

**Haut Conseil de
la santé publique**

rapport

**Les systèmes d'information
pour la santé publique**

décembre 2009

**HAUT CONSEIL DE LA SANTE
PUBLIQUE**

**LES SYSTEMES D'INFORMATION
POUR LA SANTE PUBLIQUE**

NOVEMBRE 2009

Sommaire

Mandat.....	3
Composition du groupe de travail.....	4
Synthèse.....	5
Rapport.....	11
1 Introduction.....	
1.1 Contexte de la mise en place du Groupe de travail sur les systèmes d'information	11
1.2 Champ couvert et méthodes de travail.....	11
2 État des lieux.....	12
2.1 Les grandes sources de données	12
2.1.1 Les données de mortalité	13
2.1.2 Les données de morbidité	14
2.1.3 Handicap, santé fonctionnelle, qualité de vie.....	15
2.1.4 Recours aux soins	16
2.1.5 Déterminants de la santé.....	17
2.1.6 Les inégalités sociales et territoriales	19
2.2 Limites et enjeux	19
2.2.1 Les problèmes d'accessibilité et des délais de mise à disposition.....	19
2.2.2 Autres aspects transversaux.....	21
3 Recommandations	22
3.1 Recommandations spécifiques et à court terme	22
3.2 Recommandations générales.....	25
4 Annexes	35
4.1 Principales sources de données identifiées	35
4.2 Référence des principaux textes législatifs	36
4.3 Auditions	38
4.3.1 Grille d'entretien avec les personnes auditionnées.....	38

MANDAT DU GROUPE DE TRAVAIL

Les objectifs de santé annexés à la loi du 9 août 2004 relative à la politique de santé publique sont définis et évalués tous les 5 ans par le HCSP.

Choisis en raison de leur importance et destinés à fixer des engagements en termes de résultats, leur évaluation requiert pour certains d'adapter ou de structurer un système d'information approprié.

En effet, à ce jour, l'absence de dispositif de mesure de certains indicateurs associés aux objectifs, voire l'absence de définition de ces indicateurs, ne permet pas d'apprécier l'évolution de ces objectifs.

Ce groupe sera chargé de :

Contribuer à une réflexion globale sur les besoins, l'organisation et l'articulation des systèmes d'information nationaux et régionaux pour l'évaluation des objectifs de la loi de santé publique 2004 et des plans de santé publique.

Lancer la mise en œuvre, dans la suite des conclusions du groupe de travail sur les indicateurs manquants, des travaux nécessaires à la construction d'un système d'information approprié pour les objectifs non documentés.

Conduire une analyse critique, pour les autres objectifs, sur l'adéquation du système d'information au recueil des données nécessaires à leur évaluation.

Définir, pour les objectifs nouvellement proposés pour la prochaine loi par le HCSP, les caractéristiques du système d'information nécessaire.

COMPOSITION DU GROUPE DE TRAVAIL

Marcel Goldberg, HCSP, président

Chantal Cases, HCSP

Jean-Claude Desenclos, InVS

Éric Jouglu, HCSP

Annette Leclerc, Inserm

Yvon Merlière, Cnamts

Lucile Olier, Drees

Corinne Rouxel, Dares

Alain Trugeon, HCSP

Alain Weill, Cnamts

Avec la participation de

Paul Avillach, université Bordeaux 2

Gérard Badéyan, secrétariat général du HCSP.¹

¹ Les annexes (fiches de description des sources de données, comptes-rendus des auditions), seront fournies ultérieurement.

SYNTHESE CONCERNANT LES SYSTEMES D'INFORMATION POUR LA SANTE PUBLIQUE

CONTEXTE

Après l'installation du Haut Conseil de la Santé Publique en Mars 2007, un groupe de travail a été chargé de faire le point sur les indicateurs manquants. Le rapport de ce groupe a passé systématiquement en revue les objectifs n'ayant pu donner lieu à un suivi en raison de l'insuffisance des données disponibles. A la suite de ce rapport, un nouveau groupe de travail a été mis en place pour contribuer à la réflexion globale sur les besoins, l'organisation et l'articulation des systèmes d'information nationaux et régionaux dans le cadre de l'évaluation des objectifs de la loi de santé publique 2004 et des plans de santé publique.

Le champ couvert par le groupe de travail est restreint aux sources de données permettant d'établir des indicateurs nationaux dans les champs couverts par la loi et les plans de santé publique. Pour répondre à son mandat, le groupe a organisé son travail à partir des principaux problèmes qu'il a identifiés : absence ou insuffisance de recueil de données dans certains domaines, accessibilité des bases de données existantes et appariement entre bases, données localisées, articulation avec les systèmes internationaux. Il a procédé à une série d'auditions de responsables techniques d'institutions productrices et/ou utilisatrices de données administratives ou d'enquêtes : INED, CépiDc, Irdes, Insee, CNAMTS, Cnav, Inpes, MISS, Asip Santé), ainsi que de la Cnil. Ces auditions ont été menées à partir d'une grille d'entretien, réalisée par le groupe, abordant l'ensemble de ces problématiques et les questions de politique de diffusion, notamment les difficultés pratiques dans la mise en œuvre, le dispositif concret et les moyens qui y sont affectés. Le groupe de travail a également décidé de réaliser un répertoire des sources utiles pour renseigner les indicateurs de suivi, à partir d'une grille de description standard.

Le groupe de travail, présidé par Marcel Goldberg épidémiologiste à l'Inserm, est constitué de membres du HSCP (Chantal Cases, Eric Jouglu, Alain Trugeon), ainsi que de représentants d'institutions productrices et/ou utilisatrices (Jean-Claude Desenclos InVS, Annette Leclerc Inserm, Yvon Merlière et Alain Weill CNAMTS, Lucile Olier Drees, Corinne Rouxel Dares), avec la participation de Paul Avillach, assistant hospitalier universitaire en santé publique à l'université de Bordeaux et de Gérard Badéyan, secrétariat général du HCSP Il a tenu au total 9 réunions.

ÉTAT DES LIEUX : DES SOURCES NOMBREUSES, MAIS DIFFICILEMENT ACCESSIBLES ET INSUFFISAMMENT COORDONNÉES

Le groupe de travail a établi un état des lieux des grandes sources de données utilisées pour établir les objectifs de la loi relative à la santé publique et les plans de santé publique. Elles peuvent être classées en grands groupes : les sources à vocation exhaustive, le plus souvent issues de fichiers administratifs ou de gestion, notamment ceux de l'Assurance maladie ou les données de mortalité produites par l'Inserm à partir des certificats de décès ; les grandes enquêtes nationales en population générale (déclaratives ou avec examen et/ou prélèvement) ; les réseaux de surveillance à partir d'échantillons de professionnels de santé ; les grandes enquêtes nationales auprès des professionnels et structures de soins ; les grandes enquêtes européennes et internationales en population générale. Au total, les sources d'information sur la santé et ses déterminants sont nombreuses, mais elles ne suffisent pas à suivre et évaluer l'ensemble des objectifs de la loi en raison de diverses limites.

Les données de mortalité sont une source d'information essentielle en matière de santé publique et sont très largement utilisées. Elles souffrent cependant encore de certaines limites : elles sont incomplètes notamment en raison de l'absence de transmission des certificats de décès par certains instituts médico-légaux ; les délais de production sont encore trop importants pour certaines utilisations, notamment pour l'alerte, mais aussi pour l'évaluation des politiques ; la PCS ne figure plus dans la dernière version des certificats de décès.

Les données de morbidité sont nombreuses et variées, issues de différents types de sources : enquêtes en population générale, auprès des professionnels de santé, registres, Assurance maladie et PMSI... Cependant la diversité de ces sources est également une limite à la production de données systématiques et cohérentes ; il existe notamment trop peu de données issues d'examens de santé. L'analyse des inégalités sociales n'est possible que dans les enquêtes déclaratives, où la morbidité déclarée soumise à divers biais. Les données issues de systèmes de l'Assurance maladie et du PMSI ne permettent de repérer que la morbidité traitée par le système de soins. Enfin, certains territoires ne sont pas couverts par l'ensemble des sources : c'est notamment le cas des départements d'Outre-mer.

Handicap, santé fonctionnelle, qualité de vie : les sources d'information sont rares et peu fréquentes. La précédente enquête Handicap, Invalidité, Dépendance date de plus de 10 ans et la nouvelle enquête sera décennale, et son rythme ne permet pas d'évaluer les politiques de santé publique à l'échelle quinquennale, comme le prévoit la loi de 2004. Par ailleurs, les enquêtes nationales menées sur qualité de vie des patients ne couvrent pas l'ensemble des pathologies chroniques.

Recours aux soins : les données sont issues d'enquêtes en population générale, parfois appariées avec les données de remboursement de l'Assurance maladie, et de l'Assurance maladie utilisées isolément. Les premières permettent une analyse assez détaillée des déterminants individuels de la santé et des recours aux soins, mais la taille de leurs échantillons limite l'analyse des pathologies peu fréquentes et ne permet pas d'analyse à une échelle infranationale. Les secondes permettent une analyse sur données exhaustives, mais ne comportent pratiquement aucun élément sur les déterminants de la santé et du recours aux soins. Il n'en reste pas moins que leur utilisation est encore trop peu répandue et doit être développée.

Déterminants de la santé : les enquêtes en population sont assez riches en la matière, mais elles laissent de côté certains déterminants. C'est partiellement le cas pour la santé au travail, malgré l'existence de grandes enquêtes sur l'exposition aux risques, ou la production d'indicateurs d'exposition par le département Santé Travail de l'InVS. C'est pour l'exposition des populations aux nuisances environnementales que le manque de données est le plus criant. Une partie de ces manques peuvent être comblés par le repérage de la zone géographique des sujets inclus dans des enquêtes et l'utilisation de données physiques externes sur ces zones. Les caractéristiques physiques et sociales de l'habitat ne font pas l'objet de dispositifs d'observation de taille suffisante. Les déterminants professionnels et sociaux de la santé sont également documentés de façon insuffisante : les données de source médicale (Assurance maladie notamment) ne recueillent pas d'informations sur les déterminants de la santé. Par contre, la Cnav et la Cnaf disposent de données de situation socioprofessionnelle exhaustives particulièrement riches, et l'appariement de données de soins et de situation socioprofessionnelle avec des données d'enquêtes permettrait de largement documenter les déterminants professionnels et sociaux de la santé. On ne dispose pas en France d'indicateurs collectifs de l'environnement social, du type indicateurs de « défavorisation », qui devraient pouvoir être établis à des échelles territoriales suffisamment fines.

Les inégalités sociales et territoriales : on ne dispose pas d'indicateurs d'inégalités sociales et territoriales suffisamment fins, en termes à la fois de problèmes de santé spécifiques et de territoires couverts, et il manque en France un véritable « système de surveillance » pérenne des inégalités de santé.

Le groupe de travail a auditionné les principaux grands producteurs de données utilisées pour l'établissement d'indicateurs de santé sur leur politique de mise à disposition et de diffusion. De manière générale, le groupe a constaté une réelle volonté des interlocuteurs auditionnés de mettre à disposition de la recherche publique les données qu'ils produisent. Le cadre juridique permettant d'accéder à ces données est généralement clair, malgré quelques certaines difficultés ponctuelles, et les organismes chargés de les appliquer et de contrôler leur mise en œuvre ne s'opposent pas à l'accès aux données des bases médico-administratives pour la santé publique. Il est néanmoins apparu qu'en pratique, l'accessibilité à ce type de données reste un problème majeur. La diffusion de ces données pour la santé publique et la recherche ne fait en général pas explicitement partie des missions de l'institution qui les produit, et les moyens nécessaires pour mettre à disposition les données dont l'accessibilité effective est fortement contingentée.

Les efforts récents de certains organismes se heurtent à des problèmes de moyens au sein des organismes producteurs et ceux utilisateurs des données. Une autre difficulté récurrente est la surcharge de la Cnil : malgré la qualité du travail d'instruction des dossiers, le manque de moyens face à l'augmentation des demandes est très préjudiciable.

