

*Maladies chroniques
et traumatismes*

Système multi sources de surveillance des cancers Croisement des données ALD et PMSI

**Description des données 2006-2008
et construction d'indicateurs**

Sommaire

Abréviations	2	3.7.2 Description de l'algorithme de sélection des cas dans le PMSI	25
Liste des tableaux	3	3.7.3 Différentes étapes de chaînage et d'élimination des cas prévalents dans le PMSI	26
Liste des figures	5	3.7.4 Différentes étapes de sélection et de rétention des ALD dans le Sniiram	26
Résumé	6	3.7.5 Croisement PMSI/ALD et élimination des cas prévalents	26
1. Introduction	11	3.7.6 Traitement des cas non appariés	26
1.1 Le système de surveillance des cancers en France	11	3.7.7 Construction des indicateurs	26
1.2 Justification de la mise en place du Système multi source de surveillance des cancers (SMSC)	11	3.7.8 Indicateurs complémentaires calculés	27
1.3 Objectif du SMSC et méthodologie générale	11	4. Résultats	28
1.3.1 Objectif	11	4.1 Description des ALD issues du Sniiram	28
1.3.2 Méthodologie générale du SMSC	12	4.1.1 Description générale des données du Sniiram transmises par la CnamTS	28
1.4 Etude pilote	12	4.1.2 Description des effectifs de bénéficiaires d'ALD cancer pour les huit localisations cancéreuses étudiées.	29
1.4.1 Déroulement de l'étude pilote	12	4.2 Description du PMSI	33
1.4.2 Axes de travail suite à l'étude pilote	13	4.2.1 Identifiants manquants	33
1.5 Le croisement ALD et PMSI	13	4.2.2 Effectifs de patients prévalents et de patients "incidents" après chaînage sur les années antérieures	33
2. Objectifs de ce travail	14	4.2.3 Effet du choix de l'algorithme de sélection de cas prévalents sur le nombre de cas "incidents" de cancer	35
2.1 Objectif général	14	4.3 Description des effectifs "incidents" croisés	37
2.2 Objectifs spécifiques	14	4.3.1 Effectifs par localisation de cancer issus des sélections dans les bases	37
3. Matériel et méthode	15	4.3.2 Effectifs croisés	37
3.1 Données	15	4.3.3 Description des cas appariés : délai d'apparition, cohérence sur le sexe, l'âge, le département et la région de résidence	50
3.1.1 Les données hospitalières	15	4.3.4 Etude des cas "incidents" non appariés	51
3.1.2 Données d'ALD cancer du Sniiram	15	4.4 Indicateurs d'incidence issus des données croisées ALD et PMSI 2006-2008, par localisation	54
3.1.3 Transmission des données et circuit d'anonymisation	16	5. Discussion	59
3.2 Population étudiée	17	5.1 Rappel des objectifs	59
3.2.1 Définition de cas "incidents" PMSI d'une année N	17	5.2 Limites et avantages des sources de données utilisées	59
3.2.2 Définition des cas "incidents" ALD d'une année N	17	5.2.1 PMSI	59
3.3 Les premières localisations cancéreuses étudiées	18	5.2.2 ALD issues du Sniiram	60
3.4 Principe méthodologique d'interrogation des bases et de dénombrement des cas	18	5.2.3 Croisement des bases	60
3.4.1 PMSI	18	5.3 Discussion des résultats	61
3.4.2 ALD du Sniiram	21	5.3.1 Croisement des données et cohérence des cas appariés	61
3.4.3 Croisement des cas "incidents" PMSI et ALD cancer du Sniiram	21	5.3.2 Construction des indicateurs	61
3.5 Description des données des bases PMSI et Sniiram et des effectifs issus du croisement	21	5.4 Travaux "princeps" de surveillance des cancers à partir des bases médico-administratives	62
3.5.1 Analyse descriptive des données ALD cancer issues de la base Sniiram	21	5.5 Perspectives	63
3.5.2 Analyse descriptive des données de cancer issues du PMSI	21	6. Conclusion	64
3.5.3 Analyse descriptive des effectifs croisés PMSI/ALD	21	Références bibliographiques	65
3.6 Construction d'indicateurs d'"incidence" issus du croisement des données PMSI et ALD	24	Annexes	66
3.6.1 Construction des indicateurs	24		
3.6.2 Comparaison aux estimations nationales d'incidence 2006, 2007 et 2008	25		
3.7 Synthèse de la démarche aboutissant à la construction des indicateurs	25		
3.7.1 Méthode d'interrogation des bases	25		

SYSTEME MULTI SOURCES DE SURVEILLANCE DES CANCERS

CROISEMENT DES DONNEES ALD ET PMSI

Description des données 2006 - 2008
et construction d'indicateurs

Rapport final d'étude

Avril 2012

Auteurs :

Institut de veille sanitaire (InVS), Département des maladies chroniques et traumatismes (DMCT) :

Yao KUDJAWU
Jérémy RUDANT
Elsa DECOOL
Arlette DANZON
Isabelle GREMY

InVS, Service des systèmes d'information (SSI) :

Arnaud MUSSET

InVS, Direction scientifique (DS) :

Javier NICOLAU

Remerciements à l'ensemble des partenaires, en particulier :

- la Caisse nationale de l'Assurance maladie des travailleurs salariés (CnamTS) ;
- le Régime social des indépendants (RSI).

Relecteurs du rapport

Philippe TUPPIN, CnamTS
Jean-Claude DESENCLOS, DS, InVS

Abréviations

ACP	Anatomie et cytologie pathologiques
ALD	Affections de longue durée
ATIH	Agence technique de l'information hospitalière
BMA	Base médico-administrative
Cavimac	Caisse d'assurance vieillesse, invalidité et maladie des cultes
CCMSA	Caisse centrale de la Mutualité sociale agricole
CMU	Couverture maladie universelle
CnamTS	Caisse nationale d'Assurance maladie des travailleurs salariés
Cnil	Commission nationale de l'informatique et des libertés
DAS	Diagnostic associé significatif
DCC	Dossier communiquant de cancérologie
DMP	Dossier médical personnel
DP	Diagnostic principal
DR	Diagnostic relié
Foin	Fonction d'occultation d'identifiants nominatifs
Francim	France incidence mortalité (Réseau français des registres de cancer)
InVS	Institut de veille sanitaire
Inca	Institut national du cancer
MSA	Mutualité sociale agricole
NIR	Numéro d'inscription au répertoire
PMSI	Programme de médicalisation des systèmes d'information
RATP	Régie autonome de transport parisien
RSA	Résumé de sortie anonymisé
RSI	Régime social des indépendants
SMSC	Système multi sources de surveillance des cancers
SNC	Système nerveux central
SNCF	Société nationale des chemins de fer
Sniiram	Système national d'information interrégimes de l'Assurance maladie

LISTE DES TABLEAUX

Tableau 1 :	Synthèse des résultats : indicateurs « d'incidence » de cancer invasif issus des données croisées ALD, PMSI et estimation nationale d'incidence	10
Tableau 2 :	Les codes CIM10 utilisés pour sélectionner les cas de cancer dans les bases PMSI et Sniiram	18
Tableau 3 :	Description des bénéficiaires d'ALD pour cancer des bases Sniiram 2006-2008	29
Tableau 4 :	Effectifs de bénéficiaires d'ALD par localisation de cancer selon l'année de début d'ALD dans le Sniiram 2006, 2007 et 2008	30
Tableau 5 :	Distribution des bénéficiaires ayant débuté leur ALD cancer en 2006, 2007 et 2008 selon la première année d'apparition dans le Sniiram	32
Tableau 6 :	Pourcentage de séjours avec identifiant manquant par localisation et type de sélection, 2004-2008	33
Tableau 7 :	Effectifs de patients avec séjours pour cancer en DP en 2006, 2007 et 2008, et effectifs après chaînage sur les séjours pour cancer en DP, DR, DAS des années antérieures	34
Tableau 8 :	Effet de l'algorithme de sélection des cas sur le nombre de cas prévalents et sur les effectifs après chaînage sur le DP, DR ou DAS, données 2008	36
Tableau 9 :	Effectifs PMSI de la population « incidente » à croiser avec les ALD	37
Tableau 10 :	Effectifs ALD de la population « incidente » à croiser avec le PMSI	37
Tableau 11 :	Cancer du sein, effectifs appariés par année et cas absents 2006-2008	49
Tableau 12 :	Cancer colo rectal, effectifs appariés par année et cas absents 2006-2008	46
Tableau 13 :	Cancer du rein, effectifs appariés par année et cas absents 2006-2008	47
Tableau 14 :	Cancer du poumon, effectifs appariés par année et cas absents 2006-2008	47
Tableau 15 :	Cancer du système nerveux central, effectifs appariés par année et cas absents 2006-2008	48
Tableau 16 :	Cancer du col de l'utérus, effectifs appariés par année et cas absents 2006-2008	48
Tableau 17 :	Cancer du corps de l'utérus, effectifs appariés par année et cas absents 2006-2008	49
Tableau 18 :	Cancer de la thyroïde, effectifs appariés par année et cas absents 2006-2008	49
Tableau 19 :	Délai entre le début d'ALD et la date d'apparition dans le PMSI des cas appariés sur 2006 à 2008	50
Tableau 20 :	Cohérence des données sur le sexe, l'âge, le département et la région de résidence entre les cas PMSI et ALD appariés en 2006, 2007 et 2008	51
Tableau 21 :	Description des cas ALD « incidents » invasifs non appariés l'année N, par localisation cancéreuse et par année	52

Tableau 22 :	Description des cas issu du PMSI « incidents » invasifs non appariés l'année N aux ALD, par localisation cancéreuse et par année	53
Tableau 23 :	Cancer invasif du sein chez la femme : indicateurs « d'incidence » issus des données croisées PMSI/ALD et données d'estimation nationale d'incidence	54
Tableau 24 :	Cancer invasif du côlon, rectum : indicateurs « d'incidence » issus des données croisées PMSI/ALD et données d'estimation nationale d'incidence	55
Tableau 25 :	Cancer invasif du rein : indicateurs « d'incidence » issus des données croisées PMSI/ALD et données d'estimation nationale d'incidence	55
Tableau 26 :	Cancer invasif du poumon : indicateurs « d'incidence » issus des données croisées PMSI/ALD et données d'estimation nationale d'incidence	56
Tableau 27 :	Cancer invasif du système nerveux central : indicateurs « d'incidence » issus des données croisées PMSI/ALD et données d'estimation nationale d'incidence	56
Tableau 28 :	Cancer invasif du col de l'utérus : indicateurs « d'incidence » issus des données croisées PMSI/ALD et données d'estimation nationale d'incidence	57
Tableau 29 :	Cancer invasif du corps de l'utérus : indicateurs « d'incidence » issus des données croisées PMSI/ALD et données d'estimation nationale d'incidence	57
Tableau 30 :	Cancer invasif de la thyroïde : indicateurs « d'incidence » issus des données croisées PMSI/ALD et données d'estimation nationale d'incidence	58

LISTE DES FIGURES

Figure 1 :	Schéma de croisement des cas « incidents » invasifs issus du PMSI et du Sniiram	7
Figure 2 :	Circuit d'anonymisation des données PMSI et ALD cancer du Sniiram pour générer le numéro anonyme unique commun « AnoSMSC »	17
Figure 3 :	Chaînage des données PMSI 2006, 2007 et 2008 sur une durée identique consécutive	20
Figure 4 :	Chaînage des données PMSI 2006, 2007 et 2008 sur le maximum de durée disponible le permettant	20
Figure 5 :	Schéma de croisement des cas « incidents » invasifs issus du PMSI et du Sniiram	22
Figure 6 :	Croisement ALD, PMSI : cas appariés et non appariés, recherche de cas prévalents probables parmi les cas non appariés	23
Figure 7 :	Construction des indicateurs d'incidence à partir des données PMSI et ALD croisées et exclusion des cas prévalents des cas non appariés	24
Figure 8 :	Effectifs croisés de cas « incidents » de cancer du sein 2006, 2007 et 2008	36
Figure 9 :	Effectifs croisés de cas « incidents » de cancer du côlon, rectum 2006, 2007 et 2008	39
Figure 10 :	Effectifs croisés de cas « incidents » de cancer du rein 2006, 2007 et 2008	40
Figure 11 :	Effectifs croisés de cas « incidents » de cancer du poumon 2006, 2007 et 2008	41
Figure 12 :	Effectifs croisés de cas « incidents » de cancer du système nerveux central 2006- 2008	42
Figure 13 :	Effectifs croisés de cas « incidents » de cancer du col de l'utérus 2006, 2007 et 2008	43
Figure 14 :	Effectifs croisés de cas « incidents » de cancer du corps de l'utérus 2006, 2007 et 2008	44
Figure 15 :	Effectifs croisés de cas « incidents » de cancer de la thyroïde 2006, 2007 et 2008	45

Résumé

Introduction

L'objectif du Système multi sources de surveillance des cancers (SMSC) est de produire des indicateurs de surveillance des cancers avec réactivité et sur tout le territoire, à partir du croisement de trois sources de données : données issues du programme de médicalisation des systèmes d'information (PMSI), de l'Assurance maladie relatives aux exonérations du ticket modérateur pour cancer, correspondant à l'affection longue durée n° 30 ou ALD N°30, données d'anatomie et de cytologie pathologiques (ACP). Le SMSC est destiné à compléter le dispositif des registres, en permettant une surveillance plus réactive et pour tout le territoire.

Les enseignements de l'étude pilote, réalisée entre 2004 et 2009, et l'évolution du contexte relatif au Dossier communicant de cancérologie (DCC) avaient orienté le projet selon deux axes : la production d'indicateurs de surveillance des cancers à partir du croisement des deux sources PMSI et ALD d'une part (1^{er} axe) et la création de novo d'une base des ACP constituant une troisième source d'autre part (2^e axe).

Ce rapport présente les résultats de l'analyse du croisement des données ALD et PMSI du 1^{er} axe. Il s'agit des résultats descriptifs du croisement individuel des données PMSI et ALD, et du principe de construction d'indicateurs de surveillance de l'incidence des cancers à partir de ces données.

Objectifs de ce travail

L'objectif était de décrire à l'échelle nationale, les données ALD cancer issues du Système national d'information interrégime de l'Assurance maladie (Sniiram) et les données de séjours PMSI pour cancer ainsi que les données issues du croisement de ces deux sources de données, pour huit localisations cancéreuses (sein, côlon rectum, col de l'utérus, corps de l'utérus, rein, poumon, système nerveux central et thyroïde). Il s'agissait aussi d'élaborer une méthode de construction d'indicateurs de surveillance à partir des données ALD et PMSI croisées.

Matériel et méthode

Sources des données

Les données utilisées étaient celles des données hospitalières du PMSI 2004-2008 et d'ALD cancer issues du Sniiram des années 2006-2008. A l'exception de quelques régimes d'assurance maladie notamment la Mutualité sociale agricole (MSA) qui n'alimentait pas encore le Sniiram, la banque de France, la chambre de commerce et d'industrie de Paris, l'Assemblée nationale, le Sénat, le Port autonome de Bordeaux et les Français de l'étranger, les données d'ALD cancer provenaient des régimes d'assurance maladie. L'utilisation de ces données fait suite à l'autorisation de la Commission nationale de l'informatique et des libertés (Cnil) obtenue au cours de la phase pilote du SMSC.

Un numéro d'anonymat commun aux données du PMSI 2006-2008 et du Sniiram 2006-2008 spécifique au SMSC a permis de croiser les données individuelles des deux sources entre elles. Ce numéro d'anonymat n'était pas présent dans les données du PMSI 2004-2005 qui n'ont pas pu être croisées avec les données du Sniiram. Cependant ces données 2004-2005 ont pu être chaînées aux données du PMSI 2006-2008 dans le cadre du repérage des cas prévalents de cancer grâce à un identifiant anonyme usuel.

Population décrite

Le croisement des données PMSI et ALD cancer concernait principalement les nouveaux cas de cancer (cas médico-administratifs) enregistrés dans les bases PMSI et Sniiram de 2006, 2007 et 2008. Ces nouveaux cas médico-administratifs de cancer seront désignés dans la suite du rapport sous le terme cas « incidents » PMSI ou cas « incidents » ALD. Ils ont été définis comme suit :

- cas « incidents » PMSI d'une année N : patients ayant eu un premier séjour hospitalier pour cancer invasif de la localisation étudiée en diagnostic principal (DP) l'année N et ne présentant pas de séjour pour la même pathologie en diagnostics principal, relié (DR) ou associé (DAS) les deux années antérieures ;
- cas « incidents » ALD de cancer d'une année N : bénéficiaires d'une ALD pour cancer invasif de la localisation étudiée, débutée l'année N et enregistrée dans la base Sniiram de l'année N.

Description des données

Les données étudiées ont été décrites en trois étapes : 1) une description générale des données de la base Sniiram et pour chacune des localisations cancéreuses étudiées, 2) une description des

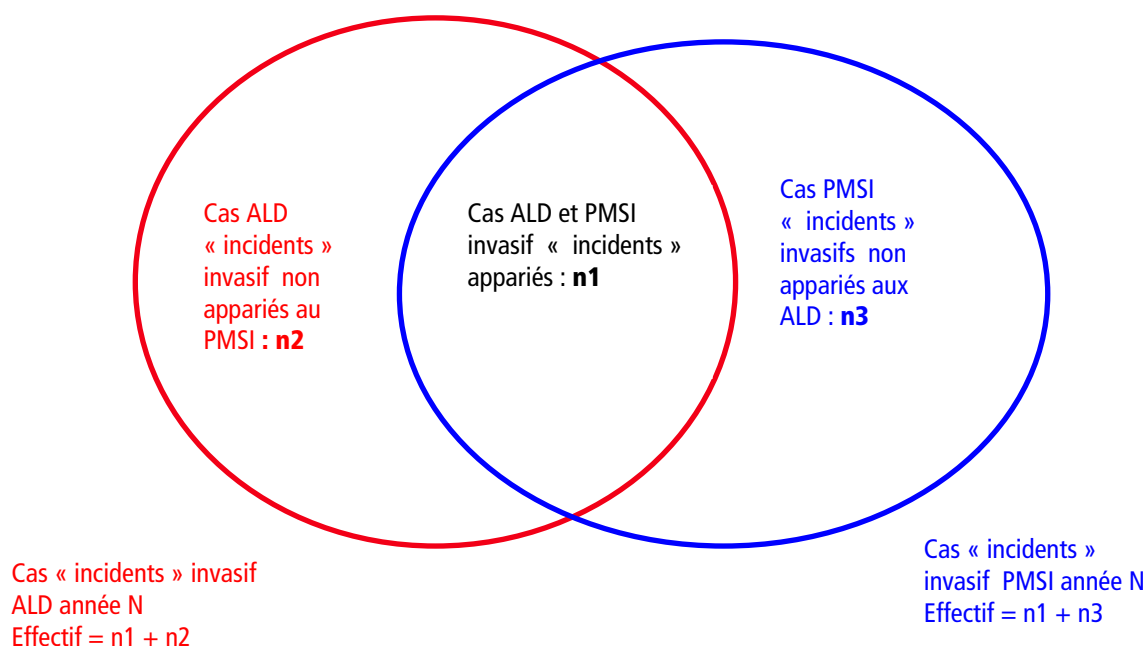
données du PMSI pour chacune des localisations cancéreuses étudiées, 3) une description par localisation de cancer des effectifs croisés issus du PMSI et des ALD (figure 1).

Les effectifs des cas appariés (n1), non appariés (n2 et n3) et de l'union (n1+n2+n3) ont été calculés.

Le calcul des principaux effectifs a été effectué grâce aux requêtes basées sur le « structured query language » (SQL) des bases croisées.

I Figure 1 I

Schéma de croisement des cas « incidents » invasifs issus du PMSI et du Sniiram



Le délai d'apparition des cas appariés dans la deuxième source a été calculé. La recherche de cohérence des informations sur le sexe, l'âge, le département et la région de résidence parmi les cas appariés entre les deux sources a été réalisée.

En ce qui concerne les cas ALD et PMSI non appariés pendant l'année N, une recherche de leur présence dans les bases PMSI et ALD des années antérieures ou postérieures a été effectuée. Elle avait pour objectif de voir si ces cas invasifs non appariés étaient des cas *in situ* ou prévalents dans ces sources ou s'ils étaient survenus ou enregistrés les années suivantes.

Construction des indicateurs

Les indicateurs construits dans ce rapport avaient pour objectif de mesurer le niveau d'« incidence » médico-administrative des cancers au cours de l'année d'étude.

Le principe de construction de ces indicateurs reposait sur l'ensemble des cas PMSI et ALD cancer croisés (union). A partir des effectifs issus de l'union des bases croisées (premier niveau d'indicateur), un affinage a été réalisé. Il consistait à retirer des effectifs de l'union, les cas non appariés suivants, qui ne seraient pas *a priori* des cas incidents :

- les cas d'ALD cancer invasif non appariés ayant la même année pour le cancer étudié, un séjour hospitalier avec un code *in situ* en DP dans le PMSI ;
- les cas d'ALD cancer invasif non appariés ayant un séjour hospitalier dans les PMSI des années antérieures (cas « prévalents ») ;
- les cas de cancer invasif PMSI de l'année N ayant une ALD cancer avec une date de début antérieure à l'année N (cas « prévalents »).

A l'issue de l'analyse, quatre niveaux d'indicateurs ont été construits :

- le premier niveau correspondait à l'effectif de l'union des cas « incidents » ALD et PMSI ;
- le deuxième niveau correspondait à l'indicateur du premier niveau duquel ont été exclus

les cas prévalents et *in situ* ;

- le troisième niveau correspondait à l'indicateur du deuxième niveau duquel ont été exclus des cas prévalents supplémentaires (cas PMSI « incidents » non appariés retrouvés parmi les cas ALD dans les bases Sniiram des années antérieures, cas ALD « incidents » non appariés retrouvés parmi les séjours PMSI des années antérieures). Le calcul de cet indicateur n'était possible que pour les années 2007 et 2008 ;

- le quatrième niveau correspondait à l'indicateur du troisième niveau mais avec un chaînage sur quatre ans des données PMSI. Son calcul n'était possible que pour l'année 2008.

Résultats

Description des données

Pour chaque localisation de cancer, la proportion de cas ALD et PMSI appariés était similaire sur les trois années. Elle était inférieure à 30 % pour le cancer du col de l'utérus, autour de 33 % pour les cancers du rein et du poumon, autour de 40 % pour les cancers du côlon, rectum et du corps de l'utérus et près de 50 % pour le cancer du sein.

La majorité des cas appariés sur 2006-2008 pour une même localisation cancéreuse, apparaissait plus tôt dans les ALD que dans le PMSI. Ainsi, entre 89 % (cancer du corps de l'utérus) et 97 % (cancer du sein) des cas appariés avaient une date de début d'ALD antérieure à la date du séjour hospitalier.

Les pourcentages de cas appariés ayant la même information dans le PMSI et dans les ALD en 2006, 2007 et 2008 étaient élevés. Ils étaient de 100 % pour le sexe et l'âge et de 99 % pour la région de résidence, pour les localisations étudiées. Mais ils variaient de 94 % à 97 %, selon la localisation cancéreuse pour le département de résidence.

Indicateurs d' « incidence » issus des données croisées ALD et PMSI 2006, 2007 et 2008

Le tableau 1 présente les effectifs des premier, deuxième et quatrième niveaux d'indicateurs. Il rappelle aussi les effectifs nationaux estimés à partir des données des registres du réseau Francim (actualisation 2011 des projections nationales).

Les effectifs issus de l'union (indicateur de l'union) des cas incidents détectés par ALD ou PMSI étaient nettement supérieurs aux valeurs issues des projections nationales d'incidence de l'année concernée suggérant l'existence d'un nombre résiduel de cas prévalents non identifiés.

Les effectifs obtenus après exclusion des cas *a priori* non incidents (prévalents) étaient toujours élevés - mais moins que les effectifs de l'union - comparés aux valeurs de l'estimation nationale d'incidence. L'écart entre les effectifs et les valeurs des estimations nationales avait diminué environ de moitié par rapport à l'union.

Le croisement des données du PMSI chaînées sur quatre ans avec les données d'ALD, associé à l'exclusion supplémentaire des cas prévalents retrouvés dans les bases des années antérieures impactait peu les effectifs (dernière colonne).

Au final, pour 2008, les niveaux des indicateurs issus du croisement ALD/PMSI étaient très proches du niveau de l'incidence nationale estimée pour le cancer du sein (écart relatif¹ de +5 %) et du poumon (-2 %), peu éloignés pour les cancers de la thyroïde (+12 %), du côlon rectum (+19 %), du rein (+24 %) et très éloignés pour les cancers du corps de l'utérus (+31 %), du col de l'utérus (+47 %) et du SNC (+58 %) (tableau 1).

Discussion

Cette première étape des analyses a consisté à décrire les données d'ALD cancer et du PMSI cancer et leur croisement de même qu'à définir les principes de construction d'indicateurs d' « incidence » à partir des données croisées.

Le PMSI présentait une complexité liée à la difficulté de repérer les séjours incidents de cancer parmi l'ensemble des séjours pris en charge. La définition d'un cas « incident » dans le PMSI était la même pour toutes les localisations et l'algorithme de sélection des cas incidents (séjour en DP l'année N sans séjour en DP, DR ou DAS les années antérieures) était identique pour toutes les localisations étudiées. Cependant cet algorithme pourrait faire l'objet d'affinage dans une phase ultérieure du projet.

¹ Ecart relatif = (Indicateur 2008 – Estimation 2008) / Estimation 2008

Si la définition de cas « incident » a paru moins complexe pour les données d'ALD issues du Sniiram, les difficultés liées à l'exploitation des données se sont révélées plus importantes en raison de la présence dans la base pour un même bénéficiaire, de plusieurs dates de mise en ALD et de la présence de certains bénéficiaires dans la base Sniiram de l'année N+1 alors qu'ils ont été mis en ALD l'année N.

La deuxième étape des analyses consistera à évaluer les indicateurs en les comparant aux données de cancers observées dans les départements avec registres. En fonction des résultats obtenus, la production des indicateurs estimés de surveillance de l'incidence des cancers à l'échelle nationale et infranationale pourrait être envisagée. Cette deuxième étape de travail fera l'objet d'un autre rapport.

I Tableau 1 I

Synthèse des résultats : indicateurs « d'incidence » de cancer invasif issus des données croisées ALD, PMSI et estimation nationale d'incidence

Localisation (cancer invasif)	Nombre de cas issus de l'estimation nationale d'incidence ^a			Union des cas « incidents » du PMSI et des ALD ^b			Indicateur commun aux trois années ^c (écart relatif ^e par rapport aux estimations nationales d'incidence 2006, 2007 ou 2008)			Indicateur 2008 ^d chaînage des données PMSI jusqu'en 2004 et croisement avec les données ALD du Sniiram (écart relatif ^e par rapport aux estimations nationales d'incidence 2008)
	2006	2007	2008	2006	2007	2008	2006	2007	2008	2008
Sein	50 800	52 500	54 100	65 450	66 366	69 298	54 684 (+8 %)	54 319 (+3 %)	57 591 (+6 %)	56 751 (+5 %)
Côlon, rectum	38 100	38 600	39 100	48 103	48 948	50 851	44 977 (+18 %)	45 822 (+19 %)	47 693 (+22 %)	46 480 (+19 %)
Rein	9 500	9 800	10 100	12 914	13 702	13 906	12 084 (+27 %)	12 777 (+30 %)	12 959 (+28 %)	12 523 (+24 %)
Poumon	33 400	34 500	35 700	36 621	36 858	38 021	34 135 (+2 %)	34 181 (-1 %)	35 471 (-1 %)	34 836 (-2 %)
SNC	4 400	4 400	4 500	7 266	7 538	7 928	6 641 (+51 %)	6 885 (+56 %)	7 240 (+61 %)	7 126 (+58 %)
Col de l'utérus	3 100	3 000	3 000	4 734	4 900	4 922	4 242 (+37 %)	4 419 (+47 %)	4 462 (+49 %)	4 417 (+47 %)
Corps de l'utérus	5 900	6 000	6 000	8 211	8 445	8 537	7 717 (+31 %)	7 908 (+32 %)	7 989 (+33 %)	7 847 (+31 %)
Thyroïde	6 648	7 013	7 391	8 918	9 065	9 607	7 885 (+19 %)	8 058 (+15 %)	8 622 (+17 %)	8 303 (+12 %)

^a Estimations calculées lors de la mise à jour des projections nationales d'incidence 2011 de Francim.

