

Direction scientifique – BdD et SI en santé

Contribution de l'Institut de veille sanitaire
au débat public organisé par la Conférence Nationale de santé
Comment utiliser les données de santé ?

Présentation générale de l'Institut de veille sanitaire

La loi n°98-535 du 1er juillet 1998 modifiée par la loi n°2004-806 du 9 août 2004 relative à la politique de santé publique, confère à l'InVS les missions suivantes (article L1413-2 du code de la santé publique) :

- la surveillance et l'observation permanentes de l'état de santé de la population. A ce titre, il participe au recueil de données sur l'état de santé de la population à des fins épidémiologiques en s'appuyant notamment sur des correspondants publics et privés faisant partie d'un réseau national de santé publique ;
- la veille et la vigilance sanitaires. A ce titre, l'InVS doit, entre autres, détecter de manière prospective les facteurs de risque susceptibles de modifier l'état de santé de la population ou de certaines de ses composantes, ..., d'étudier et de répertorier pour chaque type de risque les populations le plus fragiles ou menacées ;
- l'alerte sanitaire ;
- une contribution à la gestion des situations de crise sanitaire.

En vue d'assurer l'accomplissement de ces missions, l'InVS :

- effectue dans son domaine de compétence, toutes études, recherches, actions de formation ou d'information ;
- met en place des systèmes d'information lui permettant d'utiliser, dans les meilleurs délais les données scientifiques, climatiques, sanitaires, démographiques et sociales, notamment en matière de morbidité et de mortalité, qui sont nécessaires à l'exercice de ces missions (art L1413-3 -2).

L'article L1413-4 du code de la Santé publique prévoit que l'état et les collectivités territoriales, leurs établissements publics, les organismes de sécurité sociale, les services de protection civile ou d'urgences, les observatoires régionaux de santé, etc., et tous correspondants publics et privés de l'InVS, appartenant au Réseau national de santé publique, lui transmettent les informations nécessaires à l'exercice de ses missions. Depuis juillet 2009, il est également précisé dans cet article que les professionnels de santé transmettent à l'institut les données individuelles nécessaires à l'exercice de ses missions dans des conditions préservant la confidentialité de ces données à l'égard des tiers.

Les données nécessaires

Pour répondre à ses missions, l'InVS est amené à développer des systèmes d'informations reposant sur des données individuelles, au sens d'unité statistique. Globalement, la finalité de ces systèmes est de dénombrer des pathologies diagnostiquées dans une population définie, d'en décrire les tendances évolutives, les principaux déterminants, les inégalités géographiques, sociales, de contribuer à l'évaluation des politiques en matière de santé et d'accès aux soins, etc. Les

informations nécessaires à la surveillance sont donc relatives à l'état de santé de l'individu au sens large (morbidité, recours aux soins et prise en charge, facteurs de risque connus), à ses caractéristiques socio-démographiques mais aussi à son environnement, son activité professionnelle, etc., afin d'explorer, de décrire puis de suivre dans le temps l'état de santé et ses déterminants.

Les données « de santé » pertinentes pour la surveillance, comme pour la recherche en santé publique, dépassent largement le cadre de l'état de santé proprement dit. Les sources de données à mobiliser pour obtenir ces informations sont multiples et ont souvent été conçues pour d'autres finalités que la santé publique. Les problématiques que cela soulève concernent la disponibilité et les conditions d'utilisation de ces sources pour une finalité autre que celle pour laquelle elles ont été créées, les besoins d'interactions entre elles, c'est-à-dire entre systèmes d'informations, et les conditions de sécurité et de confidentialité que cela impose vis-à-vis des personnes. Ces problématiques s'inscrivent, de plus, pour l'InVS, dans de fortes contraintes de temps inhérentes à ses missions, notamment en cas d'alerte et de saisine par les ministères de tutelle ou leurs représentants locaux. Cette contrainte est d'ailleurs inscrite dans le code de la santé publique puisque l'article L1413-3 -2 (cf. ci-dessus) précise que la mise en place des systèmes d'information doit intervenir dans les « meilleurs délais ».

En annexe sont listées les principales sources de données utilisées actuellement par l'InVS.

L'objet de la présente contribution est d'exposer les problématiques liées à la transmission et à l'utilisation par l'InVS, dans des conditions compatibles avec l'exercice de ses missions, de données individuelles relatives à la santé issues de sources non dédiées initialement à la surveillance épidémiologique.

