

Sources et mesures

Jean-Paul SARDON

Observatoire démographique européen - Institut national d'études démographiques (ODE-INED)

L'ambition de cette sixième séance était, comme il est écrit dans le programme :

- de traiter, « sous l'angle démographique, des méthodes de mesure de l'état de santé, de l'incidence des maladies et de leur impact sur la mortalité » ;
- de s'intéresser « aux indices spécifiques et de synthèse, mais aussi aux principes autorisant leur utilisation non biaisée, et aux situations permettant de s'affranchir de ces principes » ;
- de s'attacher à « lier ces méthodes d'analyse et de mesure aux outils de collecte des données existant ou susceptibles d'être imaginés dans ce but ».

Il était également précisé dans le programme que, « dans cet esprit, les communications décrivant les sources existantes et les exploitations auxquelles elles peuvent aboutir, seront les bienvenues ».

La présentation qui était ainsi faite insistait, en dépit du titre de la séance, plus sur le versant *mesures* que sur le versant *sources*.

Les 14 communications qui vont nous être présentées, vont-elles dans le sens souhaité par les organisateurs du colloque ?

Pour répondre nous allons examiner, successivement, chacun des deux piliers sur lesquels repose cette séance, même s'il est un peu artificiel de les séparer, dans la mesure où, le plus souvent, ils sont étroitement imbriqués :

- les sources
- les mesures

En nous posant, pour chacun, les mêmes questions.

En effet, le premier problème qui se pose au spécialiste de la santé, démographe, épidémiologiste ou économiste est triple :

- **Quelles sont les sources disponibles ?**
- **Quelles sont les limitations de chacune d'elles ?**
- **Comment peut-on pallier leurs défauts et améliorer leur utilisation ?**

Nous nous trouvons ainsi dans une thématique qui se rapproche de celle qui avait été développée dans la séance 3, consacrée à « L'utilisation des sources disponibles » du XII^{ème} colloque national de démographie qui s'est tenu en 2002 à Amiens, bien qu'ici le champ d'application soit spécifique : la santé.

Ensuite, pour analyser les données contenues dans les sources dont la critique a été précédemment faite, se pose la question de la mesure la plus pertinente. Ce qui amène à un autre groupe de questions symétrique à celui envisagé pour les sources :

- **Quelles est l'éventail des mesures possibles ?**
- **Quelles sont les limitations de chacune d'elles ?**
- **Comment peut-on pallier leurs défauts et améliorer leur utilisation ?**

1. Les sources

Sans chercher à dresser un inventaire des sources disponibles pour traiter d'un sujet aussi vaste que « Démographie et santé », et en l'absence d'un texte qui se serait attaché à les présenter, nous allons nous limiter à quelques-unes, essentiellement celles qui ont été utilisées par les auteurs de la séance, car on peut en avoir une idée un peu plus précise après avoir pris connaissance de tous les textes.

Le seul inventaire est celui entrepris par Alain Ayerbe et Philippe Cordazzo, mais il ne traite que de la base de données d'Eurostat.

1.1. Sources disponibles

Chacune des cinq grandes catégories dans lesquelles on peut les classer :

1. les enregistrements permanents, comme l'état civil,
 2. les fichiers de gestion,
 3. les registres,
 4. les enquêtes, rétrospectives ou continues, sur échantillon,
 5. le recensement,
- fait ici l'objet d'au moins une communication.

1.1.1. Les enregistrements permanents

L'enregistrement administratif permanent par excellence dans le domaine de la santé, si je puis dire, c'est celui des décès dans l'état civil et son complément indispensable le fichier des causes de décès individuelles géré par le Centre d'épidémiologie sur les causes médicales de décès (CépiDc) de l'Inserm. Ces données sont ici utilisées par de nombreux auteurs, tous ceux qui traitent de la mortalité quelle que soit la façon dont elle est abordée.

Nadine Scandella, de l'Observatoire régional de santé d'Aquitaine, et Christophe Bergouignan, de l'Université Montesquieu-Bordeaux IV, utilisent de leur côté les informations collectées par un système qui fut pilote, permettant, à l'échelle de quelques régions françaises, d'enrichir les données collectées dans le cadre de la déclaration obligatoire du Sida, dans leur texte intitulé « L'enregistrement des sérologies positives au VIH en Aquitaine entre 1989 et 1998. Analyse des apports d'un système pilote pour appréhender la dynamique régionale de l'épidémie. ».