^b Effectifs issus de l'union des cas « incidents » invasifs des sources PMSI et ALD de l'année N, c'est-à-dire appartenance à l'une ou l'autre des deux groupes de patients suivants :
 - cas « incidents » PMSI de l'année N : patients ayant eu un séjour avec un diagnostic de cancer invasif de la localisation étudiée en position DP l'année N mais sans séjour avec un diagnostic de cancer invasif de cette localisation en position DP, DR ou DAS les deux années antérieures ;
 - cas « incidents » ALD de cancer d'une année N : bénéficiaires d'une ALD pour cancer invasif de la localisation étudiée, enregistré dans le Sniiram de l'année N, et débutée l'année N.

^c Effectifs de l'union PMSI/ALD desquels ont été retirés les cas non appariés suivants :
 - PMSI « incidents » de l'année N non appariés apparaissant dans la base Sniiram de l'année N mais avec une date de début d'ALD antérieure à l'année N ;
 - ALD « incidents » de l'année N non appariés retrouvées parmi les cas du PMSI prévalents de l'année N (séjours avec cancer invasif de la localisation étudiée en position DP l'année N, et séjour pour ce cancer invasif en position DP, DR, DAS les années N-1 ou N-2) ;
 - les cas d'ALD « incidents » non appariés au PMSI invasif ayant un séjour PMSI avec un code *in situ* en DP pour le cancer étudié.

^d Effectifs de l'union des ALD incidentes avec le PMSI invasif chaînés sur quatre ans, desquels ont été retirés les cas suivants :
 - les cas non appariés listés en c ;
 - les cas prévalents supplémentaires retrouvés dans les bases des années antérieures (et non présents dans les bases de l'année d'étude) :
 - PMSI « incidents » non appariés retrouvés parmi les cas d'ALD pour cancer invasif des bases Sniiram des années antérieures ;
 - ALD « incidents » non appariés retrouvés parmi les séjours PMSI prévalents des années antérieures.

^e Ecart relatif = (valeur de l'indicateur de l'année N – valeur de l'estimation nationale d'incidence de l'année N) / valeur de l'estimation nationale d'incidence de l'année N.

1 Introduction

1.1 Le système de surveillance des cancers en France

Le système de surveillance des cancers en France repose principalement sur les registres des cancers généraux et spécialisés (par organe) [1], qui assurent l'enregistrement exhaustif des nouveaux cas de cancers dans une zone géographique délimitée (en général le département). Les registres fournissent des données d'incidence observée dans les zones couvertes.

Des estimations nationales et régionales d'incidence sont produites à partir des données des registres et des données de mortalité par une modélisation statistique [2,3] ainsi que des projections nationales pour l'année en cours [4,5]. Les données des registres permettent également d'estimer la prévalence des cancers et d'étudier la survie [1].

La couverture de la population par les registres généraux est d'environ 20 %. La population des enfants (moins de 15 ans) est, quant à elle, couverte à 100 % par deux registres nationaux, le registre national des tumeurs solides de l'enfant (RNTSE) et le registre national des hémopathies malignes de l'enfant (RNHE). Depuis le 1^{er} janvier 2011, la couverture des registres pédiatriques a été étendue à 17 ans inclus.

1.2 Justification de la mise en place du Système multi sources de surveillance des cancers (SMSC)

Le développement d'un système en complément de celui des registres, basé sur un croisement automatisé et passif des bases médico-administratives du programme de médicalisation des systèmes d'information (PMSI) et des affections longue durée (ALD) ainsi que des comptes-rendus anatomopathologiques a paru pertinent pour trois raisons principales :

- la nécessité de disposer d'indicateurs capables de repérer des ruptures de tendances récentes de l'incidence. En effet, la publication des données issues des registres est réalisée avec un délai de trois ans environ car les données d'incidence issues de ce dispositif de référence doivent être exhaustives et validées, ce qui nécessite du temps ;
- la nécessité de disposer en dehors des zones couvertes par les registres d'indicateurs d'incidence à des niveaux territoriaux régionaux, départementaux ou infra-départementaux. La couverture par les registres n'a pas en effet vocation à être nationale. Une couverture nationale nécessiterait trop de moyens et ne serait pas coût-efficace. Le Plan Cancer 2009-2013 [6] recommande d'ailleurs le renforcement du dispositif de registres existants et non son extension. Le dispositif des registres permet en effet de réaliser des estimations d'incidence aux niveaux national et régional basées sur le rapport morbidité/mortalité mais il ne permet pas de disposer d'incidence à des niveaux territoriaux plus fins (régions, départements, ou autres zonages) d'où la nécessité du développement du SMSC ;
- le SMSC n'a pas vocation à remplacer le dispositif des registres mais à le compléter.

1.3 Objectif du SMSC et méthodologie générale

1.3.1 Objectif

L'objectif du SMSC est de renforcer la surveillance épidémiologique des cancers, en permettant de disposer pour tout le territoire de données de surveillance des cancers récentes et à une échelle géographique fine. Ce dispositif devrait permettre d'identifier des ruptures de tendance récentes des cancers et tenter de répondre à des questions environnementales sur tout le territoire. Il n'est pas destiné à remplacer le dispositif des registres mais à le compléter.

1.3.2 Méthodologie générale du SMSC

Le SMSC repose sur le croisement de trois sources de données :

- les données issues du PMSI ;
- les données de l'assurance maladie relatives aux exonérations du ticket modérateur pour cancer, correspondant à l'affection longue durée « tumeur maligne » N° 30 ou ALD N°30 ;
- les données issues des comptes-rendus d'anatomie et cytologie pathologiques (ACP).

Le croisement de ces trois sources permettrait d'éliminer un maximum de cas prévalents, à condition de disposer de données historiques pour chacune d'elles. Le croisement des trois sources nécessite l'utilisation d'un identifiant anonyme unique par patient qui soit commun aux trois bases.

1.4 Etude pilote

Une étude de faisabilité de ce nouveau système de surveillance a été réalisée entre 2004 et 2009 [7].

1.4.1 Déroulement de l'étude pilote

Elle avait ciblé deux régions, l'Ile-de-France (IdF) et le Nord-Pas-de-Calais (NPdC) définies par le Plan cancer au regard des priorités de santé publique dans ces deux régions et du développement, en parallèle, de deux études de faisabilité pour la mise en place de registres de cancers en zones urbaines (Val-de-Marne en région IdF et région lilloise dans le NPdC). Huit structures ACP, cinq en IdF (Institut Gustave Roussy, Groupe hospitalier Pitié-Salpêtrière, Hôpital Ambroise-Paré, Cabinet de pathologie Tolbiac et Institut de pathologie de Paris) et trois dans le NPdC (CHRU de Lille, Groupe hospitalier de l'institut de Lille et le Laboratoire de Valenciennes) ont été volontaires pour participer au pilote.

➤ La localisation cancéreuse étudiée

La localisation cancéreuse étudiée était le cancer de la thyroïde. Il présentait des caractéristiques pertinentes pour tester le dispositif. Ces caractéristiques reposaient sur sa faible fréquence associée à sa faible létalité de même qu'un diagnostic histologique systématique (donc passage par les ACP) et une prise en charge chirurgicale (donc séjour présent dans le PMSI).

➤ Sources de données, recueil d'informations et anonymisation des données

Le recueil des données individuelles ACP de cancer de la thyroïde a été réalisé rétrospectivement grâce à un support informatisé dont le contenu avait été établi et validé par les experts de cette pathologie. Ce formulaire a été installé sur des postes dédiés dans chacune des huit structures ACP. Les cas de cancer de la thyroïde correspondaient aux patients résidant en IdF et dans le NPdC pour lesquels un cancer de la thyroïde était diagnostiqué au sein de l'une des huit structures en 2004. Le recueil des données a commencé en 2006.

Les données individuelles des patients domiciliés en IdF ou dans le NPdC et bénéficiant en 2004 d'une exonération du ticket modérateur au titre de l'ALD pour un cancer thyroïdien (code CIM10 : C73) ont été extraites du Système national d'information interrégime de l'Assurance maladie (Sniiram).

Les résumés de sortie anonyme (RSA) dont le code CIM10 correspondait à un cancer de la thyroïde (C73) en diagnostic principal, associé ou relié (DP, DR ou DAS), de patients domiciliés en IdF ou dans le NPdC ont été extraits de la base 2004 du PMSI national.

Les données ACP recueillies ont été anonymisées grâce à la fonction Foin² déjà utilisée pour les données PMSI et ALD cancer. La transmission sécurisée de ces données entre les différents niveaux du projet était assurée par le logiciel Camel³. Le croisement des trois sources de données a été réalisé grâce à l'identifiant anonyme commun.

➤ Le réseau de partenaires impliqués et remontée des données à l'InVS

Les partenaires impliqués dans l'étude pilote se situaient à trois niveaux géographiques.

- Au niveau local, avec les huit structures ACP. Leur choix était basé sur les caractéristiques d'activité en cancérologie thyroïdienne, de représentativité des différents secteurs d'activité (public, privé) mais également en termes de taille de la structure (importante/petite), et de logiciels informatiques variés (principales sociétés équipant les ACP).

Elles avaient pour tâches de rassembler rétrospectivement les comptes-rendus ACP de cancer de la thyroïde diagnostiqué en 2004 puis de les transmettre au niveau régional *via* le circuit informatique sécurisé mis en place.

- Au niveau régional, le Centre de regroupement en informatique et statistique en ACP d'IdF (Crisapif) et les directions régionales des affaires sanitaires et sociales d'IdF et du NPdC (Drass, devenues Agences régionales de santé). Les tâches dédiées à ces structures ont consisté essentiellement à l'anonymisation (de premier niveau) des données ACP, à la centralisation des flux de données puis à leur transfert au niveau national.

² Foin (Fonction d'occultation d'informations nominatives) est une librairie d'anonymisation développée par la CnamTS, qui comprend un algorithme informatique associé à un code ou secret d'anonymisation.

³ Camel (Chiffrement assurance maladie élargie) logiciel qui permet le contrôle, le cryptage et le transport des données.

- Au niveau national, on distinguait d'une part la Caisse nationale d'assurance maladie des travailleurs salariés (CnamTS) qui était en charge de l'anonymisation (deuxième niveau) des données identifiantes transmises par le niveau régional, puis de leur retransmission vers l'InVS, d'autre part l'Inserm (IFR69) qui jouait le rôle de tiers de confiance (mobilisable dans le cadre d'une alerte pour le retour aux cas).

Un troisième niveau d'anonymisation irréversible était réalisé par l'InVS sur les données des trois sources grâce à une clé propre à l'InVS.

1.4.2 Axes de travail suite à l'étude pilote

Les enseignements de l'étude pilote [7] et l'évolution du contexte relatif au Dossier communicant de cancérologie (DCC) [8] ont fait orienter le projet selon deux axes : le croisement des deux sources PMSI et ALD d'une part (premier axe) et la création de novo d'une base des ACP constituant une troisième source d'autre part (deuxième axe).

Initialement, la constitution de la source ACP devrait se faire à partir des comptes-rendus des laboratoires ACP. Mais les difficultés techniques de remontée d'informations rencontrées au cours du pilote dans les deux régions tests avaient rendu impossible le déploiement en l'état du dispositif.

Ainsi la constitution de cette source ACP relève des travaux menés conjointement par l'Institut national du cancer (Inca) et l'Agence des systèmes d'information partagé en Santé (Asip-Santé) sur le DCC. Ce projet prévoit la création d'un entrepôt national de données issues du DCC qui sera alimenté par les données des ACP.

La constitution de la troisième base de données du SMSC à croiser avec les deux autres est donc dépendante de l'avancée des travaux de l'Inca et de l'Asip-Santé. Actuellement le projet est en cours de développement dans des régions pilotes. L'InVS est attentif dans ce projet à ce que la base de données issues de cet entrepôt puisse être utilisée pour le SMSC, c'est-à-dire qu'elle comporte le même identifiant patient que les bases ALD et PMSI, afin d'être croisées et qu'elle comporte les variables essentielles pour répondre aux objectifs.

1.5 Le croisement ALD et PMSI

Le projet en cours (premier axe) piloté par l'InVS porte sur la production d'indicateurs de surveillance des cancers à partir du croisement des bases de données ALD et PMSI.

L'analyse de cette étape est réalisée à partir des données PMSI 2004-2008 et ALD cancer issues de la base Sniiram des années 2006-2008. Ces données d'ALD sont celles de tous les régimes à l'exception de la MSA qui n'alimente pas encore la base Sniiram, de la banque de France, de la Chambre de commerce et d'industrie de Paris, de l'Assemblée nationale, du Sénat, du port autonome de Bordeaux et des Français de l'étranger.

Le rapport présente :

- 1) les résultats descriptifs du croisement individuel des données PMSI et ALD ;
- 2) la construction d'indicateurs de surveillance de l'incidence des cancers.

Ce présent travail a porté sur huit localisations cancéreuses : col de l'utérus, côlon rectum, corps de l'utérus, poumon, rein, sein, système nerveux central et thyroïde.

Dans une deuxième étape, un travail qui sera ultérieurement publié évaluera les indicateurs construits en les comparant aux données de cancers observées dans les départements avec registres, puis, en fonction de ces résultats, à produire des indicateurs de surveillance de l'incidence des cancers à l'échelle nationale et infranationale. Cette étape fera également l'objet d'un rapport.

2 Objectifs de ce travail

2.1 Objectif général

L'objectif général était de décrire à l'échelle nationale, les données ALD cancer issues du Sniiram des années 2006-2008, les données de séjours PMSI 2004-2008 pour cancer ainsi que les données issues du croisement de ces deux sources sur la période 2006-2008, par localisation cancéreuse. Il s'agissait aussi d'élaborer une méthode de construction d'indicateurs de surveillance à partir des données ALD et PMSI croisées.

2.2 Objectifs spécifiques

Plus spécifiquement, il s'agissait de :

- décrire les données des bénéficiaires d'ALD cancer issues du Sniiram 2006-2008 ;
- décrire les séjours hospitaliers pour cancer sélectionnés dans le PMSI 2006-2008 ;
- définir par localisation et pour chaque source de données les cas médico-administratifs « incidents », à décrire et à croiser ;
- décrire par localisation les effectifs issus du croisement entre les sources PMSI et ALD, en particulier :
1) les cas appariés et non appariés, 2) le délai d'apparition des cas appariés dans chaque source, 3) la cohérence des informations démographiques entre les cas appariés ;
- construire des indicateurs de surveillance nationale à partir des données croisées pour chaque localisation, et comparer le niveau de ces indicateurs aux données d'estimation nationale d'incidence des cancers.

3 Matériel et méthode

3.1 Données

Les données utilisées sont de deux types : données hospitalières du PMSI et données ALD issues du Sniiram.

3.1.1 Les données hospitalières

➤ Le PMSI

Le PMSI enregistre de manière exhaustive les séjours hospitaliers (publics et privés) en France métropolitaine et dans les départements d'outre-mer (DOM). La base PMSI contient des informations démographiques du patient pour chaque séjour (âge, sexe, code postal de résidence) et des informations médicales pour chaque séjour hospitalier (diagnostics et actes réalisés).

Les diagnostics de la pathologie qui a motivé l'hospitalisation sont codés grâce à la classification internationale des maladies (CIM). Le DP était, jusqu'en 2009, « le motif de prise en charge qui a mobilisé l'essentiel de l'effort médical et soignant au cours de l'hospitalisation ». Le DR correspondait au type de cancer des patients hospitalisés pour séances de chimiothérapie ou de radiothérapie (car les séances sont codées en DP). Les DAS étaient les diagnostics, symptômes et autres motifs de recours significatifs d'une majoration de l'effort de soins et de l'utilisation des moyens par rapport aux DP et DR.

Depuis 2002, un numéro anonyme permet d'identifier un même patient parmi les différents séjours, mais compte tenu du changement des règles d'anonymisation intervenu après 2003, il n'est possible de chaîner d'une façon continue les séjours PMSI qu'à partir de 2004.

Chaque année, dans le cadre de sa mission de surveillance et d'alerte, l'InVS reçoit de l'Agence technique d'information hospitalière (ATIH) une copie de la base nationale des résumés de sorties anonymisés (RSA) du PMSI.

➤ Extraction des données du PMSI

Les données PMSI utilisées sont celles des années 2004-2008. Les séjours initialement extraits sont les séjours avec un des codes diagnostics suivants en position DP, DR ou DAS :

- cancer invasif (C00 à C97) ;
- cancer *in situ* (D00 à D09) ;
- tumeur bénigne (D10 à D48) ;
- kystes cérébraux (G930).

L'utilisation des codes de cancer *in situ*, de tumeur bénigne (pour le SNC) et de kystes cérébraux permettait de repérer parmi les cas ALD cancer non appariés mis en ALD pour tumeur invasive les cas qui s'apparieraient aux données du PMSI codées cancer *in situ* ou tumeur bénigne ou kystes cérébraux.

3.1.2 Données d'ALD cancer du Sniiram

➤ Les ALD issues du Sniiram

Le Sniiram est un entrepôt de données géré par la CnamTS. Les bases de données du Sniiram sont constituées à partir des données d'activité et de dépenses détaillées, individualisées et chaînées couvrant les soins de ville, les soins hospitaliers sanitaires et sociaux, publics et privés.

Les données d'ALD cancer provenant du Sniiram contiennent des informations démographiques des bénéficiaires (année et mois de naissance, sexe, code postal de résidence, régime d'affiliation) et des données médicales (code CIM de la pathologie qui a entraîné l'exonération du ticket modérateur, la date de début ou date d'effet⁴ de l'ALD, la date de fin de l'ALD).

Les données d'ALD cancer de la base Sniiram sont des données prévalentes (anciennes et nouvelles mises en ALD pour cancer au moment de l'extraction). En ce qui concerne les données ALD cancer de la MSA qui n'alimentaient pas le Sniiram, une demande de mise à disposition et d'utilisation de ces données, formulée par l'InVS vient de recevoir un avis favorable de la part de la Cnil. Ces données seront intégrées aux futures analyses dès leur réception.

⁴ La date d'avis est la date de traitement de la demande par l'assurance maladie. La date d'effet (ou date de début) est la date à partir de laquelle les soins sont remboursés à 100 %. Elle peut être antérieure à la date d'avis.

➤ Extraction des données d'ALD

Pour ce rapport, les données d'ALD cancer du Sniiram dataient de 2006 à 2008. Les bénéficiaires ont été extraits de la base Sniiram grâce aux codes CIM10 de pathologie pour laquelle l'exonération du ticket modérateur avait été accordée. Les codes CIM10 C00 à C97 (cancer invasif) et D00 à D09 (cancer *in situ*) ont été extraits pour ce travail.

Pour chaque année, une table de données bénéficiaires et une table de données médicales ont été transmises à l'InVS par la CnamTS.

Contrairement à la base PMSI, il n'existe qu'un type de diagnostic (motif de l'exonération du ticket modérateur) dans la base Sniiram contre trois dans la base PMSI : DP, DR et DAS.

3.1.3 Transmission des données et circuit d'anonymisation

➤ Autorisations Cnil

L'utilisation des données PMSI 2004-2008 et d'ALD cancer issues du Sniiram 2006-2008 pour le croisement avait reçu l'avis favorable de la Cnil en date du 17 août 2009, dans le cadre d'une demande d'extension du projet pilote SMSC.

Conformément à l'autorisation Cnil, les données PMSI et ALD cancer anonymes transmises respectivement par l'ATIH et la CnamTS ont subi à l'InVS un niveau supplémentaire d'anonymisation avec des clés propres à l'InVS.

➤ La clé d'appariement : le numéro anonyme Foin

La procédure d'anonymisation a été réalisée à partir du logiciel Foin qui comprend un algorithme informatique auquel sont adjoints deux codes ou secrets d'anonymisation, chacun détenu par une autorité différente et indépendante. Il s'agit du principe de double anonymisation irréversible. Le premier secret est détenu par l'organisme émetteur et permet de générer un premier numéro d'anonymat (anoSNIIR1) au départ des données. Ce numéro d'anonymat anoSNIIR1 est généré à partir de trois variables : le numéro de sécurité sociale de l'ouvrant droit, le sexe et la date de naissance de l'ouvrant droit ou de l'ayant droit. Le second est détenu par l'organisme récepteur et permet, à la réception des données, de générer l'anonymat définitif (anoSNIIR2) sous lequel sont transmis les données PMSI et ALD cancer à l'InVS.

➤ Mise à disposition et architecture d'anonymisation des données PMSI et d'ALD transmises par l'ATIH et la CnamTS (figure 1)

- Données du Sniiram : les données d'ALD cancer du Sniiram 2006-2008 (sans la MSA) anonymisées avec l'AnoSniir2 avaient été transmises par la CnamTS à l'InVS, en juillet 2010. L'InVS à réception de ces données les a réanonymisées grâce à son propre secret Foin (Foin3). Cette anonymisation a permis de générer l'identifiant anonyme « AnoSMSC » spécifique à ces données. Une fois l'AnoSMSC généré, les données d'AnoSNIIR2 ont été détruites.

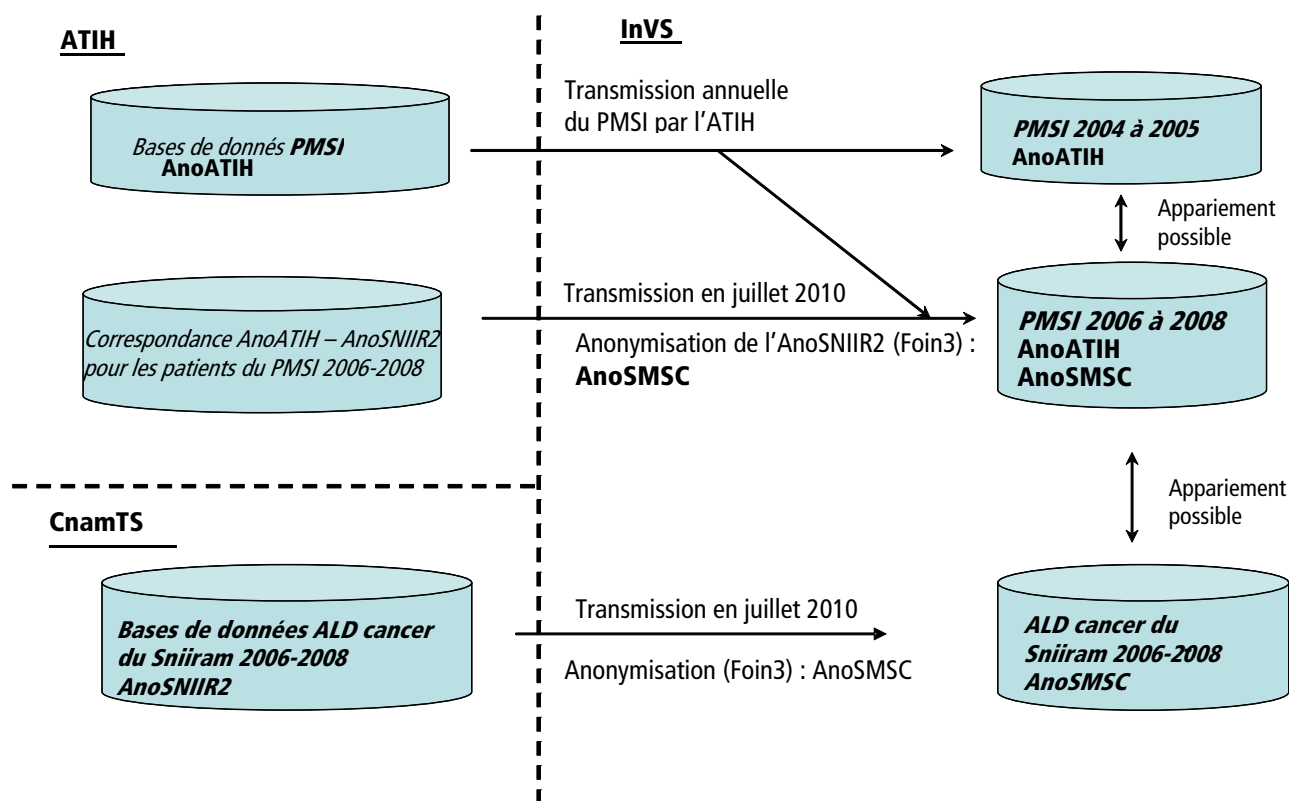
- Données du PMSI : les données PMSI 2004-2008 transmises annuellement par l'ATIH à l'InVS étaient anonymisées avec le numéro Foin de l'ATIH (AnoATIH) à partir de l'anoSNIIR2. L'ATIH gardait chez elle les identifiants avec AnoSNIIR2.

Afin de permettre le croisement des données PMSI 2006-2008 avec celles des ALD cancer 2006-2008 via l'anoSMSC commun, l'InVS a demandé et obtenu auprès de l'ATIH les identifiants AnoSNIIR2 du PMSI 2006-2008 de même que leurs correspondances avec l'AnoATIH. La transmission de ces informations a permis à l'InVS grâce à son propre secret Foin3 de générer sur les données PMSI 2006-2008 l'identifiant anonyme AnoSMSC commun aux données d'ALD cancer du Sniiram et de les apparier aux données médicales correspondantes.

Une fois l'AnoSMSC généré, les données d'AnoSNIIR2 du PMSI 2006-2008 ont été détruites.

I Figure 2 I

Circuit d'anonymisation des données PMSI et ALD cancer du Sniiram pour générer le numéro anonyme unique commun « AnoSMSC »



➤ Transmission pérenne et anonymisation des prochaines données pour le croisement (annexe 1)
Afin de simplifier le circuit d'anonymisation des données PMSI et ALD cancer à l'avenir, il a été prévu, après accord Cnil, que les données PMSI et ALD cancer de 2009 et plus soient transmises par la CnamTS à l'InVS avec l'AnoSNIR2 (et non plus par l'ATI pour le PMSI). Les données ALD cancer de la MSA seront centralisées à la Caisse centrale de la Mutualité sociale agricole (CCMSA) et anonymisées, en attendant leur intégration aux données de la base Sniiram.

3.2 Population étudiée

Dans ce travail, la description et le croisement des données PMSI et ALD cancer a concerné principalement les nouveaux cas médico-administratifs de cancer traités à l'hôpital (PMSI) ou bénéficiant de l'exonération du ticket modérateur pour cancer (ALD) et résidant en France métropolitaine ou dans les DOM.

Dans la suite du rapport ces nouveaux cas médico-administratifs de cancer seront désignés sous le terme de cas « incidents » PMSI et cas « incidents » ALD.

3.2.1 Définition de cas « incidents » PMSI d'une année N

Les cas « incidents » PMSI de cancer d'une année N ont été définis comme :

- des patients avec identifiant anonyme valide ayant eu un séjour avec un diagnostic de cancer invasif de la localisation étudiée en position DP l'année N ;
- et sans séjour avec un diagnostic de cancer invasif de cette localisation en position DP, DR ou DAS les années antérieures.

3.2.2 Définition des cas « incidents » ALD d'une année N

Les cas « incidents » ALD de cancer d'une année N ont été définis comme :

- des bénéficiaires d'ALD pour cancer invasif de la localisation étudiée, enregistrés dans le Sniiram de l'année N ;
- et qui ont débuté leur ALD l'année N (date à partir de laquelle les soins sont remboursés).

3.3 Les premières localisations cancéreuses étudiées

Huit localisations cancéreuses ont été retenues pour cette première analyse. Il s'agissait des localisations de cancer du sein invasif, cancer du côlon/rectum/anus, cancer du rein, cancer du poumon, cancer du système nerveux central (SNC), cancer du col, cancer du corps de l'utérus et cancer de la thyroïde.

Le choix des sept premières localisations (en concertation avec l'Inca, les départements santé travail et environnement de l'InVS) était basé sur des critères de santé publique tels que :

- cancer faisant l'objet d'une politique de santé publique (actions de prévention primaire et secondaire, évaluation des besoins de soins à l'échelle régionale/départementale) ;
- cancer peu fréquent et difficultés pour le repérage de signaux anormaux ou faibles (peu de sources d'information de qualité disponibles actuellement pour la localisation) ;
- cancer à incidence élevée, à forte létalité ;
- importance des variations géographiques d'incidence ;
- liens établis ou suspectés avec des facteurs de risque environnementaux ou professionnels.

Le choix du cancer de la thyroïde faisait suite à l'étude pilote SMSC.

D'autres localisations cancéreuses seront étudiées dans un second temps.

