

SANTÉ  
ENVIRONNEMENT  
TRAVAIL

SEPTEMBRE 2021

ÉTUDES ET ENQUÊTES

FAISABILITÉ DE LA MISE EN PLACE  
D'UN SYSTÈME DE SURVEILLANCE DE  
L'INCIDENCE DES CANCERS EN LIEN  
AVEC L'ACTIVITÉ PROFESSIONNELLE :  
ÉTUDE PILOTE SICAPRO (2010-2014)

En partenariat avec :

## Résumé

### **Faisabilité de la mise en place d'un système de surveillance de l'incidence des cancers en lien avec l'activité professionnelle : étude pilote Sicapro (2010-2014)**

Ce rapport présente les résultats de l'étude de la faisabilité de la mise en place d'un système de surveillance de l'incidence des cancers en lien avec l'activité professionnelle. Il n'existe actuellement pas de système de surveillance de ce type en France. Les estimations de l'incidence des cancers par métier ou secteur d'activité représentent des informations d'intérêt pour la description et la compréhension des relations entre facteurs de risque professionnels et apparition de cancer dans la population française. Elles permettraient d'identifier les métiers ou secteurs d'activité à forte incidence de cancer et les évolutions au cours du temps, afin d'orienter les politiques publiques destinées à améliorer et d'évaluer la prévention et la reconnaissance des cancers d'origine professionnelle.

Le système envisagé de surveillance de l'incidence des cancers en lien avec l'activité professionnelle (Sicapro), dont l'objectif principal est d'identifier des groupes professionnels à risque accru de cancer, se base sur le couplage au niveau individuel, des données médicales issues des registres des cancers avec les données sur la carrière professionnelle provenant de la Caisse nationale d'assurance vieillesse (Cnav).

Dans un premier temps, afin de tester la faisabilité de la mise en place d'un tel système de surveillance, une étude pilote a été réalisée avec deux registres volontaires (Registre général des cancers de Lille et de sa région et Registre des tumeurs du Doubs et du Territoire de Belfort) qui ont fournis les données concernant l'ensemble des cas de cancers enregistrés pour les années 2010-2014.

Les objectifs de cette étude pilote étaient d'étudier la performance de l'appariement à partir des données de l'état civil des personnes atteintes de cancer issues des registres et celles figurant dans les données de la Cnav ; d'identifier les données d'historiques professionnels disponibles dans les bases de la Cnav et d'étudier la qualité de ces données administratives dans un but de surveillance épidémiologique ; de mettre au point les différents indicateurs qui pourront être construits en routine pour le système de surveillance final et d'étudier les limites de la démarche, notamment dans l'identification des expositions professionnelles.

Il existe en France depuis les années 1970, des registres de cancers généraux ou spécialisés qui permettent une surveillance épidémiologique locale et nationale des cancers en enregistrant de manière exhaustive l'ensemble des cas de cancers survenant parmi la population résidant dans le territoire couvert par chaque registre. Les registres de cancers généraux recueillent des informations sur tous les cancers alors que les registres de cancers spécialisés recueillent de nombreuses informations sur certains cancers (appareil digestif, hémopathies malignes, sein et autres cancers gynécologiques, système nerveux central, thyroïde, mésothéliome ou concernant des populations spécifiques (enfants et adolescents). Les deux registres des cancers volontaires pour la participation à cette étude pilote (Lille et sa région ; Doubs et Territoire de Belfort) sont des registres généraux, intégrant l'ensemble des localisations de cancers.

Les données recueillies par la Cnav contiennent l'ensemble des données inhérentes à la carrière des assurés afin de calculer leur retraite. Sont inclus l'ensemble des personnes ayant au moins une période d'activité donnant lieu à cotisation au régime général quelle que soit la durée et concerne l'ensemble des individus ayant travaillé ou travaillant en France métropolitaine ou ultramarine, y compris la population étrangère. Les remontées d'informations provenant des autres régimes d'affiliation (Mutuelle sociale agricole, Régime

social des indépendants, etc.) sont remontées afin de permettre le calcul des droits à la retraite, généralement à l'approche de celle-ci, en fin de carrière.

L'appariement des données a été réalisé par la Caisse nationale d'assurance vieillesse sur les données d'état-civil (nom, prénom, date et commune de naissance) en l'absence du recueil du numéro d'inscription au répertoire national d'identification des personnes physiques (NIR) par les registres de cancers. La performance de cet appariement était satisfaisante puisque 96,9% des cas de cancers ont été retrouvés dans les bases de la Cnav.

La reconstitution des historiques de carrières a été réalisée à partir des numéros employeurs, le secteur d'activité directement accessible via le code d'activité principale exercée n'étant renseigné de manière systématique que pour les années les plus récentes. Le numéro d'entreprise (Siret) recueilli dans la base permet de retrouver le secteur d'activité (NAF) de l'entreprise via la base Sirene® de l'Insee.

La faisabilité de la reconstitution de l'historique professionnel dans la perspective de réaliser des mesures d'associations entre survenue du cancer et secteur d'activité s'est révélée également satisfaisante à partir de l'année 1999. À partir des cas incidents de 2019, nous disposerons donc d'un historique professionnel fiable sur les vingt dernières années de carrières, ce qui semble suffisant y compris pour des études portant sur des maladies avec des durées de latence longues comme les cancers.

Le renseignement de la catégorie socio-professionnelle (CSP) par les employeurs est obligatoire à partir de l'année 2008 mais des problèmes de remontées d'information pour les années 2009 à 2011 n'ont pas permis de fournir des indicateurs dans le cadre de cette étude pilote. Néanmoins ces informations sont exploitables à partir de l'année 2012 et permettront de fournir des indicateurs d'incidence de cancer par CSP au moment du diagnostic pour le système de surveillance étendu

Enfin, deux exemples d'utilisation possibles de ces données sont présentés ici. La première étude présente les ratios standardisés d'incidence de cancer du poumon selon le secteur d'activité occupé au moment du diagnostic chez les cas âgés de 20 à 64 ans et enregistrés entre 2010 et 2014 par le registre du cancer du Doubs et du territoire de Belfort. La seconde est une étude cas-témoins estimant le risque de la survenue d'un cancer du poumon pour l'ensemble de la carrière pour les deux registres en considérant comme exposés les individus ayant travaillé au moins un an versus les individus ayant travaillé moins d'un an pour chaque secteur d'activité. Ces résultats sont mis en perspective avec le calcul d'indices de disparité d'incidence attendue du fait de différences dans la prévalence de tabagisme par secteurs d'activité, le tabac restant le principal facteur de risque pour le cancer du poumon.

La principale limite identifiée lors de cette étude est liée à la population couverte par la Caisse nationale d'assurance vieillesse. En effet la Cnav ne couvre que les salariés au régime général et les remontées des autres régimes ne sont donc complètes que pour les individus retraités. De plus les informations remontées par les autres régimes ne permettent pas de reconstituer les secteurs d'activité occupés au cours de la carrière ou les CSP pour les périodes où les individus étaient à leur compte ou salariés cotisants à un autre régime. Cette limite pourrait être levée par la mise en œuvre par la Cnav du répertoire de gestion des carrières unique (RGCU) visant à mettre en commun les données relatives à la carrière de chaque assuré issues de l'ensemble des régimes en vue d'assurer la complétude et la cohérence de ces données. L'alimentation du RGCU en temps réel par les autres régimes, fixée au 31 décembre 2022, pourrait ainsi permettre d'étendre la population couverte par le système de surveillance à l'ensemble des actifs, quel que soit le régime d'affiliation.

En dépit des limites mises en évidence lors de cette phase pilote, le système de surveillance Sicapro possède les propriétés d'un système de surveillance efficace : il ne nécessite pas de

recueil de données complémentaires, une forte sensibilité car les registres recueillent de façon exhaustive les cas incidents de leur zone et il est représentatif de la population salariée au régime général.

L'inclusion de l'ensemble des registres de cancers au système de surveillance, couvrant près de 20% de la population, permettra une plus grande puissance permettant une étude de l'incidence pour des localisations de cancers plus rares et/ou pour des secteurs d'activité à un niveau plus fin. Le système de surveillance permettra également de produire des indicateurs à un niveau départemental pour les départements intégralement couverts par un registre, y compris pour certains départements d'outre-mer (Guadeloupe, Martinique et Guyane).

**MOTS CLÉS :** PROFESSIONS, TRAVAILLEURS, SURVEILLANCE DE LA POPULATION, INCIDENCE, CANCER, CANCER DU POUMON, FRANCE

**Citation suggérée :** Lapostolle A. *Faisabilité de la mise en place d'un système de surveillance de l'incidence des cancers en lien avec l'activité professionnelle : étude pilote Sicapro (2010-2014)*. Saint-Maurice : Santé publique France, 2021. 60 p. Disponible à partir de l'URL : [www.santepubliquefrance.fr](http://www.santepubliquefrance.fr)

ISSN : 2609-2174 - ISBN-NET : 979-10-289-0734-1 - RÉALISÉ PAR LA DIRECTION DE LA COMMUNICATION, SANTÉ PUBLIQUE FRANCE — DÉPÔT LÉGAL : SEPTEMBRE 2021

## Abstract

### Feasibility of the implementation of a monitoring system of cancer incidence among employees in France

Occupational exposure to carcinogens is estimated to be one of the most important causes of cancer in France. Because this problem leads to major issues in terms of prevention, for recognition of occupational diseases and for reparation of the prejudice of occupational disease, a monitoring system of cancer among employees is needed in France.

This study aims at evaluating the feasibility of the implementation of a monitoring system of incidence of occupational cancer based on matching cases from cancer registries to occupational data issued to administrative data from Caisse nationale d'assurance vieillesse (CNAV), a public administration handling retirement pensions among employees.

A pilot study was performed using cancer data during 2010 to 2014 from two cancer registries: Lille area, and Doubs et territoire de Belfort (n=37 516). Data were matched with the CNAV database using first name, last name, sex and date and city of birth, to identify occupations of patients diagnosed with cancer. In parallel, data for 580 000 controls were selected with frequency-matched by sex, age, year and registry in CNAV databases to conduct case-controls studies.

Among cases, 36 365 incident cases (96.9%) were matched. Among occupational data, only activity sectors at the diagnosis and for the entire career can be retrieved. Profession was not available for the pilot study due to an issue in data collection for years 2009 to 2011. Among the cases, an employer identification number was filled out for 585 012 persons-year since 1972 and activity sector can be retrieved for 366 941 persons-year (67%) from this identification number. After 1999, we can clearly see an improvement in data collection as this proportion was 96% (138 551 out of 144 780 persons-year).

Two examples of possible use of this data are presented in this report. The first study presents the standardized incidence ratios (SIR) for lung cancer according to activity sector occupied at the time of diagnosis for cases aged 20 to 64 years old and recorded between 2010 and 2014 by the registry of Doubs et Territoire de Belfort. These results are highlighted with the use of indicators of expected incidence disparities due to differences in tobacco consumption by activity sector, tobacco remaining the main risk factor for lung cancer.

The second is a case-control study estimating the risk of lung cancer for the entire career for the two registries. Cases were lung cancer patients diagnosed in 2010-2014 recorded in the registries. Controls were selected at random from the files of the National Pension Fund (CNAV) among individuals living or working in the same area as the cases at the time of cancer diagnosis. Exposure was defined as being employed in a given activity sector for at least one year during entire work history. Logistic regressions were conducted separately for men and women for each activity sector, adjusting for age, year and registry.

Using CNAV data in order to determine profession and activity sectors for employees seems to be an effective way to collect occupational data that were not available in cancer registries databases, especially since 1999. The same method will be applied repeatedly to others cancer registries in France, covering nearly 20% of the population, in order to monitor the incidence of cancers by occupational activity sector and profession. The monitoring system will also make it possible to produce indicators at a departmental level for the departments fully covered by a register, including for certain overseas departments.

**KEY WORDS:** OCCUPATIONS, WORKERS, POPULATION SURVEILLANCE, INCIDENCE, CANCER, LUNG CANCER, FRANCE

## Auteur

**Annabelle Lapostolle**, Santé publique France, Direction santé environnement et travail

## Membres de l'équipe projet

**Pascal Guenel**, Centre de recherche en épidémiologie et santé des populations, Inserm  
**Julie Chesneau**, Santé publique France, Direction appui, traitements et analyses de données  
**Delphine Jezewski-Serra**, Santé publique France, Direction appui, traitements et analyses de données  
**Pascale Bernillon**, Santé publique France, Direction appui, traitements et analyses de données

## Membres du groupe de travail

**Alain Monnereau**, Registre des hémopathies malignes de la Gironde  
**Sandrine Plouvier**, Registre général des cancers de Lille et de sa région  
**Anne Sophie Woronoff**, Registre des tumeurs du Doubs, CHRU Besançon

## Remerciements

- À la Caisse nationale d'assurance vieillesse (Cnav)
- À toutes les sources de données qui contribuent à l'enregistrement des cancers par les registres :  
les laboratoires d'anatomie et de cytologie pathologiques, les départements de l'information médicale (DIM) des établissements de santé publics et privés, les échelons locaux des services de l'Assurance maladie, les médecins généralistes et spécialistes.

## Sigles et abréviations

<b>APE</b>	Activité principale exercée
<b>Cnav</b>	Caisse nationale d'assurance vieillesse
<b>CIM-O-3</b>	Classification internationale des maladies pour l'oncologie, 3 <sup>e</sup> édition
<b>DNT</b>	Déclaration nominative trimestrielle
<b>Francim</b>	France cancer incidence et mortalité, réseau français des registres des cancers
<b>InVS</b>	Institut de veille sanitaire (devenu Santé publique France en 2016)
<b>MSA</b>	Mutuelle sociale agricole
<b>NAF</b>	Nomenclature française des activités
<b>NAP</b>	Nomenclature d'activités et de produits
<b>NIR</b>	Numéro d'inscription au répertoire national d'identification des personnes physiques
<b>OR</b>	Odd ratio
<b>PCS</b>	Nomenclature des professions et catégories socioprofessionnelles
<b>RGCU</b>	Répertoire de gestion des carrières unique
<b>RIA</b>	Rapports d'incidence attendus
<b>RNIPP</b>	Répertoire national d'identification des personnes physiques
<b>RSI</b>	Régime social des indépendants
<b>SIR</b>	Ratios standardisés d'incidence ( <i>Standardized incidence ratio</i> )
<b>Siret</b>	Système Informatique pour le répertoire des entreprises sur le territoire
<b>SNGC</b>	Système national de gestion des carrières
<b>SNGI</b>	Système national de gestion des identités

# Sommaire

Résumé .....	2
Abstract .....	5
Auteur, équipe projet, groupe de travail, remerciements .....	6
Sigles et abréviations .....	6
<b>1. INTRODUCTION .....</b>	<b>8</b>
1.1 Contexte.....	8
1.2 Nécessité de développer un système pérenne de surveillance de l'incidence des cancers en lien avec l'activité professionnelle (Sicapro).....	10
1.3 Objectifs : Étudier la faisabilité d'un système de surveillance de l'incidence des cancers en lien avec activité professionnelle par la mise en place d'une étude pilote .....	11
<b>2. MÉTHODES .....</b>	<b>13</b>
2.1 Sources de données .....	13
2.2 Production des indicateurs de surveillance .....	21
<b>3. RÉSULTATS DE L'ÉTUDE PILOTE .....</b>	<b>24</b>
3.1 Les données disponibles pour la surveillance des cancers en lien avec l'activité professionnelle.....	24
3.2 Les indicateurs de surveillance .....	30
<b>4. DISCUSSION .....</b>	<b>47</b>
4.1 Performance de l'appariement et qualité des données.....	47
4.2 Étude d'incidence des cancers du poumon par secteur d'activité pour les cas du registre du Doubs et du territoire de Belfort .....	49
4.3 Étude cas témoin sur les cancers du poumon.....	50
<b>5. CONCLUSION ET PERSPECTIVES.....</b>	<b>53</b>
<b>Références bibliographiques .....</b>	<b>55</b>
<b>ANNEXE. Structure des données de la Cnav .....</b>	<b>58</b>
Base identité.....	58
Base Salaire .....	58
Base retraite .....	59
Base trimestres autre régimes .....	59
Base période assimilée.....	60
Base déclarations nominatives annuelles (DNA).....	60
Base assurance volontaire.....	60
Base allocation amiante.....	60

# 1. INTRODUCTION

## 1.1 Contexte

### 1.1.1 Épidémiologie des cancers professionnels

En France, en 2010, 12,0 % des salariés avaient été exposés à au moins une nuisance cancérigène (cancérigènes chimiques, rayonnements ionisants ou travail de nuit pour les femmes) [1]. Les estimations de la proportion de cas de cancers attribuables à des facteurs professionnels se situent entre 4 et 8,5% de l'ensemble des cancers [2, 3], représentant en France entre 15 000 et 33 000 nouveaux cas annuels [4, 5]. Toutefois, rien ne distingue sur le plan clinique un cancer ayant une origine professionnelle d'un autre cancer de même type. Néanmoins du fait de la longue période de latence entre exposition et survenue du cancer, de la méconnaissance de l'exposition, de la méconnaissance du contexte réglementaire de reconnaissance des maladies professionnelles et du caractère multifactoriel des cancers avec une forte imputabilité à des facteurs de risque comportementaux comme le tabac, il est probable qu'une forte proportion de cancers ayant des causes professionnelles ne soit pas reconnue comme telle.

### 1.1.2 La surveillance des cancers professionnels

La surveillance épidémiologique est un processus continu et systématique de recueil, d'analyse, d'interprétation et de dissémination de données descriptives permettant le suivi temporel des événements de santé dans une population [6, 7].

#### 1.1.2.1 La surveillance des cancers en lien avec l'activité professionnelle dans les pays européens

Il existe actuellement en Europe deux systèmes de surveillance des cancers en lien avec l'activité professionnelle permettant de suivre l'évolution temporelle de l'incidence des cancers par secteur d'activité : Nocca dans les pays nordiques et Occam en Italie.

Le projet Nocca pour *Nordic Occupational Cancer* se base sur l'appariement des données de registres des cancers de cinq pays nordiques (Danemark, Finlande, Islande, Norvège et Suède) avec les données professionnelles issues de recensements de la population [8]. Des cohortes ont été constituées par les personnes âgées de 30 à 64 ans au moment des recensements 1960, 1970, 1980/1981 et/ou 1990 pour lesquelles l'activité professionnelle était recueillie en texte libre. Ces individus ont été suivis jusqu'en 2005 afin de détecter la survenue d'un cancer par appariement aux bases de données de registres des cancers. Ce système de surveillance est encore actif, un article récent étudiant les risques de cancers chez les pompiers a été publié avec un suivi de l'incidence parmi les personnes des cohortes jusqu'en 2009 uniquement pour la Suède [9]. L'inconvénient de ce système de surveillance est qu'il se base sur des mesures transversales répétées de l'activité professionnelle et que les activités exercées entre deux recensements ne sont donc pas enregistrées.

Le projet Occam pour *Occupational Cancer monitoring* se base sur l'appariement des données de registres des cancers en Italie avec les bases de données de la sécurité sociale dans lesquelles les secteurs d'activité passés sont renseignés pour les personnes travaillant dans le secteur privé depuis 1974 [10]. En parallèle, des témoins sont tirés au sort dans les bases des services régionaux de santé. La population étudiée est constituée des personnes résidant en Italie âgées de 35 à 69 ans. Des matrices-emploi exposition ont été utilisées afin d'estimer les expositions professionnelles à des substances cancérigènes [11]. Les dernières

données publiées issues de ce système de surveillance concernaient les cas incidents de cancer du sein en Lombardie pour les années 2002-2009 [12].

