



HAL
open science

[Epidemiological cohorts: methodological, ethical and practical aspects. Report of the Scientific Seminar of the ADELFF (Nancy, 12-4 January 2000)]

Marcel Goldberg, Françoise Clavel-Chapelon, Francis Derriennic, Hélène Jacqmin-Gadda, Béatrice Larroque, Laurence Meyer, Jean-Jacques Moulin, Margot Tirmarche

► **To cite this version:**

Marcel Goldberg, Françoise Clavel-Chapelon, Francis Derriennic, Hélène Jacqmin-Gadda, Béatrice Larroque, et al.. [Epidemiological cohorts: methodological, ethical and practical aspects. Report of the Scientific Seminar of the ADELFF (Nancy, 12-4 January 2000)]. *Epidemiology and Public Health* = *Revue d'Epidémiologie et de Santé Publique*, Elsevier Masson, 2000, 48 (5), pp.503-8. inserm-00175207

HAL Id: inserm-00175207

<https://www.hal.inserm.fr/inserm-00175207>

Submitted on 27 Sep 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Les cohortes épidémiologiques : aspects méthodologiques, éthiques et pratiques. Compte-rendu du Colloque scientifique de l'ADELF (Nancy, 12-14 janvier 2000)

Epidemiological cohorts: methodological, ethical and practical aspects

M. Goldberg¹, F. Clavel-Chapelon², F. Derriennic³, H. Jacqmin-Gadda⁴, B. Larroque⁵, L. Meyer⁶, J.J. Moulin⁷, M. Tirmarche⁸ (membres du Comité Scientifique du Colloque).

¹INSERM Unité 88 ; ²INSERM Unité 521 ; ³INSERM Unité 170 ; ⁴INSERM Unité 330 ; ⁵INSERM Unité 149 ; ⁶INSERM Unité 292 ; ⁷INRS ; ⁸IPSN.

L'intérêt des cohortes épidémiologiques

La cohorte est un moyen d'investigation qui fait partie de la panoplie de base de l'épidémiologie. Les cohortes sont largement utilisées depuis fort longtemps dans des domaines très divers, qu'il s'agisse de « cohortes historiques » (reconstituées rétrospectivement à partir d'une source de données existantes), ou de cohortes « contemporaines », les premières étant habituellement mises en place pour répondre rapidement à une question spécifique (rôle potentiel de l'exposition à certains facteurs de risque), alors que les secondes ont généralement des objectifs plus larges en termes de santé et de facteurs de risque, à plus long terme.

Les avantages principaux des cohortes sont la possibilité d'analyses longitudinales permettant de tenir compte au mieux de phénomènes liés au temps : séquence temporelle exposition-effet, effet génération, effet période. Elles sont aussi adaptées à l'étude simultanée de plusieurs problèmes de santé, et les données d'exposition sont recueillies avant la survenue des effets analysés.

Les cohortes historiques concernent presque uniquement des problèmes de santé se traduisant par un décès, car la mortalité fait depuis longtemps l'objet d'un enregistrement exhaustif et permanent à l'échelle de la population générale ; toutefois, l'existence de registres de maladies (cancer essentiellement) dont certains ont maintenant une ancienneté suffisante dans plusieurs départements français, permet la réalisation d'études de cohortes historiques basées sur l'analyse de cas incidents, et non plus seulement de décès. En France, leur usage a été extrêmement limité jusqu'à une période récente, du fait d'obstacles légaux concernant l'accès aux données de mortalité. Une première amélioration est venue de la loi 94-548 relative au traitement de données nominatives ayant pour fin la recherche dans le domaine de la santé et modifiant la loi 78-17 relative à l'informatique, aux fichiers et aux libertés. Mais si cette loi a levé l'interdiction d'accès aux données de décès nominatives, il subsistait de considérables difficultés pratiques pour réunir les informations démographiques permettant l'accès effectif à ces données ; la publication récente du Décret 98-37 du 16 janvier 1998 permet aujourd'hui l'accès au Répertoire National d'Identification des Personnes Physiques de l'INSEE (RNIPP) dans le cadre des recherches dans le domaine de la santé. On peut donc considérer qu'il est dorénavant possible, sous réserve d'un accord de la CNIL, d'utiliser sans restriction cet outil de base.

