

# Surveillance épidémiologique

de la mortalité et investigation  
d'agrégats spatio-temporels  
en entreprise

PRINCIPES GÉNÉRAUX ET DONNÉES NÉCESSAIRES

### AVERTISSEMENT AU LECTEUR

Le but de ce guide est de montrer comment un service de santé au travail peut mettre en place, sans trop de difficultés, le recueil systématique de données indispensables à l'analyse épidémiologique de la mortalité ou à une investigation de clusters. Il ne s'agit pas d'un manuel d'épidémiologie à l'usage des médecins du travail, comme il en existe d'excellents par ailleurs. Une analyse épidémiologique, même simple requiert une compétence et une expérience que la lecture d'un document comme celui-ci ne saurait apporter.

## Abréviations

<b>Cim</b>	Classification internationale des maladies
<b>Cnil</b>	Commission nationale de l'informatique et des libertés
<b>Francim</b>	Réseau des registres français des cancers
<b>InVS</b>	Institut de veille sanitaire
<b>INRS</b>	Institut national de recherche et de sécurité
<b>Insee</b>	Institut national de la statistique et des études économiques
<b>Inserm</b>	Institut national de la santé et de la recherche médicale
<b>SIR</b>	Standardized Incidence Ratio (Ratio standardisé d'incidence)
<b>SMR</b>	Standardized Mortality Ratio (Ratio standardisé de mortalité)
<b>IC95 %</b>	Intervalle de confiance à 95 %



# Sommaire

<b>1. Pourquoi ce guide méthodologique ?</b> .....	<b>7</b>
1.1 La place de la surveillance épidémiologique dans la santé au travail .....	7
1.2 Quelle surveillance épidémiologique ? .....	8
1.3 Contenu du guide .....	8
1.4 Objectifs et limites de ce guide .....	9
<b>2. Première partie : la surveillance épidémiologique de la mortalité en entreprise</b> .....	<b>11</b>
2.1 L'outil de base : la cohorte des salariés .....	11
2.2 Constitution de la cohorte : aspects pratiques .....	12
2.2.1 Définition de la cohorte .....	12
2.2.2 Données individuelles nécessaires .....	12
2.2.3 Limites liées à l'effectif des salariés – Cas des petites entreprises .....	14
2.3 Analyse épidémiologique de la mortalité .....	14
2.3.1 Comparaison de la mortalité dans l'entreprise à celle d'une population de référence .....	14
2.3.2 Données complémentaires .....	19
2.4 Données nécessaires : récapitulatif .....	20
<b>3. Deuxième partie : les enquêtes de "clusters" (ou agrégats spatio-temporels de cas)</b> .....	<b>21</b>
3.1 Evaluation de l'importance du problème .....	21
3.2 Constitution de la cohorte (population dont sont issus les cas observés) .....	22
3.2.1 Définition de la cohorte .....	22
3.2.2 Données indispensables sur les cas observés .....	22
3.2.3 Analyse épidémiologique de l'incidence : le ratio standardisé d'incidence (SIR) .....	22
3.2.4 Données de référence .....	22
3.3 Recherche de la cause d'un excès de cas .....	23
<b>4. Exemples</b> .....	<b>25</b>
4.1 Surveillance de la mortalité chez des hommes ayant travaillé à EDF-GDF .....	25
4.2 Etude de mortalité dans l'industrie productrice de métaux durs en France .....	26
4.2.1 Objectif de l'étude .....	26
4.2.2 Etablissement de la cohorte .....	26
4.2.3 Population de référence externe .....	26
4.2.4 Résultats .....	27

4.2.5 Etude cas – témoins complémentaire .....	27
4.2.6 Conclusion de l'étude.....	28
<b>4.3 Etude d'incidence : investigation d'une suspicion d'agrégat de cancers parmi le personnel d'un établissement de recherche .....</b>	<b>28</b>
4.3.1 Données sur la population de l'entreprise.....	28
4.3.2 Données sur les cas .....	28
4.3.3 Données de référence .....	29
4.3.4 Résultats .....	29
4.3.5 Difficultés.....	29
<b>5. Conclusion.....</b>	<b>30</b>
<b>6. Pour en savoir plus .....</b>	<b>30</b>
<b>7. Références .....</b>	<b>31</b>
<b>8. Annexes .....</b>	<b>32</b>
Annexe 1 - Procédure d'accès aux causes de mortalité .....	32
Annexe 2 - Calcul d'un ratio standardisé de mortalité ou d'incidence .....	34
8.2.1 Nombre de décès ou cas observés .....	34
8.2.2 Calcul de SMR (ou SIR) .....	34
Annexe 3 - Formule du SMR, du SIR et de leur intervalle de confiance .....	36
8.3.1 Formule du SMR et du SIR .....	36
8.3.2 Formule de l'intervalle de confiance du SMR et du SIR (niveau de confiance 95 %) .....	36
Annexe 4 - Taux de mortalité par cancer en France en 1995 et 2000.....	37
Annexe 5 - Taux d'incidence des cancers en France en 1995 et 2000.....	38
Annexe 6 - Lexique.....	39

# 1. Pourquoi ce guide méthodologique ?

## 1.1 La place de la surveillance épidémiologique dans la santé au travail

Sur le plan scientifique, l'épidémiologie est la discipline qui est en première ligne pour la connaissance des risques professionnels à l'échelle collective. Elle a pour but de décrire la répartition et la fréquence des problèmes de santé et des facteurs de risque, d'expliquer les relations entre les expositions aux nuisances et les pathologies, d'évaluer l'efficacité des actions de prévention et de réparation. Parmi les approches épidémiologiques des risques professionnels, une place à part doit être faite à la "surveillance épidémiologique".

En entreprise, la surveillance épidémiologique permet d'apporter des connaissances indispensables concernant la santé des populations de travailleurs, inaccessibles à la surveillance individuelle "ordinaire" de la médecine du travail, si bien faite soit-elle. La raison en est essentiellement de nature statistique, car la très grande majorité des problèmes de santé sont "rares" au sens statistique, et ont peu de chance d'être observés de façon suffisamment fréquente et régulière par la surveillance individuelle exercée par le médecin du travail.

Considérons, pour illustrer ce problème, l'exemple des cancers, à partir d'une étude réalisée dans une usine sidérurgique [1]. La collectivité suivie plus de 20 ans dans cette entreprise comptait plus de 5 000 hommes salariés, et y a dénombré pendant cette période 105 décès par cancer du poumon, soit 5 cas par an en moyenne, ce qui correspond, en fonction de l'âge des sujets, au nombre que l'on attend "normalement". De plus, la plupart des cas de cancer se sont produits après 20 ans de latence donc plus souvent chez des sujets retraités ; il s'est donc produit chaque année en moyenne 1 à 2 cas environ de cancer du poumon parmi les salariés en activité, c'est-à-dire soumis à la surveillance de la médecine du travail. Imaginons que dans cette entreprise, on ait en réalité eu un doublement du risque de cancer pulmonaire du fait de l'exposition à un cancérogène (une telle augmentation correspond, par exemple, à une exposition importante à l'amiante). Ceci se serait traduit, en moyenne annuelle, par la survenue de 2 à 4 cas de cancer pulmonaire parmi les salariés en activité ; en admettant (ce qui est loin d'être toujours vrai), que le médecin du travail soit effectivement informé de la survenue de la totalité de ces cancers, que pourrait-il en déduire ? Observer quelques cas de cancers du poumon une année donnée est en effet tout à fait courant, d'autant que la grande majorité des sujets considérés sont par ailleurs des fumeurs. Il est clair que dans de telles conditions, l'excès de cancer serait passé inaperçu. La seule façon de pouvoir identifier un doublement du risque est d'enregistrer systématiquement, année après année, la totalité des décès par cause, y compris chez les retraités, et de calculer le risque de cancer dans la population concernée. La mise en œuvre d'une surveillance épidémiologique de la mortalité, comme on le verra dans ce guide, peut permettre de mettre en évidence un risque élevé (risque plus élevé que dans la population).

La surveillance épidémiologique permet de répondre à diverses questions qui peuvent se poser à tout moment au médecin du travail, au responsable d'une entreprise ou aux acteurs sociaux. Des informations importantes pour la prévention proviennent de systèmes de surveillance épidémiologique :

- La production régulière de données sur la fréquence des problèmes de santé parmi le personnel d'une entreprise permet de comparer ceux-ci selon les professions, les secteurs ou les ateliers, d'observer éventuellement un risque plus élevé de telle pathologie parmi une catégorie particulière de salariés, de rendre compte de l'augmentation de tel problème au fil des années.
- Il est possible également de vérifier si l'exposition à une nuisance connue est bien contrôlée, en étudiant systématiquement l'incidence des maladies ou les causes médicales de décès parmi les personnels exposés à cette nuisance.
- Lorsqu'il arrive qu'on ait l'attention attirée par une fréquence qui semble excessive d'un certain problème de santé (plusieurs cas de cancer rapprochés dans une petite collectivité de salariés, par exemple), la vérification de l'existence d'un véritable excès n'est possible qu'en employant des données et des méthodes relevant de la surveillance épidémiologique (comparaison de la fréquence de ce problème dans l'entreprise à celle d'une population de référence).

## 1.2 Quelle surveillance épidémiologique ?

Les techniques de surveillance épidémiologique sont variées, allant de la simple analyse épidémiologique de données enregistrées en routine (cf. § 4.3 *Etude d'incidence – Investigation d'un agrégat de cancers dans un établissement de recherche*), jusqu'à des systèmes de surveillance ad hoc pouvant être très sophistiqués, et impliquant des ressources matérielles et méthodologiques importantes, particulièrement lorsqu'on souhaite prendre en compte les effets des conditions de travail et des expositions à des nuisances variées (cf. § 4.1 *Surveillance de la mortalité chez des personnels d'EDF-GDF*).

Cependant, une façon simple et économique permettant de répondre rapidement à de nombreuses questions consiste à mettre en place une surveillance épidémiologique de la mortalité. En effet, du fait de l'existence d'un recueil exhaustif des causes médicales des décès survenant en France, et d'un système formalisé d'accès aux données de mortalité, il est possible dans toute entreprise, de connaître la mortalité par cause des salariés (y compris des personnes ayant changé d'entreprise ou retraitées) et de la comparer à la population de référence adéquate. Aujourd'hui, pratiquement inexistante dans les entreprises françaises y compris les plus importantes (alors qu'elle est une pratique courante dans la plupart des pays développés), cette méthode privilégiée ne nécessitant que des moyens accessibles à chaque service de santé au travail, mérite incontestablement de devenir une activité régulière de ceux-ci. Cette surveillance épidémiologique de la mortalité des personnels d'une entreprise, qui constitue une approche systématique, devrait permettre au médecin de repérer des événements de santé anormaux dans une entreprise, et de répondre le cas échéant à diverses questions qui peuvent lui être posées.

Rien ne s'oppose, en principe, à la mise en place de systèmes de surveillance comparables pour l'incidence de maladies ne se traduisant pas par un décès. Cependant, on ne dispose pas dans notre pays, sauf exception, de système national de recueil de données d'incidence de maladies identique à ce qui existe pour la mortalité, ce qui empêche la mise en œuvre de systèmes de surveillance épidémiologique simples dans la plupart des situations. Néanmoins, il peut arriver que dans des cas particuliers, par exemple si l'entreprise se situe dans un département disposant d'un Registre des cancers<sup>1</sup>, il soit possible d'envisager de construire des systèmes de surveillance épidémiologique de l'incidence de maladies en tous points comparables à ce qui est présenté ici pour la mortalité, bien que l'exhaustivité ne soit pas garantie dans le cas de l'incidence contrairement à la mortalité. Dans ce guide, on emploiera les termes de "mortalité" ou "d'incidence", selon qu'il s'agit de décès ou de cas de maladie lorsque les données nécessaires sont disponibles.

## 1.3 Contenu du guide

Ce guide méthodologique comporte deux parties principales :

- on envisagera tout d'abord les principes de la mise en place d'une surveillance épidémiologique systématique de la mortalité en entreprise ;
- on décrira aussi le principe des enquêtes de "clusters"<sup>2</sup> (agrégats spatio-temporels de cas parmi le personnel d'une entreprise), destinées à vérifier l'existence d'un excès de cas et le cas échéant à en identifier la cause.

Ces principes seront illustrés par des exemples : l'analyse épidémiologique de la mortalité parmi le personnel d'Electricité De France - Gaz De France, l'étude de mortalité par cancers broncho-pulmonaires parmi les salariés de l'industrie productrice de métaux durs en France, et l'analyse d'un cluster de cas de cancers dans un établissement de recherche.

<sup>1</sup> Un registre est une structure épidémiologique enregistrant en permanence et de façon exhaustive tous les cas d'une maladie se produisant dans un territoire donné (le département le plus souvent en France). La plupart des registres concernent le cancer ; certains pays disposent de registres nationaux des cancers, ce qui n'est pas le cas de la France.

<sup>2</sup> Un "cluster", mot anglais signifiant agrégat, est un regroupement inhabituel d'un problème de santé dans un espace – temps défini. On parle aussi "d'agrégat spatio-temporel".

## 1.4 Objectifs et limites de ce guide

L'exercice d'une surveillance de la mortalité en entreprise et la réalisation d'enquêtes de "clusters" sont des activités de nature épidémiologique. Leur mise en œuvre implique deux étapes successives :

- le recueil des données nécessaires à l'analyse de la mortalité et à l'investigation de clusters ;
- l'analyse épidémiologique proprement dite.