3.4 Principe méthodologique d'interrogation des bases et de dénombrement des cas

3.4.1 PMSI

- Numéro d'anonymat et chaînage des données

Deux numéros d'anonymat (AnoATIH et AnoSMSC) étaient disponibles dans les bases de données PMSI 2006-2008, mais un seul (AnoATIH) était présent dans les bases PMSI 2004- 2005 (figure 1). L'AnoATIH permettait de chaîner les données PMSI 2006-2008 avec les données PMSI 2004-2005 tandis que l'AnoSMSC permettait de croiser les données PMSI 2006-2008 avec les données ALD cancer du Sniiram mais aussi de chaîner entre elles les données PMSI 2006-2008.

Quelques patients (moins de 0,01%) avaient deux numéros AnoATIH pour un numéro AnoSMSC.

Il a été attribué aux patients du PMSI 2004 et 2005 présents dans le PMSI 2006 à 2008 le numéro AnoSMSC retrouvé dans le PMSI 2006-2008.

Le chaînage entre les bases PMSI 2004 à 2008 a ensuite été effectué sur l'AnoSMSC.

- Sélection initiales des séjours

Pour chaque année d'étude et pour chaque localisation étudiée, les séjours des patients résidant en France (métropole et DOM) avec un diagnostic correspondant aux codes CIM10 de cancer « invasif » (tableau 1) en position DP, DR ou DAS ont été sélectionnés. Les séjours avec un diagnostic correspondant aux codes CIM10 de cancer « *in situ* » ou « tumeurs bénignes/kystes » pour le SNC (tableau 1) en position DP ont aussi été sélectionnés pour les années 2006, 2007 et 2008 (à des fins de croisement avec les données d'ALD 2006 à 2008).

I Tableau 2 I

Les codes CIM10 utilisés pour sélectionner les cas de cancer dans les bases PMSI et Sniiram

Localisation de cancer	Code CIM
Sein Cancer invasif Carcinome <i>in situ</i>	C50 D05
Poumon Cancer invasif Carcinome <i>in situ</i>	C33 C34 D021 D022
Côlon-rectum-anus Cancer invasif Carcinome <i>in situ</i>	C18, C19, C20, C21 D01
Col de l'utérus Cancer invasif Carcinome <i>in situ</i>	C53 D06
Corps de l'utérus Cancer invasif Carcinome <i>in situ</i>	C54, C55 D070
Rein Cancer invasif Cancer <i>in situ</i>	C64, C65, C66, C68 D091
SNC Tumeur maligne Tumeur bénigne et kystes	C70, C71, C72, D32, D33, D353, D42, D43, D444, G930
Thyroïde Cancer invasif Cancer <i>in situ</i>	C73 D093

➤ Constitution des bases d'analyse issues du PMSI

- Les bases des cas prévalents de cancers

Bases avec les cancers « invasifs » ou « in situ » en diagnostic principal (DP) (Base 1)

Pour les années 2006, 2007 et 2008, et pour chaque localisation étudiée, des tables de cas prévalents avec un seul séjour par patient ont été constituées. Le séjour hospitalier conservé était le premier séjour de l'année avec un code cancer « invasif » retenu par l'algorithme de sélection en position DP. Si le patient n'avait pas de séjour avec un code cancer « invasif », le premier séjour avec code cancer « *in situ* » ou code de tumeur bénigne ou de kyste pour le SNC a été retenu. C'est la base 1.

Bases avec les cancers « invasifs » en diagnostic principal (DP), relié (DR) ou associé significatif (DAS) (Base 2)

Pour les années 2004, 2005, 2006 et 2007, et pour chaque localisation étudiée, des tables de cas prévalents de cancer invasif avec un seul séjour par patient ont été constituées. Le séjour hospitalier conservé était le premier séjour de l'année avec un code cancer « invasif » retenu par l'algorithme de sélection en position DP, DR ou DAS. C'est la base 2.

- Les bases des cas « incidents » de cancers

Pour les années 2006, 2007 et 2008, et pour chaque localisation, une base de patients « incidents » a été constituée à partir des bases des patients prévalents (Bases 1 et 2).

Les patients prévalents invasifs en DP de la base 1 de l'année N ont été chaînés par l'AnoSMSC aux patients prévalents en DP, DR ou DAS de la base 2 des années antérieures via des requêtes SQL.

La base des cas incidents invasifs de l'année N était ainsi composée des patients de la base 1 de l'année N qui n'ont pas été retrouvés parmi les patients de la base 2 des années précédentes après chaînage.

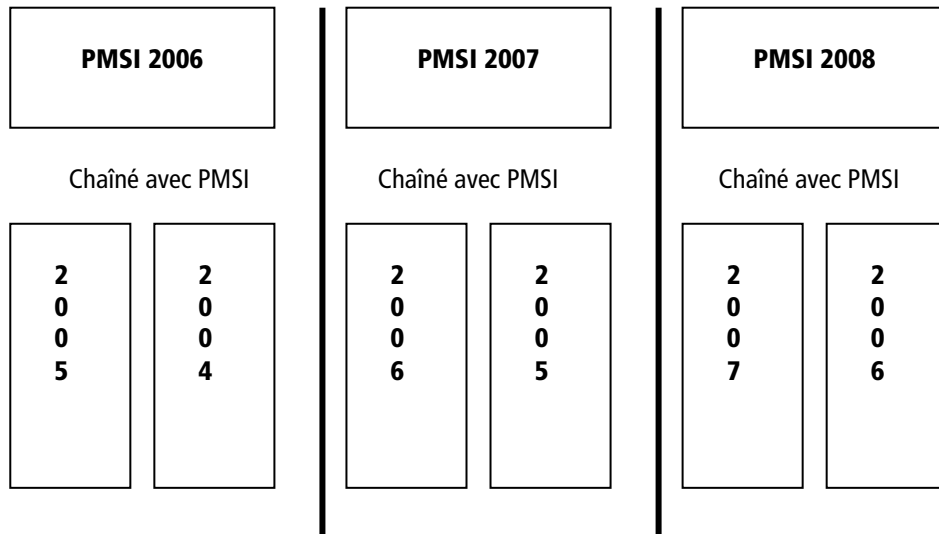
Deux options de chaînage en lien avec le nombre d'années de recul de données disponibles et les objectifs spécifiques de l'étude ont été réalisées :

1) un chaînage sur une durée identique consécutive de deux ans pour toutes les années d'étude 2006-2008 (figure 2). Ce chaînage a été utilisé pour décrire les effectifs croisés et construire des indicateurs comparables entre les années 2006, 2007 et 2008 ;

2) un chaînage sur quatre ans, correspondant au maximum d'années de recul disponibles, pour l'année 2008 (figure 3).

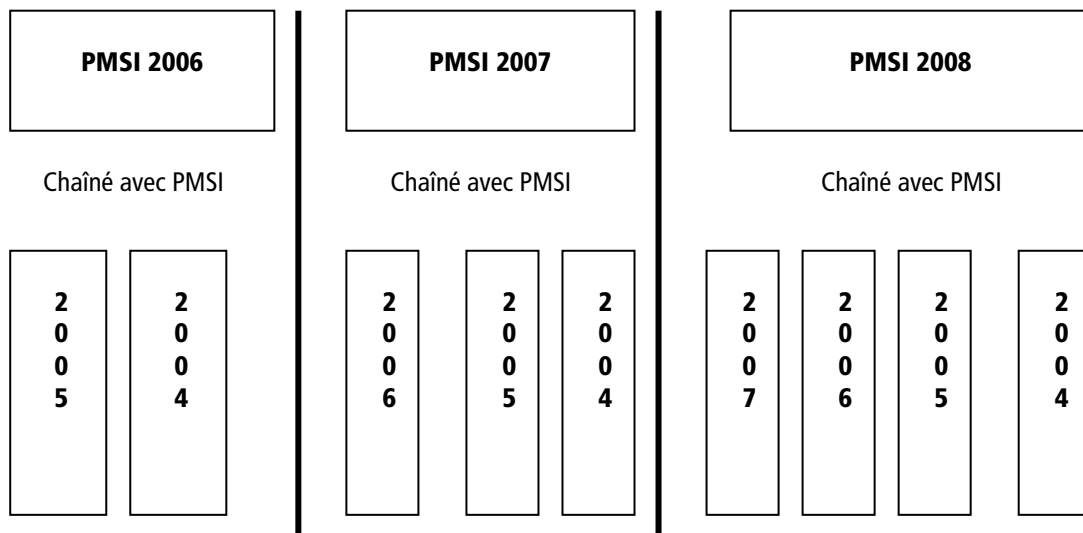
I Figure 3 I

Chaînage des données PMSI 2006, 2007 et 2008 sur une durée identique consécutive



I Figure 4 I

Chaînage des données PMSI 2006, 2007 et 2008 sur le maximum de durée disponible le permettant



3.4.2 ALD du Sniiram

Contrairement aux bases PMSI 2006-2008, les bases Sniiram d'ALD cancer 2006-2008 ne contenaient qu'un seul numéro anonyme, l'AnoSMSC, commun au PMSI. Il a été utilisé pour chaîner les bases Sniiram entre elles.

- Les bases de cas prévalents de cancers

Pour chaque année d'étude et pour chaque localisation étudiée, la sélection des bénéficiaires d'ALD pour cancer a reposé sur le code CIM10 de diagnostic de cancer pour lequel ils ont été exonérés du ticket modérateur (tableau 1). L'année de la date de début d'ALD retenue par localisation correspondait à l'année d'enregistrement de la base Sniiram.

- Les bases de cas « incidents » de cancers

Pour chaque année d'étude, la sélection des nouveaux bénéficiaires parmi les cas prévalents de la base Sniiram de l'année N reposait sur la date de début d'ALD la plus ancienne renseignée dans le Sniiram de l'année N pour le cancer étudié, certains patients ayant plusieurs dates de début d'ALD pour une même localisation.

3.4.3 Croisement des cas « incidents » PMSI et ALD cancer du Sniiram

Les cas « incidents » PMSI et ALD cancer retenus dans chaque base pour chaque année d'étude ont été croisés à l'aide de l'AnoSMSC commun, en utilisant des requêtes SQL.

Les effectifs croisés faisaient apparaître deux types de populations : 1) la population des cas appariés, 2) la population des cas PMSI ou ALD non appariés (figure 4).

3.5 Description des données des bases PMSI et Sniiram et des effectifs issus du croisement

Les données étudiées ont été décrites en trois étapes : 1) une description générale des données de la base Sniiram et pour chacune des localisations cancéreuses étudiées, 2) une description des données du PMSI pour chacune des localisations cancéreuses étudiées, 3) une description par localisation de cancer des effectifs croisés issus du PMSI et des ALD.

3.5.1 Analyse descriptive des données ALD cancer issues de la base Sniiram

Pour chaque base de données Sniiram, les informations suivantes ont été décrites :

- ensemble des bénéficiaires d'ALD cancer ;
- nombre de localisations cancéreuses par bénéficiaire d'ALD cancer ;
- le régime d'assurance maladie d'affiliation des bénéficiaires.

Pour chaque localisation étudiée, la répartition des effectifs selon l'année de mise en ALD et l'année d'apparition dans les bases Sniiram a été réalisée.

3.5.2 Analyse descriptive des données de cancer issues du PMSI

Les pourcentages par année et par localisation des séjours avec identifiants manquants ont été présentés. Les effectifs de patients prévalents et après chaînage ont été calculés par année et par localisation.

L'effet du choix des algorithmes (DP, DP-DR ou DP-DR-DAS) sur le nombre de patients sélectionnés a été également présenté.

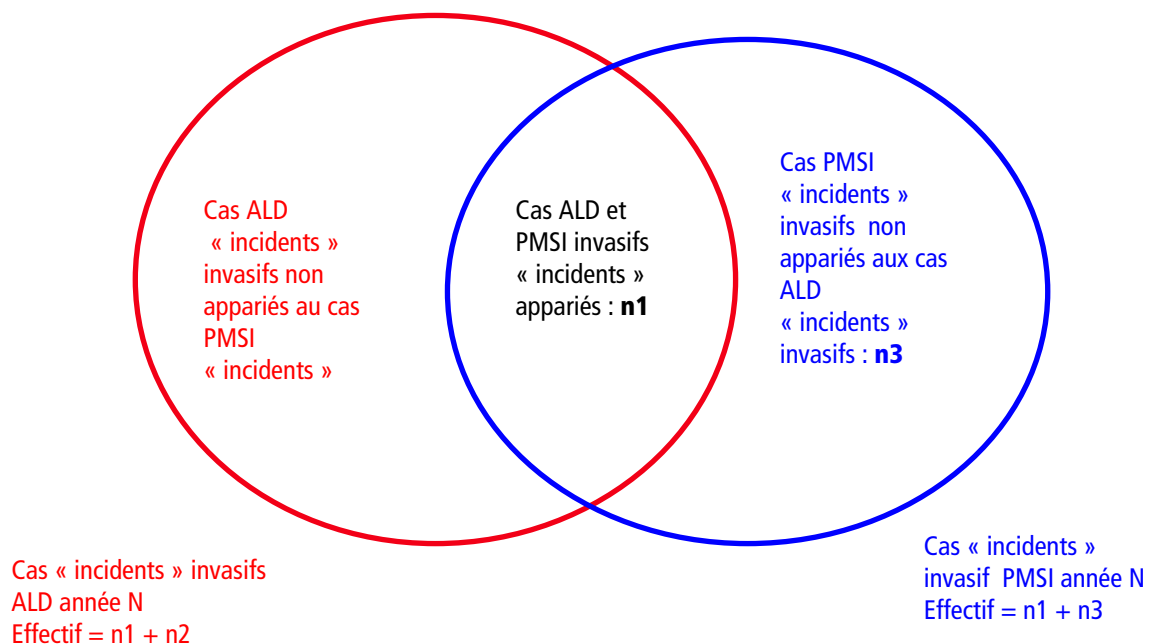
Les effectifs de patients « incidents » à croiser avec les bénéficiaires « incidents » d'ALD ont été ensuite calculés.

3.5.3 Analyse descriptive des effectifs croisés PMSI/ALD

Pour chacune des localisations, une description par année de croisement des cas invasifs « incidents » issus du PMSI et des ALD a été effectuée (figure 4) et les effectifs des cas appariés (n1), non appariés (n2 et n3) et de l'union (n1+n2+n3) ont été calculés.

I Figure 5 I

Schéma de croisement des cas « incidents » invasifs issus du PMSI et du Sniiram



Concernant les cas appariés sur la période 2006-2008, l'étude du délai d'apparition des cas « incidents » entre les deux sources a été réalisée et la cohérence des informations sur le sexe, l'âge, le département et la région de résidence entre les deux sources a été étudiée.

En ce qui concerne les cas « incidents » invasifs du PMSI ou des ALD non appariés, une recherche de leur présence parmi les cas prévalents, *in situ* ou des années postérieures de la source complémentaire a été réalisée (figure 5) :

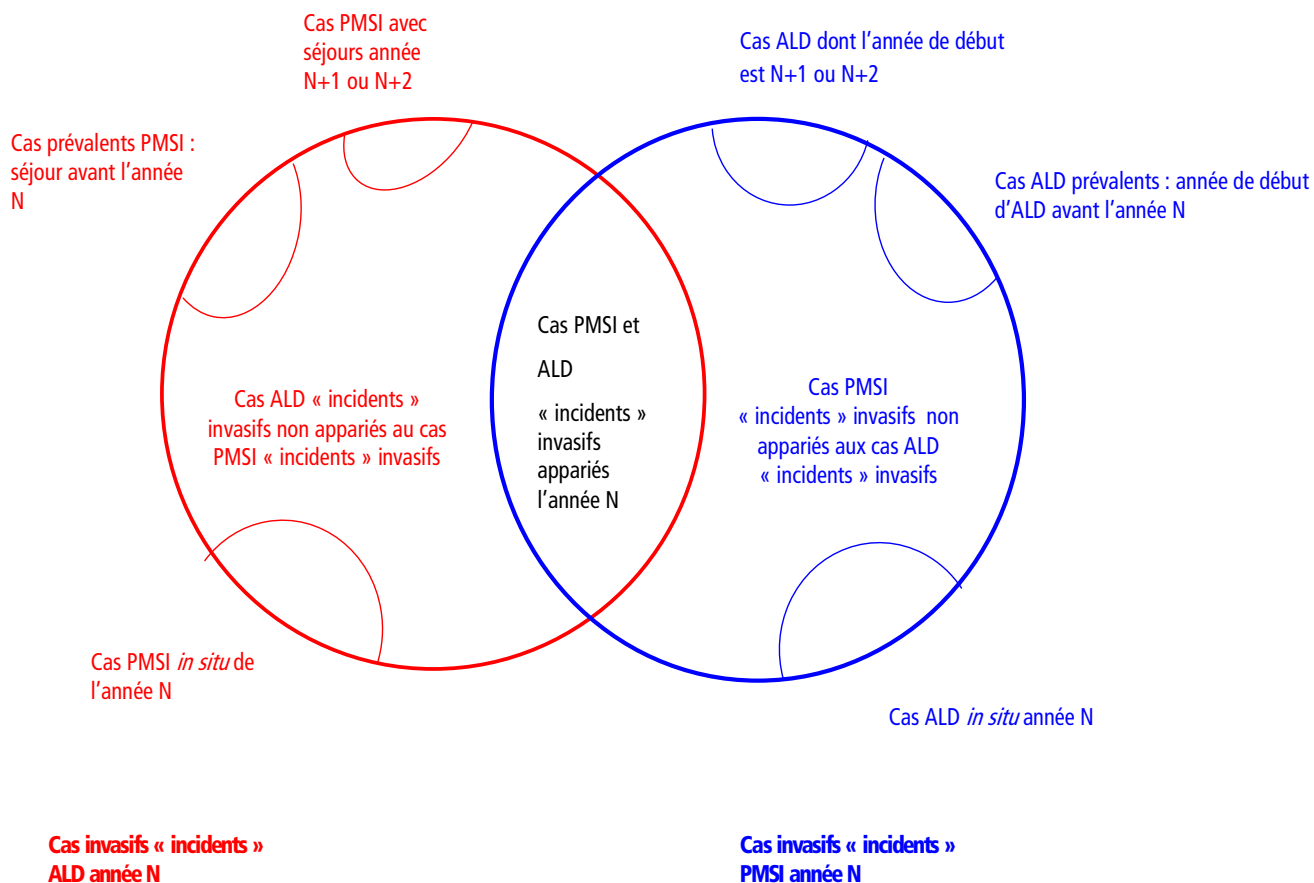
- les cas PMSI « incidents » invasifs non appariés aux cas d'ALD « incidents » invasifs de l'année N ont été recherchés parmi :
 - les cas d'ALD *in situ* de l'année N : cas en ALD pour un cancer *in situ* de la localisation, débutée l'année N et enregistré dans le Sniiram de l'année N ;
 - les cas d'ALD « prévalents », en distinguant deux situations :
 - cas d'ALD prévalents l'année N : les cas en ALD dans la base Sniiram de l'année N dont la date de début d'ALD est antérieure à l'année N ;
 - cas d'ALD prévalents les années antérieures : les cas supplémentaires apparaissant dans les bases Sniiram des années antérieures (cette recherche n'a été possible que pour les cas PMSI non appariés des années 2007 et 2008).
 - les cas d'ALD dont la date de début est postérieure à l'année N (donc issus des bases Sniiram des années N+1 ou N+2). Cette recherche n'a été possible que pour les cas PMSI non appariés des années 2006 et 2007.
- les cas ALD « incidents » invasifs non appariés au cas du PMSI « incidents » invasifs ont été recherchés parmi :
 - les cas PMSI *in situ*, c'est à dire les cas ayant un séjour PMSI pour cancer codé *in situ* en DP pour le cancer étudié, l'année N. Pour le cancer du système nerveux central, cette recherche a été effectuée sur les séjours codés comme tumeur « bénigne » ou « kystes cérébraux » car ces lésions sont souvent comptabilisées parmi les tumeurs du SNC à surveiller en raison de leur caractère imprévisible ;
 - les cas « prévalents » PMSI, en distinguant deux situations :
 - cas prévalents PMSI de l'année N, c'est-à-dire séjour pour cancer invasif de la localisation étudiée codé en DP l'année N, et séjour pour ce cancer invasif en position DP, DR, DAS les années N-1 ou N-2. Ces cas étaient des cas exclus du PMSI « incidents » suite au chaînage sur les années antérieures ;
 - cas prévalents PMSI des années antérieures (cas supplémentaires apparaissant les années antérieures dans le PMSI) c'est-à-dire séjour pour cancer invasif de la localisation étudiée en position DP, DR, DAS les années N-1 ou N-2, et sans séjour l'année d'étude pour ce cancer (cette recherche n'a été possible que pour les ALD non appariés des

années 2007 et 2008, l'année 2006 ne pouvant être chaînée avec le PMSI 2004 ou 2005).

- les cas PMSI ayant un séjour pour le cancer invasif en DP postérieur à l'année d'étude (cette recherche n'a été possible que pour les cas d'ALD des années 2006 et 2007).

I Figure 6 I

Croisement ALD, PMSI : cas appariés et non appariés, recherche de cas prévalents probables parmi les cas non appariés



3.6 Construction d'indicateurs d'« incidence » issus du croisement des données PMSI et ALD

3.6.1 Construction des indicateurs

Les indicateurs construits dans ce rapport ont pour objectif de mesurer le niveau d'« incidence médico-administrative » de cancer une année donnée. Le principe de construction des indicateurs est de partir de l'effectif de l'ensemble (union) des cas « incidents » ALD et PMSI de l'année N croisés (figure 6), puis d'affiner les données en retirant des effectifs les cas non appariés qui ne seraient pas *a priori* des cas incidents.

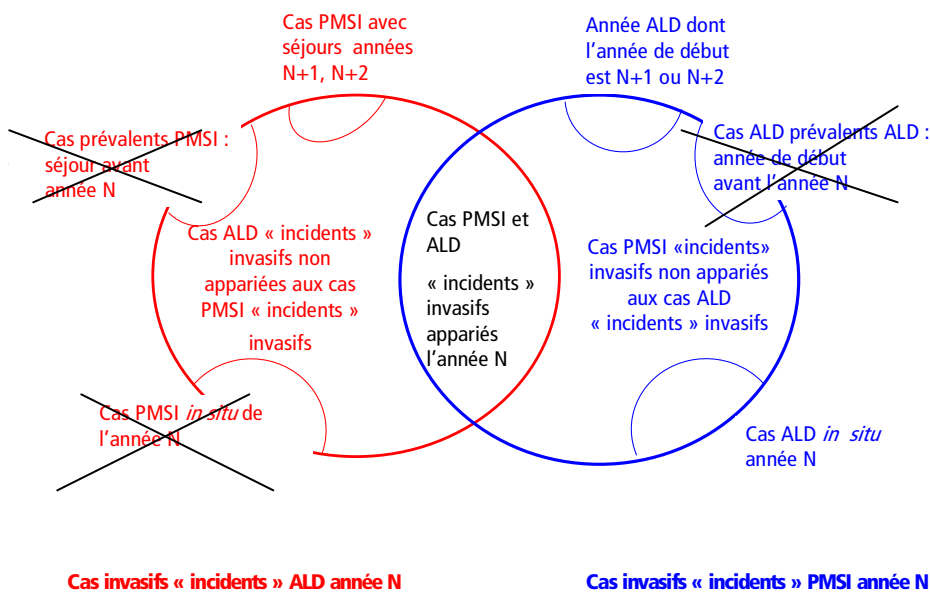
Pour chaque localisation, les cas non appariés suivants ont été considérés comme n'étant pas *a priori* des cas incidents (figure 6) :

- les cas prévalents :
 - . les cas d'ALD ayant un séjour PMSI avant l'année N ;
 - . les cas du PMSI avec une année de début d'ALD antérieure à l'année N.
- les cas d'ALD non appariés ayant un séjour PMSI avec un code *in situ* en DP pour le cancer étudié ou un code de tumeur bénigne ou de kystes cérébraux pour le cancer du SNC. Ces cas ont en effet été considérés comme des cas de cancer *in situ* mis en ALD avec un code invasif pour des raisons médico-administratives, en particulier pour le cancer du sein.

Les cas prévalents et les cas d'ALD non appariés ayant un séjour PMSI avec un code diagnostic *in situ* ou de tumeur cérébrale bénigne ou de kystes cérébraux (population indiquée par une croix sur la figure 6) ont été exclus de la construction des indicateurs.

I Figure 7 I

Construction des indicateurs d'incidence à partir des données PMSI et ALD croisées et exclusion des cas prévalents des cas non appariés (la croix indique la population de données exclue de la construction des indicateurs).



Quatre niveaux d'indicateurs ont été construits :

- le premier niveau correspondait aux effectifs issus de l'union des cas « incidents » invasifs des sources PMSI et Sniiram (ALD) de l'année étudiée ;
- le deuxième niveau correspondait à l'indicateur de premier niveau duquel ont été retirés les cas non appariés suivants : 1) PMSI invasifs « incidents » non appariés retrouvés parmi les ALD prévalents l'année N, 2) ALD invasifs « incidents » non appariés retrouvés parmi les cas PMSI prévalents l'année N, 3) ALD invasifs « incidents » non appariés ayant un séjour PMSI avec un code *in situ* en DP pour le cancer étudié.
Le calcul des indicateurs des 1^{er} et 2^e niveaux était possible pour toutes les années étudiées (2006-2008). Ces indicateurs peuvent être comparés sur cette période ;
- le troisième niveau correspondait à l'indicateur de deuxième niveau duquel ont été retirés les cas prévalents supplémentaires de l'union PMSI/ALD retrouvés dans les bases des années antérieures (et non présents dans les bases de l'année d'étude). Il s'agissait des cas non appariés suivants :
 - cas d'ALD prévalents les années antérieures : PMSI invasifs « incidents » non appariés retrouvés parmi les cas d'ALD pour cancer invasif dans les bases Sniiram des années antérieures ;
 - cas PMSI prévalents les années antérieures : ALD invasifs « incidents » non appariés retrouvés parmi les séjours des bases PMSI des années antérieures.Le calcul des indicateurs de troisième niveau n'était possible que pour les années 2007 et 2008 ;
- le quatrième niveau correspondait à l'indicateur de troisième niveau mais avec un chaînage sur le maximum d'année de recul (quatre ans) des données PMSI. La construction de cet indicateur ne concernait que l'année 2008.

3.6.2 Comparaison aux estimations nationales d'incidence 2006, 2007 et 2008

Le niveau des indicateurs produits pour chaque localisation a été confronté aux estimations nationales d'incidence de cancer des années 2006, 2007 et 2008 réalisées en 2011 par Francim et l'InVS basées sur des projections qui tiennent compte des scénarii d'évolution du cancer étudié. Pour les cancers du rein, du SNC, du côlon-rectum et du col de l'utérus le scénario de prolongation de la tendance récente a été retenu, tandis que pour les cancers du sein et du corps de l'utérus, le scénario de la constance des taux d'incidence depuis 2005 a été privilégié.

3.7 Synthèse de la démarche aboutissant à la construction des indicateurs

3.7.1 Méthode d'interrogation des bases

Requête des bases réalisée par SQL

- dans le PMSI : sélection des cas, avec identifiants anonymes valides, de cancer invasif et *in situ* ou kyste cérébral et tumeur bénigne (pour le SNC) en DP, DR, DAS pour chaque année de 2004 à 2008 ;
- dans le Sniiram : sélection des lésions de cancer invasif et *in situ* ou kyste cérébral et tumeur bénigne du cerveau, ayant motivé la mise en ALD au cours des années 2006, 2007 et 2008.

3.7.2 Description de l'algorithme de sélection des cas dans le PMSI

Année N d'étude : algorithme exclusivement basé sur la sélection des cas en DP. Deux bases sont constituées pour chaque localisation, celle des lésions invasives et celle des lésions *in situ* ou kyste cérébral ou tumeur bénigne du cerveau ;

Années antérieures : pour chaque année et localisation, un algorithme de sélection des cas en DP, DR ou DAS a été réalisé.

Un seul séjour par patient a été retenu et concernait le premier séjour de l'année avec un code cancer « invasif » en position DP, DR ou DAS.

- Les bases des cas « incidents » de cancers

Pour les années 2006, 2007 et 2008, et pour chaque localisation, une base de patients « incidents » a été constituée à partir des bases des patients prévalents (Bases 1 et 2).