Le caractère prospectif du suivi des sujets et du recueil des données, qui caractérise les cohortes contemporaines permet de planifier le recueil de données d'exposition nombreuses et précises (y compris dans le cadre de banques de matériel biologique, où celui-ci est recueilli avant la survenue des pathologies étudiées), variables dans le temps, actualisées en fonction des questions qui peuvent se poser. Il est également possible de prendre en compte des problèmes de santé très divers, appréhendés en termes de morbidité, et non plus simplement de mortalité. La littérature épidémiologique internationale a rendu familières certaines grandes cohortes, comme celle de Framingham, mise en place il y a plus de 50 ans. En France, on a vu depuis quelques années se développer, à la suite de l'Étude Prospective Parisienne qui a longtemps été isolée de ce point de vue, plusieurs cohortes contemporaines aux objectifs divers.

Le Colloque de Nancy

Dans les dernières années les cohortes ont connu, dans des domaines très variés, un essor important en France. Cela a permis d'acquérir une expertise logistique et scientifique certaine, mais aussi de mieux percevoir les difficultés inhérentes à ces enquêtes, d'ordre pratique (du fait notamment de leur durée puisque certaines peuvent s'étendre sur une voire plusieurs décennies, de la multiplicité des sources de données et du volume de celles-ci, ainsi que de la nécessité d'employer du personnel stable et formé), méthodologique (épidémiologique et/ou statistique, par exemple portant sur l'analyse des données répétées, ou sur la prise en compte des données manquantes) ou éthique, du fait par exemple que l'analyse des données peut prendre place de nombreuses années après leur recueil pour étudier des hypothèses qui n'avaient parfois pas été proposées à l'origine.

L'objectif du Colloque était donc de faire émerger le savoir-faire sur ces différents aspects grâce à l'expérience acquise par les équipes dans ce domaine. Il s'agissait aussi de faire se rencontrer des personnes qui partagent, sans parfois le savoir, des méthodes et des problèmes voisins.

Environ 220 personnes ont assisté aux communications orales et affichées, qui étaient structurées selon plusieurs thèmes : aspects épidémiologiques et logistiques (recrutement — inclusion, suivi des sujets, gestion des données, causes de décès), banques de matériel biologique, aspects statistiques, aspects déontologiques et éthiques.

Les communications ont concerné des cohortes touchant à tous les domaines de la santé (cancer, reproduction humaine, vieillissement, maladies infectieuses, etc.), et prenant place dans des populations variées (professionnelles, personnes âgées, nouveau-nés, etc.) quel que soit le pays (notamment les pays en voie de développement).

Recrutement-inclusion

Plusieurs études de cohorte ont été présentées. Dans certaines présentations les différentes étapes du recrutement ont été décrites de façon précise avec les difficultés rencontrées et les manières de les résoudre. Les aspects pratiques, rarement présentés en détails dans les congrès en général et encore moins dans les articles, ont été ainsi mis en évidence.

Études en population

Les études en population sont nombreuses. Leur justification réside dans leur validité externe présumée c'est-à-dire l'applicabilité des résultats de cette enquête à une population source, avec la possibilité de donner des estimations de la distribution et de la prévalence dans la population source des maladies et des facteurs de risque étudiés. Au cours des présentations, ont été discutés particulièrement les effets de sélection des populations et les implications de cette sélection sur les résultats. Le recrutement de cohortes de personnes âgées a été illustré par les présentations des cohortes POLA et EVA. Dans cette dernière, où les personnes âgées ont été recrutées à partir de listes électorales mais sans aucune relance, on observe un très bon taux de suivi. Par contre, les sujets inclus avaient une éducation, une santé et un niveau cognitif meilleurs que la population générale.