En dehors des aspects sectoriels, le Groupe de travail a identifié certains problèmes transversaux qui concernent la plupart des sources citées. On a souligné le manque de dispositif de collecte de données dans certains domaines correspondant à des objectifs de la loi. On renvoie au rapport du Groupe de travail du HCSP sur les indicateurs manquants pour une analyse détaillée de ces manques. Un autre problème concerne le calendrier des recueils de données qui ne sont pas coordonnés avec les évaluations à relativement court terme de la loi et des plans de santé publique. Par ailleurs, les délais de mise à disposition de données provenant de certaines sources permanentes sont trop longs. Concernant les aspects géographiques, il faut rappeler que les dispositions en cours de mise en place au niveau de l'Europe vont amener à réviser les modalités de réalisation de plusieurs enquêtes nationales. Enfin, le Groupe de travail souligne la quasi absence de dispositifs de recueil de données longitudinales, indispensables pour l'étude des trajectoires de santé, de soins, de situation socioprofessionnelle, etc.

PRINCIPALES RECOMMANDATIONS

Le Groupe de travail donc des recommandations pour remédier aux limites identifiées. Certaines des recommandations qui suivent sont spécifiques de certaines sources, d'autres sont plus générales et concernent la gouvernance des systèmes d'information pour la santé publique.

RECOMMANDATIONS SPECIFIQUES ET A COURT TERME

RS-1. Régler de façon urgente certains problèmes juridiques : transmission des données des Instituts médicolégaux au CépIdc, accès à l'EDP, et entreprendre une réflexion sur les textes qui encadrent l'utilisation du NIR à des fins d'appariement de données d'enquête avec des sources médico-administratives.

RS-2. Veiller à ce que la mise en place de l'Identifiant national de santé ne rende pas impossible les appariements de données avec sources utilisant le NIR comme identifiant.

RS-3. Initier ou développer certains systèmes d'information relevant d'organismes publics dans des domaines cruciaux et actuellement non couverts, notamment en médecine ambulatoire (projet PROSPERE notamment).

RS-4. Soutenir la mise en place d'une structure technique d'interface entre producteurs et utilisateurs de données médico-administratives.

RS-5. Renforcer les moyens de la CNIL dans le secteur santé.

RECOMMANDATIONS GENERALES

RG-1. Mieux coordonner les différentes actions de pilotage et de suivi des systèmes d'information pour la santé publique. Les principales missions de pilotage et de suivi des structures concernées sont :

Poursuivre l'identification des données manquantes pour les indicateurs de la loi et des Plans de santé publique et les sources potentielles.

Tenir à jour un annuaire opérationnel des sources utilisables pour documenter indicateurs.

Animer une réflexion avec les producteurs de données en vue de faire évoluer conjointement leur système d'information.

Proposer des évolutions des recueils de données existants pour les adapter aux besoins, et la mise en place de recueils nouveaux.

Veiller à la pérennité des systèmes d'information et à la répétition des enquêtes à visée nationale pertinentes.

Constituer un « guichet conseil » pour les utilisateurs.

Recommander aux pouvoirs publics une politique de développement et de maintenance des systèmes d'information pour la santé publique en coordination avec les agences sectorielles concernées (ATIH, ASIP).

Élaborer une politique pour les données à vocation régionale, en liaison avec les ARS et les ORS.

Élaborer une méthodologie et une réflexion juridique pour la production d'indicateurs à l'échelle de zones géographiques de petite taille à partir des sources exhaustives, tout en respectant la confidentialité des données à caractère personnel.

Adapter les dispositifs de recueil de données pour faciliter les comparaisons à l'échelle européenne et internationale.

RG-2. Inciter les organismes producteurs et utilisateurs de données à dégager les ressources nécessaires à une meilleure utilisation des données pour la santé publique, de façon concertée avec les professionnels et ces organismes :

Inscrire dans la loi de santé publique et dans les contrats entre l'État et les organismes de protection médicosociale producteurs de données une mission explicite de mise à disposition de données, accompagnée d'une obligation de moyens.

Inscrire dans la loi de santé publique et dans les contrats entre l'État et les organismes utilisateurs de données de santé publique une obligation de moyens concernant l'accès aux données à visée de santé publique.

Inscrire dans les conventions entre l'Assurance maladie et les professionnels de santé une obligation de fourniture de données aux organismes de santé publique concernés (agences, ORS, registres, organismes de recherche, etc.)

RG-3. Faciliter les travaux techniques et méthodologiques nécessaires pour l'accès aux données des organismes producteurs :

Développer une (ou plusieurs) structure(s) spécialisée(s) réalisant l'interface entre les producteurs et les utilisateurs fournissant des prestations complètes et dotée(s) de moyens suffisants.

Développer la recherche sur les systèmes d'information pour la santé publique (notamment : méthodes de validation de données médicales, méthodes d'appariement probabiliste, cryptage, géolocalisation).

RG-4. Faciliter l'appariement à l'échelle individuelle de sources différentes, médicales et non médicales en créant un (des) centre(s) d'appariement sécurisés autorisé(s) sous conditions à avoir accès aux identifiants nationaux (NIR, INS).

RG-5. Développer des dispositifs d'observation longitudinale en s'appuyant sur des cohortes existantes et en veillant à ce que les plans d'échantillonnage et les données collectées permettent de renseigner des indicateurs de la loi de santé publique.

RG-6. Mettre en place un dispositif de surveillance systématique et permanente des inégalités sociales et territoriales de santé, incluant des indicateurs de défavorisation à une échelle fine.

RG-7. Entreprendre une réflexion sur l'acceptabilité des recueils de données, leur gestion et leur utilisation à des fins de santé publique et de recherche, associant les producteurs et utilisateurs de données et les représentants de la société civile.

RAPPORT

INTRODUCTION

1.1. CONTEXTE DE LA MISE EN PLACE DU GROUPE DE TRAVAIL SUR LES SYSTEMES D'INFORMATION

Dès la présentation du projet de loi concernant la politique de santé publique, un groupe de travail a été mis en place par la Direction de la recherche, des études, de l'évaluation et des statistiques (Drees) et la Direction générale de la Santé (DGS) pour réaliser le travail technique de rédaction et de validation des définitions opérationnelles des indicateurs de suivi des objectifs proposés. Ce groupe de travail a été assisté par un groupe projet constitué de représentants des principales institutions productrices de données et/ou futures utilisatrices des indicateurs produits. Ce travail a abouti à un rapport méthodologique présentant ces indicateurs et les sources de données, existantes ou pour certaines en cours de réalisation, permettant de les renseigner en juillet 2005 :

http://www.sante.gouv.fr/htm/dossiers/losp/rapport_indicateurs_drees.doc

Sur cette base, des rapports de suivi des objectifs de la loi ont été produits sous l'égide de la Drees, par les différents producteurs, avec des fiches réactualisées par objectif d'année en année par suite de la mise à disposition de nouvelles sources :

<http://www.sante.gouv.fr/drees/santepop2008/santepop2008.htm>

Après l'installation du Haut Conseil de la Santé Publique (HCSP) en Mars 2007, un groupe de travail s'est réuni pour faire le point sur les indicateurs manquants. Le rapport de ce groupe a passé systématiquement en revue les objectifs ne pouvant donner lieu à un suivi pour l'instant, en distinguant les problèmes selon leur nature (indicateurs de processus liés à des programmes d'action, objectifs sans indicateur de suivi défini). Le groupe a en particulier réalisé une analyse spécifique de deux problèmes transversaux : la question des limitations fonctionnelles, restrictions d'activité et qualité de vie et celle des données sur la qualité de l'habitat (air, bruit, radon, plomb, à terme moisissures, ...) en lien avec les indicateurs socio-économiques des personnes occupant le logement.

A la suite de ce rapport, un nouveau groupe de travail a été mis en place en janvier 2009 par décision du Président du Haut conseil de la santé publique, pour contribuer à la réflexion globale sur les besoins, l'organisation et l'articulation des systèmes d'information nationaux et régionaux dans le cadre de l'évaluation des objectifs de la loi de santé publique 2004 et des plans de santé publique.

1.1. CHAMP COUVERT ET METHODES DE TRAVAIL

Le domaine couvert par le groupe de travail est restreint aux sources de données permettant d'établir des indicateurs nationaux dans les champs couverts par la loi et les plans de santé publique. Par contre, il n'a pas initié le lancement de travaux pour les objectifs non documentés, estimant que ceci dépassait ses possibilités et ne pouvait être réalisé que par un groupe à vocation technique et administratif, à l'instar de celui mis en place en 2004-2005.

Pour répondre à son mandat, le groupe a organisé son travail à partir des principaux problèmes qu'il a identifiés : absence ou insuffisance de recueil de données dans certains domaines, accessibilité des bases de données existantes et appariement entre bases, données localisées, articulation avec les systèmes internationaux.

Il a procédé à une série d'auditions de responsables techniques d'institutions productrices et/ou utilisatrices de données administratives ou d'enquêtes (Ined, CépiDc, Irdes, Insee, Cnamts, Cnav, Inpes, MISS, Asip Santé), ainsi que de la Cnil. Ces auditions ont été menées à partir d'une grille d'entretien, réalisée par le groupe, abordant l'ensemble des problématiques et des questions de politique de diffusion, notamment les difficultés pratiques dans la mise en œuvre, le dispositif concret et les moyens qui y sont affectés. Le groupe de travail a également décidé de réaliser un répertoire des sources utiles pour renseigner les indicateurs de suivi, à partir d'une grille de description standard.

Le groupe de travail, présidé par Marcel Goldberg épidémiologiste à l'Inserm, est constitué de membres du HSCP (Chantal Cases, Éric Jouglà, Alain Trugeon), ainsi que de personnes appartenant à des institutions productrices et/ou utilisatrices (Jean-Claude Desenclos InVS, Annette Leclerc Inserm, Yvon Merlière et Alain Weill Cnamts, Lucile Olier Drees, Corinne Rouxel Dares), avec la participation de Paul Avillach, assistant hospitalier universitaire en santé publique à l'université de Bordeaux II et de Gérard Badéyan, secrétariat général du HCSP. Ce groupe s'est réuni au total à 9 reprises.

ÉTAT DES LIEUX

1.2. LES GRANDES SOURCES DE DONNEES

Le groupe de travail a établi un état des lieux des grandes sources de données utilisées pour le suivi des objectifs de la loi relative à la santé publique et les plans de santé publique. Le champ couvert est restreint aux sources de données permettant d'établir des indicateurs nationaux dans les domaines couverts par la loi et les plans de santé publique. Les données doivent donc correspondre à une couverture nationale concernant l'ensemble de la population. Ceci n'implique pas obligatoirement que les données doivent être exhaustives pour la population française, ni même qu'elles nécessitent une stricte représentativité quand il s'agit de données provenant d'échantillons. Il est par contre nécessaire qu'elles aient une « vocation nationale » et permettent d'estimer de façon fiable des indicateurs nationaux. Idéalement, elles doivent également permettre l'établissement d'indicateurs à des échelles territoriales plus fines, notamment pour ce qui concerne les objectifs de la loi visant à lutter contre les inégalités territoriales.

Les sources de données peuvent être classées en grands groupes :

- les sources à vocation exhaustive, le plus souvent issues de fichiers administratifs ou de gestion, notamment ceux de l'Assurance maladie et des hôpitaux (ATIH), les données de mortalité produites par l'Inserm à partir des certificats de décès et les données issues des systèmes de déclaration obligatoire ;
- les grandes enquêtes nationales en population générale (déclaratives ou avec examen et/ou prélèvement) ; on peut également mentionner le recensement général de la population qui permet de travailler sur de nombreux indicateurs à divers niveaux géographiques et permet également des redressements d'échantillon ;
- les réseaux de surveillance à partir de données issues d'échantillons de professionnels de santé ;

- les grandes enquêtes nationales auprès des professionnels et structures de soins ;
- les grandes enquêtes européennes et internationales en population générale.