Les patients prévalents invasifs en DP de la base 1 de l'année N ont été chaînés par l'AnoSMSC aux patients prévalents en DP, DR ou DAS de la base 2 des années antérieures *via* des requêtes SQL.

La base des cas incidents invasifs de l'année N était ainsi composée des patients de la base 1 de l'année N qui n'ont pas été retrouvés parmi les patients de la base 2 des années précédentes après chaînage.

3.7.3 Différentes étapes de chaînage et d'élimination des cas prévalents dans le PMSI

- Chaînage des cas : les cas invasifs en DP de l'année N sont successivement chaînés sur deux ans grâce aux identifiants anonymes avec les cas invasifs en DP, DR ou DAS des années antérieures.

- Cas prévalents : ce sont tous les cas en DP de l'année N retrouvés au cours du chaînage parmi les cas des années antérieures. Ces cas sont exclus.

- Base de cas « incidents » : elle est constituée de cas invasifs en DP de l'année N non retrouvés au cours du chaînage, dans les bases des années antérieures. Ces cas sont retenus pour le croisement avec les ALD.

3.7.4 Différentes étapes de sélection et de rétention des ALD dans le Sniiram

- Plusieurs dates de début d'ALD étaient observées chez certains bénéficiaires pour une seule localisation cancéreuse. La date la plus ancienne de mise en ALD était retenue par la sélection.

- Dans la base Sniiram de l'année N, les ALD de l'année N déclarées en année N+1 (4 %) ne sont pas sélectionnées.

- La sélection des cas portait sur les codes diagnostics « invasifs » et « *in situ* » avec des dates correspondant à l'année N.

3.7.5 Croisement PMSI/ALD et élimination des cas prévalents

- La base de cas invasifs « incidents » du PMSI de l'année N était croisée avec celle des bénéficiaires d'ALD de l'année N.

- Trois résultats possibles étaient obtenus : 1) cas PMSI/ALD appariés, 2) cas PMSI non appariés, 3) cas ALD non appariés.

3.7.6 Traitement des cas non appariés

- Les cas invasifs « incidents » PMSI de l'année N non appariés sont recherchés parmi les cas ALD prévalents l'année N, les cas ALD « *in situ* » de l'année N et dans les bases Sniiram des années antérieures.

- Les cas invasifs « incidents » ALD de l'année N non appariés sont recherchés parmi les cas PMSI prévalents de l'année N, les cas PMSI « *in situ* » de l'année N et les cas PMSI prévalents des années antérieures.

3.7.7 Construction des indicateurs

Plusieurs niveaux sont distingués dans la construction des indicateurs. Ils sont fonction du degré d'exclusion des cas supposés prévalents ou d'année de recul dans le chaînage des données du PMSI :

- le niveau 1, correspondait aux indicateurs issus de l'union des cas invasifs PMSI et ALD sans exclusion ;
- le niveau 2, correspondait à l'indicateur de l'échelon 1 sans :
 - les cas PMSI invasifs « incidents » non appariés retrouvés parmi les ALD prévalents l'année N,
 - les cas ALD invasifs « incidents » non appariés retrouvés parmi les cas PMSI prévalents l'année N,
 - les cas ALD invasifs « incidents » non appariés ayant un séjour PMSI avec un code *in situ* en DP pour le cancer étudié ;
- le niveau 3, correspondait à l'indicateur de l'échelon 2 sans :
 - les cas PMSI invasifs « incidents » non appariés retrouvés parmi les cas d'ALD pour cancer invasif dans les bases Sniiram des années des années antérieures ;
 - les cas ALD invasifs « incidents » non appariés retrouvés parmi les séjours des bases PMSI des années antérieures ;
- le niveau 4, correspondait à l'indicateur de l'échelon 3 mais avec un chaînage sur le

maximum d'année de recul (quatre ans) des données PMSI. La construction de cet indicateur ne concernait que l'année 2008.

3.7.8 Indicateurs complémentaires calculés

- Délai d'apparition des ALD dans le PMSI parmi les cas appariés.
- Cohérence des informations sur le sexe, l'âge, le département et la région de résidence parmi les cas appariés.
- Pourcentage par année de cas appariés pour chaque localisation.
- Pourcentage par année de cas PMSI et ALD non appariés.

4 Résultats

4.1 Description des ALD issues du Sniiram

4.1.1 Description générale des données du Sniiram transmises par la CnamTS

Le tableau 2 présente l'ensemble des données d'ALD cancer transmises par la CnamTS. 1 753 305 bénéficiaires prévalents (anciennes et nouvelles mises en ALD pour cancer) ont été enregistrés dans le Sniiram en 2006. Parmi eux, 224 718 (13 %) étaient des nouveaux bénéficiaires en 2006 (début d'ALD cancer en 2006).

1 851 273 bénéficiaires prévalents ont été enregistrés en 2007 et 1 942 487 en 2008, dont respectivement 233 194 (13 %) et 234 730 (12 %) nouveaux bénéficiaires. Le tableau 2 présente également les années de début de l'ALD (minimum des dates de début d'ALD renseignées dans la table Sniiram).

Près de 5 % des bénéficiaires présentaient plus d'une localisation cancéreuse. Les bénéficiaires d'ALD cancer enregistrés dans les bases Sniiram provenaient de 10 caisses d'assurance maladie (tableau 2), en majorité des caisses du régime général (78 %). 11 % des bénéficiaires provenaient de caisse d'assurance maladie codée « inconnue », mais représentant *a priori* les sections locales mutualistes du régime général, et les autres régimes spéciaux.

L'annexe 2 décrit en détail les données reçues du Sniiram, et compare les effectifs cumulés aux données d'ALD agrégées.

I Tableau 3 I

Description des bénéficiaires d'ALD pour cancer des bases Sniiram 2006-2008

Bénéficiaire	Année de la base Sniiram		
	2006 (%)	2007 (%)	2008 (%)
Effectifs prévalents^a d'ALD	1 753 305	1 851 273	1 942 487
Effectifs incidents^b d'ALD	224 718 (13)	233 194 (13)	234 730 (12)
Année de début d'ALD			
Avant 2006	1 528 570 (87)	1 397 345 (75)	1 288 638 (66)
2006	224 718 (13)	220 721 (12)	190 503 (10)
2007	3 ^c	233 194 (13)	228 604 (12)
2008	0	0	234 730 (12)
Après 2008 ^c	14	13	12
Nombre d'ALD cancer par bénéficiaire			
Un cancer	1 670 767 (95,3)	1 763 295 (95,2)	1 848 987 (95,2)
Plus de un cancer	82 538 (4,7)	87 978 (4,8)	93 500 (4,8)
Effectif par régime d'assurance maladie			
Régime général	1 373 154 (78,3)	1 458 697 (78,8)	1 540 434 (79,3)
Régime agricole ^d	3 662 (0,2)	3 940 (0,2)	3 636 (0,2)
Régime social des indépendantes	147 047 (8,4)	143 950 (7,8)	141 648 (7,3)
SNCF ^e	176 (0,01)	196 (0,01)	185 (0,01)
RATP ^f	2 920 (0,17)	3 010 (0,16)	2 700 (0,14)
Gens de la mer	275 (0,02)	310 (0,02)	344 (0,02)
Mines	135 (0,01)	150 (0,01)	119 (0,01)
Militaires de carrière	26 447 (1,5)	27 491 (1,5)	28 715 (1,5)
Clercs et employés de notaire	175 (0,01)	194 (0,01)	186 (0,01)
Cavimac (cultes et religion)	543 (0,03)	553 (0,03)	545 (0,03)
Inconnu	198 771 (11,3)	212 782 (11,5)	223 975 (11,5)

^a Anciennes et nouvelles mises en ALD pour cancer, ^b nouvelles mises en ALD pour cancer, ^c erreur de codage

^d Liquidation des actes de certains bénéficiaires du régime agricole (dans certaines zones) réalisée par la CnamTS

^e Société nationale des chemins de fer, ^f Régie autonome de transport parisien

Source : Sniiram.

4.1.2 Description des effectifs de bénéficiaires d'ALD cancer pour les huit localisations cancéreuses étudiées.

Le tableau 3 présente les effectifs de bénéficiaires en ALD dans les bases Sniiram 2006, 2007 et 2008 (cas prévalents), pour chaque localisation étudiée.

Étaient dénombrés ainsi :

- moins de 20 000 cas de cancer du système nerveux central ;
- entre 29 000 et 31 000 cas de cancer du col de l'utérus ;
- entre 40 000 et 45 000 cas de cancer du corps de l'utérus ;
- entre 49 000 et 56 000 cas de cancer du rein ;
- entre 60 000 et 69 000 cas de cancer de la thyroïde ;
- entre 76 000 et 80 000 cas de cancer du poumon ;
- entre 190 000 et 211 000 cas de cancer colorectal ;
- et entre 490 000 et 550 000 cas de cancer du sein.

Les effectifs de bénéficiaires prévalents sont aussi présentés selon l'année de début d'ALD (tableau 3).

La proportion de nouveaux bénéficiaires en ALD cancer (i.e. année de début d'ALD égale à l'année d'enregistrement dans le Sniiram) parmi les cas prévalents d'ALD cancer était similaire ou presque dans les trois bases Sniiram 2006, 2007 et 2008, pour chaque localisation (tableau 3). Elle était de 8 % pour le cancer de la thyroïde, de 9 % pour les cancers du sein et du col de l'utérus, de 11 % pour le cancer du corps de l'utérus, de 13 % pour les cancers du rein et du côlon rectum, de 17 % pour le cancer du système nerveux central et de 27 % pour le cancer du poumon (tableau 3).

I Tableau 4 I

Effectifs de bénéficiaires d'ALD par localisation de cancer selon l'année de début d'ALD dans le Sniiram 2006, 2007 et 2008

Note de lecture : les effectifs de cas ALD « incidents » des bases Sniiram pour une année donnée, correspondent aux valeurs situées à l'intersection de l'année N de la base Sniiram (en colonne) et de l'année N de début de l'ALD (en ligne)

Année de début de l'ALD, par localisation de cancer invasif	Nombre de bénéficiaires d'ALD dans la base Sniiram		
	2006	2007	2008
1- Sein	N = 492 634 (%)	N = 520 237 (%)	N = 546 827 (%)
≤ 2005	445 757 (89,5)	423 494 (81,4)	400 700 (73,5)
2006	46 873 (10,5)	48 791 (9,4)	47 225 (8,6)
2007	1	47 949 (9,2)	49 961 (9,1)
2008	0	0	48 938 (8,8)
> 2008	3	3	3
2- Côlon, rectum	N = 193 687 (%)	N = 202 066 (%)	N = 211 354 (%)
≤ 2005	168 875 (85,3)	151 548 (75,0)	137 006 (64,8)
2006	24 810 (14,7)	24 490 (12,1)	21 158 (10,0)
2007	2	26 028 (12,9)	25 561 (12,1)
2008	0	0	27 629 (13,1)
> 2008	0	0	0
3- Rein	N = 49 452 (%)	N = 52 822 (%)	N = 56 079 (%)
≤ 2005	42 805 (86,6)	39 095 (74,0)	36 054 (64,3)
2006	6 647 (13,4)	6 542 (12,4)	5 708 (10,2)
2007	0	7 185 (13,6)	7 093 (12,6)
2008	0	0	7 224 (12,9)
> 2008	0	0	0
4- Poumon	N = 72 955 (%)	N = 76 315 (%)	N = 79 778 (%)
≤ 2005	53 060 (72,7)	38 806 (50,9)	31 484 (39,6)
2006	19 895 (27,3)	16 481 (21,6)	9 012 (11,1)
2007	0	21 028 (27,5)	17 291 (21,7)
2008	0	0	21 991 (27,6)
> 2008	0	0	0
5- Système nerveux central	N = 18 156 (%)	N = 18 686 (%)	N = 19 487 (%)
≤ 2005	15 073 (83,0)	12 791 (68,5)	11 452 (58,8)
2006	3 083 (17,0)	2 640 (14,1)	1 695 (8,7)
2007	0	3 255 (17,4)	2 885 (14,8)
2008	0	0	3 455 (17,7)
> 2008	0	0	0
6- Col de l'utérus	N = 29 154 (%)	N = 29 902 (%)	N = 30 735 (%)
≤ 2005	26 423 (90,6)	24 262 (81,2)	22 446 (73,0)
2006	2 730 (9,4)	2 732 (9,1)	2 365 (7,7)
2007	0	2 907 (9,7)	2 915 (9,5)
2008	0	0	3 008 (9,8)
> 2008	1	1	1
7- Corps l'utérus	N = 40 966 (%)	N = 43 057 (%)	N = 45 122 (%)
≤ 2005	36 393 (88,8)	33 540 (77,9)	31 103 (68,9)
2006	4 573 (11,2)	4 608 (10,7)	4 134 (9,2)
2007	0	4 909 (11,4)	4 911 (10,9)
2008	0	0	4 974 (11,0)
> 2008	0	0	0
8- Thyroïde	N = 60 494 (%)	N = 64 605 (%)	N = 68 902 (%)
≤ 2005	55 216 (91,3)	53 507 (82,8)	51 870 (75,3)
2006	5 276 (8,7)	5 509 (8,5)	5 361 (7,8)
2007	0	5 587 (8,7)	5 851 (8,5)
2008	0	0	5 818 (8,4)
> 2008	2	2	2

Source : Sniiram.

Le tableau 4 présente les effectifs de bénéficiaires des bases Sniiram 2006, 2007 et 2008 ayant débuté leur ALD en 2006, 2007 et 2008, pour chacune des 8 localisations étudiées. La distribution des effectifs selon la première année d'apparition dans le Sniiram est présentée, et permet d'apprécier le décalage entre année de début d'ALD et année d'apparition dans le Sniiram. Ainsi, sur les 49 891 bénéficiaires ayant eu une ALD pour cancer du sein débutée en 2006, 46 873 (94 %) sont apparus dans le Sniiram dès l'année 2006, 5 % sont apparus pour la première fois dans le Sniiram en 2007 et 1 % en 2008. Cette distribution est similaire pour les autres localisations.

Les effectifs des cas ayant débuté leur ALD en 2007 ou en 2008 et apparaissant pour la première fois respectivement dans le Sniiram 2007 et 2008 étaient très légèrement inférieurs aux effectifs rapportés dans le tableau 3. La différence correspondait aux cas apparaissant aussi dans le Sniiram des années antérieures, et correspondaient donc à des incohérences (ou « doublons »). Ces cas ont été exclus pour la suite des analyses.

I Tableau 5 I

Distribution des bénéficiaires ayant débuté leur ALD cancer en 2006, 2007 et 2008 selon la première année d'apparition dans le Sniiram

Note de lecture : pour chaque localisation de cancer (en ligne), l'effectif total de début d'ALD une année donnée (en colonne) se décompose en effectifs par première apparition dans les bases Sniiram (en ligne).

Année d'enregistrement du cancer (invasif) dans la base Sniiram	Effectif de nouveaux bénéficiaires d'ALD cancer		
	ALD cancer débutée en 2006	ALD cancer débutée en 2007	ALD cancer débutée en 2008
1- Sein	N = 49 891 (%)	N = 50 585 (%)	N = 48 727 (%)
Sniiram 2006	46 873 (94)	1	-
Sniiram 2007	2 687 (5)	47 834 (95)	-
Sniiram 2008	331 (<1)	2 750 (5)	48 727 (100)
2- Côlon, rectum	N = 26 374 (%)	N = 27 478 (%)	N = 27 541 (%)
Sniiram 2006	24 810 (94)	2	-
Sniiram 2007	1 464 (6)	25 995 (95)	-
Sniiram 2008	100 (<1)	1481 (5)	27 541 (100)
3- Rein	N = 7 092 (%)	N = 7 609 (%)	N = 7 212 (%)
Sniiram 2006	6 647 (94)	-	-
Sniiram 2007	423 (6)	7176 (94)	-
Sniiram 2008	22 (<1)	433 (6)	7212 (100)
4- Poumon	N = 21 297 (%)	N = 22 372 (%)	N = 21 971 (%)
Sniiram 2006	19 895 (93)	-	-
Sniiram 2007	1 334 (6)	21 021 (94)	-
Sniiram 2008	68 (<1)	1 351 (6)	21 971 (100)
5- Système nerveux central	N = 3 313 (%)	N = 3 475 (%)	N = 3 450 (%)
Sniiram 2006	3 083 (93)	-	-
Sniiram 2007	205 (6)	3 253 (94)	-
Sniiram 2008	25 (1)	222 (6)	3 450 (100)
6- Col de l'utérus	N = 2 897 (%)	N = 3 085 (%)	N = 3 004 (%)
Sniiram 2006	2 730 (94)	-	-
Sniiram 2007	153 (5)	2 904 (94)	-
Sniiram 2008	14 (<1)	181 (6)	3004 (100)
7- Corps l'utérus	N = 4 868 (%)	N = 5 192 (%)	N = 4 965 (%)
Sniiram 2006	4 573 (94)	-	-
Sniiram 2007	270 (6)	4 903 (94)	-
Sniiram 2008	25 (<1)	289 (6)	4 965 (100)
8- Thyroïde	N = 5 628 (%)	N = 5 883 (%)	N = 5 777 (%)
Sniiram 2006	5 276 (94)	-	-
Sniiram 2007	320 (6)	5 558 (94)	-
Sniiram 2008	32 (<1)	325 (6)	5 777 (100)

Source : Sniiram

4.2 Description du PMSI

4.2.1 Identifiants manquants

Le pourcentage de séjours pour cancer avec identifiant (i.e. numéro d'anonymat) manquant variait avec l'année et le mode de sélection des séjours. Pour les sélections effectuées exclusivement sur le DP, le pourcentage d'identifiants manquants se situait entre 1 % et 5 % selon la localisation de 2004 à 2006 et était inférieur à 0,1 % pour toutes les localisations en 2007 et 2008 (tableau 5).

Lorsque les DR et DAS étaient associés aux sélections des séjours, une plus grande variation du pourcentage d'identifiants manquants était observée entre les localisations. De 2004 à 2008, le pourcentage variait de 1 % à 12 % (tableau 5). Le pourcentage élevé de séjours avec identifiants manquants, observé à partir des sélections incluant à la fois DP, DR ou DAS pourrait s'expliquer par le nombre important de codes séances (chimiothérapie, radiothérapie...) accompagnant les DR. Néanmoins, l'exhaustivité du codage de cet identifiant s'améliore.

I Tableau 6 I

Pourcentage de séjours avec identifiant manquant par localisation et type de sélection, 2004-2008

Localisation cancéreuse invasive	2004	2005	2006	2007	2008
Sélectionnée exclusivement en DP					
1- Sein	4,8	2,6	4,6	0,05	0,03
2- Côlon rectum	2,4	2,7	3,0	0,05	0,02
3- Rein	1,9	2,0	2,6	0,02	0,02
4- Poumon	2,9	4,2	4,7	0,07	0,05
5- Système nerveux central	2,8	3,3	5,4	0,05	0,05
6- Col de l'utérus	5,0	3,9	3,6	0,09	0,07
7- Corps de l'utérus	3,2	3,3	3,1	0,04	0,03
8- Thyroïde	1,9	1,4	2,1	0,02	0,05
Sélectionnée en DP, DR ou DAS					
1- Sein	11,6	5,0	8,9	6,4	5,2
2- Côlon rectum	3,5	3,7	3,9	1,6	1,1
3- Rein	4,1	3,4	3,8	1,3	0,8
4- Poumon	5,1	4,2	5,0	2,9	2,1
5- Système nerveux central	7,0	4,9	7,4	8,9	7,2
6- Col de l'utérus	9,8	7,3	5,6	7,4	5,8
7- Corps de l'utérus	7,6	5,6	6,5	6,2	4,0
8- Thyroïde	3,5	2,2	3,1	2,8	1,5

Source : PMSI national.

4.2.2 Effectifs de patients prévalents et de patients « incidents » après chaînage sur les années antérieures

Les effectifs de cas prévalents de cancer sélectionnés en DP dans le PMSI 2006-2008 sont décrits dans le tableau 6 (première ligne de résultat pour chaque localisation). Ces effectifs de cas « prévalents » correspondaient au nombre de patients ayant eu un séjour hospitalier en 2006, 2007 et 2008 avec la localisation étudiée codée en DP.

Le tableau 6 présente aussi les effectifs « d'incidence » après exclusion des patients « prévalents » ayant eu aussi un séjour codé en DP, DR ou DAS pour la localisation considérée les années (une à quatre selon l'année considérée) antérieures (« Effectifs après chaînage en DP, DR ou DAS avec les données PMSI des années antérieures »). Le chaînage des séjours prévalents de l'année considérée avec les séjours PMSI de la première année antérieure permettait ainsi de repérer selon la localisation de 8 % (cancer du rein, du corps de l'utérus) à 20 % (cancers du poumon, du SNC, de la thyroïde) de patients avec anciens séjours parmi les patients avec un séjour l'année considérée.

Le chaînage consécutif des séjours issus du premier chaînage sans les anciens séjours, avec les données PMSI de la deuxième année antérieure a permis de repérer à nouveau selon la localisation 1 % à 4 % de patients avec des anciens séjours (tableau 6). En d'autres termes, le chaînage sur deux années consécutives de séjours pour cancers permettait d'obtenir environ 96 % à 99 % de séjours pouvant être

considérés comme des patients avec premier séjour hospitalier pour la localisation de cancer considérée i.e. cas « incidents ». Le chaînage sur trois et quatre ans aboutissait à 99 % et plus de patients avec premier séjour hospitalier pour cancer (tableau 6).

I Tableau 7 I

Effectif de patients avec séjours pour cancer en DP en 2006, 2007 et 2008 et effectifs après chaînage sur les séjours pour cancer en DP, DR, DAS des années antérieures

Note de lecture : pour chaque localisation, les effectifs et pourcentage de premier séjour de obtenus après chaque chaînage sont calculés à partir des effectifs de la ligne immédiatement au dessus.

Effectifs de patients par localisation de cancer et par année (pourcentage premier séjour hospitalier)			
Localisation cancéreuse	2006	2007	2008
1- Sein			
- Cas prévalents sélectionnés en DP	55 742	56 996	58 176
- Effectifs ^a après chaînage en DP, DR ou DAS avec les données PMSI des années antérieures :			
1 an	49 457 (88,7)	51 064 (89,6)	52 357 (90)
2 ans	48 448 (98,0)	49 962 (97,8)	51 421 (98,2)
3 ans	-	49 285 (98,6)	50 734 (98,7)
4 ans	-	-	50 227 (99,0)
2- Côlon rectum			
- Cas prévalents sélectionnés en DP	51 345	52 138	53 234
- Effectifs ^a après chaînage en DP, DR ou DAS avec les données PMSI des années antérieures :			
1 an	42 955 (83,6)	43 768 (84,0)	45 009 (84,5)
2 ans	42 034 (97,8)	42 895 (98,0)	44 245 (98,3)
3 ans	-	42 450 (99,0)	43 795 (99,0)
4 ans	-	-	43 556 (99,5)
3- Rein			
- Cas prévalents sélectionnés en DP	11 456	12 235	12 362
- Effectifs ^a après chaînage en DP, DR ou DAS avec les données PMSI des années antérieures :			
1 an	10 521 (91,8)	11 262 (92,0)	11 364 (91,9)
2 ans	10 394 (98,8)	11 090 (98,5)	11 190 (98,5)
3 ans	-	11 018 (99,3)	11 084 (99,0)
4 ans	-	-	11 023 (99,4)
4- Poumon			
- Cas prévalents sélectionnés en DP	37 159	35 982	36 352
- Effectifs ^a après chaînage en DP, DR ou DAS avec les données PMSI des années antérieures :			
1 an	29 835 (80,3)	29 283 (81,4)	29 905 (82,3)
2 ans	29 245 (98,0)	28 726 (98,1)	29 388 (98,3)
3 ans	-	28 484 (99,1)	29 142 (99,2)
4 ans	-	-	28 987 (99,5)
5- Système nerveux central			
- Cas prévalents sélectionnés en DP	7 577	7 976	8 144
- Effectifs ^a après chaînage en DP, DR ou DAS avec les données PMSI des années antérieures :			
1 an	6 117 (80,7)	6 435 (80,7)	6 661 (81,8)
2 ans	5 977 (97,7)	6 229 (96,8)	6 491 (97,4)
3 ans	-	6 132 (98,4)	6 400 (98,6)
4 ans	-	-	6 329 (98,9)

6- Col de l'utérus

- Cas prévalents sélectionnés en DP	4 076	4 099	4 072
- Effectifs ^a après chaînage en DP, DR ou DAS avec les données PMSI des années antérieures :			
1 an	3 390 (83,2)	3 429 (83,6)	3 382 (83,0)
2 ans	3 336 (98,4)	3 378 (98,5)	3 332 (98,5)
3 ans	-	3 354 (99,3)	3 311 (99,4)
4 ans	-	-	3 291 (99,4)

7- Corps de l'utérus

- Cas prévalents sélectionnés en DP	7 265	7 551	7 547
- Effectifs ^a après chaînage en DP, DR ou DAS avec les données PMSI des années antérieures :			
1 an	6 661 (91,7)	6 862 (91,0)	6 935 (91,9)
2 ans	6 606 (99,2)	6 791 (99,0)	6 867 (99,0)
3 ans	-	6 767 (99,7)	6 828 (99,4)
4 ans	-	-	6 811 (99,7)

8- Thyroïde

- Cas prévalents sélectionnés en DP	9 024	9 001	9 460
- Effectifs ^a après chaînage en DP, DR ou DAS avec les données PMSI des années antérieures :			
1 an	7 197 (79,7)	7 362 (81,8)	7 804 (82,5)
2 ans	6 894 (95,8)	7 070 (96,0)	7 554 (96,8)
3 ans	-	6 938 (98,1)	7 427 (98,3)
4 ans	-	-	7 369 (99,2)

^a Effectifs déduits du nombre de cas pour une même localisation retrouvée en DP, DR les années antérieures.

Source : PMSI national.

4.2.3 Effet du choix de l'algorithme de sélection de cas prévalents sur le nombre de cas « incidents » de cancer

Les effectifs de cas « prévalents » de cancer sélectionnés à partir du PMSI augmentaient lorsque le DR et le DAS étaient introduits en plus du DP dans l'algorithme (tableau 7). Les effectifs de patients avec séjours pour cancer obtenus avec les algorithmes DP, DR ou DAS après quatre années de chaînage étaient largement supérieurs aux effectifs issus des algorithmes basés sur la sélection exclusive des cas en DP pour lesquels les valeurs sont plus proches des estimations d'incidence nationale. Ainsi, les effectifs obtenus par les algorithmes basés sur DP ou DR étaient supérieurs de 9 % à 38 % selon la localisation, aux effectifs issus de l'algorithme basé exclusivement sur la sélection des séjours en DP (tableau 7).

Le chaînage sur un an des séjours sélectionnés exclusivement en DP montrait, pour le PMSI 2008, un pourcentage de 82 % à 92 % de patients avec premier séjour hospitalier pour cancer selon la localisation (tableau 7). Ce pourcentage variait de 65 % à 83 % pour les séjours sélectionnés à partir des algorithmes en DP, DR ou DAS. Il atteignait sur quatre années de chaînage successif, plus de 99 % pour les sélections de séjours réalisées en DP, plus de 98 % pour les sélections réalisées en DP, DR et 97 % pour les sélections réalisées en DP, DR ou DAS (tableau 7). Ce second algorithme permettrait de mieux sélectionner les cas prévalents et par conséquent les « incidents ». Mais finalement seules les deux années antérieures ont été retenues pour sélectionner les « incidents » par la suite pour des motifs liés à la comparabilité entre les données.

