



## **Les relations dose-effet**

*Phase 1 :*

*Analyse critique des méthodes et des modèles*

Document présenté au CEA

(à l'attention de M. Bertrand Mercier)

par la

Société de Calcul Mathématique S. A.

en application du Marché no 40003007131, notifié le 13 octobre 2007

rédaction Bernard Beauzamy et Francis Jolly

version 2, 2 janvier 2008 (tient compte des commentaires de A. Aurengo)

## **A. Présentation générale**

Notre contrat concerne l'étude des méthodes mathématiques employées pour déterminer le lien entre la dose et l'effet :

- Par "dose", on entend la quantité de radiations reçue, soit de manière normale (par les travailleurs du nucléaire), soit de manière accidentelle : exposition post-Tchernobyl, par exemple.
- Par "effet", on entend l'apparition d'une maladie (par exemple un cancer d'un certain type) ou le décès en général ou secondaire à une cause déterminée.

Le présent document analyse les méthodes existantes, de manière critique. Les documents suivants proposeront des méthodes nouvelles, plus robustes, pour remédier aux défauts que nous mettons en évidence ici.

## **B. Organisation du présent document**

Le présent document est organisé comme suit :

- Nous commençons par définir les objectifs ;
- Nous examinons en quoi consistent les données disponibles ;
- Nous expliquons ce qu'est un "modèle" et comment les modèles sont utilisés ;
- Nous passons en revue les différents indicateurs ;
- Nous faisons une analyse critique des différentes études ou articles. Ils se décomposent en deux catégories : ceux qui ne mentionnent pas les hypothèses faites et ceux qui les mentionnent.

## **C. Conclusions méthodologiques**

La situation que nous rencontrons ici est pour nous tout à fait habituelle : il y a peu de données, beaucoup de facteurs à prendre en considération et beaucoup d'incertitudes, à la fois sur les données et sur les facteurs. Mais nous avons rencontré des situations bien pires, où les données étaient beaucoup plus rares : tout ce qui touche au nucléaire fait l'objet d'un suivi attentif et les rayonnements ionisants sont le facteur de risque de cancer qui a fait l'objet du plus grand nombre d'études épidémiologiques, chez l'animal ou biologiques.

Par exemple, lorsque nous avons travaillé avec l'INERIS (1997-2000) sur les relations dose-effet en écotoxicologie, les données étaient beaucoup moins nombreuses, les phéno-

mènes physiques moins connus, et les paramètres à prendre en considération moins bien définis.

Un modèle est une nécessité pour pallier l'absence de données : on fait par exemple l'hypothèse que l'effet est fonction linéaire de la dose reçue, et il ne reste alors que deux paramètres à caler, qui sont les coefficients  $a$  et  $b$  de la relation linéaire  $effet = a \times dose + b$ . Si on fait l'hypothèse que l'effet est fonction quadratique de la dose, il y aura trois paramètres à caler, etc. A l'évidence, cela simplifie considérablement le problème et permet l'utilisation de méthodes statistiques simples : les paramètres sont en général calés par ajustement quadratique : on recherche par exemple la droite qui s'ajuste au mieux aux données, au sens des moindres carrés. N'importe quel logiciel de statistique fait cela très bien, et aucune réflexion n'est requise.

Mais on voit ici l'importance que revêt le choix du modèle. Selon qu'on l'aura choisi linéaire, linéaire quadratique, quadratique, logistique ou n'importe quoi, les résultats seront évidemment différents. Le résultat que l'on attend est généralement dans des plages de valeur (faibles doses) pour lesquelles on ne dispose pas d'observations ; on cale le modèle sur les valeurs disponibles, et on extrapole : on voit ici que le résultat sera tout différent selon qu'on extrapolera avec une droite, une exponentielle, une parabole,... Ce phénomène est apparu très clairement dans notre travail récent pour le CEA en sismologie.

Dans certains cas, le modèle utilisé repose sur une loi physique : on sait par exemple que la chute d'un corps suit une accélération constante, jusqu'à ce que la vitesse limite soit atteinte. Mais ce n'est pas le cas ici - cette remarque est importante - les modèles choisis le sont uniquement sur considérations de simplicité statistique. Ils n'ont pas de fondement biologique.