### *1.1.2.2 La surveillance des cancers professionnels en France*

En France, la surveillance des cancers professionnels a fait l'objet dans le passé de mesures dédiées dans les différents plans cancer gouvernementaux :

- Le premier plan cancer 2003-2007, dans la mesure 13 qui préconisait de « mieux impliquer la santé au travail dans la prévention des cancers », annonçait une « systématisation de la surveillance épidémiologique des personnes exposées à des risques cancérigènes sur leur lieu de travail... ». Il prévoyait notamment de « mettre en place dans trois à cinq départements pilotes couverts par un registre général du cancer, un dispositif d'évaluation des risques liés aux expositions professionnelles, appuyé sur une coopération entre l'Institut de veille sanitaire (InVS), les entreprises, et les registres du cancer ». De plus, le contrat cadre passé le 13 février 2004 entre le ministre chargé des affaires sociales, du travail et de la solidarité d'une part, et le ministre chargé de la santé, de la famille et des personnes handicapées d'autre part, prévoyait qu'« une étude serait réalisée par l'InVS sur l'outil que pourraient représenter les registres du cancer pour la surveillance épidémiologique des cancers professionnels (étude de faisabilité sur deux registres pour deux types de cancer) ».
- Le deuxième plan cancer (2009-2013) préconisait, dans l'objectif d'améliorer le recensement des cancers d'origine professionnelle, de s'appuyer sur la coordination des bases de données existantes.
- Enfin, l'objectif 12 du troisième plan cancer (2014-2017) portait sur la prévention des cancers liés au travail ou à l'environnement, et soutenait en particulier par son action 12.4, la surveillance épidémiologique et la recherche pour « améliorer la connaissance des cancers par profession et secteur d'activité ».

Si la surveillance des cancers professionnels constitue une priorité reconnue au niveau national, le dispositif de surveillance existant actuellement en France, basé pour l'essentiel sur l'étude de la mortalité - et non de l'incidence - des cancers en fonction de la profession s'avère très insuffisant. En effet, la production d'indicateurs d'incidence des cancers par profession et par secteur d'activité au niveau national, qui seule permettrait d'identifier de façon fiable les groupes professionnels à risque accru de cancer, est actuellement irréalisable faute de bases de données médico-administratives contenant l'ensemble des informations nécessaires.

La surveillance épidémiologique des cancers d'origine professionnelle de la population française est assurée par Santé publique France à travers différents programmes. Le Programme national de surveillance des mésothéliomes (PNSM) porte uniquement sur la surveillance d'un cancer rare lié à l'exposition à l'amiante, complété depuis 2012 par la déclaration obligatoire (DO) des mésothéliomes [13]. Dans la suite de ces systèmes de surveillance, Santé publique France coordonne la mise en place d'un dispositif national de surveillance des mésothéliomes et de leurs expositions intégrant la surveillance médico-sociale (DNSM). Ce nouveau dispositif, qui sera mis en place par Santé publique France, garantira la conservation des acquis et répondra aux nouveaux enjeux, notamment au renforcement des connaissances sur les cas extrapleuraux et leurs expositions [14].

Le projet COSMOP, précédemment mis en œuvre par santé publique France, a permis quant à lui de produire des indicateurs de mortalité par cancer en fonction du secteur d'activité dans une grande cohorte de salariés constituée à partir de l'échantillon démographique permanent de l'Insee ou des déclarations annuelles des données sociales effectuées par les employeurs [15, 16]. Toutefois, la mortalité n'est pas un indicateur bien adapté pour la surveillance des

cancers à survie élevée (sein, prostate, hémopathies malignes, etc.), et reflète également des différences sociales dans l'accès aux soins ou aux programmes de dépistage du cancer, accentuant l'effet des inégalités sociales entre professions et secteurs d'activité.

## 1.2 Nécessité de développer un système pérenne de surveillance de l'incidence des cancers en lien avec l'activité professionnelle (Sicapro)

L'une des principales difficultés à la mise en place d'un système de surveillance performant réside dans le temps de latence de plusieurs années entre l'exposition à des cancérogènes sur le ou les lieu(x) de travail et l'apparition éventuelle d'un cancer, nécessitant l'utilisation de bases de données de grande taille comportant à la fois des informations sur les cas de cancer incidents et sur les professions exercées de nombreuses années auparavant.

En France, l'enregistrement continu et exhaustif des nouveaux cas de cancer survenant sur une zone géographique donnée, est réalisé par des registres des cancers, dispositif de référence<sup>1</sup>. Afin d'assurer la surveillance épidémiologique, un cadre de travail partenarial est mis en place entre le réseau français des registres des cancers Francim, le Service de Biostatistique-Bioinformatique des Hospices civils de Lyon (HCL), Santé publique France et l'Institut national du cancer (INCa).

Les registres des cancers ne recueillent pas en routine de données professionnelles et leurs données ne peuvent être utilisées isolément pour la surveillance de l'incidence des cancers par profession. La profession au moment du diagnostic du patient peut figurer dans le dossier médical parmi les renseignements généraux le concernant, mais cette information n'est renseignée que de façon inconstante et souvent peu fiable. Ce constat est également fait dans d'autres pays, comme aux États-Unis où une étude montrait que 75 % des informations professionnelles recueillies par les registres des cancers étaient inexploitable à des fins épidémiologiques [17]. Une étude pilote destinée à mettre en place un recueil complémentaire d'informations professionnelles auprès des patients inclus dans les registres généraux des cancers de l'Hérault et de la Vendée menée par Santé publique France, a conclu à l'impossibilité de baser un système de surveillance des cancers d'origine professionnelle sur le dispositif des registres tel qu'il est conçu actuellement [18]. L'utilisation des bases de données de la Caisse nationale d'assurance vieillesse (Cnav) est ainsi apparue comme le moyen le plus approprié de recueillir des informations professionnelles chez les patients atteints de cancer.

La mise en place d'un tel dispositif de surveillance des cancers en lien avec l'activité professionnelle s'impose donc comme une priorité, afin :

- d'identifier les secteurs d'activité dans lesquels l'incidence des cancers est élevée ;
- de suggérer des pistes pour les interventions de santé publique, notamment pour la prévention des cancers en lien avec l'activité professionnelle ;
- de contribuer à l'évaluation au cours du temps de l'impact des interventions publiques.

Les estimations de l'incidence des cancers par métier ou secteur d'activité, représentent des informations d'intérêt pour la description et la compréhension des relations entre facteurs de risque professionnels et apparition de cancer dans la population française. Elles permettront d'identifier les métiers ou secteurs d'activité à forte incidence de cancer, et les évolutions au

---

<sup>1</sup> Article 2, arrêté du 6 novembre 1995 : « Un registre est défini comme un recueil continu et exhaustif de données nominatives, intéressant un ou plusieurs événements de santé dans une population géographiquement définie à des fins de recherche et de santé publique, par une équipe ayant les compétences appropriées ».

cours du temps. Ces éléments permettront d'orienter les politiques publiques destinées à améliorer et évaluer la prévention et la reconnaissance des cancers d'origine professionnelle.

Les résultats obtenus par les estimations d'expositions à des facteurs de risque grâce à l'utilisation de matrices emploi-exposition et le calcul de fractions de risque qui en découlent, permettront également de générer des hypothèses de recherche, qui pourront être testées dans un second temps avec l'objectif d'établir des relations causales.

### 1.3 Objectifs : Étudier la faisabilité d'un système de surveillance de l'incidence des cancers en lien avec activité professionnelle par la mise en place d'une étude pilote

Le système envisagé de surveillance de l'incidence des cancers en lien avec l'activité professionnelle (Sicapro), dont l'objectif principal est d'identifier des groupes professionnels à risque accru de cancer, se base ainsi sur le couplage au niveau individuel, des données issues des registres des cancers avec les données sur la carrière professionnelle provenant de la Cnav. Ce système de surveillance permettra de produire des indicateurs d'incidence des cancers (tous cancers et pour certains cancers) par profession et par secteur d'activité de façon répétée au cours du temps.

Dans un premier temps, afin de tester la faisabilité de la mise en place d'un tel système de surveillance, une étude pilote a été réalisée avec deux registres volontaires (Registre général des cancers de Lille et de sa région et Registre des tumeurs du Doubs et du Territoire de Belfort).

Les objectifs de cette étude pilote sont :

- de mettre au point les modalités de croisement des données des registres avec les bases de la Cnav ;
- d'étudier la performance de l'appariement à partir des données de l'état civil des personnes atteintes de cancer issues des registres et celles figurant dans les données de la Cnav ;
- d'identifier les données d'historiques professionnels disponibles dans les bases de la Cnav et d'étudier la qualité de ces données administratives dans un but de surveillance épidémiologique ;
- de mettre au point les différents indicateurs qui pourront être construits en routine pour le système de surveillance final ;
- et d'étudier les limites de la démarche, notamment dans l'identification des expositions professionnelles.

Ce rapport de faisabilité résume les enseignements tirés de cette étude pilote. Afin de ne pas alourdir celui-ci, il a été choisi d'illustrer les résultats par une étude portant sur les cancers du poumon uniquement. Les cancers du poumon ont été choisis en raison i) de l'incidence élevée pour cette localisation et ii) de l'existence de nombreux facteurs de risques professionnels, en dehors du tabagisme (actif ou passif). Les analyses portant sur l'ensemble des localisations de cancer feront l'objet de valorisation scientifique ultérieure.

Les cancers du poumon représentent en effet une part importante de l'ensemble des cancers en France. En 2017, chez les hommes, cette localisation se plaçait en première place quant au nombre de nouveaux cas et de décès. Le taux d'incidence standardisé sur la population mondiale était de 54 pour 100 000 personnes-années et le taux de mortalité de 33 [19]. La mortalité par cancer du poumon a augmenté entre 1980 et 1990, avec des taux standardisés pour 100 000 personnes-années passant de 43,2 à 47,7 puis diminue depuis les années 90 [20].

Le cancer du poumon est une maladie multifactorielle dont le principal facteur de risque est la consommation de tabac. Néanmoins de nombreux facteurs de risque professionnels ont été documentés [21-23]. On estimait en 2012 que 13 à 29% des cancers broncho-pulmonaires étaient d'origine professionnelle [2, 11]. Le Centre international de recherche sur le cancer (CIRC) a évalué comme cancérigènes (certains ou probables pour l'homme) du cancer du poumon, un certain nombre d'agents ou de circonstances d'exposition dont les rayonnements ionisants, l'amiante, la silice cristalline, le cadmium, le chrome hexavalent, les composés du nickel, l'arsenic, les gaz d'échappement des moteurs diesel [24]. En France, plusieurs tableaux de maladies professionnelles ouvrent droit à reconnaissance du cancer du poumon comme maladie professionnelle pour de nombreuses étiologies (amiante, arsenic, rayonnement ionisant, silice, goudron, etc.). En 2014, en France, 897 cancers broncho-pulmonaires avaient été reconnus en maladie professionnelle au titre d'un tableau sur un total de 1 685 cancers soit 53,2% des cancers reconnus. Parmi ces reconnaissances, la principale étiologie était l'amiante (n=857) avec 95,5% des cas reconnus [25].

## 2. MÉTHODES

Ce projet est couvert par la décision Cnil DR-2016-498 du 30 décembre 2016 autorisant Santé publique France à mettre en œuvre un traitement de données ayant pour finalité la surveillance de l'incidence des cancers par profession et secteur d'activité (Sicapro).

### 2.1 Sources de données

#### 2.1.1 Données des registres de cancers

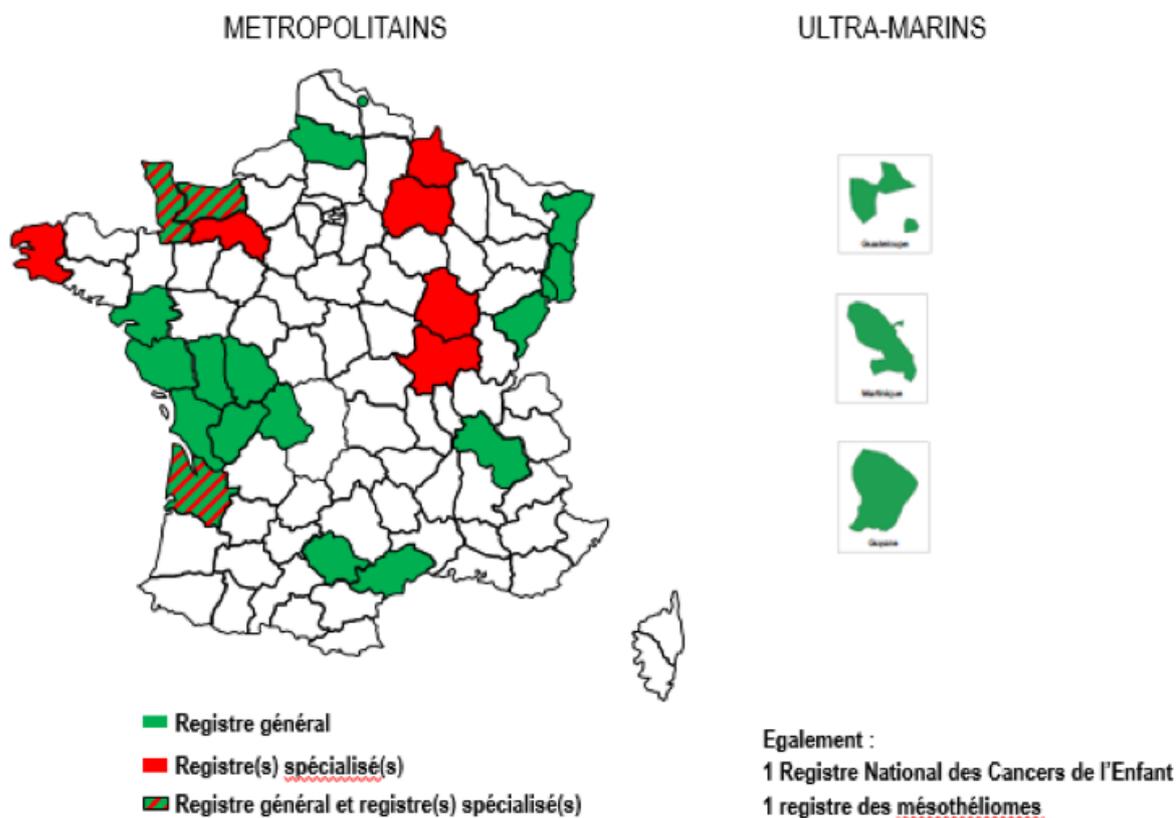
##### *2.1.1.1 Les registres des cancers en France*

Il existe en France depuis les années 1970, des registres de cancers généraux ou spécialisés qui permettent une surveillance épidémiologique locale et nationale des cancers. Actuellement il existe 14 registres généraux couvrant 18 départements, 9 registres spécialisés enregistrant les cancers pour des organes précis et couvrant 12 départements, 1 registre multicentrique du mésothéliome pleural et 5 registres généraux pour les DROM (Figure 1). En métropole leur couverture est d'environ 20% de la population.

Les registres de cancers généraux recueillent des informations sur tous les cancers alors que les registres de cancers spécialisés recueillent de nombreuses informations sur certains cancers (appareil digestif, hémopathies malignes, sein et autres cancers gynécologiques, système nerveux central, thyroïde, mésothéliome ou concernant des populations spécifiques (enfants et adolescents). Les deux registres des cancers volontaires pour la participation à cette étude pilote (Lille et sa région ; Doubs et Territoire de Belfort) sont des registres généraux, intégrant l'ensemble des localisations de cancers.

## I FIGURE 1 I

Carte des départements couverts par un registre des cancers général ou spécialisé (métropole : registres dont les données sont utilisées pour la surveillance nationale des cancers, juin 2019)



### *Les registres ayant participé à l'étude pilote*

Le registre des tumeurs du Doubs est l'un des deux registres généraux français de cancer les plus anciens. Créé en 1976, il a étendu son recueil au Territoire de Belfort en 2006, il couvre deux départements et une population de près de 700 000 habitants.

Ce registre enregistre, de manière continue et exhaustive, sans double compte, tous les nouveaux diagnostics de cancer survenus chez des personnes domiciliées dans l'un de ces deux départements, quel que soit le lieu de leur prise en charge. Les données d'incidence sont disponibles depuis 1978 dans le Doubs et depuis 2007 dans le Territoire de Belfort. Les données concernent le diagnostic, la localisation, l'histologie, les marqueurs, le stade, les traitements... Ce recueil actif est favorisé par les collaborations, nombreuses et anciennes, avec les partenaires régionaux et par l'organisation régionale de la cancérologie en Franche-Comté.

Le Registre général des cancers de Lille et de sa région est né du Plan Cancer 2003-2007 qui prévoyait notamment la création de nouveaux registres en milieu urbain pour améliorer la couverture des registres des cancers en France. Il est rattaché au groupement de coopération sanitaire « C2RC centre de référence régional en cancérologie » (union du Centre hospitalier régional Universitaire de Lille et du Centre Oscar Lambret, centre de lutte contre le cancer). Il couvre la zone de proximité de Lille qui est constituée de 101 villes de la région lilloise dans le Nord de la France et qui totalise environ 795 000 habitants. Les cancers, ainsi que certaines

tumeurs borderline et tumeurs bénignes qui surviennent chez les habitants de cette zone sont colligés au registre.

### *Les données utilisées pour l'étude pilote*

Les données extraites des registres du cancer pour cette étude sont de deux types :

- les données d'état civil (nom, prénoms, sexe, date et commune de naissance, date de décès, le cas échéant) utilisées pour effectuer le croisement avec les bases de données de la Cnav (cf. Figure 2). Seuls le sexe et la date de naissance (mois/année) ont été transmis à Santé publique France.
- les données caractérisant les cas incidents de cancer (date de diagnostic, date de décès, codes topographique et morphologique CIM-O-3).

Pour les deux registres concernés, 39 208 cas incidents de cancers correspondant à 37 516 individus ont été enregistrés entre 2010 et 2014, un individu pouvant sur la période considérée déclarer plusieurs cancers primitifs de localisations différentes (les métastases et cancers secondaires ne sont pas pris en compte).

## 2.1.2 Données de la Cnav

### *2.1.2.1 Les systèmes d'informations mis en œuvre par la Cnav*

La Cnav pilote plusieurs systèmes d'information dont le système national de gestion des identités (SNGI) et le système national de gestion des carrières (SNGC) [26].

Le Système national de gestion des identités (SNGI) comprend l'ensemble des éléments permettant l'identification des personnes : numéro d'inscription au répertoire (NIR), nom, prénom, sexe, date et lieu de naissance, date et lieu de décès. Le SNGI intègre toute personne née en France métropolitaine ou dans les DOM et enregistrée par l'Insee dans le Répertoire national d'identification des personnes physiques (RNIPP) ainsi que les ayants droit nés à l'étranger ou dans les TOM et résidant sur le territoire français dont la Cnav est chargée de réaliser l'enregistrement par délégation de l'Insee depuis 1981.

Le Système national de gestion des carrières (SNGC) contient l'ensemble des données inhérentes à la carrière des assurés afin de calculer leur retraite. Sont inclus dans le SNGC, l'ensemble des personnes ayant au moins une période d'activité donnant lieu à cotisation au régime général quelle que soit la durée. Le SNGC concerne l'ensemble des individus ayant travaillé ou travaillant en France métropolitaine ou ultramarine, y compris la population étrangère.

Le SNGC contient des données pour environ 100 millions de personnes. Les éléments de carrière les plus anciens remontent à 1930, lorsqu'ils étaient déclarés. Les données de sécurité sociale sont intégrées à partir de 1947. Pour la période antérieure à 2003, le système n'est pas complet. Les mises à jour réalisées au moment de la reconstitution de carrière pour la régularisation des retraites n'étaient en effet pas intégrées au SNGC avant 2003.