1.1.2. Les fichiers de gestion

Les fichiers de gestion, qu'ils soient administratifs ou médicaux, forment une catégorie très fournie, et sans doute la plus riche et la plus prometteuse. En effet, nombre de fichiers avec des données individuelles existent dans le domaine de la santé, que ce soit dans le secteur public ou dans le secteur privé. Même si tous ces fichiers ne sont pas spécialement organisés pour permettre la réalisation d'analyses, il est souvent possible d'en tirer parti. On peut par exemple citer les registres d'activité hospitalière et les répertoires professionnels.

C'est ainsi que nous seront présentés divers fichiers, parfois de taille impressionnante comme celui des Centres d'examen de santé qu'examinent René Gueguen, Caroline Dupré, Virginie Dauphinot et Florence Naudin, du CETAF, « Géographie de la santé dans les centres d'examens de santé ». En effet, la base de données couvre un peu plus de 6 millions d'examens assurés dans les centres d'examen de santé, même si les auteurs se sont particulièrement intéressés à la population examinée en 2005. Il s'agit là d'une source de premier ordre puisque au-delà de leurs nombres les examens comportent près de 300 paramètres, issus de questionnaires, tests ou examens biologiques et cliniques.

Une autre source majeure est en cours de constitution, c'est le Dossier médical personnel qui devrait être mis en œuvre au 1^{er} juillet prochain et dont vont nous parler Catherine Quantin, du CHU de Dijon, Benoît Riandey, de l'Ined, et Olivier Cohen, de HC-forum, dans leur communication intitulée : « Traitement épidémiologique et démographique du Dossier médical personnel. L'identifiant santé ». Ce fichier aura l'immense privilège de l'exhaustivité. Malheureusement, les auteurs nous diront que le sort de ce fichier n'est pas encore scellé, en particulier pour définir qui pourra accéder à quoi.

Cette source s'ajoutera, en le complétant largement sur les pathologies, au système d'information de l'assurance maladie, le SNIIR-AM (Système National d'Information Inter-Régimes de l'Assurance Maladie), qui est un panel (1/100) de bénéficiaires de l'assurance maladie et n'enregistre que des consommations. Ce panel est en fait une rénovation de l'échantillon permanent des assurés sociaux (EPAS, 1/600) et est un échantillon tiré du Répertoire National Inter-régimes des bénéficiaires de l'assurance maladie (RNIAM)¹.

Pour le cancer, il existe un dossier spécifique le Dossier Communiquant de Cancérologie (DCC), né du plan national cancer.

À côté de ces fichiers nationaux, il ne faut pas oublier la multiplicité des fichiers hospitaliers, comme celui de l'Institut Gustave Roussy qu'utilisent Isabelle Fournel, M. Schwarzing, E. Benzenine, C. Binquet, Benoît Riandey et Catherine Quantin, dans leur communication « Déterminer le statut vital des patients domiciliés en France métropolitaine et dans les départements d'Outre Mer en chaînant données hospitalières et données de mortalité de l'INSEE anonymisées ».

Dans le colloque d'Amiens, à propos des dossiers médicaux collectés dans les archives hospitalières, Laurence Tain, de l'Université Lumière-Lyon2, s'était attachée à un aspect tout à fait particulier, important et antérieur à tous les autres, la manière dont l'information se constitue. Elle met, ainsi, en évidence la nécessité pour l'utilisateur de maîtriser ce processus afin de pouvoir regrouper une information dispersée, dans le cas qu'elle étudie, dans plusieurs services et par conséquent dans plusieurs lieux d'archivage. Ainsi le démographe peut-il être conduit à créer lui-même le fichier par fusion de sous-ensembles séparés, sans pour cela intervenir sur le contenu. Sans ce travail préalable de compréhension de l'influence de la pratique médicale sur la production statistique l'analyse qu'il opère pourrait le conduire à des conclusions biaisées par les diverses sélections que pourrait opérer cette pratique médicale.

1.1.3. Les registres

Nous venons d'évoquer les registres, à partir du Répertoire National Interrégimes des bénéficiaires de l'assurance maladie (RNIAM). Nous pourrions également parler du Répertoire National d'Identification des Personnes Physiques (RNIPP), évoqué dans la communication d'Isabelle Fournel, qui permet de préciser si une personne est en vie ou décédée. Le registre de population est certainement le système d'observation le plus achevé. C'est, en particulier, un instrument très utile à l'observation des migrations, puisque lorsqu'un individu quitte une commune, il est tenu non seulement de signaler son départ et d'indiquer sa commune de destination, mais également de déclarer son installation dans la commune d'arrivée. La supériorité d'un tel système ne doit toutefois pas faire oublier ses limites. En Belgique, par exemple, où les registres de population fonctionnent depuis de nombreuses années, il n'y a pas de concordance absolue entre la statistique des départs et celle des arrivées, pas plus qu'il n'y a d'égalité entre les nombres d'événements enregistrés à l'état civil et inscrits dans le registre.

