

3.2

SURVEILLANCE, MODÉLISATION MATHÉMATIQUE ET ANALYSE CONTEXTUELLE

Les maladies infectieuses sont l'un des domaines où les fondements théoriques ont été les plus développés en épidémiologie. La théorie mathématique des épidémies fournit ainsi un cadre de référence pour la reconstitution historique des pandémies passées, contribuant à une meilleure compréhension des mécanismes de transmission, une alerte plus précoce vis-à-vis des phénomènes émergents, et désormais la prévision de la diffusion épidémique dans le temps et l'espace. A la différence des sciences physiques cependant, il manque encore de lourdes plateformes d'observation qui permettraient la validation de pans entiers de cette théorie et de nouveaux développements.

La connaissance fine des inégalités de santé (âges, appartenance socio-professionnelle, sexes...) et des mécanismes qui conduisent à l'exposition des populations à des risques environnementaux, sanitaires et professionnels, mobilise l'épidémiologie, la modélisation mathématiques, les sciences de la vie et de la nature et les sciences humaines et sociales (telles que la géographie, la sociologie, la démographie, l'économie, l'anthropologie). Les configurations de facteurs qui caractérisent ces questions impliquent de renouveler les approches utilisées. L'approche contextuelle offre des perspectives pour analyser simultanément des données collectées aux niveaux individuels et collectifs, les variations socio-spatiales. Autant de connaissances qui sont déterminantes pour élaborer des politiques de prévention adaptées.

Connaissance des populations et dimensions économiques et sociales

Les données : spatialisation et intégration – Système de surveillance en temps réel et modélisation mathématique des épidémies : alerte précoce, prévision, et évaluation des stratégies de prévention

Antoine Flahault, Inserm unité 707, Université Pierre et Marie Curie

Contexte et dynamique scientifique à l'étranger

Des associations statistiques ont été retrouvées entre des facteurs environnementaux (climatiques, météorologiques, de couverture végétale) et des séries épidémiologiques (fièvre de la vallée du Rift, dengue hémorragique, grippe). L'originalité de ces méthodes tenait moins ici en la mise en œuvre de techniques mathématiques ou statistiques très innovantes, qu'en la mise en réseau de données issues de disciplines qui n'avaient pas l'habitude de se rencontrer : des équipes issues des sciences de l'observation de la Terre (par satellite ou de terrain), des épidémiologistes, des virologues, des entomologistes. Ainsi, la température de surface (mesurable en temps-réel par satellite) de la mer du Golfe du Bengale semble fortement liée aux épidémies de choléra au Bangladesh¹. La température de l'eau aurait un impact important sur le développement du plancton, réservoir du vibron cholérique. Linthicum et coll.² ont montré que des observations satellitaires effectuées en temps-réel permettaient de prédire des épidémies de fièvre de la vallée du Rift au Kenya. Il s'agit d'une fièvre hémorragique virale transmise (à l'homme et au bétail) par un moustique. Les auteurs ont pu observer qu'une couverture végétale anormalement abondante dans la corne de l'Afrique de l'Est (visible par satellite) due à un excès de précipitations locales que les climatologues ont montré être associées aux oscillations climatiques El Nino/La Nina mesurées dans l'Océan Pacifique Sud, favorisaient au Kenya la prolifération des insectes vecteurs de la maladie et le risque épidémique.

L'accumulation des séries épidémiologiques en France, disponibles sans restriction d'accès sur Internet grâce au travail continu et bénévole des médecins généralistes du réseau Sentinelles et de l'Inserm depuis novembre 1984, a permis par exemple de mettre au point une méthode de détection et de prévision des tendances épidémiques de grippe et de gastro-entérites à l'échelle régionale. Cette méthode, dite des « analogies »³, empruntée à la météorologie, permet de prédire avec fiabilité à partir de l'expérience passée, à plus de trois semaines, une courbe épidémique de syndromes grippaux et de diarrhées aiguës en France.

Avec FluNet et l'expérience récente du SRAS, nous savons désormais que la prochaine pandémie de grippe sera monitorée avec Internet⁴. Aujourd'hui, les données accumulées chaque jour dans le système FluNet servent à la mise au point par l'OMS des recommandations annuelles sur la composition vaccinale d'une saison hivernale à l'autre dans chaque hémisphère. L'exemple du SRAS nous a enseigné à quel point un système d'information fonctionnant en temps-réel sur Internet avait été utile au printemps 2003 pour maîtriser la diffusion pandémique de la maladie.

Ainsi, d'une part ces systèmes d'information en temps réel permettent de proposer des prévisions épidémiologiques et contribuent à une prévention et un contrôle plus efficace de ces pathologies. D'autre part, la mise en place de grandes bases de données d'observation du vivant, couplées aux observations de la Terre et de son environnement permettra de mieux comprendre les mécanismes d'émergence et de diffusion des épidémies, notamment en apportant des possibilités de valider les théories sur la dynamique des épidémies.

Dans l'étude de la transmission des maladies contagieuses émergentes, les modèles mathématiques, statistiques et informatiques, basés sur une division de la population en catégories correspondant au

¹ Lobitz B, Beck L, Huq A, et al. From the cover: climate and infectious disease: use of remote sensing for detection of *Vibrio cholerae* by indirect measurement. *Proc Natl Acad Sci U S A*. 2000;97:1438-43.

² Linthicum KJ, Anyamba A, Tucker CJ, Kelley PW, Myers MF, Peters CJ. Climate and satellite indicators to forecast Rift Valley fever epidemics in Kenya. *Science*. 1999;285:397-400.

³ Viboud C, Boelle P, Carrat F, Valleron AJ, Flahault A. Prediction of the spread of influenza epidemics by the method of analogues. *Am J Epidemiol*. 2003 (sous presse).

⁴ Flahault A, Dias-Ferrao V, Chaberty P, Esteves K, Valleron AJ, Lavanchy D. FluNet as a tool for global monitoring of influenza on the Web. *JAMA* 1998 Oct 21;280(15):1330-2

statut susceptible, infectieux ou immun des sujets, occupent une place de premier plan.⁵ L'impact potentiel d'une attaque bioterroriste avec le virus de la variole ne peut être étudié autrement, pas plus que les conséquences d'un recul de la vaccination contre la poliomyélite, ou celle d'une épidémie de fièvre aphteuse. Pour des maladies contagieuses endémiques, dont le contrôle est ou pourrait être réalisé par la vaccination, cette approche est également nécessaire.

Aujourd'hui, trois formulations de ces modèles sont couramment utilisées : 1) déterministe, basée sur un système d'équations différentielles ordinaires ou partielles ; 2) stochastique, basée sur des équations différentielles stochastiques *ou* sur une modélisation markovienne par processus de saut ; 3) individu centré, basée sur une description de règles locales d'évolution individuelle.

L'histoire naturelle de chaque type d'infection conduit généralement à proposer une division originale de la population en compartiments ; voire à s'affranchir de la discrétisation compartimentale pour admettre que certaines caractéristiques des individus ou des agents pathogènes puissent se répartir selon un continuum.

Le réalisme de la description dépend également du choix de la structure des contacts dans la population étudiée : du simple mais peu réaliste pan-mixage où les contacts sont équiprobables entre tous sujets de la population, on peut passer à une structuration des contacts par âge, puis à des réseaux de complexité croissante : spatialisation sur un treillis, structure en « petits mondes ».

L'étude théorique (mathématiquement ou par simulation numérique) des propriétés des modèles développés permet alors l'investigation de scénarios dans l'évolution et le contrôle des maladies.

Un second intérêt du développement de tels modèles est qu'ils fournissent un cadre pour l'estimation de paramètres épidémiologiques d'intérêt dans les maladies transmissibles, lorsqu'une approche statistique plus classique se heurterait à la non indépendance des données observées, ainsi qu'à la nature partielle des observations (par exemple uniquement l'incidence : les données collectées par le réseau *Sentinelles* rentre évidemment dans cette catégorie).

Avec ces modèles réalistes, l'estimation statistique reste cependant problématique. Ainsi, il est par exemple nécessaire de "marginaliser" la fonction de vraisemblance par rapport à toutes les trajectoires possibles correspondant aux événements non observables. Bien que complexe, ce programme peut être réalisé de manière approchée grâce aux méthodes d'intégration stochastique de type Monte Carlo par Chaînes de Markov (MCMC)⁶. Récemment sont même apparues des méthodes de simulation numérique, dites de « Las Vegas » qui permettraient théoriquement d'obtenir des résultats exacts, par couplage depuis le passé⁷.

Le travail d'une équipe américano-soviétique⁸ a permis de reproduire avec précision la dynamique mondiale de diffusion du nouveau virus grippal de type A (H3N2) identifié à Hong-Kong en 1968-69. Un modèle mathématique analogue, appliqué au cas du virus du SIDA dix ans après son identification, mais avant le recensement de son extension planétaire avait permis de prévoir la propagation rapide de l'épidémie vers l'Asie du Sud-Est⁹. En utilisant des techniques statistiques de rétrocalcul, à partir des seules données acquises et d'hypothèses sur la distribution attendue de la durée d'incubation, il a été possible de prédire dès le début de la pandémie du SIDA la durée moyenne d'incubation en l'absence de traitement entre 5 et 10 ans selon les auteurs^{10 11}, à une époque où certains pensaient qu'il y avait des porteurs sains qui ne développeraient jamais la maladie.

Les épidémiologistes n'ont pas pu prévoir l'émergence du syndrome respiratoire aigu sévère (SRAS) et sa diffusion dans le monde en 2003^{12 13}. Les méthodes de prévisions reposant sur l'analyse de séries chronologiques passées ne sont évidemment pas adaptées à la problématique des maladies émergentes. Les modèles mathématiques disponibles se sont montrés peu performants dans le cas du SRAS, car très sensibles aux paramètres de transmission et aux conditions initiales qui restent largement méconnus, et peu enclins à rendre compte de phénomènes aléatoires, comme celui des

⁵ Anderson RM, May RM. Oxford Science Publications, Oxford, 1991

⁶ Gibson GJ, Renshaw E. IMA J Math Appl Med Biol 1999

⁷ Propp J, Wilson D. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms* 1996, 9:223-52.

⁸ Rvachev L, Longini I. A Mathematical Model for the Global Spread of Influenza. *Math Biosci* 1985;75:3-22

⁹ Flahault A, Valleron AJ. A method for assessing the global spread of HIV-1 infection based on air travel. *Math Pop Stud* 1992, 3:161-71.

¹⁰ Costagliola D, Mary JY, Brouard N, Laporte A, Valleron AJ. Incubation time for AIDS from French transfusion-associated cases. *Nature*. 1989;338:768-9.

¹¹ Medley GF, Anderson RM, Cox DR, Billard L. Incubation period of AIDS in patients infected via blood transfusion. *Nature*. 1987;328:719-21.

¹² Vogel G. Sars outbreak : Modellers struggle to grasp epidemic's potential scope (News). *Science*. 2003. 300 :558-59.