Globalement, il apparaît que les sources d'information sur la santé et ses déterminants sont nombreuses, mais elles ne suffisent pas à suivre et évaluer l'ensemble des objectifs de la loi en raison de diverses limites qu'on examinera.

1.2.1. Les données de mortalité

Le CépiDc (Inserm) produit la statistique annuelle des causes médicales de décès en France (le délai actuel de production est de deux ans) et réalise des études et recherches sur les causes de décès et sur la méthodologie de codage. Les données sont issues du certificat de décès qui indique la cause initiale et les autres causes de décès (immédiates, intermédiaires, associées) ainsi qu'un certain nombre d'informations démographiques sur le sujet décédé (sexe, âge, lieu de décès, lieu de domicile...). Le certificat de décès est régi par le code des collectivités territoriales et a été modifié par la loi de santé publique de 2004 (article 25) qui fixe les conditions de confidentialité et l'objet de la collecte (à but de santé publique) ainsi que les destinataires du certificat.

Les données produites sont une source d'information essentielle en matière de santé publique et sont très largement utilisées. Elles souffrent cependant encore de certaines limites :

- **les données sont incomplètes** en raison de l'absence de transmission des certificats de décès par certains instituts médico-légaux, ce qui induit une sous-déclaration très importante des causes de décès des jeunes, notamment en Ile-de-France, mais aussi, de manière marginale, en raison des décès non identifiés sous X (morts dans la rue). On note aussi l'absence d'information sur les décès à l'étranger. Par ailleurs, le système de recueil des données de mortalité français mis en place en 1968 n'inclut pas l'enregistrement des causes médicales de décès des enfants morts-nés, ce problème est d'autant plus important que le taux de décès foetal est particulièrement élevé en France ;
- **la profession (PCS)** était dans le passé codée par l'Insee pour l'ensemble des sujets décédés. Elle n'est disponible actuellement que pour les sujets actifs (absence d'information sur l'ancienne PCS des retraités par exemple). Ce manque est d'autant plus dommageable que l'accès pour appariement à l'échantillon démographique permanent (EDP) de l'Insee, qui permettrait l'analyse des différentiels sociaux de causes de décès, est impossible dans l'état actuel de la réglementation ;
- **les délais de production** sont encore trop importants pour certaines utilisations, notamment pour l'alerte, mais aussi pour l'évaluation des politiques. Pour y remédier, le CépiDc développe actuellement la certification électronique en ligne, à l'hôpital mais aussi en libéral. Sur la base des résultats des études pilotes, cette évolution aurait également pour conséquence d'améliorer la qualité du codage. Si l'application est bien reçue par les médecins qui l'utilisent, son déploiement est encore trop limité.

1.2.2. Les données de morbidité

Elles sont nombreuses et variées, issues de différents types de sources : enquêtes en population générale du système statistique public (Insee, Drees, Dares, etc.), systèmes de veille, de surveillance, de collecte et enquêtes des agences et instituts sous tutelle du ministère de la Santé - InVS, Inpes, ATIH, etc. - et d'organismes de recherche - Irdes et Inserm notamment - ; enquêtes auprès des professionnels de santé sur leurs clientèles ; registres de pathologies ; données issues des systèmes d'information et de gestion de l'Assurance maladie (notamment les codifications des affections de longue durée et la morbidité traitée, recalculée à partir de populations consommant certains soins traceurs)...

Cependant, malgré la diversité de ces sources, on ne dispose pas aujourd'hui de bases de données rassemblant de façon systématique et cohérente, à la fois de l'information sur la morbidité et sur les déterminants médicaux et sociaux de la santé. La plupart des données disponibles sur la morbidité en population générale sont issues d'enquêtes par interview, déclaratives et entachées de ce fait de nombreux biais et d'une imprécision de mesure. La taille des échantillons utilisés ne permet en général pas de repérer les pathologies peu fréquentes. Elle ne permet pas non plus, sauf exception, la production d'une information localisée. *A contrario*, les données issues d'examens de santé autorisent une mesure précise de la morbidité mais sont en général relatives à des populations et/ou des territoires particuliers ou de faibles échantillons. Les grandes enquêtes en population générale permettent une description sociodémographique riche de la population, tandis que les données issues d'examens de santé contiennent en général peu d'informations de ce type sur les patients, interdisant l'analyse des inégalités sociales. Les données issues des systèmes de gestion de l'Assurance maladie (SNIIR-AM) ou de l'ATIH (PMSI) pour la morbidité hospitalière fournissent des données quasiment exhaustives sur de nombreux aspects de la morbidité et à un niveau géographique fin, mais avec des limites : elles ne permettent de repérer, à travers l'analyse des traitements, de la reconnaissance d'une affection de longue durée ou de la prise en charge hospitalière, que la morbidité traitée par le système de soins. Ces bases de données n'étant pas constituées initialement dans une perspective épidémiologique posent des problèmes des questions d'interprétation souvent complexes.

De plus, ces données sont très pauvres en descripteurs sociodémographiques (ces caractéristiques sont appréhendées essentiellement, de façon très indirecte, à travers l'affiliation à la CMUC ou le lieu de la résidence). Enfin, il faut souligner que certaines populations (comme les personnes vivant en institution), et certains territoires sont particulièrement mal couverts par l'ensemble des sources : c'est notamment le cas des départements d'outre-mer, pour lesquels la connaissance est en général très insuffisante. Si la statistique publique s'efforce, suite aux recommandations du Cnis, de couvrir ces départements dans ses enquêtes, et notamment l'enquête Santé, elle n'est en général pas en mesure, pour des raisons de coût, de le faire sur des échantillons représentatifs de chacun de ces départements, alors même que la situation sanitaire est très différente d'un territoire à l'autre. Par ailleurs, certaines statistiques administratives posent des problèmes de qualité spécifiques.

Il faut noter que la production de données statistiques sur la morbidité et ses déterminants sociaux est une préoccupation qui n'est pas propre à la France. Le champ de la statistique sanitaire n'était pas jusqu'à une période récente structuré au niveau européen comme le sont déjà depuis de nombreuses années les statistiques économiques et sociales. L'adoption, en décembre 2008, d'un règlement cadre sur les statistiques de santé publique et de santé au travail vise à améliorer la production de statistiques harmonisées au niveau européen sur les causes de décès, la morbidité et les systèmes de soins. En particulier, des enquêtes nationales quinquennales harmonisées sur la santé (*European Health Interview Survey*) comportant des questions sur les caractéristiques sociodémographiques et sur les maladies et les limitations fonctionnelles seront réalisées à partir de 2014. L'un des objectifs de ces enquêtes est précisément l'analyse des inégalités sociales de santé. Par ailleurs, la direction générale de la santé de la Commission européenne (DG SANCO) finance actuellement la réalisation d'enquêtes pilotes par examen dans des pays volontaires. Compte tenu de leur coût, il n'est cependant pas prévu pour l'instant de rendre ces enquêtes obligatoires par voie de règlement.

1.2.3. Handicap, santé fonctionnelle, qualité de vie

En la matière, des efforts importants ont été engagés ces dernières années par la statistique publique pour développer l'information disponible. Suite aux recommandations du Cnis, l'Insee et la Drees ont décidé de fusionner les anciennes enquêtes décennales Handicap et Santé afin de mesurer la prévalence des différents handicaps en France, leur prise en charge, l'insertion sociale des handicapés et leur état de santé. Le volet Ménages de cette nouvelle enquête a été collecté en 2008 et les personnes en institution ont été enquêtées en 2009. Compte tenu de leur complexité et de leur coût, ces enquêtes, destinées à alimenter les travaux de recherche sur toutes les dimensions du handicap, sont réalisés à intervalle relativement long. La précédente enquête HID (handicap, invalidité, dépendance) datait de plus de 10 ans. Ce type d'enquête, essentielle pour la connaissance de ces sujets, n'est pas produit à un rythme qui permet d'en faire un outil pour le suivi et l'évaluation des lois de santé publique, comme le prévoit la loi de 2004.

Par ailleurs, des enquêtes nationales innovantes ont été menées de façon ponctuelle sur de grandes pathologies (VIH, cancer...) et ont permis d'évaluer la qualité de vie des patients. Le plan Cancer 2 prévoit la réédition de l'enquête sur la qualité de vie des patients deux ans après le cancer.

Le groupe de travail sur les indicateurs manquants du HCSP a identifié cette question parmi les sources d'information transversales à créer et pérenniser (http://www.hcsp.fr/docspdf/avisrapports/hcspr20080915_indicmanquants.pdf):

« Un objectif général (35) et toute une série d'objectifs par pathologies (64 Parkinson, 65 sclérose en plaque, de 83 à 87: troubles musculo-squelettiques, 75 BPCO) abordent la question des limitations fonctionnelles et limitations d'activité, avec des modulations dans la formulation et les relations (restrictions d'activités induites par des limitations fonctionnelles, limitations fonctionnelles et restrictions sévères...). Une mesure générale a été fournie à partir de l'enquête Santé Insee 2002-2003, ce qui a d'ores et déjà permis d'établir une fiche pour l'objectif 35. Au-delà de ce résultat, pour répondre à cette problématique de mesure des limitations fonctionnelles et des restrictions d'activité commune à plusieurs objectifs, la DGS avait lancé un appel d'offres remporté par l'Ined.

Celle-ci a donc présenté le rapport résultant de son travail au cours d'une réunion commune au groupe de travail sur les indicateurs manquants et au groupe de travail sur les indicateurs transversaux².

Ce travail fait apparaître que les pathologies examinées sont très diverses en termes de répercussion sur la santé. Il semble ainsi nécessaire de juxtaposer les indicateurs, chacun ayant son sens propre, ce qui est une réponse à la question de la justification de la présentation d'objectifs par pathologies. Le travail sur l'analyse des déterminants des limitations fonctionnelles prouve que les trajectoires pour passer de la maladie à la limitation fonctionnelle puis à l'incapacité ou au défaut de participation sociale sont diverses et varient notamment selon les pathologies.

Il semble possible de retirer de ce travail une vision générale de cadrage par grands groupes de pathologies et de présenter différents indicateurs et leurs déterminants dans une partie illustrant l'objectif 35 ainsi que dans la partie globale sur l'analyse transversale de l'état de santé. Ceci est donc fait cette année pour l'objectif 35 et peut être programmé pour l'exercice 2009 pour les indicateurs transversaux ».

Cependant pour des raisons liées à la faible prévalence de certaines pathologies ou à leur repérage qui peut nécessiter des moyens d'investigations spécifiques ou multiples, il sera difficile d'atteindre, en règle générale, le niveau des pathologies particulières. Le groupe de travail sur les indicateurs manquants concluait que « *La question de la mise en place d'une enquête généraliste sur toutes les maladies chroniques avec retentissement fonctionnel ou d'enquêtes spécifiques reste donc posée. Il conviendra d'examiner quelles réponses pourra apporter l'enquête HID-Santé actuellement en cours de collecte par l'Insee. Par ailleurs, une formulation plus cohérente des objectifs concernant les répercussions fonctionnelles de maladies chronique serait souhaitable à l'avenir.* »

Le groupe de travail sur les systèmes d'information souscrit entièrement à cette analyse.

1.2.4. Recours aux soins

Les données sont issues d'enquêtes en population générale, parfois appariées avec les données de remboursement de l'Assurance maladie (enquête ESPS de l'Irdes notamment, enquête Handicap-santé de la statistique publique), ainsi que de l'Assurance maladie utilisées isolément. Les premières permettent une analyse assez détaillée des déterminants individuels de la santé et des recours aux soins (caractéristiques sociales et économiques, comportements, à terme localisation géographique, ...), mais la taille de leurs échantillons (20 à 30 000 personnes au maximum à l'heure actuelle) limite l'analyse des pathologies peu fréquentes et ne permet pas d'analyse à une échelle infranationale.

