

Usine d'incinération d'ordures ménagères de Gilly-sur-Isère (Savoie)

Retour d'expérience sur le recensement rétrospectif des cas de cancers



Ce rapport a été rédigé par :

- Françoise Mégas, médecin enquêteur
- Laurent Voisin, moniteur d'étude
- Marielle Schmitt, Cire Rhône-Alpes
- Alexandra Thabuis, Cire Rhône-Alpes

Résumé

L'étude rétrospective d'incidence des cancers autour de l'usine d'incinération d'ordures ménagères de Gilly-sur-Isère, décidée dans un contexte de crise lors de sa fermeture fin 2001, avait pour objectif de déterminer s'il existait un excès de cancers autour de l'installation. Elle a consisté à recenser de la manière la plus exhaustive possible les cancers survenus entre 1994 et 2002 dans la zone d'étude (exposée aux retombées atmosphériques de l'incinérateur) afin de comparer l'incidence observée à celle établie par les registres des cancers français. Ce retour d'expérience décrit les principales difficultés rencontrées et les solutions apportées lors du recensement des cas de cancers.

Le recueil des données médicales a été réalisé auprès de multiples sources d'information : laboratoires d'anatomo-cyto-pathologie et d'hématologie, départements d'information médicale (DIM) et services spécialisés des hôpitaux et cliniques, caisses primaires d'assurance maladie (CPAM), médecins libéraux, registres des cancers spécialisés. Les 59 fichiers informatiques recueillis ont été traités : recherche des adresses manquantes (en utilisant des données administratives des bureaux des entrées des établissements de soin et de la Caisse nationale d'assurance maladie), sélection des patients de la zone d'étude, homogénéisation des codages des cancers, concaténation des fichiers en une seule base, analyse des informations disponibles sur chaque cancer, dédoublonnage. La majorité des cancers a ensuite été validée par consultation des dossiers médicaux, afin d'éliminer les faux positifs (métastases de cancer primitif connu, récidives...).

Sur 2 360 cancers ayant fait l'objet d'une validation, 34 % ont été exclus ; au final, la base comprenait 2 055 cas de cancers. Trois types de sources ont permis d'identifier à eux seuls 94 % des cas : laboratoires, DIM des hôpitaux et CPAM. Avec une moyenne de 3,2 sources par cas et un taux de 91 % des cas avec confirmation histologique, la base pouvait être considérée comme ayant la même exhaustivité que celle d'un registre.

Le recours à des données administratives et la consultation des dossiers médicaux se sont révélés nécessaires compte tenu des incertitudes sur les adresses des patients au moment du diagnostic, des erreurs de codage des cancers dans certains fichiers recueillis et des difficultés à identifier les faux positifs. En conclusion, une étude rétrospective d'incidence des cancers est réalisable localement à condition que des moyens humains très importants soient mis en œuvre (2,5 équivalents temps plein pendant trois ans pour recenser 2 055 cas). Des recommandations et des conseils pratiques pour la réalisation d'une telle étude sont formulés en fin de rapport.

Abstract

The retrospective cancer incidence study carried out around the municipal solid waste incinerator of Gilly-sur-Isère was ordered in a context of crisis during its closing in the late 2001. Its purpose was to determine whether or not there was an exceeding number of cancers around the incinerator. This study consisted in counting as exhaustively as possible the cancers that occurred between 1994 and 2002 in the study area, which was exposed to the atmospheric fallouts from the incinerator. Thus, it was planned to compare the observed cancer incidence to the French cancer registries'. This work describes the main difficulties encountered as well as the solutions found during the enumeration of cancer cases.

The collection of medical data was carried out thanks to multiple sources of information: pathology and hematology laboratories, hospitals and clinics medical informatics departments, clinical services, health insurance funds, liberal practitioners or specialised cancer registries. The 59 collected medical data files were dealt with: looking for the missing addresses (by using administrative data collected from hospital admission desks and the national health insurance fund for salaried workers), selecting patients from the study area, homogenizing cancers coding, merging files into a single database, analysing available information on each cancer, de-duplicating the database. Most cancers were validated by consulting medical folders so as to exclude the false cases like metastasises of a known primary cancer or recurrences.

2360 cancers were validated and 34% of them were excluded. The final database was made of 2055 cancer cases. Three types of sources only allowed to identify 94% of cases: laboratories, hospital medical informatics departments and health insurance funds. With an average of 3.2 sources per case and 91% of histologically-verified cases, the database could be considered as much exhaustive as a registry's.

Resorting to administrative data and consulting medical folders turned out to be necessary considering uncertainties about the patients' place of residence by the time of the diagnosis, errors in coding cancers in a few databases that were collected and difficulties to identify false cases. As a conclusion, a retrospective cancer incidence study can be carried out at a local scale provided a very important manpower is brought into play (2.5 full-time workers during three years to take a census of 2055 cases). Some recommendations as well as practical advice to carry out such a study are given at the end of the report.

Liste des abréviations

ACP	Anatomo-cyto-pathologie
Adicap	Association pour le développement de l'informatique en cytologie et anatomo-pathologie
BDE	Bureau des entrées
Careps	Centre Rhône-Alpes d'épidémiologie et de prévention sanitaire
CH	Centre hospitalier
CHU	Centre hospitalier universitaire
CIM	Classification internationale des maladies
CIMO	Classification internationale des maladies pour l'oncologie
Circ	Centre international de recherche sur le cancer
Cire	Cellule interrégionale d'épidémiologie
CME	Commission médicale d'établissement
CnamTS	Caisse nationale d'assurance maladie des travailleurs salariés
Cnil	Commission nationale de l'informatique et des libertés
CPAM	Caisse primaire d'assurance maladie
Crisap	Centre de regroupement informatique et statistique en anatomie pathologique
Ddass	Direction départementale des affaires sanitaires et sociales
DIM	Département d'information médicale
ENCR	European Network of Cancer Registries
EQRS	Evaluation quantitative des risques sanitaires
ERSM	Echelon régional du service médical de la CnamTS
ETM	Exonération du ticket modérateur
Francim	Réseau français des registres de cancer
HCL	Hospices civils de Lyon
Insee	Institut national de la statistique et des études économiques
InVS	Institut de veille sanitaire
OMS	Organisation mondiale de la santé
Pires	Protocole inter-régime d'examens spéciaux
PMSI	Programme de médicalisation des systèmes d'information
SNOMED	Systematized Nomenclature of Medicine
UIOM	Usine d'incinération d'ordures ménagères
VBA	Visual Basic Application
VMC	Validation médicale des cas

Sommaire

1	CONTEXTE ET OBJECTIFS	9
2	PRÉSENTATION DE L'ÉTUDE RÉTROSPECTIVE D'INCIDENCE DES CANCERS DE 1994 À 2002 AUTOUR DE L'UIOM DE GILLY-SUR-ISÈRE	11
2.1	PRÉSENTATION DE L'ÉTUDE	11
2.1.1	Objectif	11
2.1.2	Zone d'étude	11
2.1.3	Période d'étude	11
2.1.4	Population d'étude	11
2.1.5	Populations de référence	11
2.1.6	Définition de cas	11
2.1.7	Sources d'information et traitement de l'information	12
2.1.8	Aspects éthiques	13
2.2	ORGANISATION DE L'ÉTUDE	13
2.2.1	Moyens humains	13
2.2.2	Moyens matériels	13
2.2.3	Calendriers	14
3	RECUEIL ET TRAITEMENT DES DONNÉES	15
3.1	RECUEIL DES DONNÉES	15
3.1.1	Organisation du recueil des données	15
3.1.2	Description des sources d'information et des fichiers informatiques	16
3.2	TRAITEMENT ET VALIDATION DES DONNÉES	17
3.2.1	Etape 1 : sélection des patients de la zone d'étude dans les fichiers médicaux	17
3.2.2	Etape 2 : traitement de l'information médicale dans les fichiers médicaux	18
3.2.3	Etape 3 : concaténation des fichiers médicaux	18
3.2.4	Etape 4 : recherche des doublons et homonymes	18
3.2.5	Etape 5 : analyse médicale des données	18
3.2.6	Etape 6 : validation des adresses par consultation des dossiers médicaux	19
3.2.7	Etape 7 : validation médicale des cancers par consultation des dossiers médicaux	19
4	BILAN DU RECENSEMENT DES CAS	23
4.1	NOMBRE DE SOURCES PAR CAS	23
4.2	CONTRIBUTION DES SOURCES AU RECENSEMENT DES CAS	23
4.3	DISPONIBILITÉ DE L'INFORMATION MÉDICALE SELON LA LOCALISATION CANCÉREUSE	24
4.4	ANALYSE DE L'INCIDENCE DES CANCERS	25
5	DISCUSSION ET RECOMMANDATIONS	27
5.1	PERTINENCE D'UNE ÉTUDE D'INCIDENCE	27
5.2	FAISABILITÉ D'UNE ÉTUDE D'INCIDENCE	27
5.2.1	Exhaustivité et fiabilité de la base de données	27
5.2.2	Recommandations pour la réalisation d'une étude d'incidence	28
6	CONCLUSION ET PERSPECTIVES	29
7	RÉFÉRENCES	31
	ANNEXES	33
	ANNEXE 1 : COMMUNES DE LA ZONE D'ÉTUDE	33
	ANNEXE 2 : CARTE DE LA ZONE D'ÉTUDE	34
	ANNEXE 3 : GRILLE DE CORRESPONDANCE DES CODES DE COMPORTEMENTS TUMORAUX EN ADICAP ET CIMO2	35
	ANNEXE 4 : QUESTIONNAIRE SUR LES FILIÈRES DE SOINS ENVOYÉ AUX MÉDECINS DE LA ZONE D'ÉTUDE	36
	ANNEXE 5 : QUESTIONNAIRE DE RECUEIL DES CAS ENVOYÉ AUX MÉDECINS DE LA ZONE D'ÉTUDE	37
	ANNEXE 6 : CODES RETENUS POUR LES REQUÊTES	38
	ANNEXE 7 : LISTE DES FOURNISSEURS DE DONNÉES ET DES REQUÊTES RÉALISÉES	39
	ANNEXE 8 : RÉSUMÉ DES RÈGLES DE COMPTABILISATION DES TUMEURS MULTIPLES UTILISÉES DANS L'ÉTUDE	42
	ANNEXE 9 : CONSEILS PRATIQUES POUR LA RÉALISATION D'UNE ÉTUDE RÉTROSPECTIVE D'INCIDENCE DES CANCERS	43

1 Contexte et objectifs

L'usine d'incinération d'ordures ménagères (UIOM) de Gilly-sur-Isère située près d'Albertville en Savoie a démarré en 1971. En 1992 et 2001, le préfet de Savoie mettait en demeure l'exploitant afin que l'UIOM respecte la réglementation en vigueur (arrêté du 25 janvier 1991). En février 2001, les élus de Grignon (commune voisine de Gilly-sur-Isère) interpellaient le sous-préfet d'Albertville à propos du nombre de cancers qui leur semblait anormalement élevé sur cette commune, et la Direction départementale des affaires sanitaires et sociales (Ddass) était alertée par un médecin d'Albertville qui signalait une surincidence des cancers autour de l'UIOM. Les mesures à l'émission alors effectuées par l'exploitant montraient des concentrations en dioxines particulièrement élevées (1 285 ng I-TEQ_{OTAN}/Nm³ en août et 75 ng I-TEQ_{OTAN}/Nm³ en octobre, à comparer à la valeur limite fixée à 0,1 ng I-TEQ_{OTAN}/Nm³ par la directive européenne 2000/76/CE du 4 décembre 2000). Les mesures réalisées dans l'environnement proche (lait, végétaux, sols) révélèrent par ailleurs des concentrations en dioxines supérieures aux normes. Le 25 octobre 2001, le préfet décidait de fermer l'usine. Par la suite étaient organisés l'abattage massif du bétail, le retrait des produits non conformes et du foin contaminé, la surveillance de la production laitière, fromagère et carnée dans une zone de 30 km de rayon autour de l'UIOM. Ces événements spectaculaires et très médiatisés ont renforcé les inquiétudes de la population quant aux risques sanitaires liés à la présence de cette usine.

