JOURNAL DE LA SOCIÉTÉ FRANÇAISE DE STATISTIQUE

FRANÇOIS LENORMAND

Le système d'information de l'assurance maladie, le SNIIR-AM et les échantillons de bénéficiaires

Journal de la société française de statistique, tome 146, nº 3 (2005), p. 47-73

<http://www.numdam.org/item?id=JSFS_2005__146_3_47_0>

© Société française de statistique, 2005, tous droits réservés.

L'accès aux archives de la revue « Journal de la société française de statistique » (http://publications-sfds.math.cnrs.fr/index.php/J-SFdS) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/

LE SYSTÈME D'INFORMATION DE L'ASSURANCE MALADIE, LE SNIIR-AM ET LES ÉCHANTILLONS DE BÉNÉFICIAIRES

François LENORMAND *

RÉSUMÉ

Dans le but de mieux comprendre, et si possible piloter, la dynamique des dépenses de santé, un nouveau système d'information unifié de l'assurance maladie est mis en place. Il isole chaque dépense de chaque bénéficiaire et tente de les lier car le pilotage de la dépense rend pertinente l'observation précise des parcours de soins. Cette masse d'information considérable conduit à privilégier le travail sur échantillon. À cet effet un échantillon inter régimes des bénéficiaires va très prochainement compléter les échantillons permanents d'assurés sociaux des grands régimes.

Mots clés : bénéficiaires, dépenses de santé, échantillon.

ABSTRACT

France's sickness-insurance fund has introduced a new information system aimed at obtaining better knowledge of health expenditures and, if possible, at steering them. The system distinguishes separate expenditure items for each benefit recipient and attempts to link the two data sets (expenditures and individuals), as expenditure steering warrants a detailed observation of individual healthcare consumption. Because of the large mass of information involved, a sample-based approach has been preferred. For this purpose, an "inter-regime sample of benefit recipients" will very shortly be available to complete the continuous samples of contributors to the main health-insurance regimes.

Keywords : benefit recipients, health expanditures, sample.

* Caisse Nationale d'Assurance Maladie des Travailleurs Salariés, 26 av. Prof. André Lemierre 75986 Paris cedex 20.
Courriel : francois.lenormand@cnamts.fr
L'auteur tient à remercier Benoît Riandey pour l'aide critique apportée à ce document. Sans ses encouragements et son soutien, cet article n'aurait pu voir le jour.

1. Une mutation impulsée par le législateur

En une dizaine d'années, le système d'information de l'assurance maladie a connu une mutation de grande ampleur. Favorisée par le progrès informatique, elle répond d'abord à la volonté du législateur de disposer d'un système unifié d'information de l'assurance maladie. Cette dynamique trouve sa source dans les ordonnances d'avril 1996 qui ont promu une transparence inter régimes de l'assurance maladie.

Celle-ci s'est d'abord appuyée sur un programme statistique commun aux quinze régimes de l'assurance maladie, puis sur un outil de gestion commun aux régimes, le Répertoire national inter régimes de l'assurance maladie (RNIAM), et le système de la carte électronique Vitale qui lui est associé.

L'année 1999 a été marquée par le vote de la loi sur la couverture maladie universelle (CMU) qui généralise le champ de l'assurance maladie. Notamment, l'aide médicale d'État (AME), étend la couverture maladie aux immigrés en situation irrégulière. La même année, la loi de financement de la sécurité sociale pour 1999 crée le Système national d'informations inter-régimes de l'assurance maladie (SNIIR-AM), destiné d'abord à permettre le suivi individuel de l'activité des offreurs de soins (professionnels de santé libéraux, structures hospitalières publiques et privées), ensuite à restituer à ces offreurs de soins une image de leur activité au travers des remboursements de l'assurance maladie.

La dernière réforme de l'assurance maladie, en août 2004, s'efforce d'accroître l'efficacité de la maîtrise de la dépense. Dans ce but, elle trace un cadre de parcours de soins plus rationnels entre médecins et organise la circulation des informations entre médecins traitants et médecins consultants.

Ces grands textes législatifs traitent d'abord des enjeux que le système d'information doit éclairer. Abordons donc ces questions de fond avant d'examiner les outils que gestionnaires et statisticiens ont développés.

2. Les enjeux du système d'information SNIIR-AM

Une question centrale commande l'évolution des systèmes d'information de l'assurance maladie : *À quel niveau de détail la consommation médicale est-elle accessible en France ? Ou, plus exactement, jusqu'à quel niveau de détail faut-il descendre, si l'on veut comprendre et piloter cette consommation ?*

On pourrait soutenir que le cadre de pensée de la régulation financière du système de santé est contingent aux informations dont on dispose pour piloter ce système, mais il semble, historiquement, que la relation entre mode de régulation et système d'informations soit inverse : on a longtemps cru à la possibilité de réguler efficacement les dépenses d'assurance maladie en le faisant de façon globale.

Pour l'essentiel, la régulation a été pensée sur le mode collectif, comme s'il suffisait de peser sur des agrégats financiers pour en maîtriser l'évolution, les agents économiques individuels prenant conscience de cette contrainte

collective et adaptant leurs comportements en conséquence. Le dernier quart de siècle est révélateur à cet égard.

Cette conviction s'est épanouie dans le mode de gestion de la dépense hospitalière : à partir de 1984, la dépense de ces établissements a cessé d'être gérée au niveau central, au profit d'une régulation de leurs recettes. La possibilité d'une maîtrise des dépenses hospitalières par le biais d'un contingentement des recettes permettant de les financer a constitué pendant vingt ans un dogme absolu. En matière de système d'informations, les conséquences ont été immédiates : le suivi des dépenses des différents établissements hospitaliers a été abandonné au profit du suivi des versements qui leur étaient faits par l'assurance maladie. Le mode de financement des hôpitaux a retardé la médicalisation naissante des systèmes d'informations de ces structures. Le PMSI [1] a commencé à exister dix ans après le budget global, mais le nombre d'études réalisées à partir de ces informations apparaît aujourd'hui extraordinairement restreint au regard des ambitions de départ et des efforts déployés.

Par contagion, ce mode de régulation à visée globalisante s'est étendu, naturellement, à d'autres secteurs : celui des cliniques privées (avec un objectif spécifique au début des années 1990), et même certain secteurs de soins de ville (activité des médecins libéraux, de 1995 à 1997).

Le moins qu'on puisse dire est qu'un tel cadre de pensée n'a pas favorisé l'émergence de systèmes d'informations détaillés.

La statistique des dépenses d'assurance maladie est restée pendant quarante ans très proche de ce qu'elle était déjà à la fin des années 1960 : une statistique décrivant, lorsque c'était possible, les dépenses des régimes d'assurance maladie au travers du système de tarification qui permet de valoriser les remboursements de consommation médicale. En matière de soins remboursés à l'acte, une nomenclature, la NGAP (nomenclature générale des actes professionnels) regroupe ces actes par classe et fixe la valeur de chacune des classes d'actes, telle qu'elle est reconnue par les régimes d'assurance maladie. Pendant longtemps, le problème essentiel de cette nomenclature a été celui de sa mise à jour, au regard tout à la fois du progrès médical et des revendications tarifaires des professionnels de santé.

Aucun autre critère n'était véritablement pris en compte : ni la nature exacte de l'acte réalisé, ni le professionnel de santé qui l'exécutait, ni, bien sûr, le patient qui en bénéficiait, encore moins la pathologie dont il souffrait ou le diagnostic porté par le médecin.

À la fin du 20e siècle, le panorama change. Tout d'abord, avec le redémarrage, évident dès 1998, des dépenses d'assurance maladie. L'échec des systèmes de régulation mis en place par les ordonnances de 1996 est moins vécu comme le résultat d'une absence de volonté politique d'appliquer les procédures de sanctions financières que comme la conséquence d'un défaut d'information et de compréhension des déterminants des dépenses.

1. Programme de médicalisation du système d'information : c'est le système d'information décrivant les dépenses hospitalières de chaque patient.

La loi de financement de la sécurité sociale pour 1999 crée le SNIIR-AM destiné d'abord à permettre le suivi individuel de l'activité des offreurs de soins : professionnels de santé libéraux, structures hospitalières publiques et privées; ensuite à restituer à ces offreurs de soins une image de leur activité au travers des remboursements de l'assurance maladie.

Dès lors, le système d'informations de l'assurance maladie sera voulu et conçu comme étant un système permettant de rendre disponible l'information issue de l'activité de remboursement des caisses d'assurance maladie à un niveau de détail le plus fin possible.