¹ Pour plus de détail voir : Riandey B., (2004), « Le nouveau système d'information de l'assurance maladie en France permettrait-il une estimation des flux d'émigration ? » Colloque de l'AIDELF, Budapest, septembre 2004.

1.1.4. Les enquêtes

Les enquêtes constituent, sans doute, le type de source le plus répandu dans le domaine de la santé avec les fichiers de gestion. Mais, elles sont assez peu utilisées dans les communications présentées. Nicolas Brouard, de l'INED, dans son travail intitulé : « Espérance de vie et espérances de santé : Sources, mesures et modèle », les évoque car les informations couplées à la mortalité générale pour donner naissance aux diverses espérances de santé, proviennent des enquêtes menées sur l'état de santé des populations. Parmi ces dernières, il convient de citer l'enquête Santé et protection sociale (ESPS) faite par l'IRDES (Institut de Recherche et Documentation en Économie de la Santé) dont les données viennent enrichir périodiquement l'échantillon permanent des assurés sociaux (EPAS) que nous avons évoqué précédemment.

Mais c'est surtout dans la communication d'Alain Ayerbe et Philippe Cordazzo, intitulée « Les données statistiques sur la santé disponibles dans la base New Cronos (Eurostat) » que ces dernières apparaissent. En effet, dans leur travail de recension et d'analyse des données mises à la disposition du public par Eurostat, à côté des sources administratives, les enquêtes, qu'elles soient nationales et internationales figurent en bonne place. Il convient ainsi de citer les enquêtes nationales de santé, qui comportent généralement deux volets, enquêtes sur entretien et enquête sur examen clinique. Malheureusement, les protocoles de ces enquêtes peuvent notablement varier d'un pays à l'autre, ce qui limite fortement la comparabilité des données qui en sont issues. Toutefois, quelques enquêtes européennes permettent de procéder à des comparaisons internationales, c'est notamment le cas du Panel communautaire des ménages, remplacé aujourd'hui par l'enquête sur le revenu et les conditions sociales, d'eurobaromètre et de l'enquête sur les forces de travail, pour reprendre le jargon d'Eurostat, qui comportent chacune des éléments sur l'état de santé et les soins.

1.1.5. Les recensements

Si le recensement général de la population n'est pas à proprement parler une source majeure de l'information sur la santé, il n'en constitue pas moins un élément incontournable à deux titres :

- fournissant l'image la moins déformée de la population totale, il constitue la base de sondage idéale pour les tirages d'échantillon, et permet de contrôler la représentativité de tous les fichiers et d'assurer les redressements nécessaires ;
- il fournit également la population soumise au risque, et permet à ce titre de définir le dénominateur pour le calcul des taux.

Si le premier aspect, redressement des échantillons, est celui qui intervient dans la communication de René Guéguen, Caroline Dupré, Virginie Dauphinot et Florence Naudin, du centre technique d'appui et de formation des Centres d'examen de santé, dans leur texte « Géographie de la santé observée dans les Centres d'examen de santé : propositions pour un tableau de bord de la santé dans les CES », c'est le second auquel s'intéresse Gérard-François Dumont. En effet, dans sa communication intitulée : « Recensement rénové et données géodémographiques : la France va-t-elle disposer de taux annuels de mortalité fiables aux échelons communaux », il s'intéresse aux conséquences du passage du recensement exhaustif traditionnel à une nouvelle méthode, qui associe un recensement exhaustif, mais tournant pour les communes de moins de 10 000 habitants à une enquête sur un échantillon de 8% de la population dans les communes de 10 000 habitants ou plus. Cette communication dépasse largement de cadre du colloque car les conséquences de ce changement ne se limitent au seul champ de la mortalité, mais à tous les domaines qui exigent une connaissance précise de la population, à un échelon infranational.

1.2. Limites des sources disponibles

À propos de la seconde question, « **Quelles sont les limitations de chacune d'elles ?** », il convient de rappeler que ces imperfections peuvent concerner l'un quelconque des divers maillons qui constituent la chaîne qui va de la collecte à la mise à disposition des données :

- le mode de collecte,
- le questionnaire,
- les réponses,
- l'exploitation,
- ...