Dans ce contexte, en novembre 2001, la cellule de crise dédiée aux dioxines créée par le préfet sollicitait la Cellule interrégionale d'épidémiologie (Cire) Rhône-Alpes et l'Institut de veille sanitaire (InVS) afin de répondre aux questions suivantes :

- Existait-il un excès de cancers dans la population exposée aux retombées atmosphériques de l'UIOM ?
- La concentration en dioxines dans le lait des mères allaitantes résidant à proximité de l'UIOM était-elle supérieure à la concentration habituellement retrouvée en France ?
- L'UIOM étant fermée, quels étaient pour la population riveraine, les risques résiduels et les conséquences sanitaires liées aux expositions passées ?

Pour répondre à ces questions, quatre études locales ont été décidées en 2001 et 2002 :

- une étude rétrospective d'incidence des cancers entre 1994 et 2002 autour de l'UIOM confiée à la Cire Rhône-Alpes [1] ;
- une étude rétrospective de mortalité par cancer entre 1968 et 1999 autour d'Albertville confiée au Registre des cancers de l'Isère [2] ;
- une évaluation quantitative des risques sanitaires (EQRS) liés aux émissions atmosphériques de l'UIOM, réalisée par le Centre Rhône-Alpes d'épidémiologie et de prévention sanitaire (Careps) [3] ;
- une étude prospective d'imprégnation en dioxines du lait maternel auprès de mères allaitant leur enfant et résidant autour de l'UIOM, également réalisée par le Careps.

L'étude rétrospective d'incidence des cancers a été menée par la Cire durant trois ans et demi, l'essentiel du temps ayant été consacré au recensement des cas. Ce retour d'expérience décrit les principales difficultés rencontrées lors de ce recensement et les solutions apportées. Il a pour objectif de :

- fournir les éléments permettant d'apprécier les difficultés d'un recensement rétrospectif des cancers dans une population générale en l'absence de registre ;
- proposer des recommandations à l'attention des autorités sanitaires pouvant être amenées à envisager ce type d'étude ;
- donner des conseils pratiques pour la réalisation d'une telle étude ;
- contribuer à la réflexion sur la faisabilité du projet de l'InVS d'un système prospectif de surveillance épidémiologique multi-sources des cancers en France.

2 Présentation de l'étude rétrospective d'incidence des cancers autour de l'UIOM de Gilly-sur-Isère

2.1 Présentation de l'étude

2.1.1 Objectif

L'objectif était de déterminer s'il existait un excès de cas de cancers autour de l'UIOM. Pour cela, il s'agissait :

- d'établir l'incidence des cancers dans la population exposée aux émissions atmosphériques de l'UIOM en recensant de la manière la plus exhaustive possible tous les cas de cancers survenus entre 1994 et 2002 ;
- de la comparer aux incidences des cancers dans des populations de référence établies à partir des données des registres des cancers.

2.1.2 Zone d'étude

La zone d'étude sur laquelle ont été recensés les cas de cancers a été déterminée à partir de la modélisation de la dispersion atmosphérique des émissions de dioxines de l'UIOM de Gilly-sur-Isère réalisée par ARIA Technologies pour les besoins de l'EQRS. Les 30 communes retenues (annexes 1 et 2) correspondaient à celles sur lesquelles les concentrations en dioxines modélisées dans l'air ambiant dépassaient, en un point au moins, le bruit de fond habituellement retrouvé en milieu rural, soit 50 fg/m³ [4].

2.1.3 Période d'étude

La période d'étude s'étendait sur 9 ans, de 1994 à 2002. Elle tenait compte de la période de fonctionnement de l'UIOM, du temps de latence de survenue des cancers chez l'adulte après exposition, et de la disponibilité des données médicales.

2.1.4 Population d'étude

La population d'étude était la population résidant sur la zone d'étude. Les données de recensement de l'Institut national de la statistique et des études économiques (Insee) indiquaient que 46 140 personnes habitaient la zone d'étude en 1990 et 47 853 en 1999.

2.1.5 Populations de référence

Trois populations de référence ont été retenues pour les comparaisons d'incidence :

- la population de l'Isère où il existe un registre général des cancers ;
- la population de 7 départements (environ 9 % de la population française) disposant d'un registre général des cancers dont les données de 1994 à 2001 étaient disponibles : Bas-Rhin, Haut-Rhin, Doubs, Hérault, Isère, Manche et Tarn ;
- la population française totale pour laquelle il existe une estimation de l'incidence des cancers.

2.1.6 Définition de cas

Un cas a été défini comme toute tumeur confirmée par l'histologie, la cytologie ou l'hématologie, ou par l'accumulation de signes évocateurs (cliniques, radiologiques, biologiques...), découverte entre le 1^{er} janvier 1994 et le 31 décembre 2002 chez une personne domiciliée dans la zone d'étude au moment du diagnostic, et répondant aux critères d'inclusion ci-dessous.

2.1.6.1 Critères d'inclusion des tumeurs

Afin de permettre la comparaison avec les trois populations de référence retenues, les règles d'inclusion des cancers pour les calculs d'incidence établies par le Centre international de recherche sur le cancer (Circ) et le Réseau français des registres des cancers (Francim) ont été appliquées [5,6] :

- toutes les tumeurs malignes et invasives ont été retenues, à l'exception des carcinomes cutanés baso-cellulaires et spino-cellulaires ;
- pour le système nerveux central, les tumeurs bénignes, de malignité incertaine et in situ ont également été considérées ;
- les récurrences de cancers déjà diagnostiqués et les métastases dont le cancer primitif était connu n'ont pas été retenues.

2.1.6.2 Choix de la date d'incidence

La règle de l'European Network of Cancer Registries (ENCR) a été suivie pour établir la date d'incidence [7]. D'une manière générale, la date d'incidence retenue correspondait à la date de première confirmation histologique ou cytologique de malignité sauf si une prise en charge (admission dans un établissement de soins ou consultation pour la pathologie maligne) avait eu lieu plus de 3 mois avant la confirmation, la date de cette prise en charge étant alors retenue comme date d'incidence.

2.1.7 Sources d'information et traitement de l'information

Les sources d'information sollicitées ont été :

- les laboratoires de diagnostic, c'est-à-dire les laboratoires publics et privés d'anatomo-cyto-pathologie (ACP) et d'hématologie susceptibles d'avoir réalisé des analyses pour des patients de la zone d'étude ;
- les établissements hospitaliers et les cliniques susceptibles d'avoir pris en charge des patients de la zone d'étude : départements d'information médicale (DIM), bureaux des entrées (BDE), services cliniques, unités de concertation ;
- les Caisses primaires d'assurance maladie (CPAM) couvrant la zone d'étude : régime général et régimes particuliers (agriculteurs, commerçants, cheminots, mineurs, militaires) ;
- les médecins libéraux, généralistes et spécialistes, de la zone d'étude ;
- les registres des cancers spécialisés couvrant la Savoie (leucémies et lymphomes de l'enfant, tumeurs solides de l'enfant, cancers de la thyroïde).

Les données recueillies auprès de ces sources ont été croisées, analysées et dédoublonnées. Les cas de cancers ont été validés directement à partir des données informatiques disponibles, ou par consultation des dossiers médicaux. Les données médicales fournies par les sources étaient codées selon différentes nomenclatures : Adicap¹, SNOMED², CIM9 et CIM10³, CIMO2 et CIMO3⁴, codifications internes. Tous les cancers de la base ont été transcodés en CIMO2, en application des règles du Circ [8]. Les codes de la CIMO2 sont composés de 4 caractères pour la topographie (localisation de la tumeur primitive), 4 caractères pour la morphologie (type histologique de la tumeur) et 1 caractère pour le comportement (bénin : /0, de malignité incertaine : /1, carcinome in situ : /2 ou malin : /3).

¹ Association pour le développement de l'informatique en cytologie et anatomo-pathologie. Cette nomenclature était utilisée par tous les anatomo-cyto-pathologistes ayant participé à l'étude, à l'exception de ceux du Centre hospitalier universitaire (CHU) de Grenoble et des laboratoires d'hématologie.

² Systematized Nomenclature of Medicine, provenant de la Systematized Nomenclature of Pathology développée par le College of American Pathologists. La version utilisée au CHU de Grenoble provient de la première version de la SNOMED (1979) à laquelle ont été ajoutés au fur et à mesure, des codes topographiques ou morphologiques internes.

³ Classification internationale des maladies, 9^{ème} et 10^{ème} révisions, publiées par l'Organisation mondiale de la santé (OMS).

⁴ Classification internationale des maladies pour l'oncologie, 2^{ème} et 3^{ème} éditions, publiées par l'OMS.

2.1.8 Aspects éthiques

Le comité consultatif sur le traitement de l'information en matière de recherche dans le domaine de la santé a émis un avis favorable et la Commission nationale de l'informatique et des libertés (Cnil) a donné son autorisation pour le traitement des données nominatives des patients de la zone d'étude sur la période 1994-2002.

Dans la mesure où les données sur les patients avaient été recueillies initialement pour un autre objet et que certaines des personnes concernées étaient perdues de vue, la Cnil a accordé une dérogation à l'obligation individuelle d'information prévue à l'article 40-5 de la loi du 6 janvier 1978 modifiée. L'information de la population concernée a été diffusée par le biais du Comité de suivi du plan d'action sanitaire local mis en place par le préfet de Savoie (regroupant, tous les six mois, associations, élus et administrations), des médecins libéraux de la zone d'étude (affichage d'une note d'information dans leur salle d'attente), et par la diffusion d'un encart dans la presse régionale (Dauphiné libéré du 18 septembre 2003).

2.2 Organisation de l'étude

2.2.1 Moyens humains

L'étude a été menée par la Cire Rhône-Alpes. Deux enquêteurs ont été recrutés pour le recensement des cas de cancers :

- un médecin enquêteur chargé de recueillir les données (contacts, relances, autorisations, cahiers des charges des requêtes), de vérifier, homogénéiser et coder les données médicales recueillies, et de valider les cas ;
- un moniteur d'étude chargé d'assister le médecin dans le recueil des données, de traiter les données recueillies (nettoyage et homogénéisation des fichiers sources, croisement des données, structuration des données, recherche des doublons et homonymes), de préparer l'analyse médicale des données et les vérifications dans les dossiers médicaux et de construire la base de données.

Le groupe technique pilotant l'étude comprenait des représentants de la Cire Rhône-Alpes, du Département santé environnement de l'InVS et du Registre des cancers de l'Isère.

2.2.2 Moyens matériels

2.2.2.1 Logiciels

- CanReg 4, logiciel conçu par le Circ pour l'enregistrement prospectif des cas incidents par les registres des cancers⁵.
- Excel et Access pour la gestion des fichiers et la création de la base de données.

2.2.2.2 Nomenclatures

- Nomenclatures CIM9, CIM10 et CIMO2 (codes relatifs aux cancers saisis informatiquement par les enquêteurs).
- Nomenclature Adicap (sous format informatique).
- Nomenclature interne dérivée de la SNOMED utilisée par les laboratoires d'ACP du CHU de Grenoble (sous format informatique).
- Nomenclature interne utilisée par le laboratoire d'hématologie du Centre hospitalier Lyon Sud des Hospices civils de Lyon (HCL) (sous format papier).

⁵ Ce logiciel, retenu au départ pour construire la base de données, n'a finalement pas été utilisé car il était inadapté à l'importation et à la manipulation de nombreux fichiers volumineux. Il a simplement servi à compléter la recherche des doublons et homonymes.