Le SNGC et le SNGI ne contiennent pas l'historique des adresses des assurés. Celle-ci est renseignée uniquement lors de l'ouverture des droits à la retraite et n'est pas historisée. Néanmoins le système est pérenne, les données étant conservées après la liquidation de la retraite et le décès.

Dans le cadre de l'étude pilote, huit bases de données anonymisées ont été transmises à Santé publique France :

- La base identité

Cette base contient les informations suivantes issues du SNGI : le sexe, la date de naissance, le statut vital de l'individu au moment des extractions par la Cnav (en 2018) et la date de décès le cas échéant.

- La base retraite

Cette base issue du SNGC contient les informations concernant la retraite du régime général : date d'effet et montant.

- La base salaire

Cette base issue du SNGC contient les informations relatives à l'activité salariée ouvrant cotisation au régime général. Une « validité » dans la base salaire correspond à une année pour laquelle la personne était salariée et cotisait au régime général.

Le SNGC a beaucoup évolué au cours du temps. Trois grandes périodes peuvent être définies :

- avant 1971, il n'y avait pas de notion d'employeurs, les salaires étaient regroupés annuellement ;
- entre 1971 et 1999, la notion d'employeur apparaît ;
- changement des bases en 1999 : apparition de nouvelles variables comme le code APE (activité principale exercée) défini par référence à la nomenclature française des activités (NAF) et la PCS (profession et catégorie socio-professionnelle), si l'employeur les fournit. Le code APE reste déclaratif. La PCS est obligatoire depuis 2008 mais cette information reste malgré tout remplie de manière parcellaire.

Le numéro d'entreprise (Siret) recueilli dans cette base permet de retrouver le secteur d'activité (NAF) de l'entreprise *via* la base Sirene® de l'Insee.

La Déclaration nominative trimestrielle (DNT) pour les salariés des particuliers (ex des gens de maison) et les Chèques emplois services universels sont également transmis à la Cnav.

- La base autre régime

Les échanges avec les autres régimes de retraite ne sont pas réguliers. Ils se produisent généralement à l'approche de la retraite (vers 53 ans). Les échanges avant cet âge sont aléatoires.

- La base période assimilée

Cette base issue du SNGC renseigne les validités pour lesquelles sont enregistrées des périodes assimilées : maladie, maternité (≥60 jours pour valider un trimestre et être dans la base), périodes de chômage (≥50 jours pour valider un trimestre et être dans la base), périodes de guerre, périodes militaires, pensions d'invalidité (≥60 %) et rentes Accident du travail.

L'alimentation des périodes assimilées dans le système se fait annuellement. Depuis 2008, les périodes de chômage et les périodes maladie apparaissent quelle que soit leur durée. Les périodes assimilées de type congé maternité ou arrêt maladie ne sont pas rattachées à un Siret car elles sont fournies par la Caisse nationale d'assurance maladie (Cnam).

Il peut y avoir des années sans validité dans l'ensemble de ces bases. L'absence d'information une année donnée peut traduire différentes situations : l'assuré peut être en activité dans un autre régime sans que l'information soit remontée pour les personnes avant la retraite ; l'assuré peut-être en recherche d'emploi sans percevoir d'indemnités ; ou il peut s'agir d'une

année d'inactivité pour l'assuré. En particulier, les périodes passées en préretraite peuvent donner lieu à une absence d'information dans les situations ne donnant pas droit à validation [27].

La structure des données extraites par la Cnav pour l'étude pilote est disponible dans l'annexe 1 en fin de ce document.

### *2.1.2.2 Détermination des secteurs d'activité*

Les codes APE n'étant renseignés dans le SNGC que depuis 1999 sans caractère obligatoire, il a été fait le choix de rechercher le secteur d'activité pour chaque validité de la base salaire en recherchant le secteur d'activité de l'entreprise dans la base Sirene®.

La base Sirene® reprend, pour chaque entreprise et chaque établissement, les informations contenues dans le répertoire Sirene. La base Sirene® rassemble ainsi des informations économiques et juridiques sur 28 millions d'établissements, dont 11 millions actifs, appartenant à tous les secteurs d'activité. Cette base est actualisée quotidiennement. La base Sirene® enregistre toutes les entreprises et leurs établissements, quelle que soit leur forme juridique et quel que soit leur secteur d'activité, situés en métropole ou dans les départements/régions d'outre-mer. Les entreprises étrangères qui ont une représentation ou une activité en France y sont également répertoriées<sup>2</sup>.

Dans la base Sirene®, les secteurs d'activité correspondants aux validités recherchées étaient codés selon différentes classifications :

- Nomenclature d'activités et de produits - NAP (en vigueur du 1<sup>er</sup> janvier 1973 au 31 décembre 1992)
- Nomenclature d'activités française - NAF 1993 (du 1<sup>er</sup> janvier 1993 au 31 décembre 2002)
- Nomenclature d'activités française - NAF rév. 1, 2003 (du 1<sup>er</sup> janvier 2003 au 31 décembre 2007)
- Nomenclature d'activités française - NAF rév. 2 en vigueur depuis le 1<sup>er</sup> janvier 2008.

Il a donc été nécessaire de procéder à un transcodage afin de disposer de l'ensemble des secteurs d'activité codés selon une même classification. L'ensemble des secteurs d'activité a donc été recodé vers la nomenclature d'activités française révision 2<sup>3</sup>.

Lors de l'application de ces tables de passage, deux cas apparaissent :

- Cas où le secteur d'activité codé selon une classification non en vigueur correspond à un seul secteur d'activité dans la NAF rév. 2 : le transcodage d'une nomenclature à l'autre est alors direct.
- Cas où le transcodage direct d'une nomenclature à l'autre n'est pas possible (à un secteur d'activité correspondent plusieurs secteurs d'activité de la NAF rév. 2) : le transcodage est réalisé à partir d'une table de passage probabiliste réalisée par l'Insee. Si par exemple 80 % des emplois d'un secteur X se retrouvent dans le secteur d'activité xx et 20 % en secteur yy en NAF rév. 2, alors la validité sera affectée avec un poids de 0,8 au secteur xx et de 0,2 au secteur yy.

---

<sup>2</sup> <https://www.sirene.fr/sirene/public/static/contenu-base-sirene>

<sup>3</sup> <https://www.insee.fr/fr/information/2579599>

### 2.1.3 Circuit des données et appariement des bases

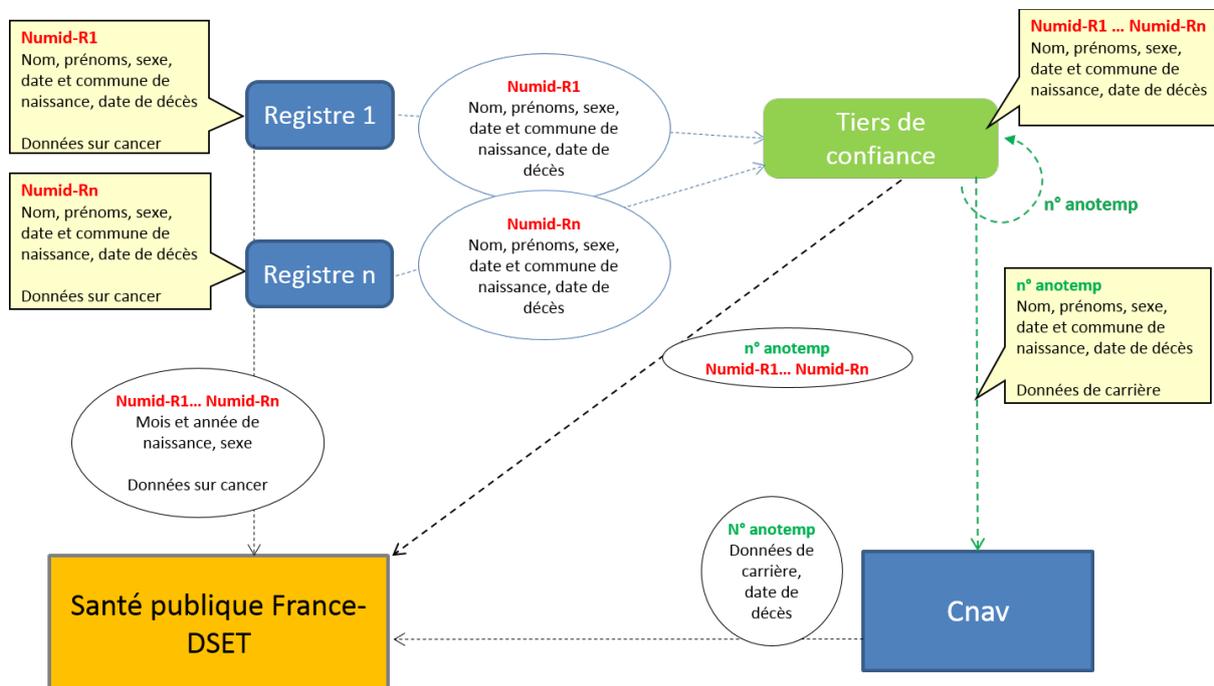
Pour ce système de surveillance, le circuit des données est organisé entre Santé publique France, les deux registres des cancers, la Cnav et un tiers de confiance (CESP : Centre d'épidémiologie et de santé publique – Inserm) de façon à constituer un fichier d'analyse ne comportant pas d'information nominative (Figure 2). Ce tiers de confiance a pour rôle de garantir la confidentialité des données tout en permettant le recueil des informations nécessaires au suivi. Il est ainsi le seul à conserver la correspondance entre les différentes données identifiantes d'un individu. Il a pour rôle la vérification de la bonne mise en forme des fichiers reçus des registres, l'établissement d'un fichier unique avec création d'une table de correspondance, et la mise en relation des identifiants avec les bases de la Cnav.

Le circuit des données comporte six étapes :

1. Les données nominatives des cas incidents de cancer, détenues par les registres et accompagnées d'un numéro d'identifiant (Numid-Rn), sont adressées au tiers de confiance.
2. Les registres adressent à Santé publique France les données sur les cas de cancers (mois et année de naissance, sexe, date de diagnostic, codes topographique et morphologique (CIM-0-3)), associées aux numéros d'identifiant des registres (Numid-Rn).
3. Le tiers de confiance crée une liste commune de sujets provenant des deux registres en lui attribuant un nouveau code de confidentialité (anotemp).
4. Le tiers de confiance transmet à la Cnav le fichier commun comportant le code de confidentialité anotemp et les données nominatives. Par ailleurs, il transmet à Santé publique France la table de correspondance entre les codes de confidentialité du fichier commun (anotemp) et les numéros d'origine dans les différentes sources (Numid-Rn).
5. L'appariement est réalisé par la Cnav. Une recherche au SNGI est réalisée sur la base des données d'état civil fournies par les registres des cancers : nom, prénom, sexe, date et lieu de naissance. Une fois l'individu identifié au SNGI, ses données de carrières sont extraites du SNGC puis transmises à Santé publique France, accompagnées de la date de décès et de leur code de confidentialité (anotemp).
6. Santé publique France, grâce à la table de correspondance, réattribue aux sujets adressés par la Cnav les numéros d'identifiants d'origine des sujets (Numid-Rn). Une fois reconstituée, la liste des sujets comportant les historiques de carrière sera appariée aux données sur les cancers, issues des registres.

## I FIGURE 2 I

### Circuit des données pour le dispositif de surveillance Sicapro



#### 2.1.4 Constitution d'un échantillon représentatif de sujets inclus dans les bases de la Cnav

La Cnav a réalisé un tirage au sort d'individus présents dans le SNGC au cours de la période d'étude afin de disposer d'une base de données de référence dans la population des départements considérés par comparaison avec la base des cas de cancer. Cette base « témoin » permettra de d'étudier l'existence d'associations entre certains secteurs d'activité et l'incidence des cancers. Le tirage au sort des témoins a été réalisé dans les bases Cnav suivantes :

- la base 'salaire' pour laquelle on dispose du département de l'employeur et qui concerne donc uniquement les individus salariés l'année de tirage au sort ;
- la base 'retraite' pour laquelle on dispose du département de résidence de la personne retraitée.

Ces deux bases ont été réunies et stratifiées par tranche d'âge quinquennale, de sexe et de département. Un pas de sondage différent par tranche d'âge a été utilisé pour le tirage au sort des témoins de façon à obtenir une distribution par âge proche de celle des cas en favorisant les classes les plus âgées.

## I TABLEAU 1 I

### Pas de sondage pour le tirage au sort des témoins pour une année donnée

Âge l'année du tirage au sort	Hommes	Femmes
Moins de 20 ans	1/100	1/100
20-24 ans	1/50	1/50
25-29 ans	1/25	1/25
30-34 ans	1/25	1/25
35-39 ans	1/20	1/20
40-44 ans	1/20	1/20
45-49 ans	1/12	1/12
50-54 ans	1/12	1/12
55-59 ans	1/10	1/10
60-64 ans	1/10	1/10
65-69 ans	1/10	1/10
70-74 ans	1/12	1/12
75-79 ans	1/12	1/12
80-84 ans	1/20	1/20
85 ans et plus	1/20	1/20

#### 2.1.5 Données de tabagisme : baromètre santé 2010

Les données de prévalence de tabagisme par secteur d'activité sont issues du Baromètre santé 2010.

Les Baromètres de Santé publique France sont des enquêtes téléphoniques dont l'objectif est de mesurer les connaissances, les attitudes, les croyances et les comportements des personnes résidant en France en matière de santé. Les baromètres santé sont des études transversales répétées avec sondages aléatoires à deux degrés (tirage d'un ménage puis d'un individu), réalisés à l'aide du système d'interview par téléphone assisté par ordinateur [28, 29]. Le Baromètre santé 2010 est la cinquième vague. Les informations sur le tabagisme sont récoltées de manière standardisée et répétée à chaque vague [30, 31]. Pour cette étude ont été utilisées les proportions de fumeurs actuels, agrégeant fumeurs quotidiens et occasionnels, d'anciens fumeurs et de non-fumeurs.

Pour le baromètre santé 2010, 27 653 individus âgés de 15 à 85 ans ont été interrogés, dont 14 835 actifs occupés. Le secteur d'activité professionnel est codé selon la classification NAF rév. 2 en 21 sections [28, 29].

## 2.2 Production des indicateurs de surveillance

### 2.2.1 Calcul des ratios standardisés d'incidence (SIR) par secteur d'activité au moment du diagnostic

La population d'étude est constituée par les cas incidents de cancer du poumon (C33-C34) diagnostiqués au cours des années 2010 à 2014, âgés de 20 à 64 ans (au moment du diagnostic), salariés au régime général et résidant dans les départements du Doubs et du Territoire de Belfort.

#### 2.2.1.1 Estimation des personnes-années de la population de référence

Le protocole de l'étude prévoyait initialement l'utilisation des données du recensement afin d'obtenir un dénominateur pour le calcul de taux d'incidence ou d'une population soumise au risque pour le calcul des SIR. Néanmoins l'utilisation des données du recensement ne permettait pas de distinguer les personnes salariées affiliées au régime général des personnes affiliées à d'autres régimes (pour lesquelles nous ne disposons pas de secteur d'activité pour les cas).

Le nombre de personnes-années par sexe et par âge a donc été reconstitué *a posteriori* à partir du tirage au sort réalisé par la Cnav dans ses bases de données en appliquant le poids correspondant au pas de sondage utilisé pour le tirage au sort. Ainsi en fonction de l'âge, chaque individu tiré au sort se voyait affecter une pondération allant de 10 chez les 55-69 ans à 50 chez les 20-24 ans.

En raison de la structure des bases de données de la Cnav, l'estimation du nombre de personnes-années n'a pas été possible pour le registre de Lille et de sa région. En effet, ce registre ne couvre pas l'ensemble du département du Nord et les données de la Cnav ne permettaient de tirage au sort que pour les départements, l'information de la commune n'étant pas disponible pour les salariés mais uniquement pour les personnes retraitées.

#### 2.2.1.2 Prise en compte des secteurs d'activité au moment du diagnostic

Pour estimer les SIR par secteur d'activité professionnelle, seule l'activité exercée au moment du diagnostic a été intégrée et seul le premier employeur de la validité considérée a été pris en compte.

Au cours d'une même année, une même personne peut avoir à la fois une activité salariée, être affiliée à un autre régime que le régime général, posséder une période assimilée et être à la retraite. Pour l'année du diagnostic, il a donc été décidé de conserver en priorité l'information de l'activité salariée. Ainsi une personne ayant travaillé une première partie de l'année et ayant été au chômage ou à la retraite en cours d'année sera comptabilisée au titre de son activité salariée. La même démarche a été appliquée au groupe de témoins tirés au sort pour déterminer l'activité exercée l'année du tirage au sort.

Une même personne pouvant avoir plusieurs validités l'année considérée (retraite, autre régime, période assimilée, etc.), l'activité salariée a été considérée en priorité puis les autres régimes, les périodes assimilées, et la retraite.

Seuls les secteurs d'activité pour lesquels l'effectif des cas était supérieur à 5 ont été conservés pour l'analyse.

### 2.2.1.3 Calcul des ratios standardisés d'incidence (SIR) par secteur d'activité

Le ratio standardisé d'incidence est le rapport du nombre de cancers incidents dans une population sur le nombre de cas attendus si l'incidence était semblable à celle observée dans la population de référence [32]. Cet indicateur standardisé permet une comparaison entre les différents secteurs d'activité.

$$SIR_{Sect\_x} = \frac{O_{Sect\_X}}{A_{Sect\_X}}$$

Avec  $SIR_{Sect\_x}$  : ratio standardisé d'incidence (Standardized incidence ratio) pour le secteur d'activité X  
 $O_{Sect\_X}$  : nombre de cas incidents de cancer du poumon observés pour le secteur d'activité X  
 $A_{Sect\_X}$  : nombre de cas attendus dans le secteur d'activité X

Les intervalles de confiance au risque 5 % des SIR ont été calculés en utilisant les formules suivantes :

$$SIR_{Inf} = \frac{O}{A} \left( 1 - \frac{1}{9 \times O} - \frac{1.96}{3\sqrt{O}} \right)^3$$

$$SIR_{Sup} = \frac{O + 1}{A} \left( 1 - \frac{1}{9(O + 1)} - \frac{1.96}{3\sqrt{(O + 1)}} \right)^3$$

Le calcul du nombre de cas attendus a été réalisé en utilisant les taux d'incidence par sexe et âge estimés pour la France à partir des données de l'ensemble des registres pour le cancer du poumon [4]. Ces taux par sexe et âge ont été appliqués aux nombres de personnes-années par sexe et âge pour chaque secteur d'activité.

## 2.2.2 Méthode de l'étude cas-témoins

L'étude cas-témoins porte sur les cas incidents de cancer du poumon pour les registres de Lille et de sa région et du Doubs et du Territoire de Belfort diagnostiqués au cours des années 2010-2014, chez des personnes salariées au régime général ou retraitées l'année du diagnostic. Se limiter à la population des individus salariés au régime général ou retraités l'année du diagnostic permet d'avoir une même définition que les témoins qui ont été tirés au sort dans cette population.

Pour les analyses portant sur le cancer du poumon, sont considérés comme témoins l'ensemble des personnes tirées au sort dans les bases de la Cnav à l'exception des cas de cancer du poumon tirés au sort comme témoins après l'année du diagnostic. Un cas de cancer du poumon est conservé comme témoin s'il est tiré au sort une année précédant son diagnostic.

Un cas incident d'une autre localisation de cancer est conservé comme témoin s'il est tiré au sort quelle que soit l'année.