¹³ Dye-C, Gay-N. Modeling SARS epidemic. *Science*. 2003. (Publié en ligne le 23 Mai ;10.1126/science.1086925).

« *superspreaders* » (personnes à haut potentiel de diffusion). On peut dire que tous les modèles SIR déterministes étaient plutôt alarmistes, à tort au cours du printemps 2003. L'introduction de technologies plus récentes (modélisation SIR stochastiques « individus-centrés ») permettant notamment la simulation - sur de très gros calculateurs - de contacts entre des millions d'individus d'une ville entière ou d'un pays pourrait à l'avenir pallier les carences actuelles des modèles plus anciens et s'avérer un moyen performant pour comprendre et prédire l'émergence de nouveaux virus. Ils commencent à être développés dans le cas de la lutte contre le bioterrorisme¹⁴.

Principales questions scientifiques

Un grand nombre de questions de santé publique ne reçoivent de réponses qu'à partir de l'utilisation de modèles mathématiques, validés lorsque l'exercice est possible auprès d'observations recueillies souvent à partir de longues séries constituées dans le temps et l'espace. Nous en citerons quelques unes à titre d'exemple parmi beaucoup d'autres :

Quel est l'impact prévisible d'une stratégie de vaccination universelle contre le méningocoque de type C ? Quels sont les conséquences de différents scénarios de contrôle d'une attaque bioterroriste par l'agent de la variole ? Quel est le rôle attendu de la vaccination contre le pneumocoque sur l'évolution des résistances aux antibiotiques de cette bactérie ? Peut-on prévoir la diffusion de l'épidémie de grippe durant l'hiver en France ? Les épidémies de grippe sont-elles liées à des facteurs climatiques ?

Ce type de questions, relevant souvent d'une actualité particulièrement sensible, génèrent des travaux multidisciplinaires, qui doivent être développés autour de plate-formes épidémiologiques ouvertes sur l'extérieur, en collaboration avec les institutions nationales et internationales. Ces travaux devront bénéficier des outils de recueil d'information les plus performants et les plus fiables et s'enrichir des méthodes d'analyse les plus pertinentes du moment.

Le potentiel de recherche et d'innovation en France

Plateformes épidémiologiques

Dès 1995, l'Inserm a mis en ligne - en accès libre de tous droits et avec l'avis favorable de la CNIL - la base de données du réseau Sentinelles créée en novembre 1984 et mise à jour continuellement et sans interruption depuis. Le site fournit à l'utilisateur un atlas épidémiologique le renseignant en temps réel sur l'épidémiologie de certaines maladies transmissibles en France (comme la grippe, les gastroentérites, la varicelle). Ainsi, sur le plan épidémiologique, la France, est restée l'un des pays les plus avancés dans le domaine de la surveillance électronique des maladies. Là où les *Centers for Diseases Control* à Atlanta aux USA ou le *Royal College of General Practitioners* au Royaume Uni continuaient à utiliser des échanges de disquettes (les CDC se sont mis tardivement à utiliser des systèmes non graphiques basés sur le téletexte pour la transmission de messages), l'Inserm proposait un système d'information téléinformatique basé sur un réseau de médecins Sentinelles. C'est ainsi que s'est constituée l'une des plus grandes bases de données mondiales sur certaines maladies transmissibles. Le réseau comportait au 1er janvier 2005, 1204 médecins généralistes libéraux en activité dont une fraction (environ 250) se connectaient chaque semaine pour recueillir des données. Le protocole de recueil prévoit une surveillance permanente et continue d'indicateurs de morbidité (grippe, diarrhées, varicelle, rougeole, oreillons, hépatites A, B, prescription de tests sérologiques pour le VHC (et suivi), urétrites masculines, suicides et tentatives de suicide, crise d'asthme, et hospitalisations demandées par le généraliste).

L'Organisation Mondiale de la Santé (OMS), disposant de nombreux émetteurs d'information, qui sont aussi de véritables « sentinelles » dispersées dans tous les Etats membres, a demandé à l'Inserm dès 1997 de développer FluNet pour la surveillance de la grippe dans le monde. Le site (JAMA, 1998) est conçu comme un système intégrant sur une même plate-forme d'ergonomie facile d'utilisation, 1) la saisie sécurisée de données à distance par des centres accrédités, 2) l'administration du site par l'OMS à Genève qui crée les comptes et mots de passe des nouveaux utilisateurs et définit leurs privilèges d'accès, qui met à jour le site et ses pages d'information et d'alerte ainsi que les liens avec

¹⁴ Legrand J, Viboud C, Boëlle P, Valleron AJ, Flahault A. Modeling responses to a smallpox epidemic taking into account uncertainty. *Epidemiol Infect.* 2003 (sous presse).

d'autres institutions, et 3) l'accès libre et sans restriction de tout utilisateur d'internet à des analyses automatisées en temps réel de la base, fournissant des cartes, des animations vidéo, des graphiques ou des tableaux de la situation épidémiologique qu'il souhaite étudier. Ce site recevait dès mars 2003 une moyenne de 70 000 *hits* par jour, soit l'un des sites sur l'épidémiologie des maladies infectieuses les plus visités au monde. Cette moyenne quotidienne était de 5 000 *hits* quatre ans plus tôt. D'autres projets ont été par la suite développés suivant une méthodologie analogue. Ainsi, les dix dernières années de l'épidémiologie de la rage humaine et animale sont en ligne depuis 1998, et les données sont mises à jour régulièrement par les centres collaborateurs de l'OMS pour ces pathologies. Parallèlement, le réseau de médecins Sentinelles allaient permettre la mise en place de projets innovants, favorisant des expériences pilotes de forte interaction entre la ville et l'hôpital, les secteurs libéraux et publics, notamment dans le cadre d'un programme d'évaluation par un essai randomisé du dépistage du cancer broncho-pulmonaire par le scanner hélicoïdal multibarrettes à faibles doses (programme Depiscan).

Modélisation mathématique

L'objectif principal de la recherche n'est pas d'assurer la veille épidémiologique du pays. Il se trouve qu'un "produit dérivé" d'une plate-forme épidémiologique comme le réseau Sentinelles est de fournir des données de surveillance à l'Institut de Veille Sanitaire qui renouvelle chaque année depuis sa création une convention avec l'Inserm pour contribuer au fonctionnement du système. Mais l'objectif principal d'une recherche dans le domaine est de disposer d'un système de recueil, le plus pérenne possible, qui permet plusieurs années plus tard d'utiliser de longues séries temporelles et spatiales. De nombreux modèles mathématiques encore aujourd'hui reposent sur des théories mais n'ont jamais pu être validés par des observations. Les données acquises par le réseau Sentinelles depuis près de 20 ans pour des pathologies comme la grippe, la rougeole, et depuis plus de 10 ans pour les diarrhées aiguës et la varicelle permettent désormais de confronter le corpus de modèles existants aux observations françaises. Ainsi, l'Inserm a développé un programme de modélisation mathématique sur les maladies transmissibles, en particulier sur la dynamique temporo-spatiale de la grippe et de l'infection par le VIH (Stat Med 1988, Lancet 1990, Am J Public Health 1991, Math Pop Stud 1992, JAIDS 1992, Eur J Epidemiol 1994). La biodiversité des souches virales de gastroentérites en France en hiver a pu être étudiée par des virologues de Paris et de Dijon (Brachet, 2003). La cohérence des épidémies de grippe observées en France, aux USA et en Australie a pu être regardée (C Viboud, Emerg Infect Dis, 2003). Les liens entre des indicateurs climatiques, les souches virales circulantes et la taille et la sévérité des épidémies de grippe a été étudiée avec une équipe de la NASA et d'Ann Arbor University (Viboud, Eur J Epidemiol, 2004). Le même type de programme, orienté sur les liens entre séries épidémiologiques et climatiques a été développé avec le CNES. Fin 2002, la Direction Générale de l'Armement et l'InVS ont contracté avec l'Inserm pour le développement de modèles de diffusion d'agents biologiques d'origine malveillante (variole, peste, fièvres hémorragiques chez l'homme (Legrand, Epidemiology & Infection, 2004), ou en collaboration avec l'INRA et l'AFSSA sur la fièvre aphteuse chez l'animal (Le Ménac'h, Vet Res 2005). Ces travaux avaient pour but d'aider à mieux comprendre les mécanismes de diffusion des épidémies, et de fournir des projections sur les niveaux de prévalence à venir.

Des collaborations avec une équipe d'économistes de Harvard, a permis, à partir des données Sentinelles, l'évaluation médico économique de l'intérêt d'une immunisation sélective des adultes contre la varicelle (Hanslik 2003). Des données issues d'enquêtes menées en population, en collaboration avec l'Institut Pasteur de Paris, a permis d'étudier par modélisation mathématique la sélection de la résistance aux antibiotiques, plus particulièrement dans le cas du pneumocoque (Temime 2003). Des travaux théoriques portant sur la modélisation SIR, l'estimation de paramètres épidémiques à l'aide de méthodes MCMC et l'analyse des séries chronologiques non linéaires ont également été menés en collaboration étroite avec des mathématiciens et des statisticiens de l'Université Pierre et Marie Curie.

L'hypothèse de travail

La mise en oeuvre de l'ensemble de ces outils, leur confrontation avec des recueils de données épidémiologiques plus précises, plus fiables et plus rapides, permettront de proposer des cartes de risque pour les maladies émergentes ou ré-émergentes dans les différentes régions concernées du globe contribuant à la prévention et au contrôle des épidémies. L'utilisation de ces méthodes génère des questions théoriques qui font l'objet de développements associés.

Perspectives scientifiques et priorités envisageables

Constituer des banques de données de qualité

Un grand nombre de modèles mathématiques ne peuvent pas être confrontés aux données réelles, voire ne peuvent pas être opérationnels faute de système d'information fonctionnant en temps-réel. Le réseau Sentinelles, mais aussi les systèmes de recueil d'information développés avec l'Organisation Mondiale de la Santé offrent un ensemble de bases de données uniques¹⁵, géo-référencées, mises à jour sans interruption, permettant de tester un grand nombre d'hypothèses et de modèles issus de la théorie mathématique des épidémies, souvent ancienne mais peu validée, faute de données appropriées.

Développer un programme de modélisation autour de la plateforme épidémiologique

En renforçant les compétences en modélisation en France, en disposant de plate-formes épidémiologiques nationales et mondiales on disposera d'une recherche capable de traiter les questions que seuls les modèles mathématiques peuvent espérer aborder en santé publique. Des questions théoriques n'ont pas tardé à émerger lors de la confrontation des modèles aux données réelles : quel est l'impact de l'hypothèse de panmixage sur la dynamique épidémique (c'est-à-dire de l'hypothèse d'une probabilité équivalente pour tous les individus de se rencontrer dans une population) ? Quelle famille de modèles répond le mieux aux observations faites sur l'épidémiologie du SRAS, par exemple à la co-existence de « superspreaders » (personnes à haut potentiel contagieux) et de personnes assez faiblement contagieuses ? Comment estimer les paramètres d'un modèle SIR à partir d'une approche stochastique ou individu centrée (lorsque classiquement on utilise des modèles déterministes pour l'estimation des paramètres) ?

