt, 2003), des méta-analyses (QUOROM statement ; Moher, 1999), dans la modification de présentation des articles de recherche clinique (introduction des résumés structurés ; Ad hoc Working Group for Critical Appraisal of the Medical Literature, 1987), dans l'amélioration de l'expression des résultats des essais thérapeutiques (*Number Needed to Treat* ou définition de seuils de progression cliniquement pertinents permettant de « dichotomiser » des données quantitatives ; Laupacis, 1988), dans les méthodes de recherche et de synthèse de la littérature (notion de revue systématique, méthodologie des méta-analyses, création et validation d'échelles de qualité des essais, hiérarchisation des niveaux de preuve ; Cook *et al.*, 1997 ; Haynes, 1994 ; Mulrow, 1987 ; Mulrow et Cook, 1998).

La nécessité d'un enregistrement systématique des essais cliniques en cours peut également être mise au crédit de l'épidémiologie clinique. Trente ans après que cette idée ait été proposée par Thomas Chalmers (1977), alors que l'on sait que 50 % des études présentées dans des congrès scientifiques ne sont jamais publiées, et que les études non publiées diffèrent bien évidemment des études publiées (biais de publication en faveur des études positives qui sont publiées plus souvent et plus précocement ; Dickersin, 2002), il n'existe toujours pas de registre international exhaustif des essais en cours ou réalisés. Récemment, l'International Committee of Medical Journal Editors a décidé d'imposer comme une condition de publication que les essais soient enregistrés dans un registre public d'essais cliniques avant que ne débute l'inclusion des malades dans un essai (De Angelis, 2004). Cette mesure s'appliquera à tout essai clinique débutant son recrutement après le 1^{er} juillet 2005.

Enfin les épidémiologistes cliniciens ont eu plus généralement une influence importante dans les programmes visant à répondre à l'enjeu majeur, auquel nous sommes confrontés, de translation dans la pratique médicale quotidienne des données de la recherche clinique (Lenfant, 2003). Plusieurs programmes gouvernementaux aux États-Unis (Programme TRIP pour Translating Research Into Practice de l'Agency for Health Care Research and Quality) ou au Royaume-Uni (Programme « Our Healthier Nation », saving 300 000 lives) (<http://www.archive.official-documents.co.uk>) visent ainsi à améliorer l'efficacité de notre système de soins en favorisant ou en accélérant la transposition des résultats de la recherche dans la pratique clinique.

4 | Le développement et l'enseignement de l'épidémiologie clinique

Un élément important dans l'internationalisation de l'épidémiologie clinique a été la création en 1980, avec l'aide de la fondation Rockefeller, d'un réseau appelé INCLEN (International Clinical Epidemiology Network <http://www.inclen.org>). Ce réseau, devenu une organisation indépendante à but non lucratif, a pour objectif d'aider les cliniciens des pays en voie de développement à créer des unités d'épidémiologie clinique, destinées à promouvoir la recherche et la formation dans cette discipline. Le principe fondateur est que, pour combattre efficacement les maladies dans ces pays (comme ailleurs) dans un contexte de limitation des ressources financières, il est indispensable d'intégrer les principes de l'épidémiologie clinique à la prise de décision en santé. Les jeunes chercheurs sont formés soit dans des unités de pays développés, soit sur site. Ce réseau regroupe actuellement 64 unités d'épidémiologie clinique et 19 centres de recherche clinique et de formation dans 26 pays (www.inclen.org) en Amérique latine, Afrique, Inde, Asie, et des réseaux régionaux semi-autonomes ont été créés (INCLENAfrica, ChinaCLEN, IndiaCLEN, INCLEN-Southeast Asia, LatinCLEN ; Macfarlane, 1999). À ce jour, des unités d'épidémiologie clinique ont été créées dans la plupart des hôpitaux universitaires dans le monde et l'épidémiologie clinique est devenue une discipline à part entière dans de nombreux pays où sont nommés des professeurs d'épidémiologie clinique (États-Unis, Canada, Angleterre, Hollande, Australie, ...).

5 | La situation française

La France a pris dans le domaine de l'épidémiologie clinique un retard considérable vis-à-vis des États-Unis et du Canada, mais également vis-à-vis des autres pays européens (Angleterre, Pays-Bas, Italie, Espagne, pays scandinaves). Ce retard est probablement lié au manque d'intérêt marqué des épidémiologistes « classiques » pour cette nouvelle forme d'épidémiologie. Ce conflit d'intérêt épidémiologie-épidémiologie clinique n'est pas spécifique à la France mais semble avoir été réglé plus rapidement dans les autres pays. La méfiance typiquement française envers des sciences appliquées, moins valorisantes que les sciences plus fondamentales, a probablement participé également à ce phénomène, en particulier pour l'Inserm.

Malgré cela, des unités d'épidémiologie clinique existent maintenant dans différents sites hospitalo-universitaires ou hospitaliers ; ces unités sont le plus souvent insérées dans des services d'épidémiologie, de biostatistiques ou de santé publique. Un réseau français des unités d'essais cliniques, regroupant les unités

appartenant à des institutions publiques coordonnées par un méthodologiste et réalisant des essais randomisés, a été récemment créé. Ce réseau, qui regroupe 38 unités, a pour ambition de promouvoir la rigueur méthodologique dans la recherche clinique institutionnelle. Un nombre très limité d'unités Inserm (moins de 5) affiche l'épidémiologie clinique comme un de leurs thèmes de recherche. Cependant, dans le cadre d'un appel d'offres conjoint Dhos-Inserm, sept centres d'investigation clinique-épidémiologie clinique ont été créés en 2003.

La participation française à la Cochrane collaboration reste très limitée en dehors de quelques individus impliqués à titre personnel. Le centre Cochrane français, qui n'a jamais eu l'envergure des autres centres Cochrane européens, a dû interrompre ses activités en 2000. Une tentative timide de création d'un réseau Cochrane francophone est actuellement en cours.

En termes d'enseignement, un pas important a été réalisé avec l'introduction dans le second cycle des études médicales de plusieurs items tels que l'analyse critique de la littérature médicale ou la médecine fondée sur les faits prouvés. Malheureusement, l'épreuve d'analyse critique, représentant 20 % de la note de l'examen national classant, devait initialement être effective en 2003 ne le sera pas au mieux avant 2008. De plus, les articles concernant des méta-analyses ont été exclus de cet enseignement. Dans le cadre de l'enseignement licence-master-doctorat (LMD), plusieurs formations de type M2 intégrant l'épidémiologie clinique ont été mises en place récemment.

Conclusion

Restée ignorée pendant un siècle et demi, la médecine numérique prônée par Louis a pris une revanche éclatante en étant à l'origine de la révolution que constitue, en médecine, l'épidémiologie clinique et ses dérivés tels la Cochrane collaboration et *Evidence-Based Medicine*. Ces évolutions conceptuelles majeures récentes, bien qu'ayant fait initialement l'objet de vives critiques, ont dès à présent profondément marqué à la fois la recherche clinique, l'enseignement de la médecine (formation initiale et formation continue) et la pratique médicale.

Bibliographie

Ad Hoc Working Group for critical appraisal of the medical literature (1987). A proposal for more informative abstracts of clinical articles. *Ann Intern Med*, **106** : 598-604.

Bossuyt P. *et al.* (2003). Standards for reporting of diagnostic accuracy : explanation and elaboration. *Ann Intern Med*, **138** : W1-12.

- Chalmers I. (1977). Randomize the first patient. *N Eng J Med*, **296** : 107.
- Chalmers I., Hetherington J., Newdick M., Mutch L., Grant A., Enkin M., Enkin E., Dickersin K. (1986). The Oxford database of perinatal trials : developing a register of published reports of controlled trials. *Control Clin trials*, **7** : 306-324.
- Chalmers I. (1993). The Cochrane collaboration : preparing, maintaining, and disseminating systematic reviews of the effects of health care. *Ann N Y Acad Sci*, **703** : 153-163.
- Clinical Evidence. The international source of the best available evidence for effective health care. *BMJ Publishing Group*, issue **10** : décembre 2003.
- Cook D. *et al.* (1997). Systematic reviews. Synthesis of best evidence for clinical decisions. *Ann Int Med*, **126** : 376-380.
- Davidoff F. *et al.* (1995). Evidence-based medicine : a new journal to help doctors identify the information they need. *Br Med J*, **310** : 1085-1086.
- De Angelis C. *et al.* (2004). Clinical trial registration. A statement from the international committee of medical journal editors. *JAMA*, **292** : 1363-1364.
- Dickersin K. *et al.* (2002). Association between time interval to publication and statistical significance. *JAMA*, **287** : 2829-2831.
- Evidence-based medicine working group (1992). Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA*, **268** : 2420-2425.
- Feinstein A. (1968a). Clinical epidemiology. I. The populational experiments of nature and of man in human illness. *Ann Intern Med*, **69** : 807-820.
- Feinstein A. (1968b) Clinical epidemiology. II. The identification rates of disease. *Ann Intern Med*, **69** : 1037-1061.
- Feinstein A. (1968c). Clinical epidemiology III. The clinical design of statistics in therapy. *Ann Intern Med*, **69** : 1287-1312.
- Feinstein A. (1985). Clinical epidemiology. The architecture of clinical research. Philadelphia : W. Saunders Eds.
- Fletcher R., Feinstein A. (2001). The father of clinical epidemiology, 1925-2001. *J Clin Epidemiol*, **54** : 1188-1190.
- Gail M., Benichou J. (2000). *Encyclopedia of epidemiologic methods*. Wiley reference series in biostatistics. P. Armitage and T. Colton Eds.
- Grant A., Chalmers I. (1981). Register of randomised controlled trials in perinatal medicine. *Lancet*, **1** : 100.

- Grimshaw J. (2004). So what has the Cochrane collaboration ever done for us? a report card on the first 10 years. *Can Med Ass*, **171** : 747-751.
- Guyatt G. *et al.* (2004). Evidence based medicine has come a long way. *BMJ*, **329** : 990-991.
- Haynes R. *et al.* (1994). Developing optimal search strategies for detecting clinically sound studies in Medline. *J Am Med Inform Assoc*, **1** : 447-458.
- <http://www.archive.official-documents.co.uk>
- <http://www.inclen.org>
- Jenicek M. (1997). Epidemiology. Evidence-based medicine and evidence-based public health. *J Epidemiol*, **7** : 187-197.
- Laupacis A. *et al.* (1988). An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med*, **318** : 1728-1733.
- Lenfant C. (2003). Clinical research to clinical practice lost in translation? *New Engl J Med*, **349** : 868-874.
- Macfarlane S. *et al.* (1999). Global health research and INCLEN. International clinical epidemiology network. *Lancet*, **353** : 503.
- Mallett S., Clarke M. (2003). How many Cochrane reviews are needed to cover existing evidence on the effects of health care interventions? *ACP J Club*, **139** : A11-A12.
- Moher D. *et al.* (1999). Improving the quality of reports of meta-analyses of randomised controlled trials : the QUOROM statement. Quality of reporting of meta-analyses. *Lancet*, **357** : 1896-1900.
- Moher D. *et al.* (2001). The CONSORT statement : revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet*, **357** : 1191-1194.
- Morabia A. (1996). P.C.A. Louis and the birth of clinical epidemiology. *J Clin Epidemiol*, **49** : 1327-1333.
- Mulrow C. (1987). The medical review article : state of the science. *Ann Intern Med*, **106** : 485-488.
- Mulrow C., Cook D. (1998). *Systematic reviews : synthesis of best evidence for health care decisions*. American College of Physicians, Philadelphia.
- Nash B., Hicks C., Dillner L. (2003). Connecting doctors, patients and the evidence. *BMJ*, **326** : 674.

- Naylor CD. (1995). Grey zones of clinical practice : some limits to evidence-based medicine. *Lancet*, **345** : 840-842.
- Paul JR. (1938). Clinical epidemiology. *J Clin Invest*, **17** : 539-541.
- Sackett D. (1969). Clinical epidemiology. *Am J Epidemiol*, **89** : 125-128.
- Sackett D. *et al.* (1985). *Clinical epidemiology : a basic science for clinical medicine*. 1st ed. Boston : Little, Brown, p. XIII.
- Sackett D. (2002). Clinical epidemiology : what, who and whither. *J Clin Epidemiol*, **55** : 1161-1166.
- Straus SE., Jones G. (2004). What has evidence based medicine done for us? *BMJ*, **329** : 987-988.
- Vandenbroucke J. (1996). Evidence-based medicine and « médecine d'observation ». *J Clin Epidemiol*, **49** : 1335-1338.
- Wolf S. (1999). The need for perspective in evidence-based medicine. *JAMA*, **282** : 2358-2365.

SOUS-CHAPITRE 8.2

Épidémiologie et décision en santé publique

MICHEL SETBON

En tant que discipline « *qui vise à identifier les déterminants des variations de fréquence des maladies et la compréhension de leurs causes* », l'épidémiologie possède des liens étroits avec la décision en santé publique. Par sa nature et ses objectifs populationnels, elle s'avère en mesure, parfois de déterminer directement l'action publique, souvent d'être mobilisée pour éclairer des questions de santé publique. L'intérêt récent et croissant du public et des pouvoirs publics pour la maîtrise et la prévention des risques sanitaires n'a fait que renforcer ces liens en faisant de l'épidémiologie l'une des principales sources de connaissances face au risque. Cette fonction de l'épidémiologie correspond au développement de l'expertise en santé publique à laquelle elle fournit une contribution déterminante. Quels sont ses apports et limites en matière de risque sanitaire ? Comment s'opère la transition de la connaissance épidémiologique produite à la décision publique ? Questions essentielles pour rendre compte de ces liens.

Les principaux apports de l'épidémiologie (en dehors de sa contribution aux essais thérapeutiques) se répartissent selon trois grands axes problématiques, selon les questions auxquelles elle est amenée à répondre :

- le premier qui correspond à l'émergence de nouveaux risques sanitaires candidats est illustré par la question clé : y a-t-il risque ? À partir d'un danger identifié, existe-t-il des effets sanitaires néfastes conséquents ou bien, à partir d'événements inhabituels constatés, peut-on incriminer une source à leur origine ? Par exemple existe-t-il un lien causal entre un excès de leucémies constatées et la proximité d'un complexe industriel nucléaire ? La consommation de viande de bœuf infecté par l'agent de l'ESB détermine-t-elle une forme humaine d'ESST, etc. ? La réponse, comme la question, est de nature qualitative, positive, négative ou incertaine et demande la mobilisation de techniques spécifiques dont l'enquête cas-témoins est la plus souvent utilisée ;
- le deuxième axe concerne l'estimation du risque, si le risque est connu ou s'il est qualifié comme tel à partir de la réponse à la précédente question : l'estimation vise à fournir la fréquence et la probabilité d'occurrence de la pathologie en relation avec la source identifiée. Les outils observationnels comme les cohortes permettent d'y répondre dans certains cas (traitement hormonal substitutif de la ménopause chez la femme), mais dans d'autres il est nécessaire de recourir à la modélisation ;
- le troisième axe concerne le choix des actions envisageables pour réduire le risque : tel test de dépistage est-il utile ? Telles modalités de vaccination sont-elles bénéfiques ? Faut-il interdire tel produit, changer les indications de tel médicament, etc. ? Axe relevant pour l'essentiel de la modélisation qui permet de fournir différents scénarios associant une action donnée avec ses conséquences.

Si l'épidémiologie s'est affirmée comme une discipline nécessaire à l'expertise pluridisciplinaire et organisée en santé publique, les réponses qu'elle apporte n'en sont pas pour autant toujours traduites sous forme de décision. Tout d'abord, du fait de ses limites propres à révéler la réalité et l'importance d'un risque. Les contraintes de temps, de faisabilité et de disponibilité ou d'accès à des données fiables réduisent souvent la portée des réponses marquées par une importante incertitude, tant en termes de causalité que d'estimation du risque. Ensuite, l'expérience montre qu'il n'existe pas une forte cohérence entre la réponse apportée à la question posée et l'action (ou l'inaction) décidée. Ainsi, si la réponse peut être « *Non, il n'y a pas risque compte tenu des connaissances disponibles, ou bien le risque est possible, mais il est très faible* », les pouvoirs publics peuvent néanmoins décider d'agir sur la base, soit du principe de précaution dans le premier cas, soit parce qu'ils jugent que, même faible, le risque serait politiquement inacceptable dans le second. À l'inverse, la démonstration de la réalité d'un risque et sa quantification ne débouchent pas toujours sur un programme d'action. De nombreux paramètres, parfois explicites et souvent implicites, pèsent tant sur la décision d'(in)action que sur la rationalité qui

prévaut dans la sélection d'un programme. Cette discontinuité potentielle entre connaissance et action publique s'explique par la nature politique du changement qu'entraînera l'action publique, c'est-à-dire par les conséquences de toute nature que les gestionnaires politico-administratifs du risque anticipent : le plus souvent les conséquences importantes et immédiates sont d'ordre social et économique, alors que les impacts sanitaires attendus sont ténus, incertains et reportés dans des temps éloignés.

Enfin, la connaissance épidémiologique produite n'est pas toujours, loin s'en faut, considérée comme valide par le public. D'une part, l'existence du risque perçu, entendu comme « *une évaluation subjective mélangeant des informations issues de multiples sources et des sentiments suscités par le phénomène* », vient concurrencer l'estimation du risque observé, quand celle-ci a pu être produite. D'autre part, les changements de comportements que promeuvent certains programmes de santé publique fondés sur la connaissance épidémiologique sont récusés sur la base de perceptions et d'enjeux individuels : la non-observance de comportements visant à réduire des risques construits à l'échelle individuelle (tabac, drogues illicites, MST, etc.) concrétise une autre limite de l'impact de l'épidémiologie sur la décision en santé publique.

SOUS-CHAPITRE 8.3

« Mais pourquoi résistent-ils ? » Conditions de réalisation d'actions de santé publique sur une base épidémiologique

DIDIER TORNY

Un regard naïf laisserait penser que le discours rationnel de l'épidémiologie n'aurait qu'à se diffuser pour changer les comportements individuels et bâtir des politiques de santé publique. Sous cette hypothèse, au vu des résultats concrets parfois mitigés, une question émerge rapidement : « pourquoi les populations ne suivent-elles pas les recommandations issus de la connaissance épidémiologique ? ». Une réponse simple apportée par les sciences sociales a été largement diffusée hors de son champ académique : tout ne serait qu'une question de « perception » différentielle entre connaissance savante et connaissance populaire des maladies. Largement issus du savoir anthropologique sur le risque (Douglas et Wildavsky, 1982) ou le sens de la maladie (Herzlich et Augé, 1984), ces résultats montrent par exemple la difficulté à faire exister pour certaines populations une menace temporellement éloignée. Loin d'être négligeables, ces

résultats peuvent promouvoir l'idée d'une relation directe entre « irrationalité » du public et refus du discours épidémiologique.

Cependant, cette idée suppose trois éléments qui sont loin d'être toujours réunis : l'existence d'un savoir épidémiologique unifié, l'appui explicite ou implicite des structures politiques et administratives, la capacité à s'adresser directement au grand public. Ces trois facteurs constituent des éléments essentiels pour comprendre l'absence éventuelle de succès du discours épidémiologique. Nous allons brièvement analyser leur importance pour comprendre les résistances au discours épidémiologique : on soulignera d'abord l'intensité des controverses scientifiques engageant la discipline et leur caractère de plus en plus public, qui peuvent diminuer la crédibilité du discours scientifique de l'épidémiologie. Ensuite, on rappellera la nécessité pour l'épidémiologie de construire des alliances afin de pouvoir surmonter des intérêts antagonistes à ses résultats. Enfin, on indiquera l'importance des opérations de traduction indispensables à la transformation du discours scientifique de l'épidémiologie en messages politiques de santé publique.

1 | Une science en débat public

Le succès du raisonnement épidémiologique l'a rendu extrêmement populaire : pour s'en convaincre, il suffit d'ouvrir un magazine et d'y apercevoir des données extraites d'études de cohortes. De par sa diffusion massive et son apparente simplicité, il sert ainsi de point d'appui à l'argumentation dans l'espace public, que ce soit pour s'opposer à la mise à l'écart des séropositifs (Pollak, 1988), classer la nocivité des drogues légales et illégales, ou dénoncer la dangerosité de thérapeutiques ne se soumettant pas aux procédures encadrées d'essais thérapeutiques (Barbot, 2002).

Cette circulation a une contrepartie : la critique savante de la connaissance épidémiologique. Celle-ci peut d'abord provenir de mouvements, d'associations qui mettent en cause les conditions de production, de diffusion et d'application du savoir épidémiologique. Les critiques peuvent insister sur l'encadrement plus strict du recueil des données des maladies à déclaration obligatoire (Torny, 2001), l'oubli des déterminants socio-économiques dans l'apparition des pathologies (Leclerc *et al.*, 2000), ou la nécessité de la réouverture du *design* des essais pour y introduire plus de souplesse (Dodier, 2003) en raison d'un contexte d'urgence sanitaire.

D'autres critiques proviennent de professionnels de santé : rappelons que les médecins ont toujours été au cœur des combats anti-vaccinaux (Murard et

Zylberman, 1996), critiquant les preuves populationnelles apportées par certains hygiénistes et soutenant des techniques de surveillance et d'isolement des cas comme alternative à la vaccination antivariolique (Fraser, 1980). On peut également citer la longue polémique sur les causes du sida, où les « dissidents du sida » étaient issus du monde de la recherche biologique et médicale (Fujimura et Chou, 1994).

De même, les controverses sur les effets indésirables des traitements hormonaux (contraceptifs ou post-ménopause) ont vu une forte opposition entre différents groupes à l'intérieur même du monde de la santé. L'importance de ces controverses est essentielle dans la modification des comportements : que doit faire une femme ménopausée des derniers résultats indiquant que le traitement hormonal de substitution aurait un effet cancérigène ? Arrêter son traitement en écoutant certains responsables de santé publique ou suivre l'avis de son gynécologue qui lui indique de le poursuivre ?

Dans chacune de ses controverses et polémiques, il ne s'agit pas de repousser le raisonnement épidémiologique, mais d'en adapter les modalités en fonction des objets dont il s'empare peu à peu : conditions d'inclusion ou d'exclusion dans les cohortes, modalités de comparaison dans les méta-études, raisonnement aux limites sur les faibles doses, modélisations épidémiologiques à partir d'un tout petit nombre de cas. Pour ses critiques, il ne faut pas moins, mais plus d'épidémiologie, quitte à redéfinir ses procédures et ses contours.

2 | L'épidémiologie au risque de la politique

En admettant qu'une connaissance épidémiologique stabilisée soit effectivement produite et qu'elle recueille un relatif consensus dans les communautés scientifiques concernées, d'autres facteurs s'opposent à son extension hors du savoir académique. En effet, du point de vue d'une sociologie politique, les épidémiologistes n'en sont pas moins des porteurs d'intérêt comme les autres, des « entrepreneurs de morale » qui cherchent à imposer un certain regard de santé publique au détriment d'autres causes.

Ainsi, la lutte contre l'alcool au volant (Gusfield, 1981) ou la reconnaissance de la dangerosité du tabagisme passif (Berlivet, 2000) font partie de ces causes qui se constituent sur une base épidémiologique. Il est facile de comprendre que des intérêts industriels s'opposent aux constats épidémiologiques et mettent en avant l'importance des emplois en jeu ou la nécessité de séparer les usages « socialement dangereux » des substances de ceux qui relèveraient des libertés individuelles. Aussi, il n'est pas étonnant que l'épidémiologie ne soit pas le seul déterminant de la santé publique : des intérêts catégoriels, en particulier ceux

des professionnels de santé, ou des jeux de pouvoir dans la haute administration (Setbon, 1993) peuvent faire concurrence au cadre épidémiologique.

De manière plus fondamentale, diverses modalités d'appréciation des dangers peuvent s'imposer : faut-il prendre en compte les victimes futures plutôt que la prévalence ou l'incidence actuelle (Chateauraynaud et Torny, 1999 ; Torny, 2001) ? Que la décision de santé publique soit alors fondée sur l'application du principe de précaution (Callon *et al.*, 2001) ou sur un « principe de parapluie » engendré par l'expérience judiciaire et médiatique du sang contaminé (Hermitte, 1996), il n'y a pas à considérer que l'épidémiologie, comme toute autre expertise technique, puisse être le moteur de l'action publique sans capacité à agréger autour d'elle divers intérêts.

3 | De l'épidémiologie de laboratoire à la pratique de santé publique

Même en admettant que le consensus épidémiologique a atteint un certain degré d'acceptation parmi les groupes mobilisés et que le pouvoir politique a été conquis, il existe toujours un écart entre les représentations épidémiologiques et les nécessités de la santé publique. Le passage de l'un à l'autre exige une traduction, particulièrement visible dans le cas des campagnes de prévention (Paicheler, 1994) qui doivent tenir compte d'une multitude de publics et des effets propres des messages générés. Même si l'accord s'est fait sur la promotion du préservatif, à la fois contre le choix nataliste passé et les pressions religieuses et politiques présentes, les modalités de cette promotion restent totalement ouvertes et extérieures au savoir académique.

Les épidémiologistes connaissent eux-mêmes la difficulté de cette traduction et ils orientent la présentation de leurs résultats en fonction des utilisations futures en santé publique (Torny, 2001) : privilégier les années de vie perdues, le risque attribuable ou le risque relatif, constituer deux ou trois classes de population pour la consommation tabagique sont des choix essentiels dans la détermination des futures actions de santé publique. Ces anticipations, encadrées par les contraintes de la production de connaissance¹, s'adressent à différents publics : en effet, décideurs, professionnels de santé, industrie pharmaceutique ou grand public constituent autant de groupes différents qu'il faut convaincre non seulement du bien-fondé des résultats, mais aussi de l'importance de mettre en œuvre des mesures de régulation ou de prévention. Faire entendre la nécessité des

¹ En épidémiologie, comme dans toute autre science, la fraude est toujours possible, qu'elle soit motivée par la recherche de la reconnaissance académique, l'anticipation du profit ou la nécessité de plaire à ses promoteurs.

lourdes procédures des essais thérapeutiques aux cliniciens (Marks, 1999) est fondamentalement différent que de mobiliser autour de la détection du cancer (Fujimura, 1994) ; convaincre les députés de l'urgence à interdire la publicité de l'alcool (Berlivet, 2000) n'apprend pas nécessairement à produire des messages acceptables et efficaces dans la lutte contre le sida (Paicheler, 1994).

4 | Quand l'épidémiologie réussit

Les contraintes que nous avons décrites existent, à des degrés plus ou moins importants, pour toute connaissance épidémiologique. Elles n'en constituent pas moins des obstacles qui sont souvent franchis et on peut affirmer que, peu à peu, un style épidémiologique s'est formé, introduisant en santé publique les essais thérapeutiques (Marks, 1999) et les facteurs de risque (Berlivet, 2001) comme des entités naturelles du monde qui nous entoure, comme d'autres sciences avaient inclus les microbes (Latour, 1984). Ce style, avec ses dispositifs et ses pratiques, s'est largement imposé à partir du milieu du XX^e siècle pour constituer le fondement scientifique de la santé publique, que ce soit pour la détermination des priorités de santé ou l'évaluation des mesures et des produits. Aussi, en conclusion, on peut proposer d'inverser la perspective : que se passe-t-il quand la croyance en la connaissance épidémiologique est fermement partagée ? Évoquons la puissance de l'explication par les facteurs de risques à travers deux exemples.

À partir des années 1940, des classes de maladies — telles les maladies cardiovasculaires — vont être considérées comme résultant d'une série de facteurs « non médicaux » tels que l'alimentation, la consommation de tabac ou d'alcool, l'activité physique, ensuite dénommés « facteurs de risque » à partir du début des années 1960 (Aronowitz, 1999). Cette approche particulière de la maladie a des conséquences importantes : les facteurs de risque sont la base d'une politique nationale de prévention des maladies coronariennes, qui recommande un changement de style de vie et un traitement des anomalies révélées par la pratique de bilans. Mais que faire des « changements de style de vie » lorsque des gènes de prédisposition, par essence hors du comportement des individus, sont découverts ? Le second exemple concerne précisément les consultations de dépistage génétique, directement issues de la recherche épidémiologique. Même si certaines d'entre elles n'ont pas connu le succès espéré par leurs promoteurs, comme dans le cas de la chorée de Huntington (Terrenoire, 1993), d'autres ont montré la spectaculaire efficacité du discours épidémiologique. Le poids du dépistage génétique est si fort qu'il a pu entraîner des décisions radicales, telles que de nombreuses mammectomies préventives chez des femmes à qui on a fait l'annonce de détection de mutation sur les gènes *BRCA*.

Ces deux exemples, loin d'être marginaux, montrent la capacité d'orientation politique et sociale des représentations et dispositifs produits par l'épidémiologie. La montée en puissance contemporaine de la santé comme « bien en soi » (Dodier, 2003) ou support d'une utopie politique (Sfez, 2001) ne peut qu'augmenter cette faculté. Néanmoins, elle s'accompagne de la prolifération des discours concurrents sur ce qu'est la santé, comment l'obtenir, la conserver ou l'améliorer. Aussi, si les épidémiologistes veulent que leur science soit utile à la santé publique, il leur faut plus que jamais savoir trancher des controverses impliquant de multiples protagonistes, apprendre à recueillir l'assentiment de gouvernants arbitres de multiples intérêts, ne pas négliger la connaissance du terrain nécessaire à la mise en œuvre de véritables actions de santé publique. Loin d'être des formes de « résistance irrationnelle », les comportements individuels et collectifs apparaîtront dès lors comme des choix culturels, sociaux et politiques à intégrer au cœur même de la pensée épidémiologique, dès lors qu'elle souhaite s'incarner dans une politique de santé publique.

Bibliographie

- Aronowitz R. (1999). *Les maladies ont-elles un sens ?* Synthélabo. Les empêcheurs de penser en rond, Paris.
- Barbot J. (2002). *Les malades en mouvements. La médecine et la science à l'épreuve du sida*. Balland. Voix et regards, Paris.
- Berlivet L. (2000). *Une santé à risques. L'action publique de lutte contre l'alcoolisme et le tabagisme en France (1954-1999)*, Thèse de doctorat en sociologie, Rennes, Université de Rennes I.
- Berlivet L. (2001). Déchiffrer la maladie, In : *Critique de la santé publique. Une approche anthropologique*, J.-P. Dozon et D. Fassin (eds.), Balland, Paris, 74-102.
- Callon M., Lascoumes P., Barthe Y. (2001). *Agir dans un monde incertain. Essai sur la démocratie technique*. La couleur des idées. Seuil. Paris.
- Chateauraynaud F., Torny D. (1999). *Les Sombres Précurseurs. Une sociologie pragmatique de l'alerte et du risque*. Éditions de l'EHESS, Recherches d'histoire et de sciences sociales. Paris.
- Dodier N. (2003). *Leçons politiques de l'épidémie de sida*. Éditions de l'EHESS. Cas de figure. Paris.
- Douglas M. et Wildavsky A. (1982). *Risk and Culture. An Essay on the Selection of Technological and Environmental Dangers*. University of California Press. Berkeley.

- Fraser SM. (1980). Leicester and smallpox : the Leicester method. *Medical History*, **24** (3) : 315-332.
- Fujimura JH. (1994). *Crafting Science. A Sociohistory of the Quest for the Genetics of Cancer*. Harvard University Press. Cambridge.
- Fujimura JH., Chou DY. (1994). Dissent in science : styles of scientific practice and the controversy over the cause of AIDS. *Social Sciences and Medicine*, **38**, (8) 1017-1036.
- Gusfield JR. (1981). *The Culture of Public Problems. Drinking-Driving and the symbolic order*. The University of Chicago Press. Chicago.
- Hermitte MA. (1996). *Le sang et le droit*. Seuil, Paris.
- Herzlich C., Augé M. (1984). *Le sens du mal. Anthropologie, histoire, sociologie de la maladie*. Éditions des archives contemporaines. Paris.
- Latour B. (1984). *Les Microbes : guerre et paix — suivi de Irréductions*. Métailié, Pandora. Paris.
- Leclerc A., Fassin D., Grandjean H., Kaminski M., Lang T. (eds.). (2000). *Les inégalités sociales de santé*. La Découverte/Inserm, Recherches. Paris.
- Marks H. (1999). *La médecine des preuves. Histoire et anthropologie des essais cliniques (1900-1990)*. Synthélabo, Les Empêcheurs de penser en rond. Paris.
- Murard L., Zylberman P. (1996). *L'hygiène dans la République. La santé publique en France, ou l'utopie contrariée (1870-1918)*. Fayard. Paris.
- Paicheler G. (1994). *Le public face à la menace du sida*. ANRS-Cermes.
- Pollak M. (1988). *Les homosexuels et le sida. Sociologie d'une épidémie*, Métailié, Leçons de choses. Paris.
- Setbon M. (1993). *Pouvoirs contre sida. De la transfusion sanguine au dépistage : décisions et pratiques en France, Grande-Bretagne et Suède*. Seuil. Paris.
- Sfez L. (ed.). (2001). *L'utopie de la santé parfaite*. PUF, La Politique éclatée. Paris.
- Terrenoire G. (1993). Médecine prédictive, l'épreuve de l'expérience : dépistage présymptomatique de la maladie de Huntington. *Sciences sociales et santé*, **11** (3-4) : 99-122.
- Torny D. (2001). *Un seul cas suffit. Une étude sociologique de l'engagement du futur en biomédecine*. Thèse de doctorat en sociologie. École des hautes études en sciences sociales. Paris.

SOUS-CHAPITRE 8.4

Force de conviction de la modélisation prévisionnelle épidémiologique dans l'action publique face au risque

MICHEL SETBON

Introduction

Depuis les années 1980, on assiste à un renouvellement conceptuel, politique et social du champ de l'action publique face au risque. D'une part, la « preuve » (quel qu'en soit le degré) épidémiologique est devenue le référentiel pour juger de la nature et de la valeur de multiples phénomènes susceptibles d'être l'objet d'action publique, d'autre part la demande sociale de protection s'est élargie à de nouveaux risques jusque-là peu explorés par cette discipline. L'émergence de ces nouveaux risques à faible incidence ou en relation avec des faibles doses, couplée à l'exigence d'anticipation devenue une règle de la gouvernance moderne, a, en grande partie, rendu inadaptés ou insuffisants les outils classiques de l'épidémiologie.

La modélisation prévisionnelle épidémiologique représente une alternative dont le statut apparaît aujourd'hui mal défini en termes d'aide à la décision publique. On regroupera ici sous le terme de modèle prévisionnel épidémiologique toute construction intellectuelle qui, à partir d'hypothèses jugées cohérentes avec un corpus de faits observés et de connaissances, fournit une représentation de l'état du monde présent ou futur. Son objet scientifique, la prévision rationnelle, correspond directement au besoin d'une action précoce fondée sur suffisamment de connaissance pour éclairer et guider l'action publique face au risque. Si la modélisation épidémiologique s'est largement développée dans le champ scientifique, on en sait fort peu sur son pouvoir réel de conviction et en conséquence sur les conditions de son usage au service de l'action publique. La force de conviction d'un modèle épidémiologique s'évalue en dernière analyse par son utilisation dans les décisions publiques face au risque. Si, à l'instar de l'épidémiologie en général, les critères scientifiques de la modélisation prévisionnelle épidémiologique (MPE) sont insuffisants à déterminer sa force de conviction, on peut s'interroger sur les conditions qui rendent possible son utilisation dans la décision publique. Deux questions seront abordées ici : à quelles séquences de l'action publique face au risque la modélisation épidémiologique contribue-t-elle ou est-elle en mesure de contribuer ? Quelles sont les conditions qui rendent possibles son usage dans l'action publique ?

Pour éclairer ces questions, deux études de cas seront brièvement exposées : le cas du risque de transmission du nouveau variant de la maladie de Creutzfeldt-Jakob (vMCJ) en transfusion sanguine et celui (propre aux États-Unis) prédisant la (l'im)possibilité de survenue sur le territoire américain de l'épizootie d'ESB (encéphalopathie spongiforme bovine). À la suite, seront analysés les fondements qui justifient le recours à la modélisation ainsi que ses apports potentiels en relation avec sa force de conviction.

1 | Risque de transmission du vMCJ en transfusion sanguine

Depuis 1997, la possibilité d'une transmission humaine de l'agent causal de l'ESB est considérée comme un fait scientifique. En 1999, la France recensait 2 cas de vMCJ (3 en 2000, 9 en 2004), diagnostiqués comme cas primaires, c'est-à-dire comme consécutifs à un probable passage de l'agent causal de l'ESB à l'homme. À cette date, aucun cas de contamination secondaire n'avait été rapporté de par le monde, y compris par transfusion sanguine. Néanmoins, quelques indices scientifiques rendaient théoriquement possible une transmission par le sang du vMCJ, essentiellement fondée sur la comparaison avec la MCJ classique. Cette possibilité, bien que considérée comme faible, soulève depuis 1998 la question suivante : qu'est-il possible de faire pour réduire ce « risque », aussi hypothétique soit-il ?

La décision prise par les États-Unis suivis par le Canada, pays jusqu'ici indemnes d'ESB, d'exclure tous les donneurs de sang ayant séjourné plus de 6 mois dans les Iles Britanniques interpelle directement les autorités sanitaires des pays disposant d'une transfusion sanguine moderne, dont la France. À la lecture des rapports produits par la FDA et par Health Canada (1998), deux conclusions fortes se détachent : d'une part, malgré une absence de preuve directe de la transmission sanguine du vMCJ, ce danger¹ est jugé impossible à ignorer ; d'autre part, la mesure retenue pour le prendre en compte, l'exclusion des donneurs de sang qui présentent ce qui est considéré comme un facteur de risque, est déduite d'une modélisation prenant en compte la balance des estimations de la réduction du risque et de celle des bénéfices sacrifiés consécutifs à l'exclusion d'un certain pourcentage des donneurs de sang. Ce qui explique la durée retenue de six mois, qui ne possède aucun fondement épidémiologique, ni physiopathologique. Ce n'est donc pas l'estimation quantitative du risque lui-même qui est modélisée, mais la capacité théorique de la mesure d'exclusion à le réduire en fonction de l'exposition de la population des donneurs de sang.

Les autorités sanitaires françaises se penchent sur la question en juillet 1999 en désignant un groupe d'expertise qui doit répondre aux trois questions suivantes :

- La transmissibilité sanguine de l'agent responsable du vMCJ est-elle possible ?
- Quel serait le niveau d'exposition de la population des donneurs de sang français à l'agent ESB ?
- Quelles méthodes de réduction du risque de transmission du vMCJ par les produits sanguins pourrait-on mettre en place ?

La réponse à la première question, qui se fonde pour l'essentiel sur l'identification d'un marqueur de l'infectiosité dans le tissu lymphoïde, est la suivante : « *Il s'agit d'un risque potentiel qu'il convient de prendre en considération par le maintien d'une veille scientifique et épidémiologique attentive et rigoureuse* ».