C'est dire s'ils peuvent être nombreux et variés.

La question de l'imperfection des sources étant au cœur de la séance, la plupart des textes qui vont nous être présentés s'interrogent sur la qualité des données utilisées, leur fiabilité, leur représentativité ou leur complétude.

C'est ainsi que Béatrice Geoffroy-Perez, Sylvie Julliard, Aurélie Fouquet, Marcel Goldberg et Ellen Imbernon, de l'Institut de veille sanitaire, dans leur communication sur la « Surveillance de la mortalité selon l'activité professionnelle : le projet Cosmop », confirment, par la sous-mortalité qu'ils observent, la sélection que l'emploi opère en matière de bonne santé. Ils ont également mis en évidence d'autres biais sur lesquels ils reviendront dans leur présentation.

De même, René Guéguen, Caroline Dupré, Virginie Dauphinot et Florence Naudin, du centre technique d'appui et de formation des Centres d'examens de santé, dans leur texte « Géographie de la santé observée dans les Centres d'examen de santé : propositions pour un tableau de bord de la santé dans les CES », s'interrogent sur la représentativité des personnes examinées dans les Centres d'examens de santé, en particulier une surreprésentation des catégories défavorisées, ce qui les amène à proposer des procédures de redressement des estimations obtenues.

1.3. Procédures permettant d'améliorer l'utilisation des sources disponibles

Pour répondre à notre dernière question, **Comment peut-on pallier leurs défauts et améliorer leur utilisation ?**, il apparaît que la liste des procédures possibles est aussi longue et diverse que celle des imperfections.

Dans toute analyse, il faut pouvoir évaluer la qualité de la mesure² à laquelle on procède. C'est-à-dire apprécier son exactitude, ou sa capacité à rendre compte du phénomène étudié, et sa précision, ou sa capacité à donner des résultats identiques lorsque l'on répète la mesure dans des conditions identiques.

Si le défaut d'exactitude, qui est une erreur systématique, constitue un biais, il peut être corrigé si l'on en connaît l'existence et l'ampleur. Par contre, le défaut de précision, qui est une erreur aléatoire, constitue un handicap plus difficile à surmonter.

L'arsenal des méthodes disponibles est vaste, néanmoins on peut citer trois grandes familles qui ne sont pas sans avoir des liens entre elles :

- les méthodes indirectes,
- la modélisation,
- le couplage de données,

² Voir à ce sujet L. Papoz, « Qualité des données dans les études épidémiologiques », *Gérontologie et société*, n° 99, 2001, p. 25-36.

De ces trois familles, seule la modélisation n'apparaît pas dans les communications, même si l'utilisation de modèles multiniveaux est envisagée dans la communication des chercheurs du CETAF dans le but de réduire la variabilité due à la gestion décentralisée dans chacun des centres. Il est vrai qu'en dépit d'un certain succès la modélisation n'en connaît pas moins de sévères critiques qui tiennent aux hypothèses contraignantes, comme par exemple l'absence d'interactions entre les variables utilisées, dont la validation est rarement vérifiée. Elle peut s'avérer néanmoins utile dans certains cas. Je renvoie les lecteurs intéressés à un article qui utilise les modèles log-linéaires pour mesurer l'exhaustivité d'un registre de morbidité³.

C'est à un problème de même nature que s'attache la communication de Maryse Gaymard, de l'Université Bordeaux 2, Gil Bellis et Alain Parant, de l'Ined, intitulée : « L'observatoire national de la mucoviscidose : Quelle couverture de la population ? ». Mais ils privilégient une autre approche qui s'apparente plus aux méthodes indirectes qu'à la modélisation, même s'ils ont recours aux modèles de populations stables. En effet, ils estiment la complétude de cet observatoire en intégrant, notamment, dans les paramètres des populations stables, la mesure de la mortalité des personnes répertoriées.

Ces méthodes indirectes ont sans doute pourtant un rôle important à jouer dans la multiplicité des fichiers existants dans le domaine de la santé. Ces fichiers sont, en effet, le plus souvent constitués sans que l'utilisateur, épidémiologiste ou démographe, ne soit associé à sa conception.