Perspectives de collaborations entre disciplines

L'évaluation de l'impact de certains risques sanitaires, par exemple ceux concernant les pathologies émergentes ne peut être réalisée (ou ne peut être réalisée facilement), par les méthodes épidémiologiques classiques. Pourtant les besoins aussi bien en termes de pilotage de la recherche cognitive associée, qu'en termes d'aide à la décision publique sont très importants en santé publique. Une approche par modélisation mathématique fournit le cadre nécessaire pour apporter des éléments rationnels à l'évaluation de ces risques ou à l'étude de l'efficacité des scénarios de maîtrise de ces risques^{16,17,18}.

La notion d'état de santé dans son acception la plus large est induite par des interactions complexes entre les caractéristiques intrinsèques des populations et leurs milieux. Les facteurs culturels, les statuts socio-économiques, l'accès aux systèmes de soins, jouent un rôle significatif sur la santé des individus et des populations. Ainsi, la plupart des affections sont liées aux effets conjugués de facteurs multiples plutôt qu'à un seul agent infectieux facilement identifiable. Le potentiel épidémiologique contenu éventuellement dans l'étude de l'impact du réchauffement global parce qu'il peut favoriser l'extension de maladies transmissibles (fièvres hémorragiques, paludisme, choléra), l'appréciation de la menace que représenterait dans une société moderne la diffusion malveillante d'agents du passé (variole, peste) nécessitent d'être explorés par des travaux de modélisation.

Programmes de recherche et d'innovation envisageables

15 Florey CV, Ben-Shlomo Y, Egger M. Epidemiology and the World Wide Web : Is there a Net Benefit ? » *Epidemiologic Reviews* 2000, 22:181-85.

16 Hastings, Alan; Palmer, Margaret A, A Bright Future for Biologists and Mathematicians? [*Science* Vol 299(5615) (28 March 2003).

17 M. A. Palmer et al., Accelerating mathematical-biological linkages: Report of a joint NSF-NIH workshop (2003).

18 Keeling, Matt J. et al., Dynamics of the 2001 UK Foot and Mouth Epidemic: Stochastic Dispersal in a Heterogeneous Landscape *Science* Vol 294(5543) pp 813-817 (26 October 2001)

Renforcement et création de plateformes épidémiologiques

Renforcer les capacités des plateformes, notamment dans deux directions :

l'observation du vivant, en proposant aux autorités chargées de la veille sanitaire, en France ou à l'Étranger, des modèles de recueil, d'analyse et de redistribution de l'information épidémiologique et géographique basés sur l'utilisation des nouvelles technologies de la communication et répondant aux exigences et aux normes internationales de qualité ;

l'expérimentation en santé communautaire, en proposant des systèmes d'information intégrant la médecine de ville, d'hôpital, des secteurs publics et privés autour de projets de santé publique à haute valeur ajoutée.

En s'engageant dans une démarche d'assurance qualité.

Programmes d'épidémiologie mathématique

Poursuite de développements de modèles à visée opérationnelle en santé publique. Ainsi des modèles appliqués à la simulation de stratégies vaccinales doivent être mis en œuvre pour envisager l'impact de recommandations dans le domaine notamment de la lutte contre l'hépatite B et la rougeole, la tuberculose. La simulation et l'évaluation du risque d'attaque bioterroriste par des agents comme celui de la variole, mais aussi de la peste ou de fièvres hémorragiques ou d'épizooties ré-émergentes doivent être menées en collaboration avec la DGA, l'InVS, l'INRA et l'AFSSA. Les interactions entre des indicateurs climatiques ou météorologiques et les épidémies de maladies infectieuses doivent être explorées en collaboration avec les chercheurs des sciences de la Terre et de l'Espace.

Les modèles homogènes utilisés pour étudier la dynamique des maladies infectieuses supposent que les populations et les pathogènes sont bien mélangés. Ils ignorent donc les hétérogénéités et les corrélations qui se développent dans une propagation réelle de l'infection, au travers d'un réseau de contacts. Les vraies populations n'étant ni homogènes ni des graphes ordonnés, ne rentrent que dans certains cas limites dans une de ces catégories. En s'inspirant de résultats issus des sciences sociales (Milgram, 1967)¹⁹, Watts et Strogatz (1998, 1999)²⁰ ont introduit les réseaux du type « small-worlds » (SW). Les SW ont été introduits comme un modèle intermédiaire entre les graphes ordonnés (« regular lattices ») et les graphes aléatoires (« random graphs »), étant donné qu'aucun des deux ne s'est avéré pouvoir être un modèle général capable de représenter des systèmes complexes divers, tels que les réseaux de neurones, les écosystèmes, ou l'internet. C'est en utilisant ces concepts, nouveaux en épidémiologie, mais que l'on manipule depuis plusieurs années en neurobiologie numérique, que l'on étudiera comment la dynamique de la propagation d'une maladie infectieuse dépend-elle de la structure de la population au sein de laquelle elle émerge. Il existe une littérature considérable en épidémiologie consacrée aux réseaux sociaux (Anderson & May, 1992; Eames & Keeling, 2002; Watts & Strogatz, 1999) puisque les maladies sont souvent transmises par le contact social. Un exemple important est la diffusion du SIDA et d'autres maladies sexuellement transmises, mais d'autres maladies infectieuses relèvent également de cette problématique (le tout récent cas de la Pneumonie Atypique –SRAS- est exemplaire de l'importance des connexions entre porteurs –le personnel soignant au contact des patients et leur migration). Dans ce cadre, on pourrait proposer de modéliser le rôle joué par la structure de réseau sur la dynamique temporelle de différentes épidémies. Les interactions entre les éléments de la population seront décrites par un SW, où les liens représentent le contact entre les sujets, et l'infection peut seulement procéder à travers eux. En utilisant ces modèles, on se propose d'étudier comment la structure de la population déterminera la dynamique de la propagation d'une maladie infectieuse. En conséquence, cela permettrait de suggérer les points et moments de contrôle les plus efficaces pour contrôler la propagation d'une maladie infectieuse donnée. Les estimations des paramètres dynamiques seront faites à partir des données réelles dont dispose les épidémiologistes, et aussi en comparant avec des modèles stochastiques déjà validés. Des problèmes spécifiques aux SW doivent être résolus pour que ces modèles soient applicables aux problèmes épidémiologiques. Par exemple, dans un réseau quels ensembles de sommets entraînent la diffusion d'une maladie lorsqu'ils sont infectés ? Quels sommets doivent être « vaccinés » pour s'assurer qu'une maladie ne se répande sur un grand pourcentage d'un

¹⁹ S Milgram (1967). The Small-World problem, *Psychology Today*, 2:60-67.

²⁰ DJ Watts, & SH Strogatz (1998). Collective dynamics of 'small-world' networks. *Nature* 393: 440-442.

DJ Watts, & SH Strogatz (1999). *Small Worlds*. Princeton Univ. Press, Princeton.

réseau ? Comment les réponses à ces questions dépendent-elles de la structure du réseau, ou des différentes valeurs attribuées à la probabilité d'infection, ou aux seuils d'infection des individus ? Les modèles SW appartiennent à une classe plus générale de modèles, connus comme les modèles basés sur des agents (« agent-based computational models »), lesquels représentent une population d'entités individuelles, où chaque individu est muni de règles très simples pour l'interaction avec les autres individus et avec son environnement. Ce type de modèles a été utilisé avec beaucoup de succès sur une grande diversité de disciplines (dont les neurosciences), et des lois générales pour les propriétés émergentes des réseaux, valables sur plusieurs disciplines, commencent à être mises en évidence. À partir de cette approche de modélisation, on peut donc constituer des « laboratoires secs » où les épidémies peuvent être simulées, ce qui permettrait l'évaluation de plusieurs stratégies d'intervention à différents points du réseau social, dans des délais très courts, afin d'élaborer des stratégies de prévention, ou de contrôle de la propagation plus efficaces.

L'analyse contextuelle en épidémiologie sociale : avancées et perspectives

Pierre Chauvin et Basile Chaix, Inserm U707
équipe de recherche sur les déterminants sociaux de la santé
et du recours aux soins, Paris

Contexte

Jusqu'à une période récente, l'épidémiologie sociale s'est essentiellement focalisée sur l'étude des déterminants individuels de la santé et de l'accès aux soins, dissociés du contexte géographique, économique ou social (1-3). Mais l'existence d'effets du contexte sur la santé des individus et leur accès aux soins fait l'objet d'une reconnaissance croissante (4, 5). A des fins de recherche, la simple représentation cartographique des variations géographiques des phénomènes de santé fournit, certes, des informations importantes qui permettent de générer des hypothèses sur les facteurs qui influent sur ces phénomènes. Toutefois, la simple comparaison visuelle de cartes géographiques des facteurs d'exposition et des phénomènes de santé ne permet pas d'avancer réellement dans la connaissance des déterminants des phénomènes de santé. Pour ce faire, l'approche d'analyse écologique, qui consiste à mettre en relation des variables explicatives et des données de santé agrégées aux niveaux de zones géographiques plus ou moins fines, a connu un développement important. Néanmoins, de telles approches écologiques ont été régulièrement critiquées dans la littérature épidémiologique. Observant par exemple une association positive entre taux de chômage communal et taux de mortalité communal, il est difficile de tirer des enseignements précis qui puissent être utilisés en santé publique. En effet, transférer une telle association au niveau individuel afin de conclure que les chômeurs ont un risque de mortalité supérieur revient à commettre une erreur écologique et ne permet pas non plus de conclure à l'existence d'un effet collectif ou contextuel du chômage sur l'ensemble des résidents des communes à fort taux de chômage, puisqu'elle ne distingue pas les éventuels effets du chômage aux niveaux individuel et collectif. L'approche contextuelle s'est donc développée à partir du constat qu'il est nécessaire d'analyser à la fois et simultanément des données collectées au niveau individuel et collectif pour avancer dans la compréhension des déterminants sociaux de la santé. En effet, puisque les caractéristiques démographiques, sociales, et économiques des individus sont souvent corrélées aux facteurs du contexte, il est absolument nécessaire de tenir compte des facteurs individuels lorsque l'on cherche à identifier des effets véritablement contextuels.

Dès lors, de telles études contextuelles sont utiles en santé publique pour au moins deux raisons. D'une part, elles permettent de voir si d'éventuels programmes d'intervention de santé publique peuvent être mis en oeuvre de façon complètement invariante sur le territoire, ou l'on gagne à incorporer une dimension contextuelle dans ces interventions, en les faisant varier en intensité ou en nature dans l'espace. De ce point de vue, il est important de développer des méthodes qui permettent de décrire précisément la distribution spatiale des phénomènes de santé. De même, si les facteurs du contexte de résidence sont associés au risque de survenue d'événements de santé (au sens large : comportements de santé, morbidité, mortalité, etc.), des individus que l'on ne classerait pas *a priori* dans un groupe à risque sur la base de leur profil individuel pourraient être tout de même soumis à un certain niveau de risque du fait de l'influence de leur environnement. Tenir compte des variations spatiales et des facteurs du contexte permet donc une meilleure identification des populations à risque lorsque l'on cherche à mettre en place des programmes d'intervention. D'autre part, au-delà de l'identification des populations à risque, de telles analyses permettent de tester des hypothèses relatives aux mécanismes causaux par lesquels ces effets des caractéristiques sociales de l'environnement de vie agissent sur la santé des personnes et, finalement, de progresser dans la compréhension des disparités sociales de santé constatées.