Les difficultés liées à l'accès aux sources de données non dédiées à la surveillance sanitaire :

- Les sources de données administratives :

Dans tous les pays développés, bien que recueillies pour des raisons administratives et budgétaires, les bases médico-administratives alimentent les systèmes de surveillance, parfois même de manière exclusive, comme au Québec, du fait de leur disponibilité sans coût supplémentaire.

En France, pour répondre à ses missions, l'InVS recourt aux bases de données médico-administratives, conçues pour une finalité autre que la santé publique. Il dispose ainsi d'une copie annuelle de la base du Pmsi (enregistrement de tous les séjours hospitaliers), PMSI-MCO court séjour depuis 1998, PMSI-SSR moyen séjour depuis 2005 et PMSI-PSY séjours psychiatriques depuis 2007.

Outre leur intérêt propre, ces bases de données appariées entre elles ou avec d'autres systèmes d'information existants, avec des données recueillies par enquêtes en population ou auprès de professionnels, constituent de « nouveaux » systèmes d'informations plus riches, permettant de produire des informations et connaissances inaccessibles par ailleurs. Ces croisements peuvent permettre par exemple de mettre en relation données médicales et données d'exposition, environnementale ou professionnelle, ou sociodémographiques... pour un même individu.

Or, l'un des principaux problèmes rencontrés par l'InVS dans l'exercice de ses missions concerne l'accès sécurisé en routine aux données individuelles de ces bases médico-administratives et la possibilité de pouvoir croiser de manière sécurisée et efficace ces données avec d'autres systèmes d'information ou d'autres sources quelles qu'elles soient, dans un objectif de santé publique.

L'accès aux données individuelles issues des bases médico-administratives

- La base de données relative aux affections de longue durée (ALD)
La possibilité d'utiliser à des fins de surveillance les données individuelles relatives aux exonérations du ticket modérateur pour Affection longue durée (ALD) est très importante pour l'InVS. En effet, ces données peuvent permettre d'estimer la prévalence et l'incidence de

certaines pathologies à un échelon géographique fin et d'apprécier leurs tendances évolutives, à condition toutefois de pouvoir chaîner les informations concernant un même individu. L'InVS dispose, actuellement, de ces données de manière agrégée pour les trois principaux régimes d'assurance maladie.

- SNIIRAM

Les données de l'assurance maladie, et en particulier le SNIIRAM, intéressent particulièrement l'InVS. Le SNIIRAM a été institué afin de contribuer à la connaissance des dépenses de l'ensemble des régimes d'assurance maladie et à la transmission aux professionnels de santé d'informations relatives à leur activité. Son utilisation à des fins de surveillance en santé publique par l'Institut de veille sanitaire n'est prévue que sous forme de données agrégées ce qui en limite fortement la portée et l'intérêt. Ainsi, l'InVS n'est pas autorisé à accéder en routine aux données individuelles, qui ne sont conservées que deux ans « glissant » dans le SNIIRAM et sont donc perdues ultérieurement, interdisant des analyses temporelles sur de longues périodes. En revanche, l'InVS a accès aux données individuelles de l'Echantillon généraliste de bénéficiaires au 1/97, mais cet accès ne permet pas d'étudier des pathologies peu fréquentes, et n'est pas représentatif au niveau régional ou départemental, ce qui ne permet pas d'analyse au niveau régional et infra régional, besoin rendu de plus en plus nécessaire, notamment avec la création des ARS. Chaque projet d'analyse impliquant une extraction des données individuelles du SNIIRAM doit ainsi faire l'objet d'une demande spécifique à la CNIL puis d'une extraction par la Cnam-TS, ce qui induit, dans les conditions de travail actuelles de ces 2 institutions, des délais parfois incompatibles avec l'exercice des missions de l'InVS en situation d'alerte ou de saisine de la part de ses Ministères de tutelle.

Le croisement des informations issues des bases médico-administratives avec celles issues d'autres systèmes d'information ou d'autres sources (enquêtes, professionnels de santé) dans un objectif de santé publique.

Comme nous l'avons dit, l'InVS peut être amené à croiser des données issues de bases de données distinctes afin d'obtenir un système d'informations plus riche et ainsi produire des informations et connaissances plus précises d'un point de vue surveillance épidémiologique.