La première étape de recueil de données peut (devrait) se faire au niveau de l'entreprise et/ou à celui du service de santé au travail. Elle ne nécessite que des moyens simples, mobilisant uniquement des informations facilement disponibles à ce niveau, et est accessible à tout service de santé au travail.

La seconde étape, qui concerne l'analyse épidémiologique, met en jeu des compétences et des moyens spécialisés, qui sont l'apanage de professionnels de l'épidémiologie. Sauf exception, elle n'est pas à la portée du médecin du travail, particulièrement lorsque celui-ci est isolé. De plus, les résultats d'une analyse par des personnes non conscientes des pièges associés aux méthodes utilisées (voir plus loin "healthy worker effect") pourraient apparaître faussement rassurants.

Le but de ce guide est de montrer comment un service de santé au travail peut mettre en place, sans trop de difficultés, le recueil systématique de données permettant l'analyse épidémiologique de la mortalité ou l'investigation de clusters.

Cette analyse pourra ensuite, en coopération avec le médecin du travail, être réalisée par des équipes spécialisées : Département santé travail de l'InVS, Département épidémiologie en entreprise de l'INRS, Instituts de médecine du travail, équipes Inserm, etc.

Il ne s'agit donc pas ici d'un "manuel" d'épidémiologie à l'usage du médecin du travail, comme il en existe d'excellents<sup>3</sup>. En effet, l'analyse épidémiologique, même "simple", requiert une compétence et une expérience qui n'existent qu'au sein de structures spécialisées, et qu'il serait illusoire (et dangereux !) de penser acquérir par la lecture d'un guide comme celui-ci. **Par contre, la mise en place d'un recueil de données de routine permettant la réalisation de ces analyses ne met en jeu que les techniques simples décrites ici.**

<sup>3</sup> Voir : § 6 Pour en savoir plus.



## 2. Première partie : la surveillance épidémiologique de la mortalité en entreprise

### 2.1 L'outil de base : la cohorte des salariés

La surveillance épidémiologique de la mortalité<sup>4</sup> repose sur la mise en place et le suivi de « cohortes ». Dans sa version la plus simple, suffisante pour assurer les fonctions de base de la surveillance épidémiologique, la mise en œuvre en est très aisée. On en présente ici les principes généraux, des éléments plus précis permettant de mettre en place une cohorte d'entreprise de façon opérationnelle étant présentés plus loin.

**Il s'agit tout simplement d'établir un fichier avec la liste exhaustive de tous les salariés de l'entreprise à une date donnée**, avec un minimum de données concernant chacun d'eux : des informations aussi élémentaires et accessibles que la date et le lieu de naissance, le genre, le poste occupé, sont suffisantes pour initialiser la cohorte. Puis, année après année, on met à jour le fichier de la cohorte en y incluant les nouveaux embauchés de l'année, et en actualisant les données pour chaque sujet : poste occupé, sortie de l'entreprise, etc. On peut initialiser la cohorte d'une entreprise à partir d'une année donnée. Mais la cohorte sera d'autant plus utile qu'on aura pu la faire débiter le plus loin possible dans le passé, à partir des fichiers du service du personnel, des archives de la société, etc., créant ainsi une "cohorte historique", reconstituée rétrospectivement à partir d'une date fixée dans le passé (par exemple, à partir du 01/01/1970).

Une fois la cohorte constituée, il est aisé de retrouver, avec la périodicité que l'on souhaite, tous les sujets décédés (y compris ceux qui ont quitté l'entreprise, même depuis très longtemps), et pour ceux-ci, la cause médicale du décès, en utilisant des procédures qui seront décrites plus loin. Ces procédures reposent sur l'existence du recueil exhaustif et permanent de données de mortalité en France, qui permet à l'Institut national de la santé et de la recherche médicale (Inserm) de les centraliser, de les coder et de les enregistrer dans une base de données nationale.

Ces opérations, qui ne nécessitent que des données disponibles en routine dans les services de santé au travail ou dans les services du personnel de l'entreprise, sont suffisantes pour établir et mettre à jour la cohorte.

L'existence d'une cohorte permet de répondre aux deux besoins fondamentaux de la surveillance d'une collectivité de salariés qui ont été évoqués en préambule à ce guide :

- *Analyse descriptive de la mortalité*, permettant la connaissance de la fréquence des décès et de leurs causes médicales parmi le personnel d'une entreprise (incluant les anciens travailleurs de celle-ci), selon les professions, les secteurs ou les ateliers, ainsi que de l'évolution dans le temps de ces fréquences et la recherche de l'existence éventuelle d'un excès de mortalité par comparaison à une population de référence adéquate.
- *Investigation d'un "cluster"*, c'est-à-dire la détection ou la vérification de l'existence d'un excès supposé de cas, par comparaison du nombre de cas observé dans la cohorte au nombre « attendu » si aucun excès n'existait.

Dans la suite de ce guide, on détaillera les procédures à mettre en œuvre pour ces deux volets de la surveillance épidémiologique en entreprise, l'analyse descriptive de la mortalité et l'analyse de clusters de cas.

<sup>4</sup> Comme on l'a signalé, tout ce qui concerne la mortalité peut s'appliquer à l'incidence si l'on dispose des données correspondantes.

## 2.2 Constitution de la cohorte : aspects pratiques

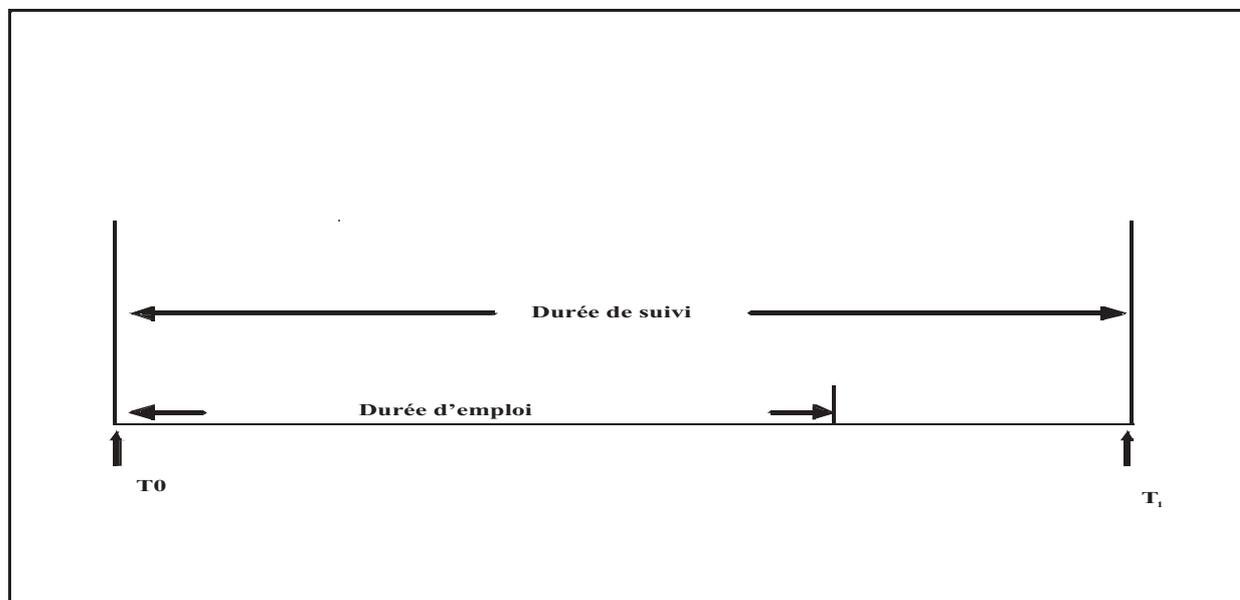
### 2.2.1 Définition de la cohorte

La cohorte étudiée doit être définie avec précision.

**Cohorte** : c'est un groupe de sujets tous vivants au début de la période d'observation ; il s'agit habituellement de **l'ensemble des personnes qui travaillent ou ont travaillé dans l'entreprise**. Il peut s'agir d'une cohorte "historique", c'est-à-dire d'une cohorte reconstituée rétrospectivement à partir d'une date fixée dans le passé grâce aux archives du service du personnel de l'entreprise. C'est dans ce groupe de sujets que seront observés tous les décès (ou les cas incidents) pendant le suivi.

**Période d'observation (ou de "suivi")** : c'est la période (plusieurs années ou décennies le plus souvent) pendant laquelle la cohorte fera l'objet d'un recueil de données concernant le statut vital et les causes de décès. On appelle souvent  $t_0$  l'année de début, et  $t_1$  l'année de fin d'observation. Il faut distinguer la période de suivi de la période d'emploi des salariés inclus dans la cohorte, comme le montre la figure 1. Dans cet exemple, une cohorte de travailleurs ayant été employée dans une entreprise pendant une certaine durée fait l'objet d'une analyse de mortalité pour une période plus longue que leur emploi : ceci permet d'inclure les décès parmi les salariés retraités.

Figure 1. Suivi des sujets inclus dans une cohorte - Phénomènes liés au temps



### 2.2.2 Données individuelles nécessaires

#### 2.2.2.1 Pour la recherche du statut vital et des causes de décès

Il existe une procédure, codifiée par un décret (décret 98-37 du 16 janvier 1998), qui précise les modalités d'accès aux données permettant de connaître le statut vital des personnes incluses dans une cohorte, et les causes médicales de décès le cas échéant. Cette procédure, décrite en détail dans l'annexe 1, permet à l'investigateur qui dispose du fichier de la population étudiée (données individuelles indispensables détaillées plus bas), d'obtenir le statut vital des personnes, puis la cause de décès des personnes décédées. Cette procédure, qui nécessite une autorisation de la Commission nationale de l'informatique et des libertés (Cnil), est conçue afin qu'aucun fichier informatique ne contienne simultanément l'identité des personnes et leur cause de décès, afin de garantir la confidentialité des données.

Pour y avoir recours, il est indispensable de disposer de l'exhaustivité des effectifs de la cohorte durant la période d'emploi de chaque sujet, année par année. Pour chacun de ces sujets, un certain nombre d'informations d'état civil sont indispensables et doivent être incluses dans le fichier de la cohorte.

**DONNEES INDISPENSABLES À INCLURE POUR CHAQUE SUJET DE LA COHORTE  
POUR LA RECHERCHE DU STATUT VITAL ET DES CAUSES DE DECES**

**Nom patronymique** (c'est-à-dire **nom de jeune fille** pour les femmes)

**Prénom**

**Genre**

**Jour, mois et année de naissance**

**Commune et département de naissance** (pays pour les personnes nées à l'étranger)

Remarque : disposer des premiers chiffres du numéro Insee ne suffit pas, car il ne comporte pas le jour de naissance.

### **2.2.2.2 Pour analyser la mortalité quel que soit l'emploi exercé**

Pour effectuer les calculs des indices décrivant la mortalité d'une cohorte, quel que soit l'emploi exercé, il est nécessaire de connaître à quel moment de la période de suivi le sujet est entré et sorti de l'entreprise.

Les données indispensables pour décrire la mortalité d'une cohorte sont, d'une part les données permettant la recherche du statut vital et des causes de décès, et d'autre part, les dates d'embauche et de sortie des sujets de l'entreprise.

**DONNEES INDISPENSABLES A INCLURE  
POUR LE CALCUL DES INDICES DECRIVANT LA MORTALITE D'UNE COHORTE**

**Date d'embauche**

**Date de sortie de l'entreprise**

Ces informations sont en principe disponibles auprès des services du personnel des entreprises (ou leur équivalent).

### **2.2.2.3 Pour analyser la mortalité selon les caractéristiques de l'emploi**

Pour analyser la mortalité de façon plus précise, il est nécessaire de disposer de quelques données supplémentaires concernant les emplois de chaque sujet inclus dans la cohorte. **Au minimum, il s'agit des emplois occupés au cours de la carrière au sein de l'entreprise, avec les dates correspondantes.** À cet historique de carrière au sein de l'entreprise pourront être ajoutées, si elles sont disponibles, des informations plus détaillées concernant les postes de travail occupés, les ateliers ou l'exposition à des nuisances particulières.

**DONNEES A INCLURE POUR CHAQUE SUJET DE LA COHORTE POUR L'ANALYSE DE LA  
MORTALITE SELON LES CARACTERISTIQUES DE L'EMPLOI**

**Postes occupés : date de début et fin pour chaque poste depuis l'entrée dans l'entreprise  
(on pourra utiliser les intitulés de poste propres à l'entreprise ou des intitulés de fonction)**

*Ateliers ou secteurs d'affectation\* : date de début et de fin pour chaque atelier ou secteur  
(on utilisera les intitulés propres à l'entreprise)*

*Expositions à des nuisances\**

*Catégorie socioprofessionnelle (PCS de l'Insee)\**

*\*Facultatif*

Les services du personnel des entreprises sont en mesure de fournir l'ensemble de ces données, qui constituent une information de base. Toute entreprise est en effet tenue de transmettre à l'Urssaf chaque année les *déclarations annuelles des données sociales* (DADS), où figurent notamment la catégorie socioprofessionnelle, le genre, l'âge, le département de naissance, la nationalité (française ou étrangère), le temps de travail, les dates de début et de fin d'emploi.

### 2.2.3 Limites liées à l'effectif des salariés – Cas des petites entreprises

Le suivi épidémiologique de la mortalité des travailleurs d'une entreprise est une activité qui devrait faire partie de la surveillance médicale régulière et systématique, en association entre les services de santé au travail et des équipes d'épidémiologistes spécialisées. Elle se heurte cependant à des difficultés dans le cas des petites entreprises du fait de la faiblesse des effectifs de salariés. Il est évident que ce qui précède n'a pas beaucoup de sens dans un tel contexte.