Mais, en même temps que le SNIIR-AM commence à être mis en œuvre, la problématique de la régulation continue d'évoluer.

Une question posée de longue date est celle de la juste rémunération des professionnels de santé et du juste financement des structures sanitaires et médico-sociales. Des éléments de réponse sont apportés au travers de deux réformes qui débouchent en 2005 : une tarification à l'activité des dépenses des hôpitaux et des cliniques, et une classification détaillée des actes médicaux, et d'abord des actes techniques, deux systèmes qui doivent permettre de gérer de façon plus efficiente la dépense. Cette exigence est intégrée aujourd'hui dans le SNIIR-AM. Elle nécessite de disposer d'informations, que l'assurance maladie ne détient pas forcément, sur la nature exacte des soins dispensés et sur leur coût véritable.

Une autre question, également très ancienne, est celle des déterminants de fond de l'évolution des dépenses de santé, en fait de l'évolution des besoins de santé, et de l'efficacité de la réponse que le progrès médical peut apporter à cette évolution. Engagé en 2002, le débat sur l'efficience de la dépense de santé (et de l'assurance maladie) a débouché, en réaction à la période antérieure, sur le concept, encore difficile à mettre en œuvre aujourd'hui, de *maîtrise* médicalisée des dépenses. Une maîtrise qui nécessite de disposer d'informations dont l'assurance maladie dispose rarement sur la maladie traitée.

On voit bien qu'au travers de ces deux questions, le contenu médical du traitement et son coût, l'assurance maladie s'efforce de mesurer l'efficience de sa dépense. De façon toute pragmatique, se trouve posée la question de l'efficacité des parcours de soins dans lesquels s'engagent, ou se trouvent engagés, les patients. Cette efficacité est supposée aujourd'hui insuffisante, sans du reste qu'on soit capable de mesurer la réalité et l'importance de ce gaspillage des ressources.

La dernière réforme de l'assurance maladie (août 2004) entérine cette idée, en traçant un cadre de parcours de soins plus rationnels entre médecins, et en organisant la circulation des informations entre médecins traitants et médecins consultants. L'idée de base est bien qu'une rationalisation de ces parcours serait, à terme, génératrice d'économies. Pour vérifier cette idée, il faut bien sûr pouvoir suivre le malade tout au long de son traitement.

Le suivi et l'analyse des parcours de soins répondent ainsi à une problématique ancienne et une contrainte législative nouvelle : le nomadisme médical qui constituerait un facteur important de soins inutiles, redondants ou dispensés

à mauvais escient, justifierait la nécessité d'une régulation économique du système, la solution résidant dans le choix d'un médecin traitant. La loi vient aujourd'hui privilégier l'adressage de médecin à médecin, et se réserve les moyens d'infléchir les comportements actuels en pénalisant financièrement l'accès direct à un médecin spécialiste.

On peut donc résumer ainsi les principales contraintes, ou les principales ambitions, autour desquelles tente de se structurer l'architecture des nouveaux systèmes d'informations de l'assurance maladie :

- *Produire une information détaillée sur les patients qui font l'objet de soins et sur les professionnels de santé qui les dispensent : « détaillée » veut dire ici caractérisant les individus en cause, tout au moins dans les limites liées à la nécessaire garantie de l'anonymat des bénéficiaires*;
- *Disposer d'une information précise sur la nature de ces soins et sur leur coût*;
- *Savoir comment les professionnels de santé ont été rémunérés à l'occasion de ces soins, ce qu'ils ont coûté au patient et ce que les régimes d'assurance maladie ont pris à leur charge*;
- *Rassembler des informations sur les épisodes de soins, tenter de les relier à des pathologies et pouvoir comparer l'efficacité et le coût des traitements.*

3. Le RNIAM, colonne vertébrale du système

Avant de décrire les principes et les modalités de constitution du SNIIR-AM et de l'échantillon de bénéficiaires qui se met en place, il est nécessaire de comprendre ce qu'impliquent ces différentes contraintes, et comment le SNIIR-AM tente d'y faire face. Mais auparavant, il convient d'évoquer l'outil préalable à ce nouveau regard porté sur les dépenses médicales, non plus du point de vue d'un agent économique, l'assuré, mais de celui du bénéficiaire.

Les ordonnances de 1996 ont donc instauré le RNIAM, répertoire destiné à identifier tous les bénéficiaires de l'assurance maladie. Géré à Tours par la Caisse nationale d'assurance vieillesse, ce répertoire s'appuie sur les sources de l'état civil et des fichiers de l'immigration. Il enregistre ainsi les enfants dès leur naissance, *via* le répertoire RNIPP [2] de l'Insee, et les immigrés à leur entrée en France. Le RNIAM a permis le développement de la carte électronique Vitale et donc le remboursement automatisé des dépenses d'assurance maladie. Néanmoins, la seconde étape du développement de la carte Vitale reste à faire car elle s'appuie encore sur le numéro de sécurité sociale de l'assuré, le NIR [3] , et non sur celui du bénéficiaire. On le verra, cet état provisoire a des conséquences importantes sur la cohérence des données du SNIIR-AM

2. RNIPP : Répertoire nationale d'identification des personnes physiques, gérant le numéro d'inscription au répertoire.

3. Numéro d'inscription au répertoire, plus classiquement désigné par « numéro de sécurité sociale ».

en interdisant la totalisation tout à fait systématique des dépenses de chaque bénéficiaire.

4. Principes et modalités de constitution du SNIIR-AM

4.1. Les contraintes de gestion et de taille

Les informations rassemblées sont des informations issues de la gestion du système de remboursement. Cette gestion met en œuvre de multiples organismes, au travers d'une organisation complexe.

Le régime général d'assurance maladie dispose de 131 caisses primaires pour assurer le remboursement des prestations qu'il verse. Il s'appuie sur 16 caisses régionales et 24 unions régionales. Près de 80 000 personnes assurent le traitement des dossiers, leur liquidation, les remboursements, les contrôles, et sont en définitive chargées de gérer plus de 100 milliards d'euros de dépenses. Adossées à ce régime, cinquante fédérations de mutuelles jouent le rôle de sections locales du régime général, c'est-à-dire assurent pour le compte de celui-ci le traitement des dossiers et le remboursement qui est à la charge du régime de base. Si l'on considère qu'à côté du régime général, qui couvre plus de 83 % de la population française, coexistent quatorze autres régimes de base, on commence à entrevoir la complexité d'une telle organisation, qui nécessite des passerelles entre ces membres dès qu'il s'agit de suivre des assurés qui peuvent changer de régime, ou des professionnels de santé dont la clientèle déborde évidemment du cadre étroit d'un seul régime. Il faut également ajouter à ce panorama les mutuelles, institutions de prévoyance et assurances complémentaires.

Quelques chiffres...

Au-delà de la complexité institutionnelle de l'assurance maladie, il y a le nombre de ceux, personnes et établissements, directement impliqués dans les soins : près de soixante millions de Français (la probabilité est très forte d'avoir affaire au moins une fois par an à l'assurance maladie), 350 000 professionnels de santé libéraux, 1 800 hôpitaux publics, un millier de cliniques privées, près de 10 000 établissements médico-sociaux pour enfants inadaptés, adultes handicapés et personnes gées dépendantes.

La qualité des informations

L'assurance maladie traite des milliards d'événements. Par exemple, il faut se souvenir que 2,5 milliards de boîtes de médicaments sont délivrées chaque année dans les pharmacies de ville et remboursées par la sécurité sociale. Pour chacune de ces boîtes, le code qui permet d'identifier le médicament est conservé. Les données enregistrées lors de chaque consommation médicale ne décrivent pas seulement les biens ou les services médicaux consommés, elles indiquent aussi quel est le consommateur, quel a été le producteur des soins, et, lorsqu'il s'agit d'une prescription, quel en a été l'exécutant (le pharmacien, lorsqu'il s'agit de médicaments).

S'agissant des actes pratiqués par les quelque 120 000 médecins libéraux qui exercent aujourd'hui en France, il faut savoir qu'on en compte environ 500 millions par an et, qu'au-delà de ces actes, il y a tous ceux réalisés par les dentistes, les sages-femmes, les infirmières, les kinésithérapeutes, les laboratoires d'analyse médicale, sans compter les interventions des transporteurs sanitaires...

Toutes ces informations sont rassemblées et traitées sous une contrainte qui est celle de l'environnement réglementaire : l'information produite est avant tout une information de gestion, qui comprend essentiellement les éléments nécessaires au remboursement d'un soin (il n'y a pas aujourd'hui de données exogènes dans les systèmes d'informations de l'assurance maladie).