Les principales questions scientifiques et objets de recherche

L'analyse contextuelle en épidémiologie sociale

La question centrale en analyse contextuelle est la suivante : le contexte de résidence, et notamment ses caractéristiques sociales, ont-ils une importance pour la santé ? Une première étape de ces analyses consiste donc à décrire les variations contextuelles, et à réaliser des inférences statistiques à leur sujet. Pour triviale qu'elle apparaisse, cette tâche est pourtant complexe, et délicate à réaliser sur le plan technique. D'une part, la façon dont elle sera conduite dépend des données brutes dont on dispose, et de la finesse avec laquelle on est en mesure de les géocoder, c'est-à-dire de les localiser dans l'espace. Des techniques entièrement différentes devront être utilisées si l'on dispose d'informations géoréférencées au niveau individuel, si l'on dispose de données individuelles géocodées au niveau de zones administratives, ou si l'on n'a accès qu'à des données agrégées au niveau de ces zones.

Quoiqu'il en soit, lorsque l'on cherche à décrire la distribution spatiale des phénomènes de santé, il est nécessaire de distinguer la part des variations géographiques due aux processus définis que l'on cherche à identifier de celle des variations correspondant à des fluctuations aléatoires. On cherche alors en général à réaliser des inférences statistiques sur la distribution spatiale des phénomènes de santé. De façon déterminante, les variations spatiales que l'on a identifiées sont-elles assez importantes pour que l'on puisse en conclure que le phénomène varie dans l'espace, qu'il a une dimension contextuelle, ou doit-on au contraire en conclure qu'il est invariant dans l'espace (6-8) ? Au-delà de l'amplitude des variations spatiales, on verra qu'il est également utile d'obtenir des informations sur l'échelle de variations des phénomènes de santé dans l'espace, afin de déterminer si ces variations surviennent à une échelle très locale ou plus large.

Lorsque des variations spatiales sont identifiées, le phénomène de santé devient un objet d'intérêt pour l'analyse contextuelle, et l'on peut se demander si il est lié aux caractéristiques du contexte de résidence. On appelle facteurs contextuels des variables qui caractérisent les groupes d'individus et doivent comme telles être mesurées à ce niveau. Dans la littérature épidémiologique, ce terme ne renvoie pas seulement à l'environnement géographique de résidence (1), mais concernent également, entre autres exemples, le ménage, le milieu scolaire, l'organisation du travail, ou encore l'organisation des soins.

Beaucoup d'études ont souligné que la prise en compte des facteurs contextuels dans les analyses, en plus des caractéristiques individuelles, pouvait permettre une meilleure identification des populations à risque lors de l'élaboration des programmes de prévention ou de répartition des ressources. Un très grand nombre d'études dans la littérature se sont, par exemple, intéressées aux effets du niveau socio-économique du contexte de résidence. Ainsi, des travaux nord-américains - mais aussi nord-européens - ont montré que l'indigence du contexte résidentiel influait sur de nombreuses variables de santé après que l'on ait tenu compte des caractéristiques socio-économiques des individus (cet ajustement sur des caractéristiques individuelles est indispensable comme nous le soulignons ci-dessous), augmentant notamment les risques de présenter un faible poids à la naissance (9, 10), de contracter une maladie chronique (11, 12), de rapporter une mauvaise santé (13-15) et finalement de mourir (16). Mais les effets du contexte de résidence ne renvoient pas seulement à ses caractéristiques socio-économiques. Par exemple, Sampson et coll. ont montré à Chicago que l'« efficacité collective » (ou la capacité globale des résidents à intervenir pour régler les problèmes communs) avait un impact sur le niveau de violence dans le quartier, et que cet effet surpassait celui des caractéristiques socio-économiques du voisinage (17).

Une difficulté centrale en analyse contextuelle est de savoir si les écarts de morbidité, mortalité, etc. observés d'un groupe à l'autre résultent effectivement d'*effets de contexte*, ou ne dépendent que des caractéristiques variables des individus qui s'y trouvent (c'est-à-dire d'*effets de composition*). Comme on l'a expliqué ci-dessus, l'analyse écologique qui met en relation des facteurs collectifs et des niveaux moyens de morbidité, mortalité, etc. sans tenir compte des caractéristiques individuelles est incapable de mettre en évidence de véritables effets du contexte (18, 19), car elle n'est pas en mesure de les distinguer d'effets qui opèrent au niveau individuel.

La littérature illustre également la difficulté de savoir quels ajustements réaliser pour que l'effet des caractéristiques contextuelles n'exprime *que* l'effet du contexte, mais aussi *tout* l'effet du contexte (ni trop, ni trop peu). Qu'il faille ou non ajuster sur une variable individuelle dépend de son statut : facteur de confusion, ou facteur médiateur appartenant à une chaîne de causalité par laquelle la caractéristique contextuelle est supposée agir (3, 5, 16).

L'apport spécifique des modèles multiniveau dans l'étude des effets du contexte

individus (au niveau 1) se trouvent regroupés au sein d'unités plus vastes (au niveau 2). Les méthodes d'analyse qui ne tiennent pas compte de cette structure complexe de variabilité peuvent être en partie inefficaces. Apparus au milieu des années 1980, les modèles multiniveaux portent une attention particulière à cette structure hiérarchique de la variabilité, et constituent comme tels des outils utiles en analyse contextuelle. Il a été montré que les individus et leur contexte, en tant que sources de variabilité distinctes et hiérarchiquement organisées, étaient souvent mieux représentés par des modèles qui, incluant des effets aléatoires aux différents niveaux de la hiérarchie, permettent de modéliser, au-delà du vecteur des espérances, la matrice des variances et des covariances.

Les modèles classiques aboutissent souvent à des estimations incorrectes : ne tenant pas compte de la violation de l'hypothèse d'indépendance des observations, ils ont tendance à surestimer le niveau de significativité statistique des effets du contexte, et à conclure qu'un effet existe alors que ce n'est pas le cas. Au contraire, en intégrant des résidus aux différents niveaux de la hiérarchie, les modèles multiniveaux tiennent adéquatement compte de la structure complexe des données lors de l'estimation des paramètres. Mais au-delà, cette propriété fait d'eux des outils utiles lors de l'investigation des effets du contexte (15, 20, 21).

L'emploi des modèles multiniveaux à seule fin d'estimation plus juste constitue cependant une utilisation restrictive de ces techniques. En effet, il est possible de considérer la dépendance entre les individus à l'intérieur des zones de résidence comme un phénomène intéressant en soi, plutôt que comme une nuisance dont il faudrait simplement tenir compte au cours des analyses (22, 23). Divers travaux dans la littérature ont montré que les paramètres de variance des modèles multiniveaux permettaient de quantifier les variations de santé qui existent d'un endroit à l'autre sur le territoire, et fournissaient comme tels des informations utiles en santé publique en aidant à évaluer si d'éventuels programmes d'intervention doivent eux-mêmes tenir compte de cette variabilité du phénomène dans l'espace (6-8).

Après avoir décrit les variations du phénomène de santé étudié à l'aide des effets aléatoires du modèle, l'objectif est en général d'expliquer ces variations au moyen de variables individuelles et contextuelles. Les auteurs s'intéressent en général à l'évolution des paramètres de variance du modèle à mesure que diverses caractéristiques explicatives individuelles et contextuelles sont introduites dans le modèle. Une chute importante de la variance inter-zone du modèle lorsque l'on inclut un facteur du contexte de résidence des individus indique par exemple que cette caractéristique du milieu de résidence a un impact notable sur le phénomène.

Ainsi, en résumé, à l'aide des effets aléatoires qu'ils incluent aux différents niveaux, les modèles multiniveau tiennent compte de la structure hiérarchique des données lors de la procédure d'estimation des paramètres et de leurs écart-types. Mais au-delà, les modèles multiniveau permettent de traiter la variance entre zones de résidence comme une source d'information potentielle, au contraire des approches plus classiques qui la considèrent comme une nuisance (4, 24). De cette façon, ils apportent des réponses à certaines questions pertinentes en analyse contextuelle (4).

Grâce à leurs composants aléatoires, les modèles multiniveaux sont en mesure de répartir les variations résiduelles du phénomène entre les différents niveaux, et leur intérêt spécifique est de permettre d'examiner comment évolue cette répartition lors des ajouts successifs de variables individuelles et contextuelles. Un premier apport des modèles multiniveaux en analyse contextuelle est de faire la part entre les effets de composition et les effets proprement contextuels dans la variabilité inter-groupe observée, en comparant la variabilité inter-groupe avant et après inclusion des facteurs individuels. Les auteurs s'attachent ensuite en général à quantifier la réduction que connaît la variance inter-zone lorsque l'on introduit des facteurs contextuels dans le modèle. On examine ainsi si les variables contextuelles utilisées sont capables d'expliquer les variations inter-groupes observées, ou si il faut recourir à d'autres facteurs explicatifs.

En conclusion, les modèles multiniveaux constituent des outils utiles dans l'analyse des effets du contexte, même si leur utilisation est assez complexe et n'est pas toujours requise. Pour peu que les variables individuelles et contextuelles pertinentes soient disponibles ou spécifiquement recueillies, ces méthodes méritent d'avoir toute leur place dans l'actuel courant de recherche en épidémiologie sociale qui, en France, cherche à comprendre — au-delà de la simple description des inégalités socio-

économiques de santé aujourd'hui établie (25) — l'effet des déterminants sociaux sur l'état de santé et le recours au soins.

La nécessité de développer des méthodes statistiques d'analyse spatiale

Les études contextuelles qui ont eu recours à des modèles multiniveaux ont permis de réunir des informations précieuses sur l'importance des effets du contexte sur la santé.(16, 26) Toutefois, l'utilisation exclusive de l'approche multiniveaux dans cette littérature constitue, à son tour, une limite, tant du point de vue de l'un que de l'autre des deux objectifs généraux d'analyse contextuelle rappelés en introduction.

En effet, on peut d'une part douter de la capacité des modèles multiniveaux à aboutir à une description adéquate de la distribution spatiale des phénomènes de santé. Premièrement, comme beaucoup d'autres méthodes en analyse spatiale, ces modèles s'appuient le plus souvent sur un territoire fragmenté en zones aux limites administratives et fournissent des informations sur la distribution spatiale des phénomènes qui sont dépendantes du découpage du territoire utilisé. Cet état de fait a depuis longtemps été décrit dans la littérature géographique sous le nom de « modifiable areal unit problem » (27-29). L'effet d'agrégation qui intervient est d'une part dû à des phénomènes d'échelle (« scale effect ») et, d'autre part, à une échelle donnée, les frontières considérées peuvent grouper les individus d'une multitude de façons différentes. En conséquence, tant les indicateurs qui quantifient les variations d'une zone à l'autre que les mesures des effets du contexte sont dépendants du découpage en zones utilisé, et des différences importantes dans les résultats peuvent être observées d'un découpage à l'autre.