Il serait tout à fait intéressant, dans ces cas, que les services de santé au travail s'organisent pour procéder à des regroupements, qui peuvent se faire sur divers critères : géographique, par branche, etc. Les services interentreprises sont souvent particulièrement bien placés pour une telle mutualisation de données et de ressources, et pourraient sans trop de difficultés constituer des cohortes impliquant les salariés de diverses entreprises, avec l'aide d'équipes d'épidémiologie spécialisées.

## 2.3 Analyse épidémiologique de la mortalité

Comme on l'a souligné plus haut, l'analyse épidémiologique de la mortalité des salariés d'une entreprise implique des compétences spécialisées, et sauf exception, le médecin du travail n'aura pas à réaliser lui-même de telles analyses qui mettent en jeu des méthodes qui sont souvent complexes.

Il est cependant important qu'il en connaisse les éléments essentiels qui lui permettront de communiquer dans de bonnes conditions à la fois avec ses partenaires à l'intérieur de l'entreprise (direction, CHSCT, salariés, ...) et avec les épidémiologistes. C'est le but de ce chapitre, qui présente les concepts de base de l'analyse de la mortalité.

### 2.3.1 Comparaison de la mortalité dans l'entreprise à celle d'une population de référence

#### 2.3.1.1 La nécessité de méthodes épidémiologiques d'analyse de la mortalité

Les buts essentiels de l'analyse épidémiologique de la mortalité des travailleurs d'une entreprise sont la description de la fréquence des décès et leurs causes (maladies ou accidents), afin de comparer ceux-ci selon les professions, les secteurs ou les ateliers, permettant éventuellement d'observer un risque plus élevé de telle pathologie parmi une catégorie particulière de salariés, de rendre compte de l'évolution au cours du temps ou de vérifier si l'exposition à une nuisance provoque des maladies pouvant se traduire par des décès parmi les personnels exposés.

Le simple comptage du nombre de décès (toutes causes confondues ou selon les causes médicales) parmi les salariés d'une entreprise ne permet pas de répondre à ces objectifs, pour diverses raisons :

- Le nombre de décès dans une collectivité dépend de nombreux facteurs propres à celle-ci, notamment de l'effectif des personnels, de leur âge et du genre. Il existe une très forte variabilité de la mortalité selon tous ces facteurs. Ainsi, un véritable excès de décès lié à des risques spécifiques de l'entreprise peut exister et passer inaperçu si l'effectif est peu important et l'âge moyen peu élevé ; inversement, si la population de l'entreprise est âgée, on observera des décès qui peuvent paraître nombreux, alors que leur fréquence peut être "normale" à cause de l'âge moyen des sujets.
- Si l'exposition à des substances toxiques ou à des conditions de travail défavorables, s'accompagne d'effets retardés à long terme (plusieurs décennies parfois, comme pour certains cancérigènes professionnels), la majorité des décès ainsi occasionnés se produira alors que les personnes concernées ont quitté l'entreprise, soit du fait de leur maladie, soit surtout parce qu'elles sont parties en inactivité : le médecin du travail n'est habituellement pas informé de tels décès pouvant survenir très longtemps après la sortie de l'emploi. Même s'il peut avoir connaissance de certains cas, il ne peut s'assurer que d'autres ne se seraient pas produits. Il n'a pas non plus la possibilité de vérifier de façon formelle la cause du décès, ce qui demande souvent l'accès à des documents médicaux comme des comptes-rendus opératoires ou des examens d'anatomopathologie.

Les méthodes d'analyse épidémiologique de la mortalité permettent de surmonter ces difficultés, malgré certaines limites.

### 2.3.1.2 L'indicateur de base de l'étude épidémiologique de la mortalité d'une cohorte des salariés : le ratio standardisé de mortalité (SMR)

En épidémiologie, lorsqu'on veut étudier la mortalité d'une population particulière (la cohorte des salariés, dans ce cas), il est toujours nécessaire de **comparer** celle-ci à d'autres populations, car il n'existe pas dans l'absolu de norme de "bonne santé" d'une population. Ce qu'on cherche à savoir, c'est si dans cette population particulière, on meurt plus précocement ou plus souvent de certaines maladies. **L'analyse épidémiologique de la mortalité des salariés d'une entreprise consiste donc à comparer celle-ci à la mortalité d'une population de référence adéquate.**

Pour cela, on utilise une méthode qui consiste à comparer le **nombre de décès attendus** dans la cohorte de l'entreprise si la mortalité était identique à celle de la population de référence à laquelle on la compare, **au nombre de décès qui ont été réellement observés** dans l'entreprise.

La mortalité d'une population donnée étant essentiellement déterminée par sa structure démographique (répartition des classes d'âge et des genres), le nombre de décès attendus est calculé en appliquant les taux de mortalité de la population de référence aux effectifs de la population étudiée **par classe d'âge** (méthode de "standardisation dite indirecte"). La mortalité d'une population évoluant avec le temps, et la période de suivi s'étalant parfois sur plusieurs décennies, **le calcul du nombre de décès attendus doit être fait année par année, tout au long de la période de suivi.** Les taux de mortalité des hommes et des femmes étant différents, ces calculs doivent toujours être réalisés **séparément pour chaque genre** (voir exemple plus bas).

Le rapport du nombre (cumulé année après année tout au long de la période de suivi) des décès observés sur le nombre des décès attendus est appelé **ratio standardisé de mortalité** (on emploie souvent le terme SMR, du fait des initiales anglaises : *Standardized Mortality Ratio*).

Remarque : si, au lieu de la mortalité, on analyse l'incidence d'une maladie (cancer par exemple) pour laquelle les données nécessaires sont disponibles, on parle de ratio standardisé d'incidence (ou SIR : *Standardized Incidence Ratio*). Le principe et les méthodes de calcul sont strictement identiques à ceux décrits ici pour la mortalité, les "nouveaux cas" étant analysés à la place des "décès".

La méthode de calcul du SMR (et du SIR) est décrite en détail dans l'annexe 2. Les formules du SMR, du SIR et de leur intervalle de confiance figurent dans l'annexe 3.

#### RESULTATS

##### *(Valeurs du SMR et de son intervalle de confiance)*

##### **Valeur du SMR**

Si le SMR vaut 1 : la mortalité de la cohorte ne diffère pas de celle de la population de référence (absence de risque).

S'il est inférieur à 1 : la mortalité est plus faible dans la cohorte.

S'il est supérieur à 1 : la mortalité est plus élevée dans la cohorte.

Un SMR égal à 0,8 signifie que la mortalité dans l'entreprise est inférieure de 20 % à celle de la population de référence. Un SMR égal à 3 signifie que l'entreprise a une mortalité trois fois supérieure à celle de la population de référence.

##### **Intervalle de confiance du ratio**

La valeur 1 signifiant qu'il n'existe pas de différence observée entre la mortalité de la cohorte et celle de la population de référence, il faut vérifier si le SMR calculé est sur le plan statistique "significativement" supérieur ou inférieur à la valeur 1. En effet, le hasard peut faire que la valeur calculée soit plus élevée ou plus basse que 1. Pour juger si l'on peut considérer avec suffisamment de confiance que la mortalité de la cohorte est équivalente, plus faible ou plus élevée, on calcule "l'intervalle de confiance" du SMR, qui se présente sous la forme d'une plage.

Par exemple : SMR = 1,95, intervalle de confiance : 0,90 – 2,80.

Si l'intervalle de confiance du SMR contient la valeur 1 (comme dans l'exemple ci-dessus), on dit que le SMR est statistiquement non significatif, c'est-à-dire que la mortalité observée dans la cohorte ne peut être considérée comme différente de celle de la population de référence.

Si l'intervalle de confiance ne contient pas 1, on dit que le SMR est statistiquement significatif et différent de 1 (plus faible ou plus élevé).

### 2.3.1.3 Taille de la cohorte et puissance statistique

Même si le calcul d'un SMR est toujours possible, il faut rester prudent dans l'interprétation des résultats dans le cas d'une cohorte de petite taille. La puissance du test statistique (c'est-à-dire la possibilité de mettre en évidence un excès de risque qui existerait) dans une étude de cohorte dépend du nombre de "personnes-années" analysées, c'est-à-dire du nombre de personnes de la cohorte et du nombre d'années de suivi. Ainsi, un petit nombre de sujets suivi pendant quelques années entraînera une puissance statistique faible. Si le SMR d'une telle analyse n'est pas significativement différent de 1, on ne pourra pas exclure qu'un excès de risque n'a pas pu être détecté par manque de puissance statistique.

### 2.3.1.4 Variabilité du SMR en fonction de la structure de la population (répartition par âge et par genre)

Les simulations suivantes (cf encadré page 17) illustrent la variabilité de la valeur du SMR en fonction de la structure démographique de la cohorte. Les valeurs des SMR, basées sur un même nombre de cas observés dans une cohorte (6 cas observés parmi 450 personnes) diffèrent selon la répartition des salariés dans les différentes classes d'âge, et selon leur genre (on se reportera à l'annexe 2 pour comprendre le détail des calculs de SMR).

**Un calcul "global" de SMR sans tenir compte ni de la structure d'âge de la cohorte de salariés, ni de l'âge ou du genre des cas, ni du genre des salariés, serait donc faux et ne peut en aucun cas être envisagé.**

### 2.3.1.5 Quelle population de référence ? Le problème du "Healthy worker effect" et des facteurs de confusion

La population de référence habituellement utilisée comme population de comparaison est la **population générale**, c'est-à-dire l'ensemble des habitants d'une zone géographique donnée (la France, par exemple). Ceci pose deux problèmes principaux, souvent rencontrés dans les études de cohortes en milieu professionnel :

- **"l'effet du travailleur en bonne santé"** ou Healthy worker effect [2]. En effet, on observe presque toujours dans les cohortes professionnelles une mortalité toutes causes inférieure à celle de la population générale (et donc des valeurs de SMR inférieures à 1). Ceci reflète le fait que les populations de travailleurs ne sont pas strictement comparables à la population générale du même âge et du même genre. En effet, seuls des individus en assez bonne santé sont embauchés dans une entreprise et restent à des postes soumis à des expositions professionnelles fortes, ce qui écarte les sujets souffrant de problèmes de santé et donc plus à risque de décéder. En général, ce "biais de sélection" diminue avec la durée de suivi, et la mortalité toutes causes de travailleurs après 15 ans de suivi est très proche de celle de la population générale. Ceci pose évidemment un problème difficile pour l'analyse épidémiologique, car la sous-mortalité habituelle observée dans les cohortes de salariés d'entreprise peut masquer de véritables problèmes. Ainsi, dans une cohorte où un agent cancérigène entraînerait un surcroît de cancers parmi les sujets exposés, l'augmentation du risque de cancer peut tout à fait passer inaperçue si elle n'est pas extrêmement élevée, car elle peut ne pas suffire à "compenser" le déficit de mortalité dû à l'effet du travailleur en bonne santé.

Ainsi, le tableau 1, extrait d'une étude de cohorte formée des travailleurs d'une usine de production de polychlorure de vinyle [3], montre parmi ceux-ci un net déficit de mortalité, surtout en début de suivi (de 0 à 4 ans de suivi).

Tableau 1

Cause de décès	SMR	
	Durée de suivi (ans)	
	0-4	>15
Maladie Cardiovasculaire	0.21	0.91
Cancer	0.45	1.12
Maladie Respiratoire	0.21	0.93
Toutes causes	0.37	0.94

D'après Fox et Collier (1976)

## Calcul du SMR dans deux populations d'hommes présentant des structures d'âge différentes

**Situation 1** - 6 décès observés dans une cohorte de 450 hommes

Classes d'âge (ans)	Nombre de salariés dans la cohorte	Nombre de décès observés dans la cohorte	Taux de mortalité de la population de référence (/100 000)	Nombre de décès attendus dans la cohorte
20-24	28	1	30	0,0084
25-29	34	1	42,8	0,0146
30-34	80	0	62	0,0496
35-39	33	1	100,8	0,0333
40-44	85	1	186,3	0,1584
45-49	75	2	353,3	0,2650
50-54	90	0	575	0,5175
55-59	25	0	910,8	0,2277
	450	6		1,2743

$$\text{SMR} = 6/1,27 = 4,71$$

Intervalle de confiance IC95 % = [1,72 – 10,25].

On observe un SMR significativement plus élevé que 1. Un excès de risque dans la cohorte par rapport à la population de référence est mis en évidence.

**Situation 2** - 6 décès observés parmi 450 hommes, dans une cohorte ayant une structure d'âge différente et des salariés décédés à un âge différent de la situation n°1.

Classes d'âge (ans)	Nombre de salariés dans la cohorte	Nombre de décès observés dans la cohorte	Taux de mortalité de la population de référence (/100 000)	Nombre de décès attendus dans la cohorte
20-24	5	0	30	0,0015
25-29	34	1	42,8	0,0146
30-34	5	0	62	0,0031
35-39	32	1	100,8	0,0323
40-44	65	1	186,3	0,1211
45-49	15	0	353,3	0,0530
50-54	204	2	575	1,1730
55-59	90	1	910,8	0,8197
	450	6		2,2182

$$\text{SMR} = 6/2,22 = 2,70 ; \text{IC95 \%} = [0,99 - 5,89].$$

Le SMR est supérieur à 1, mais l'intervalle de confiance comprend la valeur 1, aucun excès de risque n'est donc mis en évidence.

Soit, pour 6 décès dans la population d'intérêt de 450 hommes, des résultats différents sont observés selon la structure d'âge de la population et l'âge des personnes décédées.