### 2.2.2.1 Exposition à un secteur d'activité

Une durée de travail effectué dans un secteur d'activité a été calculée pour chaque individu en sommant l'ensemble des validités renseignées par secteur d'activité et en prenant en compte les poids attribués lors de l'étape de transcodage (cf. partie 2.1.2).

À partir de l'indicateur de durée dans un secteur d'activité, une variable dichotomique par secteur d'activité a été créée : sont considérées comme ayant été exposées à un secteur d'activité toutes les personnes ayant une durée cumulée d'au moins 12 mois dans ce secteur.

Une personne peut ainsi avoir été exposée à plusieurs secteurs d'activité au cours de sa carrière professionnelle.

### 2.2.2.2 Régressions logistiques

Des régressions logistiques ont été réalisées pour chaque secteur d'activité indépendamment en prenant comme référence le fait de ne pas avoir travaillé dans ce secteur d'activité ou d'y avoir travaillé moins d'un an. Les analyses ont été stratifiées par sexe en ajustant sur la classe d'âge, l'année et le département [32].

### 2.2.2.3 Calcul de rapports d'incidence attendus associés à la consommation différentielle de tabac

La méthode de calcul des rapports d'incidence attendus (RIA) a été développée pour prendre en compte la consommation de tabac lorsque cette donnée n'était pas renseignée individuellement dans les données utilisées [33, 34] :

$$RIA_{\text{tabac}} = \frac{P_{F\text{sect}X} \times RR_F + P_{ExF\text{sect}X} \times RR_{ExF} + P_{NF\text{sect}X}}{P_{F\text{sect\_non}X} \times RR_F + P_{ExF\text{sect\_non}X} \times RR_{ExF} + P_{NF\text{sect\_non}X}}$$

Avec :  $RIA_{\text{tabac}}$  = Rapports d'incidence attendus associés à la consommation différentielle de tabac

$P_{F\text{sect}X}$  = Proportion de fumeurs dans un secteur d'activité X

$P_{ExF\text{sect}X}$  = Proportion d'ex-fumeurs dans un secteur d'activité X

$P_{NF\text{sect}X}$  = Proportion de non-fumeurs dans un secteur d'activité X

$P_{F\text{sect}}$  = Proportion de fumeurs pour l'ensemble des secteurs d'activité sauf le secteur X

$P_{ExF\text{sect}}$  = Proportion d'ex-fumeurs pour l'ensemble des secteurs d'activité sauf le secteur X

$P_{NF\text{sect}}$  = Proportion de non-fumeurs pour l'ensemble des secteurs d'activité sauf le secteur X

$RR_F$  = RR d'incidence de cancer du poumon chez les fumeurs par rapport aux non-fumeurs

$RR_{ExF}$  = RR d'incidence de cancer du poumon chez les ex-fumeurs par rapport aux non-fumeurs

Les risques relatifs (RR) utilisés dans le calcul sont issus de la littérature [35]. Le RR d'incidence de cancer de poumon pour les hommes est de 23,6 chez les fumeurs et de 7,5 chez les ex-fumeurs. Pour les femmes ces RR sont respectivement de 7,8 et 2,8.

## 3. RÉSULTATS DE L'ÉTUDE PILOTE

### 3.1 Les données disponibles pour la surveillance des cancers en lien avec l'activité professionnelle

#### 3.1.1 Résultats de l'appariement entre les cas de cancers et les données de la Cnav

Pour les deux registres volontaires de Lille et de sa région et du Doubs et du Territoire de Belfort, 39 208 cas de cancers (toutes localisations) correspondant à 37 516 individus ont été diagnostiqués au cours des années 2010 à 2014, un individu pouvant sur la période considérée déclarer plusieurs cancers primitifs de localisations différentes.

Les données identifiantes de ces 37 516 individus présents dans les deux registres des cancers ont été transmises à la Cnav afin de procéder à l'appariement après la recherche préalable des données d'état civil (nom, prénom, date et commune de naissance) au SNGI. Lors de cette opération, 36 365 individus ont été retrouvés au SNGI (96,9%) et 1 151 individus n'ont pas pu être identifiés sur les 37 516 cas de cancer des registres (3,1%).

#### I TABLEAU 2 I

##### Résultats de recherche au RNIPP pour les individus perdus de vue

	N	%
Non recherchés au RNIPP	496	43,1%
Demande identifiée sans divergence d'état civil	14	1,2%
Demande identifiée, avec divergence(s) d'état civil (nom, prénom, sexe, date et/ou lieu de naissance)	170	14,8%
Demande identifiée sur le nom d'usage avec ou sans divergence d'état civil	227	19,7%
Non identifiée au RNIPP	244	21,2%
Total	1 151	100%

Les données d'état civil des cas enregistrés dans les registres des cancers sont régulièrement vérifiées et corrigées afin de permettre la recherche du statut vital auprès du Répertoire national d'identification des personnes physiques (RNIPP). Cette recherche ne concerne donc pas les personnes pour lesquelles les registres ont déjà l'information du décès. Parmi les individus non retrouvés lors de l'appariement avec les données de la Cnav, seulement 14 individus avaient été identifiés au RNIPP sans divergence d'état civil. Une majorité des cas non retrouvés au SNGI n'avait pas fait l'objet d'une recherche au RNIPP en amont (43,1%) parmi lesquels 82% (n=407) sont des personnes décédées.

On note une forte différence de répartition par sexe parmi les personnes non retrouvées au SNGI : 29% sont des hommes (n= 339) et 71% des femmes (n=812) alors que parmi l'ensemble des cas ces proportions sont respectivement de 54% et 46%. Cette différence peut être expliquée par l'enregistrement pour certains cas du nom patronymique des femmes mariées à la place du nom de naissance.

### 3.1.2 Données des registres

Sur les 36 365 cas de cancers enregistrés pour les années 2010-2014 et pour lesquels l'appariement avec les données de la Cnav a pu être réalisé, 18 372 avaient été enregistrés par le registre de Lille et sa région et 17 993 par le registre du Doubs et Territoire de Belfort. Les cas de cancers sont répartis de manière homogène selon les années de diagnostic pour les deux registres (Tableau 3).

**I TABLEAU 3 I**

**Répartition des cas de cancers par année de diagnostic selon le registre pour l'ensemble des localisations, années 2010-2014**

Année	Lille et sa région		Doubs et Territoire de Belfort	
	N	%	N	%
2010	3 685	20,1%	3 645	20,3%
2011	3 688	20,1%	3 659	20,3%
2012	3 648	19,9%	3 576	19,9%
2013	3 641	19,8%	3 635	20,2%
2014	3 710	20,2%	3 478	19,3%
Total	18 372	100,0%	17 993	100,0%

Les cas de cancers pour le registre du Doubs et Territoire de Belfort sont plus souvent des hommes (56,3% contre 53,5% pour le registre de Lille et sa région) et sont légèrement plus âgés avec une moyenne d'âge de 66,9 ans (sd=14,8) contre 65,0 ans (sd=15,4) pour le registre de Lille et sa région et des proportions plus élevées de cas de cancers âgés de plus de 65 ans (Tableau 4 et Tableau 5)

**I TABLEAU 4 I**

**Répartition des cas de cancers par sexe selon le registre pour l'ensemble des localisations, années 2010-2014**

Sexe	Lille et sa région		Doubs et Territoire de Belfort	
	N	%	N	%
Hommes	9 837	53,5%	10 128	56,3%
Femmes	8 535	46,5%	7 865	43,7%
Total	18 372	100,0%	17 993	100,0%

## I TABLEAU 5 I

### Répartition des cas de cancers par classe d'âges selon le registre pour l'ensemble des localisations, années 2010-2014

Classe d'âges	Lille et sa région		Doubs et Territoire de Belfort	
	N	%	N	%
0-19 ans	172	0,9%	125	0,7%
20-34 ans	572	3,1%	393	2,2%
35-39 ans	311	1,7%	283	1,6%
40-44 ans	576	3,1%	461	2,6%
45-49 ans	942	5,1%	741	4,1%
50-54 ans	1377	7,5%	1205	6,7%
55-59 ans	2023	11,0%	1734	9,6%
60-64 ans	2632	14,3%	2328	12,9%
65-74 ans	4349	23,7%	4731	26,3%
75 ans et plus	5418	29,5%	5992	33,3%
Total	18 372	100,0%	17 993	100,0%

Pour les années 2010-2014, 2 150 cancers du poumon ont été enregistrés : 2110 pour le registre de Lille et sa région et 2040 pour le registre du Doubs et Territoire de Belfort. La répartition des cas par localisation selon le registre est détaillée dans le Tableau 6 ci-dessous.

## I TABLEAU 6 I

### Répartition des cas de cancers par localisation selon le registre pour l'ensemble des localisations, années 2010-2014

Localisation	CIM-O-3	Lille et sa région		Doubs et Territoire de Belfort	
		N	%	N	%
Lèvre-Bouche-Pharynx	C00-C14	916	5,0%	632	3,5%
Œsophage	C15	369	2,0%	167	0,9%
Estomac	C16	310	1,7%	355	2,0%
Colon-rectum	C18-C21	2 100	11,4%	1 913	10,6%
Foie	C22	503	2,7%	473	2,6%
Thyroïde	C73	366	2,0%	306	1,7%
Larynx	C32	211	1,1%	158	0,9%
Poumon	C33-C34	2 110	11,5%	2 040	11,3%
Sein	C50	3 225	17,6%	2 591	14,4%
Ovaire <sup>2</sup>	C56, C570, C571, C572, C573, C574	270	1,5%	235	1,3%
Corps de l'utérus	C54	344	1,9%	377	2,1%
Col de l'utérus	C53	188	1,0%	157	0,9%
Prostate	C61	2 202	12,0%	2 892	16,1%
Testicule	C62	151	0,8%	154	0,9%
Rein	C64-C66, C68	557	3,0%	545	3,0%
Vessie	C67	578	3,1%	533	3,0%
Système nerveux central <sup>3</sup>	C70-C72	239	1,3%	314	1,7%
Pancréas	C25	513	2,8%	610	3,4%
Mélanome de la peau	C44 <sup>1</sup>	78	0,4%	69	0,4%
Autres localisations		3 142	17,1%	3 472	19,3%
<b>Total</b>		<b>18 372</b>	<b>100,0%</b>	<b>17 993</b>	<b>100,0%</b>

<sup>1</sup> Morphologies 97203 à 97803

<sup>2</sup> Toutes les morphologies sauf 84423, 84513, 84613, 84623, 84723 et 84733

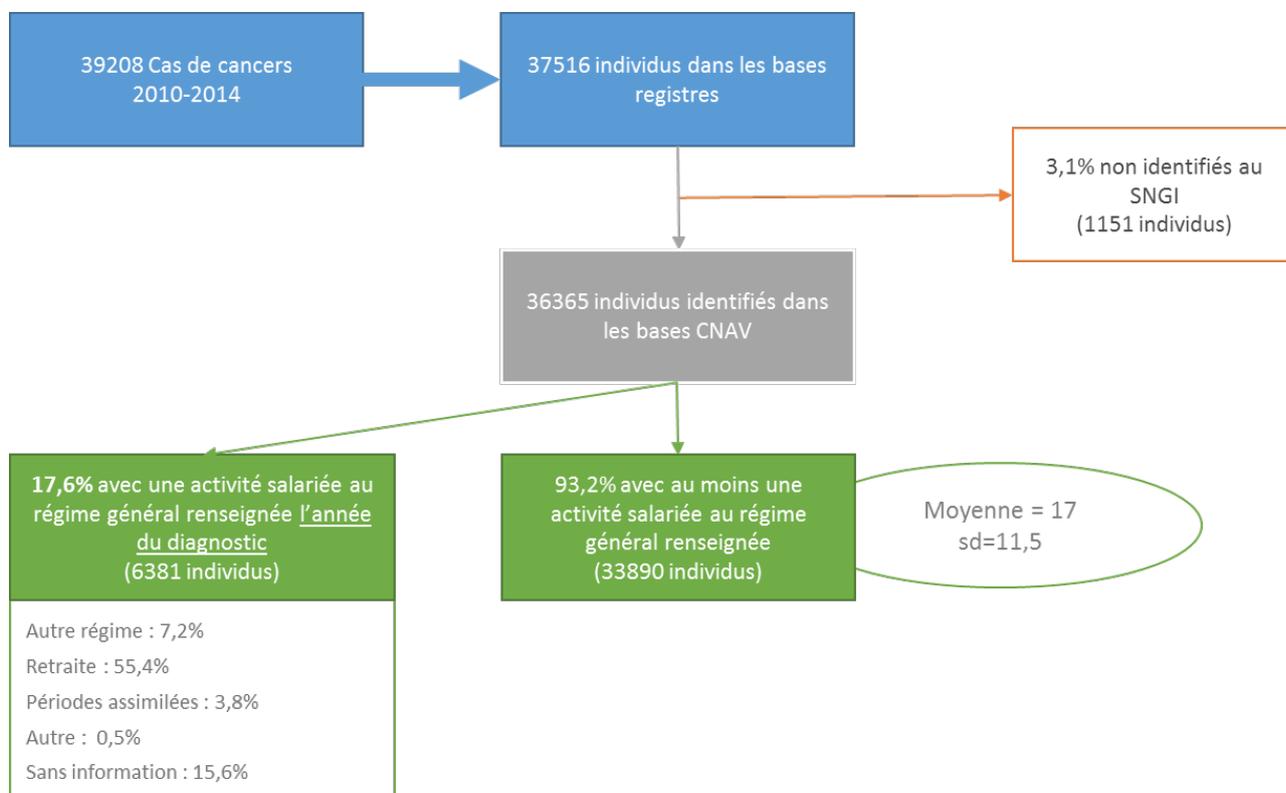
<sup>3</sup> Morphologies de 9590 à 9597, de 9670 à 9719, de 9724 à 9729 et de 9832 à 9838

### 3.1.3 Description des données d'historique professionnel disponibles

Sur les 36 365 cas de cancers (toutes localisations) appariés aux données de la Cnav, des données de validité correspondant à une activité salariée au régime général ont été retrouvées pour 33 890 individus soit 93% des cas de cancer avec une moyenne de 17 validités (sd=11,5) par individu (Figure 3). Par ailleurs, l'année du diagnostic, 6 381 individus avaient une validité en tant que salarié du régime général (17,6%), 2 608 au titre d'un autre régime (7,2%), 20 138 étaient à la retraite (55,4%), 1 398 avaient une période assimilée (3,8%) comprenant entre autres chômage et arrêts maladie, et 5 671 n'avaient aucune information l'année concernée (15,6%). Parmi ces personnes pour lesquelles aucune information n'était renseignée l'année du diagnostic, il peut s'agir de personnes sans activité pour l'année considérée, en situation de pré-retraite dans un cadre n'ouvrant pas droits à cotisation ou affiliées à un autre régime que le régime général.

### I FIGURE 3 I

#### Résultats des appariements entre les bases de données de carrières disponibles pour les cas de cancers



Les numéros Siret des entreprises, permettant de retrouver le secteur d'activité sont renseignés dans le SNGC à partir de 1972. Avant cette date, les validités pour les personnes salariées au régime général sont présentes mais sans information sur l'employeur. On peut voir dans le Tableau 7 que la présence des numéros Siret des employeurs est quasi nulle avant cette année avec seulement 4% de Siret renseignés avant 1972. La complétude des données s'améliore fortement puisque cette proportion est de plus de 90% sur les années suivantes. Au total pour les 33 890 cas de cancers avec au moins une activité salariée, nous disposons de l'information du Siret pour 69% des validités de l'ensemble des carrières.

## I TABLEAU 7 I

### Nombre et proportion de données employeur complétées dans la base salaire selon les périodes de validité pour les cas de cancers appariés

Années	Nombre de validités	Nombre de Siret renseignés		Nombre de codes APE renseignés parmi les Siret renseignés	
		N	%	N	%
Avant 1972	239 287	9 680	4,0%	9	0,1%
1972-2009	569 215	536 409	94,2%	128 980	24,0%
2010-2014	30 688	28 628	93,3%	24 055	84,0%
2014-2018	11 452	10 517	91,8%	10 481	99,7%
Total	850 642	585 234	68,8%	163 525	19,2%

On note une forte amélioration de la déclaration du code APE par les employeurs caractérisant l'activité principale par référence à la nomenclature d'activités française (NAF rév. 2) avec le temps : celui-ci était déclaré pour 24% des validités sur la période 1972-2009 contre 99,7% des validités en 2014-2018. Au total l'information était disponible pour 19,2% des validités de notre population d'étude, nécessitant de rechercher les secteurs d'activité pour les entreprises avec un Siret renseigné dans les bases Sirene de l'Insee.

#### Résultats de la recherche des secteurs d'activité pour les validités au régime général

La recherche du secteur d'activité pour les validités dans la base Sirene était uniquement possible pour les codes Siret correctement renseignés dans le SNGC. Parmi 585 234 validités pour lesquelles on disposait d'une donnée dans la variable Siret de l'employeur, on avait sur l'ensemble de la période, 194 959 codes Siret ne correspondant pas à un code valide soit 33%.

Si l'on regarde uniquement sur la période après 1999 qui correspond à la dernière modification du SNGC, 96% des codes Siret correspondent à une nomenclature reconnue et ont permis de retrouver un secteur d'activité dans la base Sirene.

## I TABLEAU 8 I

### Répartition des validités en fonction de la nomenclature utilisée

Nomenclature du secteur d'activité	1972-1998		1999-2018	
	N	%	N	%
NAF rev 2	97 789	22%	120 065	83%
NAF rev 1	24 591	6%	13 980	10%
NAF 93	106 010	24%	4 506	3%
NAP	23 112	5%	0	0%
Siret non retrouvé	188 730	43%	6 229	4%
Total	440 232		144 780	

Au total, le transcodage a été réalisé de la NAF révision 1 vers la NAF rev. 2 pour 38 571 validités, de la NAF93 vers la NAF rev. 2 pour 110 516 validités, et de la NAP vers la NAF rev. 2 pour 23 112 validités.

Les validités avec un Siret renseigné pour l'année du diagnostic, utilisées dans l'étude d'incidence (ou du tirage au sort pour les témoins) étaient toutes reconnues et codées en NAF rev. 2, elles n'ont donc pas fait l'objet d'un transcodage.

## 3.2 Les indicateurs de surveillance

### 3.2.1 Incidence des cancers du poumon par secteur d'activité pour les cas incidents du registre du Doubs et du Territoire de Belfort

En raison de l'impossibilité d'obtenir une population de personnes soumises au risque pour le calcul des ratios standardisés d'incidence pour le registre de Lille et de sa région, les résultats présentés ici concernent uniquement les cas incidents du registre du Doubs et du Territoire de Belfort.

#### 3.2.1.1 Description des cas de cancer du poumon pour les calculs de SIR

Parmi les cas de cancer du poumon (n= 862) âgées de 20 à 64 ans, 276 personnes, soit 32%, avaient une activité salariée au régime général l'année de leur diagnostic (Tableau 9). Ce pourcentage était de 89,7% parmi la population estimée pour le territoire. Cette différence dans les activités s'explique par la méthode de tirage au sort des témoins, ceux-ci étant tirés au sort uniquement parmi les personnes salariées ou retraitées.