² *Les instruments de mesure des limitations fonctionnelles dans l'Enquête santé de 2002-2003 - Évaluation de leur portée et de leurs limites à travers l'étude des répercussions des maladies chroniques sur l'état de santé fonctionnel.* Emmanuelle Cambois en collaboration avec Cyrille Cartier et Géraldine Duthé, Ined, octobre 2007.

Les secondes permettent une analyse sur données exhaustives, mais ne comportent pratiquement aucun élément sur les caractéristiques des personnes, ou les déterminants de la santé et du recours aux soins. Malgré une volonté d'ouverture de l'Assurance maladie, leur utilisation est encore limitée et assez complexe (voir ci-dessous).

1.2.5. Déterminants de la santé

Les déterminants de la santé d'une population sont nombreux et de nature diverse : biologiques et génétiques, mais aussi sociodémographiques, économiques – ce qui renvoie, mais seulement de façon partielle, aux questions d'accès financier aux soins -, comportementaux, environnementaux.... Dans ce rapport, on ne s'intéresse pas aux premiers qui relèvent d'une approche différente, mais uniquement aux déterminants sociodémographiques, économiques, comportementaux, et environnementaux qui sont l'objet des actions de santé publique.

En général, les données d'enquêtes en population sont assez riches en la matière, à la différence des autres sources de données ; cependant, comme on l'a vu, l'analyse de la morbidité associée souffre d'un certain nombre de limites. Elles laissent cependant de côté certains déterminants ce qui appelle le développement de nouvelles sources de données, ou de sources plus systématiques.

C'est partiellement le cas pour la santé au travail, avec l'existence de grandes enquêtes sur l'exposition aux risques professionnels comme Sumer, mais ignorant certains risques émergents et comportant peu de données de santé ou de déterminants individuels et encore moins d'éléments sur les trajectoires de santé et de travail. La nouvelle enquête Santé et itinéraire professionnel (SIP) a été justement conçue par la Drees et la Dares et organisée par l'INSEE, pour aborder de façon combinée les questions de conditions d'emploi et de travail et celles de santé ; elle présente l'originalité de ne pas considérer seulement les personnes ayant un emploi et comporte une dimension longitudinale, sous deux formes (les 2 vagues d'enquêtes, 2006 et 2010, et la partie rétrospective du questionnaire 2006) qui permet de mieux apprécier le sens des causalités entre état de santé et situation sur le marché du travail. Par ailleurs, suite à la dernière conférence sur les conditions de travail, un collège d'experts sur le suivi statistique des risques psychosociaux au travail a été constitué. Il est chargé de formuler d'ici fin 2010 des propositions sur la façon de faire évoluer et de compléter le dispositif d'observation statistique sur la santé au travail afin de mieux prendre en compte ces facteurs de risque. Il faut aussi noter la production d'indicateurs d'exposition à des nuisances professionnelles par le département Santé Travail de l'InVS à partir de matrices emplois-expositions (programme Matgéné) ou de réseaux de médecins du travail (notamment pour les facteurs de risque professionnels des TMS).

C'est surtout pour l'exposition des populations aux nuisances environnementales que le manque de données est le plus criant. Une partie de ces manques pourraient être comblés par le repérage de la zone géographique des sujets inclus dans des enquêtes et l'utilisation de données physiques externes sur ces zones, mais ces techniques sont encore peu utilisées et nécessitent, pour fonctionner, des normes très strictes de codage des adresses. Une autre partie des manques d'information constatés nécessiterait la constitution de nouvelles sources de données.

C'est particulièrement le cas des caractéristiques physiques et sociales de l'habitat qui ne semblent pas faire l'objet de dispositifs d'observation suffisants. Ce point a été notamment mis en évidence par le Groupe de travail sur les indicateurs manquants du HCSP

(http://www.hcsp.fr/docspdf/avisrapports/hcspr20080915_indicmanquants.pdf)

« Les membres du groupe de travail ont pu prendre connaissance de l'existence de différents organismes, tels que l'Observatoire de la qualité de l'air intérieur. Ce dernier a réalisé dans les deux dernières années l'observation en situation d'un échantillon de 567 logements, représentatif au niveau national, avec un questionnaire d'environ 600 items. Les résultats sont en cours d'exploitation ; il semble que des études descriptives aient déjà été publiées, polluant par polluant, ainsi qu'une analyse multifactorielle. Restent à développer les études portant sur le lien des différents polluants avec la santé.

Le groupe de travail a donc constaté qu'il existait déjà une source de données assez élaborée sur la qualité de l'air intérieur, mais pas sur les autres problématiques (bruit, risques dus au plomb, ...). La collecte d'informations est extrêmement coûteuse, d'autant plus qu'outre des mesures physiques, elle inclut un questionnement comportant des caractéristiques socio-économiques des personnes habitant le logement. Il serait utile que le HCSP préconise le soutien des travaux de ce type, dont le financement n'est pas a priori pérenne. »

Le groupe de travail sur les systèmes d'information souscrit à cette analyse.

Comme on l'a déjà souligné, les déterminants professionnels et sociaux de la santé sont également documentés de façon insuffisante. On ne peut attendre des professionnels qui sont les producteurs initiaux de la plupart des données de santé, qu'ils recueillent également de façon systématique des données fiables et standardisées sur les déterminants. Il en résulte donc que les données d'origine administrative et notamment celles de l'Assurance maladie, si elles sont très riches de possibilités en matière de production d'indicateurs concernant la santé et de recherche, notamment par leur capacité à être exploitées à des niveaux géographiques fins, sont pauvres en informations permettant de documenter les déterminants de la santé. Seuls quelques déterminants sociaux (le sexe et l'âge de l'assuré, ainsi que sa couverture éventuelle par la CMUC), sont disponibles dans les bases de l'Assurance maladie. Par contre, la Cnav dispose de données de situation socioprofessionnelle exhaustives particulièrement riches pour les salariés du secteur privé, mais elles ne sont pas véritablement utilisées dans des études portant sur la santé. L'appariement de données de remboursement de soins avec des données d'enquêtes permet de documenter les déterminants professionnels et sociaux de la santé et est déjà mis en œuvre pour certaines sources de données (voir supra). L'appariement de ces informations avec d'autres sources de données administratives, comme celles de la Cnav pour les situations socioprofessionnelles, permet aussi de pallier le manque de données sociales.

Il a été expérimenté par l'Irdes et est en cours à l'Inserm dans le cadre des projets Constances et Plastico. D'autres sources de données administratives (Cnaf, données fiscales...) pourraient utilement compléter ces informations, en les traitant dans l'indispensable respect de la confidentialité.

Outre les déterminants sociaux individuels, on ne dispose pas en France d'indicateurs collectifs de l'environnement social, du type indicateurs de « défavorisation », largement utilisés dans d'autres pays, et qui devraient pouvoir être établis à des échelles territoriales suffisamment fines. Des travaux de recherche récents ont proposé de tels indicateurs, mais ils ne sont pas encore stabilisés à l'heure actuelle. Ces indicateurs de défavorisation disponibles à un niveau suffisamment fin, associés à un niveau de « géolocalisation » suffisant des cas au sein des bases de données de santé permettront de produire des indicateurs de santé à l'échelle locale en fonction du niveau de défavorisation.

1.2.6. Les inégalités sociales et territoriales

Les faiblesses constatées en matière de sources de données concernant les déterminants individuels de nature professionnelle et sociale de la santé se répercutent à l'échelle populationnelle et ne permettent pas la production d'indicateurs d'inégalités sociales et territoriales suffisamment fins, en termes à la fois de problèmes de santé spécifiques et de territoires couverts.

Les connaissances actuelles reposent sur des travaux de recherche menés à partir d'enquêtes *ad hoc*, ou de l'exploitation de certaines sources nationales de données (mortalité du CépiDC, EDP de l'Insee). Ces travaux de recherche n'ont pas, par nature, vocation à être pérennisés, et il manque en France un véritable « système de surveillance » pérenne des inégalités de santé. Les difficultés déjà signalées concernant la caractérisation de la position sociale dans les certificats de décès et l'impossibilité actuelle d'accès à l'EDP ne laissent pas présager une amélioration rapide de cette situation. De fait, un système de surveillance des inégalités de santé ne peut concrètement reposer que sur l'exploitation de sources de données nationales permanentes, et la mise en œuvre systématique de procédures d'appariements entre sources de données de santé et de données administratives et sociales incluant un niveau suffisamment fin d'analyse.

1.3. LIMITES ET ENJEUX

1.3.1. Les problèmes d'accessibilité et des délais de mise à disposition

Le groupe de travail a auditionné sur leur politique de mise à disposition et de diffusion les principaux producteurs de données utilisées pour l'établissement d'indicateurs de santé ; il a également auditionné des représentants des principaux organismes utilisateurs de telles données³. Concernant les producteurs de données, le groupe a constaté de manière générale une réelle volonté de mettre à disposition de la santé publique, de la surveillance et de la recherche les données qu'ils produisent, dans le respect de la réglementation, et notamment de la protection des

³ La grille utilisée et les comptes rendus de ces entretiens figurent en annexe du rapport.

libertés individuelles s'il s'agit de données non anonymes ou indirectement nominatives.

Le cadre juridique permettant d'accéder à ces données est généralement clair⁴. Les textes actuels, s'ils rendent parfois complexes les circuits de transmission de données ce qui induit des lourdeurs techniques (inévitables pour respecter la confidentialité des données à caractère personnel), ne constituent pas un obstacle insurmontable, et les organismes chargés de les appliquer et de contrôler leur mise en œuvre (Cnil, IDS, CCTIRS essentiellement) ne s'opposent pas à l'accès aux données des bases médico-administratives pour la santé publique⁵. Ce constat global ne doit pas masquer certaines difficultés réelles :

- Comme on l'a déjà signalé, le cadre juridique actuel semble être l'obstacle interdisant l'accès effectif aux données de l'EDP de l'Insee, qui pourtant a été dans le passé une des principales sources de données sur les inégalités sociales de santé en France.
- L'interdiction de fait de pouvoir utiliser le NIR dans certaines situations⁶ rend impossible l'appariement de données d'enquêtes avec les sources nationales de données sociales et de santé.
- Le Groupe de travail s'est inquiété des conséquences potentielles de la mise en place de l'Identifiant national de santé qui pourrait aboutir à rendre impossible l'appariement de données de santé avec d'autres sources, pourtant considéré par le Groupe de travail comme une voie particulièrement privilégiée pour améliorer la situation sur de nombreux points. Les responsables de la MISS et de l'Asip-Santé ont été rassurants à cet égard lors de leur audition (il est envisagé de maintenir une table de correspondance entre NIR et INS, qui serait confié à la Cnav dans le cadre du RNIAM), mais ce dossier doit être suivi avec attention afin d'éviter des décisions aux conséquences irrémédiables.

Il est apparu qu'en pratique, l'accessibilité des données reste un problème majeur (voire insurmontable, pour nombre d'équipes), du fait de difficultés pratiques, tant du côté des organismes producteurs de données que de ceux qui ont vocation à les utiliser du fait de leurs missions.

Du côté des organismes producteurs utilisateurs de données, la diffusion de ces données pour la santé publique, la surveillance et la recherche ne fait en général pas explicitement partie des missions de l'institution qui les produit ;

⁴ Au moment de la rédaction de ce rapport, les représentants de la Cnil n'ont pas encore été auditionnés par le Groupe de travail.

⁵ On trouve en annexe la référence des principaux textes qui encadrent la collecte, la gestion et l'accès aux données de santé.