Effets associés au volontariat des sujets

Les effets du volontariat dans l'inclusion ont été discutés au cours de la conférence introductive de M. Goldberg, avec notamment l'exemple de la cohorte GAZEL en milieu professionnel. L'avantage du volontariat est une meilleure participation au suivi et une meilleure qualité des données recueillies mais avec une sélection importante des sujets sur le niveau d'éducation, l'état matrimonial et parfois la santé. Dans le cas d'enquêtes à visée étiologique, les inconvénients sont moins importants, — recherche d'associations entre facteurs de risque et maladie, — que dans l'étude de la fréquence.

Autres sources de recrutement

D'autres sources de recrutement ont été présentées, comme le recrutement à partir des acteurs du système de soins et leurs modalités pratiques discutées : l'étude VEINES à partir de patients consultant chez des généralistes, l'utilisation des fichiers d'assurance maladie pour la constitution d'une cohorte pour le dépistage du cancer du sein, les cohortes PRIME ou DESIR, organisées à partir de centres d'exams de

santé, la cohorte PRIMO à partir des services hospitaliers et la cohorte EPIPAGE à partir des maternités et services de réanimation néonatale. Un registre pré-existant peut constituer la source d'une nouvelle cohorte, exemple : une partie du recrutement de l'étude ApoEurope s'est fait à partir du registre MONICA. Les problèmes particuliers des cohortes historiques en milieu professionnel ont été illustrés par la cohorte de la COGEMA.

Suivi des sujets

Le suivi des cohortes est conditionné par les caractéristiques propres de la cohorte étudiée.

Âge des sujets suivis

Le suivi des grands prématurés pose des problèmes en raison de la mobilité des familles (perte d'adresse), du caractère parfois anxiogène des questions pour les mères inquiètes pour le devenir de leur enfant (cohorte EPIPAGE).

Le suivi des personnes âgées est soumis à plusieurs contraintes. S'il s'agit d'une étude à suivi actif, par exemple, contact régulier des personnes de la cohorte, l'augmentation de la mortalité de la morbidité aboutissent à une diminution du suivi des sujets âgés, (cohorte POLA). Un correctif peut être le déplacement à domicile des enquêteurs ou la mise à disposition d'un moyen de transport.

Structure dans laquelle la population est suivie

Le suivi en milieu professionnel est facilité dans des entreprises où la population est relativement captive, comme à EDF-GDF. En revanche, lorsque la mobilité des employés entre différents postes ne peut pas être prise en compte de manière suffisamment précise, l'analyse est limitée à un descriptif de la population en milieu professionnel, en précisant les limites de cette surveillance par rapport à une étude de cohorte.

Dans des circonstances particulières (patients séropositifs pour le VIH), il est possible d'utiliser les étapes inévitables (consultations hospitalières régulières programmées) pour assurer le suivi (cohorte PRIMO).

La possibilité de s'appuyer sur des réseaux existants est un atout, mais il peut être contrebalancé par un désinvestissement des acteurs de ce réseau (par exemple, médecins bénévoles) si l'objectif de l'étude et son intérêt ne sont pas bien perçus, d'où la nécessité d'un retour d'information (cohorte DESIR)

Conditions du suivi

D'une manière générale, le nombre et la variété des procédures de suivi mises en œuvre (courrier, nombre de contacts téléphoniques, contact avec le médecin traitant...) (cohorte PRIME) influencent son efficacité.

Dans l'étude de populations bien particulières, telles que des toxicomanes, la distribution de nourriture, de tickets de transport en commun ou d'un peu d'argent (cohorte d'utilisateurs de drogues injectables aux USA) a été utilisée pour motiver ceux-ci. Pour les personnes âgées, il faut faciliter matériellement le suivi : payer le transport, prévoir un taxi, fidéliser les sujets en gardant la même équipe tout au long du suivi... (cohorte POLA).