Au-delà de cet aspect, une limite nettement plus importante des modèles multiniveaux est de négliger les connections spatiales qui existent entre les zones, et de ne fournir, de ce fait, que des informations partielles sur la distribution spatiale des phénomènes de santé étudiés. Il a en effet été souligné que les modèles multiniveaux parvenaient à quantifier les variations qui surviennent d'une zone à l'autre, mais qu'ils étaient incapables de décrire la forme que prennent ces variations dans l'espace, notamment si elles étaient complètement aléatoires ou, au contraire, spatialement structurées (30). Or, un certain nombre de méthodes d'analyse spatiale ont récemment été développées, qui permettent de tenir compte du territoire de façon plus continue, et fournissent ainsi des informations plus complètes sur la distribution spatiale des phénomènes (31-33) : « spatial mixed models », « spatial multiple membership model », « conditional autoregressive model », etc.

Au-delà des problèmes que cela pose quand on cherche à décrire les variations spatiales au moyen modèles statistiques, le fait de s'appuyer sur un territoire fragmenté en zones administratives est également problématique quand on cherche à mesurer les facteurs du contexte résidentiel des individus. Rien n'assure en effet *a priori* que le zonage administratif utilisé pour définir les facteurs du contexte corresponde au niveau auquel opèrent les effets contextuels. De plus, même si la taille des zones considérées est adéquate, ces mesures réalisées au niveau des zones administratives sont certainement inadéquates pour les individus qui résident sur leurs marges, car incapables de décrire effectivement le milieu qui les entoure. Il reste à développer des méthodes permettant mesurer les facteurs du contexte dans un espace continu autour du lieu de résidence des individus, voire même, idéalement, dans un espace correspondant au « territoire vécu » des personnes, un territoire plus large et plus complexe à décrire que le seul quartier de résidence (comprenant, par exemple, les territoires où s'exercent la vie professionnelle et les relations sociales les plus « significantes » de l'individu).

Evaluer le gain d'informations fournies aux décideurs de santé publique lorsque l'on recourt à ces approches innovantes plutôt qu'à l'approche multiniveau couramment utilisée et comparer ces différents modèles et méthodes d'analyse spatiale sur des critères statistiques constituent, en soi, un objectif de recherche tant épistémologique que proprement épidémiologique.

La dynamique scientifique à l'étranger

La thématique de l'analyse contextuelle en épidémiologie sociale fait l'objet d'un nombre important et croissant de publications et de recherches dans revues scientifiques de référence, européennes (par exemple le *Journal of Epidemiology and Community Health*, *Social Sciences and Medicine*, ou encore l'*European Journal of Public Health*) ou nord-américaines (en particulier l'*American Journal of*

Epidemiology, Epidemiology). Une revue - *Health & Place* – est intégralement consacrée à la publication de travaux conjugant la géographie médicale, la sociologie en santé, les sciences politiques en santé, la santé publique et l'épidémiologie.

A l'étranger, pratiquement aucun centre de recherche en santé publique important ne fait l'économie de travailler sur ces questions (de même d'ailleurs que, contrairement à la France, tous ces centres – et toutes les écoles de médecine les plus importantes – affichent des départements de sciences sociales en santé). Parmi les plus en pointe sur la question et les méthodes d'analyse contextuelle, on citera par exemple :

- En Europe, le Department of Public Health, Erasmus University, Rotterdam, Pays Bas (JP Mackenbach), le MRC Social and Public Health Sciences Unit, University of Glasgow, Ecosse (AH Leyland).
- Aux USA: le RTI International, Statistics Research Division et la School of Hygiene and Public Health, Johns Hopkins University, Baltimore (G Bobashev), la Harvard School of Public Health, Boston, (SV Subramanian), le Department of Epidemiology, University of Michigan, Ann Arbor, (A Diez-Roux, J Lynch).
- Ou encore, en Nouvelle Zélande, le Department of Geography, University of Canterbury, à Christchurch.

Le potentiel de recherche en France : une interdisciplinarité qui reste à construire

En France, force est de constater qu'une telle dynamique reste à construire et à coordonner. Pour autant, le point de départ n'est pas nul : il existe de nombreuses équipes, dans différentes disciplines, qui s'intéressent déjà à la question, aussi bien en sciences sociales (les approches territoriales de la santé, en particulier en milieu urbain, font l'objet de recherches importantes, mais souvent plus qualitatives, au CNRS et à l'EHESS), en épidémiologie sociale (dans certaines unités de l'INSERM, ces approches contextuelles sont en cours de développement), en démographie (D. Courgeau est, à l'INED, un des tous premiers chercheurs à avoir travaillé sur les méthodes d'analyse multiniveau) et en géographie de la santé. Pour autant, de nombreux développements méthodologiques restent à faire et, simultanément, la description plus systématique des variations contextuelles des situations de santé (dans leur sens le plus large) reste à développer en France. Enfin, une réelle interdisciplinarité, comprenant un échange effectif de concepts sociologiques et de méthodes statistiques - qui, on l'a vu, sont assez « pointues » - reste à susciter et à construire.

Les perspectives scientifiques et les priorités envisageables

L'analyse des effets du contexte sur la santé a connu un développement considérable au cours des dix dernières années, et est aujourd'hui considérée comme l'une des principales voies à suivre pour avancer dans la compréhension des disparités sociales de santé. Les premières générations d'analyses contextuelles ont confirmé qu'il y a un réel intérêt à tenir compte des effets du contexte de résidence lors de l'étude d'un certain nombre de phénomènes de santé. Toutefois, l'analyse contextuelle est loin d'avoir connu en France le développement qu'elle a connu dans d'autres pays (34) et, d'un point de vue méthodologique, a certainement atteint ses limites dans la littérature internationale.

Ces limites renvoient : 1) à l'utilisation d'une distinction trop rigide entre effets individuels et effets contextuels ; 2) à un intérêt peut-être trop exclusif pour un petit nombre de facteurs contextuels (souvent de type socio-économique), sans que des efforts suffisants n'aient encore été mis en oeuvre pour identifier de façon plus précise les mécanismes d'influence du contexte sur la santé ; 3) à l'utilisation de données transversales, qui s'ajoute au problème des biais de confusion pour compromettre définitivement l'identification d'effets causaux du contexte; et enfin 4) à l'utilisation de l'approche d'analyse multiniveau, qui ne permet peut-être pas de décrire et de rendre compte efficacement des variations inter-zones des phénomènes de santé.

La période d'engouement pour cette thématique des effets du contexte sur la santé, qui correspond peut-être à l'enfance de ce champ d'analyse, doit maintenant faire place à une période à la fois critique et constructive, au cours de laquelle les différentes limites évoquées ci-dessus devront être discutées et résolues. Ce n'est qu'en s'attachant véritablement à surmonter ces difficultés qu'il sera possible d'accroître l'intérêt de ces analyses contextuelles pour le champ de la santé publique.

Plusieurs priorités peuvent ainsi être dégagées, auxquelles des modalités d'intervention, en terme de politique de recherche, seraient susceptibles de répondre :

1. *Développement méthodologique* : la description des variabilités contextuelles, en elle-même, et l'analyse de ces déterminants nécessite, on l'a vu, des développements méthodologiques importants et mérite sans doute un appui spécifique, par exemple sous la forme d'un appel à propositions extrêmement ciblé sur cette question.

2. *Interdisciplinarité* : le soutien à des réseaux d'équipes réunissant des disciplines différentes (sociologie, épidémiologie, santé publique, statistiques) constitue aussi une priorité. De tels réseaux devraient être, non seulement soutenus par des interventions spécifiques, mais aussi particulièrement valorisés dans les différentes institutions de recherche elles-mêmes, y compris (quand il s'agit de rapprocher des disciplines issues des sciences humaines comme la sociologie et des sciences biomédicales, comme l'épidémiologie) en prenant en compte les modalités d'évaluation scientifique, souvent différentes, des uns et des autres.

3. *Constitution de base de données* : un des obstacles majeurs au développement des travaux d'analyse contextuelle réside dans la rareté des données individuelles (et non pas agrégées), géoréférencées (au niveau le plus fin possible, dans tous les cas infracommunal), sociales et sanitaires. Si elle n'est pas un travail de recherche en soi, la constitution de telles bases de données (par le rapprochement de données issues des différents systèmes d'information existants : recensement, registres de maladies, données de l'Assurance maladie, données de mortalité, maladies à déclaration obligatoire, cohortes développées dans les EPST, données de l'IRDES, de l'INPES, etc.) est un peu l'équivalent des « grands équipements » pour la recherche fondamentale : c'est une condition nécessaire pour développer des recherches de qualité, elle nécessite des savoir-faire spécifiques et donc des financements spécifiques, elle doit pouvoir bénéficier au plus grand nombre de chercheurs et donc s'accompagner d'un effort de transparence et d'accessibilité –évidemment accompagné de conditions de confidentialité strictes – des différentes institutions propriétaires.

Références

1. Diez-Roux AV. Bringing context back into epidemiology: variables and fallacies in multilevel analysis. *Am J Public Health* 1998;88:216-22.
2. Duncan C, Jones K, Moon G. Health-related behaviour in context: a multilevel modelling approach. *Soc Sci Med* 1996;42:817-830.
3. Diez-Roux AV. Multilevel analysis in public health research. *Annu Rev Public Health* 2000;21:171-92.
4. Duncan TE, Duncan SC, Hops H. Latent variable modeling of longitudinal and multilevel alcohol use data. *J Stud Alcohol* 1998;59:399-408.
5. Diez-Roux AV. Investigating neighborhood and area effects on health. *Am J Public Health* 2001;91:1783-9.
6. Merlo J, Chaix B, Yang M, Lynch JW, Rastam L. A brief conceptual tutorial on multilevel analysis in social epidemiology – linking the statistical concept of clustering to the idea of contextual phenomenon. *J Epidemiol Community Health* 2005; in press.
7. Merlo J, Chaix B, Yang M, Lynch JW, Rastam L. A brief conceptual tutorial on multilevel analysis in social epidemiology – interpreting neighbourhood differences and the effects of neighbourhood characteristics on individual health. *J Epidemiol Community Health* 2005; in press.