2.2.2.3 Outils de transcodification

- IARCTools, logiciel du Circ permettant les transcodifications CIM9-CIMO2 ou CIM10-CIMO2.
- Outils de transcodification Adicap-CIMO2⁶ :
 - feuille Excel de transcodification du Dr Buemi du Registre des cancers du Haut-Rhin ;
 - base Access relative aux tumeurs et à leurs codifications Adicap et CIMO2 mise au point par le Pr Got (Assistance publique des hôpitaux de Paris) ;
 - fichiers de transcodification Adicap-CIMO2 téléchargés sur le site Internet de la fédération des Centres de regroupement informatique et statistique en anatomie pathologique (Crisap) ;
 - grille de correspondance entre les codes de comportements tumoraux de l'Adicap et de la CIMO2 (annexe 3), établie par la Cire et validée par le Dr Petitjean (Adicap).

2.2.3 Calendriers

La faisabilité de l'étude a été étudiée en 2002 : identification des filières de soins, prises de contacts initiales, rencontres des premiers partenaires et fournisseurs de données. Le premier semestre 2003 a été consacré à l'élaboration du protocole, à l'obtention de l'autorisation de la Cnil et à la préparation du recueil des données. Le calendrier initial de l'étude prévoyait un an pour le recensement des cas : recueil des données, validation des cas, traitement des données et construction de la base de données (figure 1).

Figure 1. Calendrier prévisionnel initial de l'étude

	2003												2004											
	01	02	03	04	05	06	07	08	09	10	11	12	01	02	03	04	05	06	07	08	09	10	11	12
Définition de la zone d'étude																								
Rédaction, validation du protocole																								
Organisation du recueil des données																								
Demande d'autorisation à la Cnil																								
Recrutement des enquêteurs																								
Recueil des données, validation des cas																								
Traitement des données, construction de la base de données																								
Analyse des données, rédaction du rapport																								
Publications																								
Coordination de l'étude (groupe technique, comité de suivi)																								

Suite aux nombreuses difficultés rencontrées, le travail de recensement des cas a pris un an et demi de plus que prévu et s'est donc terminé en janvier 2006 (figure 2).

Figure 2. Calendrier final relatif au recensement des cas

	2003					2004												2005													
	07	08	09	10	11	12	01	02	03	04	05	06	07	08	09	10	11	12	01	02	03	04	05	06	07	08	09	10	11	12	01
Recueil et traitement initial des données																															
Validation des cas par consultation des dossiers																															
Traitement des données, construction de la base																															

⁶ Il n'existe pas à ce jour d'outils officiels de transcodification Adicap-CIMO2 : les outils existants ont été mis au point par des anatomo-cyto-pathologistes français et/ou des registres mais ne sont pas validés.

3 Recueil et traitement des données

3.1 Recueil des données

3.1.1 Organisation du recueil des données

3.1.1.1 Etude de faisabilité

L'étude de faisabilité a permis d'identifier les principales filières de soins (locales, régionales voire nationales) suivies par les patients de la zone d'étude (annexe 4 : questionnaire envoyé aux médecins de la zone d'étude pour décrire les filières de soins selon le type de cancer), de recenser l'ensemble des fournisseurs potentiels de données et d'étudier la disponibilité des données.

3.1.1.2 Contact des fournisseurs de données

Tous les fournisseurs potentiels de données recensés ont été contactés et, hormis les médecins de la zone d'étude, ils ont tous été rencontrés.

Le recueil des données a nécessité l'obtention de nombreuses autorisations, notamment auprès des établissements de santé (directions, commissions médicales d'établissement (CME), services cliniques). Les réticences de certains laboratoires de diagnostic à fournir leurs données ont été levées suite aux rencontres avec le responsable national du Syndicat des médecins anatomo-cytopathologistes français (qui a adressé un courrier à tous les laboratoires syndiqués et aux Crisap concernés) et avec la présidente du Collège des laboratoires d'ACP des HCL. De multiples relances ont dû être faites auprès de l'ensemble des fournisseurs, pour les persuader de participer et accélérer le recueil.

3.1.1.3 Recueil des données

Un questionnaire de recueil des cas a été adressé aux médecins de la zone d'étude (annexe 5). Pour les autres fournisseurs de données médicales, des requêtes informatiques ont été réalisées dans leurs bases de données, si possible par les enquêteurs, sinon par le fournisseur lui-même ou un prestataire de service.

Les requêtes étaient effectuées selon trois critères :

- la pathologie (annexe 6) ;
- la date de diagnostic : date d'examen ACP, de consultation, d'hospitalisation, d'attribution de l'exonération du ticket modérateur (ETM)... ;
- le code postal de domicile du patient.

Les variables recherchées étaient les suivantes :

- variables d'identification du patient : nom, nom de jeune fille, prénom, sexe, date de naissance, libellé de l'adresse, code postal, commune de domicile ;
- variables décrivant la tumeur : topographie, morphologie, comportement, date de diagnostic, médecin et/ou établissement prescripteur pour les laboratoires.

Auprès des BDE, les requêtes étaient effectuées selon deux critères (date d'hospitalisation et commune de domicile). Les variables recherchées étaient les variables d'identification du patient, ainsi que les dates d'entrée et de sortie dans l'établissement et les services fréquentés.

Certains laboratoires hospitaliers d'ACP et d'hématologie ne codant pas les diagnostics de cancer ou n'étant pas informatisés sur la totalité de la période d'étude, il n'était pas possible d'effectuer une requête. Les patients identifiés dans les fichiers des DIM et des BDE des établissements ont été recherchés *a posteriori* dans les bases de données des laboratoires (requête indirecte).

La plupart des laboratoires de diagnostic ne renseignaient que rarement l'adresse du patient (aucune adresse disponible ou adresse du prescripteur ou de l'établissement de soin), les requêtes ont alors été faites sur deux critères au lieu de trois (tumeur et date sans tenir compte du code postal).

3.1.2 Description des sources d'information et des fichiers informatiques

Le tableau 1 présente, pour chaque type de fournisseur de données : le type de données recueillies, le nombre de fournisseurs contactés, le nombre de fournisseurs ayant participé à l'étude et le nombre de fichiers informatiques exploités.

Tableau 1. Fournisseurs de données et données recueillies

Types de sources	Types de fournisseurs de données	Types de données recueillies	Nombre de fournisseurs contactés	Nombre de fournisseurs ayant participé (taux de réponse)	Nombre de fichiers informatiques (nombre total de lignes de données)	
Médicales	Laboratoires d'ACP et d'hématologie	Diagnostics de cancer	28	28 ^a (100 %)	24 ^b (96 000)	
	Médecins libéraux	Patients suivis pour cancer	75	21 (28 %)	1 ^c (400)	
	DIM des CH et cliniques	Données du PMSI ^d avec DP ^e ou DA ^f mentionnant un cancer	22	20 (91 %)	20 (9 600)	
	CPAM	ETM pour cancer	6	6 (100 %)	5 ^b (2 200)	
	Autres sources médicales : - service clinique d'oncologie et service de radiothérapie du CH de Chambéry - Enquête permanente cancer de la Fédération nationale des centres de lutte contre le cancer - unité de concertation, service clinique d'onco-hématologie et Centre de ressources informatiques hospitalières du CHU de Grenoble		Patients vus en hospitalisation, séance ou consultation pour cancer	6	6 (100 %)	6 (2 500)
	Registres des cancers spécialisés	Cas recensés	3	3 (100 %)	3 (50)	
Administratives	BDE des CH et cliniques	Séjours	20	17 (85 %)	17 (54 800)	
Total			160	101	76	

^a Dont 7 requêtes indirectes.

^b 4 laboratoires et 1 CPAM n'ont pas fourni de fichier car il n'y avait aucun patient de la zone d'étude dans leur base de données.

^c Les 20 questionnaires papier ont été saisis dans un seul fichier informatique.

^d Programme de médicalisation des systèmes d'information.

^e Diagnostic principal.

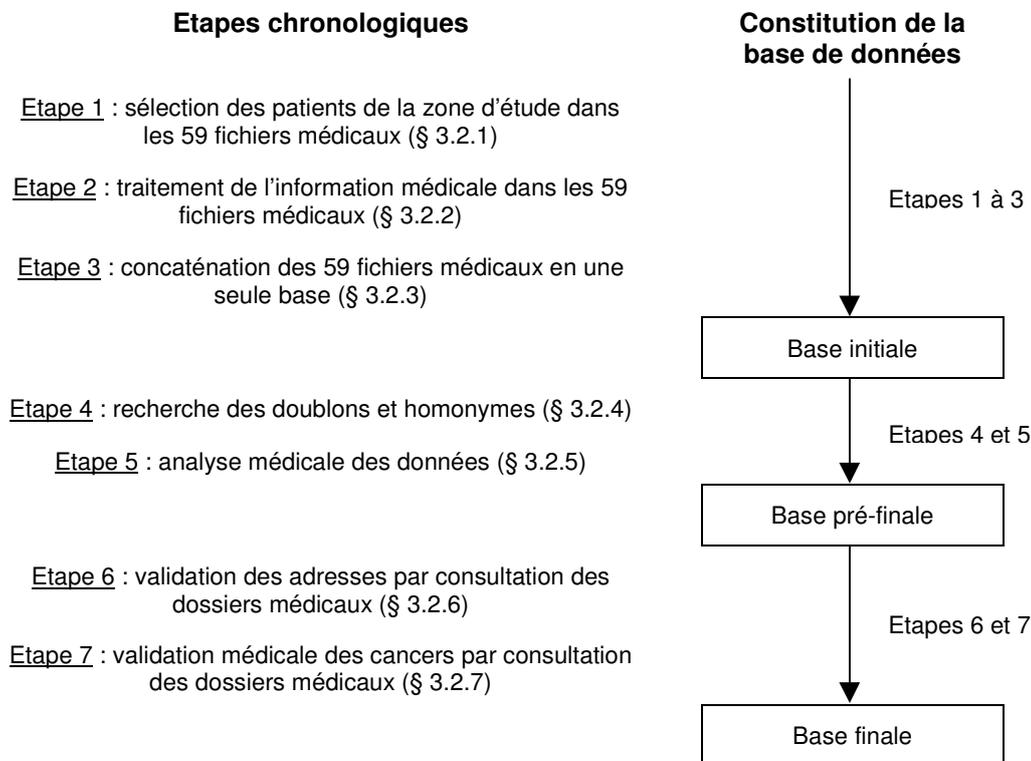
^f Diagnostic associé.

L'annexe 7 dresse la liste complète des fournisseurs de données ayant participé et décrit pour chacun d'eux le type de requête réalisé (requête directe ou indirecte, effecteur de la requête, période couverte).

3.2 Traitement et validation des données

Les étapes de traitement et de validation des données sont synthétisées dans la figure 3 et détaillées dans les paragraphes suivants.

Figure 3. Etapes de traitement et de validation des données



Au fur et à mesure de l'obtention des fichiers informatiques, leur structure a été mise en forme et homogénéisée :

- homogénéisation des formats de fichiers (PC / Macintosh, Excel / texte...) ;
- suppression des espaces inutiles, des signes de ponctuation et des accents, mise en majuscule des données ;
- homogénéisation des formats de date (jj/mm/aaaa) ;
- homogénéisation des écritures des communes (remplacement de «ST» par «SAINT», «/» par «SUR»...) ;
- homogénéisation des noms de variables ;
- ajout de nouvelles variables (ex : appartenance à la zone d'étude) ;
- codage numérique des modalités de certaines variables.

3.2.1 Etape 1 : sélection des patients de la zone d'étude dans les fichiers médicaux

Les requêtes dans les laboratoires de diagnostic ayant souvent été faites sur seulement deux critères (tumeur et date sans tenir compte du code postal), 96 000 lignes de données ont été récupérées dans les laboratoires de diagnostic dont seulement 6 200 (6 %) fournissaient un code postal de la zone d'étude.

Les deux services cliniques ont également fourni des fichiers avec peu d'adresses renseignées (390 lignes avec des codes postaux de la zone d'étude sur 1 400 lignes fournies).