À la question posée sur l'exposition de la population française le groupe d'experts, s'appuyant sur la modélisation utilisée aux États-Unis et au Canada reliant la réduction du risque en fonction du pourcentage de donneurs exclus exposés, aboutit à la conclusion suivante : « *Quelle que soit l'hypothèse retenue en termes de durée de séjour dans les Iles Britanniques et de pourcentage de viandes anglaise importées, l'impact d'une exclusion des donneurs sur le*

¹Danger (*hazard*) : agent biologique physique ou chimique présent dans un produit ou état de ce produit pouvant entraîner un effet néfaste sur la santé.

critère de séjour dans les Iles Britanniques serait extrêmement limité et jamais supérieur à 5 % du total de l'exposition » (Afssaps, 2000). Autrement dit, les données du problème français sont différentes des données nord-américaines. En effet, les données d'exposition varient selon les pays, essentiellement du fait de la consommation entre 1980 et 1996 de viandes britanniques importées en France (estimée à près de 10 % de la consommation totale). Sur cette base, la modélisation aboutissait à une réduction très limitée du risque en France : moins de 10 %, alors que, sur la base d'une exclusion supérieure à six mois, les États-Unis et le Canada obtenaient une réduction du risque de près de 90 %. Les experts décident donc de ne pas retenir la mesure.

Concernant la troisième question, les mesures appropriées pour réduire le risque, la stratégie proposée est celle d'une réduction du risque par déleucocytation, ou plus précisément d'appauvrissement en leucocytes, des produits sanguins. Elle se fonde sur l'hypothèse d'une infectiosité du sang majoritairement associée aux leucocytes (constatée en matière de MCJ).

Pour résumer, l'hypothèse d'une transmission sanguine est considérée comme possible sans preuve directe (référence explicite au principe de précaution) déterminant ainsi le besoin d'une action protectrice ; la forme d'action étant, elle, retenue par modélisation qui prend en compte les bénéfices et les pertes potentielles au niveau de la population concernée. Si le même modèle fondé sur les mêmes hypothèses a été utilisé aussi bien en France qu'outre-Atlantique, le résultat en termes d'action est différent². Dans les trois pays, les conclusions sont acceptées par les gestionnaires politico-administratifs qui décident de les mettre en œuvre. Ce n'est qu'en 2004 que plusieurs publications viendront apporter les premiers éléments de preuve de la validité de l'hypothèse d'une transmission sanguine de l'agent infectieux (Pincock, 2004 ; Aguzzi et Glatzel, 2004 ; Lewelyn *et al.*, 2004 ; Peden *et al.*, 2004) .

2 | Évaluation du risque d'ESB aux États-Unis

Alors qu'en Europe l'épidémie d'ESB a été identifiée depuis la fin des années 1980, les autorités sanitaires des États-Unis s'interrogent en 2000 sur la possibilité d'en être atteints à leur tour (Gray *et al.*, 2002). Ce cas est singulier dans la mesure où il ne s'agit pas de s'interroger sur l'incidence de l'ESB dans ce pays, ni de déterminer les mesures capables de la réduire : la question est d'anticiper en termes de probabilité la survenue de l'épidémie sur ce territoire où aucun cas d'ESB n'a été signalé durant 12 années de surveillance, ainsi que de cas

²Ce n'est qu'un an plus tard, lors d'une nouvelle révision, que sera proposée une mesure d'exclusion pour les donneurs de sang ayant séjourné plus d'un an dans les Iles Britanniques : rapport Afssaps, janvier 2001.

humains de vMCJ. Répondre à cette question est un préalable pour envisager de nouvelles mesures en plus de celles déjà prises. Le phénomène n'étant pas observé, la modélisation reste la seule méthode pour répondre à la question. Pour cela les modélisateurs s'appuient d'une part, sur les connaissances accumulées par les pays atteints, en particulier par le Royaume-Uni, et d'autre part, sur les mesures prises aux États-Unis pour prévenir l'importation et la propagation de l'épizootie.

Pour tenter de couvrir les nombreuses alternatives susceptibles de favoriser l'apparition et la propagation de l'ESB, 29 scénarios sont analysés, qui se fondent sur différentes hypothèses portant sur les caractéristiques biologiques de l'ESB, les pratiques de l'industrie d'aliments pour animaux, les techniques d'abattage, etc. Tous les points critiques de l'élevage des bovins sont pris en compte ainsi que ceux concernant la consommation humaine de tissus à risque (cervelle, moelle épinière, etc.). Ainsi, sont pris en compte dans cette modélisation tant les paramètres qui portent sur l'estimation du risque (le nombre d'animaux importés pouvant propager la maladie, le recyclage des carcasses), que l'impact et les défaillances des mesures mises en place pour prévenir l'ESB (essentiellement l'interdiction des farines animales pour les bovins) ainsi que la consommation humaine de tissus infectieux.

Les conclusions tirées de cette modélisation sont claires : la probabilité de développement d'une épidémie d'ESB aux États-Unis apparaît quasi nulle, rendant fort peu plausible cette hypothèse, même si un nombre assez conséquent d'animaux infectés par l'ESB était introduit sur le territoire, compte tenu des mesures préventives mises en place. Autrement dit, si la première barrière était franchie (malgré l'interdiction d'importer des ruminants provenant de pays atteint par l'ESB), la seconde, représentée par les mesures mises en place, rendrait quasiment impossible la diffusion de l'épidémie. Néanmoins des mesures de précaution quant à la contamination humaine pourraient être envisagées en écartant de la consommation certains matériaux à risque spécifiés.

3 | La modélisation prévisionnelle épidémiologique entre besoin et pertinence

Ces deux cas brièvement résumés ont en commun deux points essentiels : à la suite d'une demande publique d'expertise, une modélisation a été produite en l'absence de risque observé et la forme d'action qui en est issue a été acceptée par les gestionnaires politico-administratifs sur cette base. Malgré leur singularité apparente, les deux cas exposés soulèvent la question de la généralisation du recours à la modélisation face au risque, plus précisément des conditions qui la rendent possible ainsi que celle de son degré de pertinence en relation avec

sa force de conviction. La nature du problème posé et sa perception profane représentent une première dimension déterminante.

Les problèmes que figurent ces deux cas relèvent de la sécurité sanitaire. La nature de ces deux risques, véhiculés par des produits (sang et viande), fait qu'ils relèvent de la responsabilité exclusive et impérative des autorités publiques sanitaires, plus largement de l'État. Le public se trouve confronté à un risque échappant à son contrôle et dont les produits-sources sont soumis à la régulation des autorités en vue d'assurer leur sécurisation. Ce type de risque est foncièrement différent de ceux qui sont construits par les comportements individuels et pour lesquels l'exercice de la responsabilité politique est pour l'essentiel indirect, de l'ordre de l'information ou de l'incitation, tant sur les facteurs de risque que sur les moyens pour réduire le risque. Quand de plus les effets, réels ou supposés, sont effrayants (la MCJ est mortelle et son tableau clinique redoutable), reportés dans le temps (durée d'incubation estimée entre 15 et 25 ans) et susceptibles de concerner l'ensemble de la population, la responsabilité des pouvoirs publics est de les anticiper aussitôt que cela est jugé possible. Cette contrainte impérieuse rend caduque le modèle préventif fondé sur les résultats de l'observation rétrospective des effets morbides ou mortels (incluant les études cas-témoins) ou prospective (cohortes), et sur la démonstration épidémiologique et/ou biologique d'une relation de causalité avec le produit-source.

Le besoin d'action devient dans de telles conditions un impératif d'ordre social, politique et économique, quelle que soit par ailleurs sa dimension sanitaire et l'incertitude qui la caractérise. L'action ne peut plus être conditionnée à la certitude de causalité, du fait que l'absence d'intervention ne peut être justifiée sur cette base : aussi, prendre des mesures qui s'avéreront inutiles à terme est considéré comme bien plus acceptable que d'avoir à assumer les conséquences négatives de l'inaction. L'émergence et le succès sociopolitique du principe de précaution, aussi discutable soit-il dans ses applications, expriment clairement la réalité d'une demande sociale de protection à laquelle les experts scientifiques ne peuvent rester insensibles. Ce sont bien les contraintes d'ordre sociopolitique et aussi économique (les conséquences de la perception du risque par le public en tant que consommateur de viande bovine dans le cas de l'ESB) qui ont rendu possible la transition de l'exigence de démonstration causale au critère de « possibilité » sans causalité, ni même d'effets statistiques. S'y ajoutent des conditions contextuelles qui ne sont pas propres au risque lui-même ou à la connaissance disponible.

Concernant le cas du danger de transmission sanguine du vMCJ, la seconde dimension contextuelle qui permet de dépasser l'absence d'observation est qu'il s'agit d'un risque affectant la transfusion sanguine, secteur hypersensible du système de santé : l'hypothèse de la transmission est retenue malgré la fragilité des preuves disponibles, tant par les experts que par les gestionnaires. Cette posture

s'impose d'autant plus qu'elle est internationale, la France ne faisait que suivre l'exemple des États-Unis et du Canada. Cette unanimité tient à la capacité de la décision à satisfaire à la triple menace, sanitaire, politique et judiciaire propre à ce domaine dans le cas où la transmission sanguine du vMCJ s'avérerait réelle. L'important est que la mise en évidence du risque (causalité et probabilité) n'est plus un préalable à l'action. Dans le second cas, celui du risque d'ESB aux États-Unis, c'est également le souci d'anticiper une probable catastrophe économique et politique qui justifie la démarche.

Dès lors, débarrassé de l'épineuse question de la démonstration scientifique du risque, le recours à la modélisation face à ce qui correspond jusqu'ici à un danger devient à la fois pertinent et incontournable. La modélisation permet ainsi de répondre de façon rationnelle, transparente et scientifiquement fondée à ce besoin d'action (cas de la transmission sanguine du vMCJ) ou de renforcement de l'action (cas de l'ESB aux États-Unis).

Dans les deux cas, la modélisation permet de retenir les actions les mieux adaptées au contexte propre au risque, c'est-à-dire qui tiennent compte des données spécifiques du pays : exposition de la population, pratiques sociales en relation avec le produit-source, process industriel, modes de circulation des produits, etc. Comme la causalité n'est plus discutée, ni même l'estimation quantifiée du risque s'il était réel, l'apport de la modélisation est d'établir autant de scénarios possibles de mesures anticipatrices pour choisir le plus souhaitable. Il va de soi que la modélisation n'aurait plus d'intérêt si un consensus se dégageait pour supprimer le produit ou l'activité en question, ce qui a été considéré comme hors de question pour les produits en jeu dans les deux cas. En fait, la modélisation permet d'opérer un retour au modèle préventif qui prend en compte, dans le choix du mode d'action, les pertes de bénéfices et les réductions du risque. Ainsi, l'exclusion des donneurs de sang ayant séjourné plus de 6 mois dans les Iles Britanniques s'avère une action rationnelle aux États-Unis et au Canada, mais pas en France où son coût (en sang) est jugé trop élevé pour des bénéfices réduits ; protéger les receveurs de produits sanguins autant que possible conduisant à choisir la déleucocytation. Dans le second cas, la modélisation faite aux États-Unis permet de rendre inutile de nouvelles mesures face à un développement jugé improbable de l'ESB dans ce pays.

4 | Le champ de la modélisation en relation avec sa force de conviction

Dans une perspective scientifique, les différentes séquences qui composent le cadre de l'analyse du risque peuvent être en théorie l'objet de modélisation :

estimation du risque, évaluation, décision et choix des modalités d'action. Il ne semble pas en être de même dans le champ de l'action publique. Il ne s'agit pas ici de traiter de la validité scientifique des modèles en fonction de leur champ d'application mais de cerner le champ de la modélisation en relation avec sa force de conviction potentielle sur les gestionnaires du risque. Or, au regard de leur force de conviction, les critères de qualité scientifique des modèles prévisionnels élaborés face au risque sont différents de ceux qui assurent leur pertinence opérationnelle. L'indicateur rendant compte de la force de conviction d'un modèle est représenté par son impact sur l'action retenue (ou l'inaction) et non pas simplement par l'adhésion intellectuelle qu'il suscite.

L'usage d'un modèle prévisionnel épidémiologique est déterminé par le besoin d'action mais le modèle ne détermine qu'exceptionnellement l'action : il fournit une ou des réponses à un besoin construit sur d'autres bases. En tant que démarche scientifique au service de l'action, la modélisation s'inscrit bien dans le champ de l'expertise. Il est rare qu'une action publique lourde de conséquences incertaines puisse être le seul fait d'une modélisation, aussi reconnue soit-elle dans le champ scientifique. Le processus est plutôt le suivant : il faut qu'un problème émerge dans l'espace public, qu'il soit perçu (par le public comme par les pouvoirs publics) comme suffisamment important pour être inscrit sur l'agenda des gestionnaires du risque et que soit posée en conséquence la question de l'action protectrice. C'est bien parce que le problème est à la fois présent et qu'il soulève la question de l'action que le modèle devient potentiellement pertinent.

Sa pertinence vient d'abord de sa capacité à s'inscrire dans les valeurs dominantes qui encadrent la gestion du risque dans la société : le modèle, en tant que proposition d'action, ne remet pas en cause l'objectif de protection de la population, il ne vient qu'en sélectionner les modalités. Autrement dit, sa rationalité mathématique (exposition de la population, facteurs de risque quantifiés, impact réducteur des mesures) est dépendante de la rationalité politique qui justifie, dans le cas de la transmission sanguine du vMCJ, le recours à la précaution. Sa force de conviction vient de sa capacité à rendre compte de la rationalité sanitaire et des conséquences de l'action choisie : réduire le risque le mieux possible sans pour autant sacrifier des ressources sanitaires rares (le sang).

À l'inverse, un modèle élaboré en vue de produire une estimation du risque ne parvient que très rarement à convaincre de sa pertinence au point de déterminer l'action conséquente ou de modifier des actions établies sur la base d'estimations supérieures. Au cours des dernières années, le risque de vMCJ a été estimé à plusieurs reprises et les résultats publiés dans les meilleures revues scientifiques internationales ; ces estimations de l'incidence ont progressivement évolué vers de nouvelles, réduites de façon massive. Or, leur impact direct sur les

deux décisions majeures prises en 2000 s'est avéré imperceptible : l'interdiction totale des farines animales et le dépistage systématique des bovins âgés de plus de 26 mois destinés à la consommation. Est-ce pour autant que ces modèles ne sont pas crédibles ? La réponse est probablement autre : les modèles d'estimation d'incidence ne sont pas pertinents pour l'action parce que l'incidence observée ne l'est pas également. Si le critère quantitatif n'est pas déterminant pour engager l'action, il l'est encore moins pour réviser à la baisse un haut niveau de protection lorsqu'il subsiste un risque résiduel, aussi bas soit-il.

5 | Vers un développement nécessaire et problématique de la modélisation épidémiologique

La modélisation n'a pas pour vocation d'entrer en concurrence directe avec les méthodes épidémiologiques classiques, qu'elles soient observationnelles, analytiques ou quasi expérimentales. Elle représente un outil supplémentaire qui devrait être utilisable dans l'action publique face à certains risques, sur certaines séquences du processus d'analyse du risque, pour lesquelles ces méthodes sont impuissantes ou inadaptées. Mais son usage au service de la gestion du risque ne peut être ni systématique ni uniforme : il dépend de la nature du risque et du contexte sociopolitique dans lequel il s'inscrit. Son utilisation potentielle est sélective, en rapport avec certaines séquences ou questions du processus d'analyse du risque et ne peut aboutir sans interactions avec les gestionnaires du risque. Son développement à venir reste conditionné aux réponses qui seront apportées à plusieurs questions :

- Quel est son champ potentiel d'utilisation ?
- Quels objectifs accessibles peuvent être visés par la modélisation ?
- Quels sont les critères ou les conditions qui rendent son utilisation envisageable ?
- Comment sont perçus les modèles par les gestionnaires publics ?
- De quelle nature sont les réticences ou les freins à son utilisation dans l'action publique ?
- Comment développer la production de modèles épidémiologiques dans la perspective de leur utilisation ?

À ces questions il n'existe pas de réponses théoriques globales, mais tirées de l'observation et de l'analyse empirique. Cela suppose des recherches capables de rendre compte de tous les paramètres évoqués en testant leur potentiel de généralisation en fonction de différentes catégories de risques.

Conclusion

Le développement de la modélisation épidémiologique répond pour partie à l'évolution récente constatée dans les pays démocratiques développés confrontés à certains risques émergents inobservables, diffus et inquiétants. Cette tendance lourde s'explique par l'émergence d'une demande sociale d'action publique de plus en plus précoce qui se heurte à la fragilité de la connaissance disponible pour en déterminer tant le besoin que les formes concrètes. Le succès du principe de précaution, dont la nature est politique et non pas scientifique, trouve dans la modélisation un moyen d'en rationaliser l'application.

L'expérience récente (l'ESB et son risque humain de vMCJ, les OGM) indique que le besoin d'action peut coexister avec une connaissance ténue, tant en termes quantitatifs qu'en termes de causalité. Si l'on peut s'interroger sur la robustesse d'une expertise produite en réponse à une demande des autorités publiques dans un contexte cognitif fragile, la question centrale reste celle de sa pertinence et de sa qualité scientifique. La pertinence de la modélisation prévisionnelle épidémiologique se fonde sur son adéquation, d'une part au problème tel qu'il se pose à un moment donné, d'autre part à la question qu'elle prétend éclairer. Sa qualité scientifique se juge dans la cohérence des hypothèses en rapport avec les (rares) données disponibles et sur la transparence du cheminement qui en assure la reproductibilité. Néanmoins, sa force de conviction n'est pas simplement en relation linéaire avec sa qualité scientifique, du fait que le modèle acquiert une dimension décisionnelle quand il s'insère dans le processus d'expertise publique, tout comme n'importe quelle connaissance, épidémiologique ou autre.

Il est vraisemblable que l'utilisation de modèles épidémiologiques par les gestionnaires du risque est aléatoire et inconnue dans ses déterminants. La recherche de ses conditions d'usage et des réticences qui l'accompagnent devrait permettre de faire communiquer deux univers fondés sur des rationalités différentes : la science et le politique, réunis autour de l'expertise. Comprendre comment ces acteurs ponctuellement interdépendants construisent leurs produits, les uns les modèles, les autres les décisions, apparaît comme l'une des conditions pour rendre compte des critères qui structurent l'offre et la demande de modèles, préalable à une harmonisation plus féconde. Avancer dans cette direction implique de développer des recherches sur la modélisation épidémiologique associant plusieurs disciplines (épidémiologie, sciences biomédicales, sociologie, économie), ce qui suppose une prise de conscience sur leur nécessité.

Bibliographie

Afssaps, Rapport de synthèse (février 2000). *Révision des mesures de réduction du risque de transmissions des ESST par les produits sanguins.*

- Aguzzi A., Glatzel M. (2004). vCJD tissue distribution and transmission by transfusion-a worst-case scenario coming true? *Lancet*, **7** : 363 (9407) : 411-2.
- Gray G. *et al.* (2002). Harvard Center for Risk Analysis. Evaluating the risk of bovine spongiform encephalopathy in the United States, *Risk in Perspective*.
- Lewelyn CA., Hewitt PE., Knight RS., Amae K., Cousens S., Mackenzie J., Will RG. (2004). Possible transmission of variant Creutzfeldt-Jakob disease by blood transfusion. *Lancet*, 363 (9407) : 417-21.
- Peden AH., Head MW., Ritchie DL., Bell JE., Ironside JW. (2004). Preclinical vCJD after blood transfusion in a PRNP codon 129 heterozygous patient. *Lancet*, **7** : 364 (9433) : 527-9.
- Pincock S. (2004). Patient's death from vCJD may be linked to blood transfusion. Committee to discuss need for further precautions to prevent possible vCJD transmission through blood. *Lancet*, **3** : 363 (9402) : 43.

CHAPITRE 9

Les conditions nécessaires au développement de l'épidémiologie

SOUS-CHAPITRE 9.1

La formation en épidémiologie humaine à l'université

ROGER SALAMON

Nous ne parlerons ici que d'épidémiologie humaine.

Avant d'aborder cet important chapitre, il convient de préciser un certain nombre de points qui éclairent le positionnement de l'épidémiologie en France.

1 | Préalables

1.1 Le positionnement académique de l'épidémiologie

Sur un plan académique, l'épidémiologie apparaît dans une sous-section (46.01) du Comité national des universités (CNU) intitulée « Épidémiologie, économie de la santé – prévention ».

Trois autres sous-sections (46.02 : Médecine légale, 46.03 : Médecine du travail et 46.04 : Biostatistique et sciences de la communication) complètent la section 46 intitulée « Santé publique ».

De ce constat, on peut déjà tirer deux enseignements qui expliquent en partie les difficultés du développement de l'épidémiologie en France :

- l'épidémiologie apparaît exclusivement dans une section médicale du CNU ;
- l'épidémiologie n'est pas individualisée dans une sous-section du CNU mais associée à l'économie de la santé et prévention et séparée de la biostatistique.

1.1.1 Une section médicale du CNU

Les enseignants nommés par le CNU sont donc hospitalo-universitaires, tous docteurs en médecine (à quelques exceptions près), avec une activité à la fois universitaire et hospitalière.

Leurs étudiants sont des étudiants en médecine appartenant à des UFR médicales dont les programmes sont nationaux et peu flexibles.

1.1.2 L'épidémiologie n'est pas individualisée

Alors même qu'il s'agit de la discipline de base de la santé publique, elle n'est pas individualisée mais associée à l'économie de santé (en soi déjà une énorme discipline) et à la prévention (dont on comprend assez mal le sens dans un CNU).

Plus encore, elle est séparée de la biostatistique qui en est la discipline la plus proche et qui regroupe dans sa sous-section la plupart des meilleurs épidémiologistes français qui, directement ou indirectement issus de l'école du professeur Schwartz, ont une origine dans la statistique médicale.

Cela explique pourquoi, et nous le verrons dans la suite de ce chapitre, l'épidémiologie est implicitement assimilée à la « biostatistique » (cela est surtout le cas dans les UFR médicales) ou plus généralement intégrée à la « santé publique » (ce qui est retrouvé hors des UFR médicales).

Il s'agit donc en résumé d'une discipline qui semble exclure tout autre étudiant que ceux de médecine et qui se retrouve positionnée de manière fort discutable au CNU.

1.2 La faiblesse de la formation : un faux procès

Il est de bon ton en France de rendre responsable la formation des difficultés ou faiblesses des pratiques d'une discipline.

Ainsi entend on souvent des jugements sévères sur la formation en épidémiologie et plus généralement en santé publique, dont la mauvaise qualité serait responsable des difficultés à surveiller, prévoir, prédire, prévenir des phénomènes de santé inquiétants.

Les catastrophes et risques médiatisés en France dans les quinze dernières années ont bien sûr exacerbé ces reproches.

Sans pour autant considérer la formation en épidémiologie comme étant sans reproche ou d'excellente qualité (bien au contraire nous reviendrons plus bas sur ces faiblesses), il est nécessaire de cesser ce contre-sens mélangeant la cause et l'effet. C'est en grande partie parce que la « fonction » et le « statut » des épidémiologistes ont été pendant longtemps insuffisamment précisés et garantis, que la formation, suivant bien évidemment la demande, fut ce qu'elle fut.

D'ailleurs on peut aisément constater que, depuis les dernières années où un recrutement substantiel d'épidémiologistes a été réalisé, en particulier dans les agences, la formation en épidémiologie en France s'est améliorée sur les plans quantitatif et qualitatif.

1.3 La formation : une grande variété

Que l'épidémiologie soit considérée ou non comme une science, elle est surtout une discipline qui se « pratique » et cela dans les domaines variés que l'on peut distinguer en deux groupes :

- Groupe 1 : l'épidémiologie intervient comme complément naturel à d'autres disciplines (médecine, économie, gestion...) pour des fonctions autres qu'épidémiologistes (cliniciens, chercheurs, médecins de santé publique, gestionnaires de structures de soins, planificateurs...).
- Groupe 2 : l'épidémiologie est l'objet unique ou essentiel pour des chercheurs en épidémiologie, des enseignants ou des spécialistes de l'investigation ou de la surveillance.

Il y a donc très clairement deux niveaux de formation :

- l'un basique pour le premier groupe ;
- l'autre plus approfondi pour le deuxième.

Dans les deux cas, il faut noter que seule la « pratique » pourra conforter et compléter une formation initiale.

Cela signifie que l'on ne forme pas un épidémiologiste comme un physicien ou un mathématicien. La formation d'un épidémiologiste est, par obligation, de nature continue, et sans pratique, la formation initiale sera rapidement oubliée. Cela explique ainsi que l'épidémiologie soit parfois tardivement enseignée ou ré-enseignée à un public adulte souvent déjà en fonction et nous verrons la place importante des formations continues dans ce domaine.

Au terme de ces préalables, nous présenterons la situation française en distinguant :

- la formation à et par la recherche ;
- la formation en médecine ;
- la formation en santé publique, (hors de la médecine) ;
- la formation continue.

Et nous terminerons par un bref aperçu de la situation en l'étranger.

2 | En France

Le constat que l'on peut dresser en France est le suivant :

- faiblesse de structure de formation à et par la recherche de l'épidémiologie ;
- faiblesses des structures de formation en santé publique « généraliste » ;
- présence insuffisante de l'épidémiologie dans les facultés de médecine, quasi inexistante dans les autres facultés.

2.1 La formation à et par la recherche en épidémiologie

Il y a peu de DEA (ou maintenant de masters recherche) en France préparant à la recherche en épidémiologie. Les quatre principaux sont : celui de Paris VI-VII (biomathématiques), le premier créé en France en 1976 et qui est surtout orienté vers la modélisation en épidémiologie ; celui de Paris XI (santé publique), qui date des années 1980 et comprend toutes les disciplines de l'épidémiologie, y compris une ouverture aux sciences sociales ; dans les années 1990 se sont ajoutés celui de Bordeaux II (épidémiologie et biostatistique), avec une spécialisation en épidémiologie d'intervention, et le DEA de Nancy, plus axé sur l'évaluation et la recherche clinique.

Les thèses se préparent soit dans des écoles doctorales spécialisées, soit dans des écoles doctorales de site. Les deux écoles doctorales spécialisées sont à Paris (Paris VI-VII et Paris XI). Parallèlement, de nombreuses écoles doctorales de site accueillent des thèses d'épidémiologie, la principale est celle de Bordeaux II qui profite de l'amont de son DEA (maintenant master recherche).

Cette situation reflète la faiblesse du nombre d'équipes d'accueil possibles. Celles-ci ne sont en nombre suffisant qu'à Paris (Centre et Sud) et à Bordeaux. Dans les autres cas, les initiatives pédagogiques en termes de formation à la recherche sont difficiles, nécessitant l'agrégation d'équipes éloignées (c'est le cas de Nancy), ou la duplication.

Malgré cette faiblesse « structurelle », la position internationale de la recherche épidémiologique française est très loin d'être aussi négligeable que certains se complaisent trop souvent à le dire.

En termes de doctorants (il y a environ une cinquantaine de thèses passées en France par an), de publications internationales, de participation à des colloques internationaux, la situation de l'épidémiologie en France est tout à fait « convenable » et certainement digne d'éloges si l'on se réfère aux faiblesses des structures sous-jacentes.

2.2 Santé publique

Il existe aujourd'hui en France, en dehors des UFR médicales, deux structures de formation axées prioritairement sur la santé publique : il s'agit d'une école professionnelle dépendant du ministère de la Santé, l'École nationale de la santé publique (ENSP), située à Rennes, et de l'Institut de santé publique, épidémiologie et développement (Isped), situé à l'université Victor Segalen Bordeaux II.

L'ENSP ne délivre pas de diplôme universitaire et forme essentiellement des cadres hospitaliers. Elle ne comporte pas d'équipes de recherche labellisées, et est donc en dehors de la formation à et par la recherche. L'épidémiologie ne représente pas une discipline majeure de cette école (plus axée sur le droit et la gestion).

L'Isped est une structure publique, régie par l'article 33 de la loi Savary, faisant partie de l'université Victor Segalen Bordeaux II. Il est prioritairement axé sur des formations professionnelles (nous y reviendrons) mais a cependant une activité de formation à et par la recherche (cf. plus haut), grâce à son DEA (maintenant master), et à une forte implantation dans l'école doctorale de site de Bordeaux II.

Il existe néanmoins de manière fort diversifiée des enseignements universitaires qui, directement ou indirectement, intègrent l'épidémiologie ; mais, en dehors de quelques pôles sur lesquels nous reviendrons (Paris, Nancy, Bordeaux), il s'agit de formations assez éparpillées qui n'offrent pas aux étudiants un véritable parcours universitaire.

On y retrouve ainsi dans certaines villes des modules d'enseignements d'épidémiologie dans telle ou telle licence des sciences sanitaires et sociales ou dans certains modules des masters recherche (tel que le master Sciences, technologies, santé de l'université Claude Bernard Lyon I) ou des masters professionnels (tel le master professionnel Nutrition en santé publique de l'université de Montpellier I).

Tout cela donne une image d'une situation fragile et parsemée de l'enseignement de l'épidémiologie hors des UFR médicales.

À cette inquiétante fragilité, bien explicable car nous avons vu que l'épidémiologie est en France sur un plan académique une discipline « médicale », trois « sites » néanmoins émergent, sur lesquels nous pouvons mettre l'accent ; il s'agit de : Paris, Bordeaux et Nancy.

À Paris, les forces sont importantes et les efforts en matière d'enseignement ont été anciens.

Pendant longtemps ces efforts ont été faits avec une forte orientation « recherche » avec des maîtrises de Biologie humaine, des DEA et deux écoles doctorales.

C'est aussi dans ce même contexte qu'a été mis en place à Paris VI, à l'initiative du professeur Schwartz, un enseignement de statistique présentiel et par correspondance, connu et apprécié dans le monde entier : le Cesam.

Néanmoins, en dehors d'un troisième cycle de qualité et ouvert à tout candidat, quelle que soit sa formation initiale, les formations d'amont (deuxième cycle) furent longtemps offertes prioritairement aux secteurs de la santé (la MSBM, seule formation nationale de deuxième cycle, n'est permise qu'aux médecins, odontologistes et pharmaciens).

Plus récemment, d'autres formations ont été mises en place, en particulier une maîtrise des Sciences sanitaires et sociales et plusieurs DESS, en particulier dans le domaine de la biostatistique ou de l'économie.

À Bordeaux, dès 1989, une orientation claire et originale s'est affirmée sur l'idée que la santé publique devait quitter, sur un plan académique, son « exclusivité » médicale.

Il fut alors créé une UFR (au sens commun, c'est-à-dire non hospitalo-universitaire) de santé publique qui s'est transformée en 1997 en un institut, article 33 de la Loi Savary, dénommé Institut de santé publique, d'épidémiologie et de développement (Isped) ouvert à un public non médical et offrant un panel assez complet d'une formation universitaire de santé publique dans le cadre du LMD, avec une licence de Santé publique, plusieurs masters professionnels et un master recherche, et une option de l'école doctorale de l'université de Bordeaux II (site Web : isped.u-bordeaux2.fr).

Avec près de 900 étudiants par an (cela sans compter les étudiants en médecine), l'Isped représente une structure originale et malheureusement unique en France.

À Nancy, malgré un positionnement dans une UFR médicale, le Laboratoire de santé publique a réussi à s'ouvrir à des publics plus larges et s'est d'ailleurs dans cet esprit dénommé « École de santé publique ». Il a développé de fortes relations régionales (Dijon, Strasbourg. . .).

Au total, en dehors des UFR médicales et dans un cadre généraliste de santé publique, on retiendra :

- deux pôles forts : Paris et Bordeaux, sans négliger un effort régional autour de Nancy ;
- une seule structure universitaire, l'Isped à Bordeaux ;
- un éparpillement national de quelques formations.

Enfin, depuis peu de temps, certaines écoles d'ingénieur offrent des formations directement ou indirectement (par le biais des biomathématiques le plus

souvent) associées à l'épidémiologie : c'est le cas notamment de l'Agro, de Centrale, du Cnam. . .

2.3 Médecine

Dans les UFR médicales, en France, l'enseignement de l'épidémiologie est souvent rapproché, sinon confondu, avec la statistique.

Pourtant, le mot statistique n'apparaît jamais officiellement sur les programmes des premiers et deuxièmes cycles des études médicales.

En réalité, au gré des enseignants et des UFR, un enseignement de statistique est réalisé parfois dans le cadre du 1^{er} cycle (sous les intitulés divers tels que « Maths-physique » ou « Bases du raisonnement en santé publique ») alors que les bases de l'épidémiologie font partie du programme du 2^e cycle des études médicales.

En premier cycle, lorsque l'enseignement de statistiques est réalisé (pour la moitié environ des UFR), les objectifs attendus sont assez généralement les mêmes :

- expliquer la différence entre les concepts d'échantillon et de population ;
- décrire un échantillon par quelques techniques élémentaires : histogrammes, calcul de paramètres de position, de dispersion ;
- apprécier les notions de fluctuation d'échantillonnage d'une moyenne ou d'une fréquence ;
- réaliser une estimation par intervalle de confiance ;
- connaître les principes des tests statistiques et savoir réaliser quelques tests les plus simples : χ^2 , comparaison de moyennes.

En général, le domaine de la corrélation entre deux variables, l'analyse de variance, les tests non paramétriques ne sont pas traités à ce niveau.

En deuxième cycle, le programme comprend de manière officielle les éléments suivants :

- épidémiologie et santé publique (DCEM 1) ;
- recherche clinique et épidémiologique (DCEM 1) ;

- base de l'analyse critique d'un article médical (DCEM 1) ;
- santé et environnement (DCEM 2) ;

dont il faut reconnaître qu'ils ne représentent guère la discipline préférée des étudiants en médecine !

Parmi les enseignements de spécialisation, on retrouve :

- le DES de Santé publique ;
- les certificats de MSBM, en cours de disparition ;
- les DEA ou masters recherche (que nous avons déjà évoqués).

Le DES de Santé publique concerne des étudiants en médecine qui se destinent à une spécialisation en santé publique. Sur une durée de quatre ans avec des stages hospitaliers ou extra-hospitaliers, la formation, variable d'une région à l'autre, apporte généralement à des étudiants, motivés et futurs spécialistes, de bonnes bases en épidémiologie (souvent par le biais des certificats du MSBM — maîtrise des Sciences biologiques et médicales — ou DEA ou masters).

Les certificats du MSBM, organisés dans la plupart des UFR médicales par les enseignements de santé publique, sont généralement appréciés et concernent des étudiants assez divers (futurs spécialistes de santé publique ou cliniciens ou biologistes intéressés par la méthodologie et la recherche clinique).

Ces enseignements ont été soumis à d'incessantes remises en questions, dont la dernière en date (le système LMD) pourrait conduire à certaines difficultés d'organisation dans les années à venir pour certaines UFR.

En résumé on peut retenir pour les UFR médicales :

- dans le cursus normal des études médicales : un enseignement insuffisant, ponctuel et très inhomogène ;
- dans le cadre optionnel : un bilan beaucoup plus satisfaisant mais des inquiétudes pour l'avenir.

Quels sont les facteurs qui favorisent la grisaille de ce bilan et, en particulier, que reproche-t-on généralement à la statistique médicale ?

Ce sont les étudiants d'abord qui s'inquiètent d'être sélectionnés au concours de première année sur une matière leur paraissant trop théorique. Parfois, des enseignants leur donnent raison.

Ce jugement peut s'expliquer et se corriger ; il est entretenu par le souvenir de mauvais cours, trop théoriques le plus souvent, sans référence au domaine médical.

De ces cours volontiers ennuyeux (encore qu'agrémentés par de distrayants chahuts), difficiles à comprendre, aussi vite oubliés que péniblement appris, est née cette mauvaise réputation. On voit là que c'est plus l'enseignant que l'enseignement qui est jugé et nous ne retiendrons pas cette critique, d'autant moins d'ailleurs qu'elle s'adressait généralement à des enseignants qui n'étaient pas toujours issus de nos disciplines.

Une autre opinion, volontiers admise et plus grave celle-là, concerne la non-nécessité, sinon l'inutilité, de la méthodologie statistique en médecine.

Nous rentrons alors vraiment dans le fond de l'affaire. La médecine serait un art qui serait décadent s'il était dénaturé par un minimum de méthodologie. L'esprit médical serait « un esprit de finesse » ne s'accommodant pas de la méthode « numérique ».

C'est la victoire de Claude Bernard sur nos grands maîtres Pierre Charles Alexandre Louis (1787-1872) et Jules Gavarret dont l'ouvrage *Principes généraux de statistique médicale*, a été, ne l'oublions pas, publié à Paris en 1840.

À cela on peut répondre et démontrer que la méthodologie statistique est tout à fait fondamentale pour l'étudiant en médecine quelle que soit sa future activité.

Pendant sa période d'apprentissage, l'étudiant doit savoir apprécier un protocole d'étude, critiquer la notion de valeur de référence, dépister les biais et discuter les résultats d'une étude. En un mot, il doit savoir lire un article, il doit déjà avoir l'**esprit critique**, il doit apprendre, à côté de toutes les connaissances innombrables dont on veut le « gaver », la manière de manipuler ces dernières ; il doit acquérir ce que les Anglais appellent si bien « *A way of thinking* ».

À la fin de ces études, l'étudiant aura besoin, quelle que soit son activité, de la méthodologie statistique :

- s'il est chercheur, c'est évident ;
- s'il est un professionnel de santé publique, c'est aussi évident, même si ces derniers ne partagent pas encore tous cette évidence ;

- s'il est médecin praticien, il sera confronté quotidiennement à des problèmes de décision dans des conditions d'incertitude et sera soumis à l'extrême variété des cas particuliers que sont chacun de ses malades.

Tout le monde ne pourra qu'être convaincu par ces propos, mais il faut aller encore plus loin et insister sur deux autres notions :

- la médecine change : son support même devient statistique. De plus en plus, l'analyse physiopathologique cède le pas à la constatation fréquentielle et à l'analyse probabiliste ;
- pour cette raison et de la même manière qu'un des critères du caractère éthique d'un essai thérapeutique est qu'il s'appuie sur une méthodologie rigoureuse, on pourrait dire que la méconnaissance ou la mauvaise utilisation de la statistique en médecine présente toutes les caractéristiques d'une activité non éthique.

Il reste encore d'autres difficultés qui tiennent à la statistique elle-même :

- et d'abord à la difficulté qu'il y a à l'enseigner. C'est une discipline très difficile à enseigner, surtout si l'on veut dépasser les recettes et il est aussi très dangereux de mal l'enseigner, ce qui peut conduire à des situations regrettables telles que :
 - un « dégoût » : l'étudiant n'y reviendra pas. . . ,
 - un trop grand « doute » : peur de conclure, peur de décider. . . ,
 - une situation de « suffisance » : on peut tout faire tout seul. . . ;
- sans pratique, les méthodes statistiques s'oublient très vite ;
- reste enfin, si ces dernières difficultés étaient dépassées, à trouver le cadre idéal d'un tel enseignement :
 - trop tôt placé (premier cycle par exemple) il risque de ne pouvoir s'appuyer sur des exemples médicaux que ne pourraient comprendre alors les étudiants,
 - plus tard, lorsque l'étudiant baigne enfin dans la médecine, va à l'hôpital, prépare l'internat, il est difficile et parfois trop tard, de l'intéresser à une matière jugée trop fondamentale et éloignée des soins.

Sans doute la solution consiste-t-elle à introduire la statistique tout au long des études plutôt que de manière ponctuelle.

Mais, pour cela, il faut un relais au travers des autres enseignants, dans le cadre des certificats intégrés de pathologie, permettant ainsi à l'étudiant d'appliquer la méthodologie de la statistique et de l'épidémiologie aux diverses disciplines médicales au fur et à mesure de leur apprentissage.