**I TABLEAU 9 I**

**Répartition des cas de cancer du poumon et de la population estimée par type de validité pour la classe d'âge 20-64 ans, pour les années 2010-2014, Doubs et Territoire de Belfort**

Activité	Cas de cancers du poumon		Population estimée Doubs et Territoire de Belfort	
	Personnes-années	%	Personnes-années	%
Salariés affiliés au régime général	276	32,0%	1 157 167	89,7%
Autres régimes	91	10,6%	6 780	0,5%
Période assimilée	108	12,5%	9 132	0,7%
Autre situation	8	0,9%	419	0,0%
Retraités	195	22,6%	113 040	8,8%
Aucune information pour l'année concernée	184	21,3%	2 846	0,2%
	862	100,0%	1 289 384	100,0%

Les cas pour lesquels aucune information n'a été retrouvée dans les bases de la Cnav (21% des cas) peuvent correspondre à des personnes inactives l'année du diagnostic ou des personnes affiliées à un autre régime et pour lesquelles les informations ne sont pas encore remontées à la Cnav.

La répartition des cas comme celle de la population estimée est répartie équitablement selon les années concernées (Tableau 10).

## I TABLEAU 10 I

Répartition des cas de cancer du poumon et de la population estimée selon l'année de diagnostic pour les cas et l'année de tirage au sort pour les témoins pour la classe d'âge 20-64 ans, années 2010-2014, Doubs et Territoire de Belfort

Année	Cas de cancer du poumon		Population estimée Doubs et Territoire de Belfort	
	Personnes-années	%	Personnes-années	%
2010	184	21,3%	262 822	20,4%
2011	169	19,6%	263 919	20,5%
2012	170	19,7%	258 011	20,0%
2013	179	20,8%	254 200	19,7%
2014	160	18,6%	250 432	19,4%
Total	862	100,0%	1 289 384	100,0%

Chez les cas de cancer du poumon, la proportion d'hommes est plus importante que la proportion de femmes (70% d'hommes). Parmi les témoins tirés au sort dans les bases de la Cnav, 51,7% des personnes tirées au sort sont des femmes et 48,3% des hommes (Tableau 11).

## I TABLEAU 11 I

Répartition des cas de cancer du poumon et de la population estimée selon le sexe pour la classe d'âge 20-64 ans, années 2010-2014, Doubs et Territoire de Belfort

Sexe	Cas de cancer du poumon		Population estimée Doubs et Territoire de Belfort	
	Personnes-années	%	Personnes-années	%
Hommes	603	70,0%	666 544	51,7%
Femmes	259	30,0%	622 840	48,3%
Total	862	100,0%	1 289 384	100,0%

Les cas de cancer du poumon sont plus âgés que les personnes tirées au sort dans les bases de la Cnav. Les 20-34 ans représentent moins de 1% des cas de cancers du poumon contre 35,2% pour les témoins et 38 % des individus ont entre 60 et 64 ans contre 13% dans la population estimée (Tableau 12).

Parmi les individus salariés au régime général, la répartition par secteur d'activité entre cas de cancers du poumon et population estimée est globalement semblable (Tableau 13). On note une proportion légèrement plus élevée de cas dans l'industrie manufacturière par rapport à la population estimée (30,7% des cas vs. 23,3% de la population estimée) ainsi que dans le secteur de la construction (10,3% des cas vs. 6,2% de la population estimée).

## I TABLEAU 12 I

Répartition des cas de cancer du poumon et des témoins selon l'âge pour la classe d'âge 20-64 ans, années 2010-2014, Doubs et Territoire de Belfort

Classe d'âges	Cas de cancer du poumon		Population estimée Doubs et Territoire de Belfort	
	Personnes-années	%	Personnes-années	%
20-34 ans	6	0,7%	454 250	35,2%
35-39 ans	15	1,7%	131 820	10,2%
40-44 ans	28	3,2%	141 680	11,0%
45-49 ans	63	7,3%	142 656	11,1%
50-54 ans	155	18,0%	131 448	10,2%
55-59 ans	266	30,9%	119 130	9,2%
60-64 ans	329	38,2%	168 400	13,1%
Total	862	100,0%	1 289 384	100,0%

## I TABLEAU 13 I

Répartition des cas de cancer du poumon et des témoins en fonction du secteur d'activité au moment du diagnostic pour les individus salariés âgés de 20-64 ans, années 2010-2014, Doubs et Territoire de Belfort

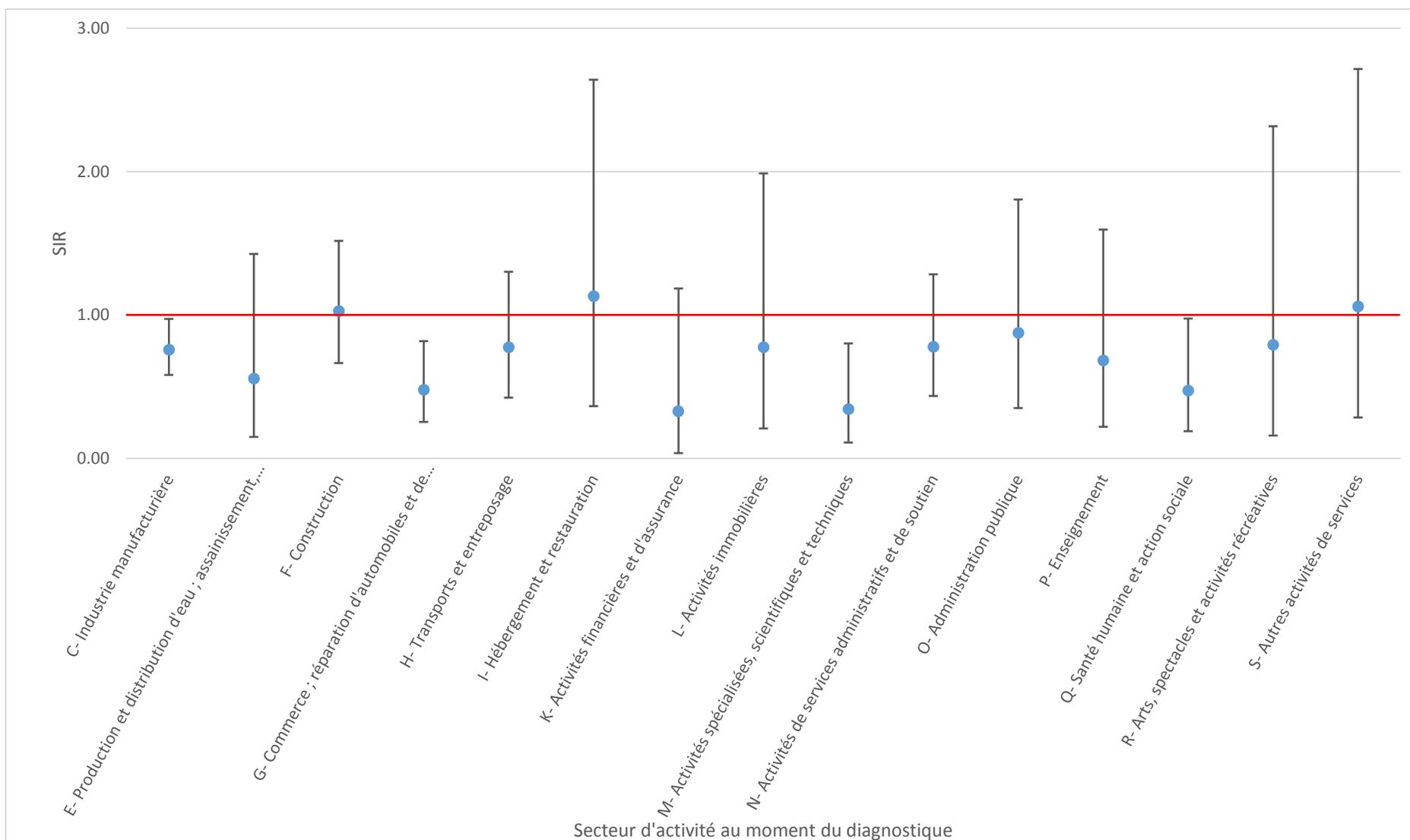
Secteur d'activité	Cas de cancer du poumon		Population estimée Doubs et Territoire de Belfort	
	Personnes-années	%	Personnes-années	%
C-Industrie manufacturière	80	30,7%	248 166	23,3%
E-Production et distribution d'eau ; assainissement, gestion des déchets et dépollution	5	1,9%	5 050	0,5%
F-Construction	27	10,3%	65 696	6,2%
G-Commerce ; réparation d'automobiles et de motocycles	17	6,5%	150 091	14,1%
H-Transports et entreposage	18	6,9%	44 268	4,2%
I-Hébergement et restauration	10	3,8%	43 453	4,1%
K-Activités financières et d'assurance	6	2,3%	28 059	2,6%
L-Activités immobilières	5	1,9%	16 759	1,6%
M-Activités spécialisées, scientifiques et techniques	9	3,4%	51 565	4,8%
N-Activités de services administratifs et de soutien	24	9,2%	140 744	13,2%
O-Administration publique	18	6,9%	59 591	5,6%
P-Enseignement	8	3,1%	27 086	2,5%
Q-Santé humaine et action sociale	25	9,6%	138 074	13,0%
R-Arts, spectacles et activités récréatives	4	1,5%	18 885	1,8%
S-Autres activités de services	5	1,9%	28 310	2,7%
Total	261	100,0%	1 065 797	100,0%

### *3.2.1.2 Ratios d'incidences standardisés en fonction du secteur d'activité au moment du diagnostic*

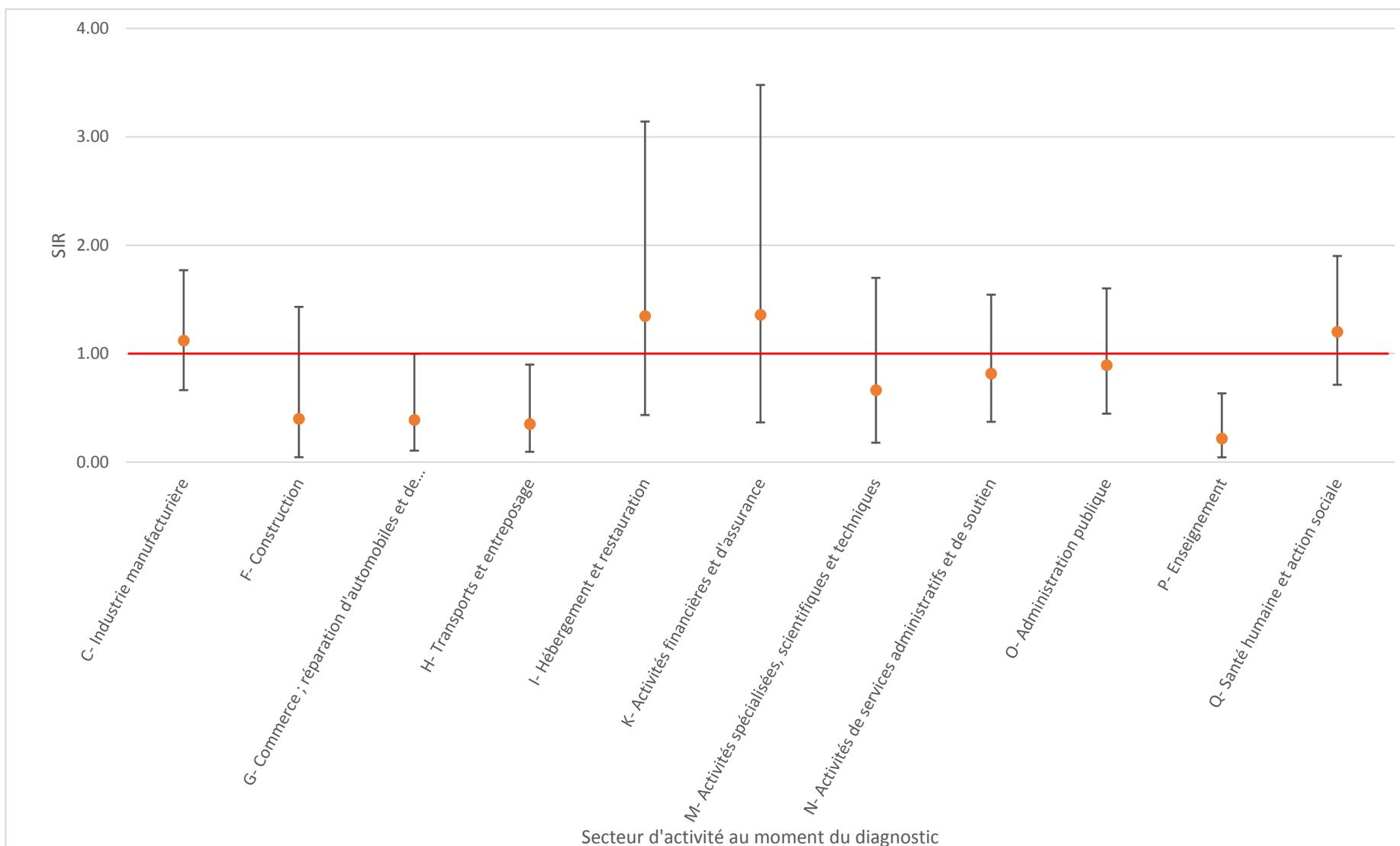
Les SIR par secteur d'activité pour les hommes salariés affiliés au régime général sont présentés dans la Figure 4 ci-dessous. Sur l'ensemble des secteurs d'activité étudiés, seuls trois avaient un ratio d'incidence standardisé très légèrement supérieur à 1 mais non significatif. Il s'agit des secteurs de la construction (SIR=1,03 [0,66-1,52]), de l'hébergement et de la restauration (SIR=1,13 [0,36-2,64]), et des autres activités de service (SIR=1,06 [0,29-2,72]).

Les secteurs d'activité pour lesquels on observe un nombre de cancers incidents observés significativement inférieur à l'attendu sont les secteurs de l'industrie manufacturière (SIR=0,76 [0,58-0,97]), du commerce, de la réparation d'automobiles et de motocycles (SIR=0,48 [0,25-0,82]), des activités spécialisées, scientifiques et techniques (SIR=0,34 [0,11-0,80]) et de la santé humaine et de l'action sociale (SIR=0,47 [0,19-0,98]).

Chez les femmes comme chez les hommes, on retrouve un SIR supérieur à 1, non significatif, pour le secteur de l'hébergement et de la restauration (SIR=1,35 [0,43-3,14]) et un SIR significativement inférieur à 1 pour le secteur du commerce, de la réparation d'automobiles et de motocycles (SIR=0,39 [0,10-1,00]) (Figure 5). Chez les femmes le nombre de cancers du poumon incidents observés est également significativement inférieur à l'attendu pour le secteur de l'enseignement (SIR=0,22 [0,04-0,63]).



**I FIGURE 4 | Ratio standardisés d'incidence du cancer du poumon chez les hommes âgés de 18 à 64 ans selon le secteur d'activité (2010-2014), Doubs et Territoire de Belfort**



**I FIGURE 5 | Ratio standardisés d'incidence du cancer du poumon chez les femmes âgées de 18 à 64 ans selon le secteur d'activité (2010-2014), Doubs et Territoire de Belfort**

## 3.2.2 Étude cas-témoins sur les cancers du poumon

### 3.2.2.1 Prévalence du tabagisme par secteurs d'activité (Baromètre santé 2010)

En France métropolitaine, en 2010, les prévalences de tabagisme actuel ou passé varient fortement selon les secteurs d'activité (tableaux 14 et 15).

Chez les hommes, la prévalence de tabagisme actuel est comprise entre 32% pour le secteur d'activité de l'information et de la communication et 55% pour le secteur des activités de services administratifs et de soutien. Cette prévalence était également élevée avec la moitié des salariés déclarant être fumeurs au moment de l'enquête pour les trois secteurs de la production et de la distribution d'eau, de l'assainissement, de la gestion des déchets et dépollution, de la construction et des activités immobilières.

C'est dans les secteurs de la production et de la distribution d'eau, de l'assainissement, de la gestion des déchets et dépollution, de l'administration publique, de l'enseignement et des autres activités de service que l'on trouve les proportions les plus élevées de non-fumeurs vie entière (35%).

Chez les femmes, le secteur de la production et de la distribution d'eau, de l'assainissement, de la gestion des déchets et dépollution se détache avec à la fois une forte proportion de fumeuses actuelles (56%) et une très faible proportion de non-fumeuses vie entière (15%). *A contrario* certains secteurs d'activité se distinguent par une faible proportion de fumeuses actuelles et une forte proportion de fumeuses vie entière : les activités financières et d'assurance, les activités de services administratifs et de soutien, l'administration publique, l'enseignement, les arts, spectacles et activités récréatives.

## I TABLEAU 14 I

### Prévalence de tabagisme parmi les hommes en activité selon le secteur d'activité, France métropolitaine, Baromètre santé 2010

NAF	Fumeurs actuels		Anciens fumeurs		Non-fumeurs		Total
	N	%	N	%	N	%	
C-Industrie manufacturière	579	41%	459	33%	363	26%	1 402
E-Production et distribution d'eau ; assainissement, gestion des déchets et dépollution	37	50%	11	15%	26	35%	74
F-Construction	469	50%	231	25%	241	26%	941
G-Commerce ; réparation d'automobiles et de motocycles	314	45%	199	29%	184	26%	697
H-Transports et entreposage	229	39%	193	33%	159	27%	581
I-Hébergement et restauration	111	48%	55	24%	66	29%	231
J-Information et communication	90	32%	103	36%	91	32%	283
K-Activités financières et d'assurance	81	37%	76	35%	62	28%	219
L-Activités immobilières	29	50%	17	29%	12	21%	57
M-Activités spécialisées, scientifiques et techniques	71	39%	54	30%	56	31%	181
N-Activités de services administratifs et de soutien	111	55%	47	23%	45	22%	203
O-Administration publique	202	34%	187	31%	210	35%	599
P-Enseignement	107	35%	91	30%	108	35%	306
Q-Santé humaine et action sociale	132	36%	133	36%	102	28%	366
R-Arts, spectacles et activités récréatives	51	42%	32	26%	38	32%	121
S-Autres activités de services	32	41%	19	25%	27	35%	79

## I TABLEAU 15 I

### Prévalence de tabagisme parmi les femmes en activité selon le secteur d'activité, France métropolitaine, Baromètre santé 2010

NAF	Fumeurs actuels		Anciens fumeurs		Non fumeurs		Total
	N	%	N	%	N	%	
C-Industrie manufacturière	193	36%	152	28%	195	36%	540
E-Production et distribution d'eau ; assainissement, gestion des déchets et dépollution	10	56%	5	28%	3	15%	18
F-Construction	38	37%	35	34%	31	30%	104
G-Commerce ; réparation d'automobiles et de motocycles	333	44%	179	23%	249	33%	761
H-Transports et entreposage	65	37%	49	27%	64	36%	178
I-Hébergement et restauration	102	47%	41	19%	73	34%	216
J-Information et communication	45	40%	27	25%	39	35%	112
K-Activités financières et d'assurance	56	27%	70	34%	81	39%	207
L-Activités immobilières	32	39%	27	33%	23	28%	82
M-Activités spécialisées, scientifiques et techniques	80	35%	71	31%	79	34%	229
N-Activités de services administratifs et de soutien	62	31%	56	27%	85	42%	203
O-Administration publique	142	31%	138	30%	174	38%	454
P-Enseignement	182	27%	204	31%	278	42%	664
Q-Santé humaine et action sociale	574	34%	496	29%	625	37%	1 695
R-Arts, spectacles et activités récréatives	23	25%	29	31%	41	45%	93
S-Autres activités de services	53	39%	39	29%	44	32%	136

### *3.2.2.2 Mesures d'associations entre la survenue d'un cancer du poumon et les secteurs d'activité occupés au cours de la carrière*

Chez les hommes (Tableau 16), après ajustement sur la classe d'âge, l'année et le département, on observe un risque d'incidence du cancer du poumon significativement augmenté pour les personnes ayant travaillé au moins une année dans les secteurs d'activité de la production et distribution d'électricité, de gaz, de vapeur et d'air conditionné (ORa=1,82 [1,10-3,00]), de la production et distribution d'eau ; assainissement, la gestion des déchets et la dépollution (ORa=1,64 [1,20-2,26]), de la construction (ORa=1,17 [1,04-1,33]), du commerce, de réparation d'automobiles et motocycles (ORa=1,18 [1,05-1,32]) de l'hébergement et de la restauration (ORa=1,70 [1,37-2,11]), des activités immobilières (ORa=1,21 [1,06-1,38]) et des activités de services administratifs et de soutien (ORa=1,21 [1,06-1,37]).