La qualité de ces informations se ressent de cette contrainte. Cette qualité est largement suffisante pour tout ce qui permet de traiter un dossier et d'en assurer le remboursement. Il s'agit bien de rembourser l'acte effectivement réalisé à la personne qui en a bénéficié, et ceci au tarif auquel avaient droit le praticien et le bénéficiaire.

Les informations qui n'ont pas un caractère bloquant au regard de cette opération sont surveillées de plus loin, même si elles font l'objet de contrôles de qualité. Ces contrôles sont d'autant plus difficiles que la qualité des informations peut être altérée à plusieurs niveaux. Il y a bien sûr le niveau de l'intervention des organismes d'assurance maladie, c'est-à-dire celui du traitement des dossiers papier (éventuellement scannérisés), mais, de plus en plus, les flux d'informations qui parviennent à l'assurance maladie sont dématérialisés, ce qui diminue les risques (aujourd'hui, les flux électroniques représentent plus des trois quarts des flux traités). Il y a aussi ce qui se passe en amont, c'est-à-dire les renseignements permettant de générer ces flux électroniques (par exemple, les renseignements fournis par les services administratifs des hôpitaux ou des cliniques privées, ou ceux concernant l'environnement de la prescription fournis par le pharmacien qui délivre le médicament). Là aussi, on peut espérer que l'informatisation des établissements et des officines simplifie ces opérations et en améliore la qualité.

Unités traitées

Le SNIIR-AM est donc un fichier de consommations. Ces consommations ne sont pas rapportées à l'assuré, mais au bénéficiaire lui-même. L'identifiant d'un bénéficiaire dans le SNIIR-AM correspond au triplet crypté {NIR de l'ouvrant droit, date de naissance du bénéficiaire, sexe du bénéficiaire}. L'ordre des paramètres du triplet est bien sûr important.

4.2. Secrets et sécurités

L'anonymisation du SNIIR-AM est assurée par le passage successif de deux fonctions d'anonymisation utilisant la procédure FOIN [4] (*cf.* l'article de Catherine Quantin *et coll.* dans ce même dossier) :

4. FOIN : fonction d'occultation d'information nominative (*cf.* Trouessin G., Allaërt FA., 1997, FOIN : a nominative information occultation function. MIE 97 ; 43 : 196-200).

- *Une fonction d'anonymisation de premier niveau, appliquée avant transmission des données d'alimentation de l'entrepôt, et donc implantée dans l'ensemble des sites alimentant le SNIIR-AM en informations (c'est-à-dire actuellement dans l'ensemble des Centres de traitement informatique, CTI du régime général, et dans l'ensemble des autres régimes). Cette fonction d'anonymisation de premier niveau sera appelée dans ce document FOIN-SNIIRAM-1* ;
- *Une fonction d'anonymisation de second niveau, appliquée après réception des données d'alimentation de l'entrepôt et implantée uniquement sur le site du gestionnaire de l'entrepôt (le CENTI, Centre national de traitement informatique). Cette fonction d'anonymisation de second niveau sera appelée dans ce document FOIN-SNIIRAM-2.*

Pour assurer la sécurité de l'anonymat du SNIIR-AM, deux règles fondamentales doivent être respectées :

- *Les deux niveaux d'anonymisation (et donc les deux secrets) ne doivent jamais se trouver sur un même site* ;
- *Le secret de second niveau doit être diffusé le moins possible : en l'occurrence, seul le gestionnaire du SNIIR-AM doit le détenir.*

Déroulement des transferts d'informations

Lorsqu'il y a transmission des informations par les régimes vers le gestionnaire du SNIIR-AM situé à vreux, au CENTI, la première fonction d'anonymisation transforme dès le départ les identifiants en clair du SNIIR-AM en des identifiants cryptés de premier niveau. À l'arrivée des données au CENTI, la seconde fonction d'anonymisation transforme ces identifiants cryptés en des identifiants cryptés de second niveau. Les identifiants cryptés de premier niveau ne sont pas visibles au cours du processus.

4.3. Richesse et pauvreté des informations du SNIIR-AM

Les données du SNIIR-AM : les principes

Le premier principe est celui de l'exhaustivité. Le SNIIR-AM concentre, au niveau national, l'information élémentaire, c'est-à-dire la quasi-totalité des informations issues de la liquidation des soins, dans un vaste entrepôt (datawarehouse). Cette information est stockée, et en principe disponible, pour l'année en cours et les deux années précédentes. Cette exhaustivité permet, comme on l'a vu, de respecter les principes posés par la loi de financement de la sécurité sociale, pour ce qui concerne l'activité des professionnels de santé et celle des établissements. Elle constitue, à l'évidence, un plus par rapport au système antérieur, qui ne fournissait que des renseignements agrégés sur l'activité de chaque professionnel de santé. Ainsi, chaque acte médical pratiqué un jour donné par un professionnel de santé peut être individualisé. L'encadré 1 fera mieux comprendre la nature des informations dont on dispose ou qui font défaut.

Encadré 1

Information disponible sur un acte remboursé

Les informations disponibles sont aussi individualisées suivant le bénéficiaire (rendu anonyme bien entendu) ce qui, cette fois-ci, va au-delà du texte de la loi mais rend heureusement possible le suivi des parcours de soins.

Ces informations devraient être dans un avenir proche partiellement **médicalisées** : elles seront enrichies, pour les patients souffrant d'une affection de longue durée, du code précisant laquelle des trente affections ouvrant droit à une prise en charge à des soins à 100 % est en cause. Plus précisément, le code de la pathologie (en nomenclature CIM10, la classification internationale des maladies) qui a entraîné cette prise en charge figurera également. Les données issues du PMSI hospitalier seront également chaînées avec les soins dispensés en ville, ce qui devrait permettre de reconstituer des trajectoires entre l'hôpital et la ville.

Le codage précis des biens médicaux est intégré dans ce système d'informations. Il en est ainsi des systèmes de codification déjà mis en œuvre à la fin des années 1990 : le codage des actes d'analyse médicale d'abord, celui des médicaments délivrés en officine de ville (ce codage n'existe pas encore pour les pharmacies hospitalières) ensuite. Le codage des autres biens médicaux (matériel d'appareillage, pansements, prothèses internes et externes) a démarré en mars 2005.

Le codage des services médicaux monte également en charge. La classification commune des actes médicaux s'applique depuis le mois d'avril aux actes techniques effectués par des médecins spécialistes dans le cadre de l'exercice d'une activité en cabinet ou en clinique.

Encadré 1 (suite)

La tarification des versements de l'assurance maladie aux cliniques privées s'appuie désormais sur la codification de leur **activité** et non plus seulement sur le nombre et la durée des séjours hospitaliers. Enfin, l'activité des hôpitaux publics donne elle-même lieu, à compter de juin 2005, à une tarification partielle.

Pour l'essentiel, les informations présentes et pertinentes dans le système sont les suivantes :

- *l'identité du patient,*
- *l'acte médical précisément effectué,*
- *l'analyse médicale prescrite et exécutée,*
- *le médicament effectivement délivré sur ordonnance,*
- *la date des soins médicaux,*
- *la date des ordonnances et de leur exécution,*
- *l'identité des professionnels de santé impliqués dans chaque acte médical et dans chaque prescription, les établissements de santé éventuellement concernés.*

Pour autant, bien des informations manqueront encore et sans doute pour longtemps, ou tout au moins elles seront très difficilement exploitables, alors même qu'elles sont cruciales pour une gestion efficace du risque par l'assurance maladie. Citons quelques exemples :

- *Les caractéristiques socioprofessionnelles des patients, et leur couverture sociale;*
- *La* ***maladie*** *qui est traitée à une* ***date*** *donnée par le médecin, et le diagnostic porté;*
- *L'****historique*** *des soins dispensés au malade;*
- *La* ***filière*** *de soins parcourue par le patient;*
- *La* ***qualité des soins*** *dispensés;*
- *L'****efficacité*** *de ces soins;*
- *Le* ***coût*** *de l'ensemble de ces soins par pathologie.*

Il s'agira donc de suppléer à ces informations, quand elles manquent ou quand elles sont inexploitables, en utilisant les informations existantes, au mieux de leurs possibilités. Les insuffisances sont doubles : tout d'abord, la maladie ne sera disponible que pour les affections de longue durée, et encore ne s'agit-il que d'une information administrative caractérisant de façon très générale le malade à une date donnée (en général celle où il a pu commencer à bénéficier d'une prise en charge intégrale de ses soins) et non la pathologie pour laquelle il est soigné aujourd'hui. Ensuite, les parcours de soins n'existent pas en tant que tels. Ils doivent être reconstitués à partir de chaque consommation médicale et des professionnels de santé impliqués, ce qui relève parfois de la déduction ou de l'intuition plus que de la certitude.