Par exemple, un séjour hospitalier repéré dans le PMSI avec un diagnostic d'une pathologie cancéreuse l'année où le patient a été inscrit en ALD permet de classer de façon fiable l'individu comme nouveau cas pour cette pathologie cette année-là.

Dans le domaine particulier des risques professionnels il n'existe pratiquement aucune source de données directement exploitable comportant à la fois des informations de qualité sur la santé et sur la vie professionnelle des personnes. Certaines investigations ou surveillances conduites par l'InVS nécessitent par conséquent d'apparier à l'échelon individuel des sources de données sur la santé (causes de décès, données de registres de cancer, ...) entre elles ou avec des sources de données sur la carrière professionnelle (échantillons de population de l'Insee, des données individuelles de carrières conservées par la CNAV), ou d'autres sources (Inspection du travail, médecine légale, données de surveillance des urgences, ...).

Enfin, pouvoir croiser données de santé et localisation fine de l'habitat permet d'explorer par exemple les liens entre état de santé et des expositions environnementales. Comme dans le domaine des risques professionnels, la surveillance de l'évolution de l'état de santé de « cohortes » de personnes exposées à un moment de leur vie à une source ou un événement environnemental nécessiterait également de croiser plusieurs sources de données sanitaires, comme les consommations de soins, et les causes de décès.

Le croisement de plusieurs bases de données individuelles nécessite un identifiant commun entre les différentes bases, afin d'obtenir un système d'information « multisources » de données individuelles dans lequel chaque individu est repérable par un numéro d'anonymat unique non réversible (ne permettant pas le retour à l'identité de la personne). L'assurance maladie et l'Agence

de traitement des informations hospitalières (l'ATIH) utilisent déjà un processus d'anonymisation commun basé sur le NIR. La fonction FOIN (fonction d'occultation des informations nominatives) génère à partir du NIR, de la date de naissance et du sexe, un numéro anonyme, spécifique et non réversible qui permet pour un individu donné de chaîner toutes ses « consommations » de soins dans les deux bases SNIIRAM et PMSI. Ces 2 bases médico-administratives constituent, avec les certificats médicaux de décès, les principales sources d'informations « médicalisées » nationales exhaustives, elles permettent de plus un suivi dans le temps. Pour la plupart des projets de SI multisources, il est et sera donc nécessaire de recourir à ces bases, donc à ce même processus d'anonymisation, ce qui suppose de disposer, dans les autres sources à croiser, des variables le permettant, notamment le NIR. La procédure pour le « montage » de tel SI impose des règles de sécurité maximale, lourdes et complexes, le NIR ne peut être recueilli dans le cadre d'une enquête sans un décret en conseil d'état, et la fonction d'anonymisation utilisée est « propriété » de la CNAMTS.

Ci-dessous des exemples précis d'appariements de différentes sources de données avec des bases médico-administratives :

- Exemple 1 : le projet de système multisources « cancer »

Le système de surveillance des cancers en France repose principalement sur les registres des cancers qui couvrent, pour les adultes 18 % de la population. L'amélioration de ce dispositif de surveillance dans le cadre du plan Cancer intègre la mise en place d'un dispositif d'enregistrement national des nouveaux cas de cancers. Ce projet « système national multisources de surveillance des cancers » (SNMSC) prévoit d'apparier les données individuelles anatomocytopathologiques, recueillies auprès des anatomopathologistes, et les données du Pmsi et des ALD, afin d'identifier tous les nouveaux cas de cancers, chacune des sources étant insuffisante à elle seule pour distinguer avec certitude entre un cas récidivant et un nouveau cas. L'impossibilité d'utiliser directement le NIR et la fonction d'anonymisation FOIN utilisée par la Cnam-TS pour anonymiser les données des affiliés et produire l'identifiant commun au PMSI et aux ALD, a conduit à un dispositif tellement compliqué (notamment nécessitant que les données recueillies transitent par l'assurance maladie pour être anonymisées) qu'après 4 années de travail, il est apparu irréalisable en routine.