⁶ Il ne s'agit pas d'une interdiction formelle, mais la procédure d'autorisation (Décret en Conseil d'État) constitue dans les faits un obstacle infranchissable pour la plupart des équipes concernées.

en conséquence, les moyens nécessaires pour mettre à disposition les données et l'information permettant de les utiliser entrent en concurrence avec les moyens à allouer aux missions de base, et l'accessibilité effective des données est fortement contingentée. Divers modes d'organisation ont commencé, parfois très récemment, à être mis en œuvre : organisation spécifique en interne, réponse au coup par coup sans large information sur la disponibilité des données, transmission des données *via* le réseau Quételet, appels d'offres ouverts pour l'exploitation secondaire d'enquêtes nationales (enquête décennale Santé, HID, Sumer, SIP notamment), procédures d'accès à distance sur des sous-ensembles de données.... Mais dans beaucoup de cas, on se heurte à des problèmes de moyens.

Du côté des organismes utilisateurs des données, on observe le plus souvent une insuffisance des ressources consacrées à l'accès aux données médico-administratives, qui requiert des moyens techniques lourds et des compétences spécialisées qui n'ont pas été suffisamment développés. En effet, les données administratives sont par nature complexes à utiliser, et l'investissement d'entrée en la matière peut être très important, le plus souvent sans commune mesure avec les ressources des équipes potentiellement utilisatrices. L'Assurance maladie a mis en œuvre des formations spécifiques qui conditionnent l'accès à certaines données, mais là encore le nombre en est nécessairement limité par le nombre de formateurs, et la complexité des données requiert un fort investissement des utilisateurs.

Une autre difficulté récurrente est la surcharge de la Cnil. Les délais de traitement des dossiers transmis ne cessent de se rallonger, dépassant fréquemment les 4 mois, voire nettement plus selon l'expérience des membres du Groupe et de certains de leurs interlocuteurs. Tous soulignent en revanche la qualité du travail d'instruction des dossiers par la Cnil et sa bonne compréhension des enjeux, ainsi que la qualité des échanges techniques avec ses services. Ce que tous interprètent comme, là encore, un manque de moyens face à l'augmentation des demandes est très préjudiciable : il met en danger certains projets (retards de collectes, voire risques d'abandon) et pourrait même inciter certains à passer outre ou à ne pas demander d'autorisation.

1.3.2. Autres aspects transversaux

En dehors des aspects examinés ci-dessus, le Groupe de travail a identifié certains problèmes transversaux qui concernent l'ensemble du système d'information pour la santé publique, la surveillance et la recherche.

On a souligné le manque de dispositif de collecte de données dans certains domaines correspondant à des objectifs de la loi. On renvoie au rapport du Groupe de travail du HCSP sur les indicateurs manquants pour une analyse détaillée de ces manques

(http://www.hcsp.fr/docspdf/avisrapports/hcspr20080915_indicmanquants.pdf)

On doit souligner ici que le seul dispositif systématique de recueil large et représentatif de données diverses sur la santé et les procédures médico-techniques auprès des professionnels de santé est le fait d'un organisme privé, et que les conditions d'accès aux données collectées les rendent de fait pratiquement hors de la portée financière des structures publiques.

Un autre problème concerne la temporalité des recueils de données. La loi ainsi que les plans de santé publique exigent des évaluations à relativement court terme (quinquennal, voire moins), alors que certaines enquêtes permettant de renseigner les indicateurs ne sont réalisées et/ou reproduites qu'avec une fréquence supérieure (la répétition de certaines enquêtes, permettant de disposer des séries longues indispensables, n'est d'ailleurs pas toujours assurée). Par ailleurs, les délais de mise à disposition de données provenant de certaines sources permanentes ne sont pas satisfaisants pour certains besoins, comme on l'a signalé plus haut pour les causes de décès.

Concernant les aspects géographiques, il faut encore insister sur l'absence déjà signalée de couverture de l'ensemble du territoire et l'insuffisance des données à l'échelle infranationale pour certaines sources. Il convient de souligner également deux évolutions récentes qui vont avoir un impact important dans les années à venir : (1) les dispositions réglementaires en cours de mise en place au niveau de l'Europe vont amener à réviser les modalités de réalisation de plusieurs enquêtes et systèmes de recueil nationaux ; (2) le renforcement du niveau régional dans la mise en œuvre des politiques de santé de la loi HPST, avec la création des Agences régionales de santé.

Enfin, le Groupe de travail souligne l'insuffisance des dispositifs de recueil de données longitudinales, pourtant indispensables pour l'étude des trajectoires de santé, de soins, de situation socioprofessionnelle, etc. Les enquêtes Santé et Itinéraire Professionnel 2006 et 2010, ainsi que la prochaine mise en place de grandes cohortes constituées d'échantillons représentatifs d'enfants et d'adultes, contribuent à y remédier

RECOMMANDATIONS

Le Groupe de travail propose un certain nombre de recommandations pour remédier aux manques et limites identifiées. Certaines sont spécifiques de certaines sources, d'autres sont plus générales et concernent la gouvernance des systèmes d'information pour la santé publique.

1.4. RECOMMANDATIONS SPECIFIQUES ET A COURT TERME

Certains des problèmes identifiés par le Groupe de travail peuvent bénéficier d'améliorations substantielles dans un délai rapide, soit parce qu'il s'agisse d'aspects ponctuels, soit qu'il s'agisse de soutenir des projets déjà commencés. Ainsi, certains problèmes très spécifiques pourraient trouver une solution rapide, les points concernés ne semblant pas poser de difficultés juridiques majeures.

Il s'agit d'harmoniser les textes qui encadrent la transmission des données collectées par les Instituts médicolegaux au CépiDc, ainsi que l'accès à l'Échantillon démographique permanent (EDP) de l'Insee : le fait que la plupart des Instituts médicolegaux transmettent actuellement leurs données au CépiDc, et que l'EDP ait déjà été utilisé dans les années passées par des équipes de l'InVS et de l'Inserm sans que cela ait suscité le moindre problème montre bien qu'il s'agit plus d'un problème de forme que de fond.

De même, il n'existe pas de bonnes raisons pour que le CépiDc ne puisse être destinataire des causes médicales des morts nés. En effet, le CépiDc utilise depuis 1997 un certificat néonatal conforme au modèle de l'OMS, qui permettrait d'enregistrer les causes médicales de décès des enfants mort-nés selon les recommandations de l'OMS (mort-nés à partir de 22 semaines d'âge gestationnel ou de poids de naissance supérieur à 500 g) : pour cela, il suffirait d'ajouter au certificat néonatal actuel une case demandant au médecin de préciser si l'enfant est né mort ou vivant.

Finalement, quelques aménagements de textes, qui ne supposent pas d'évolution de la loi, pourraient débloquent certaines situations.

RS-1. Régler de façon urgente certains problèmes juridiques : transmission des données des Instituts médicolégaux au CépiDc, accès à l'EDP, causes médicales des morts nés.

*

* *

La France dispose, par rapport à la plupart des pays, de deux avantages qui peuvent être décisifs pour la construction d'un système d'information performant pour la santé publique : (1) l'existence de grandes bases de données nationales centralisées gérées par des organismes publics, couvrant la totalité de la population dans divers domaines stratégiques : recours aux soins, hospitalisation, handicaps, prestations et position sociales ; (2) l'existence d'un identifiant individuel unique (le NIR), qui est utilisé (directement ou sous forme anonymisée et irréversible) par toutes les bases de données nationales. Comme cela a été souligné, des progrès majeurs pourraient être réalisés de manière très efficiente en systématisant les appariements individuels de données provenant de ces diverses bases, ou de données d'enquête avec ces sources. Les difficultés sont de divers ordres : problèmes techniques et méthodologiques liés à la complexité des données et à certaines des limites de celles-ci qu'on a examinées plus haut ; problèmes éthiques et juridiques, le législateur ayant à juste titre veillé à ce que de tels appariements ne puissent pas porter préjudice aux libertés individuelles et à la confidentialité des données à caractère personnel. Notamment, la décision récente de créer un Identifiant national de santé distinct du NIR, dictée par cette préoccupation, pourrait rendre à terme réellement impossible certains appariements.

On se trouve aujourd'hui dans une situation contradictoire à divers égards. D'un côté, il est parfois possible d'apparier individuellement des données d'enquête provenant de bases de données nationales, lorsque le dispositif d'enquête permet que le NIR soit « naturellement » présent dans le flux de données : la Cnil autorise de tels appariements lorsqu'ils sont justifiés, moyennant bien évidemment de sévères contraintes de confidentialité et de sécurité. Par contre, lorsque le dispositif d'enquête repose sur des sources qui n'ont pas accès au NIR (c'est le cas notamment de nombreuses enquêtes en population), les textes ne permettent pas de fait de procéder à des appariements (il faut en effet un décret en Conseil d'État pour avoir l'autorisation de recueillir et d'utiliser le NIR, seul identifiant actuellement disponible dans les sources médico-administratives). De nombreux systèmes de surveillance ou études sont ainsi impossibles à réaliser (même lorsqu'ils sont diligentés par les pouvoirs publics).

Une solution théoriquement possible pour procéder à des appariements sans disposer du NIR est l'utilisation de méthodes d'appariement « probabiliste », reposant sur l'utilisation simultanée pour chaque sujet de variables non identifiantes. C'est la méthode actuellement utilisée pour l'accès aux causes médicales de décès, puisque le volet médicalisé du certificat de décès qui est transmis au CépiDc est anonymisé. Cependant les solutions à mettre en œuvre pour des bases de données plus complexes sont beaucoup plus difficiles, et en tout état de cause nécessiteraient des développements méthodologiques et pratiques (incluant un accès aisé aux bases de données concernées) qui ne permettent pas d'envisager ce type de solution à court terme.

Le Groupe de travail considère qu'il convient d'entreprendre dès maintenant une réflexion sur l'utilisation du NIR ou d'identifiants anonymisés qui peuvent en découler, incluant les perspectives liées au déploiement de l'Identifiant national de santé. Cette réflexion doit concerner les aspects juridiques, éthiques, méthodologiques et techniques, ainsi que les ressources nécessaires, et certainement associer la Cnil, le CCNE, les organismes producteurs et utilisateurs de données de santé et les représentants des citoyens.

RS-2. Entreprendre une réflexion sur les textes qui encadrent l'utilisation du NIR, afin de faciliter l'appariement de données provenant de diverses sources médico-administratives, ainsi que de données d'enquête avec ces sources ; veiller à ce que la mise en place de l'Identifiant national de santé ne rende pas impossible les appariements de données avec des sources utilisant le NIR comme identifiant.

*

* *

La connaissance des pratiques des professionnels de santé et des comportements des patients vis-à-vis du système de soins sont des éléments essentiels. La gestion du système de soins nécessite notamment des données concernant les prescriptions et les pathologies qui ont motivé la visite auprès des médecins libéraux (âge, sexe, diagnostic posé, prescriptions, effets attendus...). Or, en dehors des données de gestion de l'Assurance maladie qui ne donnent qu'une image très partielle des pratiques et comportements, on ne dispose pas en France de dispositif relevant d'organismes publics permettant de documenter systématiquement ce domaine et d'en analyser les évolutions. La seule source existante à large échelle est un organisme privé à but lucratif, et les tarifs pratiqués rendent de fait inaccessibles ces données qui sont donc en pratique réservées aux industriels. Cet état de fait est difficilement acceptable et il serait important de développer des systèmes d'information relevant d'organismes publics ou accessibles à ceux-ci. Il existe actuellement des expériences dans ce domaine, qui doivent être soutenues.

Comme l'a souligné le Groupe de travail sur les indicateurs manquants (http://www.hcsp.fr/docspdf/avisrapports/hcspr20080915_indicmanquants.pdf) d'autres domaines, comme les caractéristiques physiques et sociales de l'habitat, sont très mal couverts par les dispositifs d'observation, et il est nécessaire d'en dresser une liste raisonnablement complète.

RS-3. Initier ou développer certains systèmes d'information relevant d'organismes publics dans des domaines cruciaux et actuellement non couverts, notamment en médecine ambulatoire.