Le troisième type de traitement permettant d'améliorer l'efficacité des sources disponibles et par conséquent le rendement d'une analyse c'est de coupler plusieurs sources entre elles. L'appariement de données est particulièrement présent dans cette séance, puisqu'il est à la base de quatre communications. C'est ainsi que Béatrice Geoffroy-Perez, Sylvie Julliard, Aurélie Fouquet, Marcel Goldberg et Ellen Imbernon, de l'Institut de veille sanitaire, présentent le projet Cosmop (cohorte pour la surveillance de la mortalité par profession) qui vise, en suivant l'évolution des causes de décès selon l'activité professionnelle, à évaluer l'impact des facteurs professionnels sur la santé et à orienter les politiques en matière de risques professionnels. Dans le cadre de cette étude, le fichier des causes médicales de décès a été couplé, dans un premier temps, avec l'échantillon démographique permanent (EDP) de l'INSEE, puis devant les conséquences de la mise en place du recensement rénové sur la pérennité de l'échantillon démographique permanent, il a été décidé d'explorer l'appariement avec un autre fichier de l'INSEE, le panel DADS (déclarations annuelles des données sociales, faites par les employeurs).

Le texte d'Isabelle Fournel, M. Schwarzinger, E. Benzenine, C. Binquet, Benoît Riandey et Catherine Quantin, « Déterminer le statut vital des patients domiciliés en France métropolitaine et dans les départements d'Outre Mer en chaînant données hospitalières et données de mortalité de l'INSEE anonymisées », montre tout le profit que l'on peut tirer dans l'évaluation des soins médicaux, de la recherche clinique et des études épidémiologiques, de la détermination du statut vital des patients par l'appariement des fichiers hospitaliers et des fichiers de mortalité de l'INSEE. En France, toutes ces opérations concernant le traitement de données à caractère personnel et en particulier les autorisations de procéder à des appariements sont soumises à l'autorisation préalable de la Commission nationale de l'informatique et des libertés. Mais nous reviendrons un peu plus loin sur ce point. L'anonymisation et le chaînage probabiliste, avec la mise en place de règles de décision, constituent ici deux étapes intermédiaires indispensables à l'analyse du statut vital des patients traités. Ce travail montre

³ P. Guillem, C. Lavergne, C. Cans, « Comment calculer l'exhaustivité d'un registre de morbidité ? L'exemple du registre des handicaps de l'enfant et observatoire périnatal de l'Isère », *Revue d'épidémiologie et de santé publique*, 48, 2000.

l'importance de la qualité des identifiants utilisés dans chacune des sources et leur compatibilité, qui découle en particulier de la méthode de *hachage* utilisée. Pour résoudre ces difficultés, les auteurs appellent de leurs vœux la mise en place de la carte vitale 2, fondée sur le NIR haché du patient dans la logique de l'identifiant qui se met en place pour le Dossier Médical Personnel.

Ce problème de l'identifiant santé du Dossier Médical Personnel fait l'objet d'une attention particulière dans la communication « Traitement épidémiologique et démographique du Dossier médical personnel. L'identifiant santé », préparée, en autres, par deux coauteurs de la communication précédente : Catherine Quantin, Benoît Riandey et Olivier Cohen. À propos de la mise en place de ce nouveau dossier médical, ils discutent des difficultés du choix de cet identifiant qui doit à la fois satisfaire les exigences du respect de l'anonymat avec le souci de choisir un identifiant efficace et si possible de partager le plus possible de sources, afin de faciliter les rapprochements tout en les sécurisant.

Dans leur communication intitulée « Une reconstitution de la trajectoire de vie des personnes décédées en Suisse, 1990-2005 », Philippe Wanner et Yannic Forney, du laboratoire de démographie et d'études familiales de l'Université de Genève, pallient l'absence de registre de population en Suisse en procédant, à partir de données anonymes et sans numéro d'identification, à un appariement des statistiques de la mortalité et le recensement de la population, visant à reconstituer des trajectoires de vie individuelle, avec pour objectif de mesurer les facteurs conduisant à un éventuel décès. Pour améliorer les scores des appariements, ils ont fait appel à une troisième source, les registres de la *centrale de compensation de l'assurance vieillesse, invalidité et survivants*, l'équivalent de nos caisses de retraite et ont testé diverses méthodes statistiques d'appariement.

Ces travaux mettent en évidence la richesse de l'appariement de sources multiples qui, par leur complémentarité, permettent de dresser un tableau plus précis de la réalité en matière de santé et de mieux comprendre les mécanismes qui déterminent l'état de santé.