Encadré 1 (suite)

Le système d'informations est en plein bouleversement et il va s'enrichir très rapidement d'informations essentielles. L'année 2006 devrait voir arriver à maturité le système. À cet horizon, l'activité des médecins et leurs prescriptions d'analyses et de médicaments pourront être suivies de la façon la plus détaillée qui puisse être (encadré 2).

Encadré 2

La reconstitution des parcours de soins

Comment arriver à reconstituer un parcours de soins à partir des informations du SNIIR-AM ?

Le plus simple est de prendre un exemple, fictif mais plausible, de parcours de soins. Il s'agit ici du cas, classique, d'une personne qui va subir une intervention chirurgicale. Elle consulte d'abord un médecin généraliste, qui lui prescrit des médicaments et une analyse médicale, avant de l'adresser à un chirurgien. Celui-ci va l'examiner, lui prescrit à nouveaux des médicaments (les mêmes ou d'autres), un complément d'analyse et une radio. Lors d'une consultation suivante, il décide de l'intervention au vu du résultat de ces examens. La consultation d'anesthésie (avec à la clé une analyse médicale spécifique) a lieu ensuite. Elle précède l'intervention, faite ici en clinique, et suivi d'une hospitalisation. L'épisode se conclut par un examen de sortie fait par le chirurgien, une prescription d'arrêt de travail, de médicaments et d'une radio de contrôle. Le généraliste de ville prend alors le relais, pour un suivi après retour à domicile.

Un exemple de parcours de soins réel ...

Voilà en fait les informations dont on dispose aujourd'hui :

... sa trace dans le système d'informations, jusqu'en 2005

Encadré 2 (suite)

Les consommations médicales sont bien répertoriées (en particulier les prescriptions - sauf les radios qui impliquent un autre membre du corps médical - sont correctement rattachées au professionnel qui les a faites), mais elles ne sont pas chaînées.

Jusqu'à présent en effet, l'adressage à un autre médecin est une situation inconnue de l'assurance maladie. Le chaînage des événements médicaux ne peut se faire aujourd'hui que sur présomption, en fonction de la datation des événements du parcours de soins du patient et de l'identification des professionnels de santé concernés. L'épisode de soins devrait être reconstitué par déduction, en fonction de la chronologie des soins.

La loi de réforme introduit explicitement l'adressage dans les relations entre médecins. L'épisode de soins précédent sera donc vu désormais de la façon suivante :

... puis sa trace dans le système d'informations, à partir de 2005

On voit sur cet exemple que l'information, sans être complète, sera, du moins lorsqu'il y aura adressage, suffisante dans la plupart des cas pour reconstituer des parcours de soins individuels.

Lorsque la pathologie sera disponible, ou, à défaut, lorsque les spécialistes impliqués dans les soins, les médicaments prescrits, les analyses médicales effectuées pourront être considérés comme des traceurs de cette pathologie, la voie sera ouverte à une analyse comparée du coût et de l'efficacité des traitements.

4.4. La mise en œuvre progressive du SNIIR-AM

De la théorie à la pratique, il y a quelques obstacles qu'il n'est pas toujours simple de franchir. Faisons ici le point sur l'état d'avancement des travaux.

Le SNIIR-AM est en fait organisé avec des magasins (datamarts) disposés autour de l'entrepôt de données et permettant d'avoir accès à des données détaillées suivant certains axes, mais partiellement agrégées sur d'autres.

Ainsi, le datamart *dépenses de santé* devient peu à peu opérationnel. C'est à partir de ce datamart que sont générées aujourd'hui les bases de données statistiques mensuelles qui permettent de mener une analyse conjoncturelles.

Ces bases sont agrégées par catégories de bénéficiaires, de professionnels de santé exécutants ou prescripteurs, par catégories d'établissements, etc.

Un autre exemple est celui du datamart *offre de soins*, qui est en cours d'achèvement et permettra un suivi individuel de l'activité des professionnels de santé. D'ores et déjà, des tableaux de bord permettent de suivre la consommation de médicaments et les analyses médicales, par catégories de bénéficiaires ou de prescripteurs, au niveau le plus détaillé du codage de ces biens ou services médicaux.

Enfin, un historique comprenant les données de pratiquement l'ensemble des régimes d'assurance maladie est disponible depuis 1983.

Deux limitations du SNIIR-AM existent cependant :

Tout d'abord, il faut avoir présent à l'esprit un fait incontournable : la **volumétrie** de cet entrepôt de données interdit de travailler directement sur celles-ci. Cela n'avait du reste jamais été envisagé. En pratique, indépendamment des habilitations qui permettent aux différentes catégories d'utilisateurs d'avoir accès aux informations avec un niveau de détail plus ou moins élevé, l'accès aux données de détail est, pour des raisons pratiques, réservé à des spécialistes et doit constituer une exception. Les temps d'accès et d'exploitation, que personne du reste n'a complètement quantifiés à ce stade, devraient de toute façon être largement dissuasifs. Étudier des parcours de soins à partir de l'entrepôt est donc probablement impossible.

Ensuite, il faut se souvenir que ces données sont issues d'un système de gestion qui a, comme tous les autres de cette taille, des **imperfections** et parfois des lacunes. Une bonne connaissance de ces imperfections et une bonne compréhension des données, de la façon dont elles ont été collectées et de leur signification réelle, sont absolument nécessaires. En d'autres termes, les exploitations sauvages de tels systèmes d'informations peuvent se révéler périlleuses pour un utilisateur non expert. De ce point de vue, les datamarts du SNIIR-AM, mieux adaptés à des utilisateurs occasionnels, présentent plus de sécurité.

5. Les échantillons de bénéficiaires

Faite de façon trop rapide, la présentation du système d'informations de l'assurance maladie pourrait laisser croire qu'antérieurement au SNIIR-AM, aucun dispositif de suivi de la consommation des bénéficiaires n'existait. Bien au contraire, le régime général a mis en place, à compter du début des années 1980, un système de suivi des consommations à partir d'un échantillon.

5.1. L'échantillon permanent d'assurés sociaux (EPAS)

La conception de l'EPAS

Il s'agissait, et il s'agit toujours, d'un échantillon de consommants. Le tirage de cet échantillon se fait à partir d'un algorithme portant sur le NIR des assurés. Une fois l'assuré sélectionné, ses ayants droit sont à leur tour sélectionnés. Il s'agit donc d'un tirage par grappe (de fait, ce tirage est au 1/600e), réalisé dans les fichiers de personnes des caisses primaires du régime général (les assurés relevant d'une section locale mutualiste en sont donc exclus).

Ce mode de tirage date de l'époque pas si lointaine où seuls les assurés étaient identifiés par leur NIR dans les fichiers des caisses et où il n'était donc pas possible de sélectionner tous les bénéficiaires à partir d'un identifiant unique. Aujourd'hui, ces fichiers sont enrichis de façon progressive avec les NIR de chaque bénéficiaire.

Une fois ce tirage effectué, les consommations des individus sélectionnés sont relevées dans les bases de données de remboursement de l'assurance maladie. Les données sélectionnées correspondent dans l'ensemble à la description qui en a été faite ci-dessus : elles permettent donc de décrire de façon très fine les consommations médicales.

Ces opérations de tirage sont réalisées tous les mois au niveau de chaque centre de traitement informatique régional (il y en a neuf aujourd'hui), et les données sont ensuite rassemblées au niveau national.

Cet échantillon sert essentiellement à quatre grandes catégories d'utilisateurs.

Tout d'abord, il permet de disposer, quasiment en temps réel, d'une vision transversale des comportements de consommation médicale, et ceci par grande catégories de bénéficiaires (essentiellement suivant le sexe, l'âge, l'existence ou non d'une affection de longue durée, la qualité de bénéficiaire – assuré ou ayant droit). La taille de l'échantillon ne permet cependant pas d'aller plus loin, par exemple de l'utiliser pour des études régionales.

Dans ce cadre, l'échantillon constitue un bon outil de mesure de l'impact de mesures prises ou envisagées dans le domaine de l'assurance maladie. La plupart des mesures proposées dans les projets de loi de financement de la sécurité sociale sont au préalable testées dans l'EPAS. C'est en effet le seul outil qui permette actuellement de juger de la dispersion de l'effet de ces mesures d'une personne à l'autre. Le dernier exemple en date est le test de l'impact financier de la mise en place d'une contribution forfaitaire à l'occasion de chaque prestation de médecin ou de laboratoire, et surtout le test de l'effet du plafonnement de cette participation, en fonction du nombre d'actes ou d'examens réalisés pour une même personne.