Accès aux données individuelles d'organismes de protection sociale

Une reconstitution rétrospective de cohorte en population générale nécessite le recours à une méthode de suivi géographique des individus, une fois qu'ils ont été identifiés comme éligibles. Le fichier du registre national des individus affiliés à l'assurance maladie (RNIAM) devrait théoriquement permettre, à terme, de localiser les individus à partir d'identifiants simples (cohorte des enfants de La Hague).

De même, pour suivre les grands prématurés, la possibilité d'accéder aux données des Caisses d'allocations familiales ou de la Caisse nationale d'assurance maladie, afin de retrouver les parents et effectuer des relances, constituerait un avantage.

Gestion des données

Les nombreux exemples d'études de cohorte présentés au colloque, convergent sur quelques caractéristiques qui conditionnent la gestion des données : nombre important de sujets ; nombreux types de données provenant en général de plusieurs sources avec des modalités de recueil spécifiques à chaque source ; recueils de données successifs dans le temps sur les mêmes sujets ; fichiers informatiques constituant des bases de données complexes ; multiplicité d'études spécifiques sur la même base de données

; résultats de différentes natures concernant la production de connaissances, mais aussi la diffusion de ces connaissances.

L'organisation de la gestion des données devient donc un chapitre essentiel du protocole d'étude où tout doit être décrit avec précision, tant en ce qui concerne les aspects techniques et logistiques que les aspects « politiques » : qui a accès aux données ? comment s'organise la mise en commun des différentes études portant sur la même base de données ?

Sur les aspects techniques, comme dans toute enquête épidémiologique, la gestion des données collectées dans les études de cohorte doit passer par différents filtres : respect de la confidentialité des données personnelles ; protection des fichiers informatiques ; procédures de correction et de validation des données enregistrées. Les aspects plus « politiques » de la gestion des données semblent passer par la notion d'analyse décentralisée des données.

C'est surtout ce dernier point, et les problèmes liés à la confidentialité des données personnelles qui ont fait l'objet de diverses communications montrant une véritable progression du « savoir-faire ».

Respect de la confidentialité des données personnelles

La difficulté consiste à mettre en place un système garantissant l'anonymat des données (tout en permettant un accès contrôlé pour les sujets enquêtés qui en font la demande) et l'appariement des données en provenance de différentes sources (questionnaires, analyses biologiques, informations issues d'autres fichiers administratifs,...) à différents moments. Une référence particulièrement intéressante pour les solutions retenues est celle de la cohorte GAZEL qui inclut 20 000 agents EDF-GDF, les données étant collectées à partir de six sources différentes, dont le registre national des causes de décès.

Partage des données

Sur ce point, la difficulté consiste à faire en sorte qu'une même base de données, constituée à titre de recherche, et appartenant au collectif qui mène la recherche, puisse être utilisée d'une façon coordonnée par différentes équipes. Des différentes expériences présentées (cohortes PAQUID, GAZEL, E3N, ESTEV, EPIPAGE, et bien d'autres), il ressort que le collectif à l'origine de la recherche fonctionne comme un véritable laboratoire de recherche « hors les murs ». La gestion des fichiers et des données qui s'y rapportent est centralisée avec des règles de fonctionnement démocratiques et écrites, tandis que les études spécifiques sont décentralisées et coordonnées au travers de réseaux de recherche (formels ou informels). L'exemple le plus détaillé présenté a été celui de la mise en place du réseau de recherche sur les relations entre l'âge, le travail et la santé qui implique des acteurs multiples dans des horizons professionnels très différents (en particulier des médecins du travail, des médecins inspecteurs du travail, des chercheurs et des universitaires), dans le cadre de l'exploitation des données de la cohorte ESTEV. Le ciment du réseau a consisté à marier des actions de recherche avec des actions de valorisation et de formation se nourrissant de la recherche. En retour, ces dernières activités ont permis aux équipes de durer dans le temps et de discuter d'hypothèses non initialement envisagées.

Statut vital et causes de décès

Une session a été organisée sur les problèmes liés à la détermination du statut vital et la recherche des causes de décès.