1.4. Problèmes de comparabilité des sources disponibles

Mais une fois ces problèmes plus ou moins ponctuels résolus, tous les problèmes ne sont pas résolus. En effet, dans le domaine de la santé, comme dans bien d'autres, on ne peut se contenter de produire des résultats, il est souvent très utile de pouvoir les comparer à d'autres, soit dans le passé pour retracer des évolutions, soit à travers l'espace pour mieux les situer. Mais on se heurte très vite à des problèmes de comparabilité. En effet, les définitions des concepts, des événements et des populations ou les procédures de collecte peuvent varier d'un pays à l'autre ou au fil du temps. Cela implique donc toute une démarche d'analyse des diverses sources et de mise au point de procédure de rapprochement des définitions, lorsque cela est possible, et de mise en garde lorsque l'écart ne peut être comblé. C'est le travail auquel se livre, avec plus ou moins de réussite l'Office statistique des Communautés européennes (Eurostat), si l'on en croit Alain Ayerbe et Philippe Cordazzo, du département de démographie de l'Université Marc Bloch Strasbourg 2. Ces derniers se sont, en effet, penchés sur les données, plus ou moins bien harmonisées, qu'Eurostat met à la disposition du public dans le domaine de la santé publique dans sa base New Cronos. Mais ce qui n'est déjà pas très simple dans le domaine de la démographie devient beaucoup plus délicat dans le domaine de la santé. Le problème touche, ici, autant les sources disponibles que le calcul d'indicateurs comparatifs standardisés. Ils ont fait un effort très utile d'inventaire et d'analyse de ces données dont la diversité donne un aperçu des difficultés auxquelles toute analyse comparative est soumise si elle se veut pertinente.

Ce problème de comparabilité des données, à travers l'espace mais également à travers le temps, est particulièrement crucial dans l'analyse de la mortalité par cause en Europe, même s'il ne s'agit pas du secteur où ce problème se pose avec le plus d'acuité. En effet, comme

l'écrivent Alain Ayerbe et Philippe Cordazzo : « Les données relatives aux causes de décès sont certainement celles qui bénéficient des plus grandes garanties de comparabilité ». Cela donne à réfléchir pour ce qui est des autres éléments statistiques de la santé. Les auteurs estiment, en effet, que « le calcul d'indicateurs comparatifs nécessite la plus grande prudence ». Il semble cependant que les choses devraient s'améliorer dans le futur proche avec la mise en œuvre d'une enquête européenne sur la santé, bénéficiant de définitions et concepts identiques et une couverture temporelle et géographique commune.

Pour revenir aux causes de décès, lors du Colloque d'Amiens, Régis Flament nous avait déjà montré qu'en dépit de règles instaurées par l'Organisation mondiale de la santé (OMS), les offices nationaux restent souvent prisonniers de leurs habitudes en matière de diagnostics et règles de codage, et n'utilisent pas forcément la dernière classification internationale des maladies (CIM). Ici, France Meslé et Jacques Vallin, de l'Institut national d'études démographiques (Paris), attirent, une nouvelle fois, notre attention sur les discontinuités engendrées par chaque nouvelle révision de la classification internationale des maladies (CIM). Ils nous montrent qu'il n'est pas toujours facile ou possible de réduire cette discontinuité, en prenant pour exemple l'adoption de 10^{ème} révision de la classification internationale des maladies (CIM-10) en France.

Dans le cas français, la difficulté de la reconstruction d'une série homogène a été aggravée par la mise en place, à la même date, d'un système de codage automatique. Il est tout à fait regrettable que les services chargés de la mise en place d'une réforme, dont chacun sait avant, que les conséquences seront loin d'être négligeables, ne mettent pas en œuvre systématiquement, pour une période transitoire et pour un échantillon, un double classement ou une double classification qui permettrait alors de disposer des coefficients de passages nécessaires. A défaut, et en dépit des trésors d'imagination qu'ils développent, les chercheurs seront contraints à un long travail fastidieux de reconstitution.

1.5. Problèmes de restriction d'accès aux sources disponibles

Enfin pour terminer évoquons une difficulté qui tend à prendre de l'importance : les limitations de l'accès aux sources. Deux éléments concourent à leur renforcement :

- le premier est d'ordre économique. Les états considèrent de moins en moins la collecte de données comme un service public, au service de la connaissance immédiate ou future, mais comme une activité qui coûte à la collectivité, et dont il faut, par conséquent, limiter les coûts. Il existe donc un risque de voir disparaître des collectes d'informations sous le prétexte fallacieux que le nombre d'utilisateurs est faible et/ou que le coût en est élevé. Parallèlement une autre attitude se développe, celle qui consiste à considérer les résultats de la collecte comme un produit, auquel certains accordent quelque utilité et qu'il convient donc de vendre pour tenter de le rentabiliser. Il convient de signaler que, au moins au niveau européen, les choses vont dans le bon sens car depuis 2004, l'accès aux bases de données d'Eurostat est devenu totalement libre.

