

Contribution de l'ITMO Santé publique à la commission "Open Data en santé"

Avril 2014

Contexte

Les objectifs de la recherche en santé publique sont la production de connaissances, ou d'outils pour l'améliorer, visant à guider les décisions en santé ou les politiques de santé publique. L'originalité et l'intérêt de cette recherche sont essentiellement évalués dans le contexte compétitif international de la publication scientifique.

La principale matière de la recherche en santé publique, menée dans les laboratoires rattachés à l'ITMO de santé publique et dépendante de différentes institutions (INSERM, CNRS, autres EPST, Universités, CHUs...), est l'exploitation de données de santé et de données susceptibles d'avoir un lien avec la santé.

C'est pourquoi les chercheurs ont besoin de pouvoir utiliser des données de natures très variées :

- clinique, génétique, biologique,
- environnementale, professionnelle, sociale, de consommation,
- économique, de parcours de soin, de consommation de soin.

La recherche en santé publique est à la fois productrice et utilisatrice de données. Trois grands types de sources peuvent être distingués :

1. les données émanant d'un protocole mis en place spécifiquement par les chercheurs pour les objectifs de la recherche,
2. les données découlant d'un protocole de recueil permanent (données cliniques, registres, causes de décès) géré par le chercheur ou son institution,
3. les données provenant d'une autre source (organismes de protection sociale, INSEE, administrative, commerciale) non initialement dédiée à la recherche.

Le plus souvent, les données traitées sont indirectement identifiantes. Les chercheurs sont donc astreints au secret. Ils doivent également veiller à ne collecter que des données ayant un intérêt potentiel pour le sujet de recherche, à cloisonner les données très identifiantes (nom, prénom, adresse de résidence notamment) des données de santé, à crypter certaines informations, et à limiter les accès et la recopie de l'information.

Dans un souci de production de connaissance scientifique, les chercheurs doivent mettre tous les moyens en œuvre pour exploiter avec la plus grande rigueur, la plus grande expertise et le plus grand esprit critique, les données à leur disposition, en particulier lorsque celles-ci n'ont pas été recueillies a priori dans le cadre d'un protocole scientifique. Par le biais de publications scientifiques, mais également de collaborations

nombreuses, la communauté des chercheurs est exercée à la mutualisation et aux échanges de méthodes permettant une utilisation optimale des bases de données. En raison de cette exigence de rigueur, de prudence et de méthode particulièrement développée dans la recherche française, celle-ci a longtemps considéré les données médico-administratives comme peu exploitables, alors que d'autres domaines de recherche (ex. les sciences économiques et sociales) pratiquent depuis de nombreuses années l'« analyse secondaire » de données recueillies par les administrations économiques et sociales ou/et les organisations internationales.

La perspective d'utiliser des données médico-administratives dans un objectif de production de connaissance utile à la santé publique est récente. Elle provient essentiellement de l'amélioration et de la centralisation des systèmes d'informations de l'assurance maladie et de l'administration hospitalière (PMSI) au sein du SNIIRAM. La fiabilisation de ces données et leur ouverture progressive aux chercheurs a généré de très nombreuses perspectives de recherches, dont les retombées pour la santé publique sont importantes. La communauté des chercheurs en santé publique y a ainsi identifié des opportunités évidentes de gain de compétitivité en matière de production scientifique.

Parallèlement, un besoin spécifique à la recherche s'est de plus en plus fait sentir : la nécessité d'apparier des données de santé produites par les institutions de recherche (enquêtes, cohortes, registres, causes de décès) avec d'autres données, dont justement les données du SNIIRAM-PMSI. Ce type d'appariement permet de capter des informations sur les consommations de soin de façon a priori plus exhaustive et moins coûteuse que dans un questionnaire d'enquête. Compte tenu de la quasi-exhaustivité de l'information, ces appariements permettent également d'avoir la puissance maximale d'analyse sur des questions parfois très spécifiques portant sur des événements rares. La problématique des appariements ne se limite pas au SNIIRAM-PMSI, et s'étend à des données caractérisant la situation professionnelle (ex. données de la CNAV), résidentielle (ex. données de la Direction Générale des Impôts) ou sociodémographique (données de l'INSEE), souvent importantes pour caractériser la survenue et les déterminants d'évènements de santé. Avec les progrès rapides des méthodes de séquençage et d'autres instruments d'analyse biologique, ce besoin dépasse désormais la seule recherche en santé publique, qui fait l'objet de la présente contribution, pour s'étendre à la recherche biomédicale dans son ensemble (comme le confirme l'expérience d'autres pays, comme celle du Danemark).

Par ailleurs, la recherche en santé publique se situe de façon croissante dans un contexte collaboratif national et international. Afin de déterminer la fiabilité et le caractère généralisable de résultats obtenus, il est devenu indispensable de pouvoir les reproduire dans plusieurs pays, contextes, et types de population. C'est pourquoi les modalités d'échange et de partage trans-frontières de données représentent un enjeu majeur de la recherche en santé publique.