Pour le secteur d'activité des industries manufacturières, on observe un odd-ratio significativement inférieur à 1 chez les hommes (ORa=0,90 [0,83-0,98]).

Chez les femmes (Tableau 17), les Odd ratios de cancer du poumon associés au fait d'avoir travaillé dans les secteurs d'activité de la production et distribution d'électricité et de la production et distribution d'eau n'ont pas pu être calculés du fait d'effectifs trop faibles.

Chez les femmes, seules les personnes ayant travaillé plus d'un an dans le secteur des activités financières et d'assurance (ORa=1,40 [1,01-1,93]) ont un risque de cancer du poumon significativement supérieur à celles ayant travaillé moins d'un an dans le secteur concerné, après ajustement. Pour les secteurs d'activité de la santé humaine et de l'action sociale, et des activités de services administratifs et de soutien, les OR ajustés sont respectivement de 1,20 et de 1,25, à la limite de la significativité (p-value respectivement de 0,06 et de 0,08).

## I TABLEAU 16 I

### Régression logistique, hommes, cas de cancer du poumon, années 2010-2014, exposition vie entière pour chaque secteur d'activité

	Témoins N = 281 470 <sup>a</sup>	Cas N = 2 318 <sup>a</sup>	OR	95% CI	p-value	OR <sup>b</sup>	95% CI <sup>b</sup>	p- value
<b>C-Industrie manufacturière</b>								
Un an ou plus (ref. <1 an)	127 831 (45,4%)	967 (41,7%)	0,86	[0,80-0,94]	<0,05	0,90	[0,83-0,98]	<0,05
<b>D-Production et distribution d'électricité, de gaz, de vapeur et d'air conditionné</b>								
Un an ou plus (ref. <1 an)	1 862 (0,7%)	16 (0,7%)	1,05	[0,64-1,71]	0,86	1,82	[1,10-3,00]	<0,05
<b>E-Production et distribution d'eau ; assainissement, gestion des déchets et dépollution</b>								
Un an ou plus (ref. <1 an)	6 286 (2,2%)	40 (1,7%)	0,77	[0,56-1,05]	0,10	1,64	[1,20-2,26]	<0,05
<b>F-Construction</b>								
Un an ou plus (ref. <1 an)	47 874 (17,0%)	309 (13,3%)	0,75	[0,67-0,85]	<0,01	1,17	[1,04-1,33]	<0,05
<b>G-Commerce ; réparation d'automobiles et de motocycles</b>								
Un an ou plus (ref. <1 an)	69 709 (24,8%)	379 (16,4%)	0,59	[0,53-0,66]	<0,01	1,18	[1,05-1,32]	<0,01
<b>H-Transports et entreposage</b>								
Un an ou plus (ref. <1 an)	30 072 (10,7%)	157 (6,8%)	0,61	[0,52-0,71]	<0,01	1,15	[0,97-1,35]	0,11
<b>I-Hébergement et restauration</b>								
Un an ou plus (ref. <1 an)	19 054 (6,8%)	91 (3,9%)	0,56	[0,46-0,69]	<0,01	1,70	[1,37-2,11]	<0,01
<b>J-Information et communication</b>								
Un an ou plus (ref. <1 an)	11 060 (3,9%)	22 (1,0%)	0,23	[0,15-0,36]	<0,01	0,73	[0,48-1,12]	0,15
<b>K-Activités financières et d'assurance</b>								
Un an ou plus (ref. <1 an)	18 033 (6,4%)	93 (4,0%)	0,61	[0,50-0,75]	<0,01	0,93	[0,75-1,15]	0,48

	Témoins N = 281 470 <sup>a</sup>	Cas N = 2 318 <sup>a</sup>	OR	95% CI	p-value	OR <sup>b</sup>	95% CI <sup>b</sup>	p-value
<b>L-Activités immobilières</b>								
Un an ou plus (ref. <1 an)	35 062 (12,5%)	262 (11,3%)	0,90	[0,79-1,02]	0,09	1,21	[1,06-1,38]	<0,01
<b>M-Activités spécialisées, scientifiques et techniques</b>								
Un an ou plus (ref. <1 an)	31 657 (11,3%)	134 (5,8%)	0,48	[0,41-0,58]	<0,01	0,96	[0,80-1,14]	0,62
<b>N-Activités de services administratifs et de soutien</b>								
Un an ou plus (ref. <1 an)	85 687 (30,4%)	319 (13,8%)	0,37	[0,32-0,41]	<0,01	1,21	[1,06-1,37]	<0,01
<b>O- Administration publique</b>								
Un an ou plus (ref. <1 an)	39 287 (14,0%)	178 (7,7%)	0,51	[0,44-0,60]	<0,01	1,05	[0,90-1,23]	0,52
<b>P-Enseignement</b>								
Un an ou plus (ref. <1 an)	18 435 (6,6%)	90 (3,9%)	0,58	[0,47-0,71]	<0,01	1,03	[0,83-1,27]	0,81
<b>Q-Santé humaine et action sociale</b>								
Un an ou plus (ref. <1 an)	29 107 (10,3%)	135 (5,8%)	0,54	[0,45-0,64]	<0,01	1,17	[0,98-1,39]	0,09
<b>R-Arts, spectacles et activités récréatives</b>								
Un an ou plus (ref. <1 an)	7 844 (2,8%)	24 (1,04%)	0,37	[0,24-0,55]	<0,01	0,84	[0,56-1,26]	0,39
<b>S-Autres activités de services</b>								
Un an ou plus (ref. <1 an)	14 715 (5,2%)	71 (3,1%)	0,58	[0,45-0,73]	<0,01	0,93	[0,73-1,18]	0,54

*a* Statistique présentée : n (%)

*b* Ajusté sur âge, département, année

**I TABLEAU 17 I**

Régression logistique, femmes, cas de cancer du poumon, années 2010-2014, exposition vie entière pour chaque secteur d'activité

	Témoins N = 287 761 <sup>a</sup>	Cas N = 689 <sup>a</sup>	OR	95% CI	p-value	OR <sup>b</sup>	95% CI <sup>b</sup>	p-value
<b>C-Industrie manufacturière</b>								
Un an ou plus (ref. <1 an)	60 301 (21,0%)	141 (20,5%)	0,97	[0,81-1,17]	0,75	1,01	[0,84-1,23]	0,90
<b>F-Construction</b>								
Un an ou plus (ref. <1 an)	6 974 (2,4%)	8 (1,2%)	0,47	[0,24-0,95]	<0,05	0,68	[0,34-1,36]	0,27
<b>G-Commerce ; réparation d'automobiles et de motocycles</b>								
Un an ou plus (ref. <1 an)	72 800 (25,3%)	104 (15,1%)	0,53	[0,43-0,65]	<0,01	0,97	[0,78-1,20]	0,76
<b>H-Transports et entreposage</b>								
Un an ou plus (ref. <1 an)	12 232 (4,3%)	23 (3,3%)	0,78	[0,51-1,18]	0,24	1,35	[0,89-2,07]	0,16
<b>I-Hébergement et restauration</b>								
Un an ou plus (ref. <1 an)	27 314 (9,5%)	36 (5,2%)	0,53	[0,38-0,74]	<0,01	1,06	[0,75-1,50]	0,73
<b>J-Information et communication</b>								
Un an ou plus (ref. <1 an)	6 381 (2,4%)	10 (1,5%)	0,61	[0,32-1,13]	0,12	1,25	[0,66-2,34]	0,50
<b>K-Activités financières et d'assurance</b>								
Un an ou plus (ref. <1 an)	20 533 (7,1%)	40 (5,8%)	0,80	[0,58-1,10]	0,18	1,40	[1,01-1,93]	<0,05
<b>L-Activités immobilières</b>								
Un an ou plus (ref. <1 an)	23 329 (8,1%)	47 (6,8%)	0,83	[0,62-1,12]	0,22	1,12	[0,83-1,51]	0,48
<b>M-Activités spécialisées, scientifiques et techniques</b>								

	Témoins N = 287 761 <sup>a</sup>	Cas N = 689 <sup>a</sup>	OR	95% CI	p-value	OR <sup>b</sup>	95% CI <sup>b</sup>	p-value
Un an ou plus (ref. <1 an)	27 700 (9,6%)	35 (5,1%)	0,50	[0,36-0,71]	<0,01	0,95	[0,67-1,34]	0,76
N-Activités de services administratifs et de soutien								
Un an ou plus (ref. <1 an)	62 942 (21,9%)	94 (13,6%)	0,56	[0,45-0,70]	<0,01	1,25	[0,99-1,58]	0,06
O-Administration publique								
Un an ou plus (ref. <1 an)	55 777 (19,4%)	87 (12,6%)	0,60	[0,48-0,75]	<0,01	1,05	[0,83-1,33]	0,68
P-Enseignement								
Un an ou plus (ref. <1 an)	27 758 (9,7%)	48 (7,0%)	0,70	[0,52-0,94]	<0,05	1,28	[0,95-1,73]	0,10
Q-Santé humaine et action sociale								
Un an ou plus (ref. <1 an)	70 666 (24,6%)	130 (18,9%)	0,72	[0,59-0,87]	<0,01	1,20	[0,98-1,47]	0,08
R-Arts, spectacles et activités récréatives								
Un an ou plus (ref. <1 an)	7 696 (2,7%)	9 (1,3%)	0,48	[0,25-0,93]	<0,05	0,98	[0,51-1,90]	0,95
S-Autres activités de services								
Un an ou plus (ref. <1 an)	22 672 (7,9%)	41 (6,0%)	0,74	[0,54-1,02]	0,06	1,18	[0,85-1,62]	0,33

<sup>a</sup> Statistique présentée : n (%)

<sup>b</sup> Ajusté sur âge, département, année

### 3.2.2.3 Prise en compte de la consommation différentielle de tabac dans l'interprétation des différences d'incidence par secteur d'activité

Les figures 6 et 7 ci-dessous représentent les odd-ratios ajustés présentés dans les tableaux de la partie précédente ainsi que les rapports d'incidence attendue associés à la consommation de tabac pour les secteurs d'activité avec des effectifs suffisants pour le calcul des RIA. Ces rapports représentent les ordres de grandeur attendus des effets du tabac par secteur d'activité en termes d'incidence du cancer du poumon. Il s'agit donc, pour les secteurs où la prévalence du tabagisme est plus élevée qu'ailleurs, des valeurs au-dessus desquelles l'incidence (ou la mortalité) relative ne pourrait être expliquée uniquement par cette plus forte prévalence mais aussi par l'existence d'autres facteurs de risques et notamment par des expositions professionnelles spécifiques.

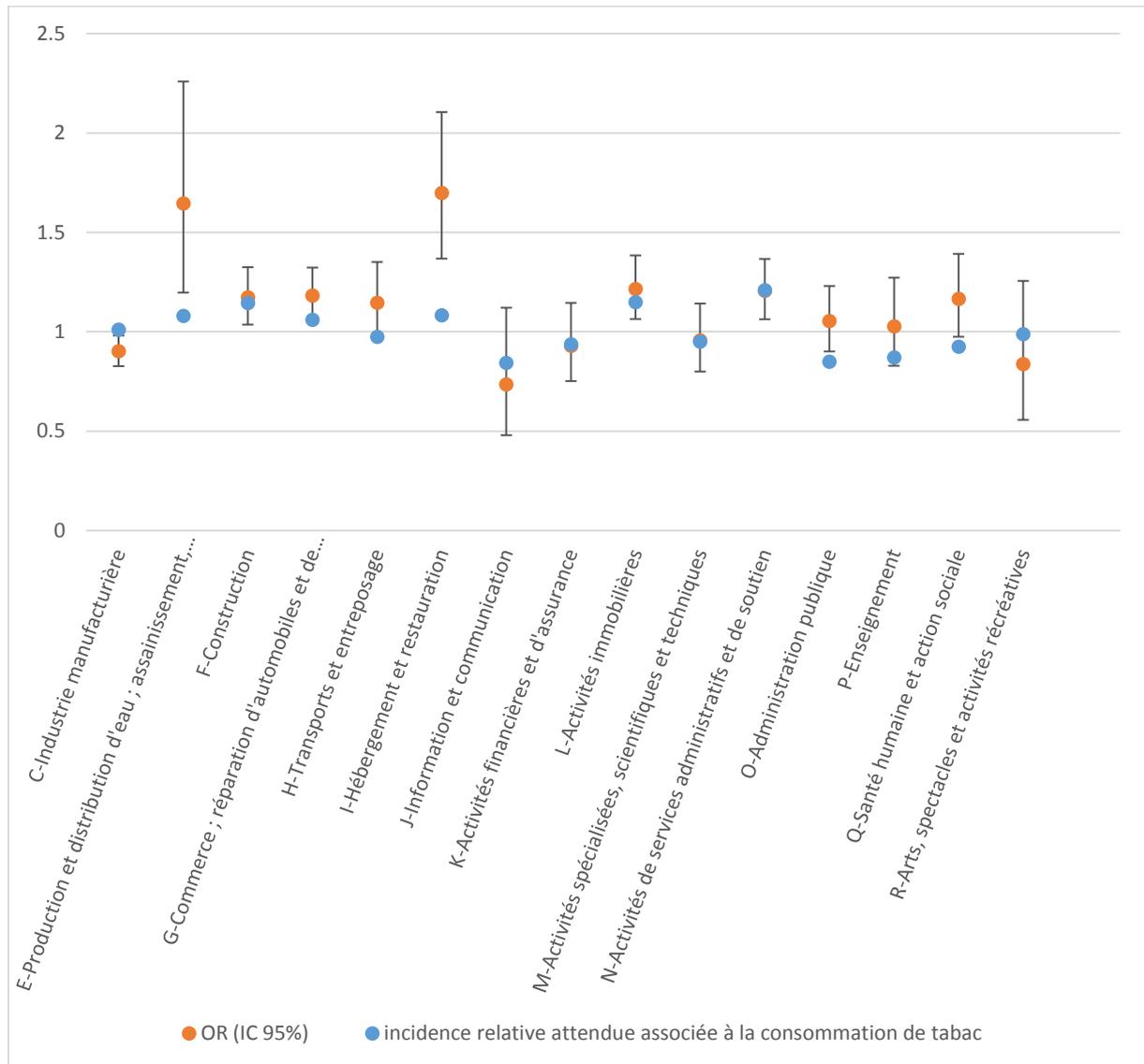
Parmi les secteurs d'activité pour lesquels on observe un sur-risque d'incidence de cancer du poumon chez les hommes (Figure 6), la consommation différentielle de tabac peut expliquer ce sur-risque dans plusieurs d'entre eux avec un intervalle de confiance se superposant au rapport d'incidence attendue : le secteur de la construction, le secteur du commerce et de la réparation automobiles, le secteur des activités immobilières et le secteur des activités de services administratifs et de soutien.

Chez les hommes les secteurs d'activités pour lesquels le risque de cancer du poumon est significativement supérieur à 1 et ne peut pas être expliqué uniquement par une prévalence élevée de tabac dans ce secteur d'activité sont : la production et distribution d'eau ; assainissement, gestion des déchets et dépollution (RIA=1,08 ; ORa=1,64) ; et le secteur de l'hébergement et la restauration (RIA=1,08 ; ORa=1,70).

Pour le secteur d'activité de l'industrie manufacturière (RIA=1,01 ; ORa=0,81 [0,75-0,88]) on observe *a contrario* un risque de cancer du poumon significativement inférieur à 1 et un OR inférieur au RIA (RIA=1,01 ; ORa=0,90).

I FIGURE 6 I

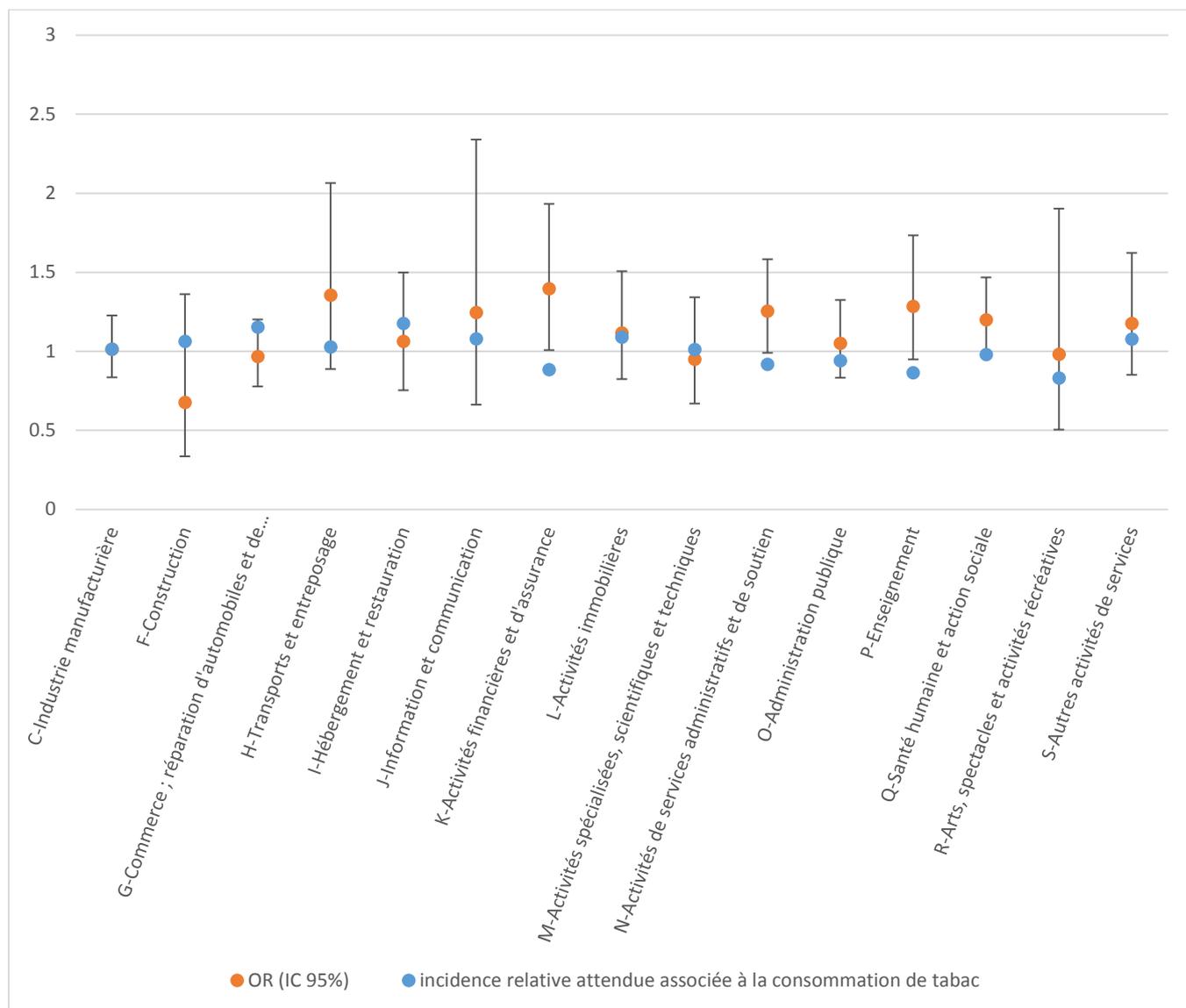
Odds-ratios d'incidence de cancer du poumon et incidence relative associée à la consommation de tabac par secteurs d'activité chez les hommes



Chez les femmes (Figure 7), le sur-risque observé de cancer du poumon dans les secteurs des activités financières et d'assurance ne peut pas être expliqué uniquement par une différence de prévalence de tabagisme avec un RIA inférieur à l'OR ajusté estimé (RIA=0,92 ; ORa=1,25).