Les fichiers des laboratoires et des services cliniques ont donc été croisés entre eux, avec les autres fichiers médicaux recueillis et avec les 17 fichiers des BDE, ce qui a permis d'identifier les lignes sans adresse correspondant à des patients de la zone d'étude. Après ces croisements, il restait encore la majorité des lignes sans adresse, qui ne correspondaient probablement pas à des patients de la zone d'étude. Pour s'en assurer, il a été demandé à l'Echelon régional du service médical de la CnamTS (ERSM), après accord de la Cnil, de fournir l'identité (nom, nom de jeune fille, prénom, sexe, date de naissance, code postal, commune) des patients résidant en Rhône-Alpes et ayant obtenu une ETM pour cancer entre 1994 et 2002 (180 000 lignes). Tous les patients dont l'adresse a été retrouvée par ce biais résidaient bien hors de la zone d'étude ; ceux qui n'ont pas été retrouvés ont par conséquent également été exclus.

Les autres sources médicales (CPAM, DIM...) ont fourni des fichiers ne contenant que des lignes indiquant un code postal de la zone d'étude. Une petite sélection a cependant dû être effectuée puisque certains codes postaux étaient communs à des communes de la zone d'étude et hors zone d'étude ; lorsque le libellé de l'adresse ou de la commune indiquait que les patients n'appartenaient pas à la zone d'étude, ils ont été exclus.

3.2.2 Etape 2 : traitement de l'information médicale dans les fichiers médicaux

Après la suppression des patients hors zone d'étude, chaque fichier médical a fait l'objet d'un traitement de l'information médicale fournie :

- homogénéisation de l'écriture des codes médicaux en CIM9 et en CIM10 (ex : C43.2 devient C432, C01 devient C019, 147.9 devient 1479, C**.8 devient C**.9...) ;
- détection des erreurs de codification par des contrôles de cohérence (ex : sexe et site tumoral, pour les organes spécifiques d'un sexe) ;
- codification en CIMO2 des tumeurs non codées ou codées par une nomenclature interne (cas des lésions hématologiques) ;
- décomposition des codes Adicap et SNOMED en codes topographiques et morphologiques ;
- transcodification en CIMO2 à l'aide des outils disponibles (CIM10-CIMO2, Adicap-CIMO2, SNOMED-CIMO2) ;
- dédoublement intra-source : il était fréquent d'avoir plusieurs lignes pour un même cancer dans les fichiers des laboratoires d'ACP (par exemple, biopsie puis exérèse d'un organe) et dans les fichiers des DIM (plusieurs séjours pour un patient, plusieurs services cliniques fréquentés et traitements successifs pour un séjour).

Ensuite, seuls les patients dont les tumeurs répondaient aux critères d'inclusion retenus ont été sélectionnés.

3.2.3 Etape 3 : concaténation des fichiers médicaux

Une fois traités, les fichiers médicaux ont été concaténés en une seule base, appelée base initiale.

3.2.4 Etape 4 : recherche des doublons et homonymes

La recherche des doublons et homonymes dans la base initiale a été réalisée sur les variables suivantes : nom, nom de jeune fille, prénom, date de naissance. Ce travail a été réalisé à l'aide des fonctions avancées de programmation VBA (Visual Basic Application) d'Excel. A titre de vérification, le fichier dédoublement sous Excel a été importé sous CanReg 4 pour effectuer, à partir des fonctions propres à ce logiciel, une dernière recherche de doublons. Aucun nouveau doublon n'a été détecté.

3.2.5 Etape 5 : analyse médicale des données

L'ensemble des informations disponibles par patient, issues des différentes sources, a été analysé pour déterminer le nombre et la nature des cancers par patient. Les tumeurs de topographies identiques ou très proches chez un patient (notamment pour les tumeurs malignes oto-rhino-

laryngologiques (ORL), digestives basses et les lymphomes) ont été regroupées. Les tumeurs multiples par patient ont été comptabilisées selon les règles du Circ (annexe 8).

Pour chaque cancer identifié, les informations provenant des différentes sources ont été décrites par plusieurs variables :

- description de la tumeur selon la CIMO2 ;
- base de diagnostic : microscopique (hématologie, cytologie ou histologie de la tumeur primitive ou d'une métastase) ou non microscopique (clinique, radiologique, biologique...) ;
- date d'incidence (§ 2.1.6.2) ;
- nombre de sources par cas.

3.2.6 Etape 6 : validation des adresses par consultation des dossiers médicaux

Les adresses fournies par la plupart des sources d'information étaient imprécises ou incertaines :

- dans les fichiers des DIM, le libellé de l'adresse était rempli pour seulement environ la moitié des lignes ;
- dans les fichiers des CPAM, des BDE (qui ont servi pour identifier des patients de la zone d'étude dans les fichiers des laboratoires de diagnostic) et de certains laboratoires, les adresses fournies étaient les dernières adresses connues (et pas forcément celles au moment du diagnostic) en raison de leur mise à jour régulière.

Pour la majorité des cas, leur appartenance à la zone d'étude a cependant été considérée comme confirmée puisque :

- deux sources au moins mentionnaient une commune de domicile dans la zone d'étude ;
- et aucune source ne mentionnait de commune de domicile hors zone d'étude (pas de discordance).

Pour 110 cas, l'adresse au moment du diagnostic n'apparaissait pas certaine car :

- une source mentionnait une commune dans la zone d'étude mais cette appartenance à la zone d'étude n'était pas confirmée par une seconde source ;
- ou certaines sources mentionnaient une commune dans la zone d'étude, d'autres hors zone d'étude (discordance).

Leurs adresses ont alors été vérifiées dans les dossiers médicaux des hôpitaux et des cliniques, ou, lorsque c'était possible, à la CPAM du régime général dans les formulaires papier de demande d'ETM pour affection longue durée. Ces formulaires, appelés Pires (Protocole interrégime d'examen spéciaux), sont remplis par les médecins diagnostiquant un cancer et mentionnent les éléments diagnostiques et de projet thérapeutique justifiant l'ETM. Alors que les adresses des patients sont mises à jour dans les bases informatiques de la CPAM (et correspondent donc à la dernière adresse connue), les Pires indiquent l'adresse du patient au moment du diagnostic.

Parmi ces 110 cas, 36 (33 %) ont été exclus car ils ne résidaient pas dans la zone d'étude au moment du diagnostic. Lors de ces vérifications, un second cancer a été identifié pour un patient.

3.2.7 Etape 7 : validation médicale des cancers par consultation des dossiers médicaux

Après la validation des adresses incertaines, il restait dans la base 2 845 cas. Parmi ces cas, il y en avait 975 (34 %) pour lesquels aucune donnée de laboratoire de diagnostic n'était disponible (pas de donnée ACP ou hématologique) et 67 (2 %) pour lesquels les données des laboratoires indiquaient une lésion cytologique seulement évocatrice.

Dans un premier temps, il a été considéré que ces 1 042 cas n'étaient pas confirmés (ne répondaient pas à la définition de cas) et nécessitaient donc d'être validés par consultation d'un dossier médical contrairement aux 1 803 autres.

Une **première série** de consultations de dossiers médicaux pour validation médicale des cas (VMC) a donc été effectuée dans les hôpitaux, cliniques, CPAM et laboratoires de diagnostic. Les demandes de consultation des dossiers ont été adressées par courrier aux directions des

établissements (et aux CME pour les CHU de Grenoble et Lyon), des CPAM et des laboratoires. Pour les établissements hospitaliers et les cliniques, les VMC ont été organisées avec les archivistes ou les secrétaires des services cliniques concernés lorsque les archives n'étaient pas centralisées.

Les cas qui n'ont pas pu être résolus par consultation d'un dossier médical l'ont été en contactant le médecin traitant du patient identifié dans le dossier.

Pour cette première série de VMC, l'objectif était de trouver dans les dossiers la preuve médicale des cancers, qui pouvait être de nature histo-cytologique, clinique, radiologique ou biologique. Sur les 1 042 cas vérifiés, 615 (59 %) ont été exclus ; il s'agissait le plus souvent d'un diagnostic hors période d'étude ou d'une récurrence, ou d'une absence de cancer (tableau 2).

Cette première série de VMC a rapidement mis en évidence une limite importante des données informatiques recueillies, à savoir l'impossibilité de connaître l'histoire naturelle de la maladie puisque les requêtes portaient sur une fenêtre de temps limitée à la période d'étude (1994-2002), d'où l'absence d'information sur ce qui s'était passé antérieurement.

Dans un premier temps, cette problématique a été identifiée pour les cancers hématologiques et pour les cancers découverts au stade métastatique. En effet, les cancers hématologiques évoluent souvent de manière chronique pendant de longues années avant une phase aiguë. Quant aux cancers découverts au stade métastatique, il apparaissait essentiel de vérifier que le cancer primitif n'avait pas été diagnostiqué avant le début de la période d'étude.

Pour ces raisons, une **seconde série** de VMC a été entreprise pour les 111 cancers hématologiques et les 118 cancers découverts au stade métastatique qui n'avaient pas déjà été validés lors de la première série de VMC.

Concernant les cancers hématologiques, 17 (15 %) ont été exclus, le plus souvent pour diagnostic hors période d'étude ou récurrence, ou absence de cancer (tableau 2).

Pour les cancers découverts au stade métastatique, 34 (29 %) ont été exclus, à 76 % du fait que le cancer primitif était connu (tableau 2).

L'analyse de l'ensemble de ces premières vérifications a montré que les données informatiques recueillies présentaient d'autres limites :

- erreurs de codage dans les fichiers des laboratoires (lésions métastatiques codées comme primitives, lésions in situ codées invasives...) ;
- erreurs de codage dans les fichiers des DIM (surveillances de cancers codées en diagnostics, tumeurs bénignes codées en cancers...) ;
- attributions d'ETM pour des cancers in situ ;
- absence de moyen d'identifier les récurrences (l'Adicap dispose d'un code pour identifier les récurrences mais celui-ci est peu utilisé, et la SNOMED n'en dispose pas).

De ce fait, il est apparu nécessaire de vérifier la majorité des cas avec, pour objectif principal, de valider la date d'incidence du cancer. Les cancers pour lesquels il a été jugé indispensable de procéder à une vérification étaient :

- les cancers décrits par une seule source ;
- les cancers pour lesquels il n'y avait pas d'ETM ;
- les cancers multiples (plusieurs cancers chez un patient) ;
- les cancers pour lesquels les différentes sources d'information fournissaient des dates éloignées (écart de plus d'un an entre la date la plus ancienne et la plus récente) ;
- les cancers dont la date d'incidence était sur le premier semestre de 1994 ;
- les cancers du sein, souvent récidivants ;
- les cancers ORL, souvent multiples et mal codés ;
- les cancers de la vessie pour lesquels des lésions in situ étaient parfois codées invasives ;
- les sarcomes des tissus mous, cancers très rares et d'intérêt particulier pour l'étude.

Ainsi, pour les 1 574 cancers n'ayant pas fait l'objet d'une VMC à ce stade, une **troisième série** de VMC a été organisée : 1 089 ont été vérifiés dont 129 ont été exclus (12 %). Les exclusions étaient principalement des diagnostics hors période d'étude ou des récives, et des cancers in situ (tableau 2). Par ailleurs 6 nouveaux cas de cancers ont été identifiés (2^{ème} ou 3^{ème} cancer pour des patients de la base).

Tableau 2. Nombre d'exclusions et répartition des motifs d'exclusion en pourcentage pour chaque série de VMC

Motifs d'exclusion	Séries de VMC (nombre d'exclusions)			
	1 ^{ère} série (615)	2 ^{ème} série (héματο) (18)	2 ^{ème} série (méta) (34)	3 ^{ème} série (129)
Diagnostic hors période d'étude ou récive	54 %	50 %	0 %	40 % ^a
Absence de cancer	26 %	38 %	3 %	6 %
Métastase d'un cancer primitif connu	5 %	0 %	76 %	11 %
Cancer in situ	4 %	0 %	0 %	20 %
Domicile hors zone d'étude au moment du diagnostic	7 %	6 %	18 %	10 %
Cancer ne répondant pas aux autres critères d'inclusion ^b	3 %	6 %	3 %	3 %
Dossier médical non retrouvé ^c	1 %	0 %	0 %	10 %

^a Dont 36 % de récives.