Dans une cohorte historique de 101 585 sujets de France Télécom, les décès ont été identifiés dans le fichier 7^{bis} de l'INSEE, à l'aide d'une procédure d'appariement basée sur nom, prénoms, sexe, date et lieu de naissance, mise au point par le Service Épidémiologie en Entreprises de l'INRS. Un total de 6 126 décès ont ainsi été repérés. Les causes de décès ont été retrouvées par le SC8 (certificat médical de décès) pour 6 001 sujets (98 %).

Une autre communication a rapporté les résultats de la détermination du statut vital dans une cohorte historique de 27 857 sujets d'une usine sidérurgique. La comparaison des résultats obtenus par la recherche dans le fichier 7^{bis} de l'INSEE (même procédure que précédemment) et par courrier aux mairies des lieux de naissance a montré que parmi les 770 décès identifiés, la recherche auprès des mairies permettait d'en retrouver 725 (94,2 %), alors que 750 (97,4 %) étaient donnés par le fichier 7^{bis}.

La comparaison des causes de décès données par deux sources, la médecine du travail et les certificats de décès (fichier du SC8), a été réalisée dans une cohorte historique de 1 785 sujets de la société COGEMA. Parmi les 612 décès enregistrés, 331 causes ont été données par les deux sources : seulement 114 (34 %) avaient un diagnostic identique. En cas de divergence, la règle de décision consiste à adopter la cause donnée par le SC8 lorsqu'elle est informative.

Une présentation détaillée de la procédure mise en place en application du Décret 98-37 du 16 janvier 1998 a été faite. Cette procédure permet de connaître le statut vital des sujets d'une cohorte et d'avoir accès aux causes de décès individuelles ; elle associe l'INSERM (IFR 69 et SC8) et l'INSEE, sur la base du RNIPP. Rappelons que cette procédure a déjà été décrite dans le Bulletin de l'ADELF (n° 16, Décembre 1999).

Aspects statistiques

L'analyse des données de cohorte soulève des problèmes statistiques complexes qui ont fait l'objet de deux sessions. Lors de la conférence introductive, D. Commenges a présenté les deux types d'analyse caractéristiques des études de cohorte : l'analyse des données longitudinales qui a pour objectif l'estimation de l'évolution de l'espérance d'une variable (souvent quantitative) mesurée de manière répétée au cours du suivi et l'analyse de données de survie où l'objectif est d'estimer le délai jusqu'à la survenue d'un événement (qualitatif).

Survie

Les problèmes spécifiques de l'analyse de données de survie ont été largement discutés. Les thèmes des données incomplètes (censure), des données sélectionnées (troncature), ou de l'estimation de la fonction de risque ont par exemple été abordés. Les méthodes pour vérifier la proportionnalité des risques dans le modèle de Cox ont été décrites et, lorsque cette hypothèse n'est pas vérifiée, des alternatives à ce modèle ont été présentées notamment dans le cadre de cohorte de reculs différents. Certains intervenants ont insisté sur la modélisation (paramétrique et non paramétrique) de l'effet des variables explicatives (dépendantes ou non du temps) lorsqu'il existe un délai entre l'exposition et l'augmentation du risque. Une communication a abordé les méthodes d'analyse d'événements multiples qui ont fait récemment l'objet de développements statistiques.

Données longitudinales

Le modèle linéaire à effets mixtes, modèle de choix pour l'analyse des données longitudinales gaussiennes, a été présenté en détail et un exemple d'utilisation des prédictions individuelles issues de ce modèle a été donné. Les méthodes de traitement des données manquantes aléatoires dans les deux grandes catégories d'approches des données longitudinales, les modèles marginaux et les modèles à effets aléatoires, ont été comparées.

Données incomplètes

Ce problème des données incomplètes dans les études de cohorte a constitué le troisième thème majeur de ces sessions. Il a été montré que des méthodes simples de recoupement de données (cold-deck et méthode déductive) permettent de récupérer de nombreuses informations manquantes et devraient être utilisées avant les méthodes d'imputation multiple. Des méthodes d'ajustement, lorsque l'information sur un facteur de confusion majeur est manquante, ont également été présentées.