Cela suppose que les cliniciens responsables de ces certificats souscrivent à cette proposition et fassent l'effort nécessaire.

Cela implique qu'ils aient eux-même conscience de l'intérêt de la statistique dans leur propre pratique et dans leur enseignement.

Cela conduit à souhaiter qu'ils soient eux-mêmes formés à cette discipline qu'en général ils ignorent.

Cet effort est à la base de tout progrès ; il s'apparente, ce qui rend aléatoire son succès, à un changement de mentalité de la profession médicale en général.

2.4 La formation continue

L'épidémiologie est une discipline très appréciée dans le cadre de la formation continue, vraisemblablement sous l'effet conjoint que les médecins ne comprennent en général que tardivement l'intérêt de cette discipline (et plus généralement de la santé publique) et qu'ils y ont été assez mal formés (en particulier parce que peu intéressés) ou qu'ils ont tout oublié car sans pratique, comme dit en préalable de ce chapitre, l'épidémiologie se perd rapidement.

Il y a bien de très nombreuses et diverses formations qui répondent à ce besoin, le plus souvent sous forme de diplômes d'université (DU) et l'exhaustivité est impossible.

Trois formations méritent d'être soulignées :

- le Cesam, déjà évoqué, est certainement le DU le plus connu, le plus ancien et le plus suivi ;
- un DU d'épidémiologie sur Internet a été mis en place en 2003 par l'Ispep à Bordeaux, avec immédiatement un très important succès dans de très nombreux pays francophones ;
- un DU d'évaluation des risques thérapeutiques (Paris VI-VII) ou un DIU de pharmaco-épidémiologie regroupant plusieurs universités et coordonné à Bordeaux, qui témoignent de l'importance croissante de cette thématique.

Enfin, dans une démarche de type anglo-saxon, de nombreuses « écoles d'été » ont inscrit en France l'épidémiologie à leur programme : la plus connue est celle mise en place à Veyrier-du-Lac par la fondation Mérieux dans le champ de ce qu'il est convenu d'appeler l'épidémiologie d'intervention, qui a conduit à initier en ce domaine deux à trois cents professionnels, travaillant généralement dans des agences ou des collectivités territoriales. Il existe aussi, parmi les écoles d'été de bonne réputation, celle de Paris-Sud et celle de l'Ispeid à Bordeaux.

3 | À l'étranger

De même que pour la France, dans les pays occidentaux comparables, l'épidémiologie est enseignée dans deux contextes forts différents : celui de la santé publique et celui de la médecine.

Il n'a pas été possible, dans le cadre de ce rapport, d'analyser comparative-ment les programmes de médecine de l'ensemble des pays auxquels la France pourrait se comparer. Dans de nombreux pays occidentaux, l'épidémiologie est mieux « considérée » qu'en France dans le cursus médical, soit pour permettre à l'étudiant d'améliorer ses connaissances (lecture critique d'articles) soit pour aider à la décision (médecine fondée sur les preuves).

En ce domaine, nous supportons encore les séquelles d'une tradition médicale très clinique (et d'ailleurs très riche et longtemps enviée) qui trop souvent négligeait les possibles apports de l'analyse décisionnelle et de l'épidémiologie.

Pour la santé publique, notre déficit est patent : il existe aux États-Unis, au Canada, au Royaume-Uni, aux Pays-Bas, en Belgique et dans certains pays d'Amérique latine des « écoles » de santé publique formant des professionnels aux disciplines de la santé publique dont l'épidémiologie est une composante majeure.

Ces écoles, souvent d'excellente réputation, ont un public volontiers international et sont majoritairement privées (et d'un coût non négligeable).

SOUS-CHAPITRE 9.2

Épidémiologie, enseignement et éducation à la santé au lycée et au collège

ANNIE MAMECIER

L'épidémiologie humaine a une place modeste dans les enseignements des écoles, des collèges et des lycées. Son apport à l'éducation à la santé des élèves et à la formation des professeurs devrait être davantage pris en compte. L'expérience française et quelques situations dans les pays étrangers sont ici rapprochées.

- 1 | L'épidémiologie humaine dans les programmes d'enseignement des écoles et des collèges en France

1.1 Une première sensibilisation à l'école primaire

La sensibilisation aux problèmes d'hygiène et de santé, abordée à l'école maternelle, est prolongée sous la forme d'une justification des principes d'hygiène personnelle et collective dans le cycle des apprentissages fondamentaux

(cycle 2) de l'école primaire : habitude quotidienne de propreté, d'alimentation, de sommeil. L'élève prend progressivement conscience de la complexité de son environnement et de l'action exercée par les hommes. Les conséquences à court et long terme de notre hygiène, les actions bénéfiques ou nocives de nos comportements (notamment l'alimentation) sont étudiées de façon adaptée au cycle des approfondissements (cycle 3) de l'école primaire.

1.2 Une formation basée sur la liaison entre éducation à la santé et éducation à l'environnement

Au collège, dès la classe de sixième, l'enseignement de sciences de la vie et de la Terre (SVT) permet de construire le concept d'environnement tout au long de l'année, dans une perspective d'éducation à la responsabilité individuelle et collective : relations des êtres vivants entre eux et avec les facteurs physicochimiques du milieu. Une contribution élémentaire à l'éducation nutritionnelle est proposée dans le chapitre du programme intitulé : « Des pratiques au service de l'alimentation humaine ».

C'est au cycle central (classes de cinquième et de quatrième) que les programmes invitent les enseignants à relier les données anatomiques, fonctionnelles et épidémiologiques à des problèmes de santé, dans le cadre d'une étude des bases physiologiques du fonctionnement du corps humain. Les effets de la consommation de certaines substances (dopants, alcool) sur le fonctionnement nerveux, des substances nocives comme le tabac, les gaz polluants . . . sur le fonctionnement de l'appareil respiratoire ou encore l'effet d'une alimentation trop riche sur les maladies cardiovasculaires sont étudiés. Les élèves sont invités, dans ce cadre scientifique, à effectuer au CDI (centre de documentation et d'information) de l'établissement, des recherches de données sur les facteurs de risques. Ces bases biologiques fondent le choix responsable des élèves dans le domaine comportemental. Éducatons à la santé et à l'environnement sont toujours liées dans ces enseignements dont la portée éducative est un objectif majeur.

1.3 Une articulation entre instruction et éducation aux responsabilités en classe de troisième

La dernière année du collège est aussi, pour la moitié des élèves d'une classe d'âge, la dernière année d'enseignement de la biologie. Le programme est centré sur l'homme et les moyens de préservation de l'organisme par rapport aux risques liés à l'environnement. Les modes de transmission de micro-organismes

(bactéries, virus), les références scientifiques permettant de comprendre l'efficacité de moyens préventifs ou curatifs mis au point par l'homme (vaccination, sérothérapie, antibiothérapie) ainsi que l'approche historique des circonstances de découvertes scientifiques qui sont à l'origine ou le résultat de ces pratiques médicales font l'objet d'un enseignement. Les recherches documentaires conseillées portent sur les MST, le sida, sur l'effet de carences et excès alimentaires, sur des maladies comme l'obésité, le diabète, les maladies cardiovasculaires, sur l'altération de la perception par les agressions (lumineuses, sonores) ou encore sur les substances chimiques comme les tranquillisants, les antidépresseurs, les drogues.

En outre, le programme de troisième comporte une préparation des élèves à la responsabilité individuelle et sociale de l'homme dans le domaine de la santé, sous la forme d'un thème à étudier avec l'enseignant de physique et chimie, quand cela est possible. Par exemple, les élèves effectuent une recherche sur la propagation d'un agent infectieux et sur les mesures prises pour l'éviter (notions d'épidémies, d'endémies, cartes de vaccinations obligatoires ou conseillées, intérêt social des mesures collectives). Un autre thème peut concerner la responsabilité de l'homme à l'égard de la santé et de l'environnement à l'échelle de la planète : gaz à effet de serre, couche d'ozone, effet de l'exposition. Ce travail concerté peut conforter l'éducation à la citoyenneté récemment introduite dans les programmes.

Donc, il y a effectivement place pour, dans ces enseignements, introduire avec profit pour les élèves des résultats épidémiologiques. Mais, des données sur les comportements individuels et les risques associés ne sont actuellement pas à la disposition des professeurs.

2 | L'épidémiologie dans les enseignements de SVT au lycée

2.1 Une imprégnation trop modeste dans les séries de culture générale

La classe de seconde de détermination, dans les lycées d'enseignement général, ne propose que quelques éléments en rapport avec le sujet : la circulation atmosphérique et la propagation corrélative des nuages de poussières et des polluants. Ces bases sont réinvesties dans les thèmes proposés aux classes de première littéraire (L) et de première économique et sociale (ES) : alimentation et environnement ou alimentation, production alimentaire, environnement, ou encore gestion de l'eau (pollutions biologiques, chimiques : nitrates, pesticides, radionucléides). Les effets néfastes des déséquilibres alimentaires sont abordés

ainsi que les aspects éthiques des rapports entre cultures intensives, pollutions et santé.

En première ES, l'enseignement de biologie aborde, dans le cadre de la communication nerveuse, les mécanismes de la dépendance dans le cas de la morphine. Dans cette classe, sont abordés les rôles respectifs des gènes et de l'environnement : « la diversité phénotypique, résultat des interactions de plusieurs gènes entre eux et avec les facteurs de l'environnement ». L'enseignement exploite les exemples de la drépanocytose, de la phénylcétonurie, du cancer (prédispositions familiales, rôles de l'environnement et de l'alimentation).

Mais les enseignants ne disposent pas de documents fiables et actualisés sur ces sujets qui pourtant passionnent les élèves. Une réflexion civique et éthique prolonge cet enseignement, dans les meilleurs cas, dans le cadre de celui d'Enseignement civique, juridique et social (ECJS) sur le diagnostic prénatal et la médecine prédictive.

2.2 Un manque important de données épidémiologiques pour la série scientifique

En première scientifique (S), l'interaction entre gènes et facteurs de l'environnement est étudiée dans le cadre de la régulation de la glycémie et des phénotypes diabétiques, en particulier de type II, en rapport avec l'obésité. Une analyse de données épidémiologiques est là aussi tout à fait fondamentale. Le concept de susceptibilité et celui de prédisposition génétique soulèvent les problèmes éthiques à aborder dans le prolongement de la classe.

En classe de terminale scientifique (TS) est étudié un exemple de relation entre mécanismes de l'évolution et génétique : mutations de l'ADN, fréquence de l'allèle βS de la globine dans le cas du paludisme. D'autre part, le VIH et le sida servent de supports à l'étude de l'immunologie dans cette classe. L'adaptabilité, la variabilité du système immunitaire sont présentées comme le résultat des interactions entre le génotype et l'environnement. Le phénotype immunitaire (diversité des lymphocytes T et B), sa sélection par les antigènes des cellules malades ou des pathogènes présents, à l'origine des clones actifs, illustrent en effet les mécanismes de l'adaptation du phénotype à l'environnement.

Dans le chapitre intitulé : « Dépistage et diagnostic génétique, biotechnologies et génétique humaine » on aborde en SVT l'identification d'une pathologie d'origine génétique, l'évaluation des risques génétiques et la prévention, le dépistage ; le diagnostic d'une maladie génique est abordé dans le cadre de la trisomie 21.

Dans ces champs scientifiques, on reconnaît à nouveau que les enseignants pourraient tirer un grand bénéfice des méthodes et des résultats de l'épidémiologie.

3 | Quelques exemples d'ancrage de l'épidémiologie dans les programmes étrangers

3.1 Données des sites gouvernementaux de l'éducation dans les pays francophones

Les programmes francophones, au sein de la Communauté européenne et à l'extérieur de celle-ci, présentent des parties susceptibles de permettre l'introduction de l'épidémiologie.

En Belgique, par exemple, le programme fondamental destiné à l'éducation des enfants de 3 à 12 ans, comporte un thème, « L'enfant gestionnaire de sa santé », qui fonde le développement des compétences premières relatives aux domaines de l'hygiène corporelle et alimentaire, des rythmes de vie (veille, sommeil), de l'importance des exercices physiques, de la biométrie (la croissance pondérale en particulier), des pathologies respiratoires ou encore de la pollution de l'air et de l'eau.

Comme en France, les enfants sont sensibilisés très tôt à la nécessité de l'équilibre alimentaire, aux déviations de la nutrition et à leurs effets sur la santé.

De manière plus marquée que dans notre enseignement français, un thème sur l'énergie sensibilise à l'extrême richesse du monde de la lumière et des sons mais aussi à la pollution sonore et aux interactions entre l'Homme et son environnement (cf. <http://www.restode.be.prges/programmes/index.htm>).

Le programme belge du premier degré de l'enseignement secondaire, destiné aux enfants de 13 et 14 ans (équivalent aux classes de cinquième et quatrième en France) comporte 11 thèmes scientifiques. L'intégration de l'épidémiologie pourrait en concerner trois : une approche de l'influence d'un facteur environnemental sur la santé d'une part, l'étude de la relation entre pollution de l'eau et pathologies d'autre part, la relation entre la qualité de l'air et les maladies respiratoires enfin. Mais cette orientation n'est pas du tout explicite (cf. <http://www.restode.be.prges/programmes/secd11A2C.htm>).

Le programme scientifique belge du second degré de l'enseignement secondaire qui concerne les élèves de 15 et 16 ans indique « ... Un important volet

quantitatif s'ajoutera à l'aspect qualitatif du cours de sciences ». Le module biologie ayant pour thème « Alimentation et santé » semble pouvoir constituer une porte d'entrée à l'épidémiologie : les maladies liées à la sous-alimentation, les maladies cardiovasculaires, l'obésité . . . sont abordées en troisième année du secondaire. En quatrième année, dans les parties « le sang et la circulation » et « la respiration », divers prolongements sur l'infarctus, le cholestérol, les effets du tabac, l'alcool, l'importance d'une activité physique peuvent être réalisés. Enfin, dans la partie « l'excrétion », le prolongement vers l'hygiène prévoit de traiter l'abus de l'alcool et la notion d'alimentation saine et équilibrée.

Le Québec, qui représente environ un quart de la population canadienne et a pour langue officielle le français, a entrepris la réforme des programmes de formation. Ceux-ci mettent l'accent sur les liens à développer entre les apprentissages des élèves et les situations de la vie courante.

Dans l'enseignement primaire, en particulier dans le troisième cycle (enfants de 10 et 11 ans), « l'éducation physique et à la santé » semble un excellent cadre pour la mise en œuvre des méthodes et la familiarisation avec des notions d'épidémiologie. L'élève est amené à établir un lien entre ses habitudes de vie et leurs effets sur sa santé. Il exerce sa pensée critique face à des opinions et à l'information disponible. Il planifie la modification d'une habitude de vie, évalue sa démarche et les résultats obtenus, identifie les améliorations souhaitables. Les savoirs essentiels à acquérir concernent notamment l'effet bénéfique de l'exercice dans la contribution à la croissance et l'hygiène corporelle associée à la pratique des exercices.

Dans l'éducation secondaire, l'enseignement obligatoire de la biologie humaine (100 heures) peut permettre l'introduction de l'épistémologie dès la classe de troisième année (jeunes de 14 ans). Le programme de 1982 propose un apprentissage qui anticipe un comportement responsable à l'égard de la santé personnelle et de la santé collective. Trois objectifs sont soulignés : comprendre les changements physiologiques qui s'opèrent chez l'élève, découvrir sa responsabilité vis-à-vis de son corps, apporter sa contribution personnelle dans le domaine de la santé publique. L'épidémiologie peut être introduite dans l'étude de la fonction de nutrition (aliments, physiologie et hygiène respiratoires, tabac et pathologies associées, hygiène cardiovasculaire, croissance et équilibre alimentation-activité. . .).

Dans les fonctions de relation, le fonctionnement nerveux et l'hygiène de vie permettent l'intégration de données épidémiologiques. Avec l'étude des MST et des méthodes de contraception, en outre, les aspects scientifiques, psychiques et éthiques peuvent être croisés. D'ailleurs, la biologie est complémentaire du programme de « formation personnelle et sociale » et de son volet « éducation à la responsabilité sexuelle ».

Toujours au Québec, le programme de biologie générale des élèves de quatrième et de cinquième années (jeunes de 15 et 16 ans) sensibilise les jeunes à la notion de qualité de la vie. Les thèmes des trois modules permettent éventuellement d'introduire des méthodes et des notions d'épidémiologie : les virus, le clonage, le stress, l'abus des médicaments, l'homéostasie. . . De plus, le programme « techniques et méthodes en sciences de la nature » correspond depuis 1992 à un cours facultatif destiné aux élèves de cinquième année. Il a le mérite de s'appliquer aux programmes de physique, chimie et biologie générale. Enfin, le programme de « formation personnelle et sociale » de 1984 s'adresse aux élèves du secondaire. Les divers aspects de la vie y sont abordés : santé, sexualité, consommation et vie en société, droits et responsabilités du citoyen, normes collectivement agréées. L'introduction de l'épidémiologie est tout à fait possible dans ce type de formation. Mais elle n'est explicitement citée.

Sites à consulter :

- base Eurydice :
http://www.eurydice.org/Eurybase/frameset_eurybase.html
- Belgique : <http://www.cfwb.be/matiere/enseig/pg002.htm>
- Luxembourg : <http://www.men.lu/secondaire>

<http://www.men.lu/edu/fre/hor>

- Suisse : <http://www.bbw.admin.ch/>
- Principauté de Monaco : <http://www.education.gouv.mc/>
- Canada : <http://www.cmec.ca/index.htm>
- http://www.meq.gouv.qc.ca/rens/brochu/sys_scol.htm
- <http://www.meq.gouv.qc.ca/publications/menu-pub-ped.htm>
- http://www.meq.gouv.qc.ca/DGFJ/dp/programmes_etudes/secondaire/biohum.htm
- http://www.meq.gouv.qc.ca/DGFJ/dp/programmes_etudes/secondaire/biogen.htm
- http://www.meq.gouv.qc.ca/DGFJ/dp/programmes_etudes/secondaire/tms532.htm

3.2 Données des sites de l'éducation dans un pays non francophone

Les programmes éducatifs du Royaume-Uni, dans les deux types d'écoles du secondaire, *comprehensive schools* et *grammar schools*, consistent en un curriculum national où la biologie est regroupée avec la physique et la chimie. Depuis la fin des années 1990, tous les enfants du cycle 3 (*key stages 3*, 11 à 14 ans) et du cycle 4 (*key stages 4*, 14 à 16 ans) reçoivent un enseignement « d'éducation personnelle sociale et de santé » sous la forme de séquences destinées à accompagner les jeunes dans leur développement psychique et physique. Citons quelques sujets : les religions, la maltraitance, les drogues, les comportements à risques (palliant notamment la non-mixité des classes), les règles communautaires, les principes de base d'une alimentation équilibrée, etc. Deux séquences de 30 à 35 minutes par semaine sont assurées, en classe entière, par l'équivalent du professeur principal en France.

Le programme de science vise le développement spirituel, avec pour support la composition du monde environnant, la place de soi dans ce monde, l'origine de la vie, et aussi l'apprentissage social, notamment la formation d'une opinion et la prise de décision à partir de données scientifiques. L'enseignement doit permettre l'utilisation des techniques d'information et de communication (TICE) dans les cycles 2, 3 et 4. L'épidémiologie peut trouver sa place dans les différents cycles d'étude où quatre domaines sont identifiés : le questionnement scientifique, les êtres vivants et les processus du vivant, les matériaux et leurs propriétés, les processus physiques.

Dès le cycle 1 (enfants de 5 à 7 ans), on montre l'importance des mesures accompagnant une observation pour répondre à une question. Dans le cycle 2 (7 à 11 ans), la nécessité de tester ses idées à partir des observations et des mesures apparaît, avec celle d'établir des liens entre les causes et les effets (exemple de l'efficacité de la vaccination de Jenner). La formation méthodologique est centrée sur la relation entre prédictions ou hypothèses et les résultats obtenus. L'effet de l'exercice et du repos sur le rythme cardiaque, les effets du tabac, de l'alcool et d'autres drogues sur l'individu servent de support à cette formation. On aborde aussi les rôles bénéfiques et les effets néfastes des micro-organismes. Les notions de hasard et de risques sont présentes, comme dans le cycle 1.

Au cycle 3 (jeunes de 11 à 14 ans), les apprentissages dans le domaine du questionnement sont enrichis par les notions de choix des facteurs à considérer lors de la collecte des données ou d'une enquête, du nombre de données à collecter ou de la taille de l'échantillon biologique, de la méthode et du matériel à utiliser, de la prise en compte des essais d'explication. Les domaines de la nutrition, de la santé, des effets des excitants, des bactéries et virus sont ainsi

abordés, ainsi que les mécanismes naturels de défense. Des exemples de chaînes alimentaires et d'accumulation de substances toxiques sont analysés et les méfaits des bruits sur l'ouïe sont abordés. L'utilisation des banques de données et de méthodes quantitatives est préconisée.

Dans le programme *single science* du cycle 4 (jeunes de 14 à 16 ans), l'apprentissage des capacités liées au questionnement, la sensibilisation aux méthodes quantitatives et aux notions de hasard et de risques se poursuivent. L'approfondissement concerne l'appréciation du degré d'incertitude des mesures et collectes de données, ainsi que l'expression de valeurs avec le degré de précision adéquat.

Les contenus supports de ces apprentissages susceptibles de faire appel à l'épidémiologie sont par exemple l'utilisation médicale des hormones (en intégrant le contrôle et l'assistance à la procréation), les effets des solvants, de l'alcool, du tabac, de drogues sur les fonctions vitales. L'influence des facteurs économiques et sociaux dans l'impact de l'homme sur son environnement (taille de la population, nature des procédés industriels, niveau de consommation et de production de déchets par exemple) fait l'objet d'un enseignement. Les effets néfastes des radiations sur les êtres vivants sont aussi abordés.

Le programme « double science » propose, dans le même cycle, un plus grand nombre d'objectifs notionnels (par exemple, l'utilisation d'hormones dans le traitement du diabète, l'homéostasie, le contrôle de la température corporelle, la transmission génétique de certaines maladies, le rôle des micro-organismes dans la décomposition de la matière organique et dans les cycles du carbone et de l'azote. . .). L'approfondissement des méthodes du questionnement scientifique apparaît aussi : juger l'incertitude d'un résultat par l'utilisation de la variation des mesures répétées, juger de la pertinence des faits collectés vis-à-vis de l'interprétation proposée représentent par exemple des objectifs affichés.

Sites à consulter :

- http://www.eurydice.org/Eurybase/frameset_eurybase.html
- <http://www.nc.uk.net/index.html>

Dans les cas étudiés, en France comme à l'étranger, l'épidémiologie ne figure donc pas, en tant que discipline, dans les enseignements scientifiques. Cependant, l'accent qui est mis, de façon d'ailleurs assez comparable d'un pays à l'autre, sur l'aspect méthodologique de la formation, notamment sur le va-et-vient rigoureux entre les idées explicatives et les faits, naturels ou provoqués par l'expérience, se prête tout à fait à une intégration des apports de l'épidémiologie dans nos enseignements. Servant à leur ancrage, les objets d'étude concernant

l'éducation à la santé et les liens entre environnement et santé, plus généralement la préparation des jeunes à la responsabilité individuelle et collective, sont bien présents dans les programmes.

4 | L'épidémiologie au service d'une éducation à la santé renouvelée

Dans la discipline d'enseignement SVT, les professeurs sont préparés à prolonger la transmission des connaissances biologiques par une éducation à la responsabilité individuelle et collective dans le domaine de la santé. Il serait bien notamment que les élèves aient des connaissances sur les facteurs environnementaux et professionnels, les inégalités sociales de santé... (par exemple, dans les filières techniques, une formation sur l'épidémiologie des risques professionnels ne serait pas inutile). Cependant, cette éducation des élèves en vue des choix comportementaux à opérer revêt un caractère global associant aux bases scientifiques des problèmes étudiés des dimensions psychiques, morales et sociales.

4.1 Une ambition nationale non satisfaite

Un cadre institutionnel de cette éducation a été proposé en 1998 par la circulaire ministérielle (C.M. 98-237 du 24 novembre 1998 ; BO n° 45 du 3 décembre 1998). L'enseignant de SVT contribue avec d'autres disciplines à la construction des capacités personnelles et relationnelles des élèves que cette éducation sous-tend. Elle se développe à travers les enseignements mais aussi l'ensemble des activités de la vie scolaire, notamment lors des « *rencontres éducatives sur la santé* » associant, dans un travail en équipe, les personnels de direction, les enseignants, les personnels de santé et d'autres partenaires, notamment des membres du comité d'éducation à la santé et à la citoyenneté (CESC). Le projet d'éducation globale fondé sur la prise en compte des attentes et des besoins des élèves devait utiliser une plage horaire banalisée, étalée sur 30 à 40 heures à deux niveaux du collège, à raison de 2 heures par semaine sur 10 semaines par exemple. Le rapport de la mission confiée aux inspections générales par les ministres de l'Éducation nationale et de la Santé (n° 2004-003 pour le ministère de la Jeunesse, de l'Éducation nationale et de la Recherche, n° 2004-013 pour le ministère des Affaires sociales, du Travail et de la Solidarité — janvier 2004) montre que la généralisation de cette action éducative n'a pas été possible en raison d'une très faible implication des enseignants dans la pratique d'une démarche de projet exigeant une participation active des élèves, et aussi parce que le temps collectif de travail collégial réunissant plusieurs catégories de personnels n'a pas pu être trouvé, peut-être en raison d'un pilotage trop peu engagé des personnels de direction.

4.2 Des recommandations utiles

Parmi les recommandations faites aux ministres, il semble utile de retenir l'idée du développement de l'éducation à la santé à partir des professeurs de SVT au collège et au lycée d'enseignement général, de Vie sociale et professionnelle (VSP) et d'Hygiène, prévention, secourisme (HPS) au lycée professionnel, de Biotechnologie santé-environnement et de Sciences techniques médicosociales (STMS) dans les lycées technologiques. Du fait de leur spécificité disciplinaire, ces professeurs pourraient élargir leur champ d'action en recherchant les collaborations nécessaires au sein de la communauté éducative et avec des partenariats, afin de développer la préparation active par les élèves d'études thématiques conduisant à des productions. *L'attribution d'une demi-heure hebdomadaire (ou d'une heure de quinzaine)* à l'horaire d'enseignement de ces professeurs et l'inscription de cet horaire à l'emploi du temps de tous les élèves (temps spécifique pour les travaux en équipe et les synthèses) permettraient de réussir la mise en place généralisée de cette action éducative en faveur de la santé et du bien-être des jeunes. L'éducation doit s'inscrire dans la durée, à la place des interventions actuelles, ponctuelles et inefficaces, trop souvent confiées à des partenaires extérieurs.

4.3 L'apport de l'épidémiologie aux innovations pédagogiques

Les méthodes et résultats de la recherche en épidémiologie, dans ce cadre, pourraient fournir un accompagnement fondamental à ce projet en constituant un fonds de données à mettre en ligne sur le site Internet des SVT auquel ont accès tous les professeurs. Ces bases de données seraient utiles aux enseignants de biologie. Elles faciliteraient en outre les travaux collectifs qui relèvent des travaux personnels encadrés (TPE) au lycée d'enseignement général, des projets pluridisciplinaires à caractère professionnel (PPCP) au lycée professionnel ou des itinéraires de découverte (IDD) au collège, travaux originaux qui associent deux ou plusieurs disciplines d'enseignement. Ces données pourraient également servir d'appui à l'enseignement civique, juridique et social (ECJS), pris généralement en charge par les historiens.

4.4 Épidémiologie et formation

Par ailleurs, une formation à l'épidémiologie devrait être proposée aux stagiaires du Capes (PLC2) et de l'agrégation de SVT et de mathématiques au sein de la deuxième année d'IUFM (Institut universitaire de formation des maîtres).

Une formation continue à distance (vidéoconférences) organisée à partir d'une académie à l'intention des regroupements interacadémiques de professeurs est tout à fait possible.

Tous ces niveaux d'intervention de l'épidémiologie : programmes d'enseignement, actions d'éducation à la santé au sein des établissements, formation initiale au moins bidisciplinaire, formation continue à distance, permettraient un enrichissement incontestable des missions des professeurs liant instruction et éducation, pour une meilleure réussite de la préparation des jeunes à leurs nouvelles responsabilités, individuelle et collective.

SOUS-CHAPITRE 9.3

Les moyens nécessaires à la recherche épidémiologique

ANTOINE FLAHAULT

On a pu croire à une certaine époque que l'épidémiologie était l'affaire d'une poignée de questionnaires remplis par un ou deux enquêteurs analysés à l'aide de quelques tests statistiques suffisamment simples pour être convaincants. Certains exemples emblématiques plaidaient pour cette vision d'une discipline encore discrète mais montante, comme en 1971, lorsque fut publiée dans le *New England Journal of Medicine* une étude cas-témoins réalisée à partir de 8 cas et 32 témoins qui allait permettre de découvrir le rôle du diéthylstilbestrol administré à des femmes enceintes dans la genèse de cancers du vagin à cellules claires chez leurs filles. Aujourd'hui, avec l'éruption des sciences de l'information dans les deux secteurs majeurs de l'épidémiologie (populationnelle et clinique), nous allons montrer que des projets pharaoniques voient actuellement le jour aux États-Unis comme en Europe, nécessitant la mise en place d'une logistique lourde pouvant mobiliser simultanément plusieurs dizaines de personnels de haute qualification autour de plateaux utilisant de nouvelles technologies entièrement dédiés à ces recherches. Nous focaliserons donc ce chapitre

principalement sur les moyens nécessaires pour conduire ces très grands projets. Nous allons aussi rendre compte du fait que le secteur privé s'est organisé pour apporter la capacité de mener ces travaux en France, et que cette capacité s'avère vitale et stratégique pour l'industrie des produits de santé. Nous allons enfin tenter de comprendre à quelles conditions les milieux académiques peuvent espérer dans notre pays mettre en œuvre des projets d'ambition similaire à ce qui se met en place dans les lieux de recherche les plus compétitifs dans ce domaine dans le monde.

1 | Bonnes pratiques en épidémiologie : le coût de la qualité

Les bonnes pratiques d'épidémiologie qui sont en place en France (<http://adelf.isped.u-bordeaux2.fr/>) s'accordent à considérer comme fondamentaux l'assurance qualité, les contrôles qualité et les audits réalisés aux différents maillons de la chaîne de recueil d'information épidémiologiques. L'organisation de telles études nécessite aujourd'hui une administration et une gestion de qualité. Elle ne peut pas être confiée, sans ressources spécifiquement allouées, aux équipes qui coordonnent ces projets, équipes le plus souvent déjà sous-dotées en matière de secrétariat d'aide aux recherches et en gestionnaires. Les compétences et la charge de travail pour réaliser des appels d'offres dans le cadre des marchés publics, puis pour gérer les contrats éventuellement obtenus nécessitent une organisation professionnalisée collaborant avec les équipes coordonnant ces projets. Enfin, les grands projets d'épidémiologie clinique nécessitent de constituer et de réunir des comités divers dont l'utilité est indiscutable (puisque prévue par les recommandations de bonnes pratiques), mais qui ne peuvent pas fonctionner sans un minimum de moyens : comité de pilotage, conseil scientifique indépendant, comités techniques, comité indépendant de suivi et de monitoring, conseil de surveillance financière. Ces comités ont besoin d'un secrétariat, d'un budget de fonctionnement. S'il est normal que les investigateurs tirent un juste profit des cohortes et des banques de prélèvement qu'ils ont contribué à recueillir, profits tant en terme de projets de recherche qu'en terme de valorisation (brevets, contrats), cela ne peut se concevoir que dans l'excellence scientifique. Or, la gestion de ces ressources constituées le plus souvent avec l'argent du contribuable doit se faire dans un esprit d'ouverture aux autres équipes (y compris à des partenaires privés) et d'indépendance vis-à-vis des investigateurs des études, indépendance qu'il convient de protéger par la mise en place de conseils scientifiques crédibles, indiscutés et disposant des moyens de favoriser l'émergence de programmes innovants. Il faut sans doute ne pas oublier les moyens dévolus à la politique de communication, toujours importante à gérer au cours de ces grandes études.

2 | L'épidémiologie populationnelle est hautement consommatrice de ressources

Nous étudierons dans ce paragraphe cinq exemples (en détaillant plus particulièrement le premier d'entre eux) qui échantillonnent différents aspects de ce domaine à l'étranger. Les projets d'envergure comparable réalisés en France ont été détaillés plus haut dans le rapport, et l'on a pu montrer qu'ils étaient largement sous-dotés tant en ressources humaines que logistiques et financières. L'objectif de cette section est de chiffrer le coût de ce type de projet dans les pays voisins du nôtre sur le plan du développement économique. Le premier projet que nous traiterons est basé sur une méthodologie d'épidémiologie descriptive de l'état sanitaire et alimentaire de la population nord-américaine (enquêtes transversales réitérées), le second est basé sur le suivi de la population d'un village entier du Massachusetts depuis l'après-guerre à nos jours, le troisième sur la mise en place d'une cohorte de nouveau-nés nord-américains qui seront suivis durant les trente prochaines années, le quatrième sur la mise en réseau de médecins généralistes britanniques utilisant un dossier informatique concernant 5 % de la population des patients du Royaume-Uni. Le cinquième montre les coûts impliqués par le recueil de données individuelles auprès d'une population de 1,6 millions de personnes (Portland, États-Unis) en vue notamment de modéliser la dynamique épidémique d'une attaque de variole d'origine bioterroriste.

Le **National Health And Nutrition Examination Survey** (Nhanes) est un programme nord-américain pluriannuel coordonné par 60 employés fédéraux des Centers for Disease Control and Prevention (CDC) d'Atlanta permettant d'établir une photographie instantanée de l'état de santé et de l'alimentation de la population à partir d'un échantillon représentatif de la population des États-Unis. Ce programme est le fruit d'une collaboration impliquant le financement notamment des CDC, des National Institutes of Health (NIH), du ministère de l'Agriculture (DoA), de la Food and Drug Administration (FDA), de l'Environmental Protection Agency (EPA). Le financement annuel du programme, incluant les coûts directs et indirects (comme l'analyse des échantillons de laboratoire, les consultations d'experts, la formation, et l'ensemble de la logistique), était de 35 millions de dollars en 2004, (*Paula Summerour-Norman, Office of Information Services, Centers for Disease Control and Prevention, communication personnelle, mars 2004*). Ce budget n'inclut pas les salaires des personnels employés par les agences fédérales citées, en particulier il n'inclut pas non plus les coûts relatifs au temps passé pour l'analyse de données. Quatre « clichés » ont été pris durant les 10 dernières années, il s'agit des enquêtes transversales suivantes :

- 1) Nhanes I (1971-1975) portant sur 31 973 personnes âgées de 1 à 74 ans ;

- 2) Nhanes II (1976-1980) sur 27 801 personnes âgées de 6 mois à 74 ans ;
- 3) Nhanes III (1988-1994) sur 33 994 personnes âgées de 2 mois et plus ;
- 4) Nhanes IV (1999-2004) portant sur 12 160 personnes en 1999-2000 et 11 039 personnes en 2001-2002.

L'observatoire équipé de centres mobiles (figure 9.3-1) qui sillonnent les routes des États-Unis, dispose d'une imagerie de haute technologie, de prélèvements biologiques et génétiques effectués auprès de personnes volontaires représentatives des différentes composantes ethniques et sociales du pays. Les entretiens au domicile des personnes enquêtées durent environ une heure et portent sur la santé, les antécédents, le régime alimentaire. Pour faciliter la participation, les sujets inclus sont rémunérés, les trajets remboursés, les frais de garde d'enfant indemnisés. Les fichiers de données d'une très grande richesse sont ensuite mis à la disposition de tous les chercheurs qui le souhaitent (données brutes disponibles sur le site Internet en format SAS). Les informations sont redistribuées aux participants, sans délai en cas d'anomalie détectée (figure 9.3-2b). L'équivalent n'existe ni en France ni en Europe. La production scientifique est impressionnante (figure 9.3-3, 8306 références dans la *National Library of Medicine* le 12 avril 2004). L'analyse des 100 derniers articles référencés dans PubMed montre les centres d'intérêt du programme : modification des apports alimentaires depuis 1970, carence en vitamine C dans la population générale, conséquences du tabagisme passif chez le petit enfant, expositions environnementales (phtalates, plomb, bore, mercure, pesticides), santé buccodentaire, couvertures vaccinales, prescriptions médicamenteuses, valeurs de référence dans les sous-groupes de la population (par exemple, chez les personnes âgées), contrôle de la glycémie chez les patients atteints de diabète de type II, évolution de la prévalence des maladies coronariennes de 1971 à 1994, IMC (indice de masse corporelle) et tour de taille comme indicateurs de risque cardiovasculaire, relation entre anti-oxydants circulants et asthme chez l'enfant, disparités raciales et insuffisance rénale chronique, relation entre fracture et densité osseuse chez les personnes âgées.

Framingham est un suivi de trois générations d'un village nord-américain (*Massachusetts*) sur le plan cardiovasculaire. Depuis 1948, 5 209 hommes et femmes âgées de 30 à 62 ans à l'inclusion ont été suivis tous les deux ans ; à partir de 1971, leurs enfants (5 124 garçons et filles) ont été suivis tous les quatre ans, et depuis 2002, 3 500 petits-enfants sont inclus dans l'étude. Le coût de cette cohorte est estimé à 4 500 dollars américains par sujet et par an. Ce budget couvre essentiellement les frais d'examen (clinique, laboratoires, imagerie lourde) et de logistique du programme (équipements, fonctionnement, frais de mission, location de salles de réunion). Ce budget inclut les cinquante personnes employées à temps plein par l'université de Boston (sur contrat des

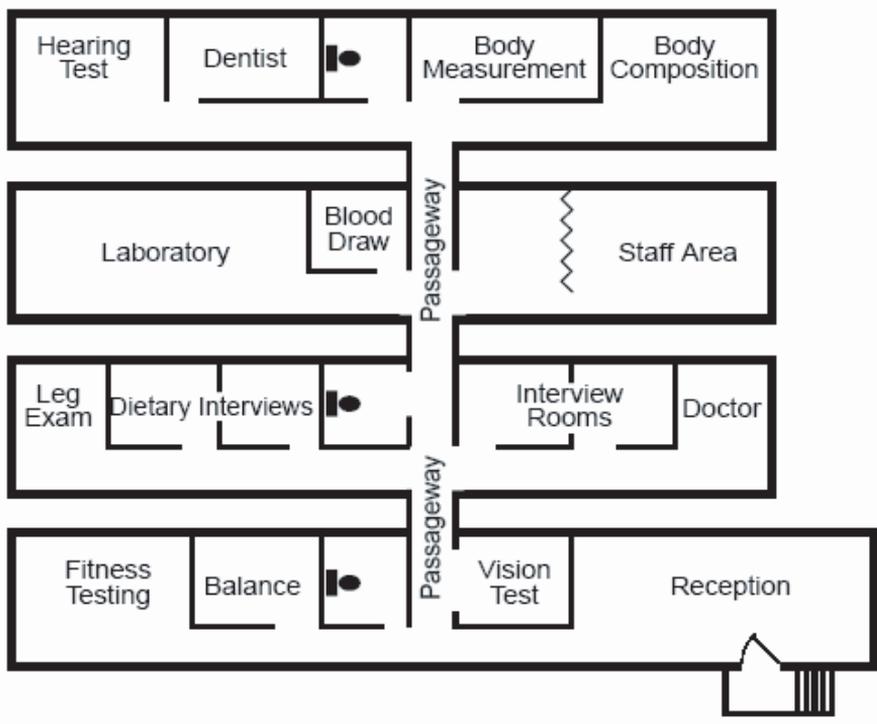


Figure 9.3-1
 Les centres mobiles d'examen des enquêtes Nhanes.