### Calcul du SMR parmi les femmes

Dans la situation d'une cohorte exclusivement féminine (450 femmes), réparties selon les mêmes classes d'âge que ci-dessus, et dans laquelle 6 cas auraient été notifiés, les valeurs des SMR calculés seraient différentes, car les taux de mortalité de référence sont différents.

Situation 1 :  $\text{SMR} = 6/1,34 = 4,49$  ;  $\text{IC95 \%} = [1,64 - 9,77]$ .

Situation 2 :  $\text{SMR} = 6/1,99 = 3,00$  ;  $\text{IC95 \%} = [1,10 - 6,54]$ .

Ici, dans les deux situations, un excès de risque est mis en évidence.

- Le fait que certains «**facteurs de confusion**» sont souvent plus fréquents parmi les travailleurs d'une entreprise qu'au sein de la population générale, des facteurs pathogènes comme des conditions de travail difficiles, la consommation de tabac et d'alcool ou de mauvaises habitudes alimentaires sont répartis de façon inégale parmi les catégories sociales, et plus fréquents parmi les travailleurs manuels que non manuels [4]. Il peut en résulter un "effet de confusion positif" qui fait apparaître un risque de mortalité plus élevé au sein de la cohorte que dans la population générale de référence, qui ne doit alors pas être attribué à une exposition professionnelle particulière.

Pour remédier à ces difficultés dues au choix de la population de référence, deux approches principales sont utilisées.

- **Utilisation d'une population générale de référence externe proche de celle de la cohorte étudiée.** Ainsi, dans certaines études épidémiologiques, la population de la zone géographique où est implantée l'entreprise est choisie comme population de référence, de préférence à la population nationale, car les taux de mortalité et les modes de vie sont généralement plus proches au sein de la même zone géographique. L'exemple suivant est tiré d'une étude concernant le risque de cancer du poumon dans une cohorte composée d'ouvriers de l'industrie métallurgique utilisant du nickel. Le tableau 2 montre les SMR pour les cancers respiratoires obtenus en comparant les sujets de la cohorte respectivement avec la population des USA, avec celle de l'État où est située l'entreprise étudiée, et avec celle des six comtés voisins d'où provenait la grande majorité des travailleurs inclus dans la cohorte : alors que le SMR obtenu avec les taux nationaux s'avère élevé de façon significative, ce n'est pas le cas avec les taux locaux [5].

Tableau 2

	Population de référence		
	USA	État	Comtés
SMR	1,3	1,1	0,93

*D'après Enterline et al. (1982)*

- **Comparaisons internes à la cohorte** : au lieu de comparer la mortalité de la cohorte à celle d'une population générale de référence externe, les épidémiologistes procèdent à des comparaisons au sein de la même cohorte (si elle a une taille suffisante), entre les salariés exerçant une profession donnée et les autres, entre les sujets travaillant dans un secteur et les autres salariés, entre les sujets exposés à un facteur de risque et les salariés non exposés. Puisque les travailleurs d'une même entreprise sont généralement sélectionnés d'une manière comparable, on s'affranchit ainsi du biais de sélection que constitue l'effet du travailleur en bonne santé. On contrôle aussi indirectement sur de nombreux facteurs de confusion socio-économiques ou liés aux modes de vie, les travailleurs d'une même entreprise étant souvent plus proches entre eux que la population générale.

Cependant, les méthodes d'analyse où l'on compare la mortalité d'un sous-ensemble de la cohorte à d'autres sujets de la cohorte sont habituellement tout à fait différentes de celles qui ont été présentées ci-dessus, et mobilisent des modèles statistiques relativement complexes<sup>5</sup>, qui ne seront pas abordés dans ce guide.

**Au total**, le choix d'une population de référence (population générale nationale ou régionale, cohortes de travailleurs) dépend du problème posé, est toujours difficile, a un impact réel sur les résultats de l'étude [6] et les données de référence "idéales" sont rarement disponibles. Les épidémiologistes spécialisés sont familiers de ces difficultés, et sont à même de faire les choix les plus appropriés.

### 2.3.1.6 Données indispensables pour le calcul d'un SMR

Les données indispensables pour ces calculs sont **exclusivement des données administratives** (aucune donnée médicale complémentaire nécessaire), avec d'une part des données sur la population de l'entreprise (cohorte d'où sont issus les décès) et d'autre part, des données sur la population de référence.

<sup>5</sup> Les deux principales méthodes utilisées dans ces circonstances sont la régression de Poisson et le modèle de Cox [2].

- Les données sur la population de l'entreprise sont celles de la cohorte qui ont été décrites plus haut :

**Nom patronymique** (c'est-à-dire **nom de jeune fille** pour les femmes)

**Prénom**

**Genre**

**Jour, mois et année de naissance**

**Commune et département de naissance** (pays pour les personnes nées à l'étranger)

**Date d'embauche**

**Date de sortie de l'entreprise**

- Les données de référence sont disponibles grâce à l'enregistrement systématique de la mortalité de la population française [7] : il est ainsi possible de disposer des taux de mortalité indispensables au calcul des décès «attendus» (taux nationaux, régionaux ou départementaux toutes causes et par causes médicales de décès dans la population générale). À titre d'illustration, les taux de mortalité par cancers en 1995 et 2000 sont présentés en annexe 4 (par genre et classes d'âge).

Comme on l'a expliqué plus haut, il est aussi parfois possible dans les cohortes de taille importante de choisir une catégorie de travailleurs comme référence «interne» pour la mortalité.

## 2.3.2 Données complémentaires

### 2.3.2.1 Histoire professionnelle

Lorsque l'on veut effectuer des comparaisons de mortalité (ou morbidité) au sein d'une entreprise (par exemple, comparer les causes de décès en fonction de la catégorie socioprofessionnelle ou du service), il est nécessaire de recueillir en plus des données de base citées ci-dessus, des informations relatives à l'activité des personnes.

Pour cela, une histoire professionnelle avec les métiers occupés au cours de la carrière professionnelle des sujets inclus dans la cohorte, avec les dates correspondantes, doit être reconstituée (voir l'exemple plus bas : § 4.1 Surveillance de la mortalité chez des hommes ayant travaillé à EDF-GDF). S'il n'est pas possible de reconstituer l'historique de carrière complet des sujets, il faut au minimum des données sur le service et le métier occupés au cours du dernier épisode professionnel.

### 2.3.2.2 Facteurs de confusion

Dans certains cas, l'analyse de la mortalité a pour objectif de vérifier si l'exposition à un agent de l'environnement professionnel potentiellement toxique est à l'origine d'une augmentation du risque de développer une maladie. On a déjà évoqué plus haut les difficultés que peut alors causer l'existence de facteurs de confusion, c'est-à-dire de facteurs qui sont associés à l'exposition et liés causalement à la maladie étudiée. Un exemple typique fréquemment rencontré est celui du tabagisme, à l'origine de nombreux problèmes de santé, et qui est habituellement plus fréquent parmi les travailleurs d'une cohorte d'entreprise que dans la population générale. Si on s'intéresse à un agent toxique pour l'appareil respiratoire, par exemple, et que le tabac n'est pas pris en compte dans les analyses, ceci peut être à l'origine d'un biais statistique qui montrerait à tort un excès de maladies dans la cohorte comparativement à la population de référence.

La meilleure façon de prendre en compte ce problème dans les analyses épidémiologiques de la mortalité est d'utiliser des méthodes statistiques dites "multivariées" qui tiennent compte à la fois de l'exposition à tous les facteurs de risque étudiés et aux facteurs de confusion. Ceci permet de calculer le risque de maladie lié au facteur professionnel d'intérêt à niveau égal d'exposition aux facteurs de confusion : si on observe un excès associé au facteur professionnel, il ne pourra alors pas être dû aux facteurs de confusion.

Cependant, pour procéder de la sorte, il faut disposer de données appropriées sur les expositions aux facteurs de confusion pertinents. Lorsque l'on veut étudier, par exemple, la mortalité par cancer broncho-pulmonaire, il est recommandé de recueillir l'information sur les habitudes tabagiques de tous les sujets de la cohorte. S'il est parfois possible d'obtenir ces données à partir des dossiers médicaux des salariés, il existe de nombreuses situations où cela n'est pas faisable, notamment lorsque la cohorte est reconstituée sur une longue période ancienne, et que les dossiers médicaux ne sont pas disponibles sur cette période ou qu'ils sont incomplets. Il existe plusieurs méthodes permettant de prendre en compte ces facteurs de confusion :

- Choisir une population de référence la plus comparable possible à la cohorte sur le plan socioprofessionnel, où les habitudes de vie sont susceptibles d'être proches : on a vu plus haut qu'il

peut parfois s'agir de l'ensemble de la cohorte elle-même, si elle est d'effectif suffisamment important ou de la population générale d'une zone géographique proche de l'entreprise étudiée. Si ces méthodes ne permettent pas de contrôler parfaitement le rôle des facteurs de confusion, elles permettent souvent d'en réduire sensiblement l'effet sur les résultats de l'analyse épidémiologique.

- Etudier les autres maladies liées au tabagisme [8], afin de prendre en compte l'effet du tabagisme dans la cohorte (dans les situations où des données individuelles sur la consommation de tabac ne sont pas disponibles). Dans l'hypothèse où un excès de mortalité par cancers broncho-pulmonaires serait dû, dans cette population, à une surconsommation de tabac, les autres pathologies liées au tabagisme (cancers du larynx, de la vessie et du rein, maladies respiratoires non cancéreuses, et les maladies ischémiques) devraient également présenter un excès. S'il apparaît, dans une étude, un excès de mortalité par cancers broncho-pulmonaires et une absence d'excès pour les autres pathologies, il est peu probable que l'excès de cancers ainsi observé soit dû au tabagisme.
- Réaliser une "étude cas-témoins au sein de la cohorte" : sans détailler ce type de méthode, on peut simplement dire qu'il s'agit de procéder sur un nombre de sujets beaucoup plus limité que celui de la cohorte dans son ensemble, à une investigation beaucoup plus précise des conditions d'exposition aux divers facteurs à prendre en compte (facteurs professionnels et facteurs de confusion personnels). Lorsqu'elle est bien maîtrisée, cette méthode est excellente.

## 2.4 Données nécessaires : récapitulatif

En résumé, l'établissement et le suivi d'une cohorte de travailleurs en vue de la surveillance épidémiologique de base de la mortalité ne nécessite que quelques données très simples et aisément disponibles.

### DONNÉES INDISPENSABLES POUR LE CALCUL D'UN SMR

**Nom patronymique** (c'est-à-dire nom de **jeune fille** pour les femmes)  
**Prénom**  
**Genre**  
**Jour, mois et année de naissance**  
**Commune et département de naissance** (pays pour les personnes nées à l'étranger)  
**Date d'embauche**  
**Date de sortie de l'entreprise**

### DONNÉES INDISPENSABLES POUR REPÉRER UN FACTEUR PROFESSIONNEL

**Cursus professionnel de chaque salarié de la cohorte** : description des épisodes professionnels  
**Exemple :**  
**PERIODE 1**  
**Poste occupé** (ou, à défaut : métier ou secteur d'activité) : Fondeur  
**Date de début d'occupation de ce poste** : jj/mm/aaaa  
**Date de fin d'occupation du poste** : jj/mm/aaaa  
  
**PERIODE 2**  
**Poste occupé** (ou, à défaut : métier ou secteur d'activité) : Chef d'atelier, Fonderie  
**Date de début d'occupation du poste** : jj/mm/aaaa  
**Date de fin d'occupation du poste** : jj/mm/aaaa  
**Etc...**

### 3. Deuxième partie : les enquêtes de "clusters" (ou agrégats spatio-temporels de cas)

Il s'agit là de décrire les principes de la démarche et non pas d'un guide opérationnel permettant de mener à bien une investigation comme il en existe par ailleurs [19].

De plus en plus de médecins du travail, de CHSCT, d'acteurs de prévention ou de représentants du personnel, font état d'observation de cas de maladies qui leur semblent excessifs dans la population de leur entreprise. La cause présumée de cet excès est parfois connue ou soupçonnée par les salariés ou les médecins (utilisation actuelle ou passée de produits ou procédés à risque). Le plus souvent, il n'y a pas d'hypothèse précise. Ces observations concernent fréquemment des pathologies cancéreuses ou d'autres pathologies graves. Les médecins du travail sont le plus souvent démunis pour explorer et vérifier ces observations, en l'absence de système de surveillance épidémiologique dans les entreprises, les branches ou les secteurs d'activité concernés.

Pour les études d'incidence, le principe de l'analyse et les méthodes de calcul sont les mêmes que pour les études de mortalité. Ce qui diffère est la nécessité de disposer de données concernant la survenue des maladies incriminées. En effet, en complément de données sur la population de l'entreprise (la cohorte), des informations validées sur les cas observés sont indispensables.

#### 3.1 Evaluation de l'importance du problème

L'observation d'un agrégat<sup>6</sup> de cas doit être interprétée avec précaution. Pris isolément, le nombre de cas peut sembler en excès alors qu'il est peut-être dû au hasard.

La vérification de la réalité d'un excès soupçonné fait appel à des méthodes épidémiologiques et statistiques qui peuvent être plus ou moins complexes selon la situation. Quelle que soit la méthode utilisée, un certain nombre d'étapes sont indispensables à l'étude statistique des données qui permettront de vérifier que l'excès de cas supposé est bien réel.