Contre-appariement

Les plans d'études dans les enquêtes de cohorte, ont été abordés lors d'une communication sur la comparaison de l'efficacité du « contre-appariement » et des études cas-témoins, nichées dans la cohorte, pour l'étude des interactions gène-environnement.

Malgré la complexité et l'actualité des problèmes statistiques abordés, la plupart des orateurs ont insisté sur des méthodes d'ores et déjà accessibles soit par les logiciels du commerce soit par des programmes diffusés gratuitement.

Les banques de matériel biologiques

Nombre de cohortes incluent désormais une banque de matériel biologique dans leur protocole, offrant des échantillons biologiques qui peuvent être analysés, en relation avec les autres données individuelles collectées, parfois longtemps après leur recueil. Trois banques, issues du sang périphérique de volontaires, ont été présentées. Le **tableau I** compare les différentes approches proposées.

Tableau I. **Principales caractéristiques de trois biothèques associées à des cohortes.**

	Cohorte E3N	Cohorte ApoE	Cohortes SEROCO/ HEMOCO
But	dosage de marqueurs de l'alimentation hormonaux recherche génétique polymorphisme	– détermination du polymorphisme de l'ApoE par géno et phénotypage – maladie d'Alzheimer génotype ApoE ; divers dosages sériques	– étude de l'histoire naturelle du VIH – étude de la réponse aux traitements
Fréquence des prélèvements	une fois	une fois	Sérum à l'inclusion puis tous les 6 mois Peripheral Blood Mononuclear Cells à l'inclusion et tous les 18 mois
Conservation	en azote liquide	en azote liquide	en azote liquide
Origine	cohorte déjà existante	cohorte déjà existante + deux autres spécifiques	cohorte prospective multi-centrique
Matériel congelé	Plasma, sérum, globules rouges, lymphocytes	Sérum et ADN	Sérum Peripheral Blood Mononuclear Cells

Cohorte E3N

Ont été exposées les contraintes, liées au protocole européen EPIC dont elle fait partie. D'autres contraintes propres à la situation française s'y ajoutent : existence de la cohorte avant la constitution de la banque biologique ; caractère « sans bénéfice individuel direct » impliquant pour les lieux de prélèvement d'être titulaires d'une autorisation délivrée par le ministère de la santé ; répartition des sujets sur tout le territoire national. La solution adoptée fut d'ouvrir des centres, dans les départements à population E3N la plus importante, et d'acheminer quotidiennement les prélèvements vers un laboratoire central qui les traitait et les congelait. Ceci a permis de réduire la distance domicile-centre d'accueil, important facteur limitatif. À l'issue du recueil, qui s'est déroulé entre 1996 et 1998, près de 25 000 prélèvements, (38 % des femmes) ont été recueillis dans 35 centres d'accueil. Le profil des femmes prélevées est quelque peu différent de celui des femmes n'ayant pas accepté ce prélèvement. La nécessité d'un protocole détaillé et rigoureux a été mise en avant.

Cohortes SEROCO/HEMOCO

Un autre aspect des banques biologiques a été abordé : la gestion des cryoconservations. Il est apparu la nécessité d'une gestion à long terme et le besoin continu de moyens humains, logistiques et financiers. Afin de gérer cette base de données constituée de 71 000 aliquots de sérums et de 16 000 aliquots de cellules (PBMC), une application informatique spécifique permet d'enregistrer la quantité prélevée, le devenir de chaque tube et la quantité restante non utilisée de certains tubes décongelés et re-stockés. À ce jour, 7 % du total des aliquots de sérums et 10 % des aliquots de lymphocytes ont été décongelés (environ 30 études). La décision du processus de décongélation doit systématiquement être argumentée par le responsable de l'étude proposée et évaluée par le conseil scientifique des cohortes. Les règles de décisions pour les décongelations ont été constituées progressivement grâce à l'expérience acquise. En particulier les objectifs des études doivent toujours être compatibles avec les objectifs des cohortes. La biothèque ne doit jamais

servir uniquement à valider une technique, et les tubes du bilan d'inclusion ne seront décongelés qu'après de longues discussions.