Difficultés rencontrées par les chercheurs en santé publique

Accès aux données individuelles

Actuellement les principales limites de l'accès aux données pour la recherche sont les suivantes :

L'accès de la recherche aux données exhaustives de remboursement du SNIIRAM (DCIR) est possible, mais limité en pratique. Il nécessite un temps d'attente conséquent résultant du grand nombre de demandes au regard des ressources disponibles et de la complexité des procédures actuelles d'autorisation d'accès. Ceci nuit à la compétitivité de la recherche.

L'accès à l'EGB (échantillon généraliste des bénéficiaires), aujourd'hui géré directement par l'INSERM pour ses structures de recherche affiliées, est très riche et générateur de résultats innovants. Toutefois, la puissance statistique peut être limitée lorsque les chercheurs souhaitent travailler sur des questions concernant des populations, des soins ou des maladies, rares.

Les différents organismes chargés de donner les avis et autorisations d'accès aux données individuelles (CCTIRS, IDS, Comité du secret statistique, CNIL) sont redondants et peu coordonnés. Il est bien souvent complexe pour le chercheur de comprendre la démarche correspondant à son besoin.

Réalisation d'appariements

Les appariements entre bases de données de recherche et bases de données médico-administratives (dont le SNIIRAM) ou statistiques nécessitent souvent l'emploi du Numéro d'Identification au Répertoire (NIR), aujourd'hui très encadré juridiquement. L'utilisation du NIR par la recherche publique exige le passage d'un décret en conseil d'Etat, dont l'obtention prend en moyenne 18 mois, au prix d'efforts importants.

Dans le cas des causes de décès, le dispositif juridique actuel est très restrictif sur les protocoles d'appariement possible et les destinataires des données (essentiellement l'INSEE, l'INSERM, l'InVS, les registres et l'administration sanitaire).

Ouverture et partage des données produites par la recherche

Les difficultés pour la recherche à ouvrir et partager ses données au niveau national et international sont les suivantes :

- la constitution de bases de données se fait sur la base de l'initiative personnelle, avec un retour sur investissement en termes de primauté sur les publications scientifiques. Il s'agit d'un des moteurs de l'innovation qui doit être considéré.
- une partie des financements obtenus par la recherche se fait à travers la monétisation de la transmission de données produites. Elle permet souvent de participer à la maintenance d'un système d'information. Une ouverture des

données qui exigerait la gratuité doit donc considérer la problématique de compensation des financements nécessaires pour le producteur.

- le partage des données de santé à l'échelle nationale et internationale doit intégrer les principes de la réciprocité et de la pertinence scientifique (intérêt et originalité potentielle des résultats attendus). La réciprocité ne doit pas seulement s'évaluer sur l'affichage d'un partage des données, mais également sur la qualité des données partagées.

Propositions

Aspects juridiques

Conformément au paragraphe 3.3.2.1 du rapport de la Commission Open Data, portant sur les accès liés aux missions permanentes, l'ITMO Santé publique souhaite qu'un ou plusieurs organismes de recherche publique aient la possibilité, en respectant toutes les spécifications de sécurité, de traçabilité, et de cloisonnement des données, de mettre en œuvre l'accès à des données individuelles, dont le SNIIRAM-PMSI, avec leur propre moyen à travers une ou des structures clairement identifiées. Il ne s'agit pas d'ouvrir de façon indifférenciée les accès aux données à l'ensemble de la communauté des chercheurs, mais de permettre :

- que les chercheurs producteurs de données (d'enquêtes ou/et de cohortes) puissent enrichir leurs propres bases, sur de longs termes pour les cohortes, avec d'autres données, dont le SNIIRAM-PMSI, et organiser la mutualisation de ces données avec d'autres équipes de recherche ou organismes,
- qu'une structure puisse faire l'interface entre les demandes d'équipes de recherche et les bases de données parfois complexes, comme le SNIIRAM-PMSI, pour orienter et conseiller techniquement les demandes initialement formulées à partir d'une question de recherche et pour garantir le respect des procédures d'accréditation et d'accès définies par les pouvoirs publics.

Conformément aux paragraphes 3.3.1 et 3.3.2.2 du rapport, l'ITMO Santé publique est favorable à la mise en place d'un guichet unique, afin de simplifier les procédures d'autorisation d'accès aux données. Il paraît hautement souhaitable que tout en respectant les spécifications générales fixées, la recherche publique puisse bénéficier d'un circuit spécifique et facilité pour ces procédures. Ceci à condition que ce guichet bénéficie de moyens suffisants pour pouvoir traiter les demandes de façon rapide et transparente, et pour remplir une fonction de conseil auprès des demandeurs pour orienter les spécifications juridiques et techniques des demandes. L'ITMO Santé publique est prêt à collaborer avec cette structure unifiée afin de mettre en place et définir des procédures d'autorisations simplifiées.