Dans un premier temps, une évaluation de l'importance du problème sera effectuée à partir de l'ensemble des éléments suivants :

- le nombre de cas signalés,
- le genre et l'âge des cas (au moment du diagnostic),
- la pathologie d'intérêt :
  - une seule pathologie est-elle signalée ?
  - s'il s'agit de cancers, est-ce d'une localisation unique ?
  - s'il y a plusieurs localisations de cancers observées, quelles sont-elles ?
- l'intervalle de temps concerné : pendant quelle période (quelles années) les cas sont-ils apparus ?
- l'existence d'un (de plusieurs) facteur(s) de risque suspecté(s) : à quelles nuisances ont été exposés les salariés ? S'il s'agit de cas de cancers, l'exposition des salariés à des substances (ou des procédés) cancérogènes connus pendant la période sera recherchée. Rappelons que le Centre international de recherche contre le cancer publie régulièrement des monographies évaluant le risque cancérogène chez l'homme de produits, mélanges et opérations industrielles. Elles sont classées en quatre groupes, et disponibles dans une liste régulièrement actualisée accessible sur [www.iarc.fr](http://www.iarc.fr).

<sup>6</sup> Rappel : un "agrégat" est un regroupement inhabituel d'un problème de santé dans un espace – temps défini. On parle aussi "d'agrégat spatio-temporel" ou de "cluster" (mot anglais). Un agrégat est observé si le nombre de cas observés est plus élevé que le nombre attendu.

Le médecin devra faire une revue de la littérature scientifique nationale et internationale portant sur le problème suspecté (pathologie observée et facteurs de risque connus ou soupçonnés).

Disposer de l'ensemble de ces informations permettra au médecin du travail d'évaluer la plausibilité de l'excès de risque suspecté et de décider s'il doit mettre en place une investigation. Si c'est le cas, il devra alors construire son étude, en définissant : la cohorte, les cas, etc...

## 3.2 Constitution de la cohorte (population dont sont issus les cas observés)

L'investigation d'un cluster nécessite, comme la surveillance épidémiologique de la mortalité, la mise en place d'une cohorte pour laquelle les mêmes données sont nécessaires.

### 3.2.1 Définition de la cohorte

Il est indispensable de disposer de l'ensemble des informations listées ci-dessus, pour pouvoir formuler une définition de la cohorte. La cohorte, qui sera définie selon des critères spatiaux et temporels, est la population à laquelle appartiennent les cas. Sa taille est très variable selon les situations. Elle peut être composée des employés d'un service ou d'un atelier, d'un bâtiment ou de l'ensemble d'une entreprise.

### 3.2.2 Données indispensables sur les cas observés

Le nombre de cas observés est le nombre de **cas incidents** (c'est-à-dire les nouveaux cas), observés année par année de la période de suivi dans la population de l'entreprise. Contrairement à la mortalité qui fait l'objet d'un enregistrement national systématique, il n'existe pas de base de données exhaustive de l'incidence des maladies pour l'ensemble de la population française. **Le médecin du travail devra donc lui-même identifier tous les cas incidents qui se seraient produits au sein de la cohorte.** Il doit de plus investiguer chaque cas observé, afin de recueillir les informations suivantes, validées de façon stricte :

- **genre ;**
- **age au moment du diagnostic ;**
- **diagnostic précis de la pathologie** (s'il s'agit d'un cancer, la localisation de la tumeur primitive est indispensable, en raison de la fréquence des métastases) ;
- **date exacte du diagnostic** (au minimum, année du diagnostic) ;
- **délai entre le début de l'exposition suspectée et le diagnostic de la maladie ;**
- **existence d'une antériorité de l'exposition par rapport au diagnostic de maladie.**

Un recueil de données (envoi de questionnaires aux salariés, les questionnant sur leurs pathologies passées) peut être effectué. L'exhaustivité des données médicales doit être atteinte.

Remarque : pour les cas de pathologies multiples chez le même sujet, il faut disposer de ces informations pour les diagnostics successifs, même si le premier diagnostic est antérieur à la période étudiée. Par exemple, si un cas de cancer a été identifié par le médecin du travail, il devra enquêter sur les antécédents personnels de cancer de cette personne.

### 3.2.3 Analyse épidémiologique de l'incidence : le ratio standardisé d'incidence (SIR)

L'analyse épidémiologique de l'incidence de cas parmi les salariés d'une entreprise consiste à comparer le nombre de cas survenus au cours d'une période de temps à l'incidence d'une population de référence. La méthode de calcul du SIR est identique à celle du SMR et est décrite en détail dans l'annexe 2. Les formules du SIR et de son intervalle de confiance sont présentées dans l'annexe 3.

### 3.2.4 Données de référence

Comme on l'a déjà souligné, il n'existe habituellement pas de données de référence concernant l'incidence des maladies pour l'ensemble de la population française. Pour les cancers cependant, on dispose d'estimations nationales des taux d'incidence des principaux cancers, effectuées à partir des données du réseau Francim des registres du cancer qui existent dans un certain nombre de

départements [9]. En annexe 5, sont présentés les taux d'incidence de référence, par genre et classe d'âge (en 1995 et 2000).

Cependant, ces estimations nationales reposent actuellement sur les données d'un petit nombre de registres départementaux : on ne dispose donc pas de taux d'incidence de référence au niveau local, départemental ou régional, et, sauf exception (estimations locales fournies par le réseau Francim), les estimations de taux nationaux sont les plus couramment utilisées dans ce type d'analyse.

### 3.3 Recherche de la cause d'un excès de cas

L'organigramme suivant (figure 2) illustre la démarche d'une investigation de clusters, de l'évaluation initiale du problème jusqu'au rapport d'étude final. L'investigation d'un cluster peut aboutir à la mise en place d'une étude d'incidence, si les données médicales exhaustives sur les cas sont disponibles, et à une étude de mortalité, si on ne dispose que des données administratives indispensables pour ce type d'étude.

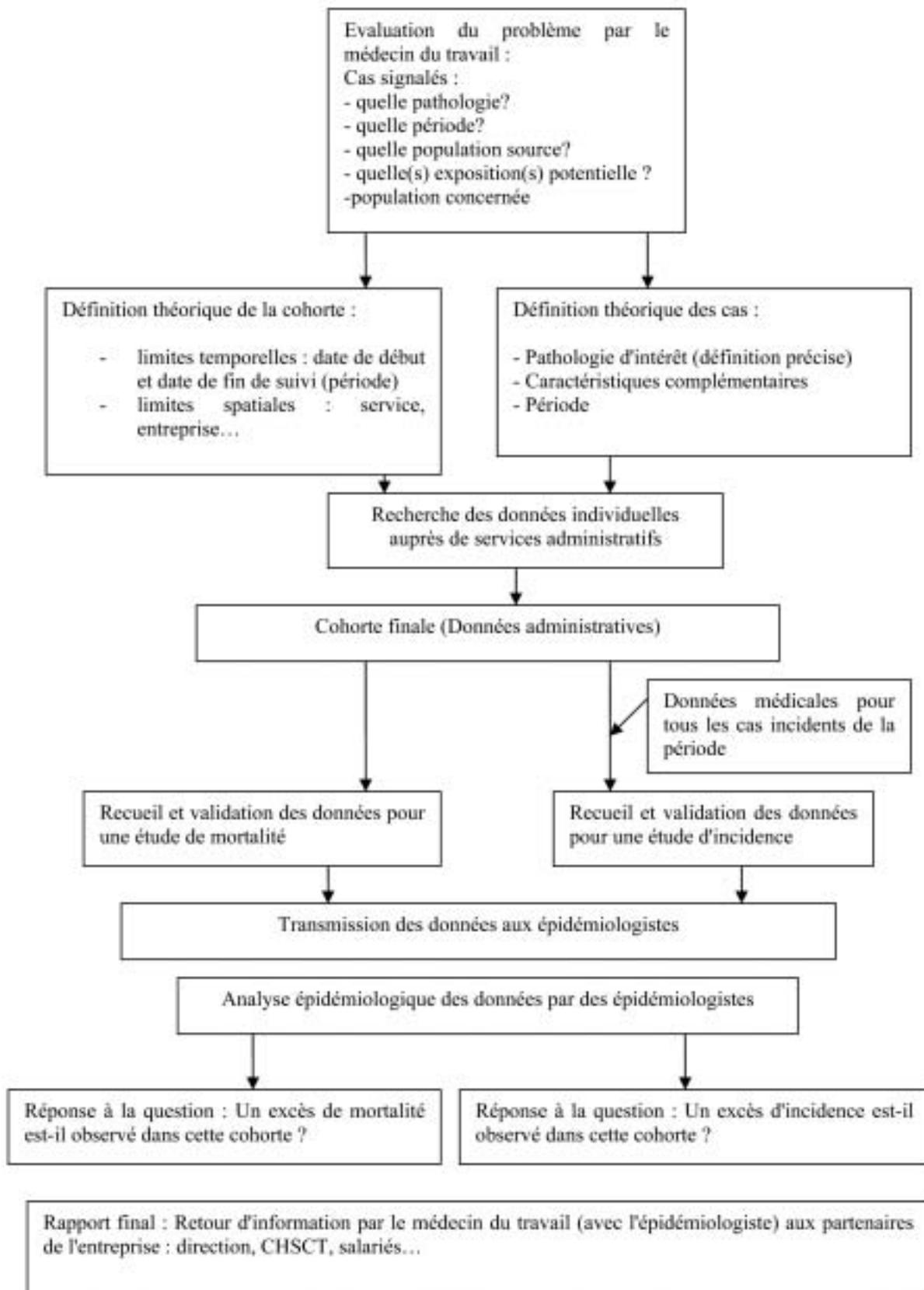
Lorsqu'un excès est mis en évidence au cours d'une étude d'incidence, la recherche de la cause de cet excès ne pourra pas être systématiquement mise en œuvre, car elle n'est pas possible dans tous les cas. En effet, cela fait appel à des méthodes épidémiologiques complexes, qui :

- nécessitent des informations complémentaires précises sur les risques potentiels existant, et ayant existé dans le passé, pour les salariés : cursus professionnels complets de tous les salariés "à risque", nature des expositions à des produits chimiques ou à des procédés "à risque", nature des expositions environnementales... ;
- dépendent d'un effectif suffisant de la cohorte, pour disposer d'une puissance statistique acceptable : mettre en place une étude complexe sur un effectif réduit n'a aucun sens, car les résultats finaux ne permettraient pas de conclure sur l'effet d'un facteur de risque ;
- appartiennent au domaine de la recherche "fondamentale" et sont en cours d'investigation par des équipes de recherche, disposant des moyens adaptés (exemples : études internationales multicentriques).

Enfin, il faut rappeler que les réponses aux questions soulevées ne seront donc jamais immédiates : une investigation d'agrégat demande un certain temps, afin d'élaborer un protocole, même simple (définir l'objectif de l'étude, préciser les moyens nécessaires, définir les rôles des partenaires éventuels, etc...), de mener l'étude, c'est-à-dire rassembler des données déjà existantes et/ou recueillir de nouvelles données, de les valider, puis de les analyser.

Au minimum plusieurs mois, et parfois plus encore, sont nécessaires pour l'analyse d'un cluster par des équipes compétentes spécialisées en épidémiologie. Bien entendu, si un système de surveillance épidémiologique est déjà en place dans l'entreprise concernée, surtout si l'on dispose d'un recul suffisant, l'investigation sera beaucoup plus rapide : c'est pourquoi, comme on l'a souligné dans le premier chapitre du guide, le développement systématique de tels systèmes est indispensable, car il permet d'apporter rapidement des réponses à des questions diversifiées.

Figure 2. Déroulement d'une étude de cluster



## 4. Exemples

### 4.1 Surveillance de la mortalité chez des hommes ayant travaillé à EDF-GDF

La division épidémiologie du Service général de médecine du travail de l'entreprise EDF-GDF a mis en place durant les années 1980 un système d'information épidémiologique concernant les agents de cette entreprise. Ce système, décrit dans diverses publications [10,11] comprenait à la fois des informations démographiques, professionnelles, décrivant les problèmes de santé et les expositions professionnelles. Ce système a permis la réalisation de nombreuses études épidémiologiques [12-14].

Une étude de la mortalité observée a été menée chez les personnes ayant travaillé à EDF-GDF depuis 1978 [15]. Ses objectifs étaient de caractériser la mortalité de cette population et en particulier d'étudier les relations entre la mortalité par cancer et les expositions professionnelles à certaines substances, puis d'évaluer la part des décès attribuables à ces expositions professionnelles.

L'étude a concerné la cohorte des hommes ayant travaillé à EDF-GDF entre janvier 1978 et mars 1995 (dite "Cohorte-78") au moins un an dans l'entreprise, c'est-à-dire 183 490 personnes. Les expositions professionnelles de chaque sujet ont été reconstituées en croisant leur histoire professionnelle détaillée et une base de données permettant d'attribuer les expositions à partir de l'histoire professionnelle du salarié (matrice emplois-expositions MATEX).

Une recherche de l'ensemble des décès survenus dans la Cohorte-78, ainsi que leurs causes médicales, a été conduite en 2000 et 2001. Plusieurs sources d'informations ont été exploitées : Service des pensions de l'industrie électrique et gazière, RNIPP et fichier 7bis de l'Insee, et service CépiDc de l'Inserm, selon la procédure décrite dans le décret 98-37 (voir annexe 1). Le statut vital de plus de 99 % des personnes de la cohorte a été retrouvé. L'analyse épidémiologique de la mortalité a porté sur les 182 053 hommes, dont le statut vital a été retrouvé.

Sur l'ensemble des décès observés (15 934 décès), la cause initiale de décès la plus fréquente était le cancer (42 % des décès), devant les maladies de l'appareil circulatoire (22 % des décès) et les morts violentes (13 % des décès). Sur les 6 629 décès par cancer, les cancers du poumon représentent 25 % de l'ensemble soit 1 642 décès.