I Tableau 8 I

Effet de l'algorithme de sélection des cas sur le nombre de cas prévalents et sur les effectifs après chaînage sur le DP, DR ou DAS, données 2008

N.B : la valeur du dénominateur qui a servi à calculer le pourcentage de premier séjour hospitalier, correspond à l'effectif à partir duquel le chaînage a été effectué pour chaque localisation (effectif de la ligne immédiatement au dessus du pourcentage considéré)

Localisation cancéreuse Effectifs de cas prévalents et effectifs après chaînage en DP, DR ou DAS sur les années antérieures	Effectifs par localisation de cancer et par algorithme de sélection (pourcentage de premier séjour hospitalier)		
	Sélection en DP	Sélection en DP ou DR	Sélection en DP, DR ou DAS
1- Sein			
Patients prévalents	58 176	100 701	126 524
Chaînés sur 1 an	52 357 (90,0)	67 626 (67,0)	85 210 (67,3)
Chaînés sur 2 ans	51 421 (98,2)	64 710 (95,7)	79 491 (93,3)
Chaînés sur 3 ans	50 734 (98,7)	62 923 (97,2)	75 829 (95,4)
Chaînés sur 4 ans	50 227 (99,0)	61 727 (98,1)	73 321 (96,7)
2- Côlon rectum			
Patients prévalents	53 234	83 987	104 233
Chaînés sur 1 an	45 009 (84,5)	55 162 (65,7)	67 425 (64,7)
Chaînés sur 2 ans	44 245 (98,3)	52 321 (94,8)	62 498 (92,7)
Chaînés sur 3 ans	43 795 (99,0)	50 672 (96,8)	59 708 (95,5)
Chaînés sur 4 ans	43 556 (99,5)	49 865 (98,4)	58 160 (97,4)
3- Rein			
Patients prévalents	12 362	15 314	22 272
Chaînés sur 1 an	11 364 (91,9)	12 776 (83,4)	17 597 (79,0)
Chaînés sur 2 ans	11 190 (98,5)	12 350 (96,7)	16 533 (94,0)
Chaînés sur 3 ans	11 084 (99,0)	12 139 (98,3)	15 946 (96,4)
Chaînés sur 4 ans	11 023 (99,4)	12 016 (99,0)	15 580 (97,7)
4- Poumon			
Patients prévalents	36 352	65 244	79 942
Chaînés sur 1 an	29 905 (82,3)	42 320 (64,9)	52 091(65,2)
Chaînés sur 2 ans	29 388 (98,3)	40 917 (96,7)	49 789 (95,6)
Chaînés sur 3 ans	29 142 (99,2)	40 386 (98,7)	48 848 (98,1)
Chaînés sur 4 ans	28 987 (99,5)	40 100 (99,3)	48 311 (98,9)
5- Col de l'utérus			
Patients prévalents	4 072	6 374	8 380
Chaînés sur 1 an	3 382 (83,0)	4 617 (72,4)	5 978 (71,3)
Chaînés sur 2 ans	3 332 (98,5)	4 497 (97,4)	5 730 (95,8)
Chaînés sur 3 ans	3 311 (99,4)	4 440 (98,7)	5 590 (97,5)
Chaînés sur 4 ans	3 291 (99,4)	4 393 (98,9)	5 499 (98,4)
6- Système nerveux central			
Patients prévalents	8 144	11 849	15 623
Chaînés sur 1 an	6 661 (81,8)	8 380 (70,7)	11 480 (73,5)
Chaînés sur 2 ans	6 491 (97,4)	8 091 (96,5)	11 045 (96,2)
Chaînés sur 3 ans	6 400 (98,6)	7 937 (98,1)	10 810 (97,9)
Chaînés sur 4 ans	6 329 (98,9)	7 836 (98,7)	10 658 (98,6)
7- Corps de l'utérus			
Patients prévalents	7 547	10 847	14 625
Chaînés sur 1 an	6 935 (91,9)	8 368 (77,1)	11 152 (76,2)
Chaînés sur 2 ans	6 867 (99,0)	8 145 (97,3)	10 643 (95,4)
Chaînés sur 3 ans	6 828 (99,4)	8 026 (98,5)	10 380 (97,5)
Chaînés sur 4 ans	6 811 (99,7)	7 960 (99,2)	10 214 (98,4)
8- Thyroïde			
Patients prévalents	9 460	11 333	13 280
Chaînés sur 1 an	7 804 (82,5)	8 670 (76,5)	10 034 (75,6)
Chaînés sur 2 ans	7 554 (96,8)	8 250 (95,2)	9 415 (93,8)
Chaînés sur 3 ans	7 427 (98,3)	8 017 (97,2)	9 069 (96,3)
Chaînés sur 4 ans	7 369 (99,2)	7 912 (98,7)	8 869 (97,8)

Source : PMSI national.

4.3 Description des effectifs « incidents » croisés

Au final, les effectifs croisés sont ceux des localisations cancéreuses invasives sélectionnées exclusivement en DP par l'algorithme dans le PMSI et les effectifs ALD cancer identifiés dans le Sniiram.

4.3.1 Effectifs par localisation de cancer issus des sélections dans les bases

Quelle que soit la localisation, les effectifs de patients issus de la sélection des séjours en DP dans le PMSI étaient plus élevés que les effectifs de bénéficiaires d'ALD issus du Sniiram (tableaux 8 et 9). La différence entre l'effectif total PMSI et l'effectif total ALD par localisation de 2006 à 2008 variait selon la localisation de 3 % (cancer du sein) à 39 % (cancer du côlon rectum) (données non montrées).

I Tableau 9 I

Effectifs « incident » PMSI à croiser avec les ALD « incidentes »

Séjour année N en DP sans séjours année N-1, N-2 en DP, DR et DAS dans le PMSI	Sein	Côlon rectum	Rein	Poumon	SNC	Col de l'utérus	Corps de l'utérus	Thyroïde
2006 (sans 2004, 2005)	48 448	42 034	10 394	29 245	5 977	3 336	6 606	6 894
2007 (sans 2005, 2006)	49 962	42 895	11 090	28 726	6 229	3 378	6 791	7 070
2008 (sans 2006, 2007)	51 421	44 245	11 190	29 388	6 491	3 332	6 867	7 554
Total	149 831	129 174	32 674	87 359	18 697	10 046	20 264	21 518

Source : PMSI.

I Tableau 10 I

Effectifs « incident » ALD à croiser avec le PMSI « incident »

Base Sniiram	Sein	Côlon rectum	Rein	Poumon	SNC	Col de l'utérus	Corps de l'utérus	Thyroïde
ALD débutée en 2006 dans la base Sniiram 2006	46 873	24 810	6 647	19 895	3 083	2 730	4 573	5 276
ALD débutée en 2007 dans la base Sniiram 2007	47 834	25 995	7 176	21 021	3 253	2 904	4 903	5 558
ALD débutée en 2008 dans la base Sniiram 2008	48 727	27 541	7 212	21 971	3 450	3 004	4 965	5 777
Total	143 434	78 346	21 035	62 887	9 786	8 638	14 441	16 611

Source : Sniiram.

4.3.2 Effectifs croisés

Le croisement des effectifs incidents entre ALD et PMSI conduisait à 3 types de cas (figures 7 à 14) :

- les cas ALD et PMSI appariés la même année ;
- les cas ALD non appariés au PMSI l'année N ;
- les cas PMSI non appariés aux ALD l'année N.

La proportion de cas ALD et PMSI appariés variait selon la localisation mais était similaire en 2006, 2007 et 2008 pour chaque localisation. Elle était de moins de 30 % pour le cancer du col de l'utérus, autour de 33 % pour les cancers du rein et du poumon, autour de 40 % pour les cancers du côlon, rectum, du corps de l'utérus et de la thyroïde, et près de 50 % pour le cancer du sein.

I Figure 8 I

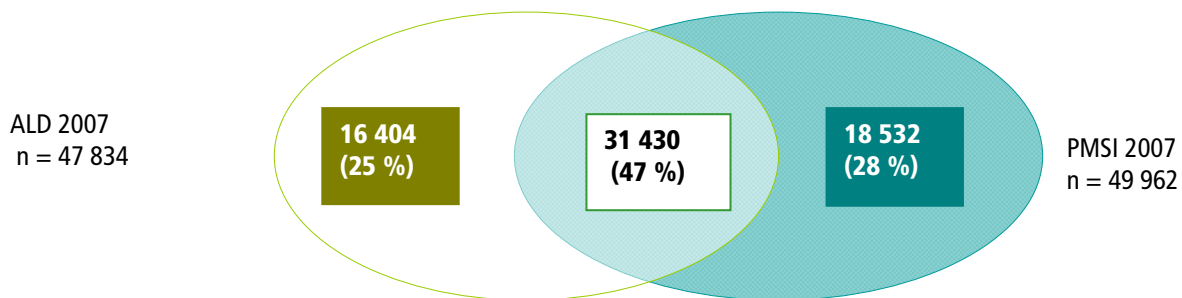
Effectifs croisés de cas « incidents » de cancer du sein 2006, 2007 et 2008



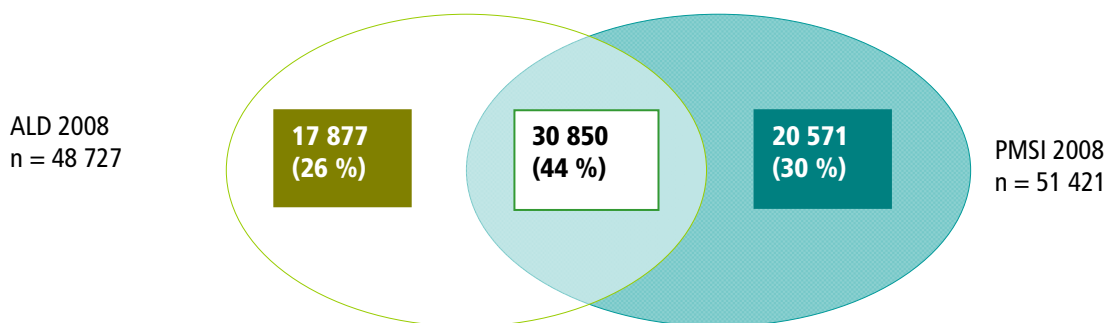
2006 : total union PMSI, ALD = 65 450



2007 : total union PMSI, ALD = 66 366



2008 : total union PMSI, ALD = 69 298

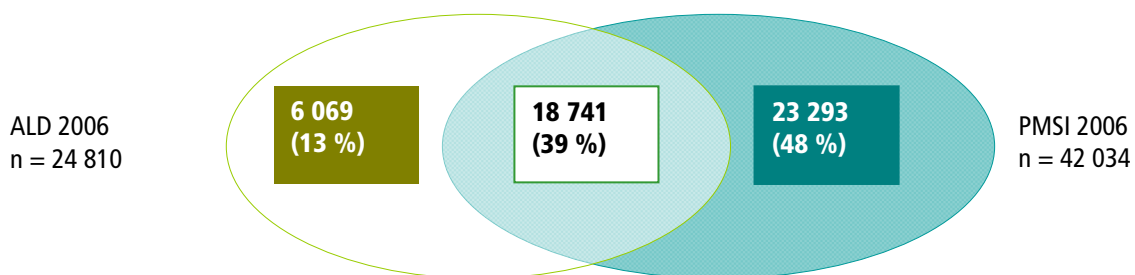


I Figure 9 I

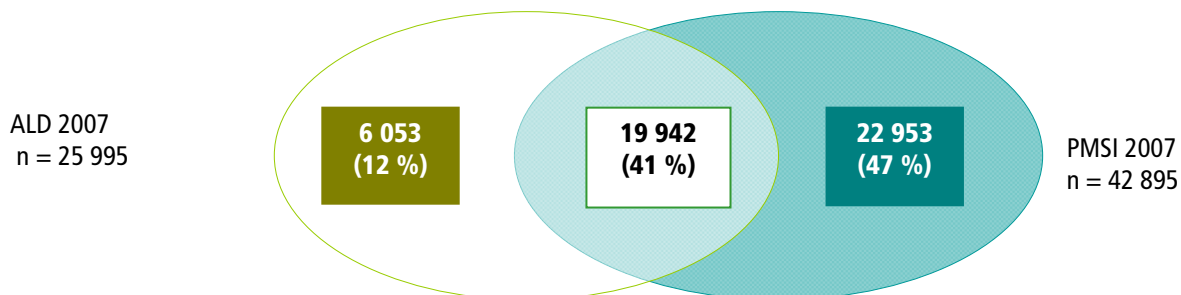
Effectifs croisés de cas « incidents » de cancer du côlon rectum 2006, 2007 et 2008



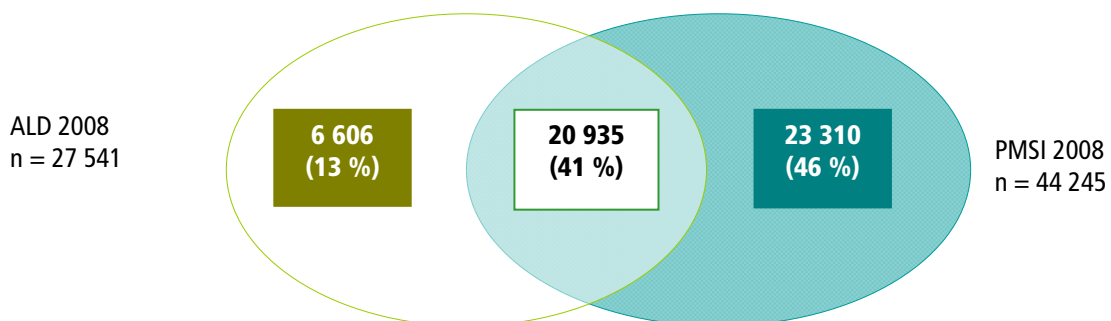
2006 : total union PMSI, ALD = 48 103



2007 : total union PMSI, ALD = 48 948



2008 : total union PMSI, ALD= 50 851

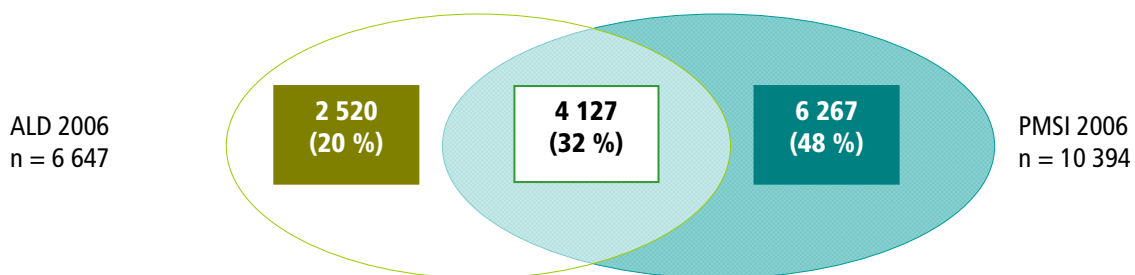


I Figure 10 I

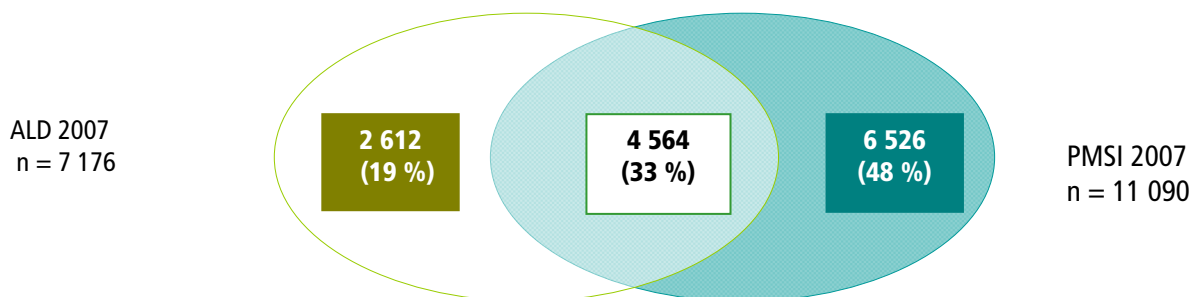
Effectifs croisés de cas « incidents » de cancer du rein 2006, 2007 et 2008



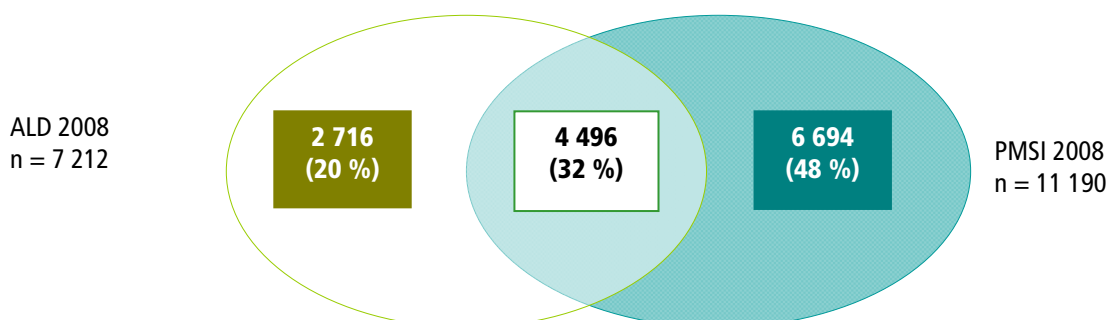
2006 : total union PMSI, ALD = 12 914



2007 : total union PMSI, ALD = 13 702



2008 : total union PMSI, ALD = 13 906

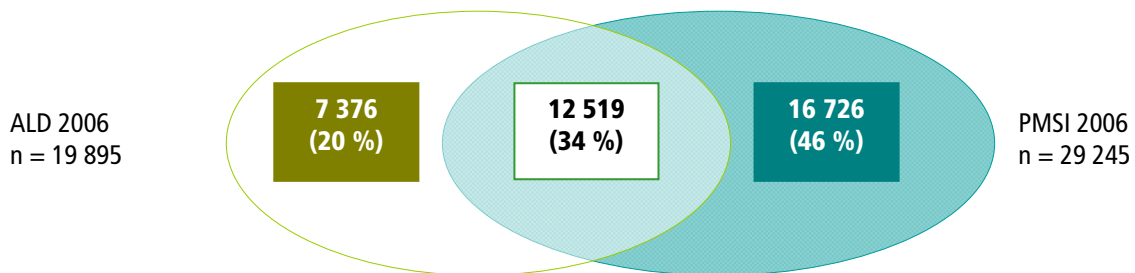


I Figure 11 I

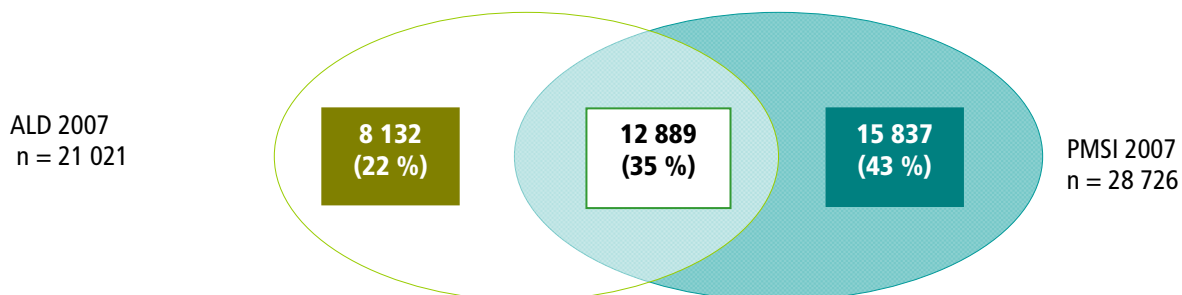
Effectifs croisés de cas « incidents » de cancer du poumon 2006, 2007 et 2008



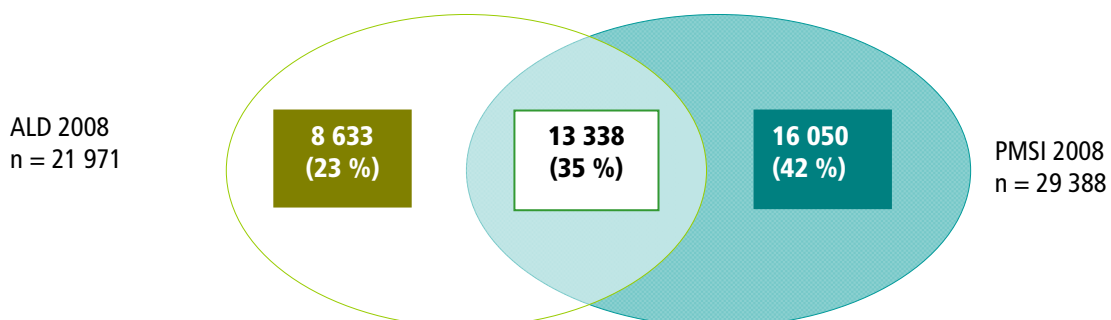
2006 : total union PMSI, ALD = 36 621



2007 : total union PMSI, ALD = 36 858



2008 : total union PMSI, ALD = 38 021

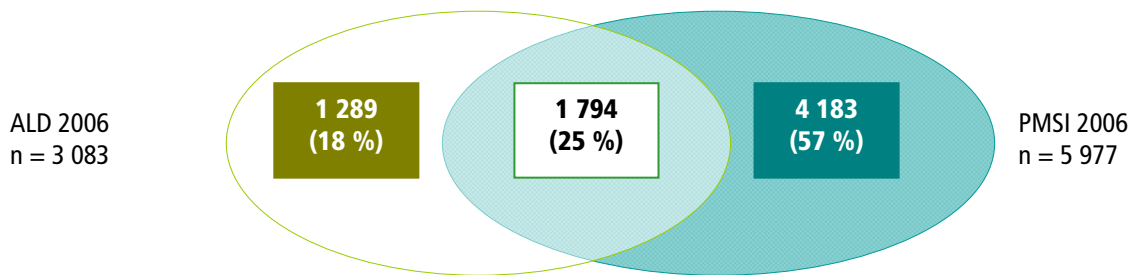


I Figure 12 I

Effectifs croisés de cas « incidents » de cancer du système nerveux central 2006, 2007 et 2008



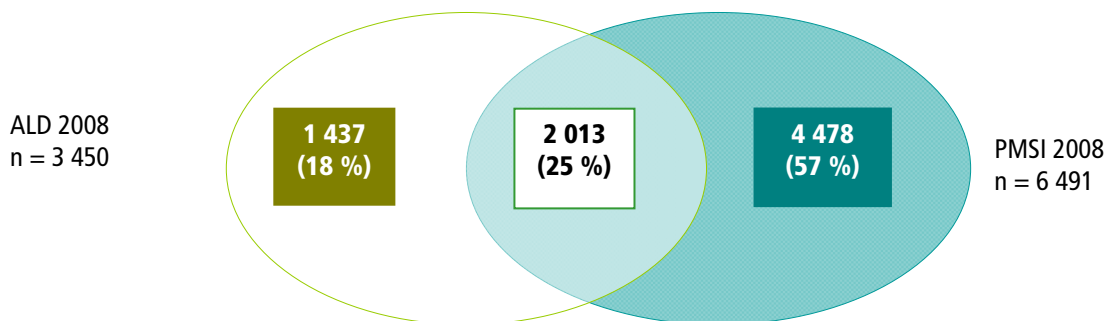
2006 : total union PMSI, ALD = 7 266



2007 : total union PMSI, ALD = 7 538



2008 : total union PMSI, ALD = 7 928

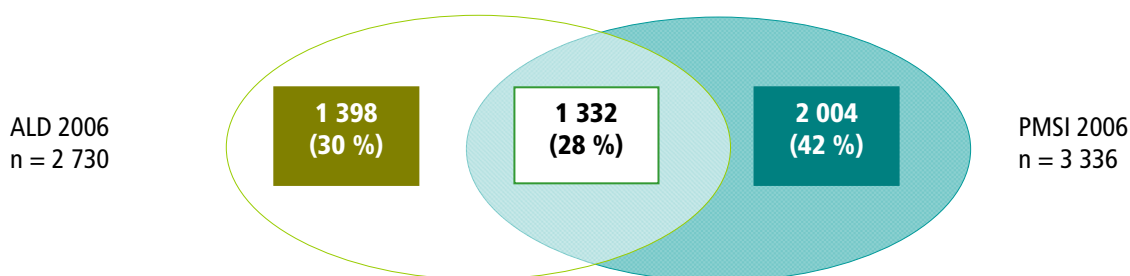


I Figure 13 I

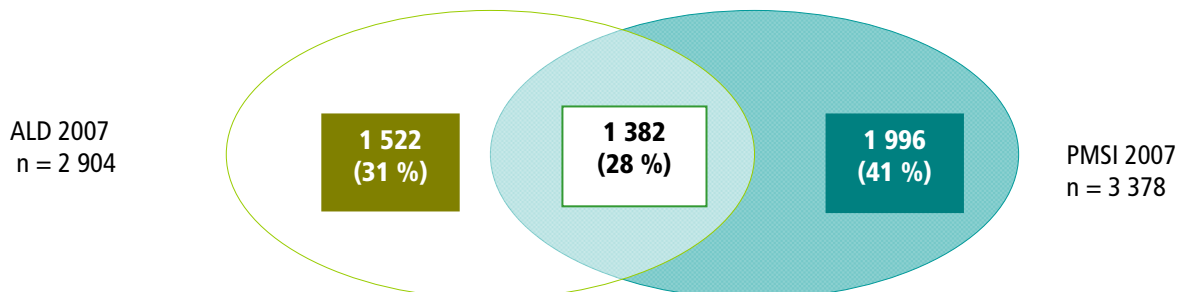
Effectifs croisés de cas « incidents » de cancer du col de l'utérus 2006, 2007 et 2008



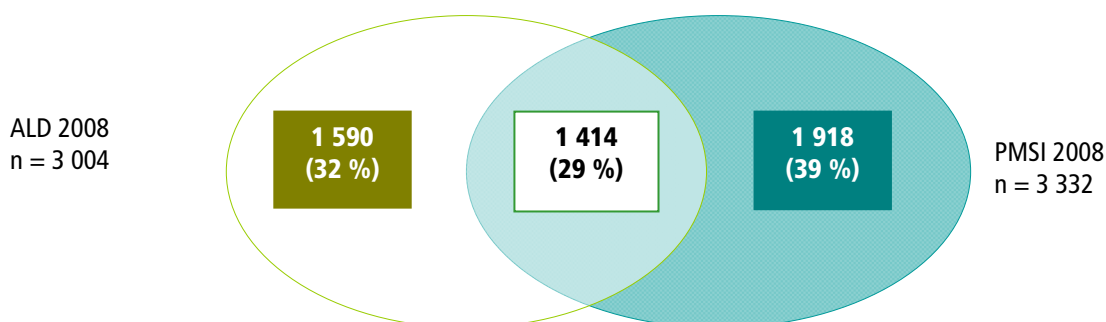
2006 : total union PMSI, ALD = 4 734



2007 : total union PMSI, ALD = 4 900



2008 : total union PMSI, ALD = 4 922



I Figure 14 I

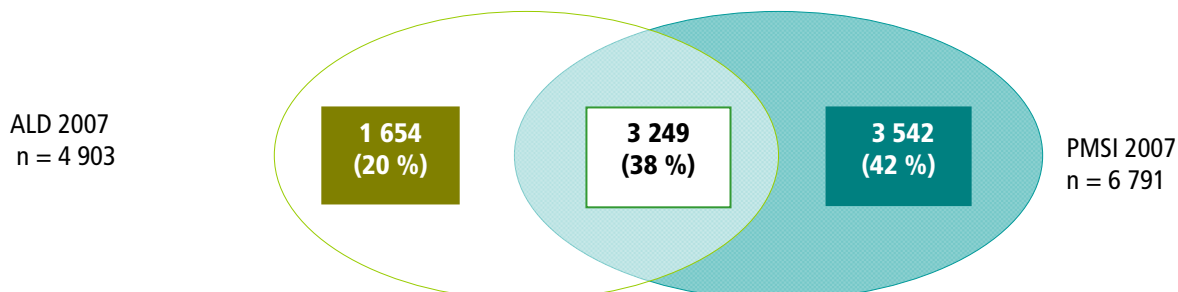
Effectifs croisés de cas « incidents » de cancer du corps de l'utérus 2006, 2007 et 2008



2006 : total union PMSI, ALD = 8 211



2007 : total union PMSI, ALD = 8 445



2008 : total union PMSI, ALD = 8 537



I Figure 15 I

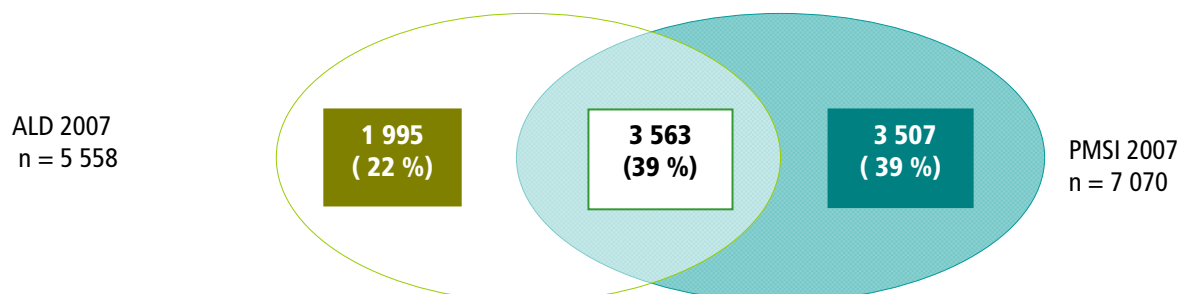
Effectifs croisés de cas « incidents » de cancer de la thyroïde 2006, 2007 et 2008



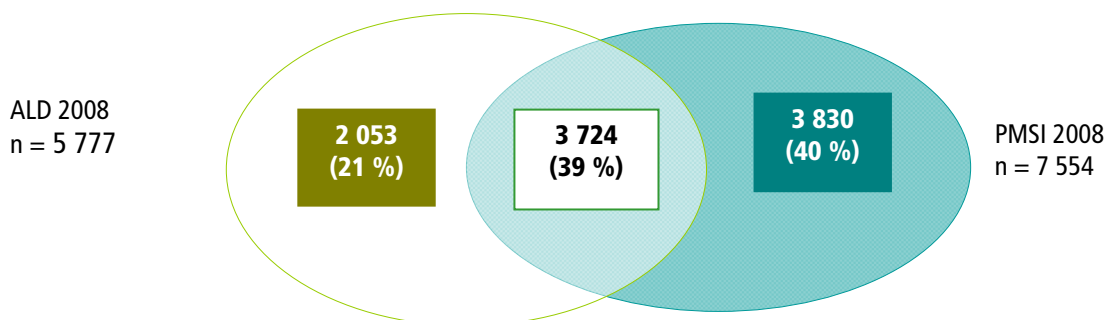
2006 : total union PMSI, ALD = 8 918



2007 : total union PMSI, ALD = 9 065



2008 : total union PMSI, ALD = 9 607



Les tableaux 10 à 17 présentent les effectifs « incidents » croisés par année. Ainsi, sur les 46 873 patientes mises en ALD pour cancer du sein en 2006, 29 871 ont eu un séjour PMSI « incident » en 2006, 3 507 un séjour « incident » en 2007, 115 un séjour « incident » en 2008, et 13 380 n'ont pas eu du tout de séjour PMSI « incident » sur la période 2006-2008. De même, sur les 48 448 patients ayant un séjour PMSI « incident » en 2006, 29 871 ont eu une ALD débutée en 2006, 290 en 2007, 145 en 2008, et 18 142 n'ont pas de mise en ALD pour cancer du sein rapportée dans le Sniiram 2006-2008.