Aucun des modèles employés dans les études consultées ici n'a jamais été validé : ils correspondent à telle ou telle hypothèse, faite par certains experts et combattue par les autres. Dans ces conditions, notre conclusion méthodologique est très simple :

*Aucune conclusion de nature épidémiologique ne peut être tirée d'une étude reposant sur un modèle choisi a priori, tout simplement parce que la conclusion dépend de la nature du modèle, et cette dépendance n'est jamais validée.*

Certains auteurs essaient plusieurs modèles sur leurs jeux de données, mais simplement pour constater que cela change ou non les conclusions. La validité des modèles n'est pas réellement analysée, par exemple quant à sa crédibilité biologique.

Dans la suite de notre travail, nous proposerons des méthodes robustes, de nature probabiliste, destinées à se substituer aux modèles factices employés dans les études que nous avons consultées.

Mais disons-le clairement : les données, dans les sujets qui nous occupent ici, sont suffisamment abondantes pour que le recours à des modèles factices ne soit pas impératif. Par exemple, les cohortes des travailleurs du nucléaire regroupent plusieurs centaines

de milliers de personnes. Les modèles a priori ne sont pas indispensables dans de telles situations. Ils ne sont employés, par les différentes équipes, que par souci de simplicité et par ignorance des méthodes probabilistes : il suffit d'appuyer sur une touche, et le logiciel fait le test correspondant. En vérité, les différentes équipes ne sont même pas conscientes du fait suivant : le résultat qu'elles obtiennent dépend de manière fondamentale du modèle qu'elles ont choisi.

Une difficulté supplémentaire tient au cloisonnement des organismes, ce qui complique l'exploitation des données. Pour beaucoup d'organismes en charge de ces questions, les données recueillies caractérisent la légitimité administrative de l'organisme et ils considèrent que cette légitimité administrative se traduit automatiquement par une compétence appropriée, ce qui est faux à l'évidence. Bien pis, le fait d'avoir un monopole d'accès induit une tentation au repli sur soi : ces organismes ne voudront pas avoir recours à des mathématiciens extérieurs, considérant qu'ils ont eux-mêmes les compétences nécessaires. Le résultat est doublement négatif :

- L'organisme ne voit pas les difficultés méthodologiques liées à l'analyse mathématique et livre des conclusions sans valeur ;
- L'exploitation globale des données ne peut être faite, faute de comparaison entre les différentes sources.

Comme disait M. Jacques Repussard, Directeur Général de l'IRSN (entretien au Figaro, 22 mars 2005) : "C'est dangereux d'avoir le monopole de l'expertise". Il est encore plus dangereux de se croire seul compétent.

Les raisons invoquées par les différents organismes pour refuser de communiquer leurs données tiennent souvent à la loi "Informatique et Libertés" et aux Comités d'Ethique. Ces raisons ne résistent pas à l'examen : pour une étude statistique, on peut parfaitement se contenter de données anonymes. La loi "Informatique et Libertés" est ici invoquée à tort, sans aucune justification, et le plus simple pour le démontrer serait que les organismes concernés déposent une demande.

Aucune des études analysées ici ne montre, de manière significative et crédible, un surcroît de risque quelconque pour les faibles doses, mais les divergences méthodologiques entre experts, reprises et amplifiées par la presse, ont un effet négatif sur la population, qui se demande toujours si on lui cache quelque chose.

Selon le baromètre IRSN 2007-p 76, une majorité de Français considèrent qu'on ne dit pas la vérité sur les dangers des radiographies médicales, centrales nucléaires, déchets radioactifs, radon dans les habitations et retombées de Tchernobyl.

De plus, nombre d'organismes voient leur légitimité dans le fait qu'ils peuvent susciter des craintes, ce qu'ils auront évidemment tendance à faire. La phrase "aucun effet décelable n'a été mis en évidence" n'est presque jamais prononcée.

Un premier travail à faire serait de définir un "format de données anonymes" permettant l'exploitation statistique. Le CEA peut prendre l'initiative d'une telle démarche et la

soumettre à la CNIL : il s'agit de savoir sous quelle forme les données d'épidémiologie doivent être mises pour que a) les traitements mathématiques puissent être faits, b) l'anonymat des différents acteurs soit préservé. Il est trop tôt pour proposer un format précis, mais nous pouvons déjà dire que, pour un traitement mathématique :

- Les renseignements nominatifs concernant chaque relevé ne nous sont pas utiles ;
- Nous n'avons besoin que d'un âge approximatif (un âge par tranche de 5 ans suffit ; il faut avoir au moins cette précision pour les enfants de moins de 15 ans)
- Nous n'avons besoin que d'un lieu approximatif. Ici, tout dépend des autres informations qui sont recueillies, par exemple l'exposition à tel ou tel endroit. Mais les êtres humains se déplacent, et la précision d'un lieu est rarement correctement établie.