La mortalité de cette population a été comparée à celle de la population générale en calculant des SMR.

Le SMR toutes causes était égal à 0,78 avec un intervalle de confiance à 95 % égal à [0,76 ; 0,79] et stable au cours de la période, illustrant l'existence d'un important "effet travailleur en bonne santé" dans cette entreprise. Le SMR par cancer était égal à 0,88 (IC95 %=[0,86 ; 0,90] et diminuait au cours du temps, passant de 0,91 pour les cinq premières années à 0,83 pour les cinq dernières..

Par contre, un gradient de mortalité toutes causes était observé selon la catégorie socioprofessionnelle. Les taux de décès étaient d'autant plus faibles que la catégorie socioprofessionnelle était élevée, et il en était de même pour les SMR. Quand on s'intéresse aux causes de décès, ce gradient n'était pas observé pour toutes les causes : il était marqué pour les cancers, les maladies de l'appareil circulatoire, les morts violentes et les maladies de l'appareil digestif, mais n'existait pas pour les maladies infectieuses et les maladies rénales.

Les résultats sont décrits et discutés en détail dans l'article du BEH cité plus haut [15]. Ces résultats ont mis en évidence de larges inégalités de mortalité selon les catégories professionnelles, avec la situation la plus défavorable chez les ouvriers non qualifiés. L'étude des relations entre certaines expositions et la mortalité a montré que certaines expositions (ex : chrome, amiante, benzène) étaient associées à une mortalité plus élevée (cancers des fosses nasales, cancers de la plèvre, leucémies). Cette étude a fait l'objet d'une thèse de doctorat [16].

## 4.2 Etude de mortalité dans l'industrie productrice de métaux durs en France

Cette étude épidémiologique a été réalisée dans 10 usines françaises impliquées dans la fabrication de métaux durs [17]. Ces usines étaient répertoriées en 1991 dans la catégorie des PME/PMI. Quatre usines employaient au 31/12/1991 moins de 100 salariés, 4 entre 100 et 200 salariés et 2 usines employaient plus de 200 personnes.

Leur activité principale était la fabrication de métaux durs à base de carbure de tungstène et de cobalt selon les procédés de métallurgie des poudres. Les métaux durs sont des pseudo-alliages utilisés essentiellement pour les activités de perçage, forage et découpage. Les qualités de dureté de ces métaux sont dues en grande partie au carbure de tungstène et à d'autres carbures métalliques qui, avec le cobalt, constituent l'essentiel de ces alliages.

### 4.2.1 Objectif de l'étude

Cette étude épidémiologique avait pour but de rechercher un éventuel excès de cancers broncho-pulmonaires (CBP) lié à des expositions professionnelles simultanées au carbure de tungstène et au cobalt.

### 4.2.2 Etablissement de la cohorte

La cohorte comprend l'ensemble des sujets, hommes et femmes, ayant été employés dans chaque usine, pendant au moins 3 mois, depuis la date de création de l'usine, ou une date ultérieure définie en fonction de la disponibilité des archives administratives, jusqu'au 31 décembre 1991.

Le recrutement des usines s'est effectué de proche en proche, par l'intermédiaire des médecins du travail et des deux chambres syndicales concernées.

Les données administratives ont été recueillies à partir des registres d'entrée et de sortie du service du personnel mis à disposition par les directions des entreprises. Pour chaque sujet, les données suivantes ont été recueillies :

- nom de naissance, prénom,
- nom marital,
- date et lieu de naissance,
- catégorie socioprofessionnelle à l'embauche,
- date d'embauche,
- date et motif de départ.

La période de suivi pour le recueil des décès et des causes de décès s'étend du 1<sup>er</sup> janvier 1968 (date de création du fichier informatisé des causes des décès de l'Inserm) au 31 décembre 1991.

Le décret 98-37, qui précise les conditions d'accès aux données de mortalité (voir annexe 1), n'était pas encore paru au moment où l'étude a été réalisée, ce qui explique que la recherche des données de mortalité a été légèrement différente de ce qui est décrit dans ce guide. L'identification des sujets décédés pendant la période de suivi a été effectuée par l'envoi d'un courrier aux services d'état civil des mairies de naissance pour les sujets nés en France et par une recherche dans le fichier 7bis de l'Insee des sujets décédés. Ce fichier recense l'ensemble des sujets décédés en France depuis 1978.

Les causes de décès ont été recherchées dans le fichier national des causes de décès géré par le CépiDc de l'Inserm. Les médecins du travail des usines ont recherché les causes non retrouvées dans ce fichier.

### 4.2.3 Population de référence externe

Les taux de mortalité utilisés pour référence ont été dans un premier temps, ceux de la population générale de la France et dans un deuxième temps, des départements dans lesquels les usines sont situées. Ces taux sont obtenus en calculant le rapport par année, classe d'âge, genre, département et cause de décès entre les nombres absolus de décès (disponibles auprès du CépiDc de l'Inserm) et les nombres absolus de résidents (disponibles auprès de l'Insee).

#### 4.2.4 Résultats

La cohorte composée des effectifs de chaque usine compte 7 459 sujets (soit 114 834 personnes-années). Le nombre total de décès survenus au cours de la période 1968-1991 est de 684. Les causes de décès ont été retrouvées pour 662 d'entre eux (96,8 %).

Pour le genre masculin, la mortalité globale observée est inférieure à celle de la population française (SMR=0,92 ; intervalle de confiance IC95 % 0,84-0,99) mais ne diffère pas de celle de la population des départements où sont situées les usines (SMR=0,99 ; IC95 % 0,91-1,07). En revanche, un excès de mortalité par cancer broncho-pulmonaire (CBP) est observé dans la cohorte étudiée comparativement aux deux références externes (France : SMR=1,29 ; IC95 % 0,99-1,66 ; Départements : SMR=1,34 ; IC95 % 1,03-1,72) (tableau 3).

**Tableau 3.** SMR en fonction des références externes - Hommes

	Références externes	Décès observés	Décès attendus	SMR	IC95 %
Toutes causes de décès	France	591	644,59	0,92	0,84-0,99
	Départements	593	600,68	0,99	0,91-1,07
Cancers broncho-pulmonaires	France	61	47,22	1,29	0,99-1,66
	Départements	61	45,47	1,34	1,03-1,72

Une sous-mortalité est observée pour un temps de latence (temps écoulé entre la date d'embauche et la date de décès ou date de fin de période de surveillance) inférieur à 9 ans traduisant un "effet travailleur en bonne santé". Cet effet diminue avec le temps.

En effet, des excès de décès non significatifs sont observés après un temps de latence compris entre 10 et 19 ans et un temps de latence de plus de 20 ans (tableau 4).

**Tableau 4.** Mortalité par CBP en fonction du temps de latence (référence externe : France)

	Décès observés	Décès attendus	SMR	IC95 %
Temps de latence (années)				
0-9	3	4,04	0,74	0,15-2,17
10-19	17	12,77	1,33	0,78-2,13
>=20	41	30,40	1,35	0,97-1,83

#### 4.2.5 Etude cas – témoins complémentaire

Les informations détaillées sur les expositions professionnelles et les habitudes tabagiques n'ont pas pu être recueillies pour tous les sujets de la cohorte. Une étude cas-témoins complémentaire incluse dans la cohorte a donc été réalisée pour rechercher ces informations sur un nombre beaucoup plus restreint de sujets. Le principe est le même que celui des études cas-témoins de population générale. Il consiste à comparer la fréquence d'exposition entre un groupe de sujets malades ou décédés (cas) et un groupe de sujets indemnes de cette maladie ou de ce décès (témoins). Les cas sont les sujets de la cohorte décédés d'un CBP. Les témoins sont un échantillon aléatoire des sujets de la cohorte.

Pour chaque sujet cas et témoins, un nouveau recueil de données sur les carrières professionnelles et les habitudes tabagiques a été réalisé dans toutes les usines. Ce recueil a été réalisé par les services du personnel (carrières professionnelles) et les médecins du travail (habitudes tabagiques) des usines à partir de plusieurs sources d'information :

- le sujet lui-même,
- la consultation du dossier du service du personnel ou du dossier de médecine du travail,
- l'interview d'un collègue de travail, d'un membre de la famille.

Les expositions professionnelles ont été évaluées à l'aide d'une matrice emplois-expositions élaborée par un groupe d'experts.

#### 4.2.6 Conclusion de l'étude

L'étude de cohorte met en évidence un excès de mortalité par CBP. L'étude cas-témoins confirme ce résultat : un risque de CBP est observé parmi les sujets exposés simultanément au cobalt et au carbure de tungstène par rapport aux non-exposés, ce risque augmentant avec la durée et la dose cumulée (niveau X durée).

### 4.3 Etude d'incidence : investigation d'une suspicion d'agrégat de cancers parmi le personnel d'un établissement de recherche

Un agrégat de cancers a été suspecté parmi le personnel d'un des bâtiments d'un établissement de recherche. Une vingtaine de cas de cancers (de diverses localisations) avait été notifiée par les personnels du bâtiment suspect, dont l'effectif était d'environ 120 personnes pendant une période de 6 ans. L'Institut de veille sanitaire a été chargé de mener cette investigation [18].

#### 4.3.1 Données sur la population de l'entreprise

Les travaux ont consisté à :

- définir la période étudiée : la période de l'étude a été définie comme la période de signalement des cas, entre janvier 1995 et décembre 2001 ;
- connaître l'exhaustivité des effectifs de la population du bâtiment suspecté de l'entreprise de 1995 à 2001, année par année, avec la structure d'âge par genre de cette population. Il a été décidé de n'inclure que les personnes ayant travaillé au moins 6 mois dans ces services.

Ces informations indispensables, qui devraient être fournies par les services du personnel des entreprises (ou établissements), ont été très difficiles à obtenir quand il s'agissait d'une cohorte rétrospective composée de personnels ayant des statuts divers : salariés gérés par plusieurs employeurs, travailleurs non salariés (boursiers, retraités), stagiaires étrangers... Le service du personnel de l'établissement n'ayant pas réussi à établir rétrospectivement une liste exhaustive des personnes ayant travaillé dans ce bâtiment entre 1995 et 2001 avec toutes les données indispensables, la cohorte n'a été reconstituée que partiellement.

#### 4.3.2 Données sur les cas

Concernant les cas, il a fallu disposer d'informations valides :

- pour chaque diagnostic, et chacun des diagnostics successifs en cas de cancers multiples :
  - le diagnostic précis : localisation du cancer primitif
  - la date précise du diagnostic
- la date de naissance (ou l'année de naissance) de chaque cas,
- le genre de chaque cas.

L'enquête a consisté à rechercher l'exhaustivité des cas de cancers (déjà signalés ou "nouveaux"), parmi la population de travailleurs du bâtiment en question, pendant la période 1995-2001.

En pratique, il n'existait pas de système d'enregistrement systématique des pathologies des salariés de l'établissement. Un **questionnaire médical**, élaboré avec l'aide de l'InVS, a été proposé par le médecin du travail aux personnels de l'établissement travaillant encore dans l'établissement, qui avaient déclaré avoir subi un cancer dans cette période. Les coordonnées du médecin traitant ont été recueillies, afin de pouvoir accéder à des informations complémentaires, le cas échéant.

Pour le personnel ayant quitté l'établissement, ce questionnaire médical leur a été envoyé (quand leurs coordonnées complètes avaient été retrouvées).

Comme plusieurs signalements de cas concernaient des personnes décédées, le médecin du travail a adressé un questionnaire médical aux proches, afin qu'ils valident les informations médicales déjà recueillies.

### 4.3.3 Données de référence

Dans cette étude, les données de référence qui ont été utilisées sont les données d'incidence de cancers toutes localisations confondues France entière, par classes d'âge, chez les hommes et les femmes, extraites des publications disponibles :

- Pour l'analyse des années 1995 à 1999 - Le cancer en France : incidence et mortalité. Situation en 1995. Evolution entre 1975 et 1995. Réseau Francim.
- Pour l'analyse des années 2000 à 2001 - Taux d'incidence estimés par genre et tranche d'âge, en France, en 2000, réseau Francim [9].

### 4.3.4 Résultats

Dans cette étude, 8 cas avaient été signalés initialement chez les hommes, et 14 cas chez les femmes. Après enquête, seulement 3 cas ont été validés chez les hommes, et 5 cas chez les femmes.

Chez les hommes,  $SIR = 3 / 3,046 = 0,98$ .

Chez les femmes,  $SIR = 5 / 1,460 = 3,42$ .

Ces résultats, basés sur des données non exhaustives, qui ne peuvent être considérés que comme indicatifs, semblaient indiquer un excès de cancers du sein chez les femmes, résultats en accord avec les données récentes de la littérature scientifique, montrant un excès de cancers, en particulier de cancer du sein, parmi le personnel de laboratoires de recherche.

### 4.3.5 Difficultés

Il a été impossible aux services administratifs de trouver l'ensemble des adresses des personnels ayant travaillé avant 2002, dans le bâtiment suspecté. Ceci n'a pas permis de retrouver l'exhaustivité des cas incidents pendant la période, donc une étude de l'incidence n'a pas été possible. D'autre part, le genre, l'année de naissance, ou le nom patronymique de certaines de ces personnes n'a pas été retrouvé, empêchant la réalisation d'une étude de mortalité.

Il a donc été recommandé au service de médecine du travail de mettre en place, en relation étroite avec les services du personnel de l'établissement, un recueil systématique des données minimales indispensables afin de permettre une surveillance épidémiologique de la mortalité de ces travailleurs, dans la perspective d'effectuer ultérieurement une étude de mortalité.