^b Ces cancers étaient des cancers de la peau autres que des mélanomes, ou des cancers hématologiques non retenus jusqu'en 2000 (codés /1 en CIMO2) et retenus à partir de 2001 (codés /3 en CIMO3).

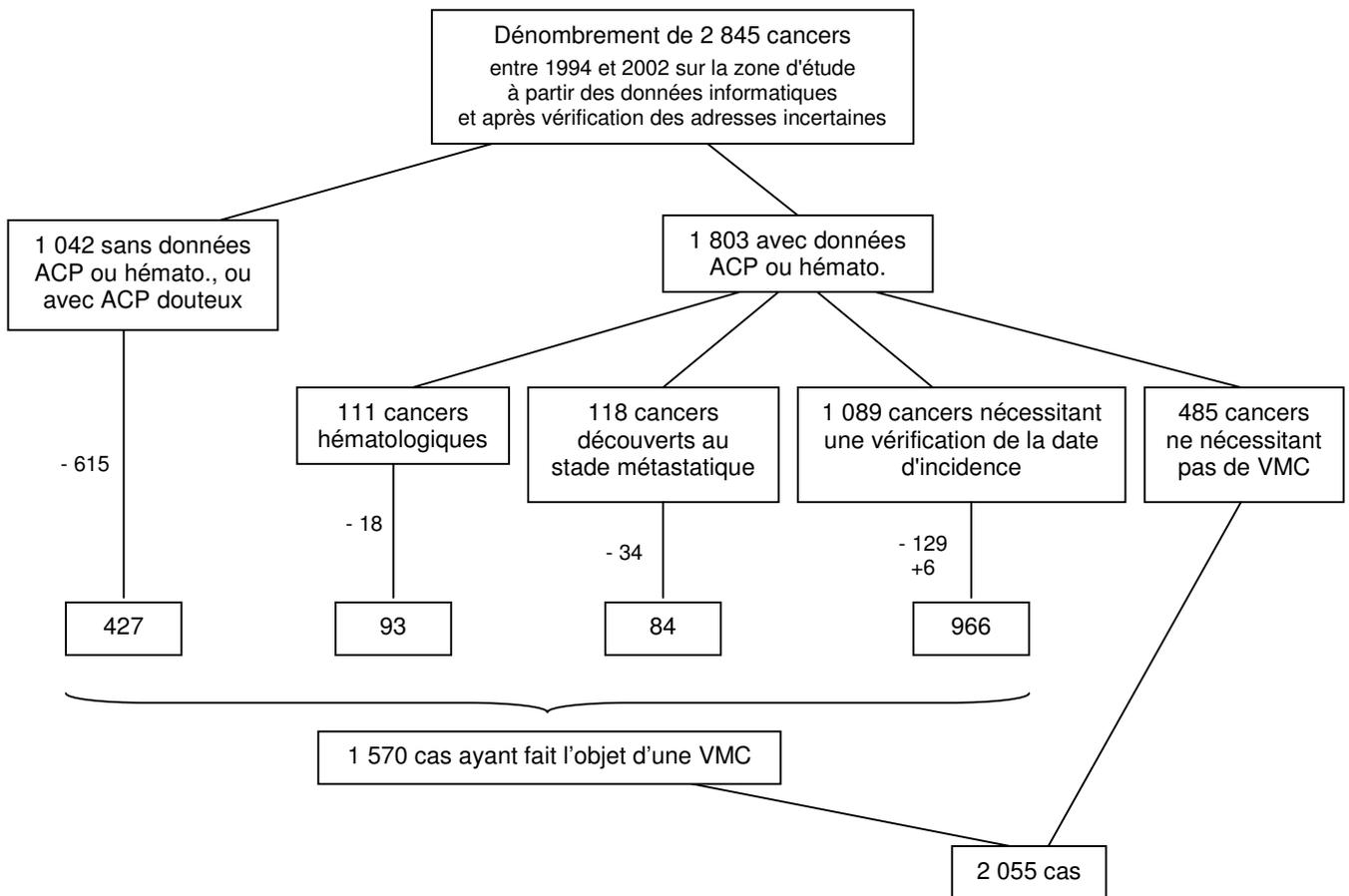
^c Lorsqu'aucun dossier médical n'a pu être retrouvé et qu'aucun médecin traitant n'a pu être contacté, la décision d'inclusion ou d'exclusion a été prise en considérant le nombre de sources médicales décrivant la tumeur, l'existence ou non d'une donnée histo-cytopathologique et d'une ETM pour le patient. Par exemple, un cancer décrit seulement par une source DIM a été exclu alors qu'un cancer décrit par une source DIM et par un laboratoire ou une CPAM a été inclus.

Au total, sur les 2 845 cas initialement recensés à partir des fichiers médicaux, 2 360 (83 %) ont fait l'objet d'une VMC. Environ 80 % de ces VMC ont été effectuées par consultation du dossier médical dans les hôpitaux et cliniques, les autres par consultation du Pires, du compte-rendu ACP ou par appel téléphonique aux médecins.

Sur les 2 360 VMC, 796 cancers ont été exclus (34 %). Il a par ailleurs été retrouvé dans les dossiers 6 cas non identifiés à partir des données informatiques. Un total de 2 055 cas a donc été recensé (figure 4).

Ce bilan montre que les données informatiques ne permettent pas, à elles seules, d'établir un dénombrement exact des cancers dans une zone et à une période d'étude définies et qu'il est indispensable de consulter les dossiers médicaux pour valider la majorité des cas.

Figure 4. Bilan des VMC



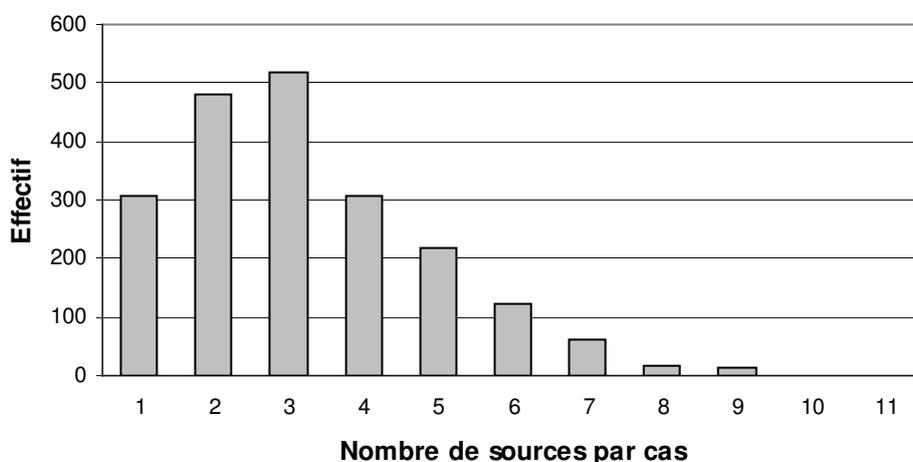
4 Bilan du recensement des cas

4.1 Nombre de sources par cas

Il s'agit du nombre de sources différentes ayant signalé un cas (ex : un DIM signalant un cas plusieurs fois compte pour une seule source) ; elle peuvent être de même type (ex : deux DIM signalant le même cas comptent pour deux sources).

Le nombre de sources par cas variait de 1 à 11 (figure 5), avec une moyenne de 3,2 et une médiane de 3,0.

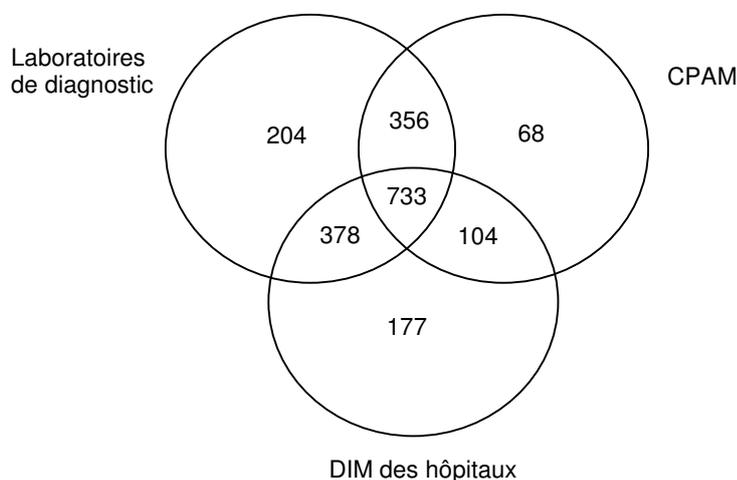
Figure 5. Distribution du nombre de sources par cas



4.2 Contribution des sources au recensement des cas

Trois types de sources ont contribué de façon majoritaire au recensement des cas : les laboratoires de diagnostic, les CPAM et les DIM des hôpitaux. La figure 6 indique le nombre de cas décrits médicalement par chacune de ces trois sources.

Figure 6. Répartition des cas selon les types de sources les ayant signalés (n=2 020)



Au total, 2 020 cas (soit 98 % des 2 055 cas recensés) étaient signalés par les laboratoires de diagnostic, les CPAM et les DIM des hôpitaux. Chacune des sources a signalé respectivement 1 671, 1 261 et 1 392 cas ; 733 des 2 020 cas (36 %) étaient signalés par les trois types de sources et 1 571 (78 %) par au moins deux sources différentes.

Les 35 cas non signalés par les trois principaux types de source ont été détectés à partir des données fournies par les médecins traitants (15 cas dont 12 uniques), les DIM des cliniques (5 cas dont 3 uniques), les registres spécialisés (3 cas uniques), les autres sources médicales (10 cas dont 7 uniques) ; 6 cas ont été identifiés lors des VMC.

Par ailleurs, pour 100 % des 1 261 cas signalés par les CPAM, leurs fichiers fournissaient une adresse dans la zone d'étude ; ce pourcentage était de 97 % pour les 1 392 cas signalés par les DIM des hôpitaux mais seulement de 72 % pour les 1 671 cas signalés par les laboratoires. Sur les 468 adresses non fournies par les laboratoires, 346 (74 %) étaient retrouvées dans les fichiers des CPAM et des DIM des hôpitaux, les autres sources médicales et administratives ayant été indispensables pour renseigner l'appartenance à la zone d'étude des 122 autres cas.

Au final, les trois principaux types de sources à eux seuls, auraient permis d'identifier 94 % des cas (disponibilité de l'information médicale et géographique).

4.3 Disponibilité de l'information médicale selon la localisation cancéreuse

Le tableau 3 présente le pourcentage de cas de cancers signalés par les trois principaux types de sources, en fonction de la localisation cancéreuse.

Tableau 3. Pourcentage de cas signalés par les laboratoires de diagnostic, les CPAM et les DIM des hôpitaux, par localisation cancéreuse

Localisation	Codes CIMO2	% de cas signalés par		
		labo.	CPAM	DIM hôp.
Lèvre-Bouche-Pharynx	C00.0 à C14.8	98 %	58 %	84 %
Œsophage	C15.0 à C15.9	92 %	72 %	90 %
Estomac	C16.0 à C16.9	94 %	52 %	81 %
Côlon-Rectum	C18.0 à C21.8	92 %	64 %	76 %
Foie	C22.0 à C22.1	47 %	36 %	87 %
Pancréas	C25.0 à C25.9	40 %	42 %	83 %
Larynx	C32.0 à C32.9	93 %	75 %	86 %
Poumon	C33.9 à C34.9	83 %	50 %	78 %
Mésothéliome de la plèvre	C38.4, M9050 à M9053	67 %	44 %	78 %
Mélanome de la peau	C44.0 à C44.9, M8720 à M8780	84 %	34 %	47 %
Sein	C50.0 à C50.9	86 %	83 %	54 %
Col de l'utérus	C53.0 à C53.9	100 %	68 %	74 %
Corps de l'utérus	C54.0 à C54.9	96 %	60 %	69 %
Ovaire	C56.9 à C57.4 sauf M8442, M8451, M8461, M8462, M8472 et M8473	86 %	71 %	81 %
Prostate	C61.9	87 %	71 %	39 %
Vessie	C67.0 à C67.9	90 %	66 %	51 %
Rein	C64.9 à C66.9, C68.0 à C68.9	82 %	71 %	56 %
Système nerveux central	C70.0 à C72.9, M≤9110 ou M≥9180	72 %	36 %	84 %
Thyroïde	C73.9	77 %	69 %	77 %
Lymphome malin non hodgkinien	C00.0 à C80.9, M9590 à M9595, M9670 à M9723, M9740 à M9741	85 %	52 %	85 %
Maladie de Hodgkin	C00.0 à C80.9, M9650 à M9667	82 %	71 %	76 %
Myélome multiple et maladie immunoproliférative	C00.0 à C80.9, M9731, M9732, M9760 à M9764, M9830	32 %	68 %	96 %
Toutes leucémies	C00.0 à C80.9, M9800 à M9941	34 %	56 %	85 %
Tous cancers	C00.0 à C80.9	81 %	61 %	68 %

Selon les localisations, certaines sources principales fournissent de l'information sur la quasi-totalité des cas ou sur une proportion beaucoup plus faible, parfois inférieure à 50 % :

- laboratoires de diagnostic : foie, pancréas, myélome multiple et maladie immunoproliférative, toutes leucémies ;
- CPAM : foie, pancréas, mésothéliome de la plèvre, mélanome de la peau, système nerveux central ;
- DIM des hôpitaux : mélanome de la peau.