Ensuite, l'EPAS est utilisé, de façon moins fréquente il est vrai, pour faire quelques suivis individuels. Ainsi, les consommations médicales des enfants dès leur naissance, ou bien celles de leur mère avant et après la naissance de l'enfant ont-elles pu être étudiées dans une perspective non plus transversale, mais longitudinale. De façon similaire, une étude sur les effets de génération et d'âge sur la consommation a pu être réalisée à partir d'un suivi longitudinal

des patients. Une même méthode a été utilisée pour suivre la consommation médicale à l'approche du décès, ou celle des personnes qui rentre dans le dispositif des affections de longue durée. Mais, pour des raisons qui seront indiquées plus loin, ces études sont limitées par le mode de constitution de cet échantillon.

Enfin, l'échantillon sert de base à l'enquête santé protection sociale (ESPS) de l'IRDES [5] , réalisée tous les deux ans, auprès d'une fraction de ces assurés. Ces enquêtes permettent de collecter des données exogènes au système de remboursement, de connaître la couverture sociale des individus, et de mieux comprendre leurs comportements en matière médicale. L'enquête ESPS est également collectée auprès des assurés des régimes agricoles et des travailleurs indépendants sur la base d'un échantillon permanent d'assurés de même type. L'enquête ESPS, réalisée quatre fois depuis 1988, est le type même de panel fondé sur des données administratives que recommandent Mylène Chaleix et Stéfan Lollivier dans leur rapport au Conseil national de l'information statistique (CNIS, voir l'article de ces deux auteurs à ce sujet dans ce même dossier). Le panel y trouve une base de sondage mécaniquement mise à jour et des informations économiques difficiles à collecter sans erreur ni lourdeur auprès des enquêtés.

Les limites de l'EPAS

L'EPAS est un instrument d'analyse irremplaçable. Mais il a un certain nombre de limites.

C'est un échantillon de **consommants**, et non de bénéficiaires. En effet, les fichiers locaux des caisses d'assurance maladie à partir desquels sont tirés au sort les individus de l'EPAS, sont des fichiers qui servent à la gestion de leurs assurés et qui comportent, de façon tout à fait volontaire, des doubles comptes. Ainsi, les assurés qui changent de caisse continuent à émarger pendant deux ans au moins dans les fichiers de la caisse d'origine (en plus de leur nouvelle caisse), au cas où des demandes de remboursements de soins relativement anciens continueraient à parvenir à celle d'origine. Ce n'est bien sûr pas la seule raison pour laquelle il y a des doubles comptes et il est clair que ces fichiers sont, de toute façon, mal historicisés. Quoiqu'il en soit, il n'est pas possible d'avoir, au vu de ces fichiers, une idée quantitativement précise du nombre de ressortissants du régime.

Les **grappes** qui composent l'échantillon se font et se défont au fil du temps. Jusqu'à présent, seul le NIR des assurés était connu et utilisé dans les caisses primaires. Le tirage par grappe des ayant droit garantit d'avoir à un moment donné la consommation de la famille de l'assuré, mais présente un inconvénient majeur. Les ayant droits changent d'ouvreur de droits au fil du temps et peuvent même devenir à leur tour assurés. Les assurés peuvent de leur côté devenir ayants droit d'un autre assuré. Ils peuvent également changer de régime. Ainsi, nombre d'individus sortent de l'échantillon au fil du temps : on peut même estimer que près de la moitié des individus ont « disparu » au

5. IRDES : Institut de recherche et de documentation en économie de la santé, anciennement CREDES.

bout d'une douzaine d'années. Les études longitudinales sont donc délicates dans certaines tranches d'âges, tout particulièrement aux âges de transition entre l'adolescence et l'état adulte, au moment des études supérieures et à l'entrée dans la vie active. Il en est de même lorsque les femmes s'arrêtent de travailler pour élever un enfant. Le mode d'identification des bénéficiaires n'est pas pérenne, et c'est la raison pour laquelle les caractéristiques de tirage de cet échantillon rendent difficiles le suivi des parcours de soins. C'est en définitive à des âges relativement élevés, en fait à partir de la cinquantaine, que les chances de ne pas perdre de vue un assuré de l'échantillon sont les plus importantes.

5.2. Le nouvel échantillon de bénéficiaires

Ce sont toutes ces raisons qui ont conduit à la conception d'un nouvel échantillon de bénéficiaires (*cf.* annexe). Le SNIIR-AM ne répond pas à lui seul aux exigences qui doivent normalement présider à la constitution d'un échantillon de bénéficiaires, et ceci, fondamentalement, pour deux raisons qui s'ajoutent aux limitations déjà évoquées ci-dessus.

La première de ces raisons est toute simple : le SNIIR-AM n'est pas une collection de bénéficiaires, mais un ensemble de consommations. Il est en effet constitué à partir des remboursements de soins qui ont été effectués : une personne qui n'a pas été soignée ne fera pas partie du SNIIR-AM. Il y manque donc une partie de la population française dont l'étude est en elle-même intéressante, celle des bénéficiaires non consommants.

La deuxième de ces raisons renvoie à un problème qui concerne aussi l'EPAS : l'identification des consommants du SNIIR-AM n'est pas pérenne. En effet, un consommant y est identifié à l'aide d'un système de cryptage qui porte aujourd'hui sur trois éléments : le NIR de l'assuré qui ouvre les droits du consommant, le sexe du bénéficiaire et sa date de naissance. Ce triplet peut évoluer dans le temps. Il est même possible qu'un même individu soit caractérisé à un instant donné par deux triplets différents (c'est le cas des enfants de parents séparés, qui peuvent être rattachés à chacun des deux parents).

Ceci dit, le SNIIR-AM présente sur l'EPAS un avantage qu'il faut conserver dans le nouvel échantillon, son côté « inter-régimes », qui permet d'une part d'étudier des trajectoires de soins en dépit de la mobilité sociale des individus et, d'autre part, de comparer les comportements de consommation médicale d'un régime d'assurance maladie à l'autre.

Les principes du nouvel échantillon de bénéficiaires

Ces principes, rappelés dans la demande de conseil adressée à la CNIL [6] découlent logiquement des développements. Ce sont en effet les suivants :

La Caisse nationale d'assurance maladie des travailleurs salariés (CNAMTS), la Caisse centrale de mutualité sociale agricole (CCMSA) et la Caisse Nationale d'Assurance Maladie des Professions Indépendantes (CANAM) se sont

6. Commission nationale de l'informatique et des libertés.

associées pour mettre en œuvre un dispositif visant à créer l'Échantillon permanent inter régimes de bénéficiaires de l'Assurance maladie, l'EPIB-AM. À terme, cet échantillon pourra être étendu à d'autres régimes d'assurance maladie.

Le Conseil pour la transparence des statistiques de l'assurance maladie, le COTSAM, a souhaité, dans son avis du 4 avril 2002 sur le SNIIR-AM, que ce type d'outils soit mis en place : il faut envisager « la constitution d'échantillons (surtout pour les bénéficiaires des soins) et de panels ; les échantillons et panels sont en effet suffisants pour beaucoup d'études ; ils écartent en partie les risques sur les libertés individuelles liés à l'exhaustivité et facilitent l'exploitation ; par ailleurs, c'est sur des panels que pourraient être stockées les informations sur une période excédant les deux ans prévus actuellement par le SNIIR-AM, permettant ainsi la réalisation d'analyses longitudinales plus pertinentes ; enfin, il est indispensable que les concepteurs du système assurent la continuité des échantillons et panels actuels. Ce dernier point concerne particulièrement les deux thèmes d'analyse qui n'ont pas encore été présentés à la CNIL : la consommation de soins des individus, d'une part, la pathologie et le parcours du patient, d'autre part. Le Conseil souligne l'intérêt pour la gestion du système de soins d'améliorer la connaissance sur ces deux thèmes. »

Le CNIS a donné le 5 mai 2003 un avis favorable pour la constitution de cet échantillon permanent inter régimes de bénéficiaires de l'assurance maladie : « Le Conseil encourage la mise en place de cet échantillon, qui permettra de réaliser des enquêtes sur panel et de constituer une base de sondage pour toutes les enquêtes sur la consommation de santé. »

À cet effet, la CNAMTS a inscrit la constitution de l'échantillon permanent inter régimes de bénéficiaires de l'assurance maladie dans son contrat d'objectifs et de gestion avec l'État.