8. Merlo J, Yang M, Chaix B, Lynch JW, Rastam L. A brief conceptual tutorial on multilevel analysis in social epidemiology – investigating contextual phenomena in different groups of individuals. *J Epidemiol Community Health* 2005; in press.
9. Ross A, Raab GM, Mok J, Gilkison S, Hamilton B, Johnstone FD. Maternal HIV infection, drug use, and growth of uninfected children in their first 3 years. *Arch Dis Child* 1995;73:490-495.
10. O'Campo P, Xue X, Wang MC, Caughy M. Neighborhood risk factors for low birthweight in Baltimore: a multilevel analysis. *Am J Public Health* 1997;87:1113-1118.
11. Malmstrom M, Johansson SE, Sundquist J. A hierarchical analysis of long-term illness and mortality in socially deprived areas. *Soc Sci Med* 2001;53:265-275.
12. Shouls S, Congdon P, Curtis S. Modelling inequality in reported long term illness in the UK: combining individual and area characteristics. *J Epidemiol Community Health* 1996;50:366-376.
13. Reijneveld SA, Verheij RA, de Bakker DH. The impact of area deprivation on differences in health: does the choice of the geographical classification matter? *J Epidemiol Community Health* 2000;54:306-313.
14. Blakely TA, Lochner K, Kawachi I. Metropolitan area income inequality and self-rated health - a multi-level study. *Soc Sci Med* 2002;54:65-77.
15. Humphreys K, Carr-Hill R. Area variations in health outcomes: artefact or ecology. *Int J Epidemiol* 1991;20:251-258.
16. Yen IH, Kaplan GA. Neighborhood social environment and risk of death: multilevel evidence from the Alameda County Study. *Am J Epidemiol* 1999;149:898-907.
17. Sampson RJ, Raudenbush SW, Earls F. Neighborhoods and violent crime: a multilevel study of collective efficacy. *Science* 1997;277:918-924.
18. O'Campo P, Rao RP, Gielen AC, Royalty W, Wilson M. Injury-producing events among children in low-income communities: the role of community characteristics. *J Urban Health* 2000;77:34-49.
19. Subramanian SV, Kawachi I, Kennedy BP. Does the state you live in make a difference? Multilevel analysis of self-rated health in the US. *Soc Sci Med* 2001;53:9-19.
20. Merlo J, Östergren PO, Broms K, Bjorck-Linne A, Liedholm H. Survival after initial hospitalisation for heart failure: a multilevel analysis of patients in Swedish acute care hospitals. *J Epidemiol Community Health* 2001;55:323-329.
21. Rabilloud M, Ecochard R, Matillon Y. Utilisation d'un modèle de régression logistique à deux niveaux dans l'analyse des variations de pratique médicale : à propos de la césarienne prophylactique. *Rev Epidemiol Sante Publique* 1997;45:237-247.
22. Burton P, Gurrin L, Sly P. Extending the simple linear regression model to account for correlated responses: an introduction to generalized estimating equations and multi-level mixed modelling. *Stat Med* 1998;17:1261-91.
23. Snijders T, Bosker R. *Multilevel Analysis. An introduction to basic and advanced multilevel modelling.* London, England: Sage Publications, 1999.
24. Rice N, Jones A. Multilevel models and health economics. *Health Econ* 1997;6:561-575.
25. Leclerc A, Fassin D, Grandjean H, Kaminski M, Lang T. *Les inégalités sociales de santé.* Paris: La Découverte, 2000.
26. Lochner K, Pamuk E, Makuc D, Kennedy BP, Kawachi I. State-level income inequality and individual mortality risk: a prospective, multilevel study. *Am J Public Health* 2001;91:385-391.

27. Fotheringham AS, Wong DWS. The modifiable areal unit problem in multivariate statistical analysis. *Environ Plan A* 1991;23:1025-1044.
28. Holt D, Steel DG, Tranmer M. Area homogeneity and the modifiable areal unit problem. *Geographical Systems* 1996;3:181-200.
29. Amrhein CG. Searching for the elusive aggregation effect: evidence from statistical simulations. *Environ Plan A* 1995;27:105-119.
30. Chaix B, Merlo J, Chauvin P. Comparison of a spatial approach with the multilevel approach for investigating place effects on health: the example of healthcare utilisation in France. *J Epidemiol Community Health* 2005; in press.
31. Banerjee S, Gelfand AE, Carlin BP. Hierarchical modeling and analysis for spatial data. Boca Raton, FL: Chapman and Hall/CRC, 2003.
32. Congdon P. Applied Bayesian modelling. Chichester, England: Wiley, 2003.
33. Chaix B, Merlo J, Subramanian SV, Lynch J, Chauvin P. Comparison of a spatial perspective with the multilevel analytic approach in neighborhood studies: the case of mental and behavioral disorders due to psychoactive substance use in Malmö, Sweden, 2001. *Am J Epidemiol.* 2005; in press.
33. Chaix B, Chauvin P. L'apport des modèles multiniveau dans l'analyse contextuelle en épidémiologie sociale : une revue de la littérature. *Rev Epidemiol Sante Publique* 2002;50:489-499.

Santé et environnement des enfants : enjeux pluridisciplinaires

*Henri Leridon (Inserm-Ined)
Christine Tichit (Ined)
Sylvaine Cordier, INSERM, U625*

Diverses circonstances physiologiques et environnementales (modes d'exposition) spécifiques expliquent la sensibilité particulière des enfants aux pollutions de l'environnement. De façon plus large, c'est dès le début de la vie intra-utérine que peuvent s'initier des atteintes morphologiques ou fonctionnelles, souvent irréversibles, qui auront des conséquences à différents âges et potentiellement jusqu'à l'âge adulte. Des indices de l'augmentation de l'incidence des certaines pathologies graves de l'enfant, atteignant différents systèmes (respiratoire, neurologique, endocrinologique, immunitaire...), ne peuvent entièrement être attribuables à l'amélioration des méthodes d'enregistrement. Ces constats expliquent que la Commission Européenne, par l'initiative SCALE (Science, Children, Awareness, Legislation, Evaluation), ait souhaité dans son plan d'action 2004-2010 et à l'issue d'un bilan des connaissances, élaborer un programme d'actions de recherche ciblées portant plus particulièrement sur les maladies respiratoires, les troubles du développement neurologique, les cancers et les perturbations du système endocrinien.

Avec le développement simultané d'outils sophistiqués, pour le diagnostic (marqueurs moléculaires...), la mesure sensible des effets (ex. test neurophysiologiques...), ou des expositions (marqueurs biologiques), l'analyse statistique (analyse cas-cas ou cas-parents) et leur mise à disposition dans le cadre d'études en population, des hypothèses auparavant inaccessibles à la démonstration épidémiologique deviennent testables. Leur apport est complémentaire à la démarche de recherche toxicologique dans le domaine, démarche nécessaire pour la démonstration de la plausibilité biologique et des mécanismes en jeu et le développement d'outils utilisables chez l'homme. L'environnement psychosocial joue un rôle important dans plusieurs des pathologies citées, soit indépendamment des pollutions de l'environnement ou comme déterminant de ces expositions. On voit bien donc que le développement des connaissances dans les relations environnement-santé de l'enfant devrait associer nécessairement biologistes, toxicologues, généticiens, pédiatres généralistes ou spécialisés (pneumologie, allergologie, neuropsychologie, neurophysiologie, endocrinologie, oncologie...), épidémiologistes, démographes et sociologues.

La situation idéale est sans doute celle associant, à partir d'un même recrutement représentatif de malades (registre) et d'un groupe témoin, les projets de recherche des cliniciens, des épidémiologistes et des biologistes. Dans le domaine des cancers de l'enfant, les outils sont bien en place. La production de connaissances nouvelles est limitée par le faible nombre de chercheurs impliqués dans cette problématique. Des moyens supplémentaires pourraient également permettre d'envisager la mise en place plus systématique de recueil de prélèvements biologiques et l'accessibilité à la communauté scientifique, des données et/ou échantillons recueillis par les registres, sur le modèle de la mise en commun qui a été initiée en milieu hospitalier par la création de Centres de Ressources Biologiques (??). Les autres maladies de l'enfance bénéficiant d'enregistrement dans certaines régions de France sont les malformations congénitales et les handicaps.

Diverses études de cohorte destinées à recueillir à différents points dans le temps les mesures de l'environnement du fœtus et de l'enfant, et les désordres ou altérations pouvant apparaître au cours de la croissance ont été mises en place en France. Dans le cadre de l'Action 26 du PNSE une grande cohorte suivant près de 20 000 enfants sur plusieurs décennies est en cours de conception. Ce projet ambitieux sera l'occasion de regrouper les chercheurs intéressés de façon que sa mise en place puisse bénéficier des techniques les plus avancées et donner aux chercheurs en retour un outil irremplaçable pour leurs travaux. Ce n'est que s'il existe des soutiens conséquents et sur le long terme, que cette initiative pourra atteindre pleinement ses objectifs.

Contexte scientifique, sanitaire et politique

On constate une demande croissante de recherches sur les déterminants de la santé des enfants. On peut ainsi lire dans la présentation du **Plan national santé environnement (PNSE)** :

« La population française est de plus en plus préoccupée par la menace que peut représenter son environnement. Deux types de raisons l'expliquent. D'une part, les scientifiques mettent régulièrement en évidence des effets sanitaires associés à des polluants biologiques, chimiques ou physiques. D'autre part, la multiplication des informations, déclarations, et appels est à l'origine d'une prise de conscience accrue qui amène le citoyen à exiger un droit à une information claire et à une participation aux décisions. » (Extrait du PNSE, 2004)

La Loi d'orientation de santé publique (2004) inclut, dans le domaine de l'environnement, des préoccupations semblables, et propose de s'intéresser prioritairement :

- à l'exposition au tabagisme passif,
- aux expositions aux allergènes et aux agents chimiques présents dans l'air (intérieur comme extérieur), dans l'alimentation et dans l'eau, et par contact.

Les inégalités sociales face à la santé sont importantes et tendent à s'aggraver, notamment pour des raisons environnementales. Ces inégalités génèrent des préoccupations sociétales et politiques de plus en plus fortes.

Dans ce contexte, la santé de l'enfant pose des problèmes spécifiques, avec des risques souvent accrus, et des enjeux pluridisciplinaires. **L'environnement sera donc défini ici dans un sens large, comme le cadre des influences physiques, chimiques, biologiques, sociales et psychosociales.**

Principaux objets de recherche dans le domaine

Environnement physique

L'enfance est une période de forte sensibilité aux conditions environnementales, à plusieurs points de vue :

- Irréversibilité des dommages causés pendant la période critique de l'enfance au cours de laquelle se développent les systèmes nerveux, immunitaires et reproductifs.
- Fragilité de l'ADN liée à la rapidité de la croissance cellulaire surtout in utero, dans la petite enfance et à la puberté.
- Besoins proportionnellement plus importants que chez l'adulte :
 - en eau et nutriments pour la croissance physiologique
 - en oxygène, compte tenu de l'intensité des activités et du métabolisme infantile
- Risques liés aux activités exploratoires des jeunes enfants, et aux comportements sociaux des adolescents.

L'exposition précoce à des agressions environnementales peut avoir non seulement des effets immédiats sur le développement de l'enfant, mais aussi des conséquences irréversibles qui se répercuteront jusqu'à l'âge adulte. Parmi les agents d'exposition toxiques les plus préoccupants pour la santé de l'enfant apparaissent les métaux, les composés organiques volatiles (cov), les pesticides, les hydrocarbures aromatiques polycycliques halogénés (haph), certains produits de consommation, la toxicité de l'air intérieur (tabac, moisissures), la pollution atmosphérique, les rayonnements solaires et ionisants, le bruit...