4.4 Analyse de l'incidence des cancers

L'analyse des données d'incidence par calcul de ratios d'incidence standardisés est présentée dans un rapport indépendant [1].

5 Discussion et recommandations

5.1 Pertinence d'une étude d'incidence

En 2001, la crise environnementale, sociale et médiatique autour de l'UIOM de Gilly-sur-Isère a conduit l'InVS à prendre rapidement la décision de lancer une étude rétrospective d'incidence des cancers dans la zone impactée par les émissions atmosphériques de l'UIOM.

Aujourd'hui, le guide pour la conduite à tenir lors d'une demande locale d'investigations sanitaires autour d'une UIOM [9] aide à définir les actions à entreprendre selon la nature du signal (sanitaire, environnemental ou social). Notamment, la réalisation d'une EQRS permet d'identifier précisément la population exposée, de quantifier les niveaux d'exposition et les risques associés afin de définir d'éventuelles mesures de réduction des expositions et d'analyser la pertinence d'engager une étude épidémiologique. Par ailleurs, la méthode décrite dans le guide pour l'évaluation et la prise en charge des agrégats spatio-temporels de maladies non infectieuses [10] prévoit, lors du signal d'un cluster, de vérifier sa plausibilité et de valider les cas avant d'engager une étude descriptive approfondie comme celle qui a été menée autour de l'UIOM de Gilly-sur-Isère.

Ainsi, l'étude rétrospective d'incidence des cancers autour de l'UIOM de Gilly-sur-Isère montre les limites d'une approche épidémiologique descriptive comme réponse aux inquiétudes de la population : limites méthodologiques et délais importants de réalisation pour une réponse partielle aux questions (pas d'étude du lien de causalité).

5.2 Faisabilité d'une étude d'incidence

5.2.1 Exhaustivité et fiabilité de la base de données

L'objectif de l'étude était de déterminer s'il existait un excès de cas de cancers autour de l'UIOM en comparant l'incidence dans cette zone aux données d'incidence des registres des cancers. Il était donc indispensable d'obtenir des données de même qualité que celles des registres.

Le travail mené s'apparentait donc à celui d'un registre, mis à part qu'il s'agissait ici d'un recueil rétrospectif, d'où des différences de méthode :

- recueil de fichiers informatiques relatifs aux cancers diagnostiqués entre 1994 et 2002, au lieu de requêtes régulières auprès des fournisseurs de données ;
- nécessité de valider le diagnostic (disponibilité de codes uniquement et non de comptes-rendus ACP, erreurs de codage), le caractère incident des cancers (pas d'historique de données pour éliminer les récurrences, effet d'œillère sur la période d'étude) et l'appartenance à la zone d'étude au moment du diagnostic.

L'exhaustivité dépend en premier lieu d'une bonne identification des filières de soin. Il est probable que quelques patients aient été pris en charge en dehors des filières identifiées. Une partie d'entre eux ont été signalés par les CPAM (qui ont fourni les ETM pour cancer des patients de la zone d'étude quel que soit le lieu de leur traitement) ; ceux qui n'ont pas bénéficié d'ETM ont échappé au recueil, cependant les registres sont confrontés au même problème.

L'exhaustivité de la base d'un registre des cancers peut être appréciée par le nombre moyen de sources par cas [8]. Cet indicateur varie de 2,0 à 3,3 pour les registres français [6] et s'élève à 3,2 dans cette étude, ce qui permet de considérer que l'exhaustivité équivaut à celle d'un registre. Elle peut également être évaluée à partir du pourcentage de cas avec confirmation histologique : un taux proche de 100 % indique un sous-enregistrement des cas (un recensement basé principalement sur les laboratoires de diagnostic entraîne une mauvaise identification des cancers

sans preuve histologique). Le taux obtenu dans cette étude est de 91 %, ce qui est légèrement inférieur au taux des registres français (96 %) [6].

Inversement, il était essentiel que la base ne comporte pas de faux positifs (cancers diagnostiqués avant la période d'étude, récurrences de cancers, absences de cancer, métastases de primitifs connus...). La majorité des cas a donc été vérifiée dans les dossiers médicaux, ce qui s'est avéré indispensable au vu du nombre d'exclusions (34 % des cas vérifiés). Ainsi, la base de données avant l'étape de validation comptait 28 % de faux positifs.

5.2.2 Recommandations pour la réalisation d'une étude d'incidence

Les sources principales d'information sont les laboratoires de diagnostic (ACP et hématologie), les CPAM et les DIM des hôpitaux puisqu'à elles seules, elles permettent l'identification de 94 % des cas. Elles sont également complémentaires en termes de nature d'informations apportées (les données des laboratoires fournissent la nature histologique du cancer, les données des DIM renseignent le type de prise en charge). En revanche, l'utilisation de ces sources sans retour aux dossiers médicaux conduit à un recensement en excès car elles ne permettent pas de repérer les faux positifs.

En fonction des objectifs assignés à une étude locale d'incidence des cancers, différentes méthodes de recensement des cas peuvent être envisagées :

- pour une comparaison à des données de référence issues des registres des cancers, le recours à tous les fournisseurs de données potentiels et la validation des cas par consultation des dossiers médicaux sont indispensables pour assurer l'exhaustivité et la fiabilité du recueil ;
- pour des études écologiques temporelles ou spatiales (comparaison des incidences dans le temps ou l'espace), le recours aux trois sources principales peut suffire et la validation des cas est à discuter en fonction de la fiabilité recherchée.

Quelle que soit la méthode retenue, des conseils pratiques permettant d'optimiser la réussite d'un recueil rétrospectif des cas de cancers dans une population donnée sont formulés en annexe 9 (étude de faisabilité, composition de l'équipe, requêtes auprès des fournisseurs de données et validation des cas dans les dossiers médicaux).

6 Conclusion et perspectives

Une étude rétrospective d'incidence des cancers, dans un objectif de comparaison à des données de référence, est réalisable localement à condition que des moyens très importants soient mis en œuvre pour le recueil, le traitement et la validation des données (2,5 équivalents temps plein pendant trois ans pour recenser 2 055 cas).

Les difficultés majeures rencontrées lors de cette étude aurait pu être évitées si :

- les laboratoires de diagnostic enregistreraient systématiquement l'adresse des patients (même lorsque l'acte est facturé au prescripteur) ;
- les fournisseurs de données conservaient l'historique des adresses des patients (pour pouvoir retrouver l'adresse au moment du diagnostic) ;
- les erreurs de codification des cancers et des actes médicaux étaient moins fréquentes ;
- des outils validés de transcodification avaient existé (en particulier Adicap-CIMO).

L'InVS travaille sur l'élaboration d'un système de surveillance nationale des cancers basé sur le recueil prospectif et systématisé des données des laboratoires d'ACP, du PMSI et des caisses d'assurance maladie [11]. L'étude d'incidence des cancers autour de l'UIOM de Gilly-sur-Isère a montré que le recours à ces trois sources aboutit à une relativement bonne exhaustivité, mais que la qualité des données dépend de la capacité à détecter les faux positifs (récidives, métastases de primitifs connus...).

Il est prévu que ce système serve à l'investigation des clusters, ce qui devrait éviter, à terme, d'engager des études locales rétrospectives d'incidence.

7 Références

- [1] Thabuis A., Schmitt M. Etude rétrospective d'incidence des cancers autour de l'UIOM de Gilly-sur-Isère. Cire Rhône-Alpes, novembre 2006.
- [2] Colonna M. Répartition spatio-temporelle des cas de cancer dans la région de Gilly-sur-Isère (Savoie). Rapport d'étape : analyse de la mortalité. Registre des cancers de l'Isère, février 2004.
- [3] Rouhan A. Evaluation quantitative des risques sanitaires liés aux émissions atmosphériques de l'usine d'incinération d'ordures ménagères à Gilly-sur-Isère en Savoie (73). Careps, EDR n°385-3, mai 2004.
- [4] Lhomann R, Jones KC. Dioxins and Furans in air and deposition : a review of levels, behaviour and processes. Science Total Environ, 1998:53-81.
- [5] Jensen OM, Parkin DM, MacLennan R, Muir CS, Skeet RG. Enregistrement des cancers : principes et méthodes. IARC Publications scientifiques No. 95, 1996.
- [6] Remontet L, Buemi A, Velten M, Jouglu E, Estève J.. Evolution de l'incidence et de la mortalité par cancer en France de 1978 à 2000. Francim, HCL, Inserm, InVS, Août 2003.
- [7] <http://www.enr.com.fr/>
- [8] Parkin DM, Chen VW, Ferlay J, Galceran J, Stm HH, Whelan SL. Comparability and quality control in cancer registration. IARC Technical report No.19;1994.
- [9] InVS. Incinérateurs et santé. Guide pour la conduite à tenir lors d'une demande locale d'investigations sanitaires autour d'un incinérateur d'ordures ménagères. InVS 2003.
- [10] InVS. Guide méthodologique pour l'évaluation et la prise en charge des agrégats spatio-temporels de maladies non infectieuses. InVS 2005.
- [11] Chérié-Challine L, Bloch J. Propositions pour la mise en place du système de surveillance épidémiologique nationale des cancers. BEH 2003;n°41-42:194-7.

Annexes

Annexe 1 : Communes de la zone d'étude

Code Insee	Commune	Code postal	Bureau distributeur
73 007	Aiton	73 220	Aiguebelle
73 011	Albertville	73 200	Albertville
73 014	Allondaz	73 200	Albertville
73 032	La Bâthie	73 540	La Bâthie
73 048	Bonvillard	73 460	Frontenex
73 061	Césarches	73 200	Albertville
73 068	Chamousset	73 390	Chamoux-sur-Gelon
73 086	Cléry	73 460	Frontenex
73 110	Esserts-Blay	73 540	La Bâthie
73 120	Fréterive	73 250	Saint-Pierre-d'Albigny
73 121	Frontenex	73 460	Frontenex
73 124	Gilly-sur-Isère	73 200	Albertville
73 129	Grésy-sur-Isère	73 460	Frontenex
73 130	Grignon	73 200	Albertville
73 153	Marthod	73 400	Ugine
73 154	Mercury	73 200	Albertville
73 162	Montailleur	73 460	Frontenex
73 170	Monthion	73 200	Albertville
73 188	Notre-Dame-des-Millières	73 460	Frontenex
73 196	Pallud	73 200	Albertville
73 202	Plancherine	73 200	Albertville
73 211	Queige	73 720	Queige
73 241	Sainte-Hélène-sur-Isère	73 460	Frontenex
73 283	Saint-Vital	73 460	Frontenex
73 292	Thénésol	73 200	Albertville
73 297	Tournon	73 460	Frontenex
73 298	Tours-en-Savoie	73 790	Tours-en-Savoie
73 303	Ugine	73 400	Ugine
73 308	Venthon	73 200	Albertville
73 312	Verrens-Arvey	73 460	Frontenex

Annexe 2 : Carte de la zone d'étude



Annexe 3 : Grille de correspondance des codes de comportements tumoraux en Adicap et CIMO2

Adicap	CIMO2
0 : bénin	/0 = bénin
1 : bénin avec dysplasie légère	
2 : bénin avec dysplasie modérée	
3 : bénin avec dysplasie sévère	
4 : malignité incertaine ou limite	/1 = indéterminé si malin ou bénin
5 : cancer in situ ou intra-épithélial	/2 = carcinome in situ
6 : cancer micro-invasif	/3 = malin, siège primitif
7 : cancer invasif	
8 : récurrence tumorale	- : Pas de code
9 : cancer primitif ou secondaire non assuré	/9 = malin, incertain si primitif ou métastatique
E : extension secondaire par contiguïté	- : Pas de code
M : cancer métastatique	/6 = malin, siège métastatique
A : absence d'envahissement par le processus tumoral (re-prélèvement après biopsie exérèse)	- : Pas de code

Annexe 4 : Questionnaire sur les filières de soins envoyé aux médecins de la zone d'étude

DRASS Rhône-Alpes
CIRE Rhône-Alpes-Auvergne

Préparation de l'étude épidémiologique d'incidence des cancers dans la région d'Albertville.