National Health and Nutrition Examination Survey

Final Report of Findings

These measurements were obtained as part of a survey and do not represent a medical diagnosis. Interpretation of these measurements must be made by a physician.

Date of Examination: April 13, 1999
Participant Name: John Q. Public
Participant Age: 43 years
Participant Gender: Male
SP ID: 123456

Body Measurements

Height/Length: 5 ft. 9 in.
 Weight: 179.0 lbs

For a person of your height, your weight is above the range of a healthy weight, and you may be overweight.

Blood Pressure and Heart Rate

	Optimal	Normal	Acceptable
Systolic Blood Pressure:	< 120	< 130	< 140
Diastolic Blood Pressure:	< 80	< 85	< 90
Resting Pulse Rate:	64 bpm		

The participant's blood pressure is mildly high. Based on the Sixth Report of the Joint National Committee on Detection, Evaluation, and Treatment of High Blood Pressure. NIH Publication, 1997.

Dental

The dental examination of the National Health and Nutrition Examination Survey is not, and is not intended to be, a substitute for the examination usually given to persons seeking care from their own dentists. Neither a dental history nor x-rays are taken, and therefore the findings are solely the result of what can be seen at the time of the examination.

The examining dentist recommends that you continue your regular routine care.

- No findings

John Q. Public, 123456, April 31, 1999
 NHANES National Center For Health Statistics, 6525 Belcrest Road, Room 900, Hyattsville, Maryland 20782.

Page 1 of 6

Figure 9.3-2a

Exemple de retour d'information aux sujets inclus dans Nhanes.

NIH) qui coordonnent le programme, le mettent en œuvre, et en analysent les données. Il n'inclut cependant pas les cinq personnes à temps plein des NIH chargées de l'étude sur place. Aucun défraiement des personnes incluses n'est

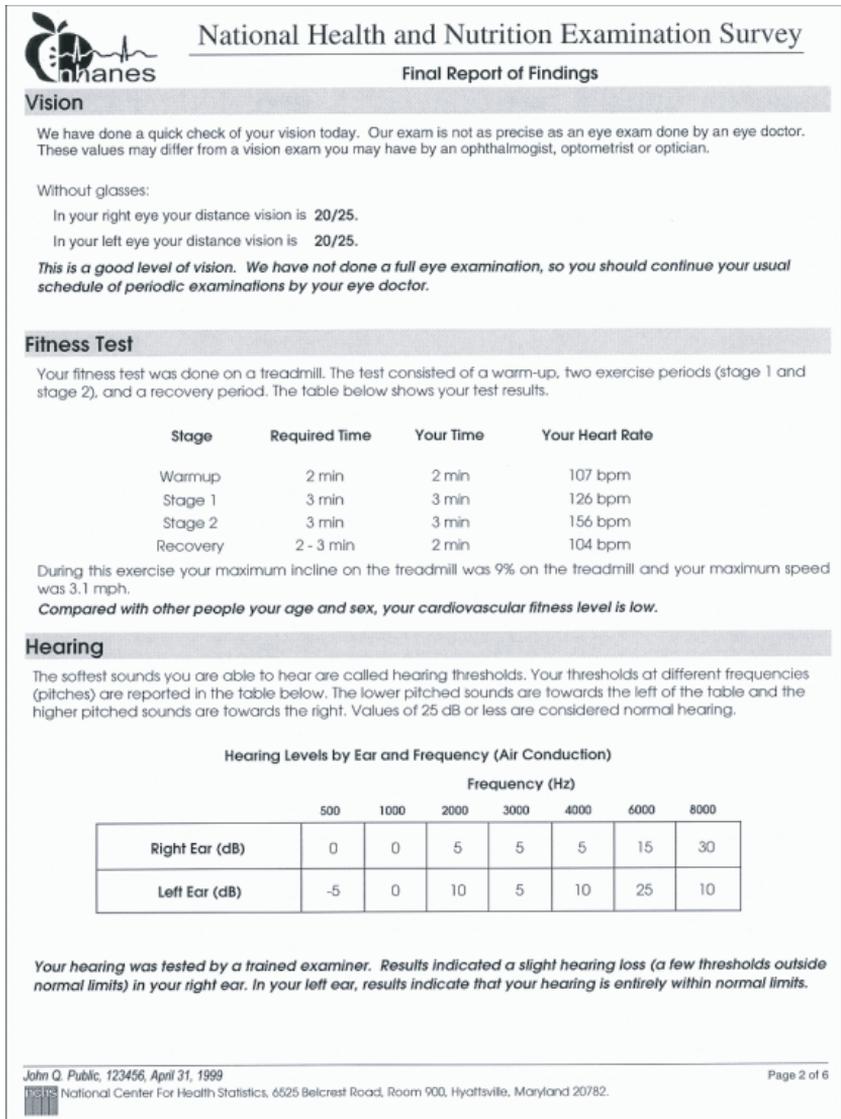


Figure 9.3-2b
 Exemple de retour d'information aux sujets inclus dans Nhanes.

prévu (à part une légère collation servie à l'occasion des examens)¹. Dirigé conjointement par le NHLBI National Institute of Health (NIH) et l'université de

¹ Il est parfois dit en France que ce programme serait directement financé par le Congrès américain. C'est inexact, il n'y a pas de ligne budgétaire réservée à l'étude Framingham lors du vote du budget par le Congrès. Le budget de Framingham est alloué par le NHLBI sur son budget propre.

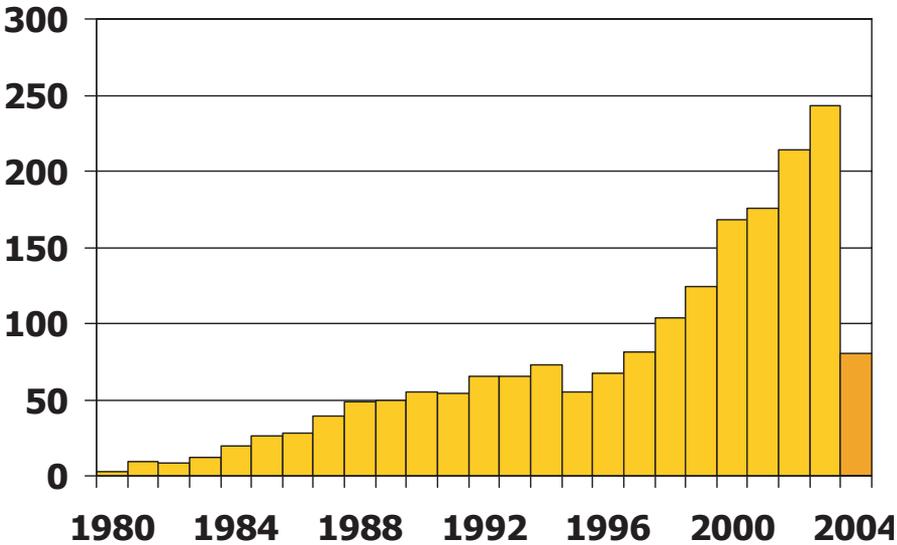


Figure 9.3-3

Production scientifique de Nhanes 1980-2004 (données provisoires arrêtées en mai).

Boston, financé depuis l'après-guerre sans discontinuité, il a permis de mesurer la plupart des risques cardiovasculaires reconnus aujourd'hui (Paul Sorlie, PhD, Project Officer, Framingham Study, NHLBI, communication personnelle, mars 2004). Ce programme est à l'origine de développements pharmaceutiques majeurs et les connaissances accumulées représentent l'une des clés de l'espérance de vie des générations à venir en termes de prévention cardiovasculaire ; il reçoit des financements complémentaires d'autres origines concernant des recherches notamment sur la démence sénile, les fonctions cognitives, et l'ostéoporose (2 483 références dans la *National Library of Medicine* le 12 avril 2004).

La **National Children Study** (NCS) est un projet de cohorte de 100 000 enfants suivis de la naissance jusqu'à 30 ans qui se met en place aux États-Unis actuellement à l'initiative des NIH (NICHD, NIEHS), en collaboration avec l'Environmental Protection Agency (EPA), et les Centers for Disease Control and Prevention (CDC). Le programme prévoit un suivi médical (grossesse et suivi post-natal, asthme, obésité et diabète, traumatismes) et psychosocial (apprentissage, comportements et santé mentale) d'un échantillon d'enfants représentatifs de la population nord-américaine sur une très longue durée. Les répercussions, les brevets industriels, la recherche et le développement que ce projet générera pour l'industrie nord-américaine sont attendus être à la mesure des investissements projetés (2,7 milliards de dollars sur 25 ans, pour lesquels un vote au Congrès est attendu ; nous n'avons pas eu de précisions sur le périmètre couvert

par un tel budget qui est cependant inférieur à celui de Framingham en proportion du nombre de sujets suivis, puisque il représente environ 1 000 dollars par sujet inclus et par an). La France cite le modèle de la National Children Study dans l'avant-projet du Plan national santé et environnement (PNSE), pour envisager d'en reprendre les principes, dans une collaboration Inserm-InVS. Il faut noter que l'Europe n'est pas partout à la traîne dans ce domaine, puisqu'en Norvège, la cohorte « MoBa » (cohorte mère et enfant) réunit un effectif supérieur à celui du NCS, avec 100 000 femmes enceintes, le suivi de leurs enfants, et la constitution d'une biobanque (sang de la mère, du père et de l'enfant au cordon ombilical). Le seul sous-projet de génomique fonctionnelle (dénommé Fuge) est financé à hauteur de 15 millions de dollars par an. Le projet MoBa qui a démarré en août 2000, avait déjà recruté 20 000 femmes enceintes en janvier 2003. Ce projet est associé à *Conor* (cohort of Norway), qui suit 200 000 personnes sur le plan de leurs habitudes de vie, pratiquant des examens cliniques et des prélèvements biologiques (ainsi, du sang total et du sérum de 175 000 personnes avait été stocké fin 2002). Ce vaste et ambitieux projet mobilise les universités de Norvège et l'institut norvégien de santé publique et s'inscrit dans une longue tradition et expérience en santé publique dans ce pays : registre du cancer depuis 1952, registre des causes de décès depuis 1964, registre médical des naissances depuis 1967, registre du diabète depuis 1989, registre des prescriptions pharmaceutiques depuis 2003.

Le **General Practitioners Research Database** (GPRD) est la mise en commun d'une base de données rassemblant les informations recueillies sur le même logiciel par un réseau de médecins généralistes du Royaume-Uni. Il s'agit de la plus grande base de données de patients au monde (environ 5 % de la population britannique couverte, soit 3 millions de personnes, ou encore 35 millions de personne-années). Le système créé en 1987 (sous le nom de VAMP research database), dont l'opérateur technique est un industriel informatique, est financé et hébergé depuis 1999 par l'agence chargée du médicament au Royaume-Uni (MHRA). Les données recueillies sont à la disposition des chercheurs. Cet accès n'est cependant pas gratuit, le gouvernement britannique facturant environ 30 000 livres (universitaires comme industriels) un fichier portant sur 50 000 dossiers informatisés depuis 1987. Cette base a été par exemple mise à contribution pour apprécier la relation entre l'exposition au vaccin contre l'hépatite B et le risque de sclérose en plaques (Marrie *et al.*, 2000). Les autorités sanitaires ne disposent pas d'un tel outil en France. Un petit « frère » développé par la Cegedim, la même société qui opère pour le GPRD, existe en France, s'appelle Thalès et fonctionne avec un but lucratif pour le compte de clients privés, industriels du médicament pour la plupart ; il réunit 1 010 médecins généralistes et de plusieurs spécialités (102 cardiologues, 61 neurologues, 100 psychiatres, 120 gynécologues, 81 pédiatres, et 100 rhumatologues). Qu'en sera-t-il du dossier médical informatisé pour tous les assurés sociaux que le ministre de la Santé appelle de ses vœux pour 2006 ? Quelles seront les utilisations

épidémiologiques possibles de ce formidable système d'information (qu'aucun autre pays ne possède au monde)? Les épidémiologistes, les statisticiens seront-ils associés à sa mise en place dès sa conception? Quelles seront les modalités d'accès à ces données à des fins de recherche clinique et épidémiologique? Ce sont les questions qui restent ouvertes à l'heure de l'écriture de ce rapport. Les Nord-Américains (au Canada, notamment dans l'État du Saskatchewan et aux États-Unis dans plusieurs HMO, dont celle du Puget Sound, dans l'État de Washington) disposent de systèmes d'informations croisant différentes bases de données : les dossiers électroniques des patients avec les prescriptions médicamenteuses et les indications pour lesquelles ces médicaments ont été prescrits, mais aussi les dispensations médicamenteuses par les pharmacies, les données chaînées des séjours hospitaliers des mêmes patients, et enfin les données de mortalité individuellement décrites. Moyennant des procédures strictes permettant de respecter l'anonymat et la confidentialité des données, ces croisements de différentes bases de données permettent de constituer très rapidement des cohortes historiques qui n'avaient jamais été bâties (ni financées) dans cet objectif, et de répondre à des questions imprévues de tous ordres. C'est ainsi que les dangers de la prescription de bêta-2-mimétiques sans association à un traitement de fond chez l'asthmatique ont pu être confirmés (Spitzer *et al.*, 1992), ou que les risques d'hémorragies digestives liés à la prise prolongée d'anti-inflammatoires non stéroïdiens ont pu être quantifiés à partir d'une cohorte « électronique » de plus de 130 000 personnes exposées (Guess *et al.*, 1988).

Le cinquième exemple est de nature différente. Il a été développé au Laboratoire National de Los Alamos aux États-Unis afin notamment d'étudier par la modélisation la dynamique d'une épidémie de variole due à une attaque bioterroriste. **EpiSims** est un logiciel de simulation d'épidémies de maladies infectieuses basé sur un modèle individu-centré. Il utilise **Transims (Transportation Analysis and Simulation System)**, un recueil géant de données mis en place avec le Laboratoire national de Los Alamos, sur des financements du ministère nord-américain de la Défense, des Transports, de l'Énergie, la National Science Foundation, et le NIH (Fogarty International Center). Ce programme a mis en place un recueil de données sur les mouvements de chaque personne de la ville de Portland (1,6 millions d'habitants). Dans le réseau social étudié, tous les lieux visités par chaque habitant étaient répertoriés (184 000 lieux possibles différents dans la ville), seconde par seconde sur une période de 24 heures. La base de données ainsi constituée, d'une taille de 12 Go, a permis le développement du modèle épidémiologique. Parmi les tout premiers résultats obtenus, le modèle suggère qu'une stratégie de vaccination ciblée — et non la vaccination de masse comme la plupart des autres modèles auparavant le laissaient suggérer — ainsi qu'une détection rapide de l'attaque terroriste seraient les deux paramètres clés du contrôle épidémique (Eubank *et al.*, 2004). Ce projet a nécessité la mise en place d'une équipe de 40 chercheurs temps-plein à Los Alamos pendant huit ans et un budget de 30 millions de dollars.

En conclusion — L'épidémiologie populationnelle est un enjeu majeur pour la gestion du système de santé, tant dans ces aspects relatifs à la santé publique que ceux relatifs à la sécurité sanitaire.

La France a su montrer qu'elle avait de réels atouts en santé publique : elle a su constituer en moins de trente ans des équipes universitaires (et mixtes) actuellement reconnues sur le plan scientifique à un niveau international. Elle dispose, de par son histoire, d'une infrastructure sanitaire et de sécurité sociale très développée et très centralisée, et *de facto* de grandes bases de données constituées à des fins de gestion, mais qui auraient pu ou dû être aussi utiles — et quasiment sans coûts additionnels — à la santé publique et à l'épidémiologie, au moins au même titre que celles utilisées dans la petite province canadienne du Saskatchewan (1 million d'habitants). Certaines bases ont heureusement été largement exploitées en France aussi, comme celles de l'EDF-GDF ou des policiers de la ville de Paris, mais trop rarement et sans assez d'investissement public. La France reste absente de la plupart des projets majeurs depuis plus de trente ans. On a évoqué dans le paragraphe précédent les projets d'essais cliniques géants qui sont à la lisière de l'épidémiologie en population et de l'épidémiologie clinique, qui mobilisent les mêmes compétences, toujours multidisciplinaires. On peut évoquer les vastes cohortes (dont certaines sont menées à des coûts dérisoires en France, en comparaison avec les projets multi-millions de dollars nord-américains dont les budgets sont souvent votés par le Congrès lui-même, donc initiés avec une volonté politique clairement affichée). Ces cohortes ont pour objectif de suivre pendant de nombreuses années des populations entières, avec bien sûr une imagerie de haute technologie, des prélèvements et des analyses biologiques de dernière génération, mais aussi des entretiens psychosociaux mobilisant des ressources coûteuses en temps humain.

Plusieurs options sont possibles pour structurer de tels projets, nous en esquisserons deux :

- 1) la première consisterait à créer des « épidémiopôles » lourds, probablement en proposant des spécialisations régionales fortes, thématiques ou méthodologiques, et faciliter ainsi la constitution de centres d'excellences en épidémiologie répondant à des critères de taille critique, de compétences, de qualité, de ressources matérielles et d'équipements, de locaux, d'ancrages universitaire, médical (hospitalier ou en population) et de recherche. Il leur serait confié (et financé) de vastes études dont le pays a besoin, éventuellement (mais pas nécessairement uniquement) décidées à un niveau parlementaire, c'est-à-dire politique (voir la proposition « Plastico » de Marcel Goldberg dans l'encadré 2 ;
- 2) la seconde option serait une approche plus intégrative de la discipline avec les autres, notamment dans le champ des recherches en biomédecine.

Ainsi, par exemple, cette approche conduirait à renforcer la capacité et la présence de l'épidémiologie au sein des cancéropôles qui se mettent en place, ou bien autour des génopôles (où l'épidémiologie n'est quasiment pas développée). En leur attribuant des moyens suffisants pour rendre ces lieux de recherches multidisciplinaires attractifs, les laboratoires de recherche en épidémiologie ainsi créés ou relocalisés se doteraient de mathématiciens, statisticiens, et bio-informaticiens, mais aussi de chercheurs en sciences humaines (notamment, des sociologues, anthropologues, et géographes) pour conduire leurs programmes dans un environnement partageant des standards et une culture de très haute qualité.

Encadré 9.3-1

« PLASTICO » : le projet d'une PLAtforme Scientifique et Technique pluri-organismes pour l'aide à la gestion de COhortes et de grandes enquêtes épidémiologiques

La raison essentielle de la relative modestie des cohortes françaises, outre les problèmes de financement, tient à l'absence en France de dispositifs destinés à surmonter les difficultés techniques et logistiques inhérentes à la gestion des cohortes épidémiologiques longitudinales ayant une longue durée de suivi (souvent plusieurs décennies). On rencontre en effet un certain nombre de problèmes communs à toutes les cohortes longitudinales, qui nécessitent toujours un lourd travail de mobilisation et de coordination sur de longues périodes : suivi des sujets inclus ; accès à des sources de données diversifiées ; coûts élevés pendant de longues durées ; implication à long terme des équipes, qui manquent de moyens suffisants et adaptés à la durée des projets, et dont la pérennité n'est souvent pas assurée ; difficultés pour disposer de personnel spécialisé stable et d'un niveau de qualification suffisant, du fait de l'absence de statut reconnu pour ce type d'activité, alors que la durée des suivis de cohorte est incompatible avec un trop fort renouvellement des personnels techniques qui doivent assurer la continuité des procédures et des recueils de données.

La plupart des difficultés opérationnelles citées pour la gestion de cohortes se retrouvent également pour d'autres types d'enquêtes de grande taille. Ainsi, certaines études cas-témoins en population incluent actuellement plusieurs milliers de sujets et le recueil de données se déroule sur plusieurs années. Elles nécessitent également des procédures de validation de diagnostics, de codage et de saisie parfois très lourdes, ont des coûts élevés et nécessitent un personnel spécialisé stable et qualifié.

Un projet, actuellement à l'étude, de créer une « plate-forme » scientifique et technique pour l'aide à la gestion de cohortes et de grandes enquêtes épidémiologiques (projet *Plastico*), ouverte à la communauté scientifique des épidémiologistes, vise à faciliter la réalisation de certaines activités liées à de telles études. Il repose sur une mutualisation de moyens pour la gestion de cohortes et d'autres types de grandes enquêtes afin de proposer des prestations diversifiées. Une telle plate-forme devrait permettre la mise en commun de ressources, la mutualisation de compétences de provenances diverses, et favoriser le développement d'activités partagées et de synergies scientifiques. Répondant à la nécessité de structures pérennes pour des opérations de très longue durée, elle devrait offrir à la collectivité épidémiologique française des prestations diverses, dans des conditions de fonctionnement apportant des solutions à certaines difficultés évoquées plus haut, tout en induisant d'importantes économies d'échelle.

Les principales activités de la plate-forme scientifique concernent le développement et la maintenance de logiciels de gestion de cohortes et de grandes enquêtes, l'accès sécurisé à des grandes bases de données nationales (fichiers de la CNAV, SNIIR-AM, DADS, PMSI, ALD, CépiDc, etc.), le « traçage » de sujets inclus dans des cohortes, l'appariement sécurisé de données individuelles en provenance de bases de données nationales, la vérification et la validation des diagnostics, le codage, la saisie automatisée de données, etc.

Le volume et la diversité des données à gérer, la variété des sources d'information sollicitées pour le suivi des sujets, la taille très importante et la très longue durée de certaines des cohortes et autres grandes enquêtes, imposent un type de fonctionnement quasi « industriel », qui implique de réunir des moyens considérables, des compétences solides dans des domaines diversifiés et des outils techniques lourds. Le caractère « sensible » des données susceptibles d'être gérées par la plate-forme, les aspects liés à la propriété intellectuelle imposent des règles strictes de confidentialité et de déontologie. *Plastico* doit donc associer plusieurs organismes dont les équipes sont susceptibles de contribuer à la plate-forme et/ou de bénéficier des prestations qui seront proposées. Dans une première étape, il est prévu d'associer des équipes qui développent des cohortes pour lesquelles la constitution de certains éléments de la future plate-forme est de toutes façons indispensable pour leurs propres besoins, et qui ont déjà des collaborations en cours. Ces équipes appartiennent à l'Inserm, à l'InVS et à la Cnamts. Il est important de souligner que le projet *Plastico* n'implique ni création d'un organisme nouveau, ni coûts supplémentaires, au contraire, puisque d'importantes économies d'échelle seront immédiatement obtenues.

Marcel Goldberg

3 | L'épidémiologie clinique repose sur des standards de qualité justifiant des coûts élevés

Les standards de qualité de l'épidémiologie clinique ont été harmonisés et standardisés à un niveau international par l'International Conference on Harmonization (<http://www.ich.org>), dans un programme associant les autorités de santé en matière de réglementation du médicament et les industriels du secteur pharmaceutique.

Notre investigation auprès d'un laboratoire pharmaceutique majeur en France nous a conduit à estimer à 2 millions d'euros le coût d'un essai randomisé concernant 500 patients et comportant trois visites (inclusion, fin de traitement et suivi), un appel téléphonique, des prélèvements biologiques (honoraires des investigateurs et traitements des patients compris).

Le travail des assistants de recherche clinique (Arc) consiste principalement à se rendre chez les investigateurs et à comparer les données du cahier d'observation avec celles du dossier médical (pour cela l'investigateur a signé un accord préalable à l'essai, autorisant l'Arc à vérifier les documents sources). Nous considérons qu'un Arc peut valider, en France, entre 5 et 8 cahiers de 10 pages par jour, ou 1 à 2 cahiers d'observation de 60 pages. Pour un essai pivot de phase III, toutes les données doivent être vérifiées. Les phases III nécessitent un Arc pour un à deux centres (d'environ 8 patients) ; les études post-AMM (autorisation de mise sur le marché) nécessitent généralement un Arc pour suivre 75 à 100 centres. Un Arc fait généralement 15 visites sur site par mois (au minimum sur chaque site, l'Arc doit procéder à l'ouverture de centre et à la clôture de centre en fin d'essai). Les frais générés par ces déplacements sont de l'ordre de 800 euros par mois et par Arc.

Parmi les autres coûts humains, le dédommagement des investigateurs représente une part non négligeable pour les industriels du médicament (jusqu'à 150 euros par patient inclus dans un essai clinique classique comportant deux à trois consultations), mais aussi pour les recherches publiques qui impliquent le secteur libéral, bien que la pratique de l'indemnisation ne soit pas systématique. Si le réseau Sentinelles de l'Inserm ne dédommage pas les médecins généralistes qui fournissent des données sans discontinuité depuis novembre 1984, le General Practice Research Database (GPRD, détaillé ci-dessus), géré par la Medicines & Healthcare products Regulatory Agency (MHRA, l'agence britannique des produits de santé) indemnise les investigateurs généralistes libéraux de 25 pences par dossier et par an. Pour un cabinet de groupe de 10 000 patients (au Royaume-Uni, les patients sont inscrits auprès de cabinets regroupant généralement plusieurs généralistes), cela correspond à un revenu annuel de 2 500 livres (ou 4 000 euros).

4 | Quel serait le coût d'un vaste projet épidémiologique mis en place en 2005 en France ?

Nous prendrons, pour détailler les coûts d'un vaste essai clinique randomisé mené en France, un exemple récent où nous avons été conduits à effectuer le chiffrage du budget d'une recherche centrée sur l'évaluation du dépistage du cancer du poumon par une imagerie de dernière génération et qui prévoirait l'inclusion de 40 000 sujets dépistés en France pendant cinq années consécutives puis suivis pendant cinq années supplémentaires (un tel essai est actuellement mené par le NCI aux États-Unis, la France n'ayant jamais à ce jour participé à aucun essai d'évaluation pour le dépistage d'aucun cancer). Nous nous sommes placés dans le contexte où nous aurions à lancer des appels d'offres pour réaliser un certain nombre de prestations nécessaires pour la réalisation de l'essai selon les standards de qualité habituels. Nous détaillons ci-dessous les surcoûts liés à la mise en œuvre du projet épidémiologique. Nous n'avons pas évoqué les coûts supportés par les organismes de recherche, les universités, ou les hôpitaux et représentés par les salaires des fonctionnaires chargés de ces recherches, leurs locaux, leurs frais de fonctionnement, leur administration et leur gestion, généralement non (ou partiellement) imputés aux budgets de ces études. Pour établir ce chiffrage, nous avons rencontré plusieurs sociétés prestataires privées de monitoring basées en France ou à l'étranger. L'ensemble de nos conclusions menait à un projet dont le coût global de réalisation semblait légèrement inférieur, mais de même ordre de grandeur que celui actuellement consenti par un projet de même nature mené aux États-Unis (167 millions d'euros en France et 220 millions de dollars aux États-Unis) sur le même sujet par le NCI (<http://cancer.gov/NLST>).

Le recrutement de 80 assistants de recherche clinique (Arc), dont 10 Arc « managers » et 5 Arc « super-managers » est nécessaire pour encadrer le monitoring en respectant les bonnes pratiques épidémiologiques, pour un budget de 40 millions d'euros (sur 10 ans, coût des infrastructures informatiques exclues). L'indemnité des investigateurs lorsque ceux-ci sont médecins libéraux et quand l'essai est promu par le secteur public peut être de l'ordre de 20 euros par consultation liée à l'essai. Dans ce cadre, il est possible de confier à l'investigateur la charge de randomiser le sujet et de saisir les données du cahier d'observation électronique directement sur Internet.

Le coût d'un essai clinique ou d'une étude épidémiologique ne se limite pas aux seuls coûts d'indemnisation des investigateurs ou de monitoring, même si ces derniers sont des facteurs favorisant, voire garantissant la qualité des données. Leur saisie, leur transmission, leur analyse et leur archivage représentent des coûts humains importants et une logistique souvent lourde. L'utilisation de

plus en plus fréquente des cahiers d'observation électroniques, si elle simplifie les procédures de saisie et de transmission des données et renforce la qualité globale du système d'information des essais cliniques et des études épidémiologiques, représente des coûts informatiques non négligeables. Des logiciels existent sur le marché et proposent des solutions intégrées, gérant les différents aspects d'un essai ou d'une étude épidémiologique. Leur mise en place, leur maintenance, et leur fonctionnement font appel à une main-d'œuvre hautement qualifiée et coûteuse. Pour l'essai envisagé ici, un budget total de 10 millions d'euros est estimé nécessaire pour répondre aux besoins et couvrir les aspects spécifiques liés à l'archivage d'imagerie (notamment des scanners d'une taille informatique de 600 Mo chacun réalisés six fois chez 20 000 sujets), du traçage de la biobanque, du cahier d'observation électronique et de l'assistance au monitoring de l'essai.

Selon les termes de la loi sur la protection des personnes lors des recherches biomédicales, dite « loi Huriet », le promoteur d'un essai doit prendre en charge tous les frais médicaux afférents à un essai clinique. Ces frais dépendent du type d'essai et de sa durée. Pour notre exemple, la location de camions équipés de scanners de dernière génération, sillonnant les routes de France la nuit, et opérant la journée six jours sur sept, représente un coût de l'ordre de 1 million d'euros par camion et par an (quatre camions sont prévus pendant cinq ans pour assumer la charge de travail). Il n'est pas possible non plus de ne pas organiser (et rémunérer) la lecture professionnelle et rapide des 40 000 résultats d'imagerie qui chaque année seront produits par cet essai. Les devis selon que cette lecture serait organisée en France ou dans un pays de l'Est de l'Europe (Hongrie) vont respectivement de 19 millions d'euros à 9 millions pour l'ensemble de l'imagerie (27 000 scanners thoraciques, 20 000 radiographies de face et de profil, et 2 000 Tep-scanners par an pendant cinq ans).

Les études cliniques menées sur de nombreuses années demandent également une logistique lourde concernant les prélèvements biologiques : collecte de prélèvements, aliquotage des paillettes, congélation, acheminement vers une banque de prélèvements fonctionnant avec tous les critères de qualité requis, mais aussi de traçabilité et d'informatisation permettant leur conservation et leur libération pour les besoins de l'étude. L'utilisation de plates-formes de génomique fonctionnelle ou de protéomique, si elle doit traiter plusieurs milliers de prélèvements par an, peut nécessiter des équipements spécifiquement dédiés, avec les consommables et le personnel correspondants. Ces coûts ont été estimés de l'ordre de 20 millions d'euros pour 40 000 prélèvements, reproduits six fois sur dix ans.

Le coût des assurances d'une recherche prises par son promoteur représente une part non négligeable du budget total. Ici, ces frais sont estimés à 7 millions d'euros pour l'ensemble de la période concernée. Ils dépendent bien sûr

des risques encourus pour les sujets inclus dans l'essai (explorations invasives, éventuelle toxicité des traitements pour les essais thérapeutiques), et de la durée d'inclusion dans l'étude.

5 | Les atouts et les faiblesses de la France comme lieu de réalisation des essais cliniques

Les effectifs de recherche et développement (R&D) des entreprises du médicament étaient de 18 227 personnes en 2000 (ils étaient de 17 748 en 1995 et 11 175 en 1990). Mais on assiste à une délocalisation de la recherche et du développement de ces entreprises de l'Union européenne vers les États-Unis notamment : la part des dépenses localisées aux États-Unis du R&D des entreprises européennes concernées est en effet passée de 26 % en 1990 à 34 % en 1999. En dix ans aux États-Unis les dépenses de R&D ont été multipliées par cinq tandis que celles de l'Union européenne ont été multipliées par deux. Et seulement 10 % des essais cliniques européens ont eu lieu en France en 2002, cette proportion accusant une baisse quasiment linéaire de l'ordre de 5 % par an depuis 1995. Les efforts mondiaux de recherche portent sur les besoins de santé publique selon le Center for Medical Research International (2001), avec près de 20 % accordés au système nerveux, 13 % au cancer, 13 % aux maladies cardio-vasculaires, 11 % à l'alimentation et au métabolisme, 11 % aux anti-infectieux (divers domaines, 33 %). Si, en 1991, 52 nouvelles molécules avaient été autorisées dans le monde, ces nouvelles entités n'étaient plus que 32 en 2000 et 31 en 2001. Si l'Europe était à l'origine de quasiment deux fois plus de nouvelles molécules entre 1987 et 1991, les États-Unis ont pris le leadership mondial dans la période 1997-2001, à quasi-égalité avec l'Europe, le Japon ayant lui diminué par deux le nombre de nouvelles entités moléculaires autorisées sur le marché. La part des biotechnologies parmi les nouvelles entités moléculaires découvertes est en pleine croissance, passant de 5 à 15 % au début des années 1990 à 39 % en 2001. Aux États-Unis en 2002, 200 maladies étaient ciblées par 371 médicaments en développement issus des biotechnologies, et déjà 155 médicaments de ce type étaient disponibles dans le monde. Dans le domaine du cancer, les laboratoires pharmaceutiques ont investi en 2002 en France 339 millions d'euros en recherche clinique, correspondant à 29 600 malades inclus dans 488 essais cliniques répartis sur quinze centres de recherche. Les partenariats Inserm-Industrie concernaient, en 2002, 222 laboratoires de l'Inserm, 350 entreprises partenaires, 544 contrats de R&D et 475 accords de transfert. Avec le CNRS, la même année, 150 entreprises étaient partenaires de 490 contrats, et la majorité des redevances perçues sur les brevets du CNRS était issue du secteur des médicaments.

Une publication sur l'attractivité de la France pour les essais cliniques porte sur une étude réalisée auprès de 11 sociétés représentant 44 % du marché du médicament français en 2001 (d'Enfert *et al.*, 2003). Le budget investi en R&D était de 131 millions d'euros, dont 53 % en phases I, II et III, avec 1 469 personnes employées (Contract Research Organisations, CRO exclues). Plus de 21 000 investigateurs avaient été impliqués dans ces essais, et 98 000 patients avaient été recrutés sur des essais locaux (2 257 investigateurs et 10 270 patients sur des essais internationaux). Les secteurs principaux en France étaient celui du cardiovasculaire (36 %), du système nerveux central (14 %), de la rhumatologie et de l'inflammation (13 %) et du cancer (12 %). Au-delà des ressources internes, les laboratoires pharmaceutiques ont fait appel à des prestataires de service, notamment pour le monitoring des essais (33 % des budgets étant extériorisés vers des CRO sur ce poste dans les essais pré-AMM et 48 % en post-AMM), la gestion de projet (19 % et 31 %). Concernant le coût des essais cliniques, la France (base 100) occupe une place intermédiaire, plus chère que la Chine (57), les pays de l'Est de l'Europe (71), l'Italie ou l'Espagne (92), mais moins chère que l'Allemagne (114), le Royaume-Uni (129) et les États-Unis (160). Une autre étude publiée par Fast Track Systems (<http://www.fast-track.com>) en 2000 soulignait de manière plus marquée le faible coût des essais cliniques en France (en base 100), par rapport aux États-Unis (485), à l'Allemagne (276), au Royaume-Uni (256). Un classement des différents pays a été réalisé par les industriels (d'Enfert *et al.*, 2003) et situe la France en tête avec les Pays-Bas en termes de simplicité des procédures administratives et d'organisation du développement clinique local, mais les États-Unis sont classés premiers en termes d'importance du marché, des autorités d'enregistrement, des leaders d'opinion, d'accessibilité des patients et de disponibilité des ressources.

En conclusion — L'épidémiologie clinique est un enjeu majeur pour l'industrie des produits de santé. Rendre le pays attractif comme terrain d'expérimentation clinique est stratégique.

L'industrie du médicament en France a hissé en 2004 l'une de ses entreprises au troisième rang mondial, dans un secteur très compétitif. Pour ses activités de recherche et de développement, cette industrie a besoin de conduire des essais cliniques, notamment de phase III, mais aussi de vastes essais randomisés après l'enregistrement des produits de santé. La France a été absente des essais sur les traitements hormonaux substitutifs de la ménopause alors que plusieurs millions de femmes étaient exposées à ces produits. Les États-Unis et le Royaume-Uni conduisaient des vastes essais depuis plusieurs années, ce qui allait amener (en 2003) à remettre en cause le rapport bénéfice sur risque de ces médicaments. La France a été absente de tous les essais de dépistage des cancers conduits dans le monde. En particulier, aucun patient n'a été inclus en France dans les nombreux grands (ni petits) essais de dépistage du cancer du

sein, du colon, de la prostate, du col de l'utérus, de l'endomètre ou de la peau. On a vu que les critères majeurs des industriels permettant la réalisation des essais étaient : le contact avec les leaders d'opinion à un niveau international, l'accès aux patients volontaires pour participer à ces recherches, l'accès à des ressources locales. On sait que la France a de réels atouts en termes de simplicité des procédures administratives (par rapport aux autres pays européens, mais surtout par rapport aux États-Unis) et d'organisation du développement clinique local. Cependant, les délocalisations des essais cliniques sont de plus en plus nombreuses ces dernières années, non pas tant vers les pays du continent asiatique ou d'Europe de l'Est, mais essentiellement vers les États-Unis. Il convient de se poser la question de savoir si notre discipline, notamment dans sa composante d'épidémiologie clinique, prend suffisamment la mesure de ce problème. S'organise-t-elle autour de plates-formes dans les grands centres hospitaliers ou de cancérologie ? Regroupe-t-elle les compétences de biostatisticiens, de mathématiciens, de bio-informaticiens capable de mettre en œuvre des projets à la hauteur des nouveaux enjeux ? À part quelques exceptions notables, dont certaines ont été rappelées dans ce document, n'est-elle pas absente des recherches d'excellence dans le domaine ? Des tentatives intéressantes ont été réalisées avec l'Inserm et les centres hospitaliers universitaires en créant il y a 20 ans les centres d'investigation clinique, mais l'implication plutôt faible des épidémiologistes et biostatisticiens en ont fait des terrains d'expérimentation réservés aux essais de phases I et II, c'est-à-dire davantage tournés vers des questions d'ordre pharmacologique que statistique ou biomathématique. Ne faut-il pas développer désormais (ou renforcer quand ils existent) des centres d'excellence en épidémiologie clinique, capables d'attirer sur notre sol les plus grands essais cliniques de phase III des industriels internationaux, capables de mener de vastes essais rivalisant avec ceux qui se déroulent depuis plusieurs dizaines d'années en Europe du Nord et aux États-Unis ?

Bibliographie

d'Enfert J. *et al.* (2003). *Thérapie*, **58** : 283-89.

Eubank S., Guclu H., Kumar VS., Marathe MV., Srinivasan A., Toroczkai Z., Wang N. (2004). Modelling disease outbreaks in realistic urban social networks. *Nature*, **429** : 180-4.

Guess HA., West R., Strand LM., Helston D., Lydick EG., Bergman U., Wolski K. (1988). Fatal upper gastrointestinal hemorrhage or perforation among users and nonusers of nonsteroidal anti-inflammatory drugs in Saskatchewan, Canada, 1983. *J Clin Epidemiol*, **41** : 35-45.