Une grande variété de pathologies infantiles sont connues pour être au moins en partie liées à l'environnement. Mais la plupart sont encore scientifiquement mal connues et appellent des recherches notamment sur les âges critiques auxquelles elles apparaissent, les facteurs de risque et de leur degré d'influence sur la santé des enfants. Les principaux types de pathologies incriminées sont les suivants :

- Les troubles d'apprentissage et neurocomportementaux, affectant les capacités intellectuelles, l'attention, ou les capacités relationnelles avec autrui. Ce problème émergent de santé publique touche de 2% à 10% des enfants en France selon les définitions et modes d'évaluation retenus. Les troubles d'apprentissage se répercutent sur les potentialités de réussite scolaire et d'insertion sociale, alors que leur dépistage précoce permet une prise en charge efficace. La méconnaissance actuelle de ces pathologies provient notamment du manque d'observations standardisées (Vaivre-Douret et Tursz, 1999);

- Les allergies et notamment l'asthme dont la prévalence, évaluée à environ 10 % chez l'enfant contre 6% chez l'adulte, est susceptible d'augmenter encore. L'allergie alimentaire du nourrisson est un facteur prédictif important des autres formes d'allergies, et constitue la première étape de la « carrière allergique » qui se structure en trois étapes pendant l'enfance : allergie alimentaire du nourrisson, dermatite atopique dans la petite enfance, asthme et infections respiratoires dans l'enfance. Les trois niveaux de prévention de la carrière allergique sont encore à explorer.

- Les perturbations endocriniennes en lien avec des malformations congénitales, des cancers de la prostate, des testicules et du sein, une altération possible de la qualité du sperme, ...
Les troubles de la croissance : retards de croissance intra utérin (RCIU), petits poids de naissance, troubles de l'immunité, désordres métaboliques (diabète).

- Les cancers de l'enfance sont préoccupants du fait de leur gravité et seraient en recrudescence dans plusieurs pays industrialisés (Steliarova-Foucher E et al, 2004).

Des travaux de recherche et de surveillance approfondis sont nécessaires pour orienter la mise en place des programmes de prévention, notamment quant à l'étude de l'exposition au tabagisme passif, aux allergènes (acariens, moisissures, animaux domestiques...) ainsi qu'aux agents chimiques présents dans l'air ambiant et intérieur, dans l'alimentation, et par contact avec des produits d'entretien, de bricolage...

Environnement psychosocial

La méconnaissance de **l'effet de la pauvreté** sur la santé des enfants est préoccupante compte tenu du nombre alarmant d'enfants concernés : environ un million d'enfants peuvent être considérés comme pauvres, et même 2 millions selon les standards généralement admis en Europe (CERC, 2004). La distribution sociale de plusieurs pathologies renvoie précisément à la grande vulnérabilité des enfants de milieux défavorisés. Le Rapport du CERC sur « Les enfants pauvres » ouvre plusieurs pistes de recherche sur le sujet.

- Les effets de la pauvreté sur l'alimentation infantile sont connus du point de vue de la question de la malnutrition, mais aussi de l'obésité qui touche aujourd'hui 17% des enfants scolarisés en ZEP à 6 ans, contre 13% des autres enfants de cet âge. Reste à comprendre le processus par lequel se mettent en place les pratiques alimentaires à risque.

- La question du renoncement aux soins, liée au niveau de ressources, se pose aussi. Dans l'ensemble les enfants défavorisés bénéficient d'une moindre prévention et d'un moindre recours aux soins, que se soit pour le suivi dentaire, auditif ou visuel. Par contre la population pauvre recourt davantage aux soins hospitaliers (Volovitch, 2003), mais il existe encore trop peu d'études sur l'accès gratuit au soin que ce soit en PMI, à l'hôpital ou en médecine d'urgence (Dumesnil et al. 2003).

- Les conséquences sanitaires de l'insalubrité des lieux de vie : un enfant pauvre sur deux vit dans un logement en mauvais état ou présentant des problèmes d'humidité ou de chauffage, contre un enfant sur trois dans l'ensemble de la population (Rizk, 2003). Ces conditions de logement se répercutent sur la santé des enfants, notamment au niveau de l'asthme qui touche plus d'enfants pauvres, et surtout de l'intoxication au plomb qui toucherait 85 000 enfants de moins de 6 ans, victimes de l'insalubrité de leur logement.

L'impact des **perturbations familiales et relationnelles** sur le développement de l'enfant appelle aussi des recherches dans plusieurs domaines. Les principes de base de comportement social sont acquis de la naissance à 3 ans, pendant que la cytoarchitecture (réseaux neuronaux) du cerveau se met en place. La qualité des interactions suscitées par l'environnement familial et relationnel de la prime enfance joue un rôle décisif dans cette phase de maturation cérébrale (Tubiana, 2005). Outre l'effet d'un déficit relationnel sur la santé mentale de l'enfant, des recherches récentes montrent des effets au niveau du développement corporel. Dans des conditions extrêmes, un exemple africain montre qu'un environnement familial altéré dans la prime enfance (indisponibilité maternelle et/ou épreuves relationnelles) joue un rôle comparable aux carences nutritionnelles dans la malnutrition infantile (Bouville, 2005).

L'apport des données de cohorte en population générale pour l'étude de la santé

Les inégalités de santé résultent généralement d'un cumul d'expositions à des conditions environnementales défavorables, de comportements et d'attitudes développés tout au long du cycle de vie. Seule une observation suivie (et non purement rétrospective) permet de repérer les étapes majeures de ces processus, de prendre en compte toutes ses dimensions (économiques familiales, scolaires) et de mieux comprendre le sens des relations causales. D'un point de vue plus strictement épidémiologique, on peut constater un intérêt croissant pour la *Life Course Approach*, qui se fonde sur la collecte, en continu, de données parfois impossibles à reconstituer rétrospectivement.

Dans ce domaine, les expériences de cohortes étrangères, notamment britanniques (NCDS de 1958, 1970, Millenium Cohort Study), ont déjà beaucoup apporté à la recherche sur les inégalités de santé et ont permis d'identifier des comportements à risque. Les deux premières enquêtes de cohortes britannique ont déjà démontré (Bynner, 2002) :

- Le risque associé au fait de fumer pendant la grossesse pour le développement ultérieur de l'enfant.
- L'intérêt de l'allaitement maternel par rapport au lait artificiel.
- L'effet de l'âge dans la valeur comparée des tests de vision et d'audition.
- Le rôle de la pollution atmosphérique sur l'asthme et les difficultés respiratoires.
- L'étude des effets positifs et négatifs des vaccinations.
- L'étiologie de maladies graves comme le cancer.

Les cohortes anglaises ont aussi renouvelé l'approche des déterminants sociaux, économiques, et familiaux de la santé. Elles ont notamment étudié la relation entre parcours éducatif et accès à la vie professionnelle, la précarité et les moyens de s'en sortir, ainsi que l'évolution de la famille pour laquelle l'approche longitudinale devient presque irremplaçable dans un contexte de transformations familiales intenses.

Perspectives scientifiques

Cohortes existantes en France

L'inventaire des travaux ou projets de recherche épidémiologique de type cohorte qui a été fait à l'Inserm en 2004 a permis de repérer, parmi les 116 recensées, 8 cohortes consacrées à un suivi d'enfants (ou de femmes enceintes) pouvant inclure une dimension environnement. Le tableau 1 présente brièvement ces études, avec 4 autres relevant du même champ. Ces cohortes, qui couvrent différentes régions françaises, ont été lancées parfois depuis plusieurs années pour répondre à une grande variété de questions épidémiologiques plus ou moins spécifiques. Elles permettent ou permettront notamment :

- *des mesures des expositions et de leur impact*, portant sur les principaux contaminants et niveaux d'exposition en prénatal ou chez les enfants ; en particulier les particules fines, les COV, les polluants organiques persistants, certains pesticides semi-persistants, ou des substances soupçonnées d'une activité type perturbation endocrinienne...

- *d'étudier l'incidence des pathologies et les associations environnement-santé chez l'enfant* : asthme, allergies, infections, croissance et obésité, désordres métaboliques, développement psychomoteur, cognitif et comportemental...

Le projet de grande cohorte figurant au PNSE

Parmi les 45 « actions » prévues au PNSE, 12 ont été retenues prioritairement, dont l'action 26 : « Réaliser une étude épidémiologique enfants en lien avec l'étude américaine National Children's Study ». L'objectif est de répondre à trois questions :

- à quel âge et à quels niveaux les enfants sont-ils imprégnés par différents polluants ?
- à quels stades de la croissance voit-on apparaître des effets sur la santé ?
- y a-t-il une relation causale entre les expositions et les effets, et quelle en est la force ?

Il est envisagé de suivre 10 à 20 000 enfants, depuis la grossesse jusqu'à l'âge adulte, avec questionnaires, prélèvements biologiques, examens médicaux. Le modèle explicite est la NCS américaine, qui vise à suivre ainsi 100 000 enfants. On trouvera au tableau 2 un récapitulatif de diverses études de cohortes (enfants) réalisées ou en cours à l'étranger.

Les objectifs de ce projet recoupant en grande partie à la fois ceux des cohortes présentées plus haut et ceux du projet présenté ci-après, des discussions sont en cours pour proposer un programme intégré sur la thématique environnement-enfants-cohortes.

Projet pluridisciplinaire de cohorte nationale enfants (CNE)

Le projet CNE (préparé par H. Leridon, Inserm et Ined) a deux spécificités :

- il fait une place importante aux dimensions démographiques, sociales, économiques et éducatives, en plus des aspects santé,
- il postule un échantillon initial construit sur une base représentative de l'ensemble de la population (par exemple toutes les naissances survenant pendant une certaine période).

Les informations concernant le champ de la santé qui pourraient être collectés dans le cadre de la CNE peuvent être classées ainsi, par ordre de difficulté croissante de collecte :

- *Des données figurant dans les dossiers médicaux disponibles au moment de l'accouchement* : caractéristiques du nouveau-né (poids, taille, âge gestationnel, malformations repérées...), circonstances de l'accouchement (présentation, analgésie, Apgar...), ainsi que *des informations obtenues par interrogation de la mère à la clinique* : caractéristiques socio-démographiques (date de naissance, nombre de grossesses antérieures, niveau d'étude, situation professionnelle...), situation avant la grossesse (délai nécessaire pour concevoir, traitements éventuels contre l'infertilité), déroulement de la grossesse (nombre de consultations et d'examen échographiques, tests de dépistage, hospitalisations...), consommation de tabac et d'alcool...

- *Des informations recueillies par enquêteur auprès de la mère*, à des durées variables après l'accouchement : allaitement maternel, croissance de l'enfant (poids et taille), vaccinations, maladies, problèmes spécifiques de vue, audition ou dentition, consultations médicales, autres actes médicaux et paramédicaux, interruptions de scolarité...