Feuille de renseignement sur les filières de soin.

Nom du médecin : _____

Type de cancer	Etablissement	Ville	Service	Médecin

A renvoyer à :
Cellule Inter-Régionale d'Epidémiologie
DRASS Rhône-Alpes
107, rue Servient
69 418 Lyon Cedex 03
Tel. : 04 72 34 31 15

Annexe 6 : Codes retenus pour les requêtes

Requêtes dans les laboratoires d'ACP

- **Adicap** : comportement tumoral :
 - ≥ 6 pour tous les organes sauf le système nerveux central
 - ≥ 0 pour le système nerveux central
- **SNOMED** : comportement tumoral :
 - ≥ 3 pour tous les organes sauf le système nerveux central
 - ≥ 0 pour le système nerveux central

Requêtes dans les DIM

- **CIM 9 (avant 1997)** :
 - 140.0 à 239.9 tumeurs
 - V76.0 à V76.9 examen spécial de dépistage des tumeurs malignes
 - V67.1 / V67.2 examen de surveillance radiothérapie / chimiothérapie
 - V58.0 / V58.1 séance de radiothérapie / chimiothérapie d'entretien
- **CIM 10 (après 1997)** :
 - C00.0 à D48.9 tumeurs
 - Z12.0 à Z12.9 examen spécial de dépistage des tumeurs
 - Z08.0 à Z08.9 examen de contrôle après traitement d'une tumeur maligne
 - Z51.0 / Z51.1 séance de radiothérapie / chimiothérapie pour tumeur

Requêtes dans les CPAM

- **CIM 9 (avant 1997)** : 140.0 à 239.9
- **CIM 10 (après 1997)** : C00.0 à D48.9

Annexe 7 : Liste des fournisseurs de données et des requêtes réalisées

Fournisseur de données	Ville	Période couverte	Requête (RI : requête indirecte, RD : requête directe, QP : questionnaire papier)	Remarque
Laboratoires d'ACP				
Laboratoire privé (Knopf)	Annecy (74)	1994-2002	RD	
Laboratoire privé (Donzel)	Chambéry (73)	1994-2002	RD (Cire)	
Laboratoire privé (Ciapa)	Grenoble (38)	1994-2002	RD (Cire)	
Laboratoire privé (Le Marc'Hadour)	Grenoble (38)	1995-2002	RD (prestataire)	
Laboratoire privé (Ney)	Grenoble (38)	1994-2002	RD (prestataire)	
Laboratoire privé (Rouault Plantaz)	Grenoble (38)	1994-2002	RD (prestataire)	
Laboratoire privé (Chouvet)	Lyon (69)	1998-2002	RD (Cire)	
CH	Chambéry (73)	1994-2002	RD	
CHU (histologie)	Grenoble (38)	1994-2002	RD (prestataire)	
CHU (cytologie)	Grenoble (38)	1994-2002	RD	
HCL Hôpital de l'Hôtel-Dieu (dermatopathologie)	Lyon (69)	2000-2002	RI	Pas de données dans la zone d'étude
HCL Hôpital Debrousse	Lyon (69)	1994-2002	RD (Cire)	
HCL Hôpital de l'Hôtel-Dieu	Lyon (69)	1994-2002	RD (Cire)	
HCL CH Lyon Sud	Lyon (69)	1994-2002	RD (Cire)	
HCL Hôpital Edouard Herriot (HEH)	Lyon (69)	1994-2002	RD (Cire)	
HCL HEH (dermatopathologie)	Lyon (69)	1994-2000	RI	
HCL Hôpital Pierre Wertheimer	Lyon (69)	1994-2002	RD (Cire)	
HCL Hôpital Louis Pradel	Lyon (69)	1994-2002	RD (Cire)	
HCL Hôpital de la Croix-Rousse	Lyon (69)	1994-2002	RD (Cire)	
Hôpital d'instruction des armées (HIA) Desgenettes	Lyon (69)	1996-2002	RD	
Centre Léon Bérard (CLB)	Lyon (69)	1994-2002	RD	
Laboratoires d'hématologie				
CHU	Grenoble (38)	1998-2002	RD	
HCL Hôpital de la Croix-Rousse	Lyon (69)	1994-2002	RI	Pas de données dans la zone d'étude
HCL Hôpital Debrousse	Lyon (69)	1994-2002	RI	
HCL HEH	Lyon (69)	1994-2002	RI	
HCL CH Lyon Sud	Lyon (69)	1999-2002	RD	
HCL Hôpital Louis Pradel	Lyon (69)	1994-2002	RI	Pas de données dans la zone d'étude
CLB	Lyon (69)	1994-2002	RI	Pas de données dans la zone d'étude
CPAM				
Régime général	Chambéry (73)	1994-2002	RD	
Mutualité sociale agricole	Chambéry (73)	1994-2002	RD	Pas de données pour les patients décédés avant 1999
Caisse maladie régionale des Alpes	Grenoble (38)	1994-2002	RD	
Sécurité sociale des mines	La Motte d'Aveillans (38)	1994-2002	RD	Pas de données dans la zone d'étude
Caisse de prévoyance et de retraite SNCF	Marseille (13)	1994-2002	RD	
Caisse nationale militaire de sécurité sociale	Toulon (83)	1994-2001	RD	

DIM des hôpitaux				
CH	Albertville (73)	1998-2002	RD	
CH	Moutiers (73)	1996-2002	RD	
CH	Aix-les-Bains (73)	1994-2002	RD	
CH	Chambéry (73)	1996-2002	RD	
CH	Annecy (74)	1994-2002	RD	
CHU	Grenoble (38)	1994-2002	RD	
HCL	Lyon (69)	1994-2002	RD	
HIA Desgenettes	Lyon (69)	1996-2002	RD	
CLB	Lyon (69)	1994-2002	RD	
DIM des cliniques				
Clinique chirurgicale Dr Herbert	Aix-les-Bains (73)	2001-2002	RD	
Clinique générale du Dr Cléret	Chambéry (73)	1997-2002	RD	
Clinique chirurgicale Saint-Joseph	Chambéry (73)	1997-2002	RD	
Clinique d'Argonay	Annecy (74)	1997-2001	RD	
Clinique Générale	Annecy (74)	2000-2002	RD	
Institut Privé de Cancérologie	Grenoble (38)	1997-2002	RD	
Clinique du Mail	Grenoble (38)	1994-2002	RD	
Clinique des Bains	Grenoble (38)	1994-2002	RD	
Clinique mutualiste des Eaux Claires	Grenoble (38)	1995-2002	RD	
Clinique des Cèdres	Grenoble (38)	1996-2002	RD	
Clinique Belledonne	Grenoble (38)	1997-2002	RD	
Médecins libéraux				
Pédiatre (Dr Baille-Barbin)	Albertville (73)	1995-2002	QP	
Oto-rhino-laryngologiste (Dr Bory)	Albertville (73)	2002	QP	
Généraliste (Dr Dechaume)	Albertville (73)	1995-2002	QP	
Généraliste (Dr Dechaume Barba)	Albertville (73)	1996-2002	QP	
Généraliste (Dr Gallot Joselzon)	Albertville (73)	1994-2002	QP	
Gynécologue-obstétricien (Dr Ghawi)	Albertville (73)	1994-2002	QP	
Pédiatre (Dr Laplane)	Albertville (73)	1997	QP	
Pédiatre (Dr Plasse)	Albertville (73)	1995-2002	QP	
Ophthalmologiste (Dr Vittot de Lagarde)	Albertville (73)	2002	QP	
Généraliste (Dr Zoudianos)	Albertville (73)	1994-2002	QP	
Généraliste (Dr Duray)	Frontenex (73)	2002	QP	
Généraliste (Dr Vincensini)	Frontenex (73)	1995-2002	QP	
Généraliste (Dr Labbaye)	Gilly-sur-Isère	2000-2002	QP	
Généraliste (Dr Péria)	Gilly-sur-Isère	1994-2002	QP	
Généraliste (Dr Estève)	Grignon (73)	1994-2002	QP	
Généraliste (Dr Lange)	Mercury (73)	1994-2002	QP	
Généraliste (Dr Lafontaine)	Sainte-Hélène-sur-Isère (73)	1994-2002	QP	
Généraliste (Dr Belmont)	Ugine (73)	1994-2002	QP	
Généraliste (Dr Lepasant)	Ugine (73)	1994-2002	QP	
Généraliste (Dr Trolong)	Ugine (73)	1998-2002	QP	
Généraliste (Dr Fenoy)	Venthon (73)	1999-2002	QP	
Registres spécialisés des cancers				
Registre Rhône-Alpes des cancers de la thyroïde	Lyon (69)	1998-2001	RD	
Association du registre des cancers de l'enfant de la région Rhône-Alpes (Arcerra)	Saint-Etienne (42)	1994-2002	RD	
Registre national des leucémies et lymphomes de l'enfant	Villejuif (94)	1994-2002	RD	

Autres sources médicales				
Service de radiothérapie, CH	Chambéry (73)	1994-2002	RD	
Service d'oncolo-hématologie, CH	Chambéry (73)	1994-2002	RD	
Service d'onco-hématologie, CHU	Grenoble (38)	1994-2002	RD (Cire)	
Centre de ressources informatiques hospitalières, CHU	Grenoble (38)	1996-2002	RD	
Unité de concertation, CHU	Grenoble (38)	1994-2002	RD	
Enquête permanente cancer, CLB	Lyon (69)	1994-2002	RD	
BDE des hôpitaux				
CH	Albertville (73)	1994-2002	RD	
CH	Moutiers (73)	1995-2002	RD	
CH	Aix-les-Bains (73)	1994-2002	RD	
CH	Chambéry (73)	1994-2002	RD	
CH	Annecy (74)	1997-2002	RD	
CHU	Grenoble (38)	1994-2002	RD	
HCL	Lyon (69)	1944-2002	RD	
BDE des cliniques				
Clinique chirurgicale Dr Herbert	Aix-les-Bains (73)	1998-2002	RD (prestataire)	
Clinique d'Argonay	Annecy (74)	1995-2002	RD (prestataire)	
Clinique Générale de Savoie	Chambéry (73)	1994-2002	RD (prestataire)	
Clinique Sainte-Marie	Chambéry (73)	1997-2001	RD (prestataire)	
Institut Privé de cancérologie	Grenoble (38)	1994-2002	RD (prestataire)	
Clinique Belledonne	Grenoble (38)	1994-2002	RD	
Clinique mutualiste des Eaux Claires	Grenoble (38)	1994-2002	RD	
Clinique du Mail	Grenoble (38)	1994-2002	RD	
Clinique des Bains	Grenoble (38)	1994-2002	RD (prestataire)	
Clinique des Cèdres	Grenoble (38)	1996-2002	RD	
Autre source administrative				
ERSM	Lyon (69)	1994-2002	RD	

Annexe 8 : Résumé des règles de comptabilisation des tumeurs multiples utilisées dans l'étude

Ces règles sont proposées par le Circ [5] :

1. L'identification de deux tumeurs primitives ou plus ne dépend pas du temps.
2. Un cancer primitif est une tumeur qui prend naissance dans un tissu ou un organe précis ; il ne s'agit pas d'une extension, d'une récurrence ou d'une métastase.
3. Une seule tumeur est prise en compte par organe, paire d'organes ou tissu. Certains groupes de codes de localisation en CIMO2 sont considérés comme correspondant à un seul organe. Les tumeurs multifocales, correspondant au développement dans un même organe ou tissu de plusieurs cancers primitifs sans apparente continuité, sont considérées comme un seul cancer.
4. Si plusieurs cancers primitifs histologiquement différents apparaissent dans un même organe ou tissu, chacune de ces tumeurs est prise en compte.
5. Dans le cas des tumeurs systémiques ou de cancers multicentriques atteignant plusieurs organes (lymphomes, leucémies, sarcomes de Kaposi), un seul cas est enregistré par individu.