La finalité du traitement est la constitution d'un échantillon permanent inter régimes de bénéficiaires de l'assurance maladie. Les raisons et les avantages de cet échantillon sont multiples et se décrivent de la façon suivante :

- *Un échantillon de bénéficiaires apporte par sa conception même une vision plus large qu'un échantillon de consommants, puisqu'il permet de suivre un échantillon représentatif des individus couverts par les régimes indépendamment de leur consommation effective. En particulier, cet échantillon permettra de disposer d'informations sur la durée des périodes de non-recours aux soins.*

- *La constitution d'un échantillon permanent de bénéficiaires permet de suivre l'évolution au fil du temps de la consommation de soins d'une population donnée. Cette population, toujours sélectionnée selon les mêmes critères, est composée de consommants et de non-consommants. Seule cette méthode permet de différencier dans l'analyse les différentes composantes influant sur le recours et la consommation de soins : modification du statut personnel par rapport aux droits ouverts, effet des couvertures complémentaires, modification du nombre d'ayants droit, effet de génération... Cette analyse n'est*

aujourd'hui pas possible avec les échantillons constitués qui sont d'abord des échantillons d'assurés (tirages par grappes).

• *Ce panel permettra également de comparer les comportements des* ressortissants en fonction de leur régime d'affiliation et en fonction de leur *parcours de protection sociale et de son évolution, le cas échéant.*

• *Il constituera un instrument de suivi des comportements de consommation dans le temps et permettra d'étudier pour la première fois les parcours de soins des malades aussi bien en termes de soins de ville que de soins hospitaliers.*

• *Il sera également un outil de veille notamment par rapport à la consommation de certains médicaments, ou le suivi de certaines affections de longues durées...*

• *Il servira d'instrument de simulation de mesures de régulation ou, de façon plus générale, de modification du système de santé ou d'assurance maladie.*

• *Par ailleurs, le panel de bénéficiaires servira de base de sondage dans le cadre d'enquêtes au sein de l'assurance maladie mais également au sein d'organismes extérieurs, comme l'IRDES et son enquête ESPS. Deux types d'enquête pourront être menés :*

- *les enquêtes* a priori : *une fraction de l'échantillon de bénéficiaires, tirée avec un algorithme précis, pourra servir de base de sondage pour des enquêtes définies* a priori.
- *les enquêtes* a posteriori : *à l'examen des données de consommation du SNIIR-AM, des enquêtes définies* a posteriori *pourront être menées sur une fraction de bénéficiaires du panel, pour en suivre les comportements; par exemple, une enquête sur une sous-population de diabétiques...*

La conception générale de l'échantillon de bénéficiaires

Le principe consiste à extraire des bases nationales des individus de chaque régime un échantillon de bénéficiaires. Dans le cas du régime général, cette base nationale rassemble l'ensemble de ses bénéficiaires, assurés ou ayant droits, qui y sont identifiés par leur NIR personnel. Il s'agit bien sûr de leur associer ensuite leurs caractéristiques individuelles et d'état de santé et, le cas échéant, leurs consommations de soins stockées dans le SNIIR-AM. Ainsi, l'échantillon permanent inter régimes de bénéficiaires de l'assurance maladie comportera des données individuelles de consommations de soins, des caractéristiques individuelles et d'état de santé. Ces données seront mises à jour de manière trimestrielle. L'historique des mouvements est limité à une période de vingt ans, à laquelle s'ajoute l'année en cours.

Chaque bénéficiaire peut avoir un ou plusieurs identifiants cryptés dans le SNIIR-AM, ceci dépendant de l'historique de sa situation au regard de l'assurance maladie. Par exemple, un bénéficiaire peut être un ayant droit d'un assuré à un instant t et devenir son propre ouvrant droit à l'instant $t+1$. À ce titre, ce bénéficiaire possède deux identifiants cryptés dans le SNIIR-AM

à deux moments différents et il est nécessaire de connaître ses deux identifiants pour pouvoir suivre sa consommation sur longue période.

L'alimentation du panel de bénéficiaires consiste donc à associer à chaque bénéficiaire l'ensemble de ses consommations de soins contenues dans le SNIIR-AM (s'il a consommé), peu importe que le bénéficiaire ait changé d'ouvreur de droit à un moment donné.

Pour permettre cette opération, il est nécessaire de créer au départ de chaque répertoire index des régimes (pour le régime général, le Centre d'exploitation national est le CEN de Valenciennes) une table de passage qui, à chaque identifiant en clair d'un bénéficiaire (NIR), associe l'ensemble de ses identifiants en clair du SNIIR-AM, le cryptage pouvant être réalisé dans un second temps.

Cet échantillon de bénéficiaires pourra servir de base de sondage ou d'échantillon maître pour tirer des sous-échantillons de personnes dans le cadre d'enquêtes.

En amont des enquêtes, un organisme tiers de confiance est nécessaire pour établir une interface entre le gestionnaire du panel et les organismes initiateurs des enquêtes. Cet organisme tiers sera le seul dépositaire de la table de passage entre les identifiants en clair des bénéficiaires (à savoir le NIR en clair) d'une part, et d'autre part, les identifiants anonymisés des bénéficiaires du panel. L'organisme tiers de confiance doit satisfaire deux critères :

- *remplir ses missions de façon indépendante au regard de cette base de données,*
- *assurer la sécurité de cette base.*

L'Insee, qui respecte ces deux critères, a accepté d'être l'organisme tiers de confiance.

Validation du panel de bénéficiaires

La bonne réussite de la constitution de l'échantillon de bénéficiaires dépend d'une validation correcte de chaque étape du projet. Cette validation doit porter sur deux aspects : le tirage des bénéficiaires, d'une part, l'appariement avec les consommations de soins du SNIIR-AM, d'autre part. Il est en effet essentiel de s'assurer de l'exhaustivité des informations de consommation tirées du SNIIR-AM.

Les utilisateurs de l'échantillon de bénéficiaires

L'utilisation de cet échantillon sera partagée par les services statistiques et médicaux de l'assurance maladie et par ceux des représentants de l'État.

Cet échantillon sera également accessible aux statisticiens des organismes de recherche (organismes d'études, chercheurs universitaires...) à des fins de recherche, en particulier comme échantillon-maître dont les données pourraient être complétées par des enquêtes ponctuelles. Des études pourront être réalisées à partir de cet échantillon pour le compte d'autres parties (les professionnels de santé).

Annexe : Constitution et validation du panel des bénéficiaires

L'interconnexion avec d'autres systèmes d'information :

L'échantillon permanent de bénéficiaires de l'assurance maladie sera connecté :

- *aux répertoires index de chaque régime, bases nationales des bénéficiaires des régimes. Pour le régime général, ce répertoire est situé au CEN situé à Valenciennes,*
- *au SNIIR-AM, base inter régimes contenant notamment les informations de consommation de soins, située au CENTI à Évreux.*

D'autres liens avec des systèmes d'information extérieurs à l'assurance maladie s'établiront, lors des enquêtes, ces dernières faisant l'objet d'une déclaration spécifique à la CNIL. Dans ce cadre, des liens de deux natures différentes seront possibles.

- *En premier lieu, le lien avec le tiers de confiance Insee*;
- *En aval des enquêtes, les données d'enquêtes pourront être restituées à l'assurance maladie par les initiateurs des enquêtes. Elles seront stockées dans des bases annexes à l'échantillon de bénéficiaires. La déclaration spécifique à la CNIL de chaque enquête précisera si les données d'enquêtes pourront être transmises à l'assurance maladie.*

Secrets et sécurités

Les principes de sécurité de l'anonymat du SNIIR-AM reposent sur la double anonymisation au départ (identifiant crypté **Id ano1**) **et à l'arrivée** (identifiant crypté **Id ano2**).