- Exemple 2 : le programme COSET

Le programme COSET a pour objectif la surveillance au long cours de l'état de santé de la population française en relation avec l'activité professionnelle. Il s'appuie sur des données longitudinales issues de cohortes d'actifs et anciens actifs couverts par les principaux régimes de protection sociale. Les données de questionnaire (relatives à l'état de santé et à la situation professionnelle de la personne) seront complétées par le recueil de données individuelles de consommation de soins et de données de carrière extraites des systèmes d'information de l'assurance maladie (Sniir-AM) et de l'assurance retraite (CNAV et autres régimes). L'intérêt de ces données est lié à leur quasi exhaustivité sur la population ciblée, et à leur disponibilité indépendamment de la participation active des sujets. Elles permettent à la fois de suivre au long cours la population d'étude, mais aussi d'analyser les biais de participation à l'inclusion et au fil du suivi. L'accès à ces données individuelles nécessite de conserver, ou de faire conserver par un tiers, sur le long terme les identifiants nécessaires à ces accès (NIR ou numéro FOIN par exemple).

- Exemple 3 : autres situations

Dans d'autres situations, il est demandé à l'InVS de documenter au niveau local les liens entre la survenue de pathologies (cancers notamment) et des expositions, soit environnementales (site pollué, incinérateurs d'ordures ménagères) soit professionnelles. Ces situations correspondent le plus souvent à des saisines, voire à des alertes. La manière la plus efficace et la plus réactive de répondre à ces préoccupations de santé publique de plus en plus

nombreuses est l'accès organisé et sécurisé aux bases de données appariées tel que décrit ci-dessus. Cette approche est généralisée dans bon nombre de pays du Nord de l'Europe.

Les missions que la loi a confié à l'InVS ne peuvent être atteintes de manière optimale pour un nombre important de questions de santé publique du fait des difficultés d'accès sécurisé aux données individuelles de santé et de leur croisement entre elles ou avec d'autres sources de données, notamment sociales. Bon nombre de ces difficultés peut être résolu techniquement avec un haut niveau de sécurité vis-à-vis de l'identité des personnes mais nécessite des évolutions réglementaires. Ces évolutions permettraient de disposer de manière beaucoup plus réactive et exhaustive d'informations sur l'état de santé de la population et son évolution dans le temps et dans l'espace, essentielles aux politiques de santé.

Outre la complexité technique, juridique et réglementaire de la procédure, les délais nécessaires à l'obtention des autorisations, des données, etc., semblent aussi liés à la multiplication des demandes de ce type, associée à un déficit de moyens humains dédiés à ces problématiques spécifiques au sein des différentes institutions concernées : producteurs de données, d'autant que la surveillance et la veille sanitaire ne font pas partie de leurs missions « princeps », CNIL et institutions utilisatrices comme l'InVS.

- Les sources de données « médicales » :

La dématérialisation croissante des sources de données médicales relevant directement de la prise en charge du patient, telles que le dossier médical, les comptes-rendus d'examen anatomopathologiques, biologiques, radiologiques, etc., ouvre des opportunités pour mieux répondre aux besoins de surveillance et de veille sanitaire et plus globalement de connaissance en santé publique. Ces besoins doivent être pris en considération dès la phase de conception des systèmes d'information correspondants. En effet, pour y répondre, ces systèmes devraient permettre une exploitation statistique des données et leur croisement avec celles d'autres sources (cf. supra). Cela suppose d'une part, une structuration et une standardisation (a minima) du recueil de l'information, d'autre part l'utilisation d'identifiants individuels permettant, ou au moins n'excluant pas, toute possibilité de croisement avec d'autres sources, notamment le SNIIRAM et le PMSI, les certificats de décès, les fichiers de la CNAV, etc. La première condition peut intervenir comme une évolution des SI mis en place mais la seconde condition doit être prise en considération dès à présent dans la définition du cadre d'interopérabilité des systèmes d'information en santé. En particulier, le DMP associé à l'identifiant national de santé, (qui est un identifiant non compatible avec les identifiants du SNIIRAM et du PMSI) est un service conçu pour contribuer à l'amélioration de la coordination de la continuité et de la qualité des soins. L'absence de finalité de santé publique, notamment de surveillance, à ce stade, ne doit pas écarter toute possibilité ultérieure d'utilisation dans une perspective de santé publique. Il en va de même pour les dossiers spécifiques comme le dossier communiquant de cancérologie ou les logiciels métiers.