Conformément aux propositions du groupe « Risque de Réidentification » de la commission, l'ITMO Santé publique est favorable à une ouverture simplifiée et fluidifiée de données issues d'échantillon, comme l'EGB. Plus spécifiquement pour ce dernier, il

encourage la création d'un échantillon présentant des caractéristiques et des droits d'accès similaires mais d'une taille correspondant à 10% de la population (au lieu de 1/97 pour l'EGB).

Conformément au paragraphe 3.3.2.2 du rapport, l'ITMO Santé publique souhaite que le cadre juridique évolue pour faciliter l'emploi du NIR. Cette évolution pourra prendre la forme d'un changement de la loi ou de l'adoption d'un ou plusieurs décrets-cadre, pour l'élaboration desquels l'ITMO Santé publique est prêt à collaborer.

Aspect opérationnel

L'organisation actuelle des systèmes d'information dans les organismes de recherche est très décentralisée, et n'est pas fondée sur un référentiel technique et organisationnel commun permettant :

- de tracer les accès aux données sensibles,
- de garantir la non recopie des données et l'usage conforme aux objectifs initiaux,
- de garantir un niveau de sécurité optimal.

A court terme, l'utilisation d'un système tel que le CASD (Centre d'Accès Sécurisé Distant aux données) devrait permettre de respecter un tel référentiel sans investissement lourd. En outre, une telle solution faciliterait vraisemblablement l'accès et l'appariement à de nombreuses bases de données statistiques et administratives.

Toutefois, compte tenu des besoins spécifiques de la recherche en santé publique, et de sa volonté d'autonomie en matière de gestion des moyens pour garantir ses propres accès, l'ITMO Santé publique insiste sur l'importance d'ouvrir juridiquement la possibilité de créer à terme autant d'unités de service directement gérées par les organismes de recherche qu'il sera nécessaire pour les besoins de la recherche publique, étant entendu que chacune de ces structures devra respecter un cahier des charges commun et précis. Il considère par ailleurs que la mise en place de plusieurs structures d'accès pérenne doit être fortement régulée, mais néanmoins encouragée pour garantir une diversité des interlocuteurs et une continuité de service.

Ouverture et partage des données produites

La mise en place d'une structure d'accès gérée par les organismes de recherche se fera dans le contexte de la définition d'une politique globale d'accès et de gestion des données. Une telle politique s'accompagnera de recommandations en matière d'ouverture des données produites au grand public, et de partage des données produites par ses équipes. L'évaluation future des laboratoires et des chercheurs devra prendre en compte le respect de ces recommandations comme un élément important.

L'ITMO santé publique, dans la continuité des textes signés par l'INSERM en 2011¹ est favorable et doit être proactif pour accélérer le partage des données produites par les

¹ Walport M, Brest P. Sharing research data to improve public health. Lancet. 2011;377(9765):537-9

entités dont il est responsable auprès des autres équipes de recherches. Il établira une charte qui clarifiera les conditions :

- de primauté des chercheurs sur les données qu'ils produisent, cette considération ne s'appliquera toutefois pas aux protocoles de recueil de données permanents (ex. registres, causes de décès),
- de réciprocité et d'équité avec les autres pays,
- d'évaluation de la pertinence scientifique des demandes de données, en veillant à ce que cette évaluation prenne en compte l'avis du producteur, mais intègre également une forte composante indépendante, de préférence internationale.

Dans le respect des conditions citées ci-dessus, l'ITMO santé publique est favorable à une politique d'Open data sur les données produites par les entités dont il est responsable, dans les limites suivantes :

- les données doivent être non ré-identifiables,
- leur documentation doit être suffisamment informative et sans ambiguïté.

Financement de la recherche

Le financement de la recherche publique est très divers, et fait intervenir dans un certain nombre de cas des acteurs privés à but lucratif.

Dans un souci de cohérence et afin d'éviter des dérives, l'ITMO Santé publique considère que les accès pérennes aux bases de données individuelles produites par une administration publique ne devront être donnés qu'aux structures fonctionnant sur financement public ou émanant d'appel d'offre géré par un organisme public.

En ce qui concerne les accès ponctuels, conformément à l'article 3.3.2.2 du présent rapport, l'ITMO santé publique est favorable à ce que ces accès soient rendus possibles sans évaluation lorsque le projet répond à une demande d'une autorité publique, indépendamment des sources de financement du projet. Il est également favorable à ce que des équipes puissent obtenir l'accès à ces données, après évaluation de la pertinence technique et de l'intérêt en santé publique par une entité indépendante et compétente, similaire au CCTIRS actuel et doté des ressources appropriées, quelle que soit la source de financement du projet. Cette entité pourrait se doter de procédures accélérées ou de méthodologies de référence dans les cas les plus simples.