Terminons cette introduction par une remarque d'ordre méthodologique : les résultats obtenus devraient systématiquement prendre en compte les incertitudes (ou les insuffisances) sur les données d'entrée. On ne devrait pas a) éliminer les données considérées comme aberrantes (comment le décide-t-on ?) et b) considérer toutes les autres comme exactes, ou faire comme si elles l'étaient. On devrait au contraire, pour chaque information, incorporer l'incertitude en entrée, et en déduire une incertitude en sortie.

Aucune des études que nous avons consultées ne satisfait à cette précaution, pourtant nécessaire si elles veulent servir de base à une démarche normative ou réglementaire.

Pour terminer ce rapide tour d'horizon, nous dirons que, certes, les lois biologiques reliant les doses reçues à l'apparition de cancers sont mal connues, mais que l'abondance des données recueillies devrait suffire à obtenir des conclusions probabilistes de bonne qualité.

\*\*\*

## Table des matières

A. Présentation générale .....	2
B. Organisation du présent document.....	2
C. Conclusions méthodologiques .....	2
II. Les données .....	7
III. Les types d'études et de modèles.....	9
A. Les types d'études .....	9
B. Les modèles .....	9
1. Modèle à seuil .....	10
2. Modèle quadratique.....	10
3. Modèle supra-linéaire.....	11
4. Modèle linéaire .....	11
IV. Analyse des différentes études .....	12
A. Epidémiologie sans valeur d'exposition .....	12
B. Etude des valeurs d'exposition.....	15
C. Etude sur les cohortes de travailleurs .....	17
D. Quantification du risque lié à l'exposition au radon .....	19
E. Etude sur les zones à forte radioactivité naturelle.....	21
F. Etude du cancer en Europe lié à l'accident de Tchernobyl.....	22

## I. L'objectif

L'objectif est de détecter une augmentation d'un taux de cancer dans une population spécifique. Un grand nombre de facteurs, qui peuvent favoriser certains types de cancer, sont pris en compte dans les modèles. Les principaux sont l'âge, le sexe, la dose reçue, le débit de dose, le tabagisme...

Nous avons analysé deux types d'études :

- les études sur des populations en contact régulier avec des sources radioactives (travailleurs dans le domaine du nucléaire) ;
- les études réalisées après un accident (par exemple à propos des liquidateurs de Tchernobyl) où la dose est reçue dans un intervalle de temps très court.

Le principal problème rencontré concerne les effets des faibles doses d'exposition. En effet, au delà d'un certain seuil, les risques pour l'homme ont été clairement identifiés, depuis les études portant sur Hiroshima–Nagasaki (pour la plupart des cancers solides et des leucémies), sur Tchernobyl (pour les cancers thyroïdiens de l'enfant) et sur les patients ayant bénéficié de traitements par radiothérapie.

## II. Les données

Les calculs réalisés pour évaluer l'excès de risque dépendent des données recueillies. Les modèles utilisés afin d'élaborer une relation dose-effet correspondent plus à une tradition intellectuelle qu'à une nécessité : les données sont en général en nombre suffisant.

Les informations sur les décès, comportant les causes de décès, sont un minimum afin d'effectuer une comparaison entre deux groupes de personnes.

Le nombre de cancers constaté dans une population est aussi utilisé. On peut alors déterminer le risque par rapport à une population de référence. Ces données sont disponibles dans des registres où l'on peut consulter le nombre et le type de cancers dans une région administrative ou un département. Ce point est délicat, car on n'a pas toujours des données fiables pour la mortalité liée à un cancer donné, à une date donnée. Souvent les études prennent pour référence non pas des données de registres, mais une fraction non exposée de la population de l'étude.

Pour étudier les relations dose-effet, la dose efficace doit être calculée en déterminant préalablement la dose absorbée pour chaque organe ou groupe d'organes. Cette détermination est simple quand il s'agit d'une exposition externe homogène ; elle est beaucoup plus délicate et repose sur de nombreux paramètres connus de manière approximative en cas de contamination cutanée ou interne.