En conclusion, la mise en place d'un système de surveillance épidémiologique dans une entreprise (ou un établissement) nécessite de disposer de la liste nominative des personnels, **année par année** (en incluant les années sans cas incident).

Il faut donc tenir à jour et conserver "l'historique" des personnels qui ont travaillé dans l'entreprise en incluant les éléments suivants :

- la liste complète des personnels de l'entreprise (nom patronymique et prénoms),
- la période d'embauche de chaque personne (pour pouvoir tenir compte de leur période de travail),
- la date de naissance complète de chaque personne (jour/mois/année),
- la commune et le département de naissance (pays pour les personnes nées à l'étranger) de chaque personne (pour permettre d'effectuer une étude de mortalité),
- le genre de chaque personne.

Pouvoir vérifier un excès d'incidence nécessite **d'être exhaustif**, dans la documentation des cas incidents. A défaut de données exhaustives et valides concernant les cas incidents, et si la pathologie est susceptible d'entraîner un décès, **les études de mortalité sont possibles**, si l'on dispose de données complètes (et valides) sur la population des personnels.

## 5. Conclusion

Il est devenu aujourd'hui nécessaire de connaître en permanence l'état de santé de la population des travailleurs. Les besoins de connaissance concernent notamment la description de l'état de santé en fonction des caractéristiques de l'emploi, des conditions de travail et des expositions professionnelles, l'identification d'éventuels excès de risque dans des populations spécifiques, la vérification d'un excès de risque présumé, et la recherche de ses causes le cas échéant.

De nombreuses méthodes, appliquées à l'ensemble de la population française ou à celle de travailleurs d'entreprises ou de secteurs spécifiques existent pour répondre à ces besoins de connaissance. Elles nécessitent, cependant, souvent d'importants moyens spécialisés. C'est pourquoi, la mise en place systématique d'une surveillance de la mortalité dans les entreprises et les secteurs économiques, qui on l'a vu, ne nécessite que des données simples, aisément accessibles à tout service de santé au travail, est une première étape indispensable, qui devrait rapidement être mise en œuvre de façon généralisée, en coopération le cas échéant avec les équipes d'épidémiologistes spécialisés de l'InVS, de l'INRS, des instituts universitaires de médecine du travail et de l'Inserm.

## 6. Pour en savoir plus

Comme on l'a dit en introduction de ce guide, l'analyse des données recueillies en routine permettant des études régulières de la mortalité par cause demande des compétences spécialisées en épidémiologie. Un soutien méthodologique peut être apporté par le Département santé travail de l'InVS, le Département épidémiologie en entreprises de l'INRS ou par d'autres équipes spécialisées.

On pourra consulter un ouvrage particulièrement utile, qui s'appuie sur de nombreux exemples concernant les risques professionnels :

Épidémiologie – Principes et méthodes quantitatives.

Bouyer J, Hémon D, Cordier S, Derriennic F, Stücker I, Stengel B, Clavel J.

Paris : Inserm, 1993.

## 7. Références

1. Moulin JJ, lafontaine M, Mantout B, Belanger A, Michel M, Wild P, Clavel T, Fournier M, Fontana JM. La mortalité par cancers broncho-pulmonaires parmi les salariés de deux usines sidérurgiques. *Rev Epidemiol et Sante Publ* 1995; 43:107-21.
2. Bouyer J, Hemon D, Cordier S, derriennic F, Stücker I, Stengel B, Clavel J. *Epidémiologie. Principes et méthodes quantitatives*. Les éditions Inserm, Paris, 1993.
3. Fox AJ, Collier FF. Low mortality rates in industrial cohort studies due to selection for work and survival in the industry. *Br J Prev Soc Med* 1976; 30:225-30.
4. Goldberg M, Melchior M, Leclerc A, Lert F. Epidémiologie et déterminants des inégalités sociales de santé. Apports récents et problèmes actuels. *Rev Epidemiol Santé Publ* 2003; 51:381-401.
5. Enterline P.E, Marsh G.M. Mortality among workers in a nickel refinery and alloy manufacturers plant in West Virginia. *J Nat Cancer Inst* 1982; 68:925-33.
6. Valenty M, Tirmarche M, Mitton N, *et al.* Causes of death of COGEMA active workers, 1980-1995. *Rev Epidemiol Sante Publ* 2003; 51(5):461-8.
7. Pavillon G, Laurent F. Certification et codification des causes médicales de décès. *Bull Epidemiol Heb* 2003; n°30-31:134-8.
8. Moulin JJ. Les études épidémiologiques de cohortes industrielles : comment prendre en compte le tabagisme ? *Arch Mal Prof* 1991; 52:319-26.
9. Remontet L, Buemi A, Velten M, Jouglu E, Estève J. Evolution de l'incidence et de la mortalité par cancer en France de 1978 à 2000. Rapport édité par l'InVS, août 2003.
10. Imbernon E, Goldberg M, Guenel P, *et al.* Matex : une matrice emplois-expositions destinée à la surveillance épidémiologique des travailleurs d'une grande entreprise (EDF-GDF). *Arch MalProf* 1991; 52:559-66.
11. Chevalier A, Goldberg M, Godard C *et al.* Incidence des cancers dans la population masculine des salariés en activité à Electricité Gaz de France. *Rev Epidemiol Sante Publ* 1996; 44 (1):25-36.
12. Luce D, Leclerc A, Goldberg M, *et al.* Facteurs de risque individuels des accidents du travail à Electricité Gaz de France. *Arch Mal Prof* 1987; 48:461-5.
13. Chevalier A, Blanc C, Goldberg M. Réorganisation du travail et absentéisme pour raison médicale dans un centre de distribution d'Electricité de France - Gaz de France. *Arch Mal Prof* 1990; 51:235-43.
14. Guenel P, Imbernon E, Chevalier A, *et al.* Leukaemia in relation to occupational exposures to benzene and other agents: a case-control study nested in a cohort of gas and electric utility workers. *Am J Ind Med* 2002; 42:87-97.
15. Marchand J-L, Imbernon E, Goldberg M. Mortalité chez les hommes ayant travaillé à EDF-GDF. *Bull Epidemiol Heb* 2003; n°30-31:150-2.
16. Marchand J-L. Analyse de la mortalité dans la population des salariés d'EDF-GDF, et des relations entre mortalité par cancer et expositions professionnelles. Thèse de doctorat de l'université René Descartes-Paris V. Discipline : sciences de la vie et de la matière. Spécialité : épidémiologie. 2003. 153 p.
17. Moulin JJ, Perdrix A, Lasfargues G, Romazini S, Bozec C, Deguerry P, Pellet F, Peltier A, Wild P. Etude épidémiologique de mortalité dans l'industrie productrice de métaux durs en France. Cahier de notes documentaires INRS 1997; 168:411-28.
18. Buisson C, Drouard S, Imbernon E. Suspicion d'un agrégat de cancers à l'observatoire de Meudon – Rapport d'étude. Département santé travail, Institut de veille sanitaire, 2003.
19. Centers for Disease Control. Department of Health and Human Services. Guidelines for investigating clusters of health events. MMWR, Morbidity Mortality Weekly Report. Recommendations and reports 39[RR-II], 1-23, 27-7-1990.

## 8. Annexes

### Annexe 1. Procédure d'accès aux causes de mortalité

Les causes médicales des décès survenus dans la cohorte seront recueillies selon les modalités fixées par le décret n° 98-37 du 16 janvier 1998. Ce décret autorise l'accès aux données relatives au décès des personnes inscrites au répertoire national d'identification des personnes physiques (RNIPP) dans le cadre des recherches dans le domaine de la santé.

Deux étapes successives doivent être réalisées :

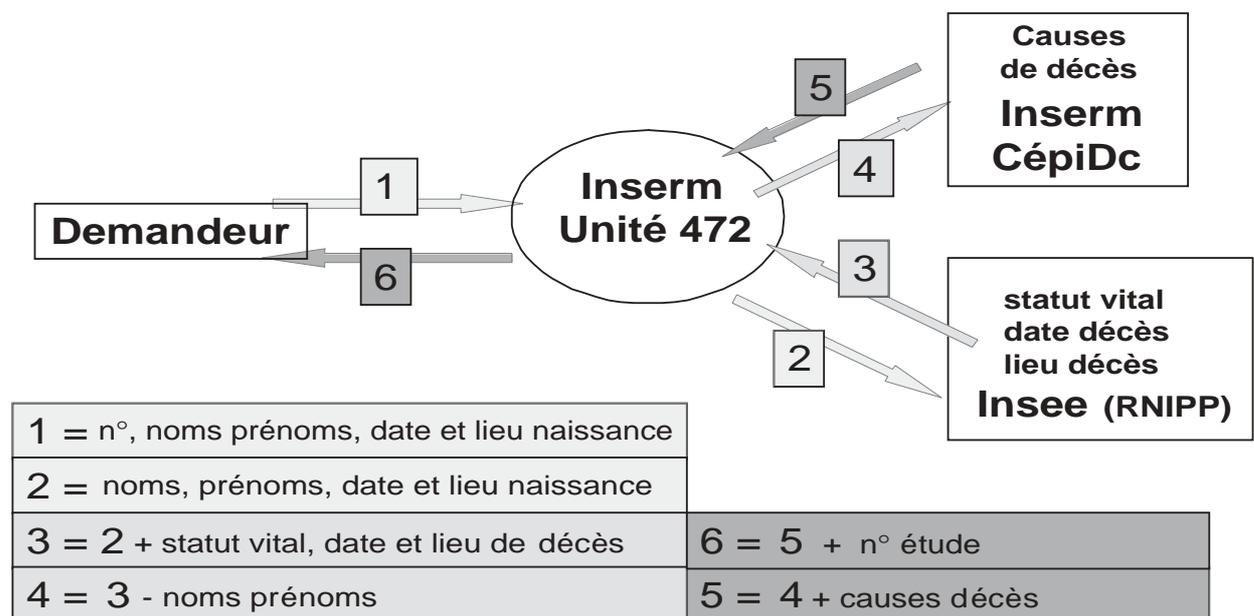
- 1 - La recherche du statut vital par interrogation du répertoire national d'identification des personnes physiques, géré par l'Insee, pour savoir si une personne est décédée.
- 2 - La recherche des causes de décès dans la base de données du service d'information sur les causes médicales de décès (CépiDc) de l'Inserm.

Le décret cité fixe les modalités d'accès et de transfert des données entre l'Insee et l'Inserm. L'Unité 472 de l'Inserm est responsable de l'application du décret et le centre de ressources informatiques (CRI) de l'IFR 69 assure la réalisation technique.

Pour déposer une demande de recherche de statut vital, le demandeur devra avoir obtenu **un avis favorable du comité consultatif** sur le traitement de l'information en matière de recherche dans le domaine de la santé, et l'**autorisation de la Commission nationale de l'informatique et des libertés** (Cnil).

La figure ci-dessous présente les étapes successives de la procédure.

#### Procédure pour l'accès aux causes de décès - Décret 98-37 -



Principales étapes de la procédure :

**1) Le demandeur envoie au CRI un fichier comprenant, pour chaque personne à étudier, les items suivants :**

- le numéro d'ordre du sujet dans ce fichier,
- le numéro identifiant le sujet dans l'étude du demandeur,
- le nom (et nom de jeune fille) et les prénoms,
- la date de naissance,
- le code du lieu de naissance (suivant le code officiel géographique),
- la commune de naissance,
- le pays de naissance,
- le genre.

**2) L'Insee, après interrogation du RNIPP, renvoie, pour chaque sujet, en plus du fichier le statut vital, et s'il y a lieu :**

- la date de décès,
- le code du lieu de décès,
- la commune de décès,
- le pays de décès,
- le numéro d'acte de décès.

**3) Le CépiDc, interrogé sur les sujets décédés du fichier, fournit leurs causes médicales de décès.**

Remarque : si le demandeur a connaissance du statut vital d'une partie des sujets du fichier, il peut interroger directement le CépiDc pour ceux-ci.

## Annexe 2. Calcul d'un ratio standardisé de mortalité ou d'incidence

Tout ce qui suit s'applique indifféremment à la mortalité (survenue de décès) ou à l'incidence (survenue de cas d'une maladie) : en effet, les méthodes utilisées pour mettre en évidence et quantifier un excès éventuel de cas d'une pathologie dans une entreprise sont identiques, qu'il s'agisse de mortalité ou de morbidité.

Lorsqu'on veut vérifier un excès de mortalité (ou d'incidence) dans la population des travailleurs d'une entreprise, on utilise une méthode qui consiste à calculer le nombre de décès (ou cas) attendus dans l'entreprise, si elle était identique à la population de référence, et à comparer ce nombre attendu au nombre de décès réellement observé.

### 8.2.1 Nombre de décès ou cas observés

Les décès (ou cas incidents) observés au cours de la période de suivi dans la population de l'entreprise sont comptabilisés.

#### ◆ Calcul des personnes-années

Chaque sujet contribue à un nombre de personnes-années défini comme le nombre d'années pendant lesquelles il est susceptible d'être enregistré comme décès (ou nouveau cas). Ainsi, chaque sujet contribue pour une unité par année à partir de la date de début du suivi jusqu'à la première des dates suivantes : la date de décès (ou diagnostic), la date où il est perdu de vue, la date de fin de suivi. La somme des contributions de chaque sujet donne le nombre total de personnes-années de la cohorte [1].

#### ◆ Calcul du nombre de décès (ou cas) attendus

Pour calculer le nombre de décès (ou cas) attendus, on applique les taux de mortalité (ou d'incidence) de la population de référence aux personnes-années de la population étudiée, subdivisée par genre (calcul séparé chez les hommes et les femmes), par classe d'âge (nombre de cas attendus calculés dans chaque classe d'âge) et par année (taux de référence de chaque année). Un exemple de taux de mortalité en 1995 et 2000 est présenté en annexe 4.