**I FIGURE 7 I**

**Odds-ratios d'incidence de cancer du poumon et incidence relative associée à la consommation de tabac par secteurs d'activité chez les femmes**



## 4. DISCUSSION

### 4.1 Performance de l'appariement et qualité des données

#### 4.1 Résultats de l'appariement

L'appariement sur les données d'état civil des cas de cancers en l'absence du NIR se révèle satisfaisant avec un taux d'appariement de 97%, seules 1 151 personnes n'ayant pas été retrouvées lors de la recherche au SNGI.

L'analyse des données de personnes non retrouvées montre l'importance d'un travail minutieux et systématique de nettoyage de données des registres préalablement au passage au RNIPP, en effet 43% des individus non retrouvés n'avaient pas eu de recherche au RNIPP, pour 35% des perdus de vue les personnes étaient décédées. La forte proportion de femmes (71%) parmi les perdus de vue confirme l'impact du nettoyage des données en amont de l'appariement, il est fort probable que cette différence dans les taux d'appariement résulte directement de différences entre usage du nom de naissance et usage du nom marital. Il sera donc nécessaire dans le cadre de la rédaction du protocole du système de surveillance de réfléchir avec les registres des cancers à la possibilité de synchroniser les appariements avec la Cnav à la recherche de statuts vitaux effectuée au RNIPP afin de disposer de données d'état civil les plus complètes et exactes possibles tout en limitant les travaux supplémentaires pour les registres. Il sera également nécessaire d'estimer le coût-bénéfice de réaliser ces vérifications des données d'état civil y compris parmi les personnes dont le statut vital est connu : le coût en terme de travail supplémentaire pour les registres, ces tâches étant particulièrement consommatrices de temps et les décès étant nombreux pour les pathologies cancéreuses ; et le bénéfice en terme de personnes retrouvées au SNGI, le taux d'appariement étant d'ors et déjà élevé.

#### 4.2 Tirage au sort des témoins

Le tirage au sort des témoins ne peut s'effectuer que dans les bases pour lesquelles nous disposons d'une information sur le département : la base « validité » pour laquelle on dispose du département de l'employeur et qui concerne donc uniquement les individus salariés l'année de tirage au sort et la base « retraite » pour laquelle on dispose du département de résidence de la personne retraitée.

Cela pose deux limites majeures :

- Les individus n'étant ni salariés ni retraités pour les années considérées ont une probabilité nulle d'être tirés au sort. Cela exclut notamment les personnes affiliées à un régime spécial (régime agricole, régime des indépendants, fonction publique, etc.) ;
- Les individus salariés ne sont pas rattachés à leur département de résidence mais à leur département de travail contrairement à la définition de cas utilisée par les registres des cancers, utilisant le département de résidence.

Néanmoins la majorité des cancers intervient chez des personnes d'un âge assez élevé (moyenne des cas de 65,8 ans, 58,5% des cas de cancers ont plus de 64 ans) donc retraités ce qui limite sûrement le biais pour l'étude cas-témoin.

Cette différence de définition entre cas et témoins implique de faire l'hypothèse que les salariés vivant dans le département (pour les cas) et les salariés travaillant dans le département (pour les témoins) étaient semblables en terme de secteurs d'activité occupés.

### 4.3 Reconstitution de l'historique professionnel

- *Absence d'information sur la PCS*

La déclaration de la profession et catégorie socio-professionnelle (PCS) est obligatoire depuis 2008 pour les entreprises de plus de 20 salariés, elle devrait donc être bien renseignée à partir de cette date-là. Cependant, pour les années 2009, 2010, 2011, selon les informations de la Cnav, un problème informatique au niveau de la base a entraîné un mauvais taux de remplissage de cette variable. Elle est de nouveau bien renseignée à partir de 2012. Pour les années 2010-2014 cette variable était renseignée pour 80% des validités salariées. La PCS dans les bases de la Cnav est codée à partir de la nomenclature PCS en vigueur au moment du report, sur 2 digits. Il est question dans l'évolution des bases de la Cnav que le code PCS soit saisi sur 4 digits. Il sera nécessaire à la mise en place du système de surveillance pérenne d'évaluer la qualité de ces données remontées dans le nouveau système de gestion des carrières pour une utilisation éventuelle dans les indicateurs d'incidence, son utilisation étant impossible sur les années concernées par l'étude pilote et sur l'ensemble de la carrière des individus en l'état actuel.

- *Absence d'information sur les employeurs avant 1972*

L'absence d'information sur les employeurs avant 1972 implique que les données sur la carrière des personnes de plus de 58 ans (sous l'hypothèse d'un début de carrière à 20 ans) sont incomplètes. Les années d'incidence de cancers dans l'étude pilote commençant pour l'année 2010, on peut supposer que l'impact de cette absence d'information est négligeable dans l'étude cas-témoins car la durée de carrière retracée est de 38 ans, suffisante pour mesurer un impact de l'activité professionnelle sur la survenue d'un cancer. De plus, cette absence d'informations impacte de la même manière cas et témoins.

Cette limite dans les données est sans effet sur l'étude d'incidence puisque l'on s'intéresse à l'activité professionnelle au moment du diagnostic.

- *Recherche du secteur d'activité à partir du Siret*

Il avait été fait le choix *a priori* de rechercher les secteurs d'activité à partir des codes Siret des établissements *via* la base Sirene car le code APE qui renseigne le secteur d'activité dans le SNGC n'était renseigné qu'à partir de 1999 et n'avait pas de caractère obligatoire.

Les résultats de l'étude pilote montrent que le code APE était renseigné pour 84% des employeurs pour notre période d'étude (2010-2014). La recherche des secteurs d'activité à partir des numéros Siret dans la base Sirene a conduit à retrouver l'ensemble des secteurs d'activité sur cette même période. La recherche du secteur d'activité *via* la base Sirene se révèle donc actuellement encore nécessaire afin de disposer du secteur d'activité pour le plus grand nombre de validités possibles. Dans les années à venir l'utilisation du code APE directement accessible au SNGC sans rajouter l'étape de recherche *via* le Siret devrait être possible, la complétude des données s'améliorant fortement. Sur la période 2014-2018, le code APE est renseigné pour 99,7% des validités avec un employeur.

Si la recherche du secteur d'activité *via* le Siret est très satisfaisante sur l'étude d'incidence portant uniquement sur l'année d'incidence, en revanche la reconstitution de l'historique professionnel n'est pas complète sur l'ensemble des années. Avant 1972 et le début de la notion d'employeur dans le SNGC, la recherche d'un secteur activité est impossible car le numéro de l'employeur n'est pas renseigné. Pour la période 1972-1998, le numéro de l'employeur est renseigné mais les résultats de cette étude montrent qu'il ne s'agit pas d'un code Siret dans 43% des cas et donc ne permet pas la recherche d'un secteur d'activité dans

la base Sirene. En revanche à partir de 1999, la qualité des données relatives au numéro de l'employeur est satisfaisante avec seulement 4% de codes employeurs ne correspondant pas à un Siret.

- *Absence d'informations sur les autres régimes*

Les systèmes d'information gérés par la Caisse nationale d'assurance vieillesse ont une visée administrative et ne concernent que les personnes cotisant au régime général pour le calcul et le versement de la retraite. La conséquence directe de cette limite dans la disponibilité des données est que l'information sur les autres régimes est parcellaire : remontée à la Cnav de l'information de cotisations uniquement au moment du calcul des droits à la retraite et sans plus d'informations que le régime de cotisation, donc une impossibilité actuellement de déterminer des secteurs d'activité pour les validités pour lesquelles l'information est disponible.

Il en résulte également que les périodes sans validités renseignées pour les personnes encore en activité ne peuvent pas être exploitées. Il est effectivement impossible de savoir s'il s'agit de périodes sans activité ou de périodes avec une activité dans un autre régime pour lesquelles l'information ne sera remontée qu'ultérieurement à la Cnav.

Les études menées avec les données du SNGC ne portent donc que sur les personnes salariées au régime général. Or les secteurs d'activité occupés et donc les expositions qui s'y rapportent sont très différents selon les régimes de cotisation : par exemple prépondérance du secteur de l'agriculture, sylviculture et pêche pour la sécurité sociale agricole (MSA) ou du secteur de la production et distribution d'électricité, de gaz, de vapeur et d'air conditionné pour le régime d'EDF/GDF - Électricité Gaz de France alors que les activités dans ce secteur d'activité sont faibles pour les salariés au régime général.

D'autre part, les indépendants ne rentrent pas non plus dans le cadre de cette étude puisque affiliés au régime social des indépendants (RSI) aux dates considérées. Or, il s'agit d'une population particulière ne disposant pas de médecine du travail ni de suivi post-professionnel. La mise en place du système de surveillance, en l'état actuel, sera donc limité aux personnes salariées au régime général en raison même de la nature des données comme c'est le cas dans l'étude pilote.

## 4.2 Étude d'incidence des cancers du poumon par secteur d'activité pour les cas du registre du Doubs et du territoire de Belfort

Le protocole de l'étude prévoyait d'utiliser les données du recensement de la population pour servir de population de référence dans le calcul des ratios standardisés d'incidence (SIR). La limitation des données de la Cnav aux seuls salariés du régime général n'a pas permis l'utilisation de cette source de données dans ce cadre car il est impossible dans les données du recensement de distinguer les individus affiliés au régime général vs. à un autre régime. L'utilisation de ces données comme population de référence aurait donc conduit à une forte sous-estimation des SIR. La solution a donc été de partir des témoins tirés au sort dans les bases de la Cnav afin de reconstituer la population de référence.

Néanmoins la reconstitution d'une population de référence pour le registre de Lille n'a pas été possible. En effet ce registre ne couvre pas l'ensemble du département du Nord mais seulement les villes de Lille et de son agglomération alors que les témoins ont été tirés au sort au niveau du département car l'information de la commune n'était pas disponible pour les salariés. Le calcul avec une population de référence couvrant l'ensemble du département aurait conduit à une forte sous-estimation des SIR pour ce registre. Parmi les registres des

cancers existants sur le territoire national, il s'agit du seul registre à ne pas couvrir la totalité d'un département. Cette limite ne constitue donc pas un frein à l'extension du système de surveillance à l'ensemble des registres.

Globalement l'étude d'incidence montre des ratios standardisés d'incidence inférieurs à 1 chez les personnes salariées au régime général par rapport à la population générale dont sont extraits les taux d'incidence permettant de calculer le nombre de cas attendus. Ce constat résulte sûrement en partie de l'effet du travailleur sain, une personne malade ayant de plus grande chance d'être exclue du marché du travail ou de partir en retraite anticipée [36, 37]. Cet effet du travailleur sain a été mis en évidence pour certaines localisations de cancer par une étude suédoise qui a montré la présence d'un effet du travailleur sain pour les cancers broncho-pulmonaires [38].

On peut également faire l'hypothèse que cette absence de sur-incidence est liée au fait que l'étude porte sur la population salariée au régime général qui constitue sûrement une population particulière au sein de l'ensemble des travailleurs. En effet, les salariés constituent une population globalement plus protégée réglementairement avec l'existence d'un suivi médical et des expositions dans le cadre de la médecine du travail. Ce résultat est concordant avec une étude récente qui montre que la population salariée était moins exposée à trois nuisances étudiées (farine, céréales et formaldéhyde) [39]. Bien que les nuisances considérées ne soient pas classées comme cancérigènes pour le cancer du poumon, on peut penser que ce constat est également valable pour une grande partie des expositions du fait du contexte réglementaire et notamment pour les cancérigènes responsables du cancer du poumon.

On observe une sous-incidence significative pour les secteurs d'activité de l'industrie manufacturière, du commerce, de la réparation d'automobiles et de motocycles, des activités spécialisées, scientifiques et techniques et de la santé humaine et de l'action sociale chez les hommes et pour le secteur du transport et de l'entreposage et le secteur de l'enseignement chez les femmes.

Le secteur d'activité de l'industrie manufacturière regroupe des situations variées et des secteurs pour lesquels les expositions à des substances cancérigènes sont très variées. Par exemple, dans l'étude Sumer menée en 2017, on note que 18% des salariés étaient exposés à un ou plusieurs agents chimiques dans le secteur de la fabrication de textiles contre 55% dans les secteurs du travail du bois ou le secteur de la fabrication de produits en caoutchouc [40]. Il est donc important dans les analyses à venir de pouvoir descendre à un niveau de la NAF révision 2 plus fin afin de pouvoir caractériser d'éventuelles disparités dans les ratios d'incidence standardisés au sein de ce secteur d'activité.

Il existe un gradient fort dans l'incidence des cancers selon la catégorie socio-professionnelle et le niveau d'éducation. Il n'a pas été possible de mesurer l'incidence selon la PCS dans cette étude pilote mais l'on peut faire l'hypothèse que la sous-incidence observée dans les secteurs d'activité des activités spécialisées et techniques ; de la santé humaine et de l'action sociale chez les hommes et pour le secteur de l'enseignement chez les femmes soient en partie liée à une surreprésentation des personnes avec une catégorie sociale et/ou une éducation plus élevée dans ces secteurs d'activité en comparaison avec la population générale.

### 4.3 Étude cas témoin sur les cancers du poumon

L'étude des données de la Cnav pour la reconstitution des carrières des individus a montré l'incomplétude des données employeur sur les informations disponibles avant 1999. Cette limite sur la reconstitution de l'historique professionnel concerne de la même manière cas et témoins et ne constitue donc pas un biais majeur.

L'étude cas-témoin menée dans le cadre de ce pilote permet de prendre en compte l'ensemble de la carrière des individus et de ne pas se limiter à l'activité au moment du diagnostic. Le principal facteur de risque de cancer du poumon est la consommation de tabac. Il était impossible de prendre en compte le tabagisme car cette information n'est disponible dans aucune des deux sources de données utilisées. Il a donc été décidé de calculer les incidences relatives attendues associées à la consommation de tabac afin de mettre en perspective les résultats par secteur d'activité. Ces incidences relatives ont été calculées en prenant en compte les données de prévalence de fumeurs et d'anciens fumeurs par secteurs d'activité issues du Baromètre santé 2010. La principale limite de cet indicateur est qu'il se base sur des données transversales et prend en compte le tabagisme passé dans un secteur d'activité en 2010. Il est possible que les prévalences de tabagisme aient fortement évolué au cours des années passées et que cette évolution diffère selon les secteurs d'activité. Cet indicateur constitue donc plutôt une aide à l'interprétation des données, le tabagisme variant fortement d'un secteur d'activité à un autre.

L'analyse réalisée en prenant en compte les disparités de prévalence du tabagisme met en exergue trois secteurs d'activités pour lesquels on observe une sur-incidence à la fois significativement supérieure à 1 et à l'incidence relative attendue associée à la consommation de tabac : le secteur de la production et distribution d'eau ; assainissement, gestion des déchets et dépollution, et le secteur de l'hébergement et la restauration chez les hommes, le secteur des activités financières et d'assurance chez les femmes.

Cette étude a également permis de mettre en évidence un risque significativement inférieur à 1 de cancer du poumon pour les personnes ayant travaillé au moins un an dans le secteur d'activité de l'industrie manufacturière. Comme pour l'étude d'incidence, il sera nécessaire dans la suite du projet d'étudier ce secteur d'activité à un niveau plus fin. Une étude de cohorte menée en Suisse sur la mortalité montre également un risque significativement inférieur à 1 pour ce secteur d'activité [41].

La même étude menée en Suisse montrait un sur-risque de mortalité pour le cancer du poumon dans le secteur d'activité de l'hébergement et de la restauration comme observé dans notre étude pour les hommes. Il s'agit d'un secteur d'activité dans lequel les salariés sont fortement exposés à agents biologiques (34,5% des salariés du secteur contre 24,9% pour l'ensemble des secteurs) et/ou à des agents chimiques (45,1% des salariés contre 32,3% des salariés tous secteurs) selon les données de l'étude Sumer [40]. On peut noter également dans ce secteur d'activité une exposition passée au tabagisme passif, qui a pu constituer pour les salariés de ce secteur une source d'exposition majeure à ce cancérogène [42].

Pour le secteur de la production et distribution d'eau ; de l'assainissement, gestion des déchets et dépollution, on note là encore une forte exposition des salariés à des agents biologiques (43,6%). Ce secteur d'activité comprend les activités liées à la gestion (comprenant les services de collecte, de traitement et d'élimination) de différents types de déchets, tels que les déchets, solides ou non, produits par les industries ou les ménages, ainsi que des sites contaminés. Ce secteur d'activité a connu au cours des dernières années des changements structurels forts avec de nouvelles filières, de nouveaux procédés et de nouveaux métiers qui se mettent progressivement en place. Il sera intéressant de surveiller les évolutions futures pour ce secteur d'activité pour lequel on dispose de peu d'études épidémiologiques.

Chez les femmes, l'enquête Sumer chez les salariés en 2010, montrait que le premier secteur d'activité exposant à un cancérogène de l'appareil broncho-pulmonaire était le secteur de la santé suivi du secteur du transport et de l'entreposage (c'étaient notamment des conductrices et des agents de transport et du tourisme) [43]. Ces secteurs étaient associés à des odd ratios supérieurs à 1 dans notre étude bien que non significatifs.

Le seul secteur d'activité pour lequel on observe un sur-risque d'incidence de cancer du poumon chez les femmes est le secteur des activités financières et d'assurance, secteur pour lequel on observe une très faible exposition des salariés à des agents biologiques (5,0%) ou chimiques (5,7%) selon l'étude Sumer en 2017. Ce sur-risque dans un secteur d'activité avec une forte proportion d'employés administratifs est difficilement explicable.

## 5. CONCLUSION ET PERSPECTIVES

Les objectifs de cette étude pilote étaient de mettre au point les modalités d'accès et de croisement des données des registres des cancers avec les bases de données de la Cnav, d'analyser la qualité des données, de mettre au point les différents indicateurs qui pourront être mis en œuvre en routine pour le système de surveillance final et d'étudier les limites de la démarche dans l'identification des expositions professionnelles.

Les résultats de cette étude pilote ont montré la faisabilité de l'appariement en l'absence de la disponibilité du NIR. Les résultats de l'appariement, basé sur les noms, prénoms, date et lieu de naissance, ont été satisfaisants avec seulement 3% de perdus de vue qui ne semblent pas avoir de caractéristiques particulières en terme de registre, âge ou année considérée.

La faisabilité de la reconstitution de l'historique professionnel dans la perspective de réaliser des mesures d'associations entre survenue du cancer et secteur d'activité s'est révélée également satisfaisante à partir de l'année 1999. Du fait de l'avancée dans le temps et de l'amélioration progressive des données renseignées au SNGC, cette limite est vouée à s'atténuer : en effet à partir des cas incidents de 2019, nous disposerons d'un historique professionnel fiable sur les 20 dernières années de carrières, ce qui semble suffisant y compris pour des études portant sur des maladies avec des durées de latence longues comme les cancers.