Conclusions

Les politiques publiques en matière de santé publique et de sécurité sanitaire impliquent de disposer de données sur les phénomènes de santé, de plus en plus réactives et fines au niveau géographique. L'Institut de veille sanitaire, principal opérateur public en matière de surveillance en santé publique et de veille sanitaire, est amené, pour assurer ses missions définies par la loi, à organiser la mise en commun efficace et sécurisée des données sanitaires individuelles disponibles dans les grandes bases de données sanitaires et sociodémographiques. La réalisation de ses missions, dans un contexte à la fois de demande sociale croissante de connaissances sur les risques et de protection de la vie privée, implique de faire évoluer l'accès et l'appariement sécurisé de ces bases de données de santé dans un cadre réglementaire adapté et en toute transparence vis-à-vis des personnes.

ANNEXE : principales sources de données utilisées par l'InVS

Les données utilisées peuvent être produites par l'InVS ou par d'autres organismes.

Les sources de données à vocation exhaustives :

- Base des causes individuelles de décès délivrée par le CépiDC de l'Inserm. L'InVS dispose d'une copie annuelle de la base depuis 1968.
- Base du PMSI : PMSI-MCO court séjour depuis 1998, PMSI-SSR moyen séjour depuis 2005 et PMSI-PSY séjours psychiatriques depuis 2007. L'InVS dispose également d'une copie des bases PMSI.
- Données de remboursement de l'Assurance maladie : accès aux données individuelles de l'Echantillon généraliste de bénéficiaires (EGB) et aux données agrégées de SNIIR-AM.
- Affections de longues durées (ALD) de l'assurance maladie
- Système de déclaration obligatoire (DO) de problèmes de santé.
- Centres de Référence (CNR)
- Registres qualifiés (nationaux, régionaux ou départementaux)
- Certificats de santé de l'enfant
- Bases de données sur les dépistages organisés des cancers (sein, colon, col)

Les réseaux de surveillance constitués à partir d'échantillon de Professionnels de santé, de laboratoires, de services hospitaliers, etc. :

- Réseaux de surveillance des infections nosocomiales (Raisin)
- Réseaux de laboratoires, de services hospitaliers ou de médecins libéraux (Sentinelles, Grog)
- Enquête permanente sur les accidents de la vie courante
- SURSAUD Surveillance sanitaire des urgences et des décès (données des services d'urgences, de SOS Médecin, ...)

Les grandes Enquêtes nationales en population générale :

- Enquêtes déclaratives
 - Enquêtes santé protection sociale (ESPS) de l'IRDES
 - Enquête sur la santé de l'INSEE/DREES
 - Baromètres santé de l'Institut national de prévention et d'éducation pour la santé (INPES)
 - Enquête nationale périnatale (DGS/DREES/PMI/INSERM)
 - Enquêtes en milieu scolaire (DREES, Ministère de l'Éducation nationale)
 - Enquête « corpulence des enfants de 7-9 ans scolarisés en CE1&CE2 » (USEN)
 - Enquêtes Surveillance médicale des risques professionnels (SUMER)
 - Enquêtes sur les Conditions de travail du Ministère du travail, des relations sociales et de la solidarité. Direction de l'animation de la recherche, des études et des statistiques (DARES)
 - Echantillon national témoin représentatif des personnes diabétiques (ENTRED)
- Avec examen de santé et/ou prélèvements biologiques
 - L'Étude nationale nutrition santé (ENNS) par l'Unité de surveillance et d'épidémiologie nutritionnelle (USEN)
 - Enquête sur la santé et les comportements alimentaires en Martinique (Escal)
 - Enquêtes de séroprévalence conduites par l'InVS (séroépidémiologie vaccinale, enquêtes hépatite B et C)

Les grandes Enquêtes nationales auprès des professionnels et structures de soins :

- Enquête nationale sur les événements indésirables liés aux soins (ENEIS)
- Enquête nationale de prévalence des infections nosocomiales (ENPIN)

Autres sources de données nécessaires à la surveillance :

- Bases de données de sécurité sanitaire :
 - Agence française de sécurité sanitaire des produits de santé (AFSSAPS)

- Agence française de sécurité sanitaire des aliments (AFSSA)
- Données de ventes de tous les acteurs de l'industrie pharmaceutique, par point de ventes et par présentation produit du Groupement pour l'élaboration et la réalisation des statistiques (GERS)
- Les mesures environnementales (eau, polluants atmosphériques, etc.)
 - Associations Agréées de Surveillance de la Qualité de l'Air (AASQA)
 - Banque de données sur la qualité de l'air (BDQA)
 - Agence de l'Environnement et de la Maîtrise de l'Energie (ADEME)
- Les données issues des recensements