Les biothèques du Centre de Médecine Préventive de Nancy

Ce centre a exposé le fruit de son expérience sur les aspects, logistiques entre autres, de la constitution d'une banque de matériel biologique. Une banque de matériel biologique associe deux types d'informations : biologiques via les échantillons sanguins ou d'autres liquides biologiques, et biocliniques recueillies au cours d'examen. Sa mise en œuvre impose la maîtrise d'une chaîne comportant plusieurs étapes : pré-analytique souvent négligée mais essentielle (tous les stades depuis le prélèvement sanguin jusqu'au stockage) ; stockage, car les conditions de conservation diffèrent suivant la nature des spécimens ; assurance qualité, impliquant la mise en place de procédures spécifiques ; gestion des stocks. L'aspect éthique est crucial ; il concerne la confidentialité, l'accès aux informations et aux échantillons biologiques. Il convient évidemment de respecter le droit des individus en accord avec la législation en vigueur. Pour mieux assurer la sécurité de la biothèque, il est fortement recommandé de prévoir deux lieux de stockage distincts.

Aspects déontologiques et éthiques

Le rôle de la CNIL

La dernière session de ces Journées, portant sur les aspects déontologiques et éthiques des études de cohorte, s'est ouverte par la communication invitée de Sophie Vuillet-Tavernier, responsable du secteur Santé et Social à la CNIL. La CNIL est de plus en plus fréquemment saisie pour des demandes en aval de la saisine, et elle examine les projets de recherche sous trois angles différents : la pertinence des données informatisées, le respect des droits des personnes, et les aspects de sécurité et de confidentialité. À ce propos, elle a rappelé que s'il existe quelque part une correspondance entre un numéro d'identification et un nom, alors les données sont considérées comme indirectement nominatives, et l'étude doit faire l'objet d'une déclaration à la CNIL. Les recherches doivent être menées dans le respect des droits des personnes, avec une information individuelle. De rares dérogations existent, lorsque les personnes sont décédées, perdues de vue, ou pour des raisons psychologiques. La spécificité des cohortes dans ce contexte est que les recherches s'étalent souvent sur de longues années : toutes les hypothèses de recherche ne peuvent ainsi être prévues au départ, et les problématiques initiales se modifient avec le temps. Le formulaire de consentement initial doit comprendre une clause informant les personnes que les données, notamment les prélèvements congelés, pourront être utilisées à des recherches non initialement prévues dans le protocole. La nécessité de demander un nouveau consentement des patients pour de nouvelles études au sein de la cohorte existante dépend de leur plus ou moins grande adéquation aux objectifs de départ de la cohorte. L'avis de la CNIL peut toujours être sollicité dans de tels cas.

Confidentialité

L'ensemble des textes qui visent à protéger les libertés individuelles à l'égard du traitement automatisé des données médicales nominatives ont pour corollaire le recours de plus en plus fréquent à des procédures de confidentialité strictes, et à des techniques cryptographiques, dont certaines ont été présentées. Les solutions retenues pour assurer la confidentialité de la base de données de la cohorte GAZEL ont entraîné des circuits de collecte et de gestion des données complexes. Le principe général repose sur la séparation physique, logistique, et institutionnelle, systématique, des données identifiantes et des données épidémiologiques ; en particulier, les fichiers d'identification des volontaires et les données contenues dans la base sont gérés dans des lieux physiquement différents sans connexion ; la base des données anonymisées est sur un serveur uniquement accessible aux personnes habilitées, qui n'est relié à aucun autre réseau.