- Marrie RA., Wolfson C., Sturkenboom MC., Gout O., Heinzlef O., Roullet E., Abenheim L. (2000). Multiple sclerosis and antecedent infections : a case-control study. *Neurology*, **54** : 2307-10.
- Spitzer WO., Suissa S., Ernst P., Horwitz RI., Habbick B., Cockcroft D., Boivin JF., McNutt M., Buist AS., Rebeck AS. (1992). The use of beta-agonists and the risk of death and near death from asthma. *N Engl J Med*, **326** : 501-6.

SOUS-CHAPITRE 9.4

Études épidémiologiques en santé : aspects législatifs et réglementaires

JEAN-CHRISTOPHE THALABARD¹

Le terme « épidémiologie » recouvre, au sens large, des situations très diverses, études descriptives en population, études analytiques, études d'intervention dont les essais cliniques. Le point commun est la constitution d'un échantillon de sujets au sein d'une population d'intérêt, sur lequel des informations (données démographiques, données de santé, données familiales, paramètres biologiques, expositions, etc.) sont recueillies soit rétrospectivement, soit prospectivement. Dans un certain nombre de cas, tout ou partie de ces données concernant l'échantillon a déjà pu être recueilli dans une finalité éventuelle initiale différente, soit au cours d'études antérieures, soit grâce à des systèmes de recueils permanents organisés (observatoires de santé, registres, voire système de soins), sous réserve de pouvoir recouper ces données.

La déclaration d'Helsinki (2000, <http://www.wma.net>) indique que « toute précaution doit être prise pour respecter l'intimité des individus et minimiser l'impact de l'étude sur l'intégrité physique et psychique ». Cette dualité se retrouve en droit français avec des principes généraux, dont l'article 16 du Code civil

qui énonce notamment que « la loi assume la primauté de la personne, interdit toute atteinte à la dignité de celle-ci et garantit le respect de l'être humain dès le commencement de la vie » et des dispositions spécifiques concernant, d'une part, la protection des données personnelles notamment dans le domaine de la santé, d'autre part, l'atteinte au corps et la nécessaire protection de l'intégrité physique de la personne dans le cas de recherches biomédicales.

1 | Les dispositifs protégeant la vie privée

Ils constituent un corps de règles résultant, outre de la loi, de règlements internes de sociétés savantes, de principes émanant d'instances représentatives et enfin de règles de bonnes pratiques entre professionnels.

1.1 La loi relative à l'informatique, aux fichiers et aux libertés (loi 78-17 ; 06/01/1978)

Elle « s'applique aux traitements automatisés de données à caractère personnel ainsi qu'aux traitements non automatisés de données à caractère personnel contenues ou appelées à figurer dans des fichiers, à l'exception des traitements mis en œuvre pour l'exercice d'activités exclusivement personnelles ».

La constitution de fichiers à des fins de recherche s'inscrit dans un cadre juridique fixé par les dispositions des articles 53 et suivants de la loi du 6 janvier 1978 modifiée, issues de la loi du 1^{er} juillet 1994 et regroupées dans le chapitre 9 relatif aux traitements de données à caractère personnel ayant pour fin la recherche dans le domaine de la santé. Sont notamment concernées par ces dispositions les études épidémiologiques et les recherches biomédicales relevant de la loi Huriet du 20 décembre 1988.

Ces dispositions ont pour principal objet de donner une base légale aux communications de données à caractère personnel médicales nécessaires à la constitution de fichiers de recherche en permettant, sous certaines conditions, une levée du secret professionnel.

Le législateur a souhaité ainsi renforcer les procédures de contrôle sur ces fichiers qui doivent faire l'objet d'une autorisation préalable quel que soit le statut juridique de l'organisme concerné. Dans un premier temps, l'avis d'un Comité consultatif sur le traitement de l'information en matière de recherche dans le domaine de la santé (CCTIRS), créé par la loi, est requis. Il est chargé d'apprécier, sur le plan scientifique, la méthodologie de chaque projet de recherche,

la nécessité du recours à des données à caractère personnel et la pertinence de celles-ci par rapport à l'objectif de la recherche.

La commission nationale de l'informatique et des libertés (Cnil) doit ensuite être saisie d'une demande d'autorisation qui comportera, outre les éléments préalablement adressés au Comité, la description des caractéristiques techniques du système informatique utilisé, les mesures de sécurité mises en œuvre pour garantir la confidentialité des données et les modalités pratiques mises en place pour assurer l'information individuelle des personnes concernées par le traitement.

Pour certaines catégories de traitements automatisés ayant pour finalité la recherche dans le domaine de la santé et portant sur des données ne permettant pas une identification directe des personnes concernées, la loi du 6 janvier 1978 modifiée par la loi du 6 août 2004 reconnaît à la Cnil la possibilité d'homologuer et de publier des méthodologies de référence, établies en concertation avec le comité consultatif ainsi qu'avec les organismes publics et privés représentatifs, et destinées à simplifier la procédure prévue. La Cnil a d'ores et déjà mis en œuvre depuis plusieurs années une telle simplification pour les recherches biomédicales relevant de la loi Huriet. Elle devrait poursuivre dans cette voie.

Une difficulté fréquente pour le chercheur concerne, d'une part, la capacité de mettre une limite aux variables d'intérêt et aux conditions d'exploitation dans le futur et, d'autre part, la description physique du système informatisé utilisé et des moyens de protection vis-à-vis de ses connections éventuelles à des réseaux, du fait du développement d'un parc micro-informatique plus ou moins anarchique, aux caractéristiques mal connues, et d'idées fausses sur les protections conférées par des machines isolées. Il en résulte encore souvent un point de vue individualiste, peu professionnel, avec une fréquente distorsion entre ce qui est annoncé et la réalité du terrain, une dispersion des supports de stockage, un archivage plus ou moins organisé dans le temps et le risque fréquent de disparition des données au départ du responsable de l'étude ou au gré des changements de matériels.

Au-delà du stockage de données propres à une étude, se pose la question de leur partage. Dans les conditions déterminées par la loi, les données de santé ne peuvent être communiquées et utilisées, sous réserve du consentement exprès de l'intéressé, que dans l'intérêt direct du patient ou pour les besoins de la santé publique. Ces points ont fait l'objet de textes de loi complémentaires : la loi n° 94-548 du 1/07/1994, relative au traitement de données nominatives ayant pour fin la recherche dans le domaine de la santé, aborde la transmission entre professionnels de santé de données personnelles de santé dans le cadre d'études épidémiologiques, observationnelles, essais cliniques et pharmacovigilance, sous réserve de respecter des règles spécifiques, réaffirmant le principe

obligatoire d'une information préalable individuelle des sujets concernés et de leur droit de refus.

La loi n° 99-641 du 27/07/1999 a complété la loi informatique et libertés d'un chapitre X particulier sur les traitements de données de santé à caractère personnel à des fins d'évaluation ou d'analyse des pratiques ou des activités de soins et de prévention, qui précise les conditions d'échanges de données indirectement nominatives quelle qu'en soit l'origine (dossiers médicaux, dossiers hospitaliers, Sécurité sociale).

Les dispositions des articles 62 et suivants de la loi du 6 janvier 1978 modifiée ont pour objet de préciser les conditions dans lesquelles les données de santé, qu'elles soient issues des professionnels de santé eux-mêmes, des systèmes d'information hospitaliers, ou des fichiers des caisses de Sécurité sociale, peuvent être diffusées et exploitées à des fins d'évaluation des pratiques de soins et de prévention.

Le principe reste l'anonymat qui doit présider à la transmission des données tant aux autorités sanitaires qu'aux tiers : la communication de ces données ne peut être effectuée que sous la forme de statistiques agrégées ou de données par patient constituées de telle sorte que les personnes concernées ne puissent être identifiées. La dérogation au principe réside dans la possibilité de transmettre des données indirectement nominatives sous réserve notamment qu'elles ne comportent ni le nom, ni le prénom du patient, ni son numéro de Sécurité sociale et que la communication des données soit autorisée par la Cnil. Ces traitements peuvent être constitués à partir des données issues des systèmes d'information visés à l'article L. 6113-7 du Code de la santé publique (c'est-à-dire des résumés de sorties anonymes du PMSI), des données issues des dossiers médicaux détenus dans le cadre de l'exercice libéral des professions de santé et des données issues des systèmes d'informations des caisses d'assurance maladie.

La communication des noms, prénoms et NIR des personnes reste exclue : les traitements concernés ne peuvent servir à des fins de recherche ou d'identification des personnes. Il convient de noter que ces dispositions n'ont pas vocation à s'appliquer aux fichiers de recherche dans le domaine de la santé.

La loi n° 2004-801 du 6 août 2004 relative à la protection des personnes physiques à l'égard des traitements de données à caractère personnel et modifiant la loi du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés correspond à la transcription en droit français de la directive européenne 95/46 (JO CE 23/11/1995) relative à la protection des données personnelles et à la libre circulation de ces données. Ce texte introduit une notion plus large d'identifiabilité, en étendant la notion de sources, puisqu'il est stipulé que « *constitue une donnée à caractère personnel toute information relative à une personne*

physique identifiée ou qui peut être identifiée, directement ou indirectement, par référence à un numéro d'identification ou à un ou plusieurs éléments qui lui sont propres. Pour déterminer si une personne est identifiable, il convient de considérer l'ensemble des moyens en vue de permettre son identification dont dispose ou auxquels peut avoir accès le responsable du traitement ou toute autre personne. » Il précise également les notions de traitement de données à caractère personnel et de fichier de données à caractère personnel.

1.2 La charte des bonnes pratiques épidémiologiques

L'Association des épidémiologistes de langue française (Adelf) a élaboré un ensemble de recommandations professionnelles concernant la pratique des études épidémiologiques, périodiquement mis à jour, largement inspiré des principes et recommandations internationales (dernière mise à jour en 2004). Concernant les données sensibles, les préoccupations des professionnels de la santé rejoignent celles des professionnels d'autres domaines (Insee, Ined, sociétés de sondage). Le respect des règles déontologiques est à la base du contrat qui lie ces professionnels à leurs employeurs, publics ou privés, pour les missions qui leur sont confiées.

Fait nouveau, la loi indique (article 11-2) que la Cnil peut donner un avis sur la conformité des règles professionnelles aux dispositions de la loi et apprécier les garanties offertes par celles-ci.

2 | La protection de l'intégrité physique des personnes

Elle concerne plus directement les études comportant une atteinte au corps de la personne et repose sur une série de dispositifs juridiques introduits progressivement depuis 1988, qui permettent, sous certaines conditions, de « *porter atteinte au corps d'autrui* » à des fins non plus thérapeutiques mais de recherche.

2.1 La loi n° 88-1138 du 20/12/1988 dite « loi Huriet-Sérusclat »

Elle reste la plus connue. Ses dispositions, initialement pensées pour les seules études expérimentales menées chez des sujets volontaires, malades ou non, dans le cadre du développement de nouvelles thérapeutiques, ont visé, secondairement, l'ensemble des recherches menées sur des sujets volontaires dès lors que

l'étude était une « recherche biomédicale » c'est-à-dire « toute recherche organisée et pratiquée sur l'être humain en vue du développement des connaissances biologiques et médicales ». Ce texte introduisait des obligations et des devoirs concernant notamment :

- une répartition précise des rôles des acteurs de la recherche ; d'une part, l'investigateur, interlocuteur auprès des personnes qui se prêtent à la recherche, obligatoirement médecin (sauf exception en chirurgie dentaire) titulaire d'une thèse, inscrit au Conseil de l'ordre, d'autre part, un promoteur, personne physique ou morale, qui commandite l'étude, en assure la responsabilité civile et pénale avec souscription d'une assurance spécifique adaptée à l'étude et interagit avec les services désignés par le ministre de la Santé ;
- l'instauration d'un regard externe sur les pratiques, par la mise en place de comités de protection des personnes se prêtant à des recherches biomédicales (CPPRB), chargés de donner un avis sur le protocole de recherche biomédicale, avant déclaration d'intention par le promoteur aux autorités de santé ;
- une distinction entre recherches « avec bénéfice direct » et « sans bénéfice direct » avec des clauses particulières pour ces dernières, visant à assurer une protection renforcée de la santé des volontaires rentrant dans ces études ;
- enfin, il était stipulé que seules les personnes bénéficiant d'un régime de protection sociale pouvaient être volontaires pour participer à une recherche biomédicale.

Ce cadre, par construction tout à fait adapté aux recherches menées par les professionnels du médicament, a marqué un changement radical dans les pratiques des recherches à support public institutionnel, rarement confrontées à un regard externe sociétal sur leurs pratiques. Les chercheurs n'ont pas manqué d'en souligner certaines difficultés et ambiguïtés, maintes fois débattues.

La recherche clinique déborde en fait largement l'expérience menée sur un groupe de sujets « instrumentalisés ». Ainsi, depuis 1988 ; les recherches à partir de collections de matériel biologique se sont multipliées. Il était difficilement compréhensible pour un chercheur du champ biomédical que la définition juridique de « recherche biomédicale » ne recouvre pas ce qu'il considérait comme une recherche biomédicale. De plus, si le législateur avait abordé explicitement en 1994 (loi n° 94-653 du 29/07/1994) l'utilisation de matériel de nature génétique, en distinguant pratique médicale et finalité de recherche, la rédaction du texte entretenait une confusion en renvoyant « pour les modalités de prélèvement et de recueil de consentement » aux dispositions concernant les recherches

biomédicales, objet d'une interprétation extensive par la plupart des acteurs de la recherche biomédicale : la constitution d'une collection d'échantillons sanguins sans hypothèse *a priori* de recherche (DNAthèque) rentrait ainsi dans le cadre de la loi du 20/12/1988, avec toutes ses dispositions (lieux habilités, assurance spécifique et sa garantie décennale, investigateur médecin, inscription sur le fichier, etc.) alors que des juristes rappelaient régulièrement la confusion entretenue entre une recherche menée **sur** la personne, sujet de droit, et une recherche menée à partir d'un prélèvement de matériel biologique, élément du corps humain.

Se placer dans ce cadre permettait également au chercheur de bénéficier d'un avis « éthique » à travers les CCPPRB, souvent exigé des financeurs, ou du régime spécial d'indemnisation pour les volontaires. . . C'est ainsi que épidémiologistes, psychologues, ergonomes ou généticiens ont été confrontés à des difficultés multiples liées à cette interprétation extensive et réduits à inventer des stratagèmes pour contourner ce qui était perçu comme un obstacle, avec tous les effets pervers que cela a pu générer.

Une autre critique a porté sur les lenteurs et lourdeurs générées, que ce soit :

- au stade initial de soumission et mise en œuvre d'un projet, avec multiplication d'étapes obligatoires séquentielles, aux formats non uniformisés, sollicitation d'experts multiples, délais de réponse mal codifiés ;
- ou lors du suivi, avec les obligations de déclaration des effets indésirables graves plus ou moins adaptées à la nature des recherches (EIG), les obligations de stockage de données, suivi laissé à l'initiative du chercheur et souvent mal organisé sur le long terme, faute de moyens.

2.2 De la propriété des données et de l'obligation d'informer

Les textes encadrant les recherches biomédicales précisent que les données appartiennent au promoteur. En pratique, l'investigateur d'une étude à promoteur institutionnel, aux financements souvent multiples et dissociés, a souvent tendance à considérer qu'elle est sa propriété, contrairement au domaine des essais thérapeutiques à promotion industrielle. La possibilité d'échanges de données est alors directement liée à la possibilité, plus ou moins stable au cours du temps, d'échanges avec l'investigateur.

La possibilité même d'échange de matériel biologique et de valorisation associés à ces études reste rarement abordée clairement dans les feuilles d'information et/ou le formulaire de consentement, alors qu'elle constitue un enjeu

actuel important, comme cela a pu être mis en évidence avec la détermination de gènes de susceptibilité à des cancers familiaux.

Plus récemment, la loi n° 2002-303 du 04/03/2002 relative aux droits des malades et à la qualité du système de santé, précise que toute personne participant à une recherche biomédicale et, plus généralement, à la collection de données de santé à des fins de recherche :

- doit avoir donné son consentement pour une telle utilisation ;
- doit pouvoir être informée, sur sa demande, non seulement des données le concernant mais également des résultats de la dite recherche ;
- doit pouvoir s'assurer que l'utilisation des informations et/ou prélèvements biologiques est conforme à ce à quoi elle a consenti.

2.3 Les dispositions récentes : loi n° 2004-806 du 09/08/2004

La révision des dispositions encadrant les recherches biomédicales a fait l'objet d'une large concertation auprès des acteurs et usagers du système de santé et/ou de la recherche à l'occasion de la transposition en droit français de la directive européenne 2001/20/CE relative aux essais cliniques de médicaments. Ce nouveau cadre correspond, très symboliquement, au chapitre II de la loi n° 2004-806 relative à la politique de santé publique. Cette loi concerne essentiellement les essais cliniques. Les éléments marquants en sont :

– *l'introduction de procédures simplifiées*. Celles-ci concernent tout particulièrement l'épidémiologie descriptive ou analytique, les incluant dans le champ de la loi mais avec des modalités simplifiées :

« Les dispositions [...] ne s'appliquent pas aux recherches dans lesquelles tous les actes sont pratiqués et les produits utilisés de manière habituelle et lorsque aucune procédure supplémentaire ou inhabituelle de diagnostic ou de surveillance n'est appliquée. »

« Par dérogation [...], les recherches biomédicales ne portant pas sur des médicaments qui ne comportent que des risques négligeables et n'ont aucune influence sur la prise en charge médicale de la personne qui s'y prête peuvent être effectuées sous la direction et la surveillance d'une personne qualifiée. »

« Par dérogation [...], les recherches biomédicales ne portant pas sur des médicaments, qui ne comportent que des risques négligeables et n'ont aucune

influence sur la prise en charge médicale de la personne qui s'y prête peuvent être réalisées sans examen médical préalable. » ;

– *des possibilités de recherche chez des personnes protégées par la loi.* Elles reflètent une logique très utilitariste. Il devient possible de déroger à la règle du consentement éclairé préalable dans des situations particulières de recherche en situation d'urgence ou d'altération de la conscience, avec l'introduction d'un tiers ou « personne de confiance ». Des clauses, certes restrictives, permettent certaines recherches chez des enfants, des femmes enceintes, des personnes privées de liberté. Enfin, l'obligation du consentement simultané des deux titulaires de l'autorité parentale dans des recherches pédiatriques peut être levée dans certaines conditions d'urgence ;

– *l'abandon de la distinction BID/SBID au profit d'une estimation de la balance « bénéfiques risques ».* Ceci traduit une évolution notable, notamment en ce qui concerne la responsabilité pour faute présumée du promoteur, au rôle renforcé. Il est supposé assumer dorénavant la gestion de l'étude et garantir son financement ;

– *la transformation des comités consultatifs de protection des personnes qui se prêtent aux recherches biomédicales (CCPPRB).* Ils deviennent des comités de protection des personnes (CPP), en s'ouvrant aux « représentants des malades et usagers du système de santé ». Saisis dorénavant par le seul promoteur, leur avis devient obligatoire et engage la responsabilité de l'État. Les recherches biomédicales passent d'un système de déclaration d'intention à un régime d'autorisation par les autorités de santé ;

– *une modification des conditions d'autorisation des lieux de recherche.* L'autorisation administrative n'est plus liée à un site mais envisagée par rapport à l'exercice habituel de l'acte pratiqué aux fins de recherche, prenant en compte la qualification des personnes impliquées ;

– *un effort de diffusion d'information sur les recherches en cours.* Une base de données nationales des recherches biomédicales autorisées est créée avec diffusion des informations auprès des associations de malades et usagers du système de santé ;

– *une limitation de l'inscription au fichier national des volontaires.* Elle ne concerne plus que les personnes se prêtant à des recherches relevant de l'Afssaps et/ou impliquant des volontaires sains ou des malades pour lesquels les bénéfiques escomptés sont sans rapport avec l'état pathologique ;

– *la possibilité pour certaines recherches de soumettre en parallèle au CCTIRS et à un CPP ;*

- *l'introduction de la possibilité de croisement de sources de données agrégées par des personnes habilitées pour une période donnée.*

3 | Éléments de comparaison internationale

En 1971 apparaissent les premiers comités d'éthique clinique dans des hôpitaux catholiques canadiens. Parallèlement, aux États-Unis, une commission nationale de protection des sujets participant à des recherches biomédicales et comportementales était mise en place. Le rapport Belmont date de 1974, tandis que la première version de la déclaration d'Helsinki est de 1975. La France s'est dotée dès 1978 d'une loi « Informatique et libertés » et en 1983 d'un Comité consultatif national d'éthique pour les sciences de la vie et de la santé (décret 83-132 du 23/02/1983), organisme purement consultatif (loi n° 94-654 du 29/07/1994 ; décret n° 97-555 du 29/05/1997).

Il apparaît que, même dans les pays réputés plus « ouverts », le simple critère de facilité et d'efficacité pour le chercheur ne justifie jamais une exception à une règle commune fondée sur un équilibre nécessaire entre l'intérêt général, qui vise à comprendre et prévenir les maladies, et la nécessité de préserver la vie privée. Cependant, les pays de culture marquée par le courant « utilitariste » ont une appréciation différente du secret médical et de l'équilibre entre respect de la vie privée et risques collectifs, ouvrant des possibilités notamment de croisement de sources de données multiples médicales et non médicales, encore difficiles dans le système de santé français.

3.1 Les pays européens (Sénat, LC n° 89, 2001)

Il convient de distinguer les pays qui ont une longue expérience des études épidémiologiques (Grande-Bretagne, pays scandinaves), des pays du sud de l'Europe (Espagne, Portugal, Italie notamment) aux règles plus récentes et moins établies. La directive européenne 95/46/CE du 24/10/1995, relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à leur libre circulation, interdit « le traitement de données à caractère personnel qui révèlent l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques, l'appartenance syndicale » ainsi que celles « relatives à la santé et à la vie sexuelle », sauf consentement exprès et explicite de l'intéressé et autorisation par la loi, le règlement et/ou l'autorité de contrôle. Cette directive 95/46 est maintenant transposée dans tous les pays européens, avec une extension, pour certains pays, aux données manuelles et à toutes données personnelles quel qu'en soit le support.

Royaume-Uni

Les comités d'éthique dans les hôpitaux et instituts de recherche remontent à 1960. Le Nuffield Council of Bioethics, organisme privé, datant de 1991, a précédé la mise en place par le National Health Service (NHS) de comités d'éthique de la recherche (REC), organisés en réseau. Aucune recherche ne peut démarrer si elle n'a pas obtenu l'accord du comité REC consulté. L'investigateur principal, non forcément médecin, doit donner les détails des coûts (personnel, matériel, médicaments, etc.) hors ceux pris en charge par le NHS, démontrer que la gestion des données est en conformité avec le « Data Protection Act 1998 » et donner son accord pour que la recherche figure dans le registre national des recherches (National Research Register), sauf secret industriel. Enfin, si l'obligation d'assurance spécifique n'apparaît pas explicitement mentionnée, il doit être capable d'expliciter les précautions et arrangements pris en cas de plainte par un participant, en distinguant la faute par négligence de la faute sans négligence.

Une commission indépendante pour la génétique humaine, mise en place en 1999, a élaboré, en 2001, des recommandations concernant l'informatisation, la protection et l'utilisation des données génétiques personnelles.

Concernant plus spécifiquement les études épidémiologiques, il persiste une zone « grise » entre la recherche et les études purement descriptives à partir de données provenant du suivi normal de patients. Il est possible, cependant, d'utiliser des données médicales à des fins de recherche, sous réserve de passer par le médecin référent (GP).

Point important, la soumission d'un dossier au REC est entièrement en ligne avec un dossier type dans lequel le demandeur coche des situations décrivant son projet de recherche, avec une grande lisibilité à chaque instant des recherches en cours.

Danemark

Le Danemark a une longue tradition de centralisation des données nominatives de santé avec possibilité de réalisation de recherches sur ces données, y compris des données sensibles. Depuis 1987, il existe un Comité national d'éthique institué par la loi avec une compétence générale. Il présente en son sein neuf membres désignés par une délégation parlementaire spécifique. Il entretient des liens privilégiés avec le Parlement.

Belgique

La Belgique s'est dotée assez tôt d'un comité d'éthique au sein du FNRS (1979), puis, depuis 1994, de comités éthiques hospitaliers et depuis 1996 d'un Comité national consultatif de bioéthique. Il existe actuellement près de 200 comités d'éthique publics et privés, avec un manque de visibilité pour le chercheur clinicien et une dilution des responsabilités.

Portugal, Espagne et Italie

Ils bénéficient d'un Conseil national d'éthique pour les sciences de la vie, strictement consultatif. Les règles reposent sur la transposition de la directive européenne n° 95/46/CE déjà accomplie.

Allemagne, Suisse

En Allemagne, chaque Land bénéficie d'une commission de bioéthique, avec une commission centrale d'éthique, indépendante, créée en 1994. Les données sensibles sont protégées par des lois fédérales transposées de la loi européenne (1997, 1990). En Suisse, une loi fédérale du 18/12/1998 a mis en place une commission nationale d'éthique pour la médecine humaine.

3.2 Pays non européens

Australie²

La législation fait une distinction claire entre études d'observation et études expérimentales. La veille en santé publique doit être distinguée de la recherche épidémiologique : les agences gouvernementales ont pour mission des actions de surveillance, et sont autorisées, le cas échéant, à mener des enquêtes à visée explicative, qui, en cas d'urgence, peuvent être dispensées d'un examen préalable par un IRB. En revanche, les chercheurs doivent solliciter un avis d'un HREC (NHMRC. Guidelines Section 95, Privacy Act, 1988). Le consentement individuel s'impose lorsque les études épidémiologiques impliquent un contact personnel entre l'investigateur et les sujets participants, mais peut ne pas s'appliquer dans des études de population.

²<http://www.health.gov.au/nhmrc/hrecbook/>

Canada

Dans les recommandations du MRC de 1987, le respect de la confidentialité est une règle stricte renforcée par le principe selon lequel les patients ne peuvent être approchés par des personnes « étrangères ». La notion d'« étranger » concerne ici autant le rapport à la personne que le fait d'appartenir à une profession médicale ou paramédicale, d'où la nécessité de passer par le médecin habituel. La possibilité d'un contact direct par le chercheur dépend de l'acceptation par le médecin traitant.

Il existe un large encouragement aux croisements de bases de données administratives et médicales : la notion de *passive approval* permet le croisement de sources de données même en l'absence de consentement formel de la part des contributeurs, sous réserve de conditions strictes, portant sur une agrégation minimale des données, de sorte que la stratification ne permette jamais d'avoir moins de cinq individus avec le même profil.

Les comités éthiques d'établissement (Institutional Review Board), accrédités par le NIH, sont tout-puissants pour accepter ou refuser qu'une étude se déroule dans un établissement, dans la mesure où tout ce qui se passe dans un lieu de soin, y compris le recueil des consentements, est sous la responsabilité de l'hôpital, sans délégation possible. L'activité de recherche est couverte par les mêmes assurances que l'activité clinique. De plus, les membres d'une équipe financée doivent suivre une formation de deux à trois heures avec certification avant de recevoir le financement.

États-Unis

Les recherches sur des personnes volontaires sont très contrôlées. Le contrôle s'exerce au niveau de l'OPRR (Office for Protection from Research Risks), de l'OHRP (Office for Human Research Protections) qui dépendent du US Department of Health and Human Services, et surtout des IRB. Les règles fédérales (Code of Federal Regulations CFR titre 45, partie 46, sous partie A — 45 CFR 46 — ou « règle commune » codifiée par le département de la santé des populations) concernant la protection des personnes se prêtant à des recherches incorporent les trois grands principes du rapport Belmont³ : respect de la personne, obligation de bienfaisance, équité et justice.

³National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. Rapport Belmont : *Ethical principles and guidelines for the protection of human subjects of research*. Washington DC : US government printing office, 1979.

Les études épidémiologiques posent des problèmes spécifiques, dans la mesure où elles utilisent souvent des documents sensibles d'ordre privé en les croisant avec d'autres sources de données comme les données de travail, d'assurances ou de police : l'épidémiologiste doit soumettre son protocole de recherche pour revue indépendante à un International Review Board (IRB), et par la suite obtenir de chaque sujet volontaire un consentement éclairé. Il s'engage à maintenir la confidentialité, sauvegarder les droits et la santé des individus et/ou des groupes étudiés.

3.3 Les IRB (International Review Board)

Groupe désigné formellement pour superviser et suivre les recherches biomédicales impliquant des sujets humains vivants, l'IRB a pour objectif de s'assurer, avant la mise en place d'une recherche et au cours de celle-ci, que les mesures appropriées sont prises pour protéger les droits et la santé des personnes participant aux recherches. L'IRB a autorité pour approuver, demander des modifications ou arrêter une recherche.

La notion d'assurance sur un projet unique ou d'assurance sur plusieurs projets ou groupe d'études dans un domaine particulier correspond à un document écrit négocié entre une institution et le département de la santé (HHS) selon les règles HHS, assurant que l'institution s'engage à satisfaire des règles particulières du CFR.

Le code fédéral isole une liste des recherches pouvant être examinées par les IRB dans le cadre d'une procédure simplifiée (45 CFR 46.110 et 21 CFR 56.110) : cette liste précise et détaillée concerne des recherches n'impliquant qu'un risque minime ou des modifications mineures sur des protocoles préalablement approuvés, sous réserve de certains critères. La procédure de recueil de consentement éclairé reste identique. Citons, à titre d'exemple, les études cliniques portant sur des médicaments déjà commercialisés ou des dispositifs médicaux homologués utilisés selon les recommandations associées, sans surcroît de risque individuel, les collections de sang par ponction, avec conditions très précises des quantités et fréquences, fonction de l'âge, du poids et de l'état ou non de grossesse, les collections prospectives d'échantillons biologiques à des fins de recherche par des moyens non invasifs, dans le cas où les modalités de prélèvements ne comportent pas de risque particulier, etc.

3.4 Protection des personnes et Health Insurance Portability and Accountability Act – 1996 (HIPAA, 1996, Medical Privacy : National Standards to protect the privacy of personal health information)

Le congrès américain a voté l'HIPAA, destiné à protéger la vie privée des personnes tout en facilitant les contacts entre professionnels de santé grâce aux échanges électroniques. Appliquée depuis 2003, cette loi a suscité de nombreuses inquiétudes et débats au sein de la communauté des épidémiologistes. Il est désormais nécessaire d'obtenir l'accord du médecin traitant pour tout sujet destiné à être inclus dans une étude avant de pouvoir soumettre un protocole à un comité.

Contrairement à la situation antérieure, où un accès était possible à des contrôles provenant des services de l'aide médicale (Medicare, Medicaid) ou des sujets hospitalisés, l'accès aux sujets hospitalisés nécessite l'accord des IRB correspondants. L'hôpital a le droit de refuser ou révoquer un accord à tout moment. Le formulaire de consentement est propre à chaque hôpital. Une grande importance est donnée aux témoins, avec nécessité d'actes notariés et preuve d'un lien familial avec un rôle accru des juges. Des copies du protocole doivent être distribuées. En revanche, une fois l'accord obtenu, les données du dossier peuvent être communiquées à d'autres entités que les médecins, à savoir les IRB eux-mêmes, des investigateurs associés, etc., avec le risque de diffusion d'informations confidentielles et la menace d'audits fédéraux et de plaintes avec engagement de la responsabilité civile et pénale.

4 | Le cas particulier des recherches gènes-environnement

Les textes ont essayé de distinguer, d'une part, les études familiales concernant les variants de gènes hautement pénétrants, souvent associés à des formes de maladies donnant des agrégations familiales, où la divulgation d'informations expose au risque de préjudice social et psychologique, et d'autre part, les études en population, plus tournées vers les interactions gènes-environnement aboutissant à une meilleure prédiction des profil de risques. Le rapport, intitulé *Research involving human biological material : ethical issues and policy guidance*, a précisé les règles à appliquer, qui tentent de concilier protection individuelle et souci de ne pas entraver la recherche en santé publique (National Bioethics Advisory Commission, 1999).

En fait, à quelques exceptions près, la règle commune s'applique partout. Des recherches financées par des sources non fédérales peuvent être soumises aux règles fédérales si la recherche est soumise à la FDA ou si elle est menée dans une institution qui a volontairement accepté de se soumettre à la règle commune pour ses recherches.

Ainsi, dans une étude prospective, la collection de nouveau matériel biologique pour analyse génétique suppose une interaction avec les sujets participants. En revanche, utiliser du matériel biologique existant, qui ne nécessite pas une interaction directe, suppose quand même le recueil d'informations identifiantes pour corrélérer les résultats génétiques avec l'état de santé. Cependant, ce niveau d'information peut dans certains cas rester anonyme.

Le risque essentiel est, ici, celui de la divulgation d'informations de nature privée, directement reliée à la possibilité de remonter à la source d'informations individuelles à partir de l'échantillon. Le NBAC a recommandé que certaines recherches soient exemptées du recueil du consentement éclairé lorsqu'elles n'utilisent que des échantillons anonymes ou des échantillons non liés par une procédure lisible, ou rendus anonymes secondairement. Ceci concerne les données uniquement disponibles sous forme anonyme, les données rendues publiques, les données de sujets décédés (*stricto sensu*, la règle commune ne concerne que les sujets vivants, mais les investigateurs doivent planifier leurs études de manière à ne pas nuire aux familles des défunts).

Le rapport introduit les notions **de respect de la confidentialité**, ou contrôle exercé par un individu sur le partage avec autrui dans le temps d'informations physique, comportementale ou intellectuelle le concernant et de **risque minimal**, défini comme la probabilité que l'amplitude du dommage au cours de la recherche ne soit pas plus grande que celle rencontrée dans la vie courante ou lors d'examen de routine. Il fournit une liste de questions que le chercheur doit se poser pour argumenter son dossier et obtenir un certificat de confidentialité. Ces questions concernent notamment :

- la facilité pour identifier la source et obtenir l'information pour d'autres personnes que l'investigateur, ainsi que le préjudice potentiel lié à la divulgation de celle-ci ;
- la possible existence d'un préjudice du groupe auquel appartient la source, justifiant d'avoir consulté les représentants de la communauté et d'exposer les procédures envisagées envers le groupe et les participants ;
- le consentement au stockage et à des recherches futures, en anticipant diverses options allant du refus d'utilisation à l'acceptation de toute utilisation future, en passant par l'acceptation d'utilisation rendue anonyme et

non liée, l'acceptation d'utilisation codée liée, l'acceptation pour l'étude en cours avec nécessité de contact préalable pour des études ultérieures ;

- le retour d'information, en prenant en compte le droit d'accès des participants aux informations médicales les concernant, sans oublier, cependant, que la qualité des déterminations biologiques peut être remise en question dès lors que les laboratoires de recherche n'ont pas vocation à fournir de résultats aux patients et sont donc exemptés d'une certification par le CLIA (Clinical Laboratory Improvement Amendments) ;
- enfin, les aspects de commercialisation, dès lors que la constitution de banques de matériel biologique « annoté » représente une richesse susceptible de valorisation. Toute possibilité de gain commercial lors du consentement doit être évoquée ainsi que les modalités de partage des profits éventuels entre collaborateurs. Une telle attitude, encore rare dans notre pays, se retrouve cependant dans le formulaire de consentement soumis aux patients pris en charge dans les centres régionaux de lutte contre le cancer.

5 | Protection des personnes et partage des données

L'American College of Epidemiology a pris une position en mai 2002 (*Policy Statement on Sharing Data from Epidemiologic Studies*) soulignant que la facilité du transfert des données ne devait pas prévaloir sur le souci de respect de la vie privée et de la confidentialité de données individuelles mais indiquait que le *Freedom of Information Act* permettait de rendre accessible, à des fins de recherche, des données, avec un double avantage sur le plan cognitif et financier de « rentabilisation » en temps et investissement.

Parallèlement, le NIH (*Final NIH Statement on Sharing Research Data* : National Institute of Health ; 2003) développe une politique active de partage de données pour toute demande de financement supérieure à 500 000 dollars par an. Depuis octobre 2003, les demandeurs doivent s'engager, lors de la soumission d'un projet, à mettre à disposition les données rendues anonymes ou expliquer en quoi leurs données ne peuvent rentrer dans cette option, avec des clauses de disponibilité respectant la priorité de ceux qui ont fait l'effort de développer la base de données.

Il définit précisément :

- les données issues d'une recherche finalisée qui correspondent au matériel enregistré et rapports de fin d'études ;

- les données « désidentifiées », c'est-à-dire des données pour lesquelles l'identifiant individuel a été enlevé mais où le lien avec l'identifiant initial est maintenu par l'investigateur du projet ;
- les données rendues anonymes sans aucune identification individuelle possible. L'usage futur de la base de données est discuté avec l'IRB dont relève la recherche ainsi qu'avec les investigateurs initiaux ayant contribué à la collection de données.

Conclusion et recommandations

Le cadre législatif, concernant les recherches biomédicales au sens large, mis en place progressivement en France depuis 1978 témoigne de la volonté d'un regard sociétal sur les pratiques de recherche avec le souci constant de protéger la liberté, la vie privée et la santé des personnes qui se prêtent à ces recherches. Si ce cadre a certainement contribué à organiser la recherche clinique et améliorer sa qualité, l'évolution des pratiques de recherche biomédicales au cours des dix dernières années a souligné l'importance des approches de type épidémiologique et l'inadaptation de certaines dispositions. Les récentes modifications ont pris en compte de nombreux points évoqués par les chercheurs et tendent à rendre le dispositif français proche du système anglo-saxon, avec un niveau de protection des personnes similaire.

Il manque encore une organisation reposant sur un guichet-recherche unique entre les différents promoteurs publics, des rapports harmonieux avec les autorités administratives ainsi que l'acceptation que la recherche en santé ne soit pas entièrement réglementée par le corps médical, empêchant l'émergence d'une recherche autonome de qualité non seulement en épidémiologie mais aussi en soins infirmiers, en psychologie, en ergonomie, comme cela peut exister dans d'autres pays. En ce sens, les procédures standardisées et pragmatiques, accessibles en ligne, mises en place en Grande-Bretagne, au Canada et aux États-Unis avec le rôle centralisateur des IRB ou REC paraissent intéressantes.

La question de l'assurance spécifique pour des recherches épidémiologiques pose la question de la nature de ce qui est réellement assuré. Les textes actuels restent muets sur ce point. Il peut être regretté que les chercheurs n'aient jamais eu de retour sur les éventuelles plaintes et leur mode de résolution selon le type de recherche biomédicale.

La volonté de fondre dans un texte unique des dispositions concernant aussi bien la thérapie génique ou les thérapies innovantes que les recherches épidémiologiques en population impose de retrouver dans les modalités de mise en

œuvre et chez tous les acteurs concernés l'intelligence et la souplesse de vue qui ont conduit les législateurs à leurs formulations, tout en sachant rester à l'écoute des praticiens du terrain, et éviter des lenteurs supplémentaires d'un système multipliant les instances et les expertises mais reposant souvent encore sur le bénévolat.