I Tableau 11 I

Cancer du sein, effectifs appariés par année et cas absents 2006-2008

		Cas ALD « incidents » de cancer du sein			Cas PMSI absents des ALD « incidentes » 2006-2008
		2006 N=46 873	2007 N=47 834	2008 N=48 727	
Cas PMSI « incidents » de cancer du sein	2006 N=48 448	29 871	290	145	18 142
	2007 N=49 962	3 507	31 430	261	14 764
	2008 N=51 421	115	3 473	30 850	16 983
	Cas ALD absents du PMSI « incidents » 2006-2008	13 380	12 641	17 471	

Total cas ALD cancer du sein « incidents » non appariés aux cas PMSI incidents en 2006 = 3 507 + 115 + 13 380 = **17 002**
 Total cas ALD cancer du sein « incidents » non appariés aux cas PMSI incidents en 2007 = 290 + 3 473 + 12 641 = **16 404**
 Total cas ALD cancer du sein « incidents » non appariés aux cas PMSI incidents en 2008 = 145 + 261 + 17 471 = **17 877**
 Total cas PMSI cancer du sein « incidents » non appariés aux cas ALD incidents en 2006 = 290 + 145 + 18 142 = **18 577**
 Total cas PMSI cancer du sein « incidents » non appariés aux cas ALD incidents en 2007 = 3 507 + 261 + 14 764 = **18 532**
 Total cas PMSI cancer du sein « incidents » non appariés aux cas ALD incidents en 2008 = 115 + 3 473 + 16 983 = **20 571**
 Source : PMSI national et Sniiram.

I Tableau 12 I

Cancer colorectal, effectifs appariés par année et cas absents 2006-2008

		Cas ALD « incidents » de cancer colorectal			Cas PMSI absents des ALD « incidentes » 2006-2008
		2006 N=24 810	2007 N=25 995	2008 N= 27 541	
Cas PMSI « incidents » de cancer colo rectal	2006 N=42 034	18 741	608	282	22 403
	2007 N=42 895	923	19 942	646	21 384
	2008 N=44 245	35	904	20 935	22 371
	Cas ALD absents du PMSI « incidents » 2006-2008	5 111	4 541	5 678	

Total cas ALD cancer colorectal « incidents » non appariés aux cas PMSI incidents en 2006 = 923 + 35 + 5 111 = **6 069**
 Total cas ALD cancer colorectal « incidents » non appariés aux cas PMSI incidents en 2007 = 608 + 904 + 4 541 = **6 053**
 Total cas ALD cancer colorectal « incidents » non appariés aux cas PMSI incidents en 2008 = 282 + 646 + 5 678 = **6 606**
 Total cas PMSI cancer colorectal « incidents » non appariés aux cas ALD incidents en 2006 = 608 + 282 + 22 403 = **23 293**
 Total cas PMSI cancer colorectal « incidents » non appariés aux cas ALD incidents en 2007 = 923 + 646 + 21 384 = **22 953**
 Total cas PMSI cancer colorectal « incidents » non appariés aux cas ALD incidents en 2008 = 35 + 904 + 22 371 = **23 310**
 Source : PMSI national et Sniiram.

I Tableau 13 I

Cancer du rein, effectifs appariés par année et cas absents 2006-2008

		Cas ALD « incidents » de cancer du rein			Cas PMSI absents des ALD « incidentes » 2006-2008
		2006 N=6 647	2007 N=7 176	2008 N= 7 212	
Cas PMSI « incidents » de cancer du rein	2006 N=10 394	4 127	236	117	5 914
	2007 N=11 090	443	4 564	210	5 873
	2008 N=11 190	17	433	4 496	6 244
	Cas ALD absents du PMSI « incidents » 2006-2008	2 060	1 943	2 389	

Total cas ALD cancer du rein « incidents » non appariés aux cas PMSI incidents en 2006 = 443 + 17 + 2 060 = **2 520**

Total cas ALD cancer du rein « incidents » non appariés aux cas PMSI incidents en 2007 = 236 + 433 + 1 943 = **2 612**

Total cas ALD cancer du rein « incidents » non appariés aux cas PMSI incidents en 2008 = 117 + 210 + 2 389 = **2 716**

Total cas PMSI cancer du rein « incidents » non appariés aux cas ALD incidents en 2006 = 236 + 117 + 5 914 = **6 267**

Total cas PMSI cancer du rein « incidents » non appariés aux cas ALD incidents en 2007 = 443 + 210 + 5 873 = **6 526**

Total cas PMSI cancer du rein « incidents » non appariés aux cas ALD incidents en 2008 = 17 + 433 + 6 244 = **6 694**

Source : PMSI national et Sniiram.

I Tableau 14 I

Cancer du poumon, effectifs appariés par année et cas absents 2006-2008

		Cas ALD « incidents » de cancer du poumon			Cas PMSI absents des ALD « incidentes » 2006-2008
		2006 N=19 895	2007 N=21 021	2008 N= 21 971	
Cas PMSI « incidents » de cancer du poumon	2006 N=29 245	12 519	268	116	16 342
	2007 N=28 726	1 147	12 889	283	14 407
	2008 N=29 388	32	1 112	13 338	14 906
	Cas ALD absents du PMSI « incidents » 2006-2008	6 197	6 752	8 234	

Total cas ALD cancer du poumon « incidents » non appariés aux cas PMSI incidents en 2006 = 1147 + 32 + 6 197 = **7 376**

Total cas ALD cancer du poumon « incidents » non appariés aux cas PMSI incidents en 2007 = 268 + 1 112 + 6 752 = **8 132**

Total cas ALD cancer du poumon « incidents » non appariés aux cas PMSI incidents en 2008 = 116 + 283 + 8 234 = **8 633**

Total cas PMSI cancer du poumon « incidents » non appariés aux cas ALD incidents en 2006 = 268 + 116 + 16 342 = **16 726**

Total cas PMSI cancer du poumon « incidents » non appariés aux cas ALD incidents en 2007 = 1 147 + 283 + 14 407 = **15 837**

Total cas PMSI cancer du poumon « incidents » non appariés aux cas ALD incidents en 2008 = 32 + 1 112 + 14 906 = **16 050**

Source : PMSI national et Sniiram.

I Tableau 15 I

Cancer du système nerveux central (SNC), effectifs appariés par année et cas absents 2006-2008

		Cas ALD « incidents » de cancer du SNC			Cas PMSI absents des ALD « incidentes » 2006-2008
		2006 N=3 083	2007 N=3 253	2008 N= 3 450	
Cas PMSI « incidents » de cancer du SNC	2006 N=5 977	1 794	23	8	4 152
	2007 N=6 229	152	1 944	28	4 105
	2008 N=6 491	14	132	2 013	4 332
	Cas ALD absents du PMSI « incidents » 2006- 2008	1 123	1 154	1 401	

Total cas ALD cancer du SNC « incidents » non appariés aux cas PMSI incidents en 2006 = 152 + 14 + 1 123 = **1 289**

Total cas ALD cancer du SNC « incidents » non appariés aux cas PMSI incidents en 2007 = 23 + 132 + 1 154 = **1 309**

Total cas ALD cancer du SNC « incidents » non appariés aux cas PMSI incidents en 2008 = 8 + 28 + 1 401 = **1 437**

Total cas PMSI cancer du SNC « incidents » non appariés aux cas ALD incidents en 2006 = 23 + 8 + 4 152 = **4 183**

Total cas PMSI cancer du SNC « incidents » non appariés aux cas ALD incidents en 2007 = 152 + 28 + 4 105 = **4 285**

Total cas PMSI cancer du SNC « incidents » non appariés aux cas ALD incidents en 2008 = 14 + 132 + 4 332 = **4 478**

Source : PMSI national et Sniiram.

I Tableau 16 I

Cancer du col de l'utérus, effectifs appariés par année et cas absents 2006-2008

		Cas ALD « incidents » de cancer du col de l'utérus			Cas PMSI absents des ALD « incidentes » 2006-2008
		2006 N=2 730	2007 N=2 904	2008 N= 3 004	
Cas PMSI « incidents » de cancer du col de l'utérus	2006 N=3 336	1 332	19	6	1 979
	2007 N=3 378	155	1 382	28	1 813
	2008 N=3 332	1	145	1 414	1 772
	Cas ALD absents du PMSI « incidents » 2006-2008	1 242	1 358	1 556	

Total cas ALD cancer du col de l'utérus « incidents » non appariés aux cas PMSI incidents en 2006 = 155 + 1 + 1 242 = **1 398**

Total cas ALD cancer du col de l'utérus « incidents » non appariés aux cas PMSI incidents en 2007 = 19 + 145 + 1 358 = **1 522**

Total cas ALD cancer du col de l'utérus « incidents » non appariés aux cas PMSI incidents en 2008 = 6 + 28 + 1 556 = **1 590**

Total cas PMSI cancer du col de l'utérus « incidents » non appariés aux cas ALD incidents en 2006 = 19 + 6 + 1 979 = **2 004**

Total cas PMSI cancer du col de l'utérus « incidents » non appariés aux cas ALD incidents en 2007 = 155 + 28 + 1 813 = **1 996**

Total cas PMSI cancer du col de l'utérus « incidents » non appariés aux cas ALD incidents en 2008 = 1 + 145 + 1 772 = **1 918**

Source : PMSI national et Sniiram.

I Tableau 17 I

Cancer du corps de l'utérus, effectifs appariés par année et cas absents 2006-2008

		Cas ALD « incidents » de cancer du corps de l'utérus			Cas PMSI absents des ALD « incidentes » 2006-2008
		2006 N=4 573	2007 N=4 903	2008 N= 4 965	
Cas PMSI « incidents » de cancer du corps de l'utérus	2006 N=6 606	2 968	122	36	3 480
	2007 N=6 791	235	3 249	99	3 208
	2008 N=6 867	4	246	3 295	3 322
	Cas ALD absents du PMSI « incidents » 2006-2008	1 366	1 286	1 535	

Total cas ALD cancer du corps de l'utérus « incidents » non appariés aux cas PMSI incidents en 2006 = 235 + 4 + 1 366 = **1 605**

Total cas ALD cancer du corps de l'utérus « incidents » non appariés aux cas PMSI incidents en 2007 = 122 + 246 + 1 286 = **1 654**

Total cas ALD cancer du corps de l'utérus « incidents » non appariés aux cas PMSI incidents en 2008 = 36 + 99 + 1 535 = **1 670**

Total cas PMSI cancer du corps de l'utérus « incidents » non appariés aux cas ALD incidents en 2006 = 122 + 36 + 3 480 = **3 638**

Total cas PMSI cancer du corps de l'utérus « incidents » non appariés aux cas ALD incidents en 2007 = 235 + 99 + 3 280 = **3 542**

Total cas PMSI cancer du corps de l'utérus « incidents » non appariés aux cas ALD incidents en 2008 = 4 + 246 + 3 322 = **3 572**

Source : PMSI national et Sniiram.

I Tableau 18 I

Cancer de la thyroïde, effectifs appariés par année et cas absents 2006-2008

		Cas ALD « incidents » de cancer de la thyroïde			Cas PMSI absents des ALD « incidentes » 2006-2008
		2006 N= 5 276	2007 N= 5 558	2008 N= 5 777	
Cas PMSI « incidents » de cancer de la thyroïde	2006 N= 6 894	3 252	171	76	3 395
	2007 N= 7 070	376	3 563	215	2 916
	2008 N= 7 554	12	357	3 724	3 461
	Cas ALD absents du PMSI « incidents » 2006-2008	1 636	1 467	1 762	

Total cas ALD cancer de la thyroïde « incidents » non appariés aux cas PMSI incidents en 2006 = 376 + 12 + 1 636 = **2 024**

Total cas ALD cancer de la thyroïde « incidents » non appariés aux cas PMSI incidents en 2007 = 171 + 357 + 1 467 = **1 995**

Total cas ALD cancer de la thyroïde « incidents » non appariés aux cas PMSI incidents en 2008 = 76 + 215 + 1 762 = **2 053**

Total cas PMSI cancer de la thyroïde « incidents » non appariés aux cas ALD incidents en 2006 = 171 + 76 + 3 395 = **3 642**

Total cas PMSI cancer de la thyroïde « incidents » non appariés aux cas ALD incidents en 2007 = 376 + 215 + 2 916 = **3 507**

Total cas PMSI cancer de la thyroïde « incidents » non appariés aux cas ALD incidents en 2008 = 12 + 357 + 3 461 = **3 830**

Source : PMSI national et Sniiram.

4.3.3 Description des cas appariés : délai d'apparition, cohérence sur le sexe, l'âge, le département et la région de résidence

➤ Délai d'apparition des cas appariés dans les bases croisées

Parmi les cas de cancer appariés sur 2006-2008 pour une même localisation, le délai médian d'apparition des bénéficiaires d'ALD dans le PMSI était d'un mois quelle que soit la localisation. 89 % (cancer du corps de l'utérus) à 97 % (cancer du sein) des cas appariés avaient une date de début d'ALD antérieure à la date du séjour hospitalier. Pour plus de 87 % des bénéficiaires, les soins hospitaliers débutaient dans un délai de cinq mois après la mise en ALD, quelle que soit la localisation (tableau 18).

I Tableau 19 I

Délai entre le début d'ALD et la date d'apparition dans le PMSI des cas appariés sur 2006 à 2008

	Localisation cancéreuse invasive 2006-2008							
	Sein N= 99 942 (%)	Côlon rectum N= 63 016 (%)	Rein N= 14 643 (%)	Poumon N= 41 704 (%)	SNC N= 6 108 (%)	Col de l'utérus N= 4 482 (%)	Corps de l'utérus N= 10 254 (%)	Thyroïde N= 11 746 (%)
Délai médian entre le début d'ALD et la date d'apparition dans le PMSI	1 mois	1 mois	1 mois	1 mois	1 mois	1 mois	1 mois	1 mois
Patients avec date de début d'ALD antérieure à la date de séjour PMSI	97 404 (97)	57 232 (91)	13 253 (91)	38 972 (93)	5 864 (96)	4 271 (95)	9 141 (89)	10 532 (90)
0 à 1 mois	51 272 (51)	43 967 (70)	7 499 (51)	23 405 (56)	4 130 (68)	2 261 (50)	5 838 (57)	6 851 (58)
2 à 5 mois	41 797 (42)	12 325 (20)	5 377 (37)	14 094 (34)	1 552 (25)	1 759 (39)	3 129 (30)	3 213 (27)
6 mois à 1 an	3 985 (4)	818 (1)	327 (2)	1340 (3)	147 (2)	240 (5)	160 (1)	430 (4)
1 à 3 ans	350 (< 1)	122 (< 1)	50 (< 1)	133 (< 1)	35 (< 1)	11 (< 1)	14 (< 1)	38 (< 1)
Patients avec date de séjour PMSI antérieure à la date de début d'ALD	2 538 (3)	5 784 (9)	1 390 (9)	2 732 (7)	244 (4)	211 (5)	1 113 (11)	1 214 (10)

Source : PMSI national et Sniiram.

➤ Cohérence des données appariées sur le sexe, l'âge, le département et la région de résidence

La comparaison individuelle des cas PMSI appariés aux bénéficiaires d'ALD montrait une cohérence parfaite sur le sexe. On notait une cohérence de 99 % pour la région de résidence et de 94 % à 97 % pour le département de résidence selon la localisation (tableau 19).

La moyenne d'âge des cas par localisation était comparable entre les deux sources (donnée non montrée). Cependant près d'un cas d'ALD sur deux avait un an d'âge de plus que les patients de la base PMSI. Ceci est dû au fait que l'âge dans la base PMSI était renseigné en âge révolu et celui des bénéficiaires d'ALD en âge atteint. Lorsque cet écart était pris en compte dans l'analyse, on observait une cohérence totale sur l'âge entre les deux sources (tableau 19).

I Tableau 20 I

Cohérence des données sur le sexe, l'âge, le département et la région de résidence entre les cas PMSI et ALD appariés en 2006, 2007 et 2008

Pourcentage de cas appariés ayant la même information dans le PMSI et les ALD en 2006, 2007 et 2008				
Localisation cancéreuse	Âge ^a	Sexe	Région de résidence	Département de résidence
Sein	100	100	99	96
Côlon, rectum	100	100	99	97
Rein	100	100	99	97
Poumon	100	100	99	96
SNC	100	99	98	94
Col de l'utérus	100	100	99	94
Corps de l'utérus	100	100	99	97
Thyroïde	100	100	99	95

^a Âge en année révolue dans PMSI et atteinte dans le Sniiram : prise en compte du décalage d'un an dans le calcul de la concordance sur l'âge.

Source : PMSI national

4.3.4 Etude des cas « incidents » non appariés

L'analyse des cas « incidents » non appariés la même année montrait plusieurs situations.

- Parmi les bénéficiaires d'ALD pour cancer invasif d'une année N non appariés aux données de cancer invasif du PMSI de l'année N, il existait des bénéficiaires pour lesquels les identifiants apparaissent dans le PMSI de la même année, mais sur des lésions cancéreuses *in situ* (bénin ou kyste cérébral pour le SNC). D'autres bénéficiaires apparaissent avec des séjours pour cancer invasif dans le PMSI des années antérieures (cas « prévalents ») ou des années postérieures. La proportion de ces bénéficiaires variait selon la localisation cancéreuse et l'année (tableau 20). Ainsi en 2006, la proportion des bénéficiaires d'ALD non appariés retrouvés parmi les cas PMSI non invasifs de l'année 2006 et des années antérieures variait entre 18 % (cancer du poumon) et 38 % (cancer du sein). Elle variait respectivement de 5 % (cancer du poumon) à 18 % (cancer du sein) et de 7 % (cancer du poumon) à 21 % (cancer du côlon et rectum) en 2007 et 2008 (tableau 20). Des bénéficiaires supplémentaires apparaissent avec des séjours pour cancer invasif dans le PMSI des années postérieures, entre 10 % (SNC, col de l'utérus) et 21 % (sein).
- Parmi les cas PMSI de cancer invasif d'une année N non appariés aux ALD pour ce cancer invasif de l'année N, certains séjours apparaissent sur des ALD de lésions cancéreuses codées *in situ* dans la base Sniiram l'année N. Leurs effectifs par localisation et par année ne dépassaient pas 2 % de tous les cas non appariés. D'autres cas apparaissent avec les identifiants des bénéficiaires du Sniiram des années antérieures ou des bénéficiaires du Sniiram de l'année N avec date de début différente de l'ALD (tableau 21). Leurs pourcentages parmi les cas hospitaliers non appariés variaient également selon l'année et la localisation. Le plus faible pourcentage quelle que soit l'année concernait le cancer du système nerveux central (autour de 10 %) et le plus fort, le cancer du sein (autour de 50 %). D'autres cas apparaissent avec les identifiants des bénéficiaires du Sniiram des années postérieures.

En résumé, sur la base des données de l'année 2007 qui présentaient l'avantage d'être chaînables avec les données de l'année antérieure (contrairement à 2006) et de l'année postérieure (contrairement à 2008) :

- parmi les cas « incidents » de cancer ALD invasifs non appariés aux cas PMSI, 20 % (poumon) à 40 % (sein) seraient des cas prévalents ;
- parmi les cas « incidents » PMSI de cancers invasifs non appariés aux cas ALD, 11 % (SNC) à 53 % (sein) seraient des cas prévalents.

I Tableau 21 I

Description des cas ALD « incidents » invasifs non appariés l'année N, par localisation cancéreuse et par année

Cas ALD « incidents » invasifs de l'année N non appariés au PMSI « incidents » invasif de l'année N (N1^a), retrouvés parmi les cas suivants :					
	Localisation cancéreuse (ALD « incidents » invasifs non appariés)	Cas avec séjour PMSI l'année N pour un cancer in situ de la localisation (tumeur bénigne ou kyste cérébral pour le SNC)	Cas prévalents PMSI		Séjour PMSI invasif (en DP) les années postérieures
			Cas prévalents l'année N^b	Cas prévalents les années antérieures^c	
2006	Sein : N1 = 17 002 (%)	2 185 (12,8)	92 (4,0)	Non disponible (pas de chaînage possible avec le PMSI 2004 et 2005)	3 622 (21,3)
	Côlon rectum : N1 = 6 069 (%)	132 (2,2)	262 (4,3)		958 (15,8)
	Rein : N1 = 2 520 (%)	8 (0,3)	56 (2,2)		460 (18,2)
	Poumon : N1 = 7 376 (%)	40 (0,5)	133 (1,8)		1 179 (16,0)
	SNC : N1 = 1 289 (%)	179 (13,9)	15 (1,2)		166 (12,9)
	Col de l'utérus : N1 = 1 398 (%)	164 (11,7)	8 (0,6)		156 (11,2)
	Corps de l'utérus : N1 = 1 605 (%)	34 (2,1)	32 (2,1)		239 (14,9)
	Thyroïde : N1 = 2 024 (%)	22 (1,1)	67 (3,3)		388 (19,2)
2007	Sein : N1 = 16 404 (%)	2 532 (15,4)	89 (0,5)	431 (2,6)	3 473 (21,2)
	Côlon rectum : N1 = 6 053 (%)	141 (2,3)	233 (3,8)	644 (10,6)	904 (14,9)
	Rein : N1 = 2 612 (%)	8 (0,3)	61 (2,3)	258 (9,9)	433 (16,6)
	Poumon : N1 = 8 132 (%)	34 (0,4)	166 (2,0)	297 (3,7)	1 112 (13,7)
	SNC : N1 = 1 309 (%)	183 (14,0)	16 (1,2)	28 (2,1)	132 (10,1)
	Col de l'utérus : N1 = 1 522 (%)	181 (11,9)	7 (0,5)	22 (1,5)	145 (9,5)
	Corps de l'utérus : N1 = 1 654 (%)	43 (2,6)	54 (3,3)	90 (5,4)	246 (14,9)
	Thyroïde : N1 = 1 995 (%)	52 (2,6)	63 (3,2)	164 (8,2)	357 (17,9)
2008	Sein : N1 = 17 877 (%)	2 434 (13,6)	87 (0,5)	579 (3,2)	PMSI 2009 non disponible
	Côlon rectum : N1 = 6 606 (%)	194 (2,9)	265 (4,0)	930 (14,1)	
	Rein : N1 = 2 716 (%)	4 (0,1)	58 (2,1)	354 (13,0)	
	Poumon : N1 = 8 633 (%)	46 (0,5)	148 (1,7)	443 (5,1)	
	SNC : N1 = 1 437 (%)	237 (16,5)	17 (1,2)	37 (2,6)	
	Col de l'utérus : N1 = 1 590 (%)	167 (10,5)	13 (0,8)	36 (2,3)	
	Corps de l'utérus : N1 = 1 670 (%)	38 (2,3)	43 (2,6)	119 (7,1)	
	Thyroïde : N1 = 2 053 (%)	40 (1,9)	87 (4,2)	274 (13,3)	

^a Effectif par localisation de cas ALD non appariés au PMSI utilisé comme dénominateur pour le calcul des pourcentages.

^b Séjour avec le cancer invasif de la localisation étudiée en position DP l'année d'étude, et séjour pour ce cancer invasif en position DP, DR, DAS les années N-1 ou N-2.

^c Séjour pour le cancer invasif de la localisation étudiée en DP, DR, DAS les années N-1 ou N-2, et sans séjour l'année d'étude pour ce cancer invasif.

Source : Sniiram.

I Tableau 22 I

Description des cas issu du PMSI « incidents » invasifs non appariés l'année N aux ALD, par localisation cancéreuse et par année

Cas PMSI « incidents » invasifs de l'année N non appariés aux cas d'ALD « incidents » invasif de l'année N (N2 ^a), retrouvés parmi les cas suivants :					
	Localisation cancéreuse (PMSI « incidents » invasifs non appariés)	Cas d'ALD <i>in situ</i> année N : cas en ALD pour un cancer <i>in situ</i> de la localisation, débutée l'année N dans le Sniiram de l'année N	Cas prévalents ALD		Cas d'ALD dont la date de début est postérieure à l'année d'étude (donc issus des bases Sniiram des années postérieures)
			Cas prévalents en ALD dans la base Sniiram de l'année N dont la date de début d'ALD est antérieure à l'année N	Cas prévalents supplémentaires apparaissant dans les bases Sniiram des années antérieures	
2006	Sein : N2 = 18 577 (%)	176 (0,9)	8 489 (45,7)	Non disponible (Sniiram 2005 non disponible)	435 (2,3)
	Côlon rectum : N2 = 23 293 (%)	6 (0,0)	2 732 (11,7)		890 (3,8)
	Rein : N2 = 6 267 (%)	1 (0,0)	766 (12,2)		353 (5,6)
	Poumon : N2 = 16 726 (%)	1 (0,0)	2 313 (13,8)		384 2,3)
	SNC : N2 = 4 183 (%)	0 (0,0)	431 (10,3)		31 (0,7)
	Col de l'utérus : N2 = 2 004 (%)	31 (1,5)	320 (16,0)		25 (1,2)
	Corps de l'utérus : N2 = 3 638 (%)	1 (0,0)	428 (11,8)		158 (4,3)
	Thyroïde : N1 = 3 642 (%)	3 (0,1)	944 (25,9)		247 (6,8)
2007	Sein : N2 = 18 532 (%)	149 (0,8)	9 426 (50,9)	18 (0,0)	261 (1,4)
	Côlon rectum : N2 = 22 953 (%)	6 (0,0)	2 752 (12,0)	17 (0,0)	646 (2,8)
	Rein : N2 = 6 526 (%)	1 (0,0)	856 (13,1)	3 (0,0)	210 (3,2)
	Poumon : N2 = 15 827 (%)	0 (0,0)	2 477 (15,6)	13 (0,0)	283 (1,8)
	SNC : N2 = 4 285 (%)	0 (0,0)	454 (10,6)	6 (0,1)	28 (0,7)
	Col de l'utérus : N2 = 1 996 (%)	36 (1,8)	293 (14,7)	0 (0,0)	28 (1,4)
	Corps de l'utérus : N2 = 3 542 (%)	1 (0,0)	440 (12,4)	1 (0,0)	99 (2,8)
	Thyroïde : N1 = 3 507 (%)	0 (0,0)	892 (25,4)	2 (0,1)	215 (6,1)
2008	Sein : N2 = 20 571 (%)	152 (0,7)	9 186 (44,6)	50 (0,2)	Sniiram 2009 non disponible
	Côlon rectum : N2 = 23 310 (%)	4 (0,0)	2 699 (11,6)	19 (0,0)	
	Rein : N2 = 6 694 (%)	0 (0,0)	885 (13,2)	10 (0,1)	
	Poumon : N2 = 16 050 (%)	0 (0,0)	2 356 (14,7)	18 (0,1)	
	SNC : N2 = 4 478 (%)	0 (0,0)	434 (9,7)	6 (0,1)	
	Col de l'utérus : N2 = 1 918 (%)	26 (1,3)	280 (14,6)	0 (0,0)	
	Corps de l'utérus : N2 = 3 572 (%)	0 (0,0)	467 (13,1)	1 (0,0)	
	Thyroïde : N1 = 3 830 (%)	1 (0,0)	858 (22,4)	8 (0,2)	

^a Effectif par localisation de cas PMSI non appariés aux ALD utilisé comme dénominateur pour le calcul des pourcentages.