### 8.2.2 Calcul de SMR (ou SIR)

Le rapport du nombre de décès observés sur le nombre de décès attendus est appelé ratio standardisé de mortalité ou "SMR" du fait des initiales anglaises de l'expression (standardized mortality ratio).

$$\text{SMR (ou SIR)} = \frac{\text{nombre de décès (ou cas) observés}}{\text{nombre de décès (ou cas) attendus}}$$

Les tableaux suivants montrent le principe du calcul d'un SMR pour comparer la mortalité dans une cohorte de travailleurs d'une entreprise à celle d'une population de référence R (par exemple, l'ensemble des habitants du département où est située l'entreprise). Le SIR se calcule exactement de la même façon.

Population de référence R	
Age	Taux de mortalité (pour 1000)
20-29	1,5
30-39	1,9
40-49	4,5
50-59	13,6

Dans chaque classe d'âge, on multiplie les taux spécifiques de la population R par les personnes-années réparties selon l'âge dans la population étudiée ce qui permet d'obtenir des nombres de décès "attendus" par classe d'âge dans la cohorte. La somme permet d'obtenir le nombre total de décès attendus pour la période.

Cohorte étudiée			
Age	Personnes-années	Nombre de décès attendus	Nombre de décès observés
20-29	2 350	3,5	8
30-39	4 400	8,4	15
40-49	2 800	12,6	22
50-59	1 750	23,8	45
<b>Total</b>	<b>11 300</b>	<b>48,3</b>	<b>90</b>

Puis, on divise le total des cas effectivement observés dans la cohorte par le nombre "attendu" si la population étudiée se comportait comme la population de référence en tenant compte de la structure d'âge.

$$\text{SMR} = 90 / 48,3 = 1,86$$

La mortalité de la cohorte est ici supérieure à celle de la population de référence de 86 %.

Ces méthodes permettent de vérifier qu'un excès de cas (ou décès) dans la population de l'entreprise étudiée est bien réel (et n'est pas lié à la structure d'âge de la population étudiée par exemple).

#### Référence :

1. Moulin JJ, Clavel F. Etudes épidémiologiques de cohortes. *Encycl Méd Chir* (Elsevier, Paris), toxicologie-pathologie professionnelle, 16-870-A-10, 1999, 6p.

## Annexe 3. Formule du SMR, du SIR et de leur intervalle de confiance

### 8.3.1 Formule du SMR et du SIR

Le SMR est le rapport entre le nombre de décès observés et le nombre de décès attendus. Le nombre de décès attendus est calculé en appliquant à chaque tranche d'âge de la population étudiée les taux de mortalité de la population de référence (dans la même tranche d'âge) :

$$\text{SMR} = \sum a_i / \sum A_i, \text{ avec } A_i = T_i P_i$$

avec :  $a_i$  : nombre de décès observés dans la classe d'âge  $i$ .

$A_i$  : nombre de décès attendus dans la classe d'âge  $i$ , sous l'hypothèse de taux de mortalité identique à ceux de la population de référence.

$T_i$  : taux de mortalité de cancers de la population de référence pour chaque classe d'âge  $i$ .

$P_i$  : nombre de personnes-années de la cohorte pour la classe d'âge  $i$ .

Les classes d'âge des données de mortalité de cancers de référence sont les suivantes : 15-19 ans, 20-24 ans, 25-29 ans, 30-34 ans, etc...

Pour obtenir la formule du SIR, il suffit de remplacer "décès" par "cas", et "mortalité" par "incidence".

### 8.3.2 Formule de l'intervalle de confiance du SMR et du SIR (niveau de confiance 95 %)

La formule utilisée (intervalle de confiance basé sur la transformation racine carrée) est la suivante [1] :

$$\text{Borne inférieure} = a/A [1 - (1/9a) - (1,96/3\sqrt{a})]^3$$

$$\text{Borne supérieure} = (a+1)/A [1 - (1/9(a+1)) + (1,96/3\sqrt{(a+1)})]^3$$

Avec :  $a$  = nombre de décès observés (dans l'entreprise).

$A$  = nombre de décès attendus (en population générale, chez des sujets de la même classe d'âge).

$\alpha$  = risque d'erreur de 1<sup>ère</sup> espèce = 0,05.

Pour obtenir la formule de l'intervalle de confiance du SIR, il suffit de remplacer le terme "décès" par "cas".

#### Référence :

1. Breslow NE, Day NE. Statistical methods in cancer research. Volume II – The design and analysis of cohort studies. IARC Scientific publications n°82, 1987.

## Annexe 4. Taux de mortalité par cancer en France en 1995 et 2000

Les tableaux ci-dessous présentent les taux de mortalité de référence tous cancers 1995 [1] et 2000 [2], taux estimés pour 100 000 personnes-années, par tranche d'âge et par genre dans la population générale française.

### HOMMES

Age (ans)	15-19	20-24	25-29	30-34	35-39	40-44
Taux de mortalité 1995	4,8	5,2	7,1	13,7	31,7	80,4
Taux de mortalité 2000	3,7	5,0	7,1	11,9	27,9	74,5

Taux de mortalité pour 100 000 personnes-années

### HOMMES (suite)

Age (ans)	45-49	50-54	55-59	60-64	65-69	70-74
Taux de mortalité 1995	180,8	303,9	457,9	727,9	1 073,6	1 388,7
Taux de mortalité 2000	171,4	301,6	472,8	712,0	1 035,0	1 425,9

Taux de mortalité pour 100 000 personnes-années

### FEMMES

Age (ans)	15-19	20-24	25-29	30-34	35-39	40-44
Taux de mortalité 1995	3,2	3,8	5,9	13,9	30,1	58,0
Taux de mortalité 2000	2,5	3,3	5,4	12,5	28,0	55,2

Taux de mortalité pour 100 000 personnes-années

### FEMMES (suite)

Age (ans)	45-49	50-54	55-59	60-64	65-69	70-74
Taux de mortalité 1995	101,8	148,3	202,1	288,8	413,4	550,0
Taux de mortalité 2000	94,4	146,5	207,8	286,7	396,1	543,8

Taux de mortalité pour 100 000 personnes-années

### Références :

1. Le cancer en France – Incidence et mortalité. Situation en 1995. Evolution entre 1975 et 1995. Réseau Francim.
2. Remontet L, Buemi A, Velten M, Jouglu E, Estève J. Evolution de l'incidence et de la mortalité par cancer en France de 1978 à 2000. Rapport édité par l'InVS, août 2003.

## Annexe 5. Taux d'incidence des cancers en France en 1995 et 2000

Les tableaux ci-dessous présentent les taux d'incidence de référence tous cancers 1995 [1] et 2000 [2], taux estimés pour 100 000 personnes-années, par tranche d'âge et par genre en France.

### HOMMES

Age (ans)	15-19	20-24	25-29	30-34	35-39	40-44
Taux d'incidence 1995	18,5	30,2	45,1	66,4	106,9	210,9
Taux d'incidence 2000	20,7	30,0	42,8	62,0	100,8	186,3

Taux d'incidence pour 100 000 personnes-années

### HOMMES (suite)

Age (ans)	45-49	50-54	55-59	60-64	65-69	70-74
Taux d'incidence 1995	278,8	470,7	1 036,5	1 496,5	1 582,2	2 084,9
Taux d'incidence 2000	353,3	575,0	910,8	1 369,2	2 015,3	2 643,0

Taux d'incidence pour 100 000 personnes-années

### FEMMES

Age (ans)	15-19	20-24	25-29	30-34	35-39	40-44
Taux d'incidence 1995	16,4	27,5	51,0	98,8	170,3	283,9
Taux d'incidence 2000	16,7	27,7	51,9	93,3	161,4	267,3

Taux d'incidence pour 100 000 personnes-années

### FEMMES (suite)

Age (ans)	45-49	50-54	55-59	60-64	65-69	70-74
Taux d'incidence 1995	394,9	504,4	650,7	798,3	788,0	897,0
Taux d'incidence 2000	412,6	534,5	662,4	795,9	919,3	1 042,9

Taux d'incidence pour 100 000 personnes-années

### Références :

1. Le cancer en France – Incidence et mortalité. Situation en 1995. Evolution entre 1975 et 1995. Réseau Francim.
2. Remontet L, Buemi A, Velten M, Jouglu E, Estève J. Evolution de l'incidence et de la mortalité par cancer en France de 1978 à 2000. Rapport édité par l'InVS, août 2003.

## Annexe 6. Lexique

Ce lexique a été rédigé principalement à partir de [1,2].

### Biais

Tout effet qui tend à produire une estimation (de la fréquence d'une maladie, de l'association entre une maladie et un facteur de risque...) différant systématiquement, en plus ou en moins, de la vraie valeur. Certains biais peuvent être contrôlés ou limités ; leur existence doit être prise en compte dans l'interprétation des résultats.

### Cas

Personne ayant la maladie ou le problème de santé que l'on cherche à étudier. La définition épidémiologique d'un cas ne correspond pas toujours à ce que serait une définition clinique.

### Cluster

Mot anglais qui signifie "agrégat". Un cluster est donc un regroupement inhabituel d'un problème de santé dans un espace – temps défini. On parle aussi "d'agrégat spatio-temporel".

### Cohorte

Ensemble de sujets nés à une même période et suivis dans temps. De façon plus large, ensemble de sujets ayant vécu une même expérience, et suivis depuis la date de cette expérience (qui peut différer d'un sujet à l'autre).

Le suivi de la cohorte est organisé de façon à repérer un ou plusieurs événements donnés, très souvent le décès avec la cause du décès, ou la survenue d'une pathologie. En pathologie professionnelle, une cohorte est le plus souvent un ensemble de personnes ayant été exposées professionnellement à une substance donnée à une date connue. Exemple : ensemble des travailleurs ayant travaillé au moins 5 ans dans un établissement entre le 01/01/1940 et 31/12/1981.

### Enquête de cohorte

Dans une **enquête de cohorte**, on mesure l'exposition des sujets à un certain moment et on étudie leur devenir du point de vue de leur maladie. Un inconvénient majeur de ce type d'enquête, appelée aussi enquête prospective, est qu'elle peut exiger un délai très long entre le début de l'étude et l'obtention des premiers résultats.

Pour parer à cet inconvénient, la même étude peut être réalisée de façon historique : dans une **étude de cohorte historique**, l'investigateur définit les groupes étudiés sur la base de l'exposition et compare leur devenir depuis cette date. Les études de cohorte historique sont probablement le type d'étude le plus fréquemment employé pour les enquêtes de morbidité et de mortalité en épidémiologie des risques professionnels.

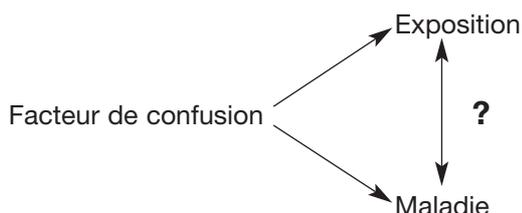
### Exhaustif / exhaustivité

Un recueil de données est exhaustif si l'échantillon est constitué par l'ensemble de la population.

### Facteur de confusion

Il s'agit d'un facteur professionnel ou non professionnel qui remplit deux conditions :

- être un facteur causal connu pour la maladie étudiée,
- être inégalement réparti dans les groupes exposés et non exposés. Un tel facteur est donc lié à la maladie et à l'exposition.



Il est susceptible de modifier la relation existant entre la maladie et l'exposition étudiée, voire créer artificiellement une relation qui n'existe pas en réalité. Il peut conduire à une interprétation erronée des résultats.

### **Incidence**

Nombre de nouveaux cas (n) observés durant une période donnée, T années par exemple, rapporté au nombre de sujets N à risque pendant cette période.

Incidence annuelle =  $n / NT$ .

Les événements comptés comme cas sont le plus souvent des maladies. S'il s'agit de décès, on parle de mortalité.

### **Population**

La population étudiée doit être définie de façon précise.

Exemple : pour la population d'une entreprise, il faut préciser : salariés entre 1995 et 2003, avec tel type de contrat, inclusion ou non des travailleurs employés de façon précaire, salariés des entreprises sous-traitants, etc...

L'effectif de la population étudiée doit être connu car il est indispensable pour calculer des taux. C'est notamment le cas où des événements de santé doivent être recensés de façon exhaustive.

### **Prévalence**

Rapport du nombre de personnes affectées par une maladie à l'effectif de la population susceptible de présenter la maladie, à un instant donné (prévalence instantanée).

### **SMR / SIR**

Rapport comparatif d'incidence : rapport entre le nombre de cas observés et le nombre de cas attendus. C'est la méthode de **standardisation indirecte**, dans laquelle on calcule un nombre d'événements attendus dans la population étudiée, en appliquant à cette population (pour chaque classe d'âge) les taux spécifiques d'une population de référence.

### **Standardisation**

Les mesures de morbidité ou de mortalité doivent être standardisées de façon à permettre des comparaisons entre populations éliminant l'effet de l'âge et du genre.

### **Taux d'incidence**

Taux : mesure de fréquence de survenue d'un phénomène dans une population par unité de temps.

Exemple : taux d'incidence annuel.

### Références :

1. Leclerc A, Lellouch J. Lexique statistique et épidémiologique, *Encycl Méd Chir* (Paris, France), Intoxications, 16980 A10, 1-1986, 13 p.
2. Moulin JJ, Wild P. L'épidémiologie en milieu professionnel. INRS, *Travail et Sécurité* 1995; 533:98-104.