*

* *

Beaucoup, sinon la plupart, des systèmes d'information utiles pour la santé publique, la surveillance et la recherche font appel à la collecte ou la consultation de données à caractère personnel. La législation française impose dans ces cas l'obtention de l'autorisation préalable de la Cnil, d'autant plus que presque toujours on a affaire à des données qualifiées de « sensibles ». On a souligné plus haut la qualité du travail d'instruction des dossiers par la Cnil et sa bonne compréhension des enjeux, ainsi que la qualité des échanges techniques avec ses services qui instruisent les dossiers en rapport avec la santé. Mais il est clair que ceux-ci sont notablement sous dimensionnés face à l'augmentation des demandes, et le délai de traitement des dossiers dépasse très fréquemment le délai légal (qui est de 2 mois, l'absence de réponse au-delà de ce délai signifiant depuis la dernière révision de la loi un refus), ce qui ne manque pas d'entraîner des difficultés diverses, aux conséquences parfois très lourdes. On peut de plus penser que l'élargissement des conditions dans lesquelles il sera possible d'accéder à des sources de données gérées par les organismes de protection socio sanitaire va multiplier les demandes en provenance d'utilisateurs diversifiés, contribuant ainsi à augmenter la charge de travail de la Cnil. Il est donc crucial que la Cnil, qui constitue un point de passage obligé de tout dispositif de recueil de données de santé à caractère personnel, voie ses moyens dédiés aux recueils de données de santé renforcés.

RS-4. Renforcer les moyens de la Cnil dans le secteur santé.

1.5. RECOMMANDATIONS GENERALES

Le bilan réalisé par le Groupe de travail du HCSP sur les systèmes d'information pour la santé publique a permis de mettre en évidence un certain nombre de lacunes, qui complète les constats du Groupe de travail sur les indicateurs manquants

(http://www.hcsp.fr/docspdf/avisrapports/hcspr20080915_indicmanquants.pdf)

Il est d'ailleurs important de rappeler que la motivation qui a présidé à la mise en place de ces Groupes de travail relevait de l'impossibilité constatée par le HCSP de remplir complètement sa mission d'évaluation des objectifs de la loi relative à la santé publique de 2004, du fait de l'insuffisance, voire de l'absence, de données pertinentes et de validité suffisante (donc des systèmes d'information correspondant), dans de nombreux domaines.

Il est donc indispensable d'améliorer la situation afin d'éviter, au moins partiellement, de rencontrer des difficultés de même nature lors de l'évaluation de la future loi de santé publique. Ceci implique que le travail réalisé par les Groupes de travail du HCSP soit poursuivi régulièrement de façon organisée, ainsi que des mesures « structurantes » relevant d'une gouvernance générale des systèmes d'information pour la santé publique, afin de dégager des solutions pérennes, et ne pas se contenter de simplement tenter de remédier de façon ponctuelle à tel ou tel problème spécifique, afin de disposer d'un système stable et performant.

Il faut tout d'abord procéder à une mise à jour régulière de la description des sources utilisables pour documenter les indicateurs, et en parallèle rendre permanente et systématique l'activité d'identification des données manquantes pour les indicateurs des lois et des Plans, afin de permettre une meilleure réactivité pour anticiper sur les besoins futurs d'information pour l'évaluation des politiques de santé publique. Le repérage des données manquantes doit s'accompagner d'une réflexion sur les sources de données potentiellement mobilisables pour remédier aux insuffisances constatées.

Dans la première partie de ce rapport, on a souligné la diversité des sources de données utilisables pour l'établissement des indicateurs nécessaires pour l'évaluation des objectifs de santé publique. Les organismes responsables de ces sources produisent des données à des fins propres (gestion, surveillance, recherche), et le bilan effectué par le Groupe de travail met en évidence que de ce fait les données disponibles ne permettent pas toujours (ou difficilement) de répondre aux besoins de l'évaluation des objectifs de la loi. Or, dans de nombreuses situations, des modifications ou ajustements limités ou modérés suffiraient pour rendre les données produites mieux utilisables pour la documentation des objectifs de santé publique. Ceci est vrai pour certaines sources, et encore plus si on considère l'ensemble de celles-ci du fait de l'absence de cohérence d'ensemble du dispositif public de production de données utiles pour la santé publique. Il apparaît donc important d'initier une réflexion impliquant les principaux producteurs de données, avec l'objectif de faire évoluer leurs systèmes d'information en y intégrant des missions de documentation des indicateurs de la loi et des plans.

À côté de la nécessaire évolution des sources permanentes de données, il est indispensable de susciter la mise en place des nouveaux recueils de données dans les domaines mal ou non couverts qui ont été identifiés, et pour cela identifier les organismes potentiellement mobilisables et réunir les ressources de compétence et de moyens pour prendre en charge ces recueils. Dans cette optique, il est également important de veiller à la pérennité des systèmes d'information et à la répétition des enquêtes à visée nationale pertinentes : en effet, chaque organisme producteur de données poursuivant ses propres objectifs et tenant compte de ses besoins propres, peut être amené à renoncer à certaines collectes de données, ou les faire évoluer d'une façon qui peut mettre en danger le suivi de l'évolution de phénomènes d'intérêt pour la santé publique. Dans cette optique, les organismes et agences sectorielles concernées doivent mettre en œuvre une politique coordonnée de développement, de déploiement et de maintenance des systèmes d'information pour la santé publique. Ceci est d'autant plus indispensable que les dispositions réglementaires en cours de mise en place au niveau de l'Europe vont amener à réviser les modalités de réalisation de plusieurs enquêtes et systèmes de recueil nationaux, comme on l'a signalé. Il est donc nécessaire d'intégrer ces contraintes dans une réflexion générale sur l'évolution des systèmes d'information afin d'adapter les dispositifs de recueil de données pour faciliter les comparaisons à l'échelle européenne et internationale.

Il faut également anticiper les besoins de données en appui de la réorganisation des structures de santé publique déconcentrées, avec notamment la création des ARS qui se voient confier des missions importantes en termes de politiques de santé à l'échelle régionale et infrarégionale. Il est donc nécessaire d'élaborer une politique pour les données à vocation régionale, en liaison avec les ARS et les ORS, incluant une réflexion sur la remontée homogène d'informations provenant des niveaux régional et infra régional, ainsi que sur le retour d'informations vers ces niveaux.

Cette réflexion doit inclure la production d'indicateurs à l'échelle de zones géographiques de petite taille à partir des sources exhaustives, tout en respectant la confidentialité des données à caractère personnel. Les collectivités territoriales, de tailles très variables, sont en effet demandeuses, à juste titre, de données à une échelle pertinente adaptée aux responsabilités de gestion que leur confère la loi. Ceci pose cependant des problèmes méthodologiques, notamment lorsqu'il s'agit de territoires de petite taille et de faible effectif de population : en effet, il faut tenir compte de problèmes de représentativité, de fluctuations d'échantillonnage, etc., qui peuvent être complexes ; la Fnors a pris l'initiative d'un travail visant à mettre au point un guide de recommandations sur les données pour les territoires infra départementaux, mais des travaux de recherche restent indispensables dans ce domaine. La production de certains indicateurs de santé à une échelle territoriale très fine pose également des problèmes juridiques et de sécurité vis-à-vis de la loi Informatique et libertés, qui interdit la diffusion de données pouvant être identifiantes dans ces conditions et mettre en danger la confidentialité des données à caractère personnel.

Au total, on constate qu'il n'existe pas actuellement de politique concertée permettant la mise en cohérence d'ensemble des systèmes d'information pour la santé publique, qu'il s'agisse de la production de données permanente et exhaustive par des agences et organismes nationaux, ou de dispositifs plus ciblés reposant sur des enquêtes *ad hoc*. Le Groupe de travail recommande donc très fortement la mise en place d'une réflexion très large, impliquant les pouvoirs publics concernés, les organismes de protection sanitaire et sociale et les agences sanitaires ayant des responsabilités dans la production et l'utilisation de données pour la santé publique. Le paysage est en effet complexe, les organismes concernés étant nombreux et indépendants, avec des tutelles parfois différentes. Le Cnis, malgré sa mission générale de pilotage et de coordination des dispositifs de statistique publique, ne s'est pas jusqu'à présent, saisi de façon systématique du domaine de la santé publique. Une réelle volonté politique de mise en cohérence et de coordination des acteurs et des différentes actions de pilotage et de suivi des systèmes d'information pour la santé publique est donc indispensable pour améliorer durablement la situation actuelle, et faire en sorte de disposer de systèmes d'information cohérents, à la fois flexibles et stables.

L'ensemble de ces considérations est résumé ci-dessous sous la forme de mesures que le Groupe de travail recommande.

RG-1. Mieux coordonner les différentes actions de pilotage et de suivi des systèmes d'information pour la santé publique. Les principales missions des structures concernées seraient de :

- Poursuivre l'identification des données manquantes pour les indicateurs de la loi et des Plans de santé publique, et des sources potentiellement mobilisables.
- Tenir à jour un annuaire opérationnel des sources utilisables pour documenter les indicateurs.
- Animer une réflexion avec les producteurs de données en vue de faire évoluer conjointement leur système d'information.
- Proposer des évolutions des recueils de données existants pour les adapter aux besoins et mettre en place des recueils nouveaux.

- Veiller à la pérennité des systèmes d'information et à la répétition des enquêtes à visée nationale pertinentes.
- Recommander aux pouvoirs publics une politique de développement et de maintenance des systèmes d'information pour la santé publique en coordination avec les organismes et agences sectorielles concernées.
- Élaborer une politique pour les données à vocation régionale, en liaison avec les ARS et les ORS.
- Élaborer une réflexion méthodologique et juridique pour la production d'indicateurs à l'échelle de zones géographiques de petite taille à partir des sources exhaustives, tout en respectant la confidentialité des données à caractère personnel.
- Adapter les dispositifs de recueil de données pour faciliter les comparaisons à l'échelle européenne et internationale.

*

* *

Un des constats récurrents du Groupe de travail est la difficulté, pour toutes les catégories d'utilisateurs, d'obtenir des données gérées par les organismes producteurs. L'expérience des membres du Groupe de travail ainsi que des personnes auditionnées montre la très faible disponibilité de ces bases de données pour la santé publique, la surveillance ou la recherche. Ainsi, le potentiel de ces bases est-il sous utilisé, alors même que les données sont d'un grand intérêt et que le nombre d'utilisateurs potentiels et les besoins sont en augmentation constante. Les récents efforts de certains organismes pour faciliter l'exploitation des données qu'ils collectent sont à souligner : la CnamTS a déployé un dispositif facilitant l'accès à l'Échantillon généraliste des bénéficiaires (EGB), des appels d'offres ouverts très fructueux ont permis à de nombreuses équipes d'exploiter selon leurs intérêts propres de grandes enquêtes nationales (enquête décennale Santé, HID, Sumer, SIP).

Cependant, la situation reste très restrictive. Comme on l'a signalé, cet état de fait n'est pas à mettre au compte de la mauvaise volonté des organismes producteurs de données, mais est avant tout due à l'insuffisance des ressources consacrées par ces organismes à la diffusion et à la mise à disposition des structures de santé publique, de surveillance et de recherche des données dont ils assurent la collecte et la gestion. En effet, leurs missions, sauf exception, n'incluant pas de telles activités, les moyens qui y sont consacrés sont sous dimensionnés, voire pratiquement inexistantes, face à la complexité des bases de données et la lourdeur des procédures d'extraction, de contrôle et de sécurisation.

On observe une situation comparable du côté des organismes utilisateurs, qui ne consacrent pas non plus, sauf exception, suffisamment de moyens à ces activités : il ne suffit généralement pas, en effet, de se contenter de recevoir un fichier informatique. Un travail, qui peut être très lourd, de contrôle et de validation, de mise en forme et de synthèse des données brutes est indispensable avant leur exploitation, toutes activités qui demandent des personnels de niveau suffisant,

formés aux bases de données nationales et suffisamment pérennes pour justifier un important investissement initial et continu. De ce point de vue, on peut se référer à l'expérience de certains pays, qui ont mis en place des ressources adéquates, comme par exemple le *Population Data Center* de l'Université de Colombie Britannique.