L'unité utilisée pour quantifier la dose efficace est le Sievert. Cette dose ne se mesure pas mais s'obtient par calcul : pour une contamination atmosphérique par exemple, les

doses efficaces sont évaluées en déterminant la quantité de radioactivité rejetée puis en évaluant sa dispersion dans l'atmosphère. Cela permet ensuite d'estimer la quantité de radioactivité inhalée par les personnes. Il est alors possible de d'évaluer la dose efficace reçue par les personnes par un calcul :

$$dose\ efficace = dose\ absorbée .WR.WT$$

où :

- $WR$  est le coefficient représentant la dangerosité du rayonnement ;
- $WT$  est le coefficient représentant la sensibilité de l'organe soumis à la radiation

Notons au passage que ce calcul fait appel à des coefficients fixés de manière approximative.



### III. Les types d'études et de modèles

#### A. Les types d'études

Dans une étude épidémiologique, les principaux éléments à considérer sont :

- le critère de qualité, la puissance statistique de l'étude ;
- les incertitudes sur la dose reçue ;
- l'interprétation des résultats dans les cas de comparaisons multiples ;
- la mesure du risque, y compris du risque vie entière ;
- la transposition du risque absolu sur d'autres population, interactions avec d'autre cancérigènes ;
- la prise en compte des variables dépendant du temps ;
- la prise en compte du débit de dose (exposition de longue durée, dose flash, dose fractionnée) ;
- la prise en compte de l'âge au moment de l'exposition.

Pour rechercher un excès de cancer radio-induit, deux types d'études existent :

- Une étude de cohorte ;
- Une étude cas témoin.

L'étude de cohorte permet de comparer le nombre de cancers au sein d'une population irradiée avec le nombre de cancers dans une population témoin (ayant la même répartition en âge et en sexe).

Dans l'étude cas-témoin, on part de la population atteinte pour rechercher un facteur de cause. Ces études comparent l'exposition antérieure à un ou des facteurs de risque, des sujets ayant développé un certain type de cancer (appelés « cas ») par rapport à des sujets indemnes de ce cancer (appelés « témoins »). A chaque cas devraient être appariés au moins 2 témoins (même âge, sexe...), mais ce n'est pas toujours fait, loin de là.

#### B. Les modèles

On appelle "modèle" une relation mathématique, prédéfinie (et donc arbitraire), reliant la dose reçue et le paramètre que l'on veut étudier (l'effet).

Dans les études que nous avons analysées, différents modèles sont utilisés, reliant l'excès de cancer avec la dose reçue :

- Les modèles avec ou sans seuil ;
- Le modèle linéaire ;
- Le modèle linéaire-quadratique ;
- Le modèle quadratique ;

- Le modèle supra-linéaire.

Aucun de ces modèles ne provient d'une loi biophysique, qui décrirait par exemple l'effet de la radioactivité sur les cellules, ou sur l'ADN, ou bien qui utiliserait un mécanisme d'apparition des cancers. Ce sont au contraire des modèles purement statistiques.

Les caractéristiques grossières de ces modèles sont représentées dans la figure ci-dessous.