Annexe 9 : Conseils pratiques pour la réalisation d'une étude rétrospective d'incidence des cancers

Etude de faisabilité

L'étude de faisabilité doit permettre de savoir si un recensement exhaustif des cas de cancers est possible, en étudiant les points suivants :

- Identifier les filières de soins locales en oncologie et en hématologie en demandant aux médecins de la zone d'étude dans quels établissements (publics et privés) ils réfèrent leurs patients atteints d'un cancer.
- Identifier l'ensemble des fournisseurs de données potentiels :
 - médecins de la zone d'étude ;
 - établissements publics et privés (DIM, services cliniques, BDE) ;
 - laboratoires publics d'ACP et d'hématologie des établissements identifiés et laboratoires privés avec lesquels les cliniques et certains spécialistes (dermatologues) travaillent (y compris en dehors de la région) ;
 - régimes d'assurance maladie couvrant la population de la zone d'étude et médecins conseils correspondants (départementaux et régionaux) ;
 - registres des cancers spécifiques couvrant la zone d'étude.
- Etudier la disponibilité des données :
 - format des données (informatique, papier), période couverte, logiciels utilisés ;
 - modalités d'archivage des dossiers médicaux dans les laboratoires de diagnostic, les hôpitaux, les cliniques et les CPAM (responsable, archiviste, archivage centralisé ou décentralisé, lieux de conservation, durée de conservation, informatisation et possibilité de consultation du dossier médical informatisé) ;
 - modalités de réalisation des requêtes : par le fournisseur lui-même, par un prestataire de service ou par l'équipe chargée de l'étude.
- Demander les autorisations incontournables auprès de la Cnil, des établissements (direction, CME, chefs de services) et des médecins conseils, et établir des contacts utiles pour débloquer certaines réticences (syndicats des médecins ACP, de l'Ordre des médecins, de l'Union régionale des médecins libéraux...).
- Prévoir un budget pour rétribuer certains fournisseurs de données :
 - laboratoires libéraux et sociétés prestataires de services (laboratoires de diagnostic, DIM de certaines cliniques...) ;
 - médecins libéraux, qui ont des difficultés à retrouver des informations rétrospectives concernant leurs patients : soit ils ne sont pas informatisés, soit leur système informatique ne permet pas de réaliser des requêtes car les pathologies ne sont pas codées, et ils n'ont pas le temps de consulter les dossiers papier (parfois déjà archivés). Une rétribution pourrait les motiver à consacrer du temps à ce travail, et augmenterait leur taux de participation.
- Identifier les systèmes de codification utilisés par les différents fournisseurs de données et se procurer les nomenclatures correspondantes (en version informatique si possible), ainsi que les outils de transcodification.

Composition de l'équipe chargée de l'étude

L'équipe doit être composée au minimum d'un responsable scientifique et de deux enquêteurs, travaillant en étroite collaboration :

- Le responsable scientifique de l'étude est un épidémiologiste chargé d'élaborer le protocole de l'étude (définition de cas, zone et période d'étude...), de réaliser l'étude de faisabilité, d'animer le comité de pilotage de l'étude, d'effectuer la demande d'autorisation à la Cnil, de recruter, former et encadrer les enquêteurs, de recueillir les données de référence, d'effectuer les analyses, de rédiger le rapport d'étude et d'assurer la communication des résultats.
- Le médecin enquêteur est un médecin formé à l'épidémiologie et à la santé publique, ayant une bonne connaissance du milieu hospitalier. Outre ses qualités relationnelles (nombreux contacts avec les fournisseurs de données et travail en équipe) et de discrétion, il doit faire preuve de rigueur, d'une bonne capacité d'organisation, d'une forte autonomie et être mobile (nombreux déplacements). Une expérience professionnelle dans un registre (connaissances en cancérologie, compétences en codification des cancers...) est souhaitable.
- Le moniteur d'étude est titulaire d'un diplôme scientifique de niveau minimum bac+3, formé en statistique et traitement informatique des données : analyse, gestion et traitement de bases de données, maîtrise de l'algorithmie et des langages de programmation VBA et SQL (Structured Query Language). Il doit faire preuve de rigueur, de discrétion et d'une grande capacité d'organisation et de travail en équipe. Une expérience dans le domaine de la santé est souhaitable.

En l'absence d'expérience professionnelle dans un registre, l'idéal est que les deux enquêteurs soient formés par un registre des cancers aux principes et méthodes de l'enregistrement des cancers, en particulier sur :

- les notions de base en cancérologie générale et spécialisée (hématologie) ;
- les classifications existantes des cancers, leurs liens et leurs évolutions ;
- les règles de codification des cancers et les outils de transcodification.

Requêtes auprès des fournisseurs de données

- Pour chaque type de fournisseur de données, rédiger au préalable un cahier des charges précis définissant la période d'étude, la zone d'étude, les cancers retenus (topographie, morphologie, comportement) et les variables à recueillir.
- Lorsque l'adresse des patients n'est pas connue (cas de nombreux laboratoires de diagnostic), réaliser une requête plus large en recueillant sur l'adresse des établissements ou médecins prescripteurs au lieu de celle des patients.
- Prévoir d'élargir les requêtes d'une ou deux années par rapport à la période d'étude, afin d'aider à choisir la date d'incidence pour les cancers en bordure de période (ex : 1992-2003 pour une période d'étude de 1994-2002).
- Ne pas oublier de requêter le nom de jeune fille en plus des identifiants de base (nom, prénom, date de naissance) : cette variable est très importante lors de la phase de recherche des doublons.
- Recueillir le nom du médecin prescripteur, l'établissement d'exercice, voire le régime d'Assurance maladie auprès des laboratoires de diagnostic. Ces variables sont particulièrement utiles pour organiser les VMC.

Validation des cas dans les dossiers médicaux

Les VMC nécessitent d'identifier, pour chaque cas, le lieu où se trouve le dossier médical le plus approprié (établissement où a été pris en charge le patient, médecin traitant, CPAM, laboratoires). Dans le cas où les archives des établissements ne sont pas centralisées et sont réparties dans les services cliniques, il faut en plus identifier les services où sont passés les patients, demander l'autorisation à chaque chef de service et organiser la sortie des dossiers avec les secrétaires. En groupant les visites sur site et après les avoir parfaitement préparées, 50 à 100 dossiers sont vérifiables par jour.

Résumé

L'étude rétrospective d'incidence des cancers autour de l'usine d'incinération d'ordures ménagères de Gilly-sur-Isère, décidée dans un contexte de crise lors de sa fermeture fin 2001, avait pour objectif de déterminer s'il existait un excès de cancers autour de l'installation. Elle a consisté à recenser de la manière la plus exhaustive possible les cancers survenus entre 1994 et 2002 dans la zone d'étude (exposée aux retombées atmosphériques de l'incinérateur) afin de comparer l'incidence observée à celle établie par les registres des cancers français. Ce retour d'expérience décrit les principales difficultés rencontrées et les solutions apportées lors du recensement des cas de cancers.

Le recueil des données médicales a été réalisé auprès de multiples sources d'information : laboratoires d'anatomo-cyto-pathologie et d'hématologie, départements d'information médicale (DIM) et services spécialisés des hôpitaux et cliniques, caisses primaires d'assurance maladie (CPAM), médecins libéraux, registres des cancers spécialisés. Les 59 fichiers informatiques recueillis ont été traités : recherche des adresses manquantes (en utilisant des données administratives des bureaux des entrées des établissements de soin et de la Caisse nationale d'assurance maladie), sélection des patients de la zone d'étude, homogénéisation des codages des cancers, concaténation des fichiers en une seule base, analyse des informations disponibles sur chaque cancer, dédoublement. La majorité des cancers a ensuite été validée par consultation des dossiers médicaux, afin d'éliminer les faux positifs (métastases de cancer primitif connu, récidives...).

Sur 2 360 cancers ayant fait l'objet d'une validation, 34 % ont été exclus ; au final, la base comprenait 2 055 cas de cancers. Trois types de sources ont permis d'identifier à eux seuls 94 % des cas : laboratoires, DIM des hôpitaux et CPAM. Avec une moyenne de 3,2 sources par cas et un taux de 91 % des cas avec confirmation histologique, la base pouvait être considérée comme ayant la même exhaustivité que celle d'un registre.

Le recours à des données administratives et la consultation des dossiers médicaux se sont révélés nécessaires compte tenu des incertitudes sur les adresses des patients au moment du diagnostic, des erreurs de codage des cancers dans certains fichiers recueillis et des difficultés à identifier les faux positifs. En conclusion, une étude rétrospective d'incidence des cancers est réalisable localement à condition que des moyens humains très importants soient mis en œuvre (2,5 équivalents temps plein pendant trois ans pour recenser 2 055 cas). Des recommandations et des conseils pratiques pour la réalisation d'une telle étude sont formulés en fin de rapport.

Abstract

The retrospective cancer incidence study carried out around the municipal solid waste incinerator of Gilly-sur-Isère was ordered in a context of crisis during its closing in the late 2001. Its purpose was to determine whether or not there was an exceeding number of cancers around the incinerator. This study consisted in counting as exhaustively as possible the cancers that occurred between 1994 and 2002 in the study area, which was exposed to the atmospheric fallouts from the incinerator. Thus, it was planned to compare the observed cancer incidence to the French cancer registries'. This work describes the main difficulties encountered as well as the solutions found during the enumeration of cancer cases.

The collection of medical data was carried out thanks to multiple sources of information: pathology and hematology laboratories, hospitals and clinics medical informatics departments, clinical services, health insurance funds, liberal practitioners or specialised cancer registries. The 59 collected medical data files were dealt with: looking for the missing addresses (by using administrative data collected from hospital admission desks and the national health insurance fund for salaried workers), selecting patients from the study area, homogenizing cancers coding, merging files into a single database, analysing available information on each cancer, de-duplicating the database. Most cancers were validated by consulting medical folders so as to exclude the false cases like metastasises of a known primary cancer or recurrences.

2360 cancers were validated and 34% of them were excluded. The final database was made of 2055 cancer cases. Three types of sources only allowed to identify 94% of cases: laboratories, hospital medical informatics departments and health insurance funds. With an average of 3.2 sources per case and 91% of histologically-verified cases, the database could be considered as much exhaustive as a registry's.

Resorting to administrative data and consulting medical folders turned out to be necessary considering uncertainties about the patients' place of residence by the time of the diagnosis, errors in coding cancers in a few databases that were collected and difficulties to identify false cases. As a conclusion, a retrospective cancer incidence study can be carried out at a local scale provided a very important manpower is brought into play (2.5 full-time workers during three years to take a census of 2055 cases). Some recommendations as well as practical advice to carry out such a study are given at the end of the report.