Enfin, il faut aussi souligner les difficultés rencontrées pour la transmission par les professionnels de santé de données aux organismes de santé publique concernés (agences, organismes de recherche, ORS, registres, demain ARS, etc.), qui ont parfois considérablement perturbé l'activité de certains de ces organismes. En effet, de nombreux professionnels de santé considèrent que la transmission des données qu'ils recueillent auprès de leurs patients ne fait pas partie de leurs missions, et certains exigent une rémunération spécifique pour ce type d'activité.

Il est d'une grande importance d'explicitier, par des voies législatives, contractuelles ou conventionnelles selon les cas, que la transmission de données de santé est une « obligation de service public », et de préciser les moyens qui doivent y être consacrés par les différents acteurs. La recommandation suivante se situe dans cette optique.

RG-2. Inciter, de façon concertée avec les professionnels de santé, les organismes et professionnels producteurs de données et les organismes utilisateurs à dégager les ressources nécessaires à une meilleure utilisation des données pour la santé publique, la surveillance et la recherche.

- Inscrire dans la loi de santé publique une mission explicite de mise à disposition de données, accompagnée d'une obligation de moyens dans les contrats entre l'État et les organismes de protection médicosociale producteurs de données.
- Inscrire dans les contrats entre l'État et les organismes utilisateurs de données de santé publique (agences, organismes de recherche, ORS, registres, ARS, etc.) une obligation de moyens concernant l'accès aux données à visée de santé publique.
- Inscrire dans les conventions entre l'Assurance maladie et les professionnels de santé une obligation de fourniture de données aux organismes de santé publique concernés (agences, organismes de recherche, ORS, registres, ARS, etc.).

*

* *

On utilise encore trop peu en France les possibilités offertes par les bases de données alimentées par les organismes de protection sociale et médicale, qui contiennent des données concernant la santé ou la vie socioprofessionnelle. L'utilisation de ces bases offrirait en théorie des avantages très importants, en termes méthodologiques (exhaustivité, absence de certains biais de sélection et d'information), et opérationnels (données déjà recueillies, effectifs immenses, etc.).

Les principales bases de données concernées sont, outre le CépiDc pour la mortalité, celles du PMSI (Programme de Médicalisation du Système d'Information) qui, pour chaque épisode d'hospitalisation, enregistre des diagnostics codés selon la CIM 10 et les actes pratiqués codés selon la Classification Commune des Actes Médicaux (CCAM), et celles de l'Assurance maladie réunies au sein du *Système national d'information inter régimes de l'assurance maladie* (SNIIR-AM). Les données du SNIIR-AM incluront à terme tous les régimes de l'Assurance maladie : Cnamts, MSA, RSI et les 16 autres régimes spéciaux, et concernent aussi bien la médecine de ville que les hospitalisations. Elles contiennent notamment les affections de longue durée (ALD) codées selon la CIM-10, et les données de remboursements de soins auprès des professionnels de santé et de médicaments. Il faut aussi noter que la CNSA met également en place un système centralisé de remontées d'information concernant les prestations pour handicap.

Concernant les données socioprofessionnelles, les bases de données de la Caisse nationale d'assurance vieillesse (Cnav) permettent de tracer pour chaque individu dès l'âge de 16 ans et jusqu'à la liquidation de ses droits à la retraite, ses différentes périodes d'activité : périodes d'activité professionnelle ou assimilées (chômage, maladie, maternité ou congés parentaux, ...). On peut donc avoir accès à l'ensemble des données inhérentes à la carrière des assurés du Régime Général, y compris les données concernant d'éventuelles périodes effectuées dans d'autres régimes de base, ainsi que dans certains régimes particuliers ou spéciaux. En l'absence de données sur la position sociale dans les sources de données de santé, l'extraction de données des bases de la Cnav et/ou leur appariement avec des données de santé permettrait notamment de contribuer à documenter le problème des inégalités sociales de santé, qui est un des thèmes majeurs de la future loi de santé publique.

Cependant, un important travail est nécessaire pour définir les procédures d'accès, de transmission sécurisée, de vérification de cohérence et de complétude, de maintien de l'intégrité des données. Celles-ci sont complexes et nécessitent des moyens lourds et des compétences spécialisées pour permettre l'utilisation de ces bases de données dans des conditions compatibles avec des contraintes de qualité suffisante. L'obstacle majeur à l'accès aux bases de données médico-administratives est donc la lourdeur des procédures techniques à mettre en œuvre du fait de la complexité des bases et des contraintes de confidentialité et de sécurité. Les ressources indispensables n'ont pas été mobilisées, ni du côté des organismes producteurs ni de celui des utilisateurs. La mise en place d'une structure scientifique et technique jouant un rôle d'interface entre les utilisateurs et les bases de données est sans doute la meilleure solution. Cette structure, de type « plateforme » incluant une fonction de conseil pour les utilisateurs, aurait comme mission de faciliter l'accès aux données individuelles et l'appariement de données individuelles issues de bases de données différentes. Il pourrait s'agir d'une structure inter-organismes, associant notamment les organismes gérant les bases de données nationales et les organismes utilisateurs.

Par ailleurs, une meilleure utilisation des bases de données implique le développement de travaux méthodologiques pour faciliter l'accès aux données et améliorer leur validité. Les méthodes d'appariement, de cryptage, de validation des données médicales sont en constante évolution et peuvent encore être améliorées pour une meilleure efficacité. Les méthodes de géocodage sont également évolutives.

RS-3. Développer une structure technique d'interface entre producteurs et utilisateurs de données médico-administratives, incluant un « guichet conseil » pour les utilisateurs, et développer les travaux méthodologiques nécessaires pour faciliter et améliorer l'accès et l'analyse des données pour la santé publique, la surveillance et la recherche.

- Développer une (ou plusieurs) structure(s) spécialisée(s) réalisant l'interface entre les producteurs et les utilisateurs fournissant des prestations complètes et dotée(s) de moyens suffisants.
- Développer la recherche sur les systèmes d'information pour la santé publique (méthodes de validation de données médicales, méthodes d'appariement probabiliste, cryptage, géolocalisation).

*
* *

On a souligné la situation juridique paradoxale qui aboutit à ce que dans certaines conditions le NIR puisse être utilisé pour procéder à des appariements avec des sources médico-administratives où les sujets sont identifiés par leur NIR, directement ou indirectement (*via* une fonction d'occultation du NIR), alors que dans d'autres cas cela est en pratique interdit par les textes. Schématiquement, lorsque le circuit de recueil de données inclut un organisme (public ou privé) qui dispose du NIR, il est alors possible (moyennant des flux qui habituellement impliquent un tiers de confiance) d'accéder aux bases de données reposant sur le NIR, d'extraire des données individuelles et de les apparier avec d'autres sources. La Cnil a déjà autorisé plusieurs opérations de ce type. Par contre, lorsque le circuit de recueil de données n'inclut pas un organisme habilité à disposer du NIR, un décret en Conseil d'État est nécessaire pour avoir le droit d'utiliser le NIR.

On comprend bien l'esprit de la loi Informatique et libertés qui a cherché à encadrer de façon stricte l'utilisation d'un identifiant aussi sensible que le NIR. Cependant, les textes actuels constituent un obstacle insurmontable pour de nombreuses études de santé publique, alors même qu'il est parfois, selon le dispositif de recueil de données, possible d'utiliser le NIR, sans qu'il soit évident de justifier cette différence.

Une solution, qui a déjà été proposée, notamment par l'Insee⁷, serait la création d'un ou plusieurs « centres d'appariement sécurisé » (CAS). Il s'agirait d'une structure jouant un rôle de tiers de confiance habilitée à recevoir, dans le cadre d'études autorisées, les NIR des sujets concernés.

⁷ Chaleix M, Lollivier S. Outils de suivi des trajectoires des personnes en matière sociale et d'emploi. Insee, N° 98/B010, Juin 2004.

L'application par le CAS de procédures d'anonymisation et de cryptage et des flux sécurisés pourrait alors permettre des opérations d'appariement dans des conditions garantissant le respect des contraintes de confidentialité des données à caractère personnel. Une analyse approfondie des conditions techniques et légales permettant la mise en place de telles structures devrait être entreprise, en collaboration avec la Cnil. En tout état de cause, il est vraisemblable que la création de centres d'appariement sécurisé nécessite une modification de la loi Informatique et libertés.

Par ailleurs, le déploiement prévu de l'Identifiant national de santé (INS), destiné à être le seul identifiant utilisé pour les applications en santé, rendrait impossible à l'avenir l'appariement de données de santé avec d'autres sources utilisant le NIR comme identifiant, ce qui, bien évidemment, poserait des problèmes insurmontables. Afin d'éviter cette situation, il est envisagé que la Cnav maintienne une table de correspondance NIR-INS. Dans ce contexte, si des centres d'appariement sécurisé sont créés, il faudra qu'ils soient également habilités à recevoir également les INS des sujets inclus dans les études impliquant un appariement avec des sources utilisant le NIR.

RG-4. Faciliter l'appariement à l'échelle individuelle de sources différentes, médicales et non médicales en créant un (des) centre(s) d'appariement sécurisés autorisé(s) sous conditions à avoir accès aux identifiants nationaux (NIR, INS).

*

* *

Des dispositifs d'observation longitudinale à l'échelle individuelle sont nécessaires pour recueillir certaines informations concourant à l'établissement de plusieurs indicateurs de la loi relative à la politique de santé publique. De telles observations sont indispensables pour l'étude des trajectoires de santé, de soins, de situation socioprofessionnelle ou de comportements individuels. Si les épidémiologistes ont mis en place, parfois depuis longtemps de nombreuses cohortes, celles-ci sont pour la plupart destinées à répondre à des questions de recherche spécifiques, et n'ont pas d'objectifs de représentativité ; les données recueillies dans de tels contextes méthodologiques ne peuvent donc pas être utilisés pour l'établissement d'indicateurs visant à décrire la population française.

Il existe quelques rares dispositifs longitudinaux reposant sur des données d'origine administrative dont le plan d'échantillonnage vise à assurer la représentativité, comme l'Échantillon démographique permanent de l'Insee ou l'Échantillon généraliste des bénéficiaires de l'Assurance maladie. C'est également le cas de l'enquête Santé et itinéraire professionnel auprès des personnes qui aborde de façon les questions de conditions de travail, d'emploi et les questions de santé. Mais ces dispositifs présentent des limites en terme de données recueillies et ne suffisent pas pour répondre à l'ensemble des objectifs de la loi. Bien entendu, il n'est pas réaliste de penser qu'une cohorte, si large soit-elle, puisse répondre à l'ensemble des besoins d'information sur la population française. Cependant, la mise en place prochaine de nouvelles grandes cohortes, notamment dans le cadre du « TGIR (Très grandes infrastructures de recherche) Cohortes », peut être une occasion de compléter les sources de données disponibles.

C'est notamment le cas des cohortes Elfe qui doit inclure un échantillon représentatif de naissances, Constances et Coset qui suivront des échantillons représentatifs d'adultes. Il importe de veiller à ce que l'échantillonnage et les recueils de données tout au long du suivi des sujets de ces dispositifs permettent l'établissement d'indicateurs utiles pour l'évaluation des objectifs de la loi.

RG-5. Développer des dispositifs d'observation longitudinale en s'appuyant sur des cohortes existantes ou à initier, et en veillant à ce que les plans d'échantillonnage et les données collectées permettent de renseigner les indicateurs de la loi de santé publique.