CONCLUSION

À l'issue de la lecture des chapitres précédents, nous espérons que le lecteur extérieur à l'épidémiologie humaine aura pris connaissance de l'objet des différents champs de cette discipline, et de ses grands résultats passés (chapitre 1 « Définition, historique et champs de l'épidémiologie »), qu'il aura constaté que des méthodes spécifiques y sont activement développées en statistique, et en mathématiques (chapitre 4 « Épidémiologie et sciences mathématiques ») et qu'une collaboration plus étroite avec les mathématiciens serait bénéfique, pour les deux parties. Le rapport a montré que le développement de l'épidémiologie nécessite aussi une relation étroite avec la biologie (chapitre 5 « Épidémiologie et sciences biologiques : quatre exemples), et avec les sciences humaines et sociales (chapitre 6 « L'épidémiologie sociale »). Bien entendu, l'épidémiologie humaine doit aussi se développer en relation avec la médecine vétérinaire et l'épidémiologie animale. L'actualité le rappelle sans cesse, à propos des maladies émergentes, et ceci doit avoir une traduction immédiate dans le monde de la recherche et de l'enseignement supérieur.

Le rapport a aussi examiné l'organisation de la recherche et de l'enseignement supérieur de l'épidémiologie en France (sous-chapitres 2.1 « L'épidémiologie humaine » et 9.1 « La formation en épidémiologie à l'université »), et a souligné la faiblesse quantitative des effectifs de chercheurs, laboratoires et thèses, qui contraste avec une heureuse montée en puissance de l'épidémiologie exercée en dehors des structures de recherche, grâce à l'existence des agences sanitaires nouvellement créées. Le manque d'information fournie sur l'épidémiologie humaine au collège, à l'université dans les disciplines non médicales, et au cours de l'exercice professionnel de la médecine a été décrit (chapitre 9 « Les conditions nécessaires au développement de l'épidémiologie »).

Le rapport a aussi analysé les relations entre la recherche épidémiologique « académique » et les opérateurs du système de santé, dans deux directions particulières : la valorisation en termes de recherche et d'information des bases de données sanitaires développées par ces opérateurs (chapitre 3 « Sources de données pour l'épidémiologie ») qui a été jugée insuffisante, et le transfert difficile des résultats de la recherche épidémiologique vers les actions de santé publique correspondantes qui ne va pas de soi, comme on pourrait naïvement le croire (chapitre 8 « L'épidémiologie : la base scientifique de la santé publique »).

Dans ce chapitre, nous reprenons en assez grand détail les principaux constats du rapport ; nous décrivons des changements institutionnels décidés, ou en discussion dans le domaine de l'enseignement et de la recherche en santé

publique, qui peuvent avoir un fort impact sur l'avenir de l'épidémiologie ; nous présentons finalement des recommandations découlant de ce constat.

1 | Les constats

1.1 Une forte demande de la société en épidémiologie

Le rapport a rappelé la demande permanente de développement, ou d'amélioration, de l'épidémiologie que la société, *via* les médias, et les décideurs exprime constamment. Cette demande concerne la connaissance de l'état de santé, jugée imparfaite, et celle de l'impact sur la santé de facteurs de risque, notamment les facteurs environnementaux, d'expositions professionnelles ; l'épidémiologie est aussi perçue comme la science de référence permettant l'évaluation de l'efficacité des politiques de santé publique, des traitements, etc., même si le passage de la connaissance scientifique à la pratique en santé publique ne va pas de soi, ainsi qu'exposé dans le chapitre 8 « L'épidémiologie : la base scientifique de la santé publique ». On ne doit pas non plus oublier la demande d'épidémiologie à visée gestionnaire : ainsi, des systèmes d'information toujours plus performants se mettent en place dans tous les hôpitaux afin de documenter les coûts et dépenses de prise en charge des malades.

Les demandes que la société fait à l'épidémiologie sont quelquefois irréalistes et, hélas, en construisent une image négative injustifiée : lui demander d'apporter la preuve que telle exposition est sans risque pour la santé, par exemple, ce qui est impossible. Enfin, aussi, certaines demandes ne sont hélas pas faites à l'épidémiologie : par exemple, effectuer des études coût/bénéfice permettant de comparer les investissements de prévention effectués dans différents secteurs de la santé et permettre ainsi de s'interroger sur la rationalité de ces choix.

1.2 Un positionnement difficile de l'épidémiologie par rapport aux autres disciplines scientifiques

Chacun est prêt à reconnaître qu'il n'y a pas de recherche médicale possible sans approche populationnelle, donc sans épidémiologie. En même temps, cette universalité de l'épidémiologie fait que tout le monde la revendique : pour caricaturer, les succès de la recherche épidémiologique sont volontiers accaparés par les différents secteurs de la médecine, tandis que ses limitations et ses échecs sont laissés à la responsabilité des chercheurs épidémiologistes ! Un autre problème étrange que rencontre l'épidémiologie est qu'elle peut aboutir à

des résultats importants sans recours à des technologies avancées ou très spécifiques. Or, beaucoup de sciences sont identifiées par les technologies avancées qu'elles utilisent. C'est souvent la technologie utilisée qui fixe les frontières entre les scientifiques du domaine et ceux des autres domaines : il est difficile de faire de la physique des hautes énergies sans avoir accès à des accélérateurs de particules, de faire de la biologie cellulaire sans avoir accès à de la microscopie confocale laser, de faire de la physiologie sans avoir accès à une plate-forme d'imagerie du petit animal, etc. Le rapport a souligné cependant que l'épidémiologie ne reposait pas que sur l'usage de méthodes statistiques élémentaires : les épidémiologistes ont développé de nombreuses méthodes spécialisées, dont certaines ont diffusé dans d'autres domaines.

La perception de l'épidémiologie change aussi dans le monde scientifique à cause de la présence de plus en plus grande de nouvelles catégories de travaux épidémiologiques publiés dans la presse scientifique généraliste lue par les chercheurs de toutes les disciplines : discussion et présentation de systèmes d'observation ambitieux combinant génotypage, phénotypage et bio-informatique ; travaux probabilistes, statistiques et de mathématiques des systèmes complexes permettant de mieux représenter, comprendre et prédire la dynamique de maladies infectieuses existantes, émergentes, voire hypothétiques dans des variétés de situations. Cela fait écho aux grandes préoccupations de la santé publique de la fin du XX^e siècle et du XXI^e siècle, ravivées par la crainte du bioterrorisme.

1.3 Les moyens nécessaires à la recherche épidémiologique largement ignorés

Faute, toujours, d'avoir identifié des technologies très visibles en épidémiologie, dont le coût serait facilement identifiable, les scientifiques des autres disciplines ne se rendent pas toujours compte du coût de la recherche épidémiologique. Le rapport (notamment dans les sous-chapitres 2.1 « L'épidémiologie humaine » et 9.3 « moyens nécessaires à la recherche épidémiologique ») a rappelé la taille considérable des moyens nécessaires à la recherche épidémiologique, notamment en citant des grands projets européens ou nord-américains. Les moyens nécessaires très spécifiques de l'épidémiologie concernent une charge très importante en personnel technique de collecte, suivi et assurance qualité des données. De nombreux protocoles de recherche épidémiologiques durent des années, voire des dizaines d'années, parce que leur durée est fixée par l'histoire naturelle des maladies, et parce que le schéma de référence est l'étude prospective. De même, les systèmes d'observation destinés à documenter l'évolution diachronique des maladies en fonction, par exemple, des variations de l'environnement doivent être pérennes.

Le constat est qu'en l'état actuel, il y a de grandes difficultés à mettre en œuvre de tels projets pérennes dans le cadre des EPST ou des universités : la mise en place d'un grand projet épidémiologique sur 20 ans nécessitant entre 20 et 50 personnes (effectif type nécessaire au suivi d'une grande cohorte, ou d'un système d'observation sentinelle des maladies infectieuses) implique non seulement des décisions de financement pluriannuel (en contradiction avec les principes de contrat quadriennal), mais aussi, soit la possibilité de nombreux engagements sur emplois d'ingénieurs, techniciens, administratifs alors que les postes fixes sont devenus quasiment inexistantes, soit la possibilité de recourir à des contrats de droit public successifs (habituellement nommés CDD, à tort, puisque ne relevant pas de la définition donnée dans le droit du travail) ; mais ceux-ci sont insuffisamment payés et souffrent de nombreuses contraintes administratives.

Notons enfin le problème de l'évaluation scientifique : il faut être capable de décider la création d'un tel grand projet sur des périodes excédant les données habituelles, et disposant des modalités d'évaluation et de réorientation au cours de son existence. . .

Que face à de telles difficultés les équipes françaises aient réussi à monter des projets compétitifs sur le plan international doit être salué en proportion. L'association avec de grands systèmes extra-universitaires (mutuelles, collectivités, agences, par exemple), le soutien de grandes associations caritatives, l'insertion dans des projets européens ont été les conditions indispensables de ces succès, et la leçon doit en être gardée.

1.4 La recherche épidémiologique intégrée à la recherche biomédicale

La recherche épidémiologique nécessite une forte intégration avec la biologie, la recherche clinique, et la médecine vétérinaire quand il s'agit de zoonoses. Des exemples démonstratifs en ont été donnés dans le chapitre 5 « Épidémiologie et sciences biologiques : quatre exemples » dans le cas particulier des maladies neurodégénératives, du cancer, des maladies cardiovasculaires, des maladies infectieuses. L'interaction avec la biologie permet aux biologistes de tester « en grandeur nature » l'impact d'hypothèses physiopathologiques, et aux épidémiologistes d'importer des concepts et surtout d'intégrer des données issues de technologie. Ceci s'applique tout particulièrement aux données génomiques : ainsi, depuis plus de cinq ans, l'*American Journal of Epidemiology*, un des tout meilleurs journaux de l'épidémiologie, publie régulièrement des *HuGe (Human Genome) reviews*, sur les principaux problèmes de la recherche médicale, qui

montrent à la fois les réalisations et les besoins de recherche épidémiologique correspondants.

1.5 Le besoin en statisticiens, probabilistes, informaticiens, s'accroît

Les chapitres 3 « Sources de données pour l'épidémiologie » et 4 « Épidémiologie et sciences mathématiques » ont été entièrement consacrés à la description d'interactions entre les épidémiologistes, les probabilistes, les statisticiens, les informaticiens et les mathématiciens. Ces chapitres ont rappelé qu'il existait tout un champ de techniques statistiques spécifiques de l'épidémiologie qui s'étaient développées dans des laboratoires de biostatistique (implantés souvent, mais pas exclusivement, dans des facultés de médecine), et de statistique mathématique. De leur côté, les probabilistes ont une longue histoire de collaboration avec les épidémiologistes des maladies infectieuses et avec les épidémiologistes généticiens ; les informaticiens trouvent également en épidémiologie des problèmes intéressants à traiter en vue — par exemple — du développement de systèmes d'information hautement sécurisés en termes de confidentialité. . . Enfin, la nouvelle donne de l'épidémiologie se caractérise par une explosion du nombre des données et de leur nature (données génomiques, imagerie, etc.), ce qui implique une intensification de la collaboration des épidémiologistes avec les statisticiens et informaticiens spécialisés dans les méthodes innovantes d'extraction de l'information de ces grandes bases de données (méthodes de « fouilles de données » ou *data mining*, décrites au sous-chapitre 4.2). De même, la modélisation des maladies infectieuses fait de plus en plus émerger le besoin de méthodologies puissantes de simulation événementielle, s'appuyant sur des nouvelles catégories de données auxquelles on n'avait pas accès jusqu'ici (par exemple, données informatisées de déplacement individuel, etc.). Elle nécessite aussi des apports importants de la part des théoriciens de la dynamique des systèmes complexes.

1.6 La recherche épidémiologique nécessite aussi des liens étroits avec les sciences humaines et sociales

Le chapitre 6 « L'épidémiologie sociale » de ce rapport a rappelé qu'on ne peut évidemment avoir l'ambition d'étudier les « déterminants de l'état de santé » sans mesurer, apprécier les dimensions sociales, et donc sans faire appel aux sciences humaines et sociales, notamment la démographie (avec laquelle l'épidémiologie a des liens naturels méthodologiques, *via* la statistique), la sociologie, la psychologie clinique. . . Un problème massif d'organisation de la

recherche se pose alors : alors que les différentes disciplines des sciences humaines et sociales sont indispensables à une bonne épidémiologie du cancer, du vieillissement, des maladies cardiovasculaires, etc., on constate que les effectifs des chercheurs à la fois compétents et prêts à faire de l'épidémiologie sociale sont minuscules, y compris en considérant l'ensemble des EPST, dont le CNRS qui est le principal à cet égard. Notons enfin que les collaborations sont parfois difficiles car beaucoup de chercheurs des sciences humaines ont des systèmes d'évaluation différents des chercheurs des disciplines scientifiques (pour en citer un seul : la publication des résultats dans des revues internationales compétitives n'est pas toujours vue comme une priorité dans ces disciplines). Enfin, plus encore que dans tous les autres domaines de l'épidémiologie, se pose la question de la causalité des relations trouvées entre maladies et facteurs sociaux, qui a fait l'objet d'un chapitre séparé de ce rapport (chapitre 7 « Épidémiologie et causalité »).

1.7 Le déficit d'interaction entre « statistiques sanitaires » et recherche épidémiologique

Le rapport a souligné (chapitre 3 « Sources de données pour l'épidémiologie ») que la France disposait de systèmes d'informations extrêmement puissants dans le domaine de la santé, notamment dans les organismes de protection sociale et dans les hôpitaux, et que d'autres, très ambitieux — tel le dossier médical partagé — étaient sur le point d'apparaître. Ces systèmes, placés sous la responsabilité d'organismes ayant en charge des missions de santé publique ou de gestion hospitalière, n'avaient bien sûr pas vocation à être développés avec une finalité de recherche. Pourtant, à l'évidence, ils contiennent des informations d'un immense intérêt pour la recherche épidémiologique, tels qu'ils sont et encore plus tels qu'ils pourraient être, au prix d'ajustements dans la nature et l'organisation du recueil des données. On a notamment rappelé l'hypothèse selon laquelle plus la valorisation des données de statistique sanitaire en termes de recherche sera grande, meilleurs seront les recueils de ces données. Les problèmes de confidentialité et de respect des données médicales — souvent cités comme limitation à la valorisation de ces systèmes d'information dans un but de recherche — sont souvent évoqués. Il a été rappelé (sous-chapitre 9.4 « Études épidémiologiques en santé : aspects législatifs et réglementaires ») qu'il existait bien sûr des réglementations précises garantissant les droits des patients, mais que celles-ci étaient réalistes et bien moins contraignantes que souvent affirmé.

Le constat est donc que les potentialités en termes de résultats épidémiologiques de ces systèmes puissants (et faisant l'objet de budget sans commune mesure avec ceux de la recherche académique épidémiologique) n'ont pas été utilisées de façon optimale, faute de liens assez forts entre les organismes

détenteurs de ces données et les chercheurs, et sans doute encore plus faute de la reconnaissance, par les uns et les autres, de la légitimité et de l'importance de ce type de démarche. Il a été souligné que les années récentes ont marqué une évolution positive très notable, avec par exemple des expériences réussies s'appuyant sur des données de sécurité sociale¹ ou hospitalières (PMSI) qui démontrent donc la possibilité de la tendance souhaitée. Enfin, l'exploitation de ces très grandes bases de données peut attirer l'intérêt de la communauté des statisticiens, notamment dans le domaine de la « fouille de données » (sous-chapitre 4.2) qui est actuellement l'objet de recherches actives et s'applique avec succès dans de nombreux autres domaines.

1.8 Le sous-dimensionnement de la recherche et de l'enseignement épidémiologique en France

Il a été démontré dans le sous-chapitre 2.1 « L'épidémiologie humaine » que les effectifs d'épidémiologistes français étaient faibles comparativement aux autres pays développés, aussi bien européens que nord-américains, qui ont été étudiés dans le rapport. Il est clair, d'après l'examen de la littérature, que les équipes françaises d'épidémiologie sont tout à fait compétitives par rapport aux autres pays dans chacun de leurs propres domaines d'intérêts. Cependant leur nombre relatif plus faible entraîne que des domaines entiers de la recherche médicale ne sont couverts que très faiblement (un exemple flagrant en est la psychiatrie) et que le nombre moyen de publications scientifiques par million d'habitants est notablement plus faible en France que dans certains pays à développement semblable qui lui ont été comparés.

La recherche épidémiologique française est développée presque uniquement au sein d'un seul EPST (l'Inserm) qui n'a pas augmenté les effectifs d'épidémiologistes au cours des dernières années au-delà du développement homothétique de celui des autres disciplines qu'il héberge (alors que celles-ci ont d'autres bassins institutionnels d'expansion). On note quelques équipes à l'IRD (effectifs en croissance), aucune équipe au CNRS, et de minuscules effectifs de chercheurs dans les autres EPST. Les enseignants chercheurs épidémiologistes sont presque uniquement hospitalo-universitaires : leur nombre reste donc lui aussi à peu près constant, et leur implantation suit nécessairement une logique dans laquelle les besoins hospitaliers de terrain sont autant — voire plus — considérés que les besoins de renforcement des équipes de recherche performantes. L'éparpillement est donc la règle dans les universités. Parallèlement, la création des agences

¹Voir par exemple un numéro thématique du *Bulletin Épidémiologique Hebdomadaire* (22/03/2005) sur les complications du diabète, qui s'est appuyé sur un échantillon de 10 000 bénéficiaires du régime général des travailleurs salariés remboursés pour un traitement du diabète, ou l'étude *a posteriori* des consommations de soins des personnes décédées lors de la canicule 2003 (in Valleron et Boumendil, *CR Biol*, 2004).

(voir ci-dessous) a mené à des recrutements massifs d'épidémiologistes dans le monde extra-universitaire (environ 100 recrutements en CDI sur les années 2003-2004, et le double en CDD au seul Institut de veille sanitaire).

En conclusion, sous-effectif de chercheurs et enseignants chercheurs en France, presque uniquement recrutés dans des disciplines médicales et organismes de recherche médicale, sans renversement de tendance prévisible. Au contraire, dynamisme de la composante « études et actions » de l'épidémiologie en réponse aux nouvelles lois de santé publique.

Un autre constat (sous-chapitre 9.1 « La formation en épidémiologie à l'université ») a été celui de la faiblesse de l'enseignement de l'épidémiologie à l'université : un étudiant de LMD de mathématiques ou de biologie n'a à peu près aucune chance d'entendre parler d'épidémiologie, même lorsqu'il étudie dans les disciplines scientifiques les plus proches (statistique, écologie, dynamique et biologie des populations. . .). L'épidémiologie est presque uniquement enseignée dans les UFR médicales par des enseignants des disciplines médicales, et l'est très insuffisamment. La formation à et par la recherche est quantitativement très faible avec seulement une quarantaine de thèses par an, dont très peu sont soutenues par des candidats ayant un cursus non médical. Ceci est une situation spécifique de l'épidémiologie française.

1.9 Faiblesse de la culture épidémiologique du public et des professionnels de santé

Le public n'a pas appris à l'école ce qu'était un risque, ou un facteur de risque, comment on évaluait l'efficacité d'une prévention, comment on déclarait qu'un traitement était plus efficace qu'un autre. L'enseignement de l'épidémiologie (ses méthodes, et ses grands résultats dans les pathologies les plus importantes) est absent de l'école, du collège et du lycée. Cela n'est guère étonnant puisque ses professeurs (en particulier les professeurs de Sciences de la Vie et de la Terre) n'ont jamais entendu parler d'épidémiologie à l'université (*cf.* plus haut).

La situation n'est pas meilleure du côté des professionnels de santé : alors que les différentes réformes du système de santé s'appuient presque toujours sur un rationnel épidémiologique, alors que la « médecine fondée sur la preuve » devient leur paradigme, les médecins et les professionnels de santé n'ont reçu à ce sujet dans leur faculté de médecine qu'une formation quantitativement faible, et mal placée dans leur cursus ; ils ne bénéficient pas de formation continue, et n'ont pas accès à des systèmes d'information gratuits sur le Web dont disposent

leurs collègues anglais ou hollandais, par exemple (sous-chapitre 8.1 « L'épidémiologie clinique, base scientifique pour la prise en charge rationnelle des malades »).

2 | Les changements institutionnels récents en santé publique à prendre en compte

On rappellera ci-après l'énorme impact sur l'épidémiologie qu'ont eu et/ou que vont avoir différents dispositifs législatifs récents relatifs à la santé publique : impact parce qu'ils placent la démarche épidémiologique à la base de la rationalisation du système de santé, impact en termes de moyens parce qu'ils sont à l'origine de nouvelles structures (notamment agences) dotées de moyens importants, impact en termes institutionnels car ils mettent en place de nouvelles organisations d'enseignements. La réflexion sur l'avenir de l'épidémiologie française et l'établissement de recommandations doivent nécessairement prendre en compte ces changements.

2.1 En préalable : distinguer santé publique et épidémiologie

Si les changements institutionnels relatifs à la santé publique décrits ci-après doivent être pris en compte pour imaginer l'avenir de l'épidémiologie, il n'en faut pas pour autant oublier la distinction entre santé publique et épidémiologie : d'une part, la santé publique est certes une discipline universitaire (correspondant à la 46^e section du CNU. . .), mais elle est avant tout une part de la « chose publique », avec ses priorités, ses administrations, ses valeurs. . . D'autre part, même en ne considérant que l'aspect universitaire, la santé publique ne se limite pas, loin s'en faut, à l'épidémiologie. Ceci est clair en examinant les disciplines présentes soit dans la Commission scientifique spécialisée 3 de l'Inserm (où sont jugés les épidémiologistes, mais qui recouvre l'ensemble de la santé publique), soit dans la 46^e section du CNU qui regroupe les enseignants de santé publique : on y trouve des informaticiens médicaux, des biostatisticiens, des économistes de la santé, des sociologues de la santé, des spécialistes de médecine légale, de médecine du travail, etc. Dans la 46^e section, les épidémiologistes et apparentés comptent seulement pour environ un quart de l'effectif total.

Enfin, l'épidémiologie n'est pas seulement une science dont le but serait d'être utile à la santé publique ; les chercheurs poursuivent aussi des programmes de

recherche non pilotés par la demande sociale exprimée, par exemple pour identifier des causes ou des réseaux causaux d'une maladie (chapitre 7 « Épidémiologie et causalité »), ou pour trouver les règles de propagation d'un agent infectieux dans des communautés en fonction de leurs structurations géographiques ou sociales. Il ne faut pas imaginer, ni souhaiter, que le développement de l'épidémiologie ne se fasse que « dans la roue » des priorités de la santé publique et de sa réorganisation.

Cette remarque liminaire est destinée à bien faire apprécier que l'organisation de la recherche et de l'enseignement en épidémiologie a une spécificité, et que (ré)organiser la santé publique ne signifie pas *ipso facto* (ré)organiser l'épidémiologie. En particulier, les dispositifs d'évaluation de la recherche et de l'enseignement en épidémiologie doivent être distincts de ceux relatifs à la santé publique, et ne pas être dissous dedans.

2.2 Les agences sanitaires issues de la loi de 1998

La loi de Sécurité sanitaire de 1998 mena à la création d'agences (InVS, Afssa, Afssaps, EFS, Établissement français des greffes) ayant des missions d'information, d'évaluation, et/ou de réglementation. L'exercice de leur mission repose largement sur l'épidémiologie, soit parce qu'elles ont à utiliser et expertiser des travaux épidémiologiques extérieurs, soit parce qu'elles sont elles-même opérateurs (c'est tout particulièrement le cas de l'Institut national de veille sanitaire). Il a déjà été signalé que le financement et les moyens en personnels de ces agences étaient très importants par rapport à ceux de la recherche académique, à l'intérieur des EPST et des universités.

Signalons enfin dans la même rubrique, bien qu'issue d'une loi antérieure (1996), la Haute autorité de santé (qui a pris la suite de l'Anaes) dont les missions concernent notamment l'accréditation des établissements de soins, avec par conséquent une grande importance dans le système de santé. La méthodologie de base de la Haute autorité est l'épidémiologie (pour remplir ses missions d'évaluation), et son paradigme est la « médecine fondée sur la preuve » (sous-chapitre 8.4 « Force de conviction de la modélisation épidémiologique au service de l'action publique face au risque »).

En plus de leur interaction avec la recherche, déjà rappelée, tous ces dispositifs sont à l'origine de débouchés pour les jeunes épidémiologistes, qui renforcent la nécessité d'amélioration (quantitative) de la formation dans ces agences. Cependant, le problème actuel est de faire face, avec le peu de personnes très bien formées, à l'afflux de recrutements; le problème futur sera

démographique, puisque tous ces recrutements ont lieu dans la tranche d'âge 30-40 ans.

2.3 L'École des hautes études en santé publique (EHESP)

Cette école a été créée par la loi de Santé publique d'août 2004 (chapitre 1^{er} du Titre V de cette loi). C'est un établissement public de l'État à caractère scientifique, culturel et professionnel, placé sous la tutelle des ministres chargés de la Santé, des Affaires sociales, de l'Éducation et de la Recherche. Son fonctionnement étant réglé par l'article L.717-1 du Code de l'éducation, il s'agit d'un « grand établissement » qui a par conséquent la capacité de recruter des enseignants chercheurs. La loi lui fixe plusieurs missions : la première reprend essentiellement les missions de l'actuelle École nationale de santé publique qui lui sera intégrée (formation des personnes ayant à exercer des fonctions de direction, de gestion, d'inspection ou de contrôle dans les domaines sanitaires, sociaux ou médico-sociaux) ; cette mission ne concerne donc l'épidémiologie que de façon assez marginale. La seconde mission, en revanche, est d'assurer un enseignement supérieur en matière de santé publique : « à cette fin, elle anime un réseau national favorisant la mise en commun des ressources et des activités des différents organismes publics et privés compétents ». La troisième mission est de « contribuer aux activités de recherche en santé publique » et la quatrième est de « développer des relations internationales dans les domaines, notamment par des échanges avec les établissements dispensant des enseignements comparables ». Ainsi, l'EHESP peut créer une nouvelle donne pour le développement de l'enseignement et de la recherche en épidémiologie, y compris en permettant le recrutement (ou l'association) de nouveaux enseignants, lesquels pourraient fort bien également relever des disciplines scientifiques, dont le besoin est le plus vif pour le développement de l'épidémiologie : mathématiques, statistiques, probabilités, informatiques, sciences humaines et sociales.

2.4 Le projet d'Institut virtuel de recherche en santé publique

Par ailleurs, à l'initiative de l'Inserm, une réflexion avancée a lieu depuis 2004 sur la création d'un « Institut virtuel de recherche en santé publique ». Dans l'état actuel de ces réflexions celui-ci² aurait pour premier objectif de développer des programmes de recherche par la coopération, autour d'une stratégie scientifique concertée, des organismes nationaux ayant des missions dans

²Ce paragraphe de description de l'IVRSP a été écrit à partir des indications fournies par A. Alperovitch, coordinatrice du groupe de réflexion sur l'IRVSP, que nous remercions.

le domaine de la veille et de la sécurité sanitaire, de la gestion du système de santé ou de la définition des politiques de santé. Chaque partenaire (acteurs de la recherche, agences, caisses d'assurance maladie, directions de ministères) serait représenté dans le comité d'orientation stratégique qui fixe les priorités de l'IVRSP. Le conseil scientifique prévu — cinq personnalités scientifiques françaises et cinq étrangères — définirait les questions scientifiques correspondant à ces priorités, et mettrait en œuvre et évaluerait les programmes de recherche permettant d'y répondre. Les autres objectifs affichés du projet d'IVRSP sont la structuration du milieu de recherche en pôles régionaux, la mise en place d'une organisation permettant l'utilisation optimale des moyens — notamment des grandes bases de données — nécessaires à la recherche et la valorisation des résultats de la recherche en santé publique.

3 | Orientations prioritaires du rapport (OPR)

À partir des constats précédents, trois orientations principales ont été identifiées au fil des différents chapitres du rapport : développer l'épidémiologie observationnelle à grande échelle ; améliorer la culture en épidémiologie ; renforcer le potentiel de recherche du pays en épidémiologie. Ce sont ces orientations prioritaires qui guident les principales propositions de l'Académie des sciences.

OPR-I – Développer l'épidémiologie observationnelle à grande échelle

Le passé montre qu'un grand nombre de facteurs causaux et de facteurs de risques des maladies chroniques les plus importantes ont été identifiés grâce à l'épidémiologie observationnelle. À l'initiative de l'Inserm, une action a été menée pour identifier, en s'appuyant sur un large comité scientifique, des cohortes existantes ou en projet justifiant d'un soutien prioritaire. Cependant, les autres composantes de l'épidémiologie observationnelle qui ont démontré leur productivité en termes de découvertes dans le passé ne font l'objet d'aucune réflexion concertée entre EPST, universités et organismes ayant des missions de santé publique s'appuyant sur, ou utilisant, l'épidémiologie.

Le développement de l'épidémiologie observationnelle, dans laquelle sont inclus l'épidémiologie descriptive et les statistiques sanitaires, nécessite une conjugaison des efforts dans quatre directions :

OPR-I.1 — Créer les conditions de fonctionnement des systèmes épidémiologiques d'observation nécessitant une longue durée

Ces « systèmes » de « longue durée » sont les cohortes, ou les systèmes de surveillance épidémiologique et les registres à finalité de recherche dans les domaines environnementaux et infectieux, menant au recueil de longues séries chronologiques. Par « longue durée » on entend au moins 10 ans, en général 20 ans ou plus. Ces projets, à l'image de ceux qui existent déjà, s'appuient nécessairement sur un système de santé ou une collectivité identifiée, et associent souvent plusieurs organismes dont des agences ou autres opérateurs du système de santé. La dimension européenne est évidemment à privilégier, quand elle est pertinente.

De tels « grands » systèmes d'observation épidémiologiques doivent :

- être conçus comme des plates-formes réellement ouvertes à des équipes différentes de chercheurs ;
- être développés en relation très étroite avec les chercheurs biomédicaux ;
- être promus auprès des mathématiciens, statisticiens et probabilistes.

La difficulté à vaincre est institutionnelle : il faut que les structures support de ces « grands systèmes d'observations » puissent être pérennes et permettent donc soit d'engager sur des postes publics des nombres conséquents de collaborateurs, soit de gérer le renouvellement continu des personnels contractuels, sans perte de continuité ou de qualité, et dans un cadre de travail valorisant. Ceci pose des problèmes de gestion administrative et scientifique qui dépassent les possibilités actuelles des équipes de recherche individuelles, y compris les plus importantes, oublierait-on même que leur horizon temporel est limité à quatre ans, voire huit à l'Inserm.

La nécessité de disposer de ressources et de personnels pérennes pour ces projets observationnels à long terme mène à recommander la création d'un petit nombre d'« épidémiopoles » dans lesquels plusieurs systèmes épidémiologiques longue durée pourraient être gérés à la fois et dont le statut permettrait de garantir la pérennité des projets hébergés.

OPR-I.2 — Créer au moins un observatoire épidémiologique de très grande taille

Cet observatoire, lui aussi pluridisciplinaire, se distingue des systèmes précédents par le fait qu'il ne serait pas construit spécifiquement en fonction d'une

hypothèse de recherche précise (comme dans les systèmes d'observation longue durée précédents) mais pour apporter des observations représentatives et de qualité sur l'ensemble d'une très grande population. Des exemples de tels observatoires existent déjà à l'étranger et ont été décrits dans le sous-chapitre 9.3 « Les moyens nécessaires à la recherche épidémiologique ». L'articulation d'un tel projet avec l'« échantillon démographique permanent » (EDP) conduit par l'Insee et dont il est envisagé qu'il comporte 2,5 millions de personnes est certainement une voie à explorer. L'Insee dispose d'ailleurs de plusieurs autres observatoires intéressant l'épidémiologie, et il est souhaitable que cet organisme, compte tenu de sa compétence, assure une part beaucoup plus forte de son activité dans le domaine de la santé.

La création d'un observatoire épidémiologique de grande taille, la coordination avec les systèmes existants à l'Insee, la nature des collaborations à établir entre les différents organismes de recherche ou de santé publique impliqués devraient faire l'objet d'une attention particulière du Conseil national de l'information statistique (Cnis), dont l'implication dans le domaine de la santé devrait être renforcée.

OPR-1.3 – Mieux rentabiliser les données existantes par leur mise en commun et par la réalisation de méta-analyses puissantes

On peut constater que, bien qu'il y ait beaucoup d'équipes en France qui pratiquent des méta-analyses et/ou ont les compétences méthodologiques nécessaires, il n'existe pas dans le pays de centre d'études et de recherche important dans ce domaine, à l'image de celui de Richard Peto à Oxford, par exemple.

Une action volontariste, par exemple à l'initiative de la Haute autorité de santé, visant à la création d'un centre puissant dans le domaine, incluant une équipe de recherche ayant vocation à intégrer des statisticiens, est nécessaire.

OPR-1.4 – Mieux rentabiliser les bases de données déjà disponibles, construites dans un autre but que celui de la recherche et de l'observation épidémiologiques

Le mouvement naissant de collaboration entre les organismes ayant pour mission la gestion du système de santé, en population ou à l'hôpital, et les chercheurs doit être salué et encouragé. Ceci devrait non seulement concerner les épidémiologistes présentant des demandes d'accès à certaines informations aux

grands opérateurs de données pour tester leurs hypothèses, mais aussi les spécialistes de la fouille de données (*data mining*) qui y trouveraient un champ d'applications d'intérêt.

Les grands opérateurs de bases de données médicales pourraient organiser les règles et conditions de collaboration avec les équipes de recherche épidémiologiques et statistiques, là aussi en s'appuyant sur l'organisation existante du Conseil national de l'information statistique. La montée en charge du Système national d'information inter-régimes de l'assurance maladie (SNIIR-AM), qui doit à terme rassembler toutes les données des différents régimes d'assurance maladie, ainsi que la base de données du PMSI, peut être une excellente opportunité, à condition que soit mise en place une concertation avec le milieu de la recherche épidémiologique.

OPR-2 – Améliorer la culture en épidémiologie dans la population à l'université, chez les professionnels de la santé

L'information du public en épidémiologie humaine est mauvaise, avec de graves conséquences dans l'expression de demandes sociales injustifiées ou erronées. Malheureusement, ceci est l'aboutissement d'une longue chaîne de défauts de transmission des connaissances en épidémiologie, qui commence à l'école, continue à l'université, et se poursuit durant la vie active, avec l'absence de toute formation continue des personnels de santé, alors même que les réformes qui leur sont de plus en plus imposées reposent largement sur des considérations épidémiologiques. Pour réformer cette situation, il faut améliorer la formation et l'information relatives à l'épidémiologie à l'université, puis en cascade à l'école et dans le public.

OPR-2.1 – Informer les scientifiques

La recherche épidémiologique est mal connue en dehors des disciplines médicales ; avant de demander aux enseignants de biologie, de statistique, ou de mathématiques d'ouvrir leurs étudiants à l'épidémiologie, il est indispensable que les enseignants de ces différentes disciplines bénéficient eux-mêmes d'une réelle information sur le champ et les méthodes de l'épidémiologie. On pourrait aussi promouvoir l'intérêt des mathématiciens, statisticiens et informaticiens pour l'épidémiologie, grâce :

- à des appels d'offres ciblés de la nouvelle Agence nationale de la recherche (ANR) ;

- à l'organisation de séminaires, séminaires communs ou écoles d'été, entre épidémiologistes, biologistes et statisticiens, qui pourraient être organisés par les sociétés savantes ou associations de ces disciplines (par exemple, Adelf pour les épidémiologistes, Cimpa pour les mathématiciens, SFDS pour les statisticiens).

OPR-2.2 – L'enseignement en LMD

Le système des crédits européens (*European Credit Transfer System : ECTS*) qui va permettre aux étudiants français — enfin à leur tour — de composer leur cursus « à la carte » devra être utilisé pour permettre à tout étudiant de licence ou de M1 en biologie ou mathématiques d'avoir une vision minimale de ce qu'est l'épidémiologie, en insistant par exemple — pour les maladies chroniques — sur les mesures du risque, et pour les maladies infectieuses sur les grands résultats théoriques (théorème du seuil ...). Bien entendu, ces ECTS pourraient prendre des formes très variées selon les cursus : les biologistes des populations intéressés naturellement par les compétitions proies-prédateurs, n'auront pas de difficulté pour approfondir les relations hôtes-parasites. Les mathématiciens centreront de leur côté leur enseignement par exemple sur les systèmes dynamiques, les probabilistes sur les applications des processus de branchement.

Des secondes années de master (M2) bidisciplinaires, notamment statistique-épidémiologie, probabilités/mathématiques-épidémiologie des maladies infectieuses doivent être encouragées.

Toutes ces expériences d'enseignement devraient pouvoir prendre appui sur la future EHESP et des enseignants des UFR scientifiques — notamment statisticiens, probabilistes, informaticiens et mathématiciens désireux de s'engager dans l'interface avec l'épidémiologie — devraient pouvoir utiliser cet appui, par le biais de conventions ou de recrutements. Un fonctionnement en réseau permettrait de faire émerger une véritable culture de l'épidémiologie dans ces disciplines, et en retour chez leurs étudiants.

OPR-2.3 – L'enseignement dans les facultés de médecine

L'épidémiologie humaine n'est enseignée à l'université qu'aux futurs médecins ; l'épidémiologie animale ne leur est quasiment pas enseignée. L'enseignement est loin d'être exempt de critiques (sous-chapitre 9.1 « La formation en épidémiologie à l'université »). Les enseignants d'épidémiologie devraient donc établir un texte détaillé à l'intention de la Conférence des doyens des facultés de médecine pour redresser cette situation. Dans l'immédiat, la place de l'épidémiologie devrait être renforcée dans l'examen national classant.

OPR-2.4 – L'enseignement au collège

Des pistes prometteuses existent (sous-chapitre 9.4 « Études épidémiologiques en santé : aspects législatifs et réglementaires »). En particulier, il semble possible d'insérer une formation aux grands résultats de l'épidémiologie, notamment ceux qui ont un impact direct sur la santé des enfants et de leur famille, et à travers ces résultats d'initier au mode de raisonnement épidémiologique (notion de risque, méthode expérimentale pour juger de l'efficacité d'un traitement, différence entre corrélation et cause, etc.) dans le cadre des horaires d'éducation à la santé déjà prévus aux programmes (30 à 40 heures à deux niveaux du collège), horaires qui sont actuellement très peu utilisés. Les intervenants naturels seraient les professeurs de SVT dans les collèges et lycées d'enseignement général, de VSP (Vie sociale et professionnelle) et d'HPS (Hygiène, prévention, secourisme), et de STMS (Sciences techniques médicosociales) dans les lycées technologiques.

Notons aussi que les programmes actuels de biologie incorporent déjà des sujets (impact de la pollution sur la fonction respiratoire, nutrition, etc.) dans lesquels la composante épidémiologique pourrait être renforcée, aussi bien en ce qui concerne l'exposé des grands résultats que celui des principes des méthodologies qui y mènent.

Tout ceci nécessite évidemment un fort investissement de collaboration entre les épidémiologistes et la communauté des professeurs de SVT (formation à l'épidémiologie au moment du Capes, de l'agrégation, de la seconde année d'IUFM, pour l'instant totalement inexistante). Il est indispensable que les enseignants et chercheurs d'épidémiologie se mobilisent sur cet aspect essentiel de la formation des plus jeunes. Cette mobilisation pourrait notamment s'appuyer sur l'Association des épidémiologistes de langue française (Adelf), si celle-ci en était d'accord.