Le même principe d'un passage successif de deux fonctions d'anonymisation sur l'identifiant du bénéficiaire est mis en place sur le panel de bénéficiaires :

- *La fonction d'anonymisation de premier niveau, appliquée avant transmission des données des répertoires index de chaque régime, et donc implantée au CEN de Valenciennes pour le régime général (et dans l'ensemble des autres régimes), transformera les identifiants en clair des bénéficiaires en un identifiant crypté noté ici Id anoEB1. Cette fonction d'anonymisation de premier niveau devra être celle utilisée pour le SNIIR-AM, soit FOIN-SNIIRAM-1. Les identifiants en clair des bénéficiaires du SNIIR-AM et du panel n'étant pas les mêmes, les identifiants cryptés de premier niveau Id ano1 et Id anoEB1 seront différents.*
- *Une fonction d'anonymisation de second niveau, appliquée après réception des données des répertoires index des régimes et implantée dans trois sites transformera les identifiants cryptés Id anoEB1 en des identifiants cryptés Id anoEB2. Les trois sites concernés sont :*
  - *le site du gestionnaire du SNIIR-AM,*
  - *le site de l'organisme tiers,*
  - *le site du gestionnaire du panel.*

Cette fonction d'anonymisation de second niveau, différente de celle du SNIIR-AM, sera appelée dans ce document FOIN-EB. Elle sera spécifique à l'échantillon permanent inter régimes de bénéficiaires de l'assurance maladie.

La confidentialité, l'intégrité et la non-répudiation des données échangées seront assurées.

Caractéristiques du traitement

La constitution du panel de bénéficiaires revient à associer, à chaque identifiant d'un bénéficiaire, l'ensemble de ses consommations de soins contenues dans le SNIIR-AM (s'il a consommé) ainsi que ses caractéristiques individuelles et d'état de santé.

Identifiant retenu pour un bénéficiaire de l'échantillon :

En sortie des centres gestionnaires, il s'agit de l'identifiant crypté *Id ano1*.

Le gestionnaire du panel de bénéficiaires, l'organisme tiers et le gestionnaire du SNIIR-AM seront en possession des identifiants cryptés Id anoEB2 des bénéficiaires.

Identifiants retenus pour un bénéficiaire dans le SNIIR-AM :

Chaque bénéficiaire peut avoir un ou plusieurs identifiants cryptés dans le SNIIR-AM, ceci dépendant de l'historique de sa situation au regard de l'assurance maladie. L'alimentation du panel de bénéficiaires consiste donc à associer, à chaque bénéficiaire, l'ensemble de ses consommations de soins contenues dans le SNIIR-AM (s'il a consommé), peu importe que le bénéficiaire ait changé d'ouvreur de droit à un moment donné.

Pour permettre cette opération, il est nécessaire de créer au départ de chaque répertoire index des régimes (CEN de Valenciennes pour le régime général) une table de passage qui, à chaque identifiant en clair d'un bénéficiaire (NIR), associe l'ensemble de ses identifiants en clair du SNIIR-AM, le cryptage pouvant être réalisé dans un second temps.

Dans les schémas proposés, seul le gestionnaire du SNIIR-AM possède l'identifiant crypté de second niveau du SNIIR-AM, noté Id ano2.

Description des circuits d'information

Deux circuits d'information sont à décrire pour comprendre d'une part le cheminement des données pour la constitution de l'échantillon de bénéficiaires et d'autre part, l'utilisation de celui-ci dans le cadre d'enquêtes.

La constitution de l'échantillon de bénéficiaires et la confidentialité des traitements

Étape 1 : Tirage des échantillons de bénéficiaires dans chaque régime d'assurance maladie

Chaque régime tire dans sa base nationale des individus un échantillon de bénéficiaires, selon un algorithme de tirage basé sur les NIR des bénéficiaires. Ce tirage est réalisé au 1/100e. Un bénéficiaire est alors identifié par son NIR en clair auquel sont associés son (ou ses) identifiant(s) en clair du SNIIR-AM, à savoir le triplet {NIR de l'ouvrant droit, date de naissance du bénéficiaire, sexe du bénéficiaire}.

Cette table de correspondance entre le NIR d'un bénéficiaire et ses identifiants du SNIIR-AM est créée en clair au niveau des gestionnaires des bases nationales des bénéficiaires de chaque régime (CEN de Valenciennes pour le régime général et dans les autres régimes). Elle est ensuite véhiculée au cours des traitements avec ses mêmes identifiants cryptés.

À ce stade du traitement, aucune donnée de consommation de soins n'est extraite des fichiers de l'assurance maladie.

Étape 2 : Transmission des identifiants en clair (NIR) et cryptés des bénéficiaires sélectionnés à l'organisme tiers.

Les régimes enverront à l'organisme tiers, pour chaque bénéficiaire sélectionné, le couple composé du NIR en clair du bénéficiaire et son NIR crypté. Au départ du transfert, FOIN-SNIIRAM-1 établira, à partir du NIR en clair d'un bénéficiaire, son NIR après premier cryptage noté Id anoEB1. Le transfert consiste à envoyer les couples (NIR des bénéficiaires en clair, Id anoEB1) des gestionnaires des bases nationales de bénéficiaires de chaque régime (CEN de Valenciennes pour le régime général) à l'organisme tiers. Les identifiants Id anoEB1 ne sont pas visibles au départ du transfert. À l'arrivée du transfert à l'organisme tiers, FOIN-EB transforme les Id anoEB1 des bénéficiaires en Id anoEB2. L'organisme tiers sera le seul dépositaire de la table de passage entre les NIR en clair des bénéficiaires et les identifiants anonymisés Id anoEB2 de ces mêmes bénéficiaires. L'organisme tiers ne possèdera que la

fonction d'anonymisation de second niveau FOIN-EB. Il ne possèdera pas les identifiants cryptés Id ano2 des bénéficiaires du SNIIR-AM.

Étape 3 : Transmission par les gestionnaires des bases nationales des bénéficiaires des régimes au gestionnaire du SNIIR-AM, des identifiants des bénéficiaires sélectionnés auxquels sont associés leurs identifiants SNIIR-AM, puis extraction des consommations des bénéficiaires sélectionnés consommants.

Chaque régime transmettra au gestionnaire du SNIIR-AM les identifiants cryptés des bénéficiaires tirés en ayant soin d'associer à chaque bénéficiaire l'ensemble de ses identifiants cryptés dans le SNIIR-AM.

Au départ du transfert, FOIN-SNIIRAM-1 opère le premier cryptage sur :

- *les NIR en clair des bénéficiaires qui deviennent des Id anoEB1* ;
- *les identifiants en clair du SNIIR-AM, à savoir le triplet {NIR de l'ouvrant droit, date de naissance du bénéficiaire, sexe du bénéficiaire} qui deviennent des Id ano1.*

La table de correspondance entre ces identifiants doit être préservée. Par ailleurs, ces nouveaux identifiants ne sont pas visibles au départ du transfert. À l'arrivée du transfert au gestionnaire du SNIIR-AM, les seconds niveaux de cryptages sont réalisés par :

- *FOIN-SNIIRAM-2 qui transforme les identifiants cryptés Id ano1 du SNIIR-AM en Id ano2,*
- *FOIN-EB qui transforme les identifiants des bénéficiaires Id anoEB1 en Id anoEB2.*

À nouveau, la table de correspondance entre ces identifiants doit être assurée. Le gestionnaire du SNIIR-AM apparie alors pour chaque bénéficiaire consommant d'identifiant Id anoEB2, l'ensemble de ses consommations de soins. Le gestionnaire du SNIIR-AM possède les logiciels de cryptage FOIN-SNIIRAM-2 et FOIN-EB.

Étape 4 : Transmission des identifiants des bénéficiaires sélectionnés par les gestionnaires des bases nationales des régimes au gestionnaire du panel de bénéficiaires.

Chaque régime transmettra au gestionnaire du panel de bénéficiaires les identifiants cryptés des bénéficiaires. Au départ du transfert, FOIN-SNIIRAM-1 transforme le NIR en clair d'un bénéficiaire en un identifiant crypté de premier niveau appelé Id anoEB1. Ces identifiants Id anoEB1 ne sont pas visibles au départ du transfert.

À l'arrivée du transfert au gestionnaire du panel de bénéficiaires, FOIN-EB transforme l'identifiant crypté Id anoEB1 d'un bénéficiaire en l'identifiant crypté de second niveau Id anoEB2. Ainsi, le gestionnaire du panel possédera l'ensemble des identifiants cryptés de second niveau des bénéficiaires sélectionnés, qu'ils consomment ou non des dépenses de santé.

Le gestionnaire du panel ne possèdera que la fonction d'anonymisation de second niveau FOIN-EB. Il ne possèdera pas les identifiants cryptés Id ano2 des bénéficiaires du SNIIR-AM, ni les NIR en clair des bénéficiaires.

Étape 5 : Transmission par le gestionnaire du SNIIR-AM au gestionnaire du panel de bénéficiaires, des identifiants cryptés des bénéficiaires ayant consommé des dépenses de santé avec leurs consommations de soins associées.