Ce problème de coût des données est abordé dans les communications d'Isabelle Fournel et de Catherine Quantin : la première écrit que bien que l'interrogation du RNIPP soit bien adaptée aux grandes bases de données, son coût relativement élevé lui fait préférer l'utilisation de la base de mortalité de l'INSEE. « L'utilisation de la base de mortalité de l'INSEE permet donc de réduire les coûts et d'éviter que des études épidémiologiques soient empêchées pour des raisons financières ». La seconde vante le Dossier médical personnel en disant que, comme tout fichier administratif, il présente l'atout de la gratuité et de l'exhaustivité.

- La seconde est plutôt d'ordre éthique ou philosophique, c'est l'extension de la notion de confidentialité. Aujourd'hui, la peur du pouvoir de l'informatique a conduit les sociétés

modernes, et particulièrement en France, à arbitrer toujours plus en faveur de la « protection des libertés fondamentales des individus » au détriment de la connaissance.

C'est à ce thème que se rattache la communication de Catherine Quantin, Benoît Riandey et Olivier Cohen, « traitement épidémiologique et démographique du Dossier médical personnel. L'identifiant santé », qui nous raconte la bataille actuelle autour de ce dossier médical en cours d'élaboration. En effet, le débat est l'accès par le chercheur aux données contenues dans ce dossier, qui n'est pas prévu dans la loi instituant le DMP, d'autant plus qu'il est prévu que l'information contenue ne sera complète, y compris pour les médecins, que si le patient en décide. D'autre part, l'intérêt de ce dossier réside avant tout dans son rapprochement avec d'autres sources. Est donc également posée la question de son éventuel appariement avec d'autres données, de santé ou démographiques. Pour permettre cet appariement, il convient de disposer d'un identifiant et se pose alors le choix de cet identifiant. Dans leur texte, les auteurs exposent la discussion à laquelle a donné lieu le choix de cet identifiant.

À cet égard, la Commission nationale de l'informatique et des libertés (CNIL) et le Comité consultatif sur le traitement de l'information en matière de recherche dans le domaine de la santé (CCTIRS) jouent un rôle de premier plan, car c'est de leur attitude que dépend la réalisation, ou non, de ces opérations indispensables au développement de la connaissance souhaitées par les scientifiques. Si pendant un temps la politique de ces organismes fut assez restrictive, il semble que l'attitude change quelque peu, sans doute parallèlement au développement des méthodes d'anonymisation et de cryptage.

Il s'agit ici d'un problème fondamental pour l'étude de la santé des populations. Car comme nous l'avons dit précédemment, c'est de la mise en relation des divers fichiers contenant des informations sur l'état de santé et l'environnement social, le statut économique des individus que viendra l'amélioration de la connaissance de la situation mais également la compréhension des mécanismes qui mènent de la bonne santé à la maladie et de la maladie à la mort, préalable nécessaire à une prise en charge plus efficace.

2. LES MESURES

Si les sources n'ont pas toujours la qualité que l'on serait en droit de souhaiter, l'autre versant est tout aussi important et la mesure doit être appropriée, tant à l'objectif que l'on s'est fixé qu'à la source que l'on utilise. Ce qui amène à un autre groupe de questions, symétrique à celui envisagé pour les sources :

- **De quelles mesures dispose-t-on?**
- **Quelles sont les limitations de chacune d'elles ?**
- **Comment peut-on pallier leurs défauts et améliorer leur utilisation ?**

Dans un domaine aux confins de plusieurs disciplines, comme l'est l'étude de la morbidité, on aurait pu croire le démographe, un spécialiste de la mesure, un peu plus présent sur ce terrain pour faire bénéficier de son expérience et, peut-être, faire des propositions à propos des indices les plus fréquemment utilisés dans ce champ. Malheureusement, parmi les papiers qui ont été préparés pour cette séance, très peu se sont proposés de faire un pont entre la démographie et l'épidémiologie dans le domaine de la mesure. Il est, en effet, un peu dommage que cette séance n'ait pas mieux réussi à établir un dialogue entre épidémiologistes et démographes. En tant qu'organisateur, j'en suis sans doute le premier responsable. Si, par ces questions je donne l'impression de m'interroger avant tout sur les méthodes utilisées en épidémiologie ou en démographie de la santé, cela ne signifie nullement que celles utilisées par le démographe soient toujours exemptes de critiques.