En dépit des limites mises en évidence lors de cette phase pilote, le système de surveillance Sicapro possède les propriétés d'un système de surveillance efficace :

- Système intégré impliquant relativement peu de surcharge de travail car se basant sur des données pré-existantes sans recueil complémentaire ;
- Acceptable car il limite au maximum toutes les mesures contraignantes pour les patients et pour les personnels de santé (absence de recueil supplémentaires des données rétrospectives notamment) ;
- Sensible, les registres de cancers recueillant de façon exhaustive les cas incidents de leur zone ;
- Représentatif de la population salariée au régime général.

L'inclusion de l'ensemble des registres de cancers au système de surveillance, couvrant près de 20% de la population, permettra une plus grande puissance permettant une étude de l'incidence pour des localisations de cancers plus rares et/ou pour des secteurs d'activité à un niveau plus fin. Le système de surveillance permettra également de produire des indicateurs à un niveau départemental pour les départements intégralement couverts par un registre, y compris pour certains départements d'outre-mer (Guadeloupe, Martinique et Guyane).

Dans la perspective de la pérennisation du système de surveillance des cancers en lien avec l'activité professionnelle étendu à l'ensemble des registres du cancer, la mise en œuvre par la Cnav du nouveau système d'information inter-régime est prometteur. Le répertoire de gestion des carrières unique (RGCU) a vocation, entre autres, à mettre en commun les données relatives à la carrière de chaque assuré en vue d'assurer la complétude et la cohérence de ces données<sup>4,5</sup>. L'alimentation du RGCU en temps réel par les autres régimes, fixée au 31 décembre 2022, pourrait permettre de lever la principale limite soulevée lors de cette étude pilote en permettant de ne pas limiter les analyses aux seules personnes salariées du régime général.

---

<sup>4</sup> Loi n° 2010-1330 du 9 novembre 2010 portant réforme des retraites.

<sup>5</sup> Décret n° 2018-154 du 1<sup>er</sup> mars 2018 relatif au répertoire de gestion des carrières unique.

Au-delà de l'estimation des incidences par secteur d'activité voire à terme par profession, en tenant compte de l'historique professionnel, ce système de surveillance Sicapro, basé sur les données des registres et de la Cnav offre également des perspectives intéressantes pour l'étude du maintien en emploi et du retour à l'emploi suite à la survenue d'un cancer. En effet sont également enregistrées les validités correspondant à des périodes d'arrêt maladie ou des périodes de chômage indemnisées. Dans cette perspective, des travaux complémentaires sont actuellement en cours pour étudier la faisabilité de produire des indicateurs de maintien ou de retour à l'emploi suite à la survenue d'un cancer à partir de ce système de surveillance Sicapro.

Une prochaine étape consistera à utiliser des matrices emploi-exposition afin d'estimer les expositions à des substances cancérigènes. Un frein à l'utilisation de ces matrices est lié à l'indisponibilité des informations sur la PCS pour l'ensemble de la carrière. En effet, la plupart des matrices emploi-exposition utilise le secteur d'activité, la profession et la période afin d'estimer la probabilité d'exposition à un facteur de risque<sup>6</sup>. La disponibilité de ces données à partir de l'année 2012 permettra d'utiliser des matrices pour évaluer les expositions dans les études d'incidence du système de surveillance pérenne.

En conclusion, avec la pérennisation de ce système de surveillance Sicapro, la France se doterait d'un système de surveillance spécifique des cancers en lien avec l'activité professionnelle, renforçant ainsi les dispositifs de surveillance des cancers d'origine professionnelle et en adéquation avec la nouvelle stratégie décennale de lutte contre les cancers (2021-2030) dans laquelle figure une action visant à mieux reconnaître les expositions professionnelles pour mieux prévenir les cancers professionnels.

La mobilisation de l'ensemble des registres des cancers et le déploiement du répertoire de gestion des carrières unique (RGCU) intégrant les informations des autres régimes, sont essentiels afin d'étendre ce dispositif pérenne de surveillance Sicapro à l'ensemble du territoire couvert par un registre des cancers et à l'ensemble des régimes et aux travailleurs non-salariés.

---

<sup>6</sup> <https://expro.santepubliquefrance.fr/matrices>

## Références bibliographiques

1. Frery N, Moisan F, Schwaab Y, Garnier R. Exposition des salariés à de multiples nuisances cancérogènes en 2010. *Bull Epidemiol Hebd* 2017(13):242-9.
2. Imbernon E. Estimation du nombre de cas de certains cancers attribuables à des facteurs professionnels en France. Institut de veille sanitaire, Saint-Maurice, 2003, 28 pages.
3. Marant Micallef C, Shield KD, Vignat J, Baldi I, Charbotel B, Fervers B, Gilg Soit Ilg A, Guenel P, Olsson A, Rushton L, Hutchings SJ, Clero E, Laurier D, Scanff P, Bray F, Straif K, Soerjomataram I. Cancers in France in 2015 attributable to occupational exposures. *Int J Hyg Environ Health* 2019;222(1):22-9.
4. Defossez G, Le Guyader-Peyrou S, Uhry Z, Grosclaude P, Colonna M, Dantony E, Delafosse P, Molinié F, Woronoff A-S, Bouvier A-M, Bossard N, Remontet L, Monnereau A. Estimations nationales de l'incidence et de la mortalité par cancer en France métropolitaine entre 1990 et 2018 - Volume 1 : Tumeurs solides : Étude à partir des registres des cancers du réseau Francim, Saint-Maurice, Santé publique France, 2019, 372 p. pages.
5. Le Guyader-Peyrou S, Defossez G, Dantony E, Mounier M, Cornet E, Uhry Z, Cowppli-Bony A, Maynadié M, Troussard X, Delafosse P, Grosclaude P, Colonna M, Woronoff A-S, Remontet L, Bossard N, Monnereau A. Estimations nationales de l'incidence et de la mortalité par cancer en France métropolitaine entre 1990 et 2018 - Volume 2 : Hémopathies malignes : Étude à partir des registres des cancers du réseau Francim, Saint-Maurice, Santé publique France, 2019, 169 p. pages.
6. Wetterhall SF, Pappaioanou M, Thacker SB, Eaker E, Churchill RE. The role of public health surveillance: information for effective action in public health. *MMWR Morb Mortal Wkly Rep* 1992;41 Suppl:207-18.
7. Choi BCK. The Past, Present, and Future of Public Health Surveillance. *Scientifica* 2012;2012:875253.
8. Pukkala E, Martinsen JI, Lynge E, Gunnarsdottir HK, Sørensen P, Tryggvadottir L, Weiderpass E, Kjaerheim K. Occupation and cancer - follow-up of 15 million people in five Nordic countries. *Acta oncologica (Stockholm, Sweden)* 2009;48(5):646-790.
9. Bigert C, Martinsen JI, Gustavsson P, Sparén P. Cancer incidence among Swedish firefighters: an extended follow-up of the NOCCA study. *Int Arch Occup Environ Health* 2020;93(2):197-204.
10. Crosignani P, Massari S, Audisio R, Amendola P, Cavuto S, Scaburri A, Zambon P, Nedoclan G, Stracci F, Pannelli F, Vercelli M, Miligi L, Imbriani M, Berrino F. The Italian surveillance system for occupational cancers: characteristics, initial results, and future prospects. *American journal of industrial medicine* 2006;49(9):791-8.
11. De Matteis S, Consonni D, Lubin JH, Tucker M, Peters S, Vermeulen R, Kromhout H, Bertazzi PA, Caporaso NE, Pesatori AC, Wacholder S, Landi MT. Impact of occupational carcinogens on lung cancer risk in a general population. *International journal of epidemiology* 2012;41(3):711-21.
12. Oddone E, Edefonti V, Scaburri A, Vai T, Crosignani P, Imbriani M. Female breast cancer in Lombardy, Italy (2002-2009): a case-control study on occupational risks. *American journal of industrial medicine* 2013;56(9):1051-62.
13. Gilg Soit Ilg A, Ducamp S, Chérié-Challine L. Programme national de surveillance du mésothéliome pleural (PNSM) : 20 années de surveillance (1998-2017) des cas de mésothéliome, de leurs expositions et des processus d'indemnisation, Saint-Maurice, Santé publique France, 2019.

14. Cherie Challine L, Gilg Soit Ilg A, Grange D, Bousquet PJ, Lafay L. Dispositif national de surveillance des mésothéliomes intégrant la surveillance de leurs expositions. État des lieux des systèmes, enjeux de surveillance et recommandations, Saint-Maurice, Santé publique France, 2017, 185 p. pages.
15. Lapostolle A, Geoffroy-Perez B, Chérié-Challine L. Mortalité par cancer du poumon chez les hommes salariés selon l'activité professionnelle. *Rev Mal Respir* 2019:1-9.
16. Geoffroy Perez B, Imbernon E, Goldberg M. Projet Cosmop : Cohorte pour la surveillance de la mortalité par profession. Premiers résultats de l'étude de faisabilité à partir de l'échantillon démographique permanent. Institut de veille sanitaire, Saint-Maurice, 2005, 6 pages.
17. Armenti KR, Celaya MO, Cherala S, Riddle B, Schumacher PK, Rees JR. Improving the quality of industry and occupation data at a central cancer registry. *American journal of industrial medicine* 2010;53(10):995-1001.
18. Lauzeille D. Recueil d'informations professionnelles auprès de patients atteints de cancer. Résultats de deux études pilotes menées avec les registres dans les départements de l'Hérault et de Vendée, Saint-Maurice, Institut de veille sanitaire, 2011, 26 p. pages.
19. Jehannin Ligier K, Molinie F, Defossez G, Daubisse Marliac L, Dantony E, Uhry Z, Remontet L, Bossard N. Projection de l'incidence et de la mortalité par cancer en France métropolitaine en 2017. Santé publique France, Saint-Maurice, 2017, 80 pages.
20. Binder Foucard F, Belot A, Delafosse P, Remontet L, Woronoff AS, Bossard N. Estimation nationale de l'incidence et de la mortalité par cancer en France entre 1980 et 2012. Etude à partir des registres des cancers du réseau Francim - Partie 1 : tumeurs solides. Institut de veille sanitaire, Saint-Maurice, 2013, 122 pages.
21. Delva F, Andujar P, Lacourt A, Brochard P, Pairon JC. Facteurs de risque professionnels du cancer bronchopulmonaire. *Rev Mal Respir* 2015.
22. Pairon JC, Brochard P, Le Bourgeois JP, Ruffie P. Les cancers professionnels, Paris, Margaux Orange, 2000, 688 pages.
23. Steenland K, Loomis D, Shy C, Simonsen N. Review of occupational lung carcinogens. *American journal of industrial medicine* 1996;29(5):474-90.
24. Cogliano VJ, Baan R, Straif K, Grosse Y, Lauby-Secretan B, El Ghissassi F, Bouvard V, Benbrahim-Tallaa L, Guha N, Freeman C, Galichet L, Wild CP. Preventable exposures associated with human cancers. *Journal of the National Cancer Institute* 2011;103(24):1827-39.
25. Direction générale du travail. Conditions de travail. Bilan 2015. Ministère du Travail, de l'Emploi, de la Formation professionnelle et du Dialogue social Paris, 2017, 552 pages.
26. Ribet C, Gente J, Zins M. Les bases de données socio-professionnelles, Inserm, 2006, 69 pages.
27. Brossard C. Description des trajectoires professionnelles des seniors à partir des données de carrière Cnav, Caisse Nationale d'Assurance Vieillesse, 2008, 18 pages.
28. Guignard R, Nguyen Thanh V, Andler R, Richard JB, Beck F, Arwidson P. Usage de substances psychoactives des chômeurs et des actifs occupés et facteurs associés : une analyse secondaire du Baromètre santé 2010. *Bull Epidemiol Hebd* 2016(16-17):304-12.
29. Beck F, Gautier A, Guignard R, Richard JB. Méthode d'enquête du Baromètre santé 2010, Saint-Denis, Inpes, 2013, 28 p. pages.
30. Guignard R, Beck F, Richard J-B, Peretti-watel P. Le tabagisme en France. Analyse de l'enquête Baromètre santé 2010, Saint-Denis, Inpes, 2014, 56p. pages.

31. Beck F, Guignard R, Richard JB, Wilquin JL, Peretti Watel P. Augmentation récente du tabagisme en France : principaux résultats du Baromètre santé, France, 2010. *Bull Epidemiol Hebd* 2011(20-21):230-3.
32. Bouyer J, Hémon D, Cordier S, Derriennic F, Stücker I, Stengel B, Clavel J. *Épidémiologie-Principes et méthodes quantitatives*. Lavoisier, Paris, 2009, 498 pages.
33. Axelson O, Steenland K. Indirect methods of assessing the effects of tobacco use in occupational studies. *American journal of industrial medicine* 1988;13(1):105-18.
34. Lauzeille D, Marchand JL, Ferrand M. Consommation de tabac par catégorie socioprofessionnelle et secteur d'activité. Outil méthodologique pour l'épidémiologie, Saint-Maurice, Institut de veille sanitaire, 2009, 208 p. pages.
35. Pesch B, Kendzia B, Gustavsson P, Jöckel K-H, Johnen G, Pohlabein H, Olsson A, Ahrens W, Gross IM, Brüske I, Wichmann H-E, Merletti F, Richiardi L, Simonato L, Fortes C, Siemiatycki J, Parent M-E, Consonni D, Landi MT, Caporaso N, Zaridze D, Cassidy A, Szeszenia-Dabrowska N, Rudnai P, Lissowska J, Stücker I, Fabianova E, Dumitru RS, Bencko V, Foretova L, Janout V, Rudin CM, Brennan P, Boffetta P, Straif K, Brüning T. Cigarette smoking and lung cancer--relative risk estimates for the major histological types from a pooled analysis of case-control studies. *Int J Cancer* 2012;131(5):1210-9.
36. Shah D. Healthy worker effect phenomenon. *Indian J Occup Environ Med* 2009;13(2):77-9.
37. Li CY, Sung FC. A review of the healthy worker effect in occupational epidemiology. *Occupational medicine (Oxford, England)* 1999;49(4):225-9.
38. Kirkeleit J, Riise T, Bjorge T, Christiani DC. The healthy worker effect in cancer incidence studies. *American journal of epidemiology* 2013;177(11):1218-24.
39. Fels A, Houot M, Garras L, Delabre L, Pilorget C. Travailleurs salariés et non salariés en France entre 2007 et 2015 : description des populations et identification de différences d'exposition professionnelle. *Bull Epidemiol Hebd* 2021(2):22-31.
40. Matinet B, Rosankis E, Tassy V. Les expositions aux risques professionnels par secteur d'activité. *Synthèse Stat'* 2020;35:1-192.
41. Bovio N, Richardson DB, Guseva Canu I. Sex-specific risks and trends in lung cancer mortality across occupations and economic activities in Switzerland (1990-2014). *Occupational and environmental medicine* 2020;77(8):540-8.
42. Alipour S, Deschamps F, Lesage FX, Lebagry F. Estimation of annual incidence of lung cancer associated with work place exposure to passive smoking in France. *J Occup Health* 2006;48(5):329-31.
43. Fréry N, Moisan F, Schwaab Y, Garnier R. Expositions professionnelles à des agents cancérigènes respiratoires chez les salariés en 2010. *Bull Epidemiol Hebd* 2018(12-13):246-51.

## Annexe. Structure des données de la Cnav

### Base identité

NOM ZONE	FORMAT	LONG.	DÉBUT	FIN
Numéro SICAPRO	X(13)	13	1	13
Type d'enregistrement (=01)	9(2)	2	14	15
Validité (=espace)	X(4)	4	16	19
Code spécifique (=espace)	X(5)	5	20	24
Code identification	9	1	25	25
Code sexe	X	1	26	26
Date de naissance (jjmmaaaa)	9(8)	8	27	34
Code certification du décès	9	1	35	35
Date de décès (jjmmaaaa)	9(8)	8	36	43
Code INSEE de naissance	X(5)	5	44	48
Année extraction témoin	9(4)	4	49	52

### Base Salaire

NOM ZONE	FORMAT	LONG.	DÉBUT	FIN
Numéro Sicapro	X(13)	13	1	13
Type d'enregistrement (=20)	9(2)	2	14	15
Validité	X(4)	4	16	19
Code spécifique (=espace)	X(5)	5	20	24
Montant salaire brut	9(8)	8	25	32
Nombre employeurs	9(2)	2	33	34
Siret employeur 1	X(18)	18	35	52
Code activité employeur employeur 1	X(4)	4	53	56
Code Insee employeur 1	X(4)	4	57	60
Code condition emploi employeur 1	X	1	61	61
Code lieu d'activité employeur 1	X(3)	3	62	64
Siret employeur 2	X(18)	18	65	82
Code activité employeur employeur 2	X(4)	4	83	86
Code Insee employeur 2	X(4)	4	87	90
Code condition emploi employeur 2	X	1	91	91
Code lieu d'activité employeur 2	X(3)	3	92	94
Siret employeur 3	X(18)	18	95	112
Code activité employeur employeur 3	X(4)	4	113	116
Code Insee employeur 3	X(4)	4	117	120
Code condition emploi employeur 3	X	1	121	121

NOM ZONE	FORMAT	LONG.	DÉBUT	FIN
Code lieu d'activité employeur 3	X(3)	3	122	124
Montant DNT	9(8)	8	125	132
Montant chèque service	9(8)	8	133	140
Montant vignette	9(8)	8	141	148
Montant apprenti	9(8)	8	149	156

## Base retraite

NOM ZONE	FORMAT	LONG.	DÉBUT	FIN
Numéro Sicapro	X(13)	13	1	13
Type d'enregistrement (=10)	9(2)	2	14	15
Validité (=espace)	X(4)	4	16	19
Code spécifique (=espace)	X(5)	5	20	24
Code retraite de droit propre	9	1	25	25
Date d'effet du droit propre (jjmmaaaa)	9(8)	8	26	33
Code retraite de droit dérivé	9	1	34	34
Date d'effet du droit dérivé (jjmmaaaa)	9(8)	8	35	42
Montant global mensuel	9(10)v99	12	43	54

## Base trimestres autre régimes

NOM ZONE	FORMAT	LONG.	DÉBUT	FIN
Numéro Sicapro	X(13)	13	1	13
Type d'enregistrement (=25)	9(2)	2	14	15
Validité	X(4)	4	16	19
Code spécifique (=code régime)	X(5)	5	20	24
Nombre de trimestres autre régime	9	1	25	25

## Base période assimilée

NOM ZONE	FORMAT	LONG.	DÉBUT	FIN
Numéro Sicapro	X(13)	13	1	13
Type d'enregistrement (=22)	9(2)	2	14	15
Validité	X(4)	4	16	19
Code spécifique (type PA)	X(5)	5	20	24
Nombre de trimestres PA	9	1	25	25

## Base déclarations nominatives annuelles (DNA)

NOM ZONE	FORMAT	LONG.	DÉBUT	FIN
Numéro Sicapro	X(13)	13	1	13
Type d'enregistrement (=21)	9(2)	2	14	15
Validité	X(4)	4	16	19
Code spécifique (type DNA)	X(5)	5	20	24
Montant DNA	9(8)	8	25	32

## Base assurance volontaire

NOM ZONE	FORMAT	LONG.	DÉBUT	FIN
Numéro Sicapro	X(13)	13	1	13
Type d'enregistrement (=23)	9(2)	2	14	15
Validité	X(4)	4	16	19
Code spécifique (=espace)	X(5)	5	20	24
Montant assurance volontaire	9(8)	8	25	32

## Base allocation amiante

NOM ZONE	FORMAT	LONG.	DÉBUT	FIN
Numéro Sicapro	X(13)	13	1	13
Type d'enregistrement (=24)	9(2)	2	14	15
Validité	X(4)	4	16	19
Code spécifique (=espace)	X(5)	5	20	24
Montant allocation amiante	9(8)	8	25	32