Pays moins développés

Les aspects déontologiques et éthiques des recherches menées dans les pays moins développés ont été abordés de façon spécifique. La conduite de recherches en santé peut être problématique dans un contexte de surmortalité, surtout quand le système de santé est défaillant. Des procédures de conduite à tenir par les équipes d'enquête sont en général élaborées *a priori*, en tenant compte de la demande de soins des personnes malades et de la gravité des situations cliniques individuelles. Le paiement des soins nécessaires est difficile à codifier, les décisions étant souvent prises au cas par cas. Par ailleurs, la nécessité d'un

consentement éclairé, *a fortiori* écrit, doit se discuter en raison des conditions offertes dans un contexte socioculturel souvent démuné. Des propositions de règles ont été avancées : question de recherche opportune et méthodologie adaptée pour le pays hôte, investissement et renforcement du système de soins existant, aval d'une autorité éthique du pays hôte, rôle important des agences de recherche nationales et internationales. Les problèmes éthiques ne doivent pas être considérés comme un frein à la recherche, l'inacceptable sur le plan éthique pouvant être la non-recherche.

Accès aux données et partage des responsabilités

L'accès aux données et le partage des responsabilités au sein des cohortes ESTEV et GAZEL ont été exposés sous l'angle des aspects déontologiques. La production de connaissances est un processus lent et parfois réducteur, face aux attentes suscitées par les moyens mis en œuvre. Les relations entre équipes de recherche et participants et/ou représentants de la cohorte, et entre équipes de recherche elles-mêmes, doivent être gérées. La mise en place d'un Comité Scientifique d'une part, et, d'autre part, l'existence des Recommandations de Déontologie et de Bonnes Pratiques en Épidémiologie, permettent de régler la plupart des difficultés éventuelles. Enfin, bien sûr, des retours d'information systématiques aux participants et/ou leurs représentants font partie des règles déontologiques de base de tout épidémiologiste, et semblent devenus des procédures usuelles dans la plupart des cohortes présentées à ces Journées.

Les suites du Colloque : la création d'un « Club Cohortes » au sein de l'ADELF

Lors du Colloque, un questionnaire a été distribué aux participants pour évaluer leur intérêt dans la création d'un « Club » centré sur les divers aspects touchant aux études de cohortes. Sur les 171 participants, 57 ont rempli un questionnaire, 56 personnes étant intéressées par la création d'un tel Club. Un groupe d'animation s'est donc créé, composé actuellement de Françoise Clavel (U521), Béatrice Larroque (U149), Laurence Meyer (U292), Jean-Jacques Moulin (INRS), Isabelle Stucker (U170) et Marie Zins (U88) ; ce groupe peut s'élargir à d'autres personnes intéressées. Lors de sa réunion du 7 mars 2000, le Conseil d'Administration de l'ADELF a entériné la création du Club.

Son rôle sera d'organiser des réunions semestrielles ou annuelles d'une journée, centrées sur des thèmes précis. Les premiers thèmes retenus sont :

- Banques de données biologiques/Législation sur les cohortes ;
- Aspects statistiques : Régression de Poisson et analyse de données répétées ;
- Suivi des sujets inclus via des bases de données existantes telles RNIAM, PMSI, RNIPP.

Suivant les thèmes, l'organisation et le public peuvent différer. Un ou deux animateurs présenteront des exposés introductifs ou de clarification, accompagnés de quelques personnes qui présenteront brièvement (un ou deux transparents) une question précise à développer.

Pour faciliter l'organisation pratique, chaque réunion de Club, pourrait se tenir dans le cadre d'une organisation existante (IFR, Unité INSERM, INRS, etc.) ou d'une manifestation concomitante. Ainsi, la première journée portant sur la législation et les banques biologiques se tiendra la veille des prochaines Journées de l'ADELF à Lyon, le 11 octobre 2000.

L'objectif du Colloque était de partager des expériences et un savoir-faire dans un domaine des études épidémiologiques en plein développement, mais particulièrement complexe et lourd à mettre en œuvre. On peut considérer qu'une première étape a été réalisée, mais cet objectif ne prend tout son sens que s'il peut être poursuivi au-delà d'une manifestation unique. La création du Club Cohortes vise à prolonger cette expérience très enrichissante pour ses participants.

Remerciements : Catherine GUIBOUT pour sa contribution sur les banques de matériel biologique