*
* *

Les connaissances concernant les inégalités sociales et territoriales en France reposent actuellement sur des travaux de recherche qui n'ont pas vocation à être reproduits, encore moins pérennisés, et on a fait le constat de l'absence d'un véritable système de surveillance des inégalités sociales et territoriales de santé (cf. conclusions du Groupe de travail Inégalités sociales du HCSP). Du fait des limites des systèmes d'information de nature médicale en matière de données sur les déterminants socioprofessionnels, le Groupe de travail recommande donc d'élaborer un système de surveillance des inégalités de santé pérenne reposant sur la mise en œuvre systématique de procédures d'appariements entre sources nationales permanentes de données de santé et de données administratives et sociales, essentiellement celles gérées par la Cnav. D'autres sources devraient également être explorées, comme celles du fisc.

Un autre manque relevé par le Groupe de travail concerne les indicateurs géographiques de « défavorisation », largement utilisés dans d'autres pays à des fins de santé publique, notamment pour la priorisation des actions, ou d'épidémiologie sociale quand il s'agit de prendre en compte des effets contextuels. On ne dispose actuellement pas d'indicateurs géographiques de défavorisation « français » (c'est-à-dire construits à partir de données économiques et sociales exhaustives disponibles à une échelle géographique fine, comme les IRIS - Îlots regroupés pour l'information statistique, produits par l'Insee), qui fassent l'objet d'un consensus. Cependant, plusieurs travaux récents ont développé de tels indices, et il serait important de mettre en place une réflexion destinée à discuter des objectifs, des méthodes et des limites de ces indicateurs, afin de définir un ou plusieurs indicateurs qui pourraient faire partie d'un système de surveillance des inégalités de santé.

RG-6. Mettre en place un dispositif de surveillance systématique et permanent des inégalités sociales et territoriales de santé, incluant des indicateurs de défavorisation appropriés à une échelle fine.

*
* *

Ce rapport plaide pour une utilisation large et systématique des données d'origine médico-administrative comme une des solutions à des problèmes divers. Cependant, l'acceptabilité sociale de l'utilisation de telles sources de données pose un véritable problème. Une partie de la population la perçoit en effet comme une intrusion et une atteinte à ses libertés, comme l'ont montré les débats suscités par exemple autour du projet de la cohorte d'enfants Elfe ou du développement du Dossier médical personnel (DMP). Les restrictions que la loi impose pour l'utilisation du NIR et la décision de développer un Identifiant national de santé sont d'ailleurs des obstacles voulus explicitement pour empêcher les abus potentiels. La politique d'appariements systématiques de données provenant de plusieurs sources que propose le Groupe de travail peut engendrer des inquiétudes légitimes à cet égard.

Il serait donc d'une importance particulière de développer une réflexion sur les dispositions permettant d'encadrer l'usage de ces données (pour quels usages ? qui peut être autorisé ? dans quelles conditions ? etc.). La Conférence nationale de santé organise actuellement un débat public sur ce thème, dont les résultats devraient être diffusés et partagés de manière pédagogique. Il serait important d'envisager les suites qui y seront données.

RG-7. Entreprendre une réflexion sur l'acceptabilité des recueils de données, leur gestion et leur utilisation à des fins de santé publique et de recherche, associant les producteurs de données, les utilisateurs et les représentants de la société civile.

ANNEXES

1.6. PRINCIPALES SOURCES DE DONNEES IDENTIFIEES

Les sources à vocation exhaustive

- Données de mortalité (INSERM-CépiDc)
- Système de déclaration obligatoire (DO) de problèmes de santé à l'InVS ainsi que les Centres de Référence
- Bases de données médico-administratives
 - o Données de remboursement : SNIIR-AM
 - o Affections de longue durée (ALD) de l'Assurance maladie
 - o Le Programme médicalisé des systèmes d'information (PMSI) : MCO, psychiatrique
 - o Base de données de la CNSA
- Bases de données de sécurité sanitaire
 - o Agence française de sécurité sanitaire des produits de santé (AFSSAPS)
 - o Agence française de sécurité sanitaire des aliments (AFSSA)
- Les données de Registres qualifiés (nationaux, régional ou départemental)
- Fichier national des accidents corporels (Onisr)
- Certificats de santé de l'enfant
- GERS (par l'intermédiaire de l'InVS)

Les réseaux de surveillance à partir d'échantillons

(Professionnels de santé, laboratoires, services hospitaliers, etc.)

- Le réseau de surveillance des bactériémies nosocomiales (BN-Raisin)
- Les mesures environnementales (eau, polluants atmosphériques, etc)
 - o Associations Agréées de Surveillance de la Qualité de l'Air (AASQA)
 - o Banque de données sur la qualité de l'air (BDQA)
 - o Agence de l'Environnement et de la Maîtrise de l'Energie (ADEME)
- Réseaux de surveillance en Santé Travail de l'InVS : TMS, Samotrace (santé mentale), MCP (maladies à caractère professionnel), Programme national de surveillance du mésothéliome)

Les grandes enquêtes nationales en population générale

- **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 sur la Santé et les Consommations lors de l'Appel de Préparation à la Défense (ESCAPAD)
 - Enquête nationale périnatale (DGS/DREES/PMI/INSERM)
 - Enquêtes en milieu scolaire (DREES, Ministère de l'Éducation nationale)
 - 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)
- Échantillon national témoin représentatif des personnes diabétiques (ENTRED)
- Enquête Cocon (contraception)
- Enquête ACSF (comportements sexuels)
- Enquête de l'UFSBD (santé dentaire)

- **Avec examen de santé et/ou prélèvements biologiques**
- Étude nationale nutrition santé - ENNS (USEN-InVS)
- Enquête sur la santé et les comportements alimentaires en Martinique (Escal)

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)
- Enquête permanente sur les accidents de la vie courante

Les grandes Enquêtes européennes et internationales en population générale

- The European Study of the Epidemiology of Mental Disorders (ESEMED)
- EUROpean CANcer REGistry-based study on survival and CARE of cancer patients (EUROCARE)

1.7. REFERENCE DES PRINCIPAUX TEXTES LEGISLATIFS

Textes fondateurs de la CNIL :

Loi n° 78-17 du 6 janvier 1978 modifiée relative à l'informatique, aux fichiers et aux libertés : <http://www.cnil.fr/en-savoir-plus/textes-fondateurs/loi78-17/>

Textes d'application : Décret n°2005-1309 du 20 octobre 2005 : <http://www.cnil.fr/en-savoir-plus/textes-fondateurs/textes-dapplication/>

Convention 108 : Convention pour la protection des personnes à l'égard du traitement automatisé des données à caractère personnel : <http://conventions.coe.int/Treaty/fr/Treaties/Html/108.htm>

Charte des droits fondamentaux de l'Union européenne (2000/C 364/01) : <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2000:364:0001:0022:FR:PDF>

Directive européenne n°95/46/CE du 24 octobre 1995 du Parlement européen et du Conseil relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données : <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:FR:HTML>

SNIIR-AM

Arrêté L161-28-1 du code de la sécurité sociale du 20 juin 2005 relatif à la mise en oeuvre du Système national d'information interrégimes de l'assurance maladie :

http://www.institut-des-donnees-de-sante.fr/telechargements/ARRETE_20juin05.pdf

CNSA

Décret n° 2008-833 du 22 août 2008 portant création du Système national d'information prévu à l'article L. 247-2 du code de l'action sociale et des familles et organisant la transmission des données destinées à l'alimenter.
http://www.institut-des-donnees-de-sante.fr/telechargements/DecretCNSA_220808.pdf

IDS

Création de l'Institut des données de santé : loi portant réforme de l'assurance maladie du 13 août 2004, article 64 I.

1.8. AUDITIONS

1.8.1. Grille d'entretien avec les personnes auditionnées

1- Politique de diffusion

Votre institution a-t-elle une politique de diffusion définie en termes de publics visés, de tarification, de sécurisation, de propriété des données et des résultats produits. Cette politique s'appuie-t-elle sur des textes juridiques ? Pouvez-vous donner des exemples ? Cette politique a-t-elle évolué au cours du temps et d'autres évolutions sont-elles prévues ?

2- Difficultés pratiques

Existe-t-il des difficultés pratiques ou juridiques qui viendraient contrecarrer la volonté de mise à disposition des données ? Si oui, quelles sont-elles et quels sont les moyens mis en œuvre pour y faire face ?

3- Dispositif concret

Concrètement quelle est la procédure à suivre pour accéder aux données ? Existe-t-il un guichet dédié ou encore un service de diffusion en tant que tel ? Les demandes sont-elles traitées en interne ou par un service externalisé ?

4- Volumes et moyens

Quel est le volume de demandes et les moyens humains et financiers affectés ?

5- Appariement de données

Avez-vous à traiter des demandes nécessitant l'appariement de données en provenance de plusieurs sources de votre institution ou d'institutions multiples ? Si oui, quelles difficultés spécifiques de telles demandes engendrent-elles et comment sont-elles résolues, en termes techniques ainsi que juridiques ?

De façon plus générale sur cette question centrale une réflexion est-elle menée au sein de votre institution et avez-vous des pistes à suggérer ?

6- Données localisées

De façon générale, les bases de données ou enquêtes que vous gérez permettent-elles de produire des données régionales ou plus localisées ? Avez-vous des obligations particulières et/ou une politique en la matière (encouragement d'extensions régionales, modèles statistiques...) ?

7- Articulation avec des systèmes internationaux

Les bases de données que vous gérez s'inscrivent-elles dans des systèmes internationaux (nomenclatures, transmissions régulières régies par des textes, contraintes de format...) ?

8- Difficultés d'accès

Réciproquement connaissez-vous des difficultés pour accéder à des données d'autres institutions qui sont d'intérêt pour vous ?

9- Rapports avec la Cnil

Les systèmes d'information pour la santé publique

Dès son installation en mars 2007, le Haut Conseil de la Santé Publique, a fait le point sur le suivi des objectifs de la loi de santé publique (LSP) et a constaté que certains indicateurs restaient manquants. Un passage en revue systématique des objectifs n'ayant pu donner lieu à un suivi en raison de l'insuffisance des données disponibles a été réalisé. À la suite de ce rapport, une réflexion globale sur les besoins, l'organisation et l'articulation des systèmes d'information nationaux et régionaux dans le cadre de l'évaluation des objectifs de la loi de santé publique 2004 et des plans de santé publique a eu lieu.

Le HCSP a établi que les sources de données étaient nombreuses mais difficilement accessibles et insuffisamment coordonnées et il a formulé une série de recommandations spécifiques et de court terme et ainsi que des propositions plus générales pour améliorer la mesure des indicateurs et l'estimation de l'atteinte des objectifs de la future loi de santé publique.

À court terme le HCSP préconise particulièrement que soit engagée une réflexion juridique sur l'utilisation du numéro d'identification NIR de sorte à faciliter l'appariement de diverses sources ainsi que d'initier ou développer certains systèmes d'information relevant d'organismes publics dans des domaines cruciaux et actuellement non couverts, notamment en médecine ambulatoire.

Parmi ses recommandations générales le HCSP préconise de mieux coordonner les différentes actions de pilotage et de suivi des systèmes d'information pour la santé publique, d'inciter les organismes producteurs et utilisateurs de données à dégager les ressources nécessaires à une meilleure utilisation des données, de développer des dispositifs d'observation longitudinale en s'appuyant sur des cohortes existantes ou à initier et de mettre en place un dispositif de surveillance systématique et permanente des inégalités sociales et territoriales de santé.

Haut Conseil de la santé publique

14 avenue Duquesne
75350 Paris 07 SP
Téléphone : 01 40 56 79 80
Télécopie : 01 40 56 79 49
Site : www.hcsp.fr
Courriel : hcsp-secr-general@sante.gouv.fr
Bureaux
10 place des Cinq-Martyrs-du-Lycée- Buffon
75014 Paris
7^e étage, accueil pièce 7442
Métro : Montparnasse, Pasteur, Gaîté
Bus : 88, 91



Rencontrez-vous des problèmes particuliers dans vos rapports avec la CNIL (délais ou autres)