OPR-2.5 – L'information des professionnels de santé

Le paradigme de la « médecine fondée sur la preuve » qui s'impose dans l'ensemble des systèmes de santé, à la fois dans le but d'améliorer la qualité des soins et dans celui de rationaliser les dépenses de santé, ne peut pas être seulement proclamé par les grandes agences sanitaires et par les universitaires. La médecine fondée sur la preuve, qui n'est rien d'autre que le passage dans la pratique des connaissances épidémiologiques acquises dans les essais thérapeutiques, les études de cohortes, etc., doit être approprié par les professionnels de santé qui sont en première ligne pour en faire usage. Une formation permanente de haute qualité est nécessaire, d'autant plus que la formation initiale

dans les facultés de médecine est — on l'a vu — faible à cet égard, sans parler des écoles d'infirmières ou des personnels paramédicaux où elle est absente.

Il est nécessaire qu'à l'image d'autres pays (notamment Grande-Bretagne, Pays-Bas, . . .) les professionnels de santé aient un accès facile et gratuit à cette connaissance, à tout le moins aux produits de la Cochrane collaboration. Une grande difficulté (surmontée par les Hollandais. . .) est que ces produits sont en langue anglaise. La mise en place d'un système actualisé équivalent en langue française (pas nécessairement la simple traduction de la Cochrane collaboration) doit être envisagé, peut-être dans le cadre de la francophonie, mais son coût risque d'être considérable.

OPR-2.6 – L'information du public

Un média simple, à destination des journalistes et du grand public, présentant les résultats récents de l'épidémiologie, éventuellement de façon critique, est souhaitable. Une expérience menée dans le domaine de la démographie pourrait servir d'exemple³ : *Population et sociétés* est un bulletin mensuel de quatre pages, publié par l'Ined, et dont le succès après 30 ans d'existence ne se dément pas. Les articles sont repris presque systématiquement dans la presse, parfois en une, voire en ouverture de journaux télévisés, et servent aussi largement de matériaux pédagogiques pour les manuels scolaires (en sciences économiques et sociales, histoire. . .). Son secret est de proposer des textes lisibles par tout journaliste ou responsable politique, en traitant chaque fois d'un seul sujet : ce sont parfois des résultats originaux, mais aussi des synthèses, des ouvertures sur des thématiques d'apparence marginale (comme la diffusion des diverses pièces en euros émises par les membres de l'Union européenne à travers les pays membres, comme indicateur de la mobilité des personnes, ou l'évolution de la distribution par âge du corps électoral). . . et tout ceci a énormément contribué à faire considérer la démographie comme une discipline « utile », tant par les décideurs que par l'opinion. Une telle initiative en épidémiologie serait à coup sûr très utile, mais elle représente malheureusement une organisation lourde, l'identification d'un organisme prêt à cette initiative, d'un rédacteur en chef, etc., ce qui mène à recommander d'abord l'étude sérieuse de la possibilité de sa mise en place.

OPR-3 – Renforcer le potentiel de recherche épidémiologique

Le constat a été fait d'un sous-développement quantitatif de la recherche et de l'enseignement épidémiologique, du besoin d'intégrer au sein même des

³Le paragraphe qui suit provient d'un texte rédigé par Henri Léridon, que nous remercions.

équipes d'épidémiologie plus de compétences mathématiques (au sens large), biologiques et en sciences humaines et sociales. Les gisements des données, déjà existants et/ou en projet, doivent être mieux exploités grâce à une meilleure collaboration entre recherche et organismes de santé publique en charge de ces données.

À côté des quelques équipes et IFR d'épidémiologie de taille critique qui existent déjà en France, deux nouvelles pistes semblent prometteuses : la création d'équipes multi-organismes avec synergie de moyens et de compétences, et le travail en réseau. Ces deux pistes s'intègrent bien dans les changements institutionnels en cours.

OPR-3.1 – Créer des emplois de chercheurs et enseignants

Le constat a été fait d'une stagnation des effectifs de chercheurs temps plein en épidémiologie humaine, lesquels relèvent presque exclusivement d'un seul EPST (l'Inserm). De plus, leur démographie fait craindre à court terme une baisse d'effectifs, sauf mesures volontaristes au bénéfice spécifique de l'épidémiologie. De même, le renforcement des effectifs universitaires des disciplines concernées par l'épidémiologie est indispensable : au premier rang, ceci concerne dans les disciplines médicales, les biostatisticiens, informaticiens, médicaux, épidémiologistes (relevant de la 46^e section du CNU), mais aussi les universitaires des disciplines cliniques ayant une fonction et activité épidémiologique. Ce renforcement nécessaire en personnel universitaire concerne aussi les disciplines non médicales, notamment les mathématiciens, biologistes et spécialistes des sciences humaines et sociales qui devraient pouvoir être recrutés sur des programmes de recherche et d'enseignement de l'épidémiologie.

OPR-3.2 – Créer des équipes multi-organismes, centrées sur un même domaine d'application de l'épidémiologie milieux

L'intérêt et la faisabilité potentielle rapide⁴ de cette approche ont été récemment illustrés dans un cas particulier, avec le projet, en cours de discussion, de regroupement d'équipes appartenant à trois organismes différents (Inserm, InVS, INRS), afin de réunir des ressources humaines et matérielles constituant une masse critique importante, dotée d'une forte visibilité, et pouvant ainsi jouer le rôle d'un « *Centre de référence en épidémiologie des risques professionnels* ». Ce système devrait être complété par la mise en place d'un « *Réseau* ».

⁴Cet exemple de faisabilité s'appuie sur des documents fournis par M. Goldberg, U 687, qui a piloté le projet décrit. Nous l'en remercions.

d'équipes associées » : équipes d'épidémiologie consacrées à l'étude d'une pathologie (cancer, affections neurologiques, etc.) ou d'une population particulière (femmes, enfants, etc.) et qui sont amenées dans certains cas à s'intéresser à des facteurs professionnels, ou structures du domaine de la santé au travail où l'épidémiologie n'est qu'une activité minoritaire. L'ensemble, ainsi formé du réseau animé et du centre de référence, permettrait d'envisager des activités de réelle ampleur, à visée nationale ou internationale. Des exemples équivalents devraient pouvoir être trouvés et mis en application dans de nombreux domaines, notamment de l'environnement, pour lesquels les demandes sont nombreuses et l'offre faible.

OPR-3.3 – Créer des équipes de recherche en réseau

Comment à la fois créer de nouvelles équipes de taille critique et répondre à la nécessité de disposer localement, dans tout ensemble conséquent de recherche médicale (par exemple dans tout CHU), d'une équipe référente de recherche épidémiologique, ce qui conduit à un éparpillement de compétences sur tout le territoire ? On doit aussi se demander comment créer des équipes incluant les compétences nécessaires, mais localisées de façon dispersée (mathématiciens, statisticiens, ... des UFR scientifiques par exemple). Enfin, on doit éviter une recherche à deux vitesses : l'une dans des grandes équipes fonctionnelles d'épidémiologie, et l'autre dans des très petits groupes isolés. Ces questions mènent forcément à l'idée de réseau : le réseau permet en principe de cumuler les avantages du travail local et du travail collaboratif. Cependant, tandis qu'on a l'expérience de ce que signifie un groupe de 150 chercheurs regroupés dans un même bâtiment, on ne sait rien de la réalité de ce que serait le travail collaboratif de 30 équipes de 5 chercheurs en moyenne. La décision de création d'une équipe de recherche épidémiologique « en réseau », associant des chercheurs de disciplines et de localisations différentes doit donc répondre à des critères de qualité au moins aussi rigoureux que ceux concernant la création d'une équipe sur site : en effet, les exigences de productivité et de qualité seront les mêmes, et les difficultés de fonctionnement seront plus grandes.

Ces équipes en réseau devront présenter des projets scientifiques ambitieux, bénéficiant de leur organisation en réseau. La multidisciplinarité (que permet plus facilement le réseau) devra être présente et argumentée. On devra éviter également autant que possible les problèmes de double appartenance : de même qu'un enseignant chercheur ne peut appartenir — sauf exception — qu'à une seule structure de recherche, un enseignant chercheur « en réseau » devrait éviter d'émarger doublement, à une structure locale et à une structure en réseau.

Un tel réseau pourrait être à l'initiative (coûteuse) d'innovations⁵ : organiser des vastes essais internationaux pilotés par des organisations virtuelles d'investigateurs (sur le modèle du réseau de médecins sentinelle de l'Inserm) ou de moniteurs d'études cliniques, fonctionnant sur des e-CRF⁶ utilisant une imagerie digitalisée interprétée à distance, des banques biologiques gérées de manière décentralisée et sécurisée ; on peut imaginer que l'on puisse ainsi réaliser « virtuellement » la version du XXI^e siècle de la cohorte de Framingham en s'appuyant sur des croisements de bases de données de prescription, de soins de santé primaire, d'hospitalisations, d'explorations paracliniques, de décès ; constituer des « entrepôts de données » internationaux et multidisciplinaires à haute valeur ajoutée. De tels systèmes « virtuels » ambitieux nécessitent évidemment des moyens financiers importants, plus importants même que ceux qu'on dédierait au même nombre de chercheurs regroupés en un même lieu permettant de couvrir le financement de la cellule de direction (laquelle doit être supérieure et non inférieure au financement de l'administration d'un système localisé équivalent !), le financement de rencontres et séminaires, le financement de bibliothèques virtuelles, et — bien sûr — le financement des projets de recherche en réseau.

Notons que l'organisation de telles équipes de recherche pourrait s'appuyer sur l'EHESP (École des hauts études en santé publique, qui pourrait contracter avec les différents établissements dont relèveraient les personnels de l'équipe en réseau) et pourrait être partie prenante de l'éventuel IRVSP (Institut de recherche virtuel en santé publique).

OPR-3.4 – L'évaluation des équipes

Les structures proposées ci-dessus comportent deux innovations par rapport aux structures actuelles de recherche épidémiologique : l'association avec des organismes à mission de santé publique (et non de recherche), l'ouverture forte sur des nouvelles disciplines (mathématiques, biologie, sciences humaines et sociales). L'évaluation des projets devra être vigilante — en miroir — sur deux points : d'abord, comme indiqué précédemment, ne pas dériver vers une recherche systématiquement finalisée, mais conserver une place importante à la recherche fondamentale. Ensuite, veiller à ce que les chercheurs d'autres disciplines venant à l'épidémiologie fassent réellement de l'épidémiologie. Ceci concerne notamment les mathématiciens : les recherches épidémiologiques de qualité utilisant les mathématiques, les probabilités, les statistiques et l'informatique doivent *in fine* être publiées dans les grandes revues de spécialité des disciplines médicales auxquelles elles s'adressent (et, bien sûr, dans les revues généralistes) ; ceci, évidemment, ne s'oppose pas à la publication simultanée

⁵Le paragraphe qui suit est dû à Antoine Flahault, que nous remercions.

⁶E-CRF : *Electronic Case Report Form* (cahier électronique de dossiers).

des innovations méthodologiques dans les revues de mathématiques. En d'autres termes, on doit éviter la création d'un *no man's land*, ni véritablement mathématique, ni réellement épidémiologique, dans lequel seule une communauté de mathématiciens « attirés par l'épidémiologie » ou d'épidémiologistes « ayant le goût des mathématiques » se retrouverait et fonctionnerait en vase clos. En tout état de cause, il suffit de suivre l'exemple existant à l'étranger, où les mathématiciens travaillant en épidémiologie sont pleinement intégrés dans leur discipline d'accueil (ce qui ne les empêche pas — éventuellement — de poursuivre **par ailleurs** des travaux de recherche mathématiques).

RECOMMANDATIONS

De nombreuses recommandations spécifiques ont été données au fil des différents chapitres du rapport, et/ou dans la section « Orientations prioritaires du rapport » de la conclusion ; l'Académie des sciences les fait siennes.

Ci-dessous est présenté — de façon non exhaustive — le résumé des principales recommandations.

L'Académie des sciences recommande :

En ce qui concerne le développement des systèmes d'information épidémiologique :

1. Que le Conseil national de l'information statistique tienne un rôle beaucoup plus actif que dans le passé dans le domaine des statistiques sanitaires, établisse un rapport détaillé sur les différentes questions évoquées dans ce rapport, et notamment sur l'apport potentiel des différents organismes, séparément et de façon synergique, à l'amélioration de l'information épidémiologique.
2. Qu'un groupe de travail permanent comprenant notamment des membres de la Drees, de l'Inserm, de l'Insee et de l'InVS soit mis en place afin de conseiller les représentants français dans les consultations européennes sur l'élaboration en cours du système européen d'enquêtes statistiques.
3. Qu'un centre multi-organismes de recherches performant sur les méta-analyses en épidémiologie soit créé.

En ce qui concerne l'organisation de la recherche :

4. Que, afin d'assurer la pérennité indispensable des systèmes d'observation à finalité de recherche épidémiologique, une politique de recrutement sur emplois fixes soit mise en œuvre dans les EPST et/ou les universités.
5. Que, par exemple dans le cadre de l'Agence nationale de la recherche, un groupe de travail soit mis en place pour décider de l'opportunité de la création et du financement d'un petit nombre d'épidémiopôles et des relations que ces épidémiopôles devraient avoir avec des structures déjà existantes du domaine des sciences de la vie (génopôles, cancérôles).

6. Que soit favorisée la création d'équipes universitaires associant EPST, agences, opérateurs de bases de données sanitaires, épidémiologistes, statisticiens, sociologues, économistes de la santé notamment dans le domaine de l'analyse, la perception et la gestion des risques liés aux nouvelles technologies.
7. Que soient créées des « équipes de recherche épidémiologique en réseau », dotées de moyens de coordination et de moyens informatiques, permettant le développement de thématiques scientifiques pour lesquelles la pluridisciplinarité est indispensable et pour lesquelles les masses critiques correspondantes n'existent pas sur un même site. Ces « équipes de recherche en réseau » devraient intégrer au moins un quart de personnels des disciplines mathématiques, ainsi que d'autres disciplines non médicales (droit, économie, sociologie, démographie, . . .). Elles devraient présenter une articulation fonctionnelle avec des équipes de biologie, de médecine humaine et vétérinaire.
8. Que soient recrutés dans les universités ou grands établissements, des probabilistes, statisticiens, mathématiciens, informaticiens explicitement « fléchés » en épidémiologie, destinés à participer à l'enseignement et la recherche en épidémiologie dans leurs disciplines, notamment dans le cadre des équipes mixtes ou en réseau décrites aux alinéas précédents.

En ce qui concerne l'amélioration de la culture épidémiologique :

9. Que les enseignants d'épidémiologie organisent, dans le cadre de la réforme LMD, des ESPT (enseignements pouvant entrer dans le système souple du ECTS — *European credit transfer system*) en collaboration avec leurs collègues mathématiciens et biologistes et vétérinaires, afin de faciliter aux étudiants de ces disciplines l'acquisition d'une culture minimale en épidémiologie.
10. Que des séminaires et écoles d'été mixtes épidémiologie-mathématiques (au sens large : incluant probabilités, statistique, informatique) soient organisés.
11. Que des groupes de travail d'épidémiologistes réalisent, en concertation avec les enseignants concernés, des « valises pédagogiques » en épidémiologie, destinées à couvrir tout ou partie de l'horaire réglementaire d'éducation à la santé à l'intention des enseignants des collèges et lycées. Que soient également développées des maquettes d'enseignement à l'attention des professeurs au moment de leur formation en IUFM.
12. Que — à l'image de plusieurs autres pays européens — la formation professionnelle continue des médecins et personnels de santé soit facilitée

grâce à un accès gratuit des personnels de santé aux informations issues de la Cochrane collaboration, ou à des systèmes analogues en langue française qu'il serait souhaitable de créer, grâce à des collaborations dans le cadre de la francophonie.

SUMMARY AND RECOMMENDATIONS

The report focuses on human epidemiology, with the aim of identifying the barriers that limit its development in France, and the solutions that could stimulate it in the future.

Key points of the report are:

- Society at large expresses an increasing demand for epidemiology: on the one hand epidemiology is positively perceived as the scientific base for public health policies. On the other hand, there are some frustrations from the public which expects answers to questions that epidemiology cannot solve, such as demonstrating an absence of risks associated with putative hazards.
- The identity of epidemiology is subject to debate: epidemiology suffers from being perceived as a low-tech science, without highly visible technologies, at a time when there is much identification between science and technological innovation. The frontiers between epidemiology and other sciences (e.g. genetics, genomics) are also ill-defined.
- The size of the budgets and the workforce necessary for competitive epidemiological research are largely underestimated.
- Research in epidemiology must be developed in close synergy with fundamental biological sciences, clinical medicine, veterinary medicine and social sciences.
- There is increasing need for professional statisticians because of the explosion of the information to analyse and for mathematicians especially in the field of infectious diseases epidemiology.
- There is insufficient interaction between academic epidemiology and the large observational databases of the hospitals and national health system: vast sources of information are under-analysed, which in turn does not encourage the data providers to increase the quality of their data.
- The numbers of epidemiologists and laboratories dedicated to epidemiology are lower than in many other European countries, especially in the UK and northern Europe. The annual number of PhDs in epidemiology is low. Epidemiology is only taught in the schools of Medicine, and there is virtually no chance that a Master student in mathematics, biology — and even population biology — receives any information on this field. As a consequence, the teachers in schools and high schools are not prepared

to educate the children to the most basic epidemiological notions, in terms of methods and public health results. A final consequence is the difficulty of communicating to the general public about the various health risks.

The report has identified several priorities in which efforts should be made:

- to rehabilitate observational epidemiology. In brief, we recommend a shift from the widespread attitude of a “problem driven” epidemiology to a “data driven” epidemiology. This implies:
 - The funding of large number of researchers in epidemiology, with permanent positions, or a relevant organisation, to guarantee a follow up which may last decades.
 - The creation of at least one observatory for the general population.
 - The epidemiological exploitation of existing medical databases created initially for purposes other than research.
 - The careful examination of the epidemiological potential of the computerised medical files (“dossier médical personnel”) the creation of which was decided upon to ensure the continuity of health care.
- to disseminate the culture of epidemiology:
 - Among biologists, mathematicians, social sciences specialists who ignore largely the techniques and accomplishments of epidemiology.
 - Among non-medical students during their Master studies by offering credits popularising the methods and results of modern epidemiology.
 - Among medical students, by fully integrating the teaching of epidemiology within the general course work of medical education.
 - At the school and high school levels, to provide a deeper scientific insight to existing programmes devoted to the teaching of Public Health (at large).
 - Among the general practitioners which should be given a free access to the Cochrane Collaboration resources, once the language barrier has been overcome.
 - Among the general public, for example by the creation of epidemiological bulletins directed to the media.
- to strengthen the organisation of epidemiological research by:
 - Opening positions for scientists in universities and medical research institutes (Inserm) to overcome the present underdevelopment of academic epidemiology.

- Ensuring that research groups participating in networks of different institutions and universities working on a same topic attain a critical mass.
- Closely associating academic researchers with health agencies and medical data providers.

Some specific recommendations were made, such as:

- opening of permanent positions in universities for professional mathematicians specializing in epidemiological research (the same could be done for other disciplines such as genetics, ecology, sociology, . . .);
- establishing summer schools bringing together epidemiologists, mathematicians, biologists and clinicians;
- . . .

Groupe de lecture critique

COMPOSITION DU GROUPE DE LECTURE CRITIQUE

Agence française de sécurité sanitaire des aliments (Afssa)

Philippe VANNIER

Directeur de la Santé animale et du bien-être
des animaux

Agence française de sécurité sanitaire des produits de santé (Afssaps)

Anne CASTOT

Responsable des Vigilances

Caisse nationale de l'assurance maladie des travailleurs salariés (Cnamts)

Pierre FENDER

Médecin conseil national adjoint

Centre national de la recherche scientifique (CNRS)

François RENAUD

Directeur de recherche

Commission nationale de l'informatique et des libertés (Cnil)

Jeanne BOSSI

Chef de la division des Affaires publiques
et sociales

Conférence des présidents d'université (CPU)

Laurent DESVILLETES

Directeur du Centre de mathématiques
et de leurs applications (CMLA)

Direction générale de la Santé (DGS)

Lucien ABENHAIM

Professeur à l'université René-Descartes —
Paris V

Direction de l'hospitalisation et de l'organisation des soins (Dhos)

Jean-Pierre DUFFET

Mission de l'observation, de la prospective et de la recherche clinique (OPRC)

Direction de la recherche, des études, de l'évaluation et des statistiques (Drees) et Conseil national de l'information statistique (Cnis)

Philippe CUNÉO

Rapporteur Commission « Santé protection sociale »

École nationale vétérinaire d'Alfort

Barbara DUFOUR

Enseignant-chercheur — Unité pédagogique Maladies contagieuses, épidémiologie et zoonoses

Haute autorité de la santé

Laurent DEGOS

Président

Institut national d'études démographiques (Ined)

Myriam KHLAT

Directeur de recherche

Institut national de la santé et de la recherche médicale (Inserm)

Florence DEMENAIIS

Directeur de recherche

Institut national de la statistique et des études économiques (Insee)

Marc CHRISTINE

Chef adjoint de l'unité Méthodes statistiques

Institut de veille sanitaire (InVS)

Jean-Claude DÉSENCLOS

Responsable du département des Maladies infectieuses

Les entreprises du médicament (LEEM)

Chrystel JOUAN-FLAHAULT

Directeur médical à la direction des Affaires scientifiques, pharmaceutiques et médicales

Les membres du Groupe de lecture critique, désignés par le président ou le directeur général de leur établissement, ont examiné le texte du rapport puis, au cours d'une réunion qui s'est tenue à l'Académie des sciences le 29 juin 2005, ont entendu la présentation de M. Alain-Jacques Valleron, animateur du groupe de travail, et se sont exprimés.

Ils ont formulé des remarques, dont certaines ont été intégrées, avec leur accord, dans le rapport ; dix commentaires font l'objet de contributions signées : elles sont présentées ci-après.

COMMENTAIRES DE L'AGENCE FRANÇAISE DE SÉCURITÉ SANITAIRE DES ALIMENTS (AFSSA)

Philippe Vannier

Directeur de la Santé animale et du bien-être des animaux

Ce rapport est extrêmement complet et de grande qualité.

L'Afssa, en raison du délai imparti pour la lecture du rapport, ne donne que des commentaires partiels concernant l'aspect très particulier qui relève de ses compétences.

Le regret le plus important que l'on peut avoir est la place très limitée réservée à l'épidémiologie animale en tant que telle, même si son importance au regard de la santé publique est par ailleurs soulignée en plusieurs endroits du rapport : il faudrait rappeler que des recherches de qualité en épidémiologie animale sont conduites au sein de l'Afssa, de l'Inra, du Cirad et des écoles vétérinaires. Des modèles importants ont été développés et les applications des travaux sont essentielles en santé animale et santé publique.

Le rapport (sous-chapitre 2.1) souligne l'apport de l'Inserm en matière d'expertise collective pour les contributions de l'épidémiologie. Il oublie les activités multiples de l'Afssa dans ce domaine au travers des activités de ses Comités d'experts spécialisés et des activités d'appui scientifique et technique de ses laboratoires en aide à la décision de ses autorités de tutelle. En effet, de très nombreuses saisines de l'Afssa, depuis sa création, ont été en rapport avec des questions d'épidémiologie : ESST, Fièvre Catarrhale Ovine, West Nile, Rage, Influenza aviaire, Peste Porcine Classique, Brucellose, Tuberculose, Salmonelles, *Listeria*, *Campylobacter*, etc.). Dans le même sous-chapitre, on parle d'une vingtaine de personnes impliquées à plein temps en épidémiologie : ce chiffre nous paraît grossièrement sous-estimé : au sein de l'Afssa, 43 personnes sont impliquées dans différentes activités en épidémiologie. Au sein de l'Inra, 35 scientifiques ont une activité en épidémiologie. Il ne faut pas oublier le Cirad, et l'IRD mais il faudrait disposer des chiffres des agents ayant une activité en épidémiologie.

Nous ne comprenons pas la classification des publications des différents instituts car l'Afssa a publié dans des journaux référencés épidémiologie et cela n'apparaît pas dans le tableau sauf sous la rubrique *Tous journaux*, qui ne nous paraît pas adaptée à la nature des revues où l'Afssa publie. Les revues dans lesquelles les scientifiques de l'Afssa publient en épidémiologie sont nombreuses : *Preventive vet Med*, *Vet Rec.*, *Res in Vet SCI*, *VET. Microbio.*, etc. elles devraient être référencées au même titre que les autres revues et non dans la catégorie *Tous journaux*.

Enfin, même si un paragraphe du rapport mentionne l'importance de la récolte des données solides, il nous semble important de mieux insister sur ce point capital, car la tendance actuelle est plus de développer des modèles en simulant les facteurs d'incertitude, ce qui est toujours plus facile sur le plan opérationnel, que de trouver les moyens de collecter des données fiables et nombreuses sur le terrain ; or le risque est de voir proliférer des modèles trop artificiels qui peu ou prou peuvent conduire les décideurs sur de mauvaises pistes.

COMMENTAIRE DE LA CAISSE NATIONALE DE L'ASSURANCE MALADIE DES TRAVAILLEURS SALARIÉS (CNAMTS)

Frédéric van Rookeghem

Directeur

Je vous prie de trouver ci-dessous, au-delà de la reconnaissance de la qualité globale du rapport, les remarques que la Cnamts fait sur celui-ci et vous les propose en tant que contribution.

Les bases de données de l'Assurance maladie

L'Assurance maladie est très sollicitée par des chercheurs pour que, dans le cadre de leurs travaux, ils accèdent aux informations de ses bases de données. Le rapport signale que les exemples d'enquêtes où l'Assurance maladie aurait répondu favorablement " sont encore trop exceptionnels et encore trop souvent limités à des sujets trop finalisés " (chapitre 1). Les cas de collaboration entre l'Assurance maladie et les organismes de recherche ou chercheurs ne sont pas aussi exceptionnels que ce qui est dit. La liste des collaborations en 2004 et 2005 le montre.

Les problèmes posés par ces collaborations sont souvent minorés et nous pouvons les recenser en quatre points :

1. Les bases de données de l'Assurance maladie n'ont pas pour finalité de répondre à des préoccupations épidémiologiques. Elles sont constituées à partir de données nécessaires pour réaliser le paiement des prestations par les caisses et en vue d'être des outils de la gestion du risque de l'Assurance maladie. Leur utilisation dans le cadre de l'épidémiologie est une fonction dérivée.

Les caractéristiques originelles de ces bases sont à l'origine :

- de complexité (onze *data marts*, plusieurs centaines de variables, variables soumises aux modifications réglementaires et tarifaires, . . .) ;

- d'accessibilité (données confiées à l'Assurance maladie par les assurés en vue d'une prise en charge financière et ceux-ci attendent qu'elles soient gérées avec sécurité ;
 - de limites en termes de champ (absence de variables sociales, le champ des soins remboursés est le seul pris en compte).
2. Le sentiment de nombreux chercheurs qui pensent avoir trouvé une source inépuisable d'informations qui résoudraient à peu de frais pour leurs unités de recherche leurs problèmes d'accès à l'information en santé. Dans la plupart des cas, ils préfèrent utiliser les bases nominatives d'accès complexe et sécurisé à la place de celles qui ont été anonymisées et prévues pour ces utilisations larges. Souvent, ils sous-estiment la complexité alors que l'interprétation des résultats nécessite, dans la majorité des cas, une connaissance poussée des règles et circuits du remboursement, ainsi que des logiques de constitution de ces bases.
 3. Les demandes arrivent à l'Assurance maladie de tous horizons sans qu'elles soient classées par ordre prioritaire par les demandeurs. Il revient à l'Assurance maladie de faire ces choix. Ceux-ci seront toujours nécessaires quels que soient les moyens qu'elle y consacrerait car l'appétence de la recherche à l'utilisation de ces bases est proche de l'infini. La transparence sur les critères de choix serait nécessaire pour que l'impression d'une mauvaise collaboration de la Cnamts à ces recherches disparaisse.
 4. Les moyens actuellement disponibles à la Cnamts pour assister les chercheurs externes sont limités en raison de ses priorités propres, qui sont la gestion des risques et le service.

Les propositions que la Cnamts pourrait faire pour améliorer les collaborations dans le cadre de l'épidémiologie sont les suivantes :

- création d'un comité comprenant des représentants des organismes de recherche, de l'État, d'associations de patients, de l'Assurance maladie, qui donnerait des avis à l'Assurance maladie pour classer les demandes par ordre prioritaire (l'Institut des données de santé pourrait-il avoir cette fonction ?) ;
- formation des chercheurs des organismes de recherche aux requêtes sur les bases anonymisées de l'Assurance maladie ;
- coopération équilibrée entre les spécialistes de l'Assurance maladie de ces systèmes informationnels et les chercheurs, épidémiologistes ou acteurs de santé publique externes.

COMMENTAIRE DE LA COMMISSION NATIONALE DE L'INFORMATIQUE ET DES LIBERTÉS

Jeanne Bossi

Chef de la division des Affaires publiques et sociales

Ce rapport est intéressant par les informations très riches qu'il apporte sur la description des différents types d'épidémiologie et les méthodes mises en œuvre pour l'analyse des données recueillies.

À cet égard, la nécessité de renforcer en France les méthodes de l'épidémiologie descriptive et de favoriser son développement, qui est affirmée à plusieurs reprises tout au long du rapport, se traduit d'ores et déjà à la Cnil à travers l'examen des demandes d'autorisation pour la mise en place de fichiers dans le domaine de la recherche. Cette tendance se traduit ainsi par la volonté exprimée par les chercheurs de pouvoir accéder à des bases de données détenues par l'assurance maladie (Sniiram) ou par les établissements de soins (données issues du PMSI) qui contiennent l'ensemble des informations sur les actes effectués et les prestations servies aux assurés.

L'idée évoquée dans le rapport d'utiliser les données du futur dossier médical personnel, créé par la loi du 13 août 2004, s'inscrit dans cette évolution mais devra sans doute être abordée avec prudence.

Le rôle de la Cnil, autorité de protection des données à caractère personnel créée par la loi, est de s'assurer, dès lors que l'accès à des bases de données constituées à l'origine pour une finalité autre que celle de la recherche est sollicité, que les mesures nécessaires sont prises pour garantir la confidentialité des données ou leur anonymisation dans certains cas. Les données de santé restent des données sensibles qui doivent à ce titre bénéficier d'une protection particulière. Il convient, à cet égard, de noter la vision positive de la protection des données à caractère personnel dans ce rapport.

COMMENTAIRE DE LA DIRECTION DE LA RECHERCHE, DES ÉTUDES, DE L'ÉVALUATION ET DES STATISTIQUES

Philippe Cunéo

Rapporteur Commission « santé protection sociale »

L'épidémiologie humaine est l'analyse de la variation des maladies et la recherche de leurs déterminants. Ceci en fait une discipline qui s'appuie largement sur l'application des méthodes statistiques en santé et dont l'exercice trouve naturellement une grande proximité avec les préoccupations du système statistique public : Institut national de la statistique et des études économiques (Insee), Conseil national de l'information statistique (Cnis) et Direction de la recherche, des études, de l'évaluation et des statistiques (Drees) du ministère de la Santé et des Solidarités.

Le rapport présenté ici est extrêmement riche et stimulant. Il apporte une description très détaillée des relations que l'épidémiologie entretient avec d'autres disciplines, que ces dernières lui apportent les techniques d'analyse et de modélisation (les mathématiques), les fondements théoriques des phénomènes étudiés (la biologie) ou encore l'étude des déterminants sociaux (les sciences sociales). Il réalise ainsi une somme de ce qui constitue aujourd'hui les différentes facettes de la recherche épidémiologique en France et nul doute qu'il comptera pour longtemps comme une solide référence.

La Drees, de par son activité dans le domaine et aussi parce qu'elle assure le secrétariat de la formation santé et protection sociale du Cnis, souhaite y apporter quelques commentaires.

Elle émet en premier lieu le souhait que les diverses recommandations qui sont formulées dans le rapport et dont l'objectif est de contribuer au développement de l'épidémiologie, s'appuient davantage sur l'évolution récente du contexte institutionnel français et sur la grande richesse des outils disponibles ou des projets en cours. Dans cet esprit par exemple, les souhaits de la mise en place de structures pérennes visant à la gestion de grandes cohortes auraient pu faire plus explicitement référence au projet de cohorte d'enfants en train de se mettre en place sous la coordination d'Henri Léridon (unité mixte Ined-Inserm) et dans un cadre pluri-institutionnel. Il aurait également été intéressant de discuter comment

des outils relativement spécifiques à notre pays, tels que l'échantillon démographique permanent de l'Insee ou le projet d'un échantillon de bénéficiaires de l'assurance maladie, sont susceptibles de s'articuler à des initiatives inspirés d'outils étrangers, voire dans certains cas de s'y substituer. Le PMSI, malgré ses imperfections actuelles, constitue ainsi dans le domaine hospitalier un formidable gisement de données d'études et de recherche d'ores et déjà utilisables.

La Drees s'interroge ensuite sur la place privilégiée conférée aux mathématiques dans le titre du rapport, au point de les présenter comme une condition du développement de l'épidémiologie. Les quatre chapitres qui leur sont consacrés occupent un espace important, mais finalement guère plus que celui consacré à la biologie ou aux sciences sociales et sans que le lecteur comprenne véritablement en quoi la relation privilégiée entre ces deux disciplines ferait problème ou constituerait un frein au développement de l'épidémiologie.

En troisième lieu, le rapport se serait enrichi si la description de la pratique épidémiologique avait été étendue à d'autres structures que celles directement insérées dans la recherche publique ou universitaire. La pratique de l'épidémiologie au sein des agences sanitaires aurait mérité ainsi de faire l'objet d'une analyse plus approfondie, de constituer en soi un objet de description, notamment du fait de ses besoins propres. La recherche, certes privée, menée au sein des laboratoires pharmaceutiques aurait également pu être prise en compte. Les efforts importants en matière d'épidémiologie descriptive menés dans le cadre du système statistique public, à la Drees ou à l'Insee, ou dans des organismes qui peuvent leur être associés, devraient également être mentionnés, notamment au titre des opportunités qu'ils sont susceptibles d'offrir.

À cet égard, la Drees souhaite insister sur les particularités du système statistique public français qui associe l'institut statistique national, l'Insee, à des services statistiques ministériels (SSM) au sein de chaque ministère. L'Insee exerce sur ces derniers une autorité fonctionnelle souple, notamment par le biais de la mise à disposition d'une grande partie de leurs cadres. Le ministère de la Santé et des Solidarités dispose ainsi de la Drees, direction d'administration centrale dont un tiers des effectifs est constitué de personnels Insee. C'est dans le contexte très particulier de cette organisation qu'il faut apprécier à la fois l'investissement de l'Insee et celui du ministère dans le domaine des enquêtes sur la santé, dans la mesure où la Drees travaille en très étroite collaboration avec l'institut.

Le rapport se termine par trois recommandations en ce qui concerne le développement des systèmes d'information épidémiologiques.

Il souhaite dans un premier temps que le Conseil national de l'information statistique (Cnis) tienne un rôle plus actif que par le passé dans le domaine des statistiques sanitaires.

S'il est effectivement souhaitable de voir un approfondissement de l'investissement du Cnis dans ce domaine, il convient cependant de noter que sa formation Santé et protection sociale est assez active et a par exemple accordé très récemment des avis d'opportunité favorables à six enquêtes dans le domaine de la santé, parmi lesquelles le projet de dispositif européen d'enquêtes quinquennales par interview (HIS). Le Conseil a également récemment adopté une recommandation visant à encourager les opérations statistiques permettant des analyses longitudinales, cadre dans lequel s'inscrit le projet de cohorte d'enfants. La formation santé et protection sociale devrait cependant prendre davantage en compte à l'avenir, et examiner, les travaux des équipes d'épidémiologistes, notamment de l'Inserm, autres que celles de sciences sociales. Ceci vaut tout particulièrement pour les gros dispositifs longitudinaux.

Le rapport émet également le souhait que soit mis en place un groupe permanent dont l'objectif serait de conseiller les représentants français aux consultations européennes sur l'élaboration du système européen d'enquêtes statistiques. À cet égard, le principe d'une consultation régulière des principales institutions concernées vient justement d'être retenu lors de la dernière réunion de la formation Santé et protection sociale du Cnis. L'intérêt d'une démarche visant à situer dans le cadre européen les réflexions nationales en matière de système d'information ne peut d'ailleurs qu'être salué, même si une place relativement faible lui est consacrée dans le rapport.

Enfin la Drees considère que l'idée de la création d'une structure nouvelle exclusivement dédiée à la méta-analyse, pour intéressante qu'elle soit, doit être approfondie. Il n'est en effet pas sûr, compte tenu des évolutions récentes du contexte institutionnel, notamment de la mise en place de l'IVRSP ou de la création de l'EHESP, que la mise en place d'une structure supplémentaire présente réellement des avantages au regard de ses coûts potentiels.

COMMENTAIRE DE L'ÉCOLE NATIONALE VÉTÉRINAIRE D'ALFORT (ENVA)

Barbara Dufour

Enseignant-chercheur

Le rapport de l'Académie des sciences présente une vision qui nous paraît claire et pertinente de la situation de l'épidémiologie humaine en France.

Toutefois, deux remarques de fond nous paraissent importantes à formuler :

Une carence du rapport sur l'épidémiologie animale

En fait, le rapport porte sur l'épidémiologie humaine, et, en conséquence, il est logique que l'épidémiologie animale ne soit pas traitée de manière exhaustive. Il aurait toutefois été nécessaire que, compte tenu des interrelations fortes entre épidémiologies humaine et animale, un chapitre complet puisse être consacré à ces relations.

Actuellement, seul l'état des forces de l'épidémiologie animale (sous-chapitre 2.2) dans le domaine des zoonoses a été dressé conformément au plan établi par l'équipe de rédaction du rapport. La contribution de l'épidémiologie animale apparaît donc beaucoup plus pauvre qu'elle n'est en réalité.

Les relations entre épidémiologies humaine et animale débordent en fait largement le cadre strict des zoonoses et il aurait été souhaitable qu'au travers d'un chapitre complet consacré à l'épidémiologie animale il puisse être fait état, même succinctement, de l'ensemble des contributions et des apports de l'épidémiologie animale.

Ainsi, les grandes réalisations de l'épidémiologie animale pourraient être citées, car elles contribuent soit directement (zoonoses) soit indirectement (en particulier sur le plan des méthodes) à enrichir la discipline.

Depuis une vingtaine d'années, l'épidémiologie animale s'est beaucoup développée en France dans différents secteurs¹ :

- celui de **la surveillance épidémiologique**. De nombreux réseaux de surveillance épidémiologique (plus d'une vingtaine) de maladies strictement animales comme la fièvre aphteuse, les pestes porcines, la fièvre catarrhale ovine, ou de zoonoses comme la rage, les infections à virus West-Nile ou les salmonelloses d'origine aviaire ou bovine ont été créés. Par ailleurs, sur le plan méthodologique, des travaux ont été conduits sur les méthodes d'évaluation de ces réseaux (audit externes ou indicateurs de performance) ; ces outils pourraient tout à fait être transposés et utilisés pour des réseaux de surveillance des maladies humaines. Enfin, l'utilisation et le développement de systèmes d'information géographique sont à souligner en surveillance épidémiologique animale ;
- celui de **l'épidémiologie analytique des maladies non transmissibles** (écopathologie) pour lequel les méthodes d'analyse factorielle des correspondances ou en composantes multiples ont été particulièrement précieuses et de ce fait très largement utilisées ;
- celui de **la modélisation pour des maladies à haut pouvoir de diffusion**, comme la fièvre aphteuse qui correspond à un modèle particulièrement intéressant, ce qui n'a pas échappé à certaines équipes d'épidémiologie médicale, ou des maladies à longue incubation comme la paratuberculose des ruminants. Ces travaux de modélisation présentent un intérêt méthodologique dépassant le cadre des maladies étudiées ;
- celui de **l'analyse de risque** qui ne fait pas partie, à proprement parler, de l'épidémiologie mais qui s'appuie en grande partie sur des méthodes utilisées en épidémiologie et qui utilise le plus souvent des données issues de l'épidémiologie. Des travaux d'analyse de risque quantitatifs ou qualitatifs ont été conduits depuis quelques années et, là encore, les échanges méthodologiques avec l'épidémiologie humaine pourraient être profitables.