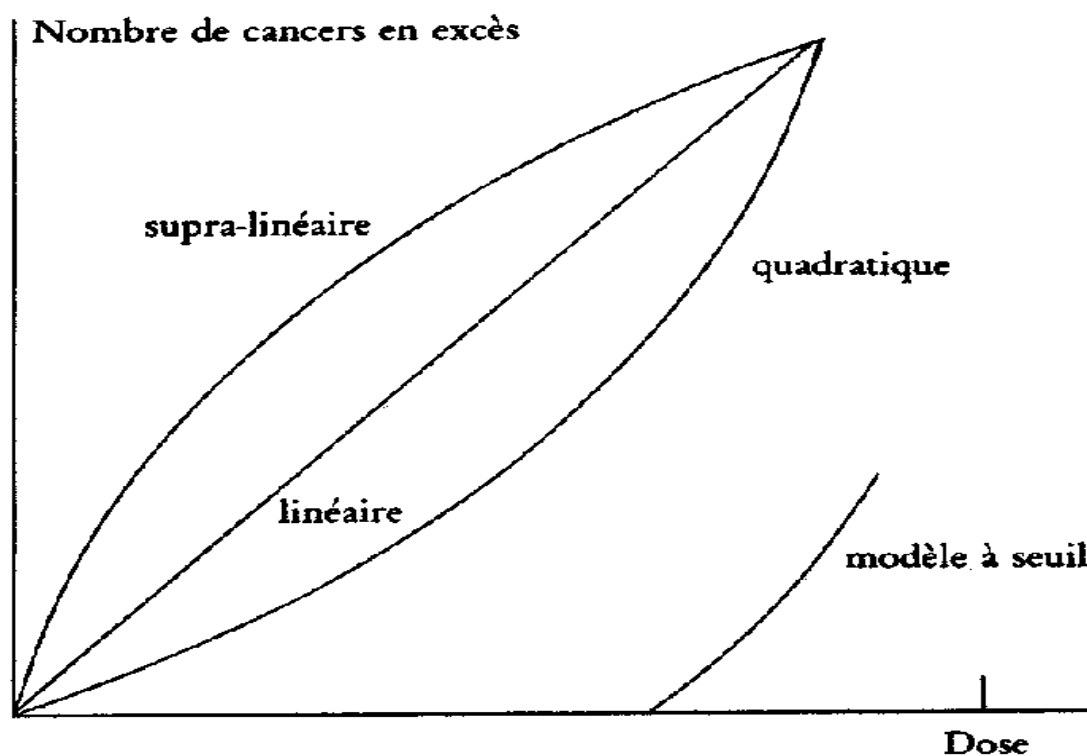


Figure 1 : représentation des différents modèles dose-effet

### 1. Modèle à seuil

Le modèle à seuil ne donne pas d'effet en dessous d'une certaine dose.

### 2. Modèle quadratique

Dans le modèle quadratique, l'effet par unité de dose est renforcé quand les doses augmentent.

L'UNSCEAR (le Comité Scientifique des Nations Unies sur les effets des radiations atomiques) a adopté un modèle quadratique basé sur le risque absolu pour la leucémie.

Le modèle de risque adopté comprend :

- une relation linéaire quadratique avec la dose ;
- une dépendance avec le sexe ;

— une dépendance avec le temps depuis l'exposition.

La formule retenue pour le risque absolu est :

$$\Delta AR(D,t,s) = \alpha_{t,s} (D + \theta D^2) e^{\beta_{t,s}(25-t)}$$

où  $\theta$  est une constante,  $D$  la dose et  $\alpha, \beta$  sont des paramètres qui dépendent du sexe et du temps écoulé depuis l'exposition.

Ce modèle est factice ; il ne peut estimer correctement le risque pour toutes les gammes de dose. Les formes retenues pour les lois sont réalisées à partir d'ajustements statistiques.

### 3. Modèle supra-linéaire

Dans le modèle supra-linéaire, l'effet par unité de dose est renforcé quand les doses diminuent.

### 4. Modèle linéaire

Dans le modèle linéaire sans seuil, toute dose produit un effet qui lui est proportionnel.

Les modèles prennent en compte les paramètres qui influencent le nombre de cancers : il s'agit de l'âge, de l'âge à l'exposition et de la dose.

Le modèle linéaire, approuvé par BEIR VII (Académie Nationale des Sciences des Etats Unis) et par le CIPR (Comité National de Protection Radiologique), est contesté par d'autres organismes (par le réseau international scientifique et par les Académie des Sciences et de Médecine) qui considèrent qu'il n'y a pas d'effet scientifiquement prouvé des faibles doses sur l'organisme.

La réponse apportée par les modèles doit être clairement définie : le modèle est-il fait pour la gestion des risques ou bien pour prédire un nombre de cancers sur une population exposée ? Pour ce dernier cas l'utilisation d'un modèle linéaire surestimerait considérablement le nombre de cancers dans une population dans les cas d'exposition à faible dose et à très faible dose.

Notons que, de manière générale, les modèles linéaires ont la faveur du public et des non-mathématiciens, toutes disciplines confondues : ils ont le mérite de la simplicité. Le fait que la nature ne soit jamais linéaire n'arrête jamais les auteurs.

Les modèles adoptés reposent généralement sur l'hypothèse d'une extrapolation linéaire vers les faibles doses.