Ainsi, l'un des indices démographiques les plus anciens et les plus couramment utilisés, l'espérance de vie transversale à la naissance, n'est-il pas, au même titre que tous les

indicateurs conjoncturels transversaux, plongé au cœur du débat sur l'influence des modifications du calendrier des cohortes sur les mesures de l'« intensité » du moment.

Sans vouloir ouvrir ici le débat, je rappelle seulement que John Bongaarts a, au cours de ces dernières années, déclenché une petite polémique en affirmant que le calcul de l'espérance de vie transversale, bien que fondé sur des quotients et non sur des événements réduits, n'en est pas moins également affecté par le recul de l'âge au décès dans les générations. Le retard du calendrier de la mortalité serait à l'origine d'une amplification des gains d'espérance de vie mesuré dans les tables transversales.

Ce débat rappellera à certains d'entre vous, un autre débat un peu plus ancien, au sujet de la mesure de la primo-nuptialité transversale. Certains démographes préférant, à la somme des événements réduits, l'intensité de table transversale de primo-nuptialité, non sujette à atteindre des valeurs impossibles dans le cadre d'une interprétation de ces indices par l'artifice de la cohorte fictive ; alors même que cet « avantage » n'est atteint que par construction, sans que cela ne supprime pour autant la distorsion introduite par des variations du calendrier.

Faute de contributions suffisantes traitant de la mesure et l'organisateur s'étant révélé défaillant pour couvrir lui-même le champ, cette partie demeurera déséquilibrée. Il n'y aura pas de réponses systématiques données aux questions que nous avons proposées. Pourtant, il y aurait sans doute des choses à dire à propos des indices utilisés dans le domaine de la santé. Il en va ainsi de la multiplicité des espérances de vie en santé qui provoque une certaine confusion : espérance de vie en santé, espérance de vie sans incapacité, espérance de vie sans limitation d'activité, espérance de vie sans dépendance, espérance de vie ajustée selon l'état de santé, ..., fondées toutes sur des concepts différents mais utilisant toutes la méthode de Sullivan qui donne des résultats biaisés si l'état de santé se modifie au fil des générations.

C'est justement sur le calcul de cette espérance de vie en santé et ses limites que Nicolas Brouard attire notre attention dans sa communication.

L'analyse des limitations de nos instruments de mesure, permet de mieux définir la manière dont il convient de les utiliser. Cette démarche analytique n'est, malheureusement pas assez développée et pourtant elle permet de résoudre certaines difficultés pratiques. C'est ainsi que l'on est en droit d'utiliser une méthode pouvant donner des résultats biaisés, si l'on est sûr que, dans les conditions dans lesquelles on se trouve, l'ampleur du biais est limité. C'est à une réflexion de ce type que se sont livrées Marie-Hélène Cazes, Myriam Khat, Emmanuelle Génin et Marguerite Guiguet. Elles se sont, en effet, penchées sur les conditions d'utilisation, sans risque d'erreur, d'une méthode très classique de l'épidémiologie : les études cas-témoins. En effet, pour que les résultats obtenus avec cette méthode ne soient pas biaisés, il faut que la population dont sont issus cas et témoins soit parfaitement homogène. On ne se trouve pas forcément dans cette configuration et, à partir de simulations les auteurs explicitent les conditions dans lesquelles la probabilité de conclure à tort à une association entre un marqueur génétique et une maladie sera élevée.

Dans sa communication, Nicolas Cauchi-Duval, de l'Université Montesquieu-Bordeaux IV, passe à un cas concret et s'interroge sur la mesure de la mortalité selon l'expérimentation du cannabis ou selon le calendrier de l'expérimentation sexuelle. Pour cela, il utilise les processus de sélection, en grande partie causés par la mortalité différentielle, intervenant entre deux enquêtes rétrospectives successives renseignant sur l'âge à l'expérimentation.

Pour conclure, en revenant à la question initiale, je dirais que c'est plutôt l'aspect *sources* qui a retenu l'attention des auteurs, même si l'aspect *mesures* n'a pas été oublié. Il est vrai que ce colloque, à la différence des précédents, avait ouvert la possibilité que des contributions méthodologiques puissent intervenir dans chacune des autres séances, de telle sorte qu'elles complètent quelque peu cette dernière. Ainsi, comme vous allez le voir le programme très ambitieux qui avait été tracé n'a été que très partiellement réalisé. Le chemin suivi a été un peu

différent mais il n'en est pas moins aussi riche et intéressant, comme vous allez le découvrir maintenant avec les présentations.