Source : PMSI national.

4.4 Indicateurs d'incidence issus des données croisées ALD et PMSI 2006-2008, par localisation

Quatre niveaux de mesure de « l'incidence » des cancers ont été définis selon le degré d'affinage des effectifs issus du croisement PMSI/ALD. Plus l'affinage est approfondi, plus les effectifs « incidents » obtenus sont faibles (tableaux 22 à 29).

1) Les effectifs du premier niveau d'indicateur (union des cas « incidents » invasifs du PMSI et des ALD de l'année N) étaient systématiquement supérieurs aux valeurs issues des projections nationales d'incidence de l'année concernée. L'écart relatif entre les valeurs de cet indicateur et les valeurs des projections variait selon la localisation de 10 % (cancer du poumon) cas à 70 % (cancer du SNC).

2) Les effectifs du deuxième niveau d'indicateur, c'est-à-dire effectif de l'union duquel sont exclus :
 - les ALD pour cancer *in situ* de l'année N retrouvés parmi les cancers invasifs du PMSI de l'année N ;
 - les cas PMSI de l'année N retrouvés parmi les ALD de la base Sniiram de l'année N à date de début d'ALD antérieure à l'année N ;
 - les cas ALD retrouvés parmi les cas PMSI prévalents de l'année N,
 étaient toujours plus élevés - mais moins que les effectifs du premier niveau - que les valeurs de la projection nationale d'incidence. L'écart entre les effectifs de l'indicateur de ce niveau et les valeurs des projections nationales avait diminué presque de moitié par rapport au premier niveau.

3) Les effectifs du troisième niveau d'indicateur étaient moins élevés comparés aux valeurs des projections pour le cancer du poumon.

4) La valeur des indicateurs de quatrième niveau (calculé que pour 2008) était proche de celles des indicateurs du troisième niveau quelle que soit la localisation.

L'écart relatif entre le niveau de l'indicateur et l'estimation nationale d'incidence était faible pour les cancers du poumon et du sein, respectivement -2 % et +5 %, un peu élevé pour les cancers de la thyroïde, du côlon rectum et du rein, +12 %, +19 %, +24 % respectivement et important pour les cancers du corps de l'utérus, du col de l'utérus et du SNC, respectivement +31 %, +47 % et +58 % (non montré).

Parmi les effectifs constituant l'indicateur de quatrième niveau, 28 % (cancer du SNC) à 54 % (cancer du sein) provenaient à la fois des deux sources de données, 11 % (cancer colorectal) à 31 % (cancer du col de l'utérus) ne provenaient que des données d'ALD, et 20 % (cancer du sein) à 56 % (cancer du SNC) ne provenaient que des données PMSI (non montré).

I Tableau 23 I

Cancer invasif du sein chez la femme : indicateurs « d'incidence » issus des données croisées PMSI/ALD et données d'estimation nationale d'incidence

Cancer invasif du sein	2006	2007	2008
1- Estimation/projections nationales d'incidence	50 800	52 500	54 100
2- Indicateurs à partir du croisement PMSI/ALD			
Cas PMSI « incidents » invasifs année N après chaînage sur deux ans	48 448	49 962	51 421
Cas ALD « incidentes » invasifs année N	46 873	47 834	48 727
Indicateur de premier niveau : union PMSI/ALD « incidents » invasifs année N	65 450	66 366	69 298
Indicateur de deuxième niveau : union PMSI/ALD après exclusion des ALD non appariées avec séjour PMSI <i>in situ</i> et des cas prévalents de cancers de l'année N	54 684	54 319	57 591
Indicateurs de troisième niveau : exclusion supplémentaire de prévalents issus du Sniiram et du PMSI des années antérieures	-	53 870	56 962
Indicateurs de quatrième niveau : union PMSI/ALD après chaînage du PMSI sur quatre ans (maximum d'années de recul)	-	-	56 751

Source : PMSI national et Sniiram.

I Tableau 24 I

Cancer invasif du côlon rectum : indicateurs « d'incidence » issus des données croisées PMSI/ALD, et données d'estimation nationale d'incidence

Cancer invasif du côlon/rectum	2006	2007	2008
1- Estimation/projections nationales d'incidence	38 100	38 600	39 100
2- Indicateurs à partir du croisement PMSI/ALD			
Cas PMSI « incidents » invasifs année N après chaînage sur deux ans	42 034	42 895	44 245
Cas ALD « incidentes » invasifs année N	24 810	25 995	27 541
Indicateur de premier niveau : union PMSI/ALD « incidents » invasifs année N	48 103	48 948	50 851
Indicateur de deuxième niveau : union PMSI/ALD après exclusion des ALD non appariées avec séjour PMSI <i>in situ</i> et des cas prévalents de cancers de l'année N	44 977	45 822	47 693
Indicateurs de troisième niveau : exclusion supplémentaire de prévalents issus du Sniiram et du PMSI des années antérieures	-	45 161	46 744
Indicateurs de quatrième niveau : union PMSI/ALD après chaînage du PMSI sur quatre ans (maximum d'années de recul)	-	-	46 480

Source : PMSI national et Sniiram.

I Tableau 25 I

Cancer invasif du rein : indicateurs « d'incidence » issus des données croisées PMSI/ALD, et données d'estimation nationale d'incidence

Cancer invasif du rein	2006	2007	2008
1- Estimation/projections nationales d'incidence	9 500	9 800	10 100
2- Indicateurs à partir du croisement PMSI/ALD			
Cas PMSI « incidents » invasifs année N après chaînage sur deux ans	10 394	11 090	11 190
Cas ALD « incidentes » invasifs année N	6 647	7 176	7 212
Indicateur de premier niveau : union PMSI/ALD « incidents » invasifs année N	12 914	13 702	13 906
Indicateur de deuxième niveau : union PMSI/ALD après exclusion des ALD non appariées avec séjour PMSI <i>in situ</i> et des cas prévalents de cancers de l'année N	12 084	12 777	12 959
Indicateurs de troisième niveau : exclusion supplémentaire de prévalents issus du Sniiram et du PMSI des années antérieures	-	12 516	12 595
Indicateurs de quatrième niveau : union PMSI/ALD après chaînage du PMSI sur quatre ans (maximum d'années de recul)	-	-	12 523

Source : PMSI national et Sniiram.

I Tableau 26 I

Cancer invasif du poumon : indicateurs « d'incidence » issus des données croisées PMSI/ALD, et données d'estimation nationale d'incidence

Cancer invasif du poumon	2006	2007	2008
1- Estimation/projections nationales d'incidence	33 400	34 500	35 700
2- Indicateurs à partir du croisement PMSI/ALD			
Cas PMSI « incidents » invasifs année N après chaînage sur deux ans	29 245	28 726	29 388
Cas ALD « incidentes » invasifs année N	19 895	21 021	21 971
Indicateur de premier niveau : union PMSI/ALD « incidents » invasifs année N	36 621	36 858	38 021
Indicateur de deuxième niveau : union PMSI/ALD après exclusion des ALD non appariées avec séjour PMSI <i>in situ</i> et des cas prévalents de cancers de l'année N	34 135	34 181	35 471
Indicateurs de troisième niveau : exclusion supplémentaire de prévalents issus Du Sniiram et du PMSI des années antérieures	-	33 871	35 010
Indicateurs de quatrième niveau : union PMSI/ALD après chaînage du PMSI sur quatre ans (maximum d'années de recul)	-	-	34 836

Source : PMSI national et Sniiram.

I Tableau 27 I

Cancer invasif du système nerveux central : indicateurs « d'incidence » issu des données croisées PMSI/ALD, et données d'estimation nationale d'incidence

Cancer invasif du système nerveux central	2006	2007	2008
1- Estimation/projections nationales d'incidence	4 400	4 400	4 500
2- Indicateurs à partir du croisement PMSI/ALD			
Cas PMSI « incidents » année N après chaînage sur deux ans	5 977	6 229	6 491
Cas ALD « incidentes » année N	3 083	3 253	3 450
Indicateur de premier niveau : union PMSI/ALD « incidents » année N	7 266	7 538	7 928
Indicateur de deuxième niveau : union PMSI/ALD après exclusion des ALD non appariées avec séjour PMSI bénin/kyste et des cas prévalents de cancers de l'année N	6 641	6 885	7 240
Indicateurs de troisième niveau : exclusion supplémentaire de prévalents issus du Sniiram et du PMSI des années antérieures	-	6 851	7 197
Indicateurs de quatrième niveau : union PMSI/ALD après chaînage du PMSI sur quatre ans (maximum d'années de recul)	-	-	7 126

Source : PMSI national et Sniiram.

I Tableau 28 I

Cancer du col de l'utérus : indicateurs « d'incidence » issus des données croisées PMSI/ALD, et données d'estimation nationale d'incidence

Cancer invasif du col de l'utérus	2006	2007	2008
1- Estimation/projections nationales d'incidence	3 100	3 000	3 000
2- Indicateurs à partir du croisement PMSI/ALD			
Cas PMSI « incidents » invasifs année N après chaînage sur deux ans	3 336	3 378	3 332
Cas ALD « incidentes » invasifs année N	2 730	2 904	3 004
Indicateur de premier niveau : union PMSI/ALD « incidents » invasifs année N	4 734	4 900	4 922
Indicateur de deuxième niveau : union PMSI/ALD après exclusion des ALD non appariées avec séjour PMSI <i>in situ</i> et des cas prévalents de cancers de l'année N	4 242	4 419	4 462
Indicateurs de troisième niveau : exclusion supplémentaire de prévalents issus du Sniiram et du PMSI des années antérieures	-	4 397	4 426
Indicateurs de quatrième niveau : union PMSI/ALD après chaînage du PMSI sur quatre ans (maximum d'années de recul)	-	-	4 417

Source : PMSI national et Sniiram.

I Tableau 29 I

Cancer invasif du corps de l'utérus : indicateurs « d'incidence » issus des données croisées PMSI/ALD, et données d'estimation nationale d'incidence

Cancer invasif du corps de l'utérus	2006	2007	2008
1- Estimation/projections nationales d'incidence	5 900	6 000	6 000
2- Indicateurs à partir du croisement PMSI/ALD			
Cas PMSI « incidents » invasifs année N après chaînage sur deux ans	6 606	6 791	6 867
Cas ALD « incidentes » invasifs année N	4 573	4 903	4 965
Indicateur de premier niveau : union PMSI/ALD « incidents » invasifs année N	8 211	8 445	8 537
Indicateur de deuxième niveau : union PMSI/ALD après exclusion des ALD non appariées avec séjour PMSI <i>in situ</i> et des cas prévalents de cancers de l'année N	7 717	7 908	7 989
Indicateurs de troisième niveau : exclusion supplémentaire de prévalents issus du Sniiram et du PMSI des années antérieures	-	7 817	7 869
Indicateurs de quatrième niveau : union PMSI/ALD après chaînage du PMSI sur quatre ans (maximum d'année de recul)	-	-	7 847

Source : PMSI national et Sniiram.

I Tableau 30 I

Cancer invasif de la thyroïde : indicateurs « d'incidence » issus des données croisées PMSI/ALD, et données d'estimation nationale d'incidence

Cancer invasif de la thyroïde	2006	2007	2008
1- Estimation/projections nationales d'incidence	6 648	7 013	7 391
2- Indicateurs à partir du croisement PMSI/ALD			
Cas PMSI « incidents » invasifs année N après chaînage sur deux ans	6 894	7 070	7 554
Cas ALD « incidentes » invasifs année N	5 276	5 558	5 777
Indicateur de premier niveau : union PMSI/ALD « incidents » invasifs année N	8 918	9 065	9 607
Indicateur de deuxième niveau : union PMSI/ALD après exclusion des ALD non appariées avec séjour PMSI <i>in situ</i> et des cas prévalents de cancers de l'année N	7 885	8 058	8 622
Indicateurs de troisième niveau : exclusion supplémentaire de prévalents issus du Sniiram et du PMSI des années antérieures	-	7 892	8 340
Indicateurs de quatrième niveau : union PMSI/ALD après chaînage du PMSI sur quatre ans (maximum d'année de recul)	-	-	8 303

Source : PMSI national et Sniiram.

5 DISCUSSION

5.1 Rappel des objectifs

Le système de surveillance des cancers basé sur les registres qui couvre 20 % de la population surveille en routine une vingtaine de localisations cancéreuses. La nécessité de disposer d'indicateurs dans un délai plus court que celui des registres et à l'échelle nationale a conduit à l'élaboration du système multi sources cancer.

Les objectifs de cette première étape d'analyses étaient de décrire les données d'ALD et du PMSI ainsi que les effectifs issus de leur croisement d'une part et de définir les principes de construction d'indicateurs « d'incidence » à partir des données croisées d'autre part.

Pour ces premières analyses de données basées sur le croisement PMSI, ALD, huit localisations cancéreuses ont été étudiées sur la base des critères de santé publique, liés à l'environnement et au travail.

Les autres localisations seront étudiées très prochainement.

5.2 Limites et avantages des sources de données utilisées

Les sources de données utilisées pour la construction des indicateurs de surveillance des cancers n'étaient pas conçues dans un objectif de surveillance épidémiologique. Elles répondaient à un objectif de tarification des séjours et de liquidation des soins (comptabilité analytique pour attribution budgétaire dans le cas du PMSI et prise en charge financière de l'assuré dans le cas des ALD). Ainsi, la nomenclature des données est soumise à des évolutions réglementaires. Leur apport à la surveillance épidémiologique des cancers avait été évalué en fonction des localisations cancéreuses et différents travaux basés sur les sources PMSI et ALD seules comparées aux données des registres de cancer avaient été menés dans ce sens [9-12].

Toutefois, les données d'ALD et de PMSI malgré ces limites possèdent des atouts majeurs :

- une couverture nationale avec structuration et codage des données standardisées ;
- un numéro patient anonyme commun des données recueillies à la fois dans le secteur public et dans le secteur privé ;
- une disponibilité relativement rapide ;
- une exhaustivité élevée (près de 100 % depuis 2007) des données du PMSI [13] ;
- l'identifiant anonyme permettait de réaliser un chaînage entre les séjours hospitaliers et de suivre le parcours hospitalier d'un patient surtout le territoire national et durant plusieurs années et ainsi de dénombrer des personnes et non plus des hospitalisations, ce qui représentait une limite majeure à l'utilisation du PMSI pour l'épidémiologie [14].

Ces atouts font de ces sources, malgré leurs caractères médico-économiques des outils potentiels de surveillance des maladies en général et des cancers en particulier.

5.2.1 PMSI

Les données du PMSI présentaient une complexité liée aux difficultés de repérage des séjours incidents pour cancer (première prise en charge du cancer) parmi l'ensemble des séjours car l'enregistrement des données ne permettait pas de différencier les nouveaux cas des anciens cas. Les données par ailleurs comportaient plusieurs catégories de diagnostics (principal, relié et associé) et d'actes ou de thérapeutique permettant d'envisager plusieurs types d'algorithmes possibles de sélection des cas en fonction des objectifs poursuivis. La nomenclature des données du PMSI est soumise à des variations réglementaires dont certaines sont susceptibles d'affecter la comparabilité dans le temps des indicateurs produits. En effet, en 2010 par exemple, la définition des codes diagnostiques a changé. Le « diagnostic principal » d'un séjour relève de la pathologie qui a motivé l'hospitalisation, alors qu'avant il était défini comme la pathologie qui a mobilisé le plus de ressources au cours du séjour hospitalier.

Dans ce travail, la sélection initiale des séjours avait privilégié pour toutes les localisations étudiées le DP dans le but de limiter la sélection des cas de cancer :

- pris en charge pour une récurrence ;
- en cours de suivi médical (code cancer en diagnostic relié) ;
- connus mais hospitalisés pour une autre pathologie (code cancer en diagnostic associé).

Les effectifs de patients pour cancer obtenus avec les algorithmes basés sur les DP, DR ou DAS une année donnée étaient largement supérieurs aux effectifs issus des algorithmes basés sur la sélection exclusive des patients en DP. Les travaux du groupe d'observation épidémiologique du cancer de la région Rhône-Alpes [11] ont montré, pour les localisations cancéreuses du sein, du côlon rectum et du rein, qu'une sélection des cas à partir du DP uniquement permettait de diminuer la variabilité géographique du rapport indicateur PMSI/incidence par rapport à une sélection incluant les actes chirurgicaux. Une sélection des cas sur le DP paraît ainsi plus adéquate pour effectuer des comparaisons d'incidence entre zones géographiques.

Le chaînage des cas de l'année N sélectionnés avec les données du PMSI des années antérieures a été réalisé sur DP, DR ou DAS afin de maximiser le repérage des cas ayant déjà fait l'objet d'un séjour hospitalier en lien avec le cancer étudié. Les cas repérés seraient vraisemblablement des anciens cas de cancer et étaient exclus de l'analyse des données de l'année d'étude.

5.2.2 ALD issues du Sniiram

Le repérage des cas « incidents » parmi les données d'ALD issues du Sniiram était moins complexe que dans le PMSI. La date de début de l'ALD (ou date d'effet, c'est-à-dire date de début de couverture des soins) étant enregistrée dans la base, son association au diagnostic (un seul type de diagnostic) de cancer qui a entraîné la mise en ALD a permis d'identifier le caractère incident du bénéficiaire pour le cancer concerné. Cependant, le traitement des données d'ALD cancer issues du Sniiram avait présenté quelques difficultés.

En effet, les données du Sniiram qui étaient utilisées pour la première fois dans un but épidémiologique avaient nécessité un temps important d'appropriation du contenu de la base. Les données générées par des flux automatiques (depuis les caisses locales vers la CnamTS) présentaient des incohérences ou anomalies susceptibles d'impacter l'interprétation des résultats. Ces anomalies étaient les suivantes :

1) une confusion sur la définition du terme « bénéficiaire d'ALD » dans la base Sniiram est apparue et concernait deux types de définitions :

- est considéré comme bénéficiaire toute personne mise en ALD une année donnée,
- ou est considéré comme bénéficiaire toute mise en ALD ayant consommé un minimum de soins une année donnée ;

2) certaines nouvelles inscriptions en ALD ne correspondraient pas à de nouveaux cas de cancer mais à des cas prévalents liés à une rechute de la maladie enregistrée comme une nouvelle exonération et non comme un renouvellement d'ALD ;

3) le changement de régime ou de caisse primaire d'assurance maladie d'un assuré déjà en ALD peut également apparaître comme une nouvelle inscription en ALD dans la nouvelle caisse ;

4) pour un même bénéficiaire et une même pathologie (code CIM10), plusieurs dates de début d'ALD ont été identifiées. Les bénéficiaires avec plusieurs dates de début d'ALD représentaient 3 % de l'ensemble des bénéficiaires ;

5) la caisse d'affiliation était manquante pour 12 % des bénéficiaires de la base. Ceci empêchait la localisation géographique du bénéficiaire ;

6) certaines dates de fin d'ALD étaient antérieures à la date de début d'ALD ;

7) il existait des cas de doublons (moins de 0,5 %) qui concernaient des bénéficiaires ayant débuté leur ALD une année donnée mais qui étaient déjà présents dans le Sniiram les années antérieures ;

8) certains bénéficiaires présentaient dans la base deux codes de départements (0,1 %) de résidence ou étaient identifiés comme affiliés à deux régimes différents pendant la même année.

Pour une meilleure compréhension et appropriation des données de la base Sniiram et une meilleure interprétation des résultats de l'analyse des données, des demandes de clarifications ont été adressées à la CnamTS.

Les données de la MSA étaient absentes de la base Sniiram et n'étaient pas prises en compte dans cette analyse. Elles représentaient 10 % des bénéficiaires de l'assurance maladie atteints de cancer. Leur absence a probablement impacté le niveau des indicateurs issus des données croisées. Elles seront intégrées dans l'analyse les phases suivantes.

5.2.3 Croisement des bases

Le choix de croiser les données du PMSI et des ALD permettait :

- 1) d'avoir un système exhaustif d'identification des cas de cancer pris en charge à l'hôpital et par l'assurance maladie : le croisement des sources permet d'avoir un système plus sensible que

l'utilisation d'une seule source. Le PMSI et les données d'ALD semblent *a priori* complémentaires puisque le PMSI concerne les prises en charge hospitalières, et les données de mise en ALD concernent aussi les soins de médecine de ville et du secteur hospitalier privé. Par exemple, en utilisant uniquement le PMSI, certains types de cancers non traités dans le secteur hospitalier échapperaient au système, mais ceux-ci devraient être repérables dans les ALD ;

2) de repérer et d'éliminer un maximum de cas prévalents : les sources de données PMSI et ALD ne permettent pas d'attester du caractère incident des cas. En effet, aucune des sources ne permet d'attester qu'un cas donné a été enregistré au moment du diagnostic du cancer. Le PMSI ne différencie pas, lors de l'enregistrement, les cas prévalents des cas incidents. Par ailleurs, la date d'effet de l'ALD peut être différée par rapport au moment du diagnostic du cancer, en particulier si le premier bilan a été pris en charge à 100 % par un autre organisme (mutuelle...). Le croisement des deux sources a permis d'éliminer un maximum de cas prévalents en chaînant chaque source avec l'autre sur les données historiques.

5.3 Discussion des résultats

5.3.1 Croisement des données et cohérence des cas appariés

Les résultats du croisement montraient que le pourcentage de cas appariés, une année donnée, entre le PMSI et les ALD était situé entre 30 % et 40 % selon la localisation. Moins d'une personne sur deux atteintes de cancer était donc à la fois mise en ALD une année donnée et repérée par l'algorithme de sélection des cas parmi les séjours hospitaliers au cours de la même année.

Une partie des cas non appariés était des cas prévalents pris en charge les années précédentes pour leur cancer. Ainsi environ 3 % à 15 % des bénéficiaires d'ALD non appariés au PMSI l'année N apparaissaient dans les bases PMSI des années précédentes, alors que 10 % à 50 % des patients du PMSI non appariés aux ALD de l'année N étaient retrouvés parmi les bénéficiaires d'ALD qui avaient une date de début d'ALD antérieure à l'année N.

Une autre partie des cas non appariés concernait ceux retrouvés parmi les données d'ALD ou hospitalières des années suivantes. En effet, 10 % à 20 % des patients en ALD non appariés au PMSI l'année N apparaissaient dans le PMSI de l'année suivante témoignant du décalage temporel entre la date de début de l'ALD et la date de fin d'hospitalisation.

Le croisement des deux sources de données nécessitait de connaître le niveau de cohérence des données appariées. Ainsi le pourcentage de cohérence sur le sexe, le département et la région de résidence entre les deux sources était très élevé et variait de 94 % à 100 % selon l'année et la localisation cancéreuse. Les discordances, certes marginales, observées pourraient s'expliquer soit par la prise en charge hospitalière d'un bénéficiaire différent de l'ayant-droit, soit par l'enregistrement d'une adresse de résidence temporaire du patient.

L'âge des données d'ALD était calculé en « âge atteint » dans l'année tandis que celui du PMSI était renseigné en « âge révolu » surestimant d'un an les données ALD. La prise en compte de cette surestimation montrait une cohérence parfaite de l'âge entre ces deux sources.

Le délai médian d'apparition des bénéficiaires d'ALD cancer dans le PMSI était de un mois.

5.3.2 Construction des indicateurs

5.3.2.1 Exclusion des cas prévalents et des cas non appariés

La construction des indicateurs reposait sur la somme des cas « incidents » appariés et non appariés. Ainsi lors de la construction, tous les cas prévalents identifiés ont été exclus de même que les cas d'ALD non appariés ayant un séjour PMSI avec un code *in situ* (ou bénin/kyste pour SNC) en position DP. Les cas avec code diagnostic *in situ* étaient négligeables pour toutes les localisations étudiées, sauf pour les cancers du sein (13 % à 15 % selon l'année) et du col de l'utérus (10 % à 12 % selon l'année). Les cas avec code diagnostic kyste cérébral ou tumeur bénigne du cerveau variaient selon l'année de 14 % à 16 %. Certains cas de cancer *in situ* du sein et du col de l'utérus ont pu en effet bénéficier d'une exonération du ticket modérateur pour cancer invasif, pour des motifs liés probablement aux politiques de prévention tandis que les tumeurs bénignes du cerveau ont pu nécessiter le même type de prise en charge qu'une tumeur invasive du cerveau et sont alors fréquemment codées en tumeur invasive dans les données ALD (mais pas dans les registres).

Les cas non appariés restants, impliqués dans la construction des indicateurs peuvent être assimilés à des cas réellement apparus dans l'une des deux sources l'année ciblée. Les hypothèses suivantes peuvent être évoquées sur leur défaut d'appariement :

- le patient n'a pas fait de demande de mise en ALD pour son cancer car il bénéficierait de la CMU-complémentaire, d'une mutuelle couvrant tous les frais... ;
- le décès précoce du patient après le diagnostic de cancer annulant toute demande de mise en ALD ;
- le patient mis en ALD n'a pas eu de prise en charge hospitalière pour son cancer ;
- le décalage entre le codage du séjour et le diagnostic histologique de la lésion. Les résultats histologiques de la lésion traitée peuvent arriver après le codage du séjour hospitalier, codage qui pourrait ne pas faire mention de la tumeur maligne empêchant son repérage parmi les données d'ALD cancer ;
- la présence de faux-positifs : patient codé cancer mais n'ayant pas en réalité développé de cancer à la fin des investigations ;
- les hospitalisations de fin d'année : certains patients diagnostiqués pour cancer sont pris en charge sur une période qui se situe à cheval entre deux années. Le PMSI ne prenant en compte que les périodes de sortie de patients, il y a un risque que les cas dont la date de découverte se situe à l'année N soient comptabilisés dans les données de l'année suivante (N+1) et donc difficilement repérables parmi les données d'ALD cancer de l'année n s'ils étaient mis en ALD ;
- les attributions d'ALD en fin d'année : 30 % à 50 % des bénéficiaires d'ALD cancer selon la localisation débutaient les soins hospitaliers dans un délai supérieur à un mois. Lorsque la période d'attribution de l'ALD et la période des soins se situent à cheval entre deux années, il y a un risque élevé de ne pas retrouver parmi les données hospitalières les données ALD cancer de l'année N ;
- 5 % des bénéficiaires d'ALD cancer avec une date de début l'année N apparaissaient dans la base Sniiram de l'année N+1 et n'étaient pas comptabilisés parmi les indicateurs de l'année N. La construction d'indicateurs « réactifs » disponibles l'année N+1 (réception des données de l'année N), et comparables d'une année à l'autre (l'année N+1 n'était pas disponible pour 2008) a milité en faveur de ce choix. Ces patients pourront toutefois être retenus pour la construction d'indicateurs moins réactifs, par exemple sur une période se terminant en N-2.

5.3.2.2 Niveau des indicateurs construits

Le niveau des indicateurs construits était très proche du niveau de l'incidence nationale estimée pour les cancers du poumon et du sein, peu éloigné pour les cancers de la thyroïde, du côlon rectum, et du rein et très éloigné pour les cancers du SNC, du corps de l'utérus et du col de l'utérus.

Le fait que les niveaux d'indicateurs soient bien au-dessus des estimations nationales d'incidence, pour les cancers du SNC, du col ou du corps de l'utérus, du rein et du côlon rectum pourrait laisser penser que le nombre de faux-positifs parmi les cas non appariés reste élevé pour ces localisations.