- *Des carnets de suivis*, remplis par la femme pendant une période déterminée et permettant d'observer plus continûment : la consommation médicale, les consultations médicales, les divers incidents de santé de l'enfant ; les pratiques alimentaires ; certains modes de vie pouvant exposer à des risques spécifiques...

- *Des tests psychologiques et psychomoteurs*, supposant la participation de personnels spécialisés (au domicile de l'enfant ou ailleurs) ;

- Des mesures environnementales au moyen de *capteurs* déposés au domicile ;

- Des *examens médicaux* organisés spécifiquement, ou dans le cadre du suivi scolaire systématique ;
- Des *données récupérées de sources extérieures* (caisses d'assurance maladie ou Mutuelles, éducation nationale) ;

- Les résultats de *prélèvements biologiques*, pour étudier le génome, la présence de polluants spécifiques, ou les effets de ces polluants : sang du cordon, sang de la mère à la naissance, prélèvements ultérieurs éventuels chez l'enfant (et ses parents ?)...

Présenté dans le cadre de l'appel à propositions Cohortes de l'Inserm, le projet a reçu l'appui d'un ensemble de partenaires institutionnels (organismes de recherche, administrations, organismes sociaux). L'Insee l'appuie fortement.

Bibliographie

BOUVILLE J.F., 2005, *La malnutrition infantile en milieu urbain africain, Etude des étiologies relationnelles*. Paris, L'Harmattan, 352 p.

BYNNER J. (2002), « British Birth Cohorts Studies : Research prospects », Paper for the *Lasmas Meeting on Longitudinal Studies*, Paris, 7-8 jan. 2002

CERC (2004), *Les enfants pauvres en France*. Paris, La Documentation française, 151 p.

CORDIER S., LEDRANS M., LERIDON H., SALINES G. (2005), *Les effets de la l'environnement sur la santé de l'enfant. Propositions d'orientations stratégiques (document interne Inserm-InVS)*

DUMESNIL S. et LE FUR P. (2003), « Etat de santé des enfants pauvres et recours aux soins en ambulatoire », in Colloque Cerc-Cnaf-CGP-Drees, *Les enfants pauvres en France*, 21 mars.

RIZK (C) (2003), « Les enfants pauvres : quartier et qualité du cadre de vie », in Colloque Cerc-Cnaf-CGP-Drees, *Les enfants pauvres en France*, 21 mars.

STELIAROVA-FOUCHER E. , STILLER C., KAATSCH P., BERRINO P., COEBERGH JW., LACOUR B., PARKIN M. (2004), Geographical patterns and time trends of cancer incidence and survival among children and adolescents in Europe since 1970s (The ACCIS Project) : an epidemiological study, *Lancet* 2004, 364, 2097-2105

TUBIANA M.(2005), « Prévention des troubles de la santé mentale, Le jeune enfant ». Communication aux *Assises nationales de la santé de l'enfant et de l'adolescent*, 13 janvier 2005.

VAIVRE-DOURET Laurence et TURSZ Anne (1999), *Les troubles d'apprentissage chez l'enfant : Un problème de santé publique ?*, ADSP n° 26 mars 1999

VOLOVITCH (P) (2003), « Une couverture santé à visée universelle à l'égard des plus démunis », *Informations sociales*, n°108, CNAF.

Tableau 1

Cohortes concernant des enfants ou des femmes enceintes, avec dimension "environnement" possible

1 - COHORTES EXISTANTES signalées dans l'appel à propositions "Cohortes" de l'Inserm en 2003-04

Responsable	INTITULE	Laboratoire	Nb équipes impliquées	Coût total (K€)	Thème : 1	Thème : 2	Durée prévue (années)	Nb de sujets	Popul. générale :	Avec prélév. biolog.:	Avec composante génétique ?
BREART G. LARROQUE B.	Etude sur les petits âges gestationnels : Cohorte EPIPAGE	Inserm U-149	9	depuis 1995 : 2100	Reproduction	Pédiatrie	8 ans	3500	OUI	NON	N
DUCIMETIERE P.	Etude des Déterminants pré et post natals de la santé et du développement de l'ENfant (EDEN)	Inserm U-258	10	1153	Reproduction	Pédiatrie	5	3000	OUI	OUI	N
CACHERA MF.	Etude Longitudinale Parisienne Croissance et Alimentation (ELPCA)	Inserm U-557	3	200	Nutrition	Santé publique	20	100	OUI	OUI	N
DELLATOLAS G.	Dépistage précoce des troubles spécifiques du développement du langage oral et écrit chez l'enfant en milieu scolaire (MATERNELLE)	Inserm U-472	4	?	Troubles du langage	Pédiatrie	4 ans	3000	OUI (+/-)	NON	N
LEROY V.	Etude épidémiologique multicentrique du devenir neurologique et psychique des grands prématurés (< 33 SA) : une cohorte prospective en Aquitaine : AQUIPAGE	Inserm U-593	8	183	Neurologie	Pédiatrie	7	180/500	NON	NON	N

2 - Autres COHORTES EN COURS

BOUYER J.	Observatoire de la fertilité. Influence de facteurs environnementaux sur le délai nécessaire pour concevoir	Inserm U569			Fertilité	Environnement	6	3000	OUI	NON	N
CORDIER S.	Perturbateurs endocriniens, grossesse et développement de l'enfant (PELAGIE)	Inserm U435			Environnement	Pédiatrie	6	2000	NON	OUI	Oui
CORDIER S.	Exposition aux pesticides, grossesse et développement de l'enfant en Guadeloupe (TI MOUN)	Inserm U435			Environnement	Pédiatrie	7	1200 femmes, 300 enfants	NON	OUI	Oui
MOMAS I.	Facteurs environnementaux, santé respiratoire et statut atopique : une cohorte de nouveau-nés franciliens	Paris V			Environnement	Pédiatrie	6	3000	NON	OUI	N

3 - Propositions de COHORTES NOUVELLES signalées dans l'appel à propositions "Cohortes" de l'Inserm en 2003-04

Responsable	INTITULE	Laboratoire	Nb équipes impliquées	Coût total (K €)	Thème : 1	Thème : 2	Durée prévue (années)	Nb de sujets	Recrut. en popul. générale ?	Avec prélév. biolog.:	Avec composante génétique ?
JAQUET D.	Relation entre la croissance ante-et post-natale : influence sur le métabolisme glucidique de l'enfant (croissancemetab, RBM 03-22)	Inserm U-457	4 (même CHU)	887	pédiatrie	métabolisme	6	600	OUI	OUI	N
LERIDON H.	Cohorte Nationale d'Enfants (CNE)	Inserm U-569 / Ined / ++	2 (mais ++ en projet)	60000	santé publique	socio-economique	20 ans	20000	OUI	OUI	Oui ?
BOUYER J.	Follow-up AMP vigilance	Inserm U-569 et U-149	4	760	santé publique	AMP	5	4500	NON	OUI	N

Tableau 2
(HL - 09-2003)

Annexe

Données sur les "Grandes cohortes" (enfants et jeunes adultes) à l'étranger

	Année de début	1ère étape	2è étape	Suivis	Types de données	Thèmes abordés	Taux de suivi
Grande-Bretagne							
	NSHD (National Survey on Health and Development)						
	1946	16 000 naissances	5400 enfants	20 enquêtes de 1948 à 2000	Child tests, questionn.	Périnatalité, Santé , scolarité	53 ans: 56%
	NCDS (National Child Development Study)						
	1958	17 000 naissances	Tous suivis	Int. à 7, 11, 16, 23, 33, 42 ans (+ registres : école, Sec.soc...)	Child tests, questionn.	Périnatalité, Santé , scolarité Insertion profess. et sociale	20 ans: 83% 42 ans : 66%
	BCS70 (British Cohort Study, 1970)						
	1970	16 000 naissances	Tous suivis	Int. à 5, 10, 16, 26, 30 ans	Child tests, questionn.	Périnatalité, Santé, scolarité Insertion profess. et sociale	10 an: 93% 16 ans: 72% 30 ans: 70%
	MCS (Millenium Cohort Study)						
	2000-01	21 000 enf. 9 mois (y c. des enfants de BCS70)		Int. à 3, 6 ans...	Child tests, questionn. Dossiers médicaux	Santé, scolarité, précarité	
	ALSPAC (Avon Longitudinal Study of Pregnancy and Children)						
	1991	15 000 f. enceintes	Tous suivis	0, 7, 8 ans (puis annuel ?)	Examens, dossiers, prél. biol., ADN	Santé et environnement	4 ans: 77%
Etats-Unis							
	NLSY79 (National Longitudinal Survey of Youth, 1979)						
	1979	14-22 ans	Tous suivis	Annuel -> 1994, bisannuel	Quest.	Emploi , scolarité, santé, alcool+	
	NLSY79 : Children (National Longitudinal Survey of Youth: Children born to mothers, 1986+)						
	1986	Env. 5000 enfants nés des mères NLSY	Tous suivis	Bisannuel	Child tests Quest.	Périnatalité, scolarité, santé	
	NLSY97 (National Longitudinal Survey of Youth, 1997)						
	1997	9000 12-16 ans	Tous suivis	Annuel	Quest.	Emploi , scolarité, santé, alcool+	

	Add Health (National Longitudinal Study of Adolescent Health)					
	1994-95	90 000 11-18 ans	20 700	1996, 2001	Quest., interv. Parents, dossiers scolaires	Scolarité, sociabilité
	ECLS-K (Early Childhood Longitudinal Study : Kindergarten Class of 1988-99)					
	1998-99	22 000 enf. en maternelle		Annuel jusqu'en 2004	Quest.	Scolarité
	ECLS-B (Early Childhood Longitudinal Study : Birth cohort 2001)					
	2001	13 500 enf. 9 mois		9, 24, 48 mois + First grade	Quest.	Développement, scolar
	NCS (National Children's Study)					
	>2001	100 000 f. enceintes		Jusqu'à l'âge adulte		Santé et environnemer
Canada						
	ELNEJ-NLSC (Enquête longitudinale nationale sur les enfants et les jeunes)					
	1994	23 000 enf. 0-11 ans		Annuel jusqu'à l'âge adulte	Quest.	Développement, scolar
	ELDEQ (Enquête longitudinale sur le développement des enfants au Québec)					
	1997	2900 f. enceintes	2100 enfants	Annuel de 0 à 5 ans		Santé (périnatalité), ali développement, modes
Danemark						
	DNBC-BHMC (Danish National Birth Cohort : Better health for mother and child)					
	1996-2000	60 000 f. enceintes (objectif : 100 000)	Tous suivis	Grossesse (2), 6 et 18 mois (+ registres). --> 20 ans	Quest., dossiers méd. prél. biol., environ.	Santé (facteurs prénat: sociaux, environnemer
Norvège						
	MOBA (Norwegian Mother and Child Cohort Study) [Programme : BioBanks for Health - Biohealth Norway]					
	> 2000	100 000 f. enceintes 70 000 pères (25 000 en 2003)			Quest., dossiers méd. prél. biol., environ.	Santé (facteurs prénat: environnementaux)