Il s'agit d'un simple transfert, sans nouveau cryptage, des Id anoEB2 des bénéficiaires du panel consommants avec leurs consommations associées. Le gestionnaire du panel de bénéficiaires procède ensuite à l'appariement avec les identifiants Id anoEB2 qu'il a reçus à l'étape précédente.

Le gestionnaire du panel ne possède pas les identifiants cryptés Id ano2 du SNIIR-AM.

Circuits d'information dans le cadre des enquêtes

Le circuit d'information décrit ci-après est identique pour les deux types d'enquête (*a priori* et *a posteriori*), tout en préservant l'anonymat des données de consommations médicales.

Étape 1 : Transmission par le gestionnaire du panel de bénéficiaires à l'organisme tiers, des identifiants cryptés Id anoEB2 des bénéficiaires sélectionnés pour l'enquête.

Le gestionnaire du panel extrait un sous-échantillon de bénéficiaires du panel, soit de manière aléatoire, soit en fonction d'un critère précis (par exemple, un sous échantillon parmi les diabétiques). Le gestionnaire du panel transmet les identifiants Id anoEB2 des bénéficiaires de ce sous échantillon à l'organisme tiers.

Étape 2 : Transmission des NIR en clair des bénéficiaires à enquêter aux organismes locaux.

L'organisme tiers, qui possède la table de passage entre les NIR en clair et les identifiants cryptés Id anoEB2 des bénéficiaires du panel, transmet aux organismes locaux les NIR en clair des bénéficiaires sélectionnés pour l'enquête.

Étape 3 : Recherche des coordonnées des bénéficiaires à enquêter.

Les organismes locaux, uniques possesseurs des adresses, transmettent en retour à l'organisme tiers les coordonnées associées aux NIR en clair des bénéficiaires à enquêter.

Étape 4 : Génération d'un numéro d'ordre N_0 pour chaque bénéficiaire, puis transmission de l'organisme tiers vers la société d'enquête des N_0 et des coordonnées associées.

À la réception des coordonnées, l'organisme tiers génère un numéro d'ordre N_0 pour chaque bénéficiaire à partir des NIR en clair des bénéficiaires à enquêter. Ce numéro est spécifique à une enquête donnée. L'organisme tiers transmet alors les N_0 et les coordonnées associées à la société d'enquête.

Sur toute la période de l'enquête et tant qu'il n'y a pas eu d'apurement des données, l'organisme tiers devra alors conserver, pour chaque bénéficiaire

enquêté, le NIR en clair, l'identifiant crypté du panel de bénéficiaires et son numéro d'ordre dans l'enquête N_0, soit le triplet (NIR bénéficiaire, Id anoEB2, N_0).

Étape 5 : Transmission de la société d'enquête à l'initiateur de l'enquête, pour chaque bénéficiaire enquêté, de son numéro d'ordre N_0 et ses données d'enquête.

Étape 6 : Récupération par l'initiateur de l'enquête, de la table de correspondance entre les numéros d'ordre N_0 et les identifiants cryptés Id anoEB2 des bénéficiaires enquêtés chez l'organisme tiers.

Étape 7 : Appariement, par l'initiateur de l'enquête, des consommations de soins stockées chez le gestionnaire du panel, à partir des identifiants cryptés Id anoEB2 des bénéficiaires enquêtés.

Étape 8 : Restitution des données des enquêtes à l'Assurance maladie par les initiateurs des enquêtes, si cette disposition est prévue dans la déclaration CNIL de la dite enquête.

La responsabilité du secret sur l'échantillon permanent de bénéficiaires est répartie ainsi entre les trois régimes :

- *La Caisse nationale d'assurance maladie des travailleurs salariés assurera la gestion du panel de bénéficiaires et notamment sa mise à disposition aux autres utilisateurs. Ce panel sera situé au CENTI, à Évreux.*
- *La Caisse centrale de mutualité sociale agricole et la Caisse Nationale d'Assurance Maladie des Professions Indépendantes seront les détenteurs du secret d'anonymisation de second niveau de l'échantillon de bénéficiaires. Selon la procédure, ce secret est généré par deux personnes. Ces deux personnes possèdent une carte avec code, chaque carte contenant la moitié du secret.*
- *Ces cartes devant être mises sous séquestre, la Caisse nationale d'assurance maladie des travailleurs salariés en sera le dépositaire.*

Qualité de la transmission des données

Les principes retenus

La bonne réussite de la constitution de l'échantillon de bénéficiaires dépend d'une validation correcte de chaque étape du projet. Cette validation doit porter sur deux aspects : le tirage des bénéficiaires, d'une part, l'appariement avec les consommations de soins du SNIIR-AM, d'autre part. Il est en effet essentiel de s'assurer de l'exhaustivité des informations de consommation tirées du SNIIR-AM ; ces tests permettront notamment d'évaluer le nombre de numéros transitoires (par exemple les NIR provisoires, NIR des nouveaux-nés...), non repérables après anonymisation.

La démarche retenue consiste à comparer les informations de l'échantillon de bénéficiaires avec celles qui existent déjà au sein de l'Assurance maladie (à savoir celles de l'EPAS). Le texte ci-après décrit les tests faits par la CNAMTS, et potentiellement par les autres régimes.

Le futur échantillon de bénéficiaires et l'EPAS n'étant pas construits selon les mêmes circuits d'information, la confrontation des informations issues de chacun d'eux permet de valider la mise en œuvre du nouvel échantillon.

Il est rappelé que le tirage des assurés de l'EPAS est réalisé par les centres informatiques des 129 caisses primaires, à partir du fichier des assurés, selon un algorithme s'appuyant sur le NIR des assurés. Les consommations de ces assurés et de leurs grappes d'ayants droit sont ensuite extraites à partir des fichiers des prestations payées, stockées dans ces mêmes centres informatiques. Pour mémoire, selon la décision du 19 août 1986 concernant le traitement automatisé de l'EPAS, la Direction des statistiques et des études (DSE) de la CNAMTS dispose notamment des NIR en clair des individus de l'EPAS, de leur état civil et de leur date de naissance (*cf.* annexe : avis favorable de la CNIL de la demande d'avis du 8 mai 1980 enregistrée sous le n° 34 et la déclaration de modification du 3 mars 1982).

Les tests de validation du tirage dans les bases nationales de bénéficiaires

Il s'agit de s'assurer que l'on retrouve bien les mêmes individus selon que la source choisie est la base nationale des bénéficiaires ou les fichiers des assurés des 129 caisses (suivant le circuit classique de l'EPAS).

La première vague de tests consiste à effectuer un tirage d'assurés dans la base nationale des individus de la CNAMTS, selon le même algorithme que celui de l'EPAS, et à comparer la population ainsi obtenue avec celle de l'EPAS. La population concernée correspond donc aux assurés dont les NIR sont sélectionnés par l'algorithme, auxquels sont associées les grappes des ayants droit de ces assurés.

Les comparaisons s'opéreront individu par individu *via* leur NIR en clair mais également de manière globale sur différents agrégats. Par exemple, il sera intéressant de regarder la structure homme/femme, la répartition géographique des bénéficiaires, les tranches d'âge...

Les tests de validation de l'appariement avec les consommations du SNIIR-AM

Il s'agit de s'assurer que l'on retrouve bien les mêmes consommations selon que la source choisie est le SNIIR-AM ou celle des fichiers de prestations payées, stockées dans les centres informatiques des 129 caisses primaires (suivant le circuit classique de l'EPAS). Il est rappelé que le SNIIR-AM et l'EPAS sont gérés au CENTI.

Après tirage des identifiants des bénéficiaires par le CEN de Valenciennes (même population que précédemment), celui-ci transmet ces identifiants au CENTI. Le CENTI extrait du SNIIR-AM les consommations de ces bénéficiaires et les associe aux identifiants de chacun d'eux. La seconde vague de tests consiste à comparer ces consommations de soins à celles issues de l'EPAS.

En attendant que la fonction d'anonymisation de second niveau de l'échantillon de bénéficiaires soit générée, la procédure suivante est proposée pour permettre la réalisation de ces tests.

À la réception des données envoyées par le CEN de Valenciennes au gestionnaire du SNIIR-AM, le CENTI génère un numéro d'ordre N_0 à chaque bénéficiaire. Il extrait du SNIIR-AM les consommations de chacun d'eux et les associe à ces numéros d'ordre. Le CENTI fournit alors à la DSE l'ensemble de ces informations, chaque bénéficiaire étant identifié par son numéro d'ordre N_0. Il est alors possible de comparer, de manière globale, les consommations de soins du nouvel échantillon à celles de l'EPAS : nombre de consultations, de visites...