Par ailleurs, ces niveaux ont été obtenus sans les données de la MSA qui représentaient 10 % des bénéficiaires d'ALD pour cancer. La future intégration des données de la MSA dans l'analyse impactera vraisemblablement en fonction de la localisation, le niveau des indicateurs produits.

5.4 Travaux « princeps » de surveillance des cancers à partir des bases médico-administratives

L'utilisation des données hospitalières ou de l'Assurance maladie comme moyen complémentaire de surveillance des cancers a été développée en France depuis un moment. Jusqu'à présent, les travaux reposaient sur les bases PMSI seules [9,11] ou ALD cancer seules [12].

La méthodologie utilisée était basée sur l'étude préalable du rapport entre le nombre de patients identifiés dans la base médico-administrative (BMA) et le nombre de cas incidents (rapport « indicateur BMA / incidence »), dans les départements avec un registre de cancers. Ce rapport est ensuite extrapolé aux départements hors zones registres, avec prise en compte de l'hétérogénéité observée dans les départements avec registres, pour estimer l'incidence dans chaque département de la zone non couverte par un registre. Toutefois, si le rapport « indicateur BMA / incidence » diffère trop entre les départements au sein de la zone registre, il ne permet pas d'estimer l'incidence de façon fiable dans les autres départements. Cette thématique d'estimation départementale de l'incidence de cancers est encore un domaine de recherche, impliquant différents acteurs. L'objectif à terme est de produire pour chaque localisation cancéreuse, et lorsqu'elles sont valides, les estimations infranationales à partir des BMA et/ou de la mortalité.

5.5 Perspectives

La deuxième phase des analyses consistera à évaluer les indicateurs en les comparant aux données de cancers observées dans les départements avec registres, puis, en fonction de ces résultats, à produire des indicateurs de surveillance de l'incidence des cancers à l'échelle nationale et infranationale. Cette deuxième étape de travail fera l'objet d'un autre rapport.

Des travaux croisant les données de registre avec les données du PMSI seront aussi nécessaires pour consolider les algorithmes de sélection de cas « incidents » dans le PMSI, et méritent d'être développés.

6 CONCLUSION

Le croisement des données individuelles du PMSI et des ALD pour la surveillance des cancers en France constitue le premier axe de la mise en place du dispositif de surveillance des cancers SMSC, complémentaire du dispositif basé sur les registres. L'utilisation des données d'anatomie et de cytologie pathologiques à travers le dossier communiquant de cancérologie en cours de construction doit compléter, à moyen terme, le dispositif et permettre de surveiller les aspects histologiques des cancers.

Ce premier croisement ne concernait pas les données ALD de la MSA qui n'étaient pas disponibles au moment de l'analyse.

Les résultats de l'étude montraient qu'il est possible de croiser anonymement les bases PMSI et ALD à condition de disposer d'un identifiant unique et commun à ces deux bases et sous réserve qu'aucune évolution réglementaire ne vienne affecter significativement ces bases. Les indicateurs calculés dans ce rapport concernaient le niveau national pour huit localisations de cancer choisies sur la base des critères de santé publique (fréquence, létalité, accessibilité à des actions de prévention primaire et secondaire, évaluation des besoins de soins à l'échelle régionale/départementale) et du prolongement de l'étude pilote SMSC.

La comparaison des effectifs issus des quatre types d'indicateurs calculés avec les estimations nationales issues des données de registres présentait trois situations selon les localisations cancéreuses : 1) résultats proches de l'estimation nationale pour les cancers du sein et du poumon, 2) résultats peu éloignés de l'estimation nationale pour les cancers de la thyroïde, du côlon rectum et du rein, 3) résultats très éloignés de l'estimation nationale pour les cancers du SNC, du corps de l'utérus et du col de l'utérus.

Le pourcentage de données appariées entre les ALD et le PMSI était constant dans le temps pour la même localisation mais variait selon la localisation de 25 % (cancer du SNC) à 47 % (cancer du sein).

La disponibilité récente des données ALD de la MSA permet leur intégration aux données du Sniiram en vue de compléter les données d'ALD. Une nouvelle analyse portant sur ces données complétées permettra d'actualiser et de consolider les résultats déjà obtenus avec les huit localisations cancéreuses et d'étendre aux autres localisations cancéreuses le calcul des indicateurs au niveau national.

A l'issue de ces analyses au cours desquelles les indicateurs calculés seront confrontés aux données d'incidence dans la zone registre, un deuxième rapport sera élaboré.

La déclinaison régionale et départementale de production des indicateurs issus du croisement PMSI, ALD sera abordée dans un deuxième temps. Elle nécessite au préalable une première étape de validation des données à partir des données de registres.

Par ailleurs, les travaux relatifs au dossier communiquant de cancérologie piloté par l'Inca et l'Asip-santé continuent d'être suivis par l'InVS dans le cadre de l'utilisation des données de ce dossier par le SMSC.

Références bibliographiques

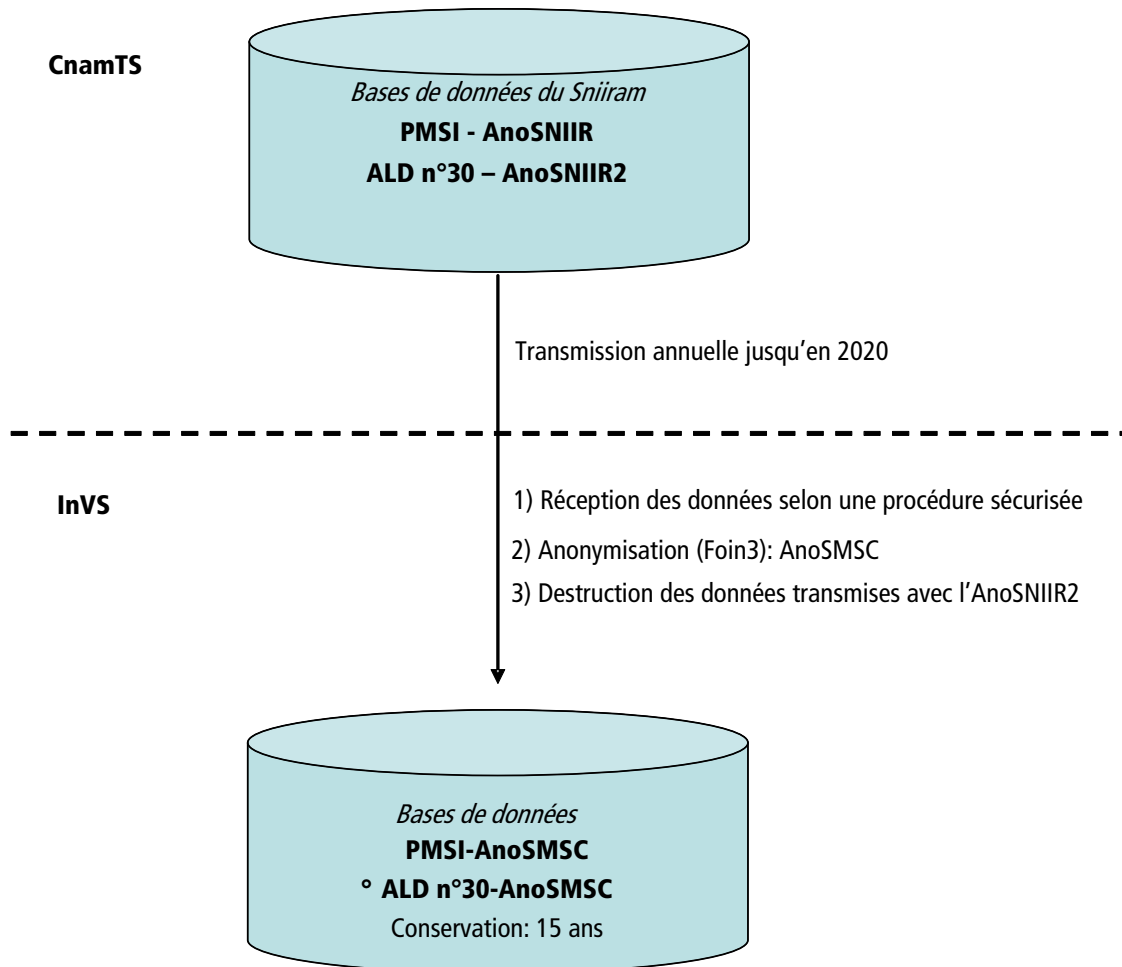
- [1] Surveillance épidémiologique des cancers en France. Institut de veille sanitaire [mis à jour le 13/07/2010; consulté le 17/06/2011]. Disponible à partir de l'URL : <http://www.invs.sante.fr>
- [2] Belot A, Grosclaude P, Bossard N, Jouglu E, Benhamou E, Delafosse P, *et al.* Estimation de l'incidence et de la mortalité par cancer en France de 1980 à 2005. InVS [mis à jour le 13/01/2011; consulté le 22/02/2011]. Disponible à partir de l'URL : <http://www.invs.sante.fr>
- [3] Colonna M, Bossard N, Mitton N, Remontet L, Belot A, Delafosse P, *et al.* Some interpretation of regional estimates of the incidence of cancer in France over the period 1980-2005. *Rev Epidemiol Sante Publique* 2008;56(6):434-40.
- [4] Projections de l'incidence et de la mortalité par cancer en France en 2011. Hospices civils de Lyon / Institut de veille sanitaire / Institut national du cancer / Francim / Institut national de la santé et de la recherche médicale [mis à jour le 13/01/2011; consulté le 17/06/2011]. Disponible à partir de l'URL : <http://www.invs.sante.fr>
- [5] Remontet L, Belot A, Bossard N, *et al.* Tendances de l'incidence et de la mortalité par cancer en France et projections pour l'année en cours : méthodes d'estimation et rythme de production. *BEH* 2009;38406-408.
- [6] Le plan cancer 2009-2013. Institut national du cancer, France [mis à jour le 25/07/2011; consulté le 19/08/2010]. Disponible à partir de l'URL : <http://www.e-cancer.fr/plancancer-2009-2013>
- [7] Caserio-Schönemann C, Kudjawu Y, Chérié-Challine L, Guillet A, Musset A, Nicolau J, *et al.* Système multi sources de surveillance des cancers (SMSC): Etude pilote sur le cancer de la thyroïde en Ile-de-France et dans le Nord-Pas-de-Calais. Institut de veille sanitaire, France [mis à jour le 31/08/2011; consulté le 07/09/2011]. Disponible à partir de l'URL : <http://www.invs.sante.fr>
- [8] Dossier communicant de cancérologie. Institut national du cancer, France [mis à jour le 11/01/2011; consulté le 03/11/2011]. Disponible à partir de l'URL : <http://www.e-cancer.fr/soins/parcours-de-soins/dossier-communicant-de-cancerologie>
- [9] Kudjawu Y, Uhry Z, Danzon A, Bloch J. Cancer du testicule : évolution nationale et variations régionales du taux de patients opérés, 1998-2008. Institut de veille sanitaire, France [mis à jour le 01/06/2011; consulté le 24/10/2011]. Disponible à partir de l'URL : www.invs.sante.fr
- [10] Lagadic C, Trétarre B, Fournier E, Molinié F, Grosclaude P. Concordance entre les données du PMSI et les données des registres de cancers. (Poster), Grell, Tolède. 2010.
- [11] Mitton N, Cólonna M, Trombert B, Olive F, Gomez F, Iwaz J, *et al.* A Suitable Approach to Estimate Cancer Incidence in Area without Cancer Registry. *J Cancer Epidemiol* 2011;418968-Epub 08/03/2011.
- [12] Uhry Z, Remontet L, Cólonna M, Belot A, Grosclaude P, Mitton N, *et al.* Estimation départementale de l'incidence des cancers à partir des données d'affection de longue durée (ALD) et des registres : évaluation pour 24 localisations cancéreuses, 2000-2005. 2011.
- [13] Les grandes sources de données sur les établissements de santé. Dress, France [mis à jour le 2011; consulté le 24/10/2011]. Disponible à partir de l'URL : http://www.sante-jeunesse-sports.gouv.fr/IMG/pdf/etabsante2007-c8_sources-2.pdf
- [14] Olive F, Gomez F, Schott AM, Remontet L, Bossard N, Mitton N, *et al.* [Critical analysis of French DRG based information system (PMSI) databases for the epidemiology of cancer: A longitudinal approach becomes possible.]. *Rev Epidemiol Sante Publique* 2011;59(1):53-58.

ANNEXE

Annexe 1 - Mise à disposition des données et architecture d'anonymisation des données ALD et PMSI, tel que envisagé de façon pérenne dans la demande Cnil n° 1492397.

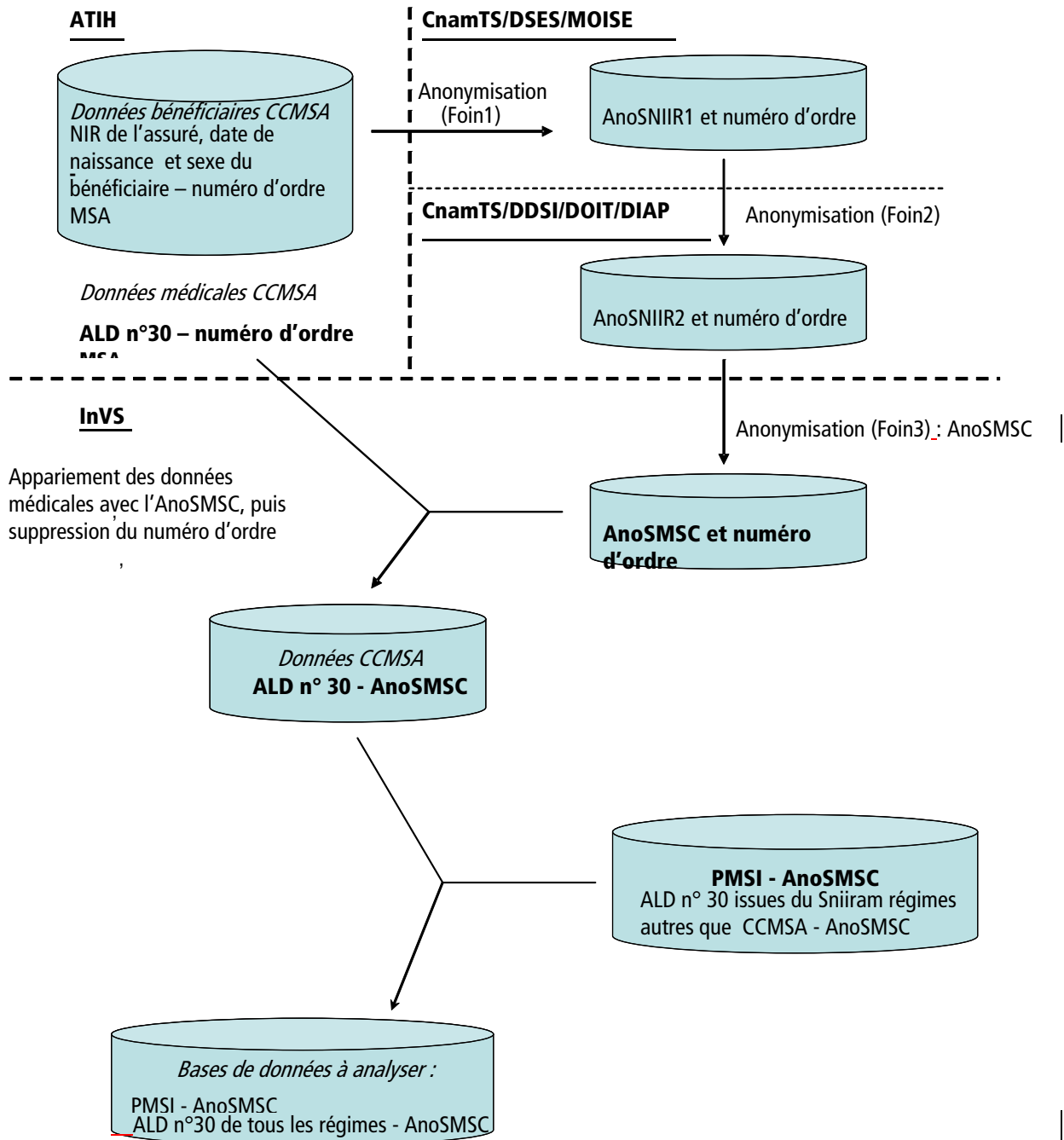
I Figure 1 I

Mise à disposition des données et architecture d'anonymisation des données PMSI et ALD cancer (ALD n° 30) pour générer le numéro anonyme unique commun



I Figure 2 I

Mise à disposition et architecture d'anonymisation des données ALD cancer (ALD n°30) de la Caisse centrale de la mutualité sociale agricole (CCMSA) pour générer le numéro anonyme unique commun (solution temporaire en attendant l'intégration de ces données dans le Sniiram)



ANNEXE 2 : Description supplémentaire des données ALD cancer issues du Sniiram 2006-2008 et comparaison aux données ALD agrégées.

Tables « bénéficiaires »

Les tables « bénéficiaires » 2006, 2007 et 2008 issues du Sniiram et transmises par la CnamTS contiennent les caractéristiques administratives des patients ayant eu une exonération du ticket modérateur pour la prise en charge d'une pathologie cancéreuse en 2006, 2007 et 2008.

Le tableau 1 décrit les effectifs des bases « bénéficiaires » 2006, 2007, 2008.

La table « bénéficiaires » 2006 totalise 1 756 642 codes identifiants « Sniiram » différents (codes BEN_BEN_ID) et 1 753 305 identifiants issus du pseudoNIR et des procédures d'anonymisation Foin, correspondant à l'AnoSMSC (BEN_NIR_PSA).

3 337 identifiants issus des procédures d'anonymisation (0,2 %) sont donc associés à deux codes bénéficiaires différents (ces bénéficiaires présentent deux codes régimes différents la même année ou deux codes départements différents la même année). Il n'y a aucun code BEN_BEN_ID avec plusieurs numéros BEN_NIR_PSA.

Les bases 2007 et 2008 présentent les mêmes caractéristiques : 0,2 % des identifiants anonymes sont associées à deux codes bénéficiaires différents soit respectivement 3 392 et 3 576 identifiants.

I Tableau 1 I

Description des effectifs des bases « bénéficiaires » 2006, 2007, 2008

	Sniiram 2006	Sniiram 2007	Sniiram 2008
Nombre de numéro BEN_BEN_ID différents	1 756 642	1 854 665	1 946 063
Nombre de numéros AnoSMSC différents (BEN_NIR_PSA)	1 753 305	1 851 273	1 942 487

Source : Sniiram.

Effectifs des tables « médicales » 2006, 2007 et 2008

Les tables « médicales » 2006, 2007 et 2008 issues du Sniiram et transmises par la CnamTS comprennent, pour chaque bénéficiaire de France (métropole et DOM-TOM), les codes CIM cancer associés à la mise en ALD pour pathologie cancéreuse, avec, pour chaque code CIM, une date de début d'ALD.

Les effectifs de cas d'ALD cumulés par code CIM étaient 1 840 871, 1 944 723 et 2 041 822 respectivement dans les bases médicales Sniiram 2006, 2007 et 2008 (tableau 2). Un patient mis en ALD pour n localisations de cancer différentes est donc comptabilisé n fois, comme c'est le cas pour les données agrégées. 5 % des bénéficiaires étaient en ALD pour plusieurs codes CIM cancer la même année. (Montré dans le rapport).

3 % des bénéficiaires avaient plus d'une date de début d'ALD pour un même code CIM renseignées dans la même année du Sniiram (tableau 2). Moins de 0,2 % des bénéficiaires (12 388) avaient des dates de début différentes renseignées dans des années du Sniiram différentes (non montré).

I Tableau 2 I

Effectif d'ALD cumulées, nombre de date de début d'ALD et année de début d'ALD par base Sniiram

	Année de la base Sniiram		
	2006 (%)	2007 (%)	2008 (%)
Effectif d'ALD cumulées par code CIM	1 840 871	1 944 723	2 041 822
Nombre de date de début d'ALD par bénéficiaire et par code CIM			
Une date	1 790 540 (97 %)	1 890 155 (97 %)	1 985 942 (97 %)
Supérieure à une date	50 331 (3 %)	54 568 (3 %)	55 880 (3 %)
Année de début d'ALD (minimum des dates de début d'ALD renseignée dans la table Sniiram, par code CIM).			
Avant 2006	1 599 579	1 456 856	1 339 105
2006	241 275	236 020	202 440
2007	3	251 834	245 830
2008	0	0	254 435
2009-2018	14	13	12

Source : Sniiram.

Le tableau 3 présente les effectifs d'ALD débutées en 2006, 2007 et 2008 par année de première apparition dans la base Sniiram 2006-2008. Ainsi, sur les 257 299 ALD débutées en 2006, 93,8 % apparaissent dans le Sniiram 2006, 5,8 % apparaissent pour la première fois dans le Sniiram 2007 et 0,4 % apparaissent pour la première fois dans le Sniiram 2008.

Parmi les bénéficiaires ayant débuté leur ALD en 2006 mais qui apparaissent pour la première fois dans le Sniiram 2007, 85,4 % ont comme mois de début d'ALD le mois de décembre 2006. De la même façon 5,6 % des ALD débutées en 2007 apparaissent pour la première fois dans le Sniiram 2008.

I Tableau 3 I

Année de première apparition dans le Sniiram 2006-2008 pour les ALD débutées en 2006, 2007 et 2008

Année de première apparition dans la base Sniiram 2006-2008	Année de début de l'ALD		
	Début de l'ALD en 2006 N=257 299 (%)	Début de l'ALD en 2007 N=266 453 (%)	Début de l'ALD en 2008 N=253 792 (%)
2006	241 275 (93,8)	-	-
2007	14 920 (5,8)	251 515 (94,4)	-
2008	1 104 (0,4)	14 938 (5,6)	253 792 (100)

Source : Sniiram.

Comparaison des ALD agrégées aux ALD issues du Sniiram

Le tableau 4 compare les ALD agrégées aux ALD issues du Sniiram. Les ALD agrégées correspondent aux données du RG et du RSI, pour les bénéficiaires affiliés à une caisse en France métropolitaine. Les effectifs d'ALD du Sniiram que nous avons retenu pour effectuer la comparaison sont les ALD des bénéficiaires résidant en France métropolitaine et affiliés aux RG, RSI mais aussi aux régimes pour lesquels les codes sont « inconnus ». Les ALD comparées concernent exclusivement les mises en ALD pour un cancer invasif (code CIM10 commençant par C).

La différence relative entre les ALD agrégées et les ALD du Sniiram est égale à 3,6 % pour 2006, 5,3 % pour 2007 et 10,2 % pour 2008.

I Tableau 4 I

Comparaison des ALD agrégées et des ALD issues du Sniiram en France métropolitaine pour toutes les localisations invasives

	Données agrégées d'ALD issues des régimes : général (RG) et indépendant (RSI) de l'assurance maladie	Sniiram 2006-2008 : Année de début d'ALD des bénéficiaires issus du RG, RSI et inconnus de l'assurance maladie	Différence relative en % entre ALD agrégées^a et ALD du Sniiram
	Mises en ALD en 2006	Début de l'ALD en 2006	2006
Homme	134 164	129 394	3,6
Femme	117 184	112 859	3,7
Total	251 348	242 253	3,6
	Mises en ALD en 2007	Début de l'ALD en 2007	2007
Homme	142 995	134 423	6,0
Femme	122 050	116 589	4,5
Total	265 045	251 012	5,3
	Mises en ALD en 2008	Début de l'ALD en 2008	2008
Homme	139 834	125 264	10,4
Femme	126 654	114 152	9,9
Total	266 488	239 416	10,2

^a Différence relative = (ALD agrégées - ALD Sniiram)*100 / ALD agrégées

Source : Sniiram et ALD.

Système multi sources de surveillance des cancers Croisement des données ALD et PMSI

Description des données 2006-2008 et construction d'indicateurs

Le croisement des données d'affection longue durée (ALD) et du Programme médical des systèmes d'information (PMSI) constitue le premier axe de réorientation du développement du système multi source de surveillance des cancers recommandé par les conclusions du pilote. L'objectif du croisement était de décrire à l'échelle nationale, les données ALD cancer issues de la base du système national d'information interrégime de l'Assurance maladie de 2006-2008, les séjours PMSI 2004-2008 pour cancer ainsi que les données issues du croisement de ces deux sources sur la période 2006-2008. Il s'agissait aussi d'élaborer une méthode de construction d'indicateurs de surveillance à partir des données croisées.

Huit localisations cancéreuses (sein, poumon, côlon-rectum, col et corps de l'utérus, thyroïde, système nerveux central (SNC) et rein) ont été étudiées.

Les séjours pour cancer de l'année d'étude ont été sélectionnés grâce aux codes CIM10 en diagnostic principal, chaînés avec les séjours identiques du PMSI des années antérieures.

Les bénéficiaires d'ALD pour cancer étaient sélectionnés grâce au code CIM10 et la date de début d'ALD la plus ancienne.

Les cas sélectionnés étaient croisés grâce à un identifiant anonyme commun. Les indicateurs mesurant leur niveau étaient confrontés aux estimations nationales d'incidence.

L'écart relatif entre indicateur issu du croisement et incidence estimée en 2008 était :

- très faible pour le poumon (-1 %) et le sein (+6 %) ;
- important pour la thyroïde (+17 %), le côlon-rectum (+22 %), le rein (+28 %) et le corps de l'utérus (+33 %) ;
- très important pour le corps de l'utérus (+33 %), le col de l'utérus (+49 %) et le SNC (+61 %).

L'écart relatif important observé pour certains cancers s'expliquerait par un nombre élevé de "faux-positifs" parmi les cas non appariés. Le PMSI et les ALD n'étaient pas conçus pour un objectif épidémiologique mais médico-économique. Toutefois, leur couverture nationale, l'exhaustivité des données et le numéro patient commun constituent des atouts majeurs pour la surveillance des cancers.

Mots clés : ALD, PMSI, cancer, surveillance, croisement

Multi source surveillance system of cancer (MSSC), cross-referencing of LTI and PMSI data

Description of 2006-2008 data and construction of indicators

The cross-referencing of data from Long Term Illness (LTI) and National Hospital Discharge Database (PMSI) represents the first line of reorientation for the development of a national multi source surveillance system of cancers (MSSC) recommended by the conclusions of the pilot study.

The objective of the cross-referencing was to describe LTI for cancer data from the National Health Insurance Cross-Schemes Information System for the year 2006-2008 at national level, hospital stays data for cancer for the period 2004-2008 and crossed data of both sources over the 2006-2008 period. It was also a question of elaborating a method for surveillance indicators construction based on cross-referenced data.

Eight cancer sites (breast, lung, colon/rectum, uterus cervix, uterus corpus, thyroid, central nervous system (CNS), and kidney) were studied. Hospital stays for cancer during the study year were selected by ICD-10 codes coded as main diagnosis, and chained with the same hospital stays from the PMSI data of previous years.

LTI's beneficiaries for cancer selection were based on ICD-10 codes and the starting date of the eldest LTI.

PMSI and LTI cases selected were cross-referenced by a common anonymous identifier. Indicators measuring the level of crossed data were compared with national incidence estimated.

The relative difference between indicators constructed and estimated incidence in 2008 was:

- very weak for lung cancer (-1 %) and breast cancer (+6 %);
- important for thyroid cancer (+17 %), colon/rectum cancer (+22 %) and kidney cancer (+28 %);
- very important for uterus corpus cancer (+33 %), uterus cervix cancer (+49 %) and CNS cancer (+61 %).

The important difference observed for certain cancers may be due to a high number of "false-positives" among unmatched cases. PMSI and LTI data were not designed for epidemiological purposes, but rather for medico-economic objectives. Meanwhile, their national coverage, the data exhaustiveness, and a common patient number, are the major assets for cancer surveillance.

Citation suggérée :

Kudjawan Y, Rudant J, Decool E, Danzon A, Grémy I. Système multi sources de surveillance des cancers, croisement des données ALD et PMSI - Description des données 2006-2008 et construction d'indicateurs. Saint-Maurice: Institut de veille sanitaire; 2012. 70 p. Disponible à partir de l'URL : <http://www.invs.sante.fr>

INSTITUT DE VEILLE SANITAIRE

12 rue du Val d'Osne

94415 Saint-Maurice Cedex France

Tél. : 33 (0)1 41 79 67 00

Fax : 33 (0)1 41 79 67 67

www.invs.sante.fr