À travers ces quelques exemples, il apparaît clairement que, au-delà des maladies zoonotiques, pour lesquelles les collaborations entre vétérinaires et médecins sont indispensables, l'épidémiologie animale est également porteuse de travaux méthodologiques pouvant enrichir l'épidémiologie humaine. Les méthodes sont, en effet, les mêmes, mais de leurs différentes applications naissent des expériences et des développements utiles à tous.

¹L'association pour l'étude de l'épidémiologie animale (AEEMA) y a contribué en produisant des livres d'épidémiologie animale, en facilitant l'organisation de formations en épidémiologie animale par l'École d'Alfort, en organisant des journées scientifiques annuelles d'épidémiologie et en publiant une revue scientifique destinées à ses 550 membres.

Sur le plan de la formation, il apparaît également que les vétérinaires ont une formation de base qui semble plus solide que celle des médecins dans ce domaine. Passant plus facilement la barrière des mentalités (la médecine des populations est en effet depuis longtemps une nécessité impérieuse pour l'exercice de la profession vétérinaire sur les animaux de rente), l'épidémiologie animale a rencontré apparemment moins de résistance pour s'implanter dans le monde vétérinaire que dans le monde médical. C'est ainsi que la formation des étudiants vétérinaires comporte depuis longtemps maintenant un solide enseignement de base en épidémiologie. Rappelons à cet égard que tous les vétérinaires praticiens (en particulier ceux pratiquant leur activité sur les animaux de rente) sont des agents de l'État, pour un certain nombre de luttes collectives contre des maladies réglementées. Dans ce cadre, en tant que « vétérinaires sanitaires », ils sont amenés à conduire des enquêtes visant à déterminer les causes et les conséquences des maladies enregistrées. La pratique de l'épidémiologie de terrain leur est donc relativement familière.

Ces vétérinaires constituent donc un vivier de personnes possiblement mobilisables pour des enquêtes plus larges, notamment dans le domaine des zoonoses.

Un besoin de formation des décideurs

Le rapport fait clairement référence à la nécessité de développer la formation en l'épidémiologie pour plusieurs catégories d'acteurs :

- les élèves des collèges et lycées et par voie de conséquence celles de leurs enseignants ;
- les médecins au cours de leur cursus de base ;
- les futurs chercheurs en épidémiologie à travers le développement de masters (les DEA actuels étant jugés trop peu nombreux).

Néanmoins, une catégorie d'acteurs est également particulièrement importante à sensibiliser et à former à l'épidémiologie : il s'agit des décideurs.

Ainsi que l'indique le rapport dans son chapitre 8, l'épidémiologie est la base scientifique de la santé publique. Ce qui revient à dire que l'épidémiologie est à l'interface de la décision de santé. Cette position est un peu particulière par rapport à celles des autres sciences.

Si le rapport présente des analyses intéressantes sur les raisons qui font que les recommandations épidémiologiques ne sont pas toujours suivies d'effets, il

n'indique pas clairement qu'un élément essentiel nécessaire à la prise en compte de l'épidémiologie dans les décisions de santé est que les décideurs techniques ou politiques soient correctement sensibilisés à l'épidémiologie. En effet, l'incertitude, qui est un des fondements de l'épidémiologie, doit être comprise et admise par les décideurs, ce qui n'est pas toujours le cas d'autant que l'aversion pour l'incertitude est plutôt la règle dans le domaine de la décision !

Seule une sensibilisation suivie d'une formation de base en épidémiologie permettrait de faire progresser ce point.

S'il est difficile d'imaginer que les décideurs politiques puissent être formés à l'épidémiologie, en revanche, les décideurs techniques devraient tous avoir reçu une formation de base solide dans cette discipline.

COMMENTAIRE DE LA HAUTE AUTORITÉ DE SANTÉ

Laurent Degos

Président

Il s'agit d'un excellent rapport, auquel les grands scientifiques de l'épidémiologie ont participé. Le document couvre le vaste champ de l'épidémiologie. Chacune des contributions est excellente.

Parmi les quelques remarques que l'on peut apporter, citons :

- Chacun des chapitres semble indépendant du voisin, et ainsi peut-on retrouver quelques redondances (de définitions, de méthodologie). Ces chapitres individualisés ont l'avantage d'apporter toute l'information sur un sujet dans un même élément mais diminuent nettement l'impact que peut apporter ce rapport.
- L'esprit critique, notamment des bases de données, des méthodologies, est peu développé au profit des résultats qui ont été obtenus par la discipline épidémiologique en France. Citons par exemple le peu de critique sur le PMSI, pour lequel le recueil des données n'est pas (loin s'en faut) parfait. Bien souvent les médecins ont pris le PMSI pour une base de données budgétaires plus que pour une base épidémiologique.
- Pour une personne étrangère, il aurait été intéressant de connaître les observatoires existants, les bases de données existantes, les grandes cohortes. On ne peut que regretter que bien souvent les données soient disjointes et un répertoire exhaustif aurait été utile pour mieux connaître l'état des lieux en France. Notamment, les tables de mortalité de morbidité, comme celles de l'occurrence des malformations sont-elles disponibles, adéquates, mises à jour et exhaustives ?
- La perception par le citoyen des résultats épidémiologiques, rapportée dans le document, montre un certain scepticisme. Ceci n'est pas propre à l'épidémiologie et aurait pu être replacé parmi les perceptions de toutes les autres activités scientifiques.
- Il est abordé par moment les méthodologies des essais cliniques, et leurs résultats. Peut-on bien individualiser ce qui est un résultat épidémiologique

dans un essai (avec des sélections de patients dans des conditions bien définies) et une épidémiologie en vie réelle qui apporte souvent des résultats différents ?

- Peut-il y avoir une épidémiologie des presque incidents ou accidents (comme cela se fait dans l'aviation civile) avec une recherche d'aide à la décision pour des mesures correctrices ? Les situations (qui sont bien différentes des situations rares) d'incidents n'ayant pas entraînés un accident, ou de circonstances, de comportements, de défauts génétiques n'ayant pas provoqué d'accident pourraient dans d'autres moments, dans d'autres circonstances, être à l'origine d'inconvénients. Cette perspective fait-elle l'objet d'études ? le presque-accident dans d'autres circonstances, est une source de renseignements et donc d'épidémiologie sans maladie. Peut-être sort-on là de l'épidémiologie ou peut-être faut-il l'inclure dans l'épidémiologie (et donc revoir la définition de l'épidémiologie) ?

COMMENTAIRES DE L'INSTITUT NATIONAL DE LA SANTÉ ET DE LA RECHERCHE MÉDICALE

Florence Demenais

Directrice de recherche

Il s'agit d'un rapport très complet et d'excellente qualité.

Nos commentaires sont les suivants :

1. Lors du bilan de l'état des forces épidémiologiques en France, un tableau synthétique des grandes études épidémiologiques françaises et de leur positionnement par rapport aux études du même type menées dans d'autres pays (par exemple, en termes d'effectifs, de champs couverts. . .) permettrait d'avoir une idée plus claire de la situation.

L'insuffisance en personnels techniques stables dans les établissements de recherche pour mener des études de grande envergure et compétitives au plan international est bien soulignée. Il pourrait être ajouté qu'il serait important de créer des nouveaux métiers, comme celui d'infirmières coordinatrices d'enquêtes, comme il en existe dans les pays anglo-saxons, et qui jouent un rôle clé dans la conduite de ces études.

2. Parmi les sources de données pour l'épidémiologie, les registres en population sont bien entendu mentionnés mais il serait intéressant d'en donner une liste plus détaillée (par exemple, sous forme de tableau).
3. Quatre exemples de liens entre épidémiologie et sciences biologiques sont présentés et montrent bien les besoins d'interface entre ces deux disciplines. Les études d'épidémiologie génétique nécessitent la constitution de collections biologiques de plus en plus importantes. Il serait souhaitable de faire le bilan des centres de ressources biologiques (CRB) existants et de souligner la nécessité de développer des procédures de contrôle de qualité des collections pouvant satisfaire les normes recommandées par l'OCDE, de faciliter les interactions entre CRB et d'assurer un soutien financier de ces CRB, qui serait indispensable à leur pérennité.
4. Les interactions entre épidémiologie et sciences de l'information sont aussi amenées à se développer. Il pourrait être mentionné qu'il existe un nombre de plus en plus important de bases de données disponibles sur Internet

(par exemple, en biologie) dont il faudrait faciliter l'accès aux épidémiologistes. Le développement de logiciels et procédures informatiques est nécessaire pour permettre d'intégrer ces bases de données et de faire face à la quantité sans cesse croissante des données disponibles dans différents domaines.

5. En ce qui concerne l'enseignement, il serait bien d'indiquer la nécessité de développer encore plus la pluridisciplinarité en intégrant, par exemple, les enseignements en épidémiologie et biostatistique avec des enseignements en biologie et bio-informatique ou dans le domaine des sciences humaines selon les filières. Ceci est en théorie rendu possible par la mise en place du LMD mais souvent difficilement réalisable en pratique. Il serait aussi important de recommander la possibilité pour des équipes de recherche travaillant dans des domaines pointus d'encadrer des étudiants d'écoles doctorales (ED) différentes et d'éviter un trop grand cloisonnement de ces ED qui pourrait être préjudiciable. Il est donc crucial de favoriser le développement de l'épidémiologie en faisant connaître plus largement cette discipline et en facilitant les interactions avec d'autres champs, aussi bien dans le cadre de l'enseignement et de la formation que dans la recherche.

COMMENTAIRE DE L'INSEE

Marc Christine

Chef adjoint de l'unité Méthodes statistiques

Le présent rapport de lecture émane d'un statisticien public qui n'est pas du tout spécialiste d'épidémiologie et qui n'a jamais travaillé personnellement sur des opérations statistiques relevant du domaine de la santé, mais qui est spécialisé dans les questions méthodologiques de traitement des enquêtes auprès des ménages. À ce titre, mon intérêt et mon apport à ce groupe de lecture critique se situent dans la question de voir, notamment, si le corpus disciplinaire que constitue la statistique (terme très global) est en bonne articulation avec les préoccupations du domaine de l'épidémiologie exprimées dans ce rapport.

Il faut souligner, en introduction, le grand intérêt que présente ce rapport de l'Académie des sciences, surtout pour quelqu'un qui ne connaît pas spécifiquement ce domaine : il trouve là matière à acquérir un volume important de connaissances et de problématiques. Cet intérêt est renforcé par le **caractère pluridisciplinaire** des approches qui sont proposées : l'épidémiologie est en effet envisagée dans ses rapports avec l'histoire, les techniques mathématiques et statistiques, la biologie, la sociologie, la sémantique, la taxinomie, sans oublier les aspects réglementaires. . .

Plusieurs chapitres ont particulièrement retenu mon attention et j'articulerai mes remarques en trois thèmes.

1 | La description des sources et, en particulier, celles relevant de la statistique publique (cf. sous-chapitre 3.1)

Ce sous-chapitre présente les différents acteurs en matière d'observation statistique. Mais, à propos du principal d'entre eux, l'Insee, il regrette qu'il ait un rôle trop faible en matière de statistiques sur la santé. Cette appréciation mérite d'être nuancée, notamment sur la participation jugée « *symbolique* » (titre du § 1.2.1) en matière d'affectation de moyens.

La contribution de la Drees au groupe de lecture critique souligne bien en effet, au contraire, le rôle important de coordination qu'exerce l'Insee à l'égard

du système statistique public et son positionnement comme interlocuteur, notamment vis-à-vis des instances statistiques européennes. De ce fait, s'il est vrai que l'Institut ne consacre que peu de moyens directs à l'observation statistique des phénomènes liés à la santé, la mise à disposition d'une quantité importante de cadres auprès des ministères et, notamment auprès de la Drees, prouve à l'évidence, au contraire, l'importance de son engagement.

De surcroît, l'Insee investit beaucoup, au cours de la période récente ou à venir, en partenariat avec d'autres administrations ou établissements publics (Ined, Drees. . .), sur des outils statistiques innovants, qui sont également évoqués dans la contribution de la Drees : cohortes d'enfants, enquêtes HID (handicap-incapacités-dépendance), échantillon démographique permanent. . . S'agissant de ce dernier outil, s'il est vrai qu'il est très « encadré » comme l'indique le rapport, il paraît exagéré de dire qu'il est peu exploité ; en tout cas, il paraît un enjeu important pour l'avenir comme source d'observation statistique, comme en témoigne à nouveau le projet actuel d'extension en termes de couverture (passage de 4 à 16 jours comme condition d'éligibilité en termes de date de naissance).

D'autres acteurs mériteraient d'être cités en tant que gestionnaires de sources d'observation, parmi ceux-ci : l'Afssa et l'InVS. À noter que l'Insee travaille en collaboration étroite avec l'Afssa, dans le cadre d'une convention, à la conception et à l'exploitation de l'enquête INCA2 sur les consommations alimentaires (le concours de l'Insee se situant principalement sur le plan méthodologique).

Au-delà de l'énumération des sources et des moyens, deux points évoqués dans le rapport doivent être mieux mis en exergue :

- d'une part, reconnaître que l'Insee a vocation à faire des enquêtes en population générale, ce qui constitue un atout pour avoir une vision de la place des phénomènes étudiés au sein de la société, mais aussi un handicap pour étudier finement des phénomènes dont la prévalence ne concerne que des sous-populations très spécifiques.

Cette dernière remarque peut être atténuée par le fait que de plus en plus d'enquêtes réalisées par l'Insee portent sur des champs très circonscrits, les possibilités offertes à l'avenir par le nouveau recensement annuel, en tant que base de sondage des enquêtes, tendant vraisemblablement à accroître cette faculté. Toutefois, existeront toujours des limites, notamment pour l'utilisation de ces enquêtes à des niveaux locaux relativement fins (sauf à mettre en œuvre des techniques statistiques adéquates, du type « petits domaines », mais qui sont tributaires du postulat de modèles comportementaux qu'on pourra toujours contester). Par ailleurs, pour l'étude de phénomènes très particuliers, l'utilisation de fichiers ciblant directement

sur les populations d'intérêt (par exemple en fonction de l'incidence de telle ou telle pathologie) restera sans conteste la seule voie appropriée ;

- d'autre part, souligner que l'accès aux données a été beaucoup facilité et développé au cours de la période récente, *via* notamment les possibilités offertes par le Centre Quételet, même si, bien entendu, cette transmission doit se faire dans le respect des règles relatives à la protection des données individuelles.

Un dernier point : il serait utile de mettre davantage en perspective le système français en fournissant quelques éléments de comparaison internationale en matière de sources de données et d'organisation de l'appareil d'observation. Les aspects internationaux sont en effet un peu diffus dans ce rapport ; or les préoccupations internationales en matière de santé sont au moins aussi importantes qu'en France. J'en donnerai pour preuve, notamment, le fait qu'au dernier colloque francophone sur les sondages (colloque *a priori* généraliste), tenu à Québec en mai 2005, une session entière eût été consacrée aux statistiques de la santé.

2 | L'articulation avec les disciplines lourdes, « conceptualisantes » (mathématiques, statistique, analyse des données, informatique. . .)

Mes remarques porteront essentiellement sur le sous-chapitre 4.1 qui est axé sur les aspects les plus en relation avec mon activité professionnelle, à savoir la méthodologie statistique.

Il me semble que l'objectif de ce sous-chapitre est de montrer ce que ces disciplines peuvent apporter à l'épidémiologie et, inversement, les développements théoriques que cette dernière a pu susciter en retour.

En filigrane, il semble y avoir un autre message qui mériterait d'être davantage explicité et mis en exergue, à savoir l'idée selon laquelle toutes les méthodes mathématiques et statistiques utilisables en épidémiologie sont très complexes et nécessitent des connaissances de niveau deuxième, voire troisième cycle (quand il ne s'agit pas de recherches théoriques en cours sur des domaines très ouverts ou non aboutis).

Le corollaire est que la maîtrise de ces outils ne serait le fait que d'un petit nombre mais qu'inversement — et ce de manière fallacieuse —, l'utilisation de techniques « presse-bouton », *via* notamment des logiciels de traitement des

données fonctionnant comme une boîte noire, risquerait d'entraîner des erreurs d'interprétation, voire de jugement, de la part d'utilisateurs qui ne connaîtraient pas le substrat théorique de ces techniques.

Ce message devrait trouver un écho dans les développements du chapitre 9 consacré à la formation (*cf. infra*).

Au-delà de ces remarques générales, ce chapitre présente une liste de « techniques » statistiques pouvant être utilisées dans le domaine de l'épidémiologie. Le terme de « technique » ne devrait pas avoir ici un sens dévalué car les outils évoqués ont la plupart du temps un niveau élevé de fondements théoriques, la qualité et le renom des auteurs de ce sous-chapitre en étant d'ailleurs une preuve éminente.

De ce fait, on pourrait regretter que ce sous-chapitre n'engendre une certaine frustration liée à la rédaction énumérative et limitative : sur un certain nombre de ces outils méthodologiques (modèle logit, méthodes de vraisemblance, *bootstrap*...), quelques lignes seulement très allusives ne permettent absolument pas de se faire une idée de l'apport des dits outils (c'est le cas notamment des § 3.1.1, 3.3 et 3.2.). Ceux qui les connaissent déjà jugeront cette présentation beaucoup trop sommaire, les autres ne verront pas quelle est leur utilité ni à quels types de question ils peuvent permettre de répondre. En corollaire de cette brièveté, certaines remarques « qualitatives » de ce chapitre manquent un peu de nuances et certains paragraphes restent un peu confus...

3 | La formation (*cf. sous-chapitre 9.3*)

Compte tenu de ce qui a été dit *supra*, il me semble que ce sous-chapitre doit faire une mention particulière des formations en statistique et de leur utilisation possible dans les différentes composantes de l'épidémiologie. Dans sa rédaction actuelle, ce chapitre n'aborde en effet que la formation à l'épidémiologie, avec quelques incursions vers les formations aux outils statistiques utilisables dans ce domaine, mais en tant que discipline intégrée à un cursus tourné essentiellement vers l'épidémiologie.

Faire mention spécifique de la formation à la statistique, dans un paragraphe approprié et distinct, donnerait sûrement plus de poids au message qui souligne que la statistique, lorsqu'elle est intégrée à un cursus médical, est mal considérée par les étudiants, dénigrée et jugée inutile, ce qui est un constat certainement très vrai. Ce d'autant plus — et l'on ne peut qu'être d'accord avec ce que dit le rapport à ce sujet — que la statistique est absolument indispensable aux chercheurs en épidémiologie, qu'elle nécessite des qualités qui ne s'improvisent

pas et qu'elle impose d'éviter des écueils dans lesquels le béotien peut tomber trop souvent.

En définitive, que le rapport présente la statistique comme une discipline en soi, avec, potentiellement, des applications spécialisées à l'épidémiologie, renforcerait, je pense, le message relatif à son importance, sans affaiblir le constat négatif actuel sur les défauts d'une discipline devenant secondaire lorsqu'elle est intégrée au sein d'un cursus médical. Cette mise en exergue procurerait sans doute une incitation claire à ne pas se contenter de formations dévalorisées de statistique, diluées au sein des études médicales, et une preuve de l'intérêt de l'apport de la statistique au domaine de l'épidémiologie.

À ce titre, il pourrait être en particulier utile que le rapport cite les formations universitaires en statistique, pouvant trouver un débouché dans l'épidémiologie ou, plus généralement, les disciplines médicales, ainsi que les établissements les dispensant : parmi ces derniers, l'une des écoles de l'Insee, l'Ensaï (École nationale de la statistique et de l'analyse de l'information) (elle est, il est vrai, évoquée au sous-chapitre 3.1), mais aussi l'Isup (Institut de statistique de l'université de Paris) et sans doute divers enseignements doctoraux à l'Université.

Ceci étant, il conviendrait sans doute aussi de souligner que la confrontation et le dialogue entre des spécialistes d'un domaine (l'épidémiologie) et des « techniciens » (sans connotation péjorative : ce peut être des doctorants en statistique de très haut niveau) ne s'exercent pas spontanément : trouver le langage commun, permettre une bonne interaction entre l'expertise scientifique et la prise de décision politique. . . , nécessitent peut-être que le statisticien ait aussi une formation médicale approfondie. Cette remarque sur les difficultés de l'acculturation mutuelle vaudrait, bien entendu, dans nombre d'autres domaines socio-économiques et pas seulement dans celui de l'épidémiologie.

COMMENTAIRE DE L'INSTITUT DE VEILLE SANITAIRE (INVS)

Gilles Brückner* et Jean-Claude Désenclos**

* Directeur général

** Responsable du département des Maladies infectieuses

Ce document de travail est une contribution extrêmement utile. Il est très complet, couvre l'historique de la discipline, ses champs d'application, son enseignement, plusieurs de ses champs en développement, fait l'état des lieux des forces en présence, donne des perspectives et fait des propositions pertinentes. Enfin il consacre à part entière le métier d'épidémiologiste.

Les remarques que nous faisons ci-dessous le sont du point de vue des missions de l'InVS, à savoir la contribution de l'épidémiologie à la santé publique et surtout à l'aide à la décision des pouvoirs publics, que ce soit en urgence où à plus long terme vis-à-vis des politiques publiques.

Le rapport, dans le chapitre « Définition, historique et champs de l'épidémiologie » traite de l'épidémiologie d'intervention. Cependant, l'application de l'épidémiologie dans l'aide à la décision est un champ bien plus large et a connu un développement important ces dix dernières années, qui s'est accéléré avec la création des agences, en particulier l'InVS. Le champ d'application s'est progressivement élargi en partant de la surveillance et l'investigation des épidémies, tout d'abord dans le champ infectieux. Il couvre maintenant des aspects méthodologiques nouveaux (analyse de risque, enquêtes en population, veille prospective, modélisation. . .) et des champs de plus en plus larges (environnement, maladies chroniques, santé-travail, vigilances, toxicovigilance. . .). La structuration institutionnelle a aussi évolué dans notre pays ces dernières années avec, notamment, une organisation régionale (Cellules interrégionales d'épidémiologie) et la création de nouvelles agences (Afsse puis Afsset. . .). De plus, la création récente du Centre européen de surveillance et de prévention des maladies, à Stockholm, à l'initiative de l'Union européenne, est un jalon supplémentaire dans le développement et l'institutionnalisation de la pratique de l'épidémiologie pour l'aide à la décision au niveau européen, voire global. Ces quelques éléments et les crises récentes de santé publique, dont la canicule, nous laissent à penser qu'un chapitre spécifique « Épidémiologie, veille sanitaire

et aide à la décision des pouvoirs publics en santé publique » aurait pu trouver sa place dans ce rapport. L'InVS pouvait, si besoin, y apporter une contribution.

Il nous semble aussi qu'un chapitre aurait pu être consacré aux risques environnementaux en complément des chapitres sur les aspects génétiques, moléculaires, sur le vieillissement et le cancer. En effet, l'épidémiologie environnementale pose de nombreuses questions méthodologiques sur lesquelles une réflexion prospective spécifique aurait été utile. En termes de veille sanitaire, de nombreuses questions de méthodes se posent de la même manière. Les approches épidémiologiques doivent prendre en compte des démarches particulières liées à des valeurs seuils de référence, à des modèles avec ou sans seuil, à des évaluations d'impact sanitaire.

Dans le rapport, il nous semble que l'approche cas-témoins est insuffisamment valorisée par rapport aux grands projets de cohortes. Il aurait été utile, dans un chapitre ou sous-chapitre, comme celui sur l'épidémiologie théorique, de rappeler en quoi cette approche méthodologique a été et reste extrêmement utile et contributive, de montrer ses évolutions conceptuelles et méthodologiques et de mettre en perspective ses applications actuelles et futures.

Le chapitre sur l'interface épidémiologie animale et humaine est décevant. Il fait un état des lieux rapide et ne traite pas fondamentalement de ce que pourrait être, au plan théorique et pratique, cette interaction que tout le monde appelle de plus en plus de ses vœux.

Les réflexions et propositions sur la formation sont très pertinentes. Il aurait été néanmoins utile d'approfondir l'analyse sur les problèmes que pose l'enseignement de cette discipline dans les facultés de médecine, très hospitalo-centrées et peu en phase avec la pratique et la recherche en santé publique communautaire. De la même manière, une analyse de l'insuffisance ou de l'incapacité de l'Université à anticiper les besoins de formation en épidémiologie vis-à-vis des besoins des agences aurait été très utile. Une recommandation pourrait en être qu'universités et agences sanitaires interagissent le plus en amont possible pour anticiper ces besoins de formation.

COMMENTAIRE DU LEEM

Chrystel Jouan-Flahault

Directeur médical des affaires scientifiques, pharmaceutiques et médicales

La part industrielle de la recherche en épidémiologie

L'épidémiologie expérimentale, décrite dans le premier chapitre de ce rapport comme une des trois composantes de l'épidémiologie, comprend notamment en son sein les essais randomisés, eux-mêmes représentés pour une large part par les essais thérapeutiques randomisés destinés à tester l'effet d'un médicament par rapport à un placebo ou à un comparateur actif.

Dans ce secteur d'activité et de recherche, les entreprises du médicament occupent une place majeure ; ainsi, en France, en 2003, 71 % des essais cliniques déclarés à l'Afssaps (Agence française de sécurité sanitaire des produits de santé) étaient promus par les industriels du médicament¹. Ces essais portaient principalement sur les médicaments d'oncologie, de neurologie, et du système cardiovasculaire.

Les essais cliniques conduits par les entreprises du médicament répondent tous à des standards de qualité très élevés, définis par des recommandations internationales (bonnes pratiques cliniques, *guidelines* européens). Le respect de ces recommandations est évalué de façon très rigoureuse par des experts ; il constitue une condition nécessaire à l'octroi des autorisations de mise sur le marché.

Pour la recherche industrielle dans le secteur du médicament, la France possède des atouts certains par rapport aux autres pays, mais le maintien de cette attractivité est actuellement en péril, notamment du fait de l'entrée en compétition de nouveaux acteurs, comme par exemple les pays de l'Europe de l'Est ou d'Amérique latine.

Pour se maintenir à la place qu'elle occupe, la France devra impérativement conserver ses atouts actuels que sont la simplicité et la rapidité des autorisations

¹Rapport annuel 2003 de l'Afssaps. L'évaluation des essais cliniques, p. 78.

administratives, et ce malgré l'entrée en vigueur de la transposition de la directive européenne sur les études cliniques (directive 2001/20/CE) dans le cadre de la loi de santé publique du 9 août 2004, qui va harmoniser l'ensemble des pays européens sur la réglementation relative aux essais cliniques.

Elle devra également développer les atouts qu'elle possède déjà, en termes de qualité des soins et d'expertise scientifique et médicale, pour créer une réelle valeur ajoutée dans un type de recherche clinique dite « exploratoire », au sein de pôles d'excellence tels que les cancéropôles récemment mis en place.

L'industrie pharmaceutique est, par essence, l'acteur majeur de la recherche clinique sur le médicament, puisqu'elle est en charge de la découverte, du développement et du suivi de l'immense majorité des molécules à usage thérapeutique. Elle possède, dans ce domaine, une expertise validée et reconnue, avec des méthodologies en constante évolution, pour répondre aux besoins de l'innovation thérapeutique et aux exigences croissantes des autorités d'enregistrement. Elle doit donc être prise en compte à sa juste place dans cet état des lieux sur l'épidémiologie en France.

Présentation à l'Académie
des sciences

par Jean-François Bach

Membre de l'Académie des sciences

Comme son nom l'indique, l'épidémiologie fut d'abord l'étude des épidémies, leur recensement, la recherche de leurs causes et des facteurs de prédisposition mais aussi leur dynamique avec la dimension de prédiction qui y est associée. Progressivement, la démarche fut étendue aux autres domaines de la pathologie et permit d'aider les médecins cliniciens à cerner la complexité des maladies liée à la variabilité considérable des individus (hétérogénéité génétique), des comportements et de l'environnement. Une démarche délibérément appuyée sur les mathématiques, au départ la statistique puis la modélisation, contribuant de façon majeure à la rigueur nécessaire du raisonnement et de l'action médicale. Les progrès méthodologiques s'étendirent rapidement à l'évaluation des nouveaux traitements, avec les essais thérapeutiques dont la méthodologie et, partiellement, la pratique, ont été confiées à des épidémiologistes. Le champ s'est encore élargi avec l'émergence de l'épidémiologie génétique, de l'épidémiologie sociale et même de l'épidémiologie animale. En bref, c'est un pan considérable de la médecine, essentiel pour l'abord des grands problèmes de santé publique, pour ce qui concerne l'analyse de la morbidité et des facteurs qui y contribuent avec la possibilité d'identifier les facteurs de risque, d'analyser, de façon prospective et donc prédictive, la dynamique des maladies et d'évaluer de façon objective les nouveaux traitements.

Le rapport RST rédigé par Alain-Jacques Valleron et ses collaborateurs aborde en profondeur tous ces aspects. L'importance de la discipline en santé publique est bien documentée, étayée par des exemples signifiants.

Le recueil des données est, bien sûr, fondamental. Il représente très souvent le maillon faible en raison des moyens considérables à mobiliser, surtout en personnels. L'importance majeure des statistiques de santé a donné lieu à la mise en place de nombreuses enquêtes, conduisant à de multiples registres ou centre de statistiques, soutenues par différents services de l'État, notamment l'Insee (Institut national de la statistique et des études économiques), la Drees (Direction de la recherche, des études, de l'évaluation et de la statistique). La multitude de ces enquêtes appelle un effort de coordination. La qualité des données recueillies est assez variable et pourrait être sans doute améliorée en mettant l'accent sur les études les plus importantes. Il faut ajouter à ces enquêtes les recueils de données réalisés dans les hôpitaux, ici aussi avec une qualité hétérogène. L'accumulation de toutes ces données représente une grande richesse. Encore faut-il pouvoir en tirer parti. La difficulté est telle que l'on a été amené à définir des stratégies informatiques sophistiquées de recherche de données au sein des bases considérables existantes, le « *data mining* », traduit de façon libre en « fouille des données ». Ces techniques, issues du milieu industriel et commercial, ont des applications majeures. Elles font appel aux techniques les plus modernes de l'informatique. Elles posent aussi de réels problèmes théoriques sollicitant les mathématiciens. Leur développement est essentiel pour faire face à la multitude croissante des données dont il serait regrettable de ne pas tirer parti. Une aide

considérable devrait être apportée par la généralisation du dossier médical informatisé. Encore faudra-t-il le rendre obligatoire pour tous les patients et trouver les réponses aux nombreuses questions posées par leur utilisation, notamment au niveau de l'analyse automatisée de certaines données, en particulier celles issues de l'imagerie. L'utilisation prospective des données est essentielle pour les décisions de santé publique.

Le principe de la modélisation n'est pas nouveau. Elle remonte en fait au XVIII^e siècle avec la présentation devant notre Académie alors Académie royale des sciences, par Daniel Bernoulli, de ses travaux de modélisation de la variolesation, une technique prophétique destinée à prévenir la variole par scarification avec des croûtes de peau de patients atteints de la variole. La modélisation moderne utilise des méthodes mathématiques et informatiques puissantes qui ont fait la preuve de leur efficacité, notamment pour la détermination de la durée d'incubation du sida et le nombre prévisible de malades atteints du nouveau variant de la maladie de Creutzfeldt-Jacob. Les techniques de modélisation trouvent de nombreuses applications dans les maladies infectieuses, plus particulièrement pour évaluer des politiques de vaccination ou de nouvelles stratégies thérapeutiques, pour anticiper le devenir des épidémies. Elles sont également appliquées dans le domaine du cancer notamment pour quantifier les effets de l'exposition aux facteurs cancérigènes.

Deux champs de la recherche clinique sont rattachés à l'épidémiologie, même si le fond de leur activité n'y est pas étroitement lié. L'épidémiologie génétique est, en fait, partie intégrante de la génétique humaine surtout pour ce qui concerne les maladies polygéniques. Dans ces maladies, d'importantes cohortes de patients doivent être constituées pour rechercher des gènes de prédisposition, soit par criblage du génome, soit par étude de gènes candidats.

Cette approche est l'objet de programmes ambitieux dans les maladies infectieuses et encore plus dans les maladies cardiovasculaires, le diabète, les cancers, les démences et le vieillissement. Le rapport présente plusieurs de ces programmes de façon très informative. Peut-être peut-on se poser la question, avec le rédacteur du rapport, de savoir quelle place donner à ces études par rapport à l'étude des facteurs comportementaux et environnementaux qui relèvent plus directement de l'activité spécifique des épidémiologistes. En d'autres termes, si l'épidémiologie génétique est à l'évidence un des domaines majeurs de la recherche médicale moderne, elle dépend sans doute plus des généticiens que des épidémiologistes, quitte, bien sûr, à ce que les équipes de généticiens recrutent des chercheurs et des cliniciens formés à l'épidémiologie.

S'agissant des facteurs environnementaux considérés au sens large ainsi que les grandes stratégies thérapeutiques ou préventives, il est nécessaire de constituer de grandes cohortes de plusieurs dizaines ou centaines de milliers de sujets.

Cela a été fait avec grand succès dans les pays anglo-saxons et en Scandinavie, à un bien moindre degré en France. On peut vivement le regretter car nombre des problèmes posés sont propres à notre pays. Or, la création et le suivi de ces cohortes demandent des moyens considérables très bien présentés dans le rapport autour de quelques exemples. Les organismes de santé et de recherche publique français sont confrontés là à un problème important qui doit les conduire à prendre des décisions rapides. Un chapitre du rapport est consacré à l'épidémiologie sociale, c'est-à-dire à l'évaluation du rôle des facteurs socio-économiques sur la santé individuelle.

Un des messages forts du rapport est la nécessité impérieuse d'intégrer l'épidémiologie dans les disciplines qui lui correspondent : la clinique, la pharmacologie, la génétique et la santé publique. Il convient aussi, et cela est discuté avec beaucoup de pertinence dans le rapport, de rapprocher, plus que cela n'est fait aujourd'hui, les épidémiologistes de l'ensemble des mathématiciens et non pas seulement des statisticiens. L'importance croissante de l'épidémiologie théorique avec toutes les applications de la modélisation l'illustre bien. En fait, la question des frontières de l'épidémiologie conçue comme discipline se pose avec acuité. Il est certes évident que la discipline s'étend au-delà des champs des maladies infectieuses. Cette extension a, cependant, conduit l'épidémiologie à s'impliquer dans quasiment toutes les activités médicales. Dès lors, le nombre d'épidémiologistes disponibles pour ces recherches apparaît insuffisant. Le chapitre du rapport analysant les moyens et les forces françaises dans la discipline insiste à juste titre sur le nombre insuffisant de chercheurs et de techniciens. L'auteur du chapitre se montre très satisfait de la qualité de ces chercheurs, ce qui resterait à vérifier dans le contexte de la compétition internationale. En tout état de cause, il est essentiel d'élargir la discipline en formant à l'épidémiologie des cliniciens, des pharmacologues et des généticiens plutôt que de confier aux épidémiologistes professionnels tous les domaines de recherche qui relèvent en théorie de leur compétence. Deux chapitres très intéressants du rapport sont consacrés aux insuffisances actuelles de la formation, tant pour ce qui concerne les études secondaires que les études supérieures, notamment les études médicales. Cette remarque s'applique aux essais thérapeutiques, sujet majeur traité de façon superficielle, qui doit être pris en charge en commun par des cliniciens et des méthodologistes biostatisticiens, indépendamment de l'appartenance de ces derniers à la discipline épidémiologique définie au sens large.

S'agissant des recommandations, elles sont pertinentes mais aussi très nombreuses en raison de l'importance et de la multitude des problèmes. Il conviendra peut-être de les hiérarchiser et de les classer en fonction de leur cible.

En conclusion, ce rapport sur un sujet central à la recherche médicale et à la santé publique vient à un bon moment. La qualité de sa présentation et de sa rédaction lui assurera une grande audience. L'introduction et les conclusions

présentent une excellente vue d'ensemble à laquelle il sera important de joindre par ailleurs un résumé assez détaillé (une dizaine de pages) qui permettra de faire ressortir les points saillants sans obliger le lecteur à lire exhaustivement l'ensemble des textes, le plus souvent excellents mais parfois un peu spécialisés.

INTERVENTION DE MAURICE TUBIANA

Membre de l'Académie des sciences

Tout en approuvant totalement le rapport, je voudrais faire trois remarques :

1. Le rapport propose, à juste titre, la création de nouveaux laboratoires d'épidémiologie. Je pense qu'il serait utile de créer des liaisons structurales entre certains de ces laboratoires et des groupes de recherche médicale, de façon à introduire des coopérations systématiques. L'expérience montre que les groupes les plus efficaces sont ceux où épidémiologistes et cliniciens ont appris à travailler en étroite liaison et se connaissent bien. Ces formes de coopérations structurées semblent plus fécondes que les coopérations épisodiques. Il faudrait les privilégier.
2. L'ampleur du nombre de sujets nécessaires dans une enquête épidémiologique est de plus en plus considérable et nécessite souvent des études multinationales. Dans l'étude EPIC sur la relation alimentation et cancer, la cohorte est de 700 000 personnes (dont 50 000 Français) provenant de sept pays différents, membres de l'Union européenne.

Dans les études de génomique, le nombre de sujets nécessaires est aussi très grand. Par exemple dans les études des relations entre génomique et évolution du cancer dans les cancers du sein, il semble qu'il faille une dizaine de milliers de malades atteints de cancer du sein, ce qui ne peut être obtenu en un temps raisonnable que dans des enquêtes multinationales.

La Commission européenne devrait être saisie de cette question car le lancement de telles enquêtes doit se faire à l'échelle de l'Union européenne.

3. Le réseau Cochrane. Le rapport souhaite, et on ne peut qu'être d'accord avec ce souhait, que ces données, extrêmement utiles et souvent indispensables en cancérologie, soient disponibles pour l'ensemble du corps médical français. Il y a donc un problème lié à la langue qu'il faut surmonter (le réseau Cochrane comprend plusieurs milliers de spécialistes parlant anglais). L'idée de créer un réseau francophone paraît irréaliste. Il n'existe pas, dans les pays francophones, les moyens humains ou financiers nécessaires pour créer un tel réseau. Les données colligées par un

réseau francophone ne pourraient donc pas supporter une comparaison avec celles du réseau Cochrane. Il serait donc préférable, au moins dans un premier temps, de trouver un moyen de lier la France à ce réseau et éventuellement d'y associer un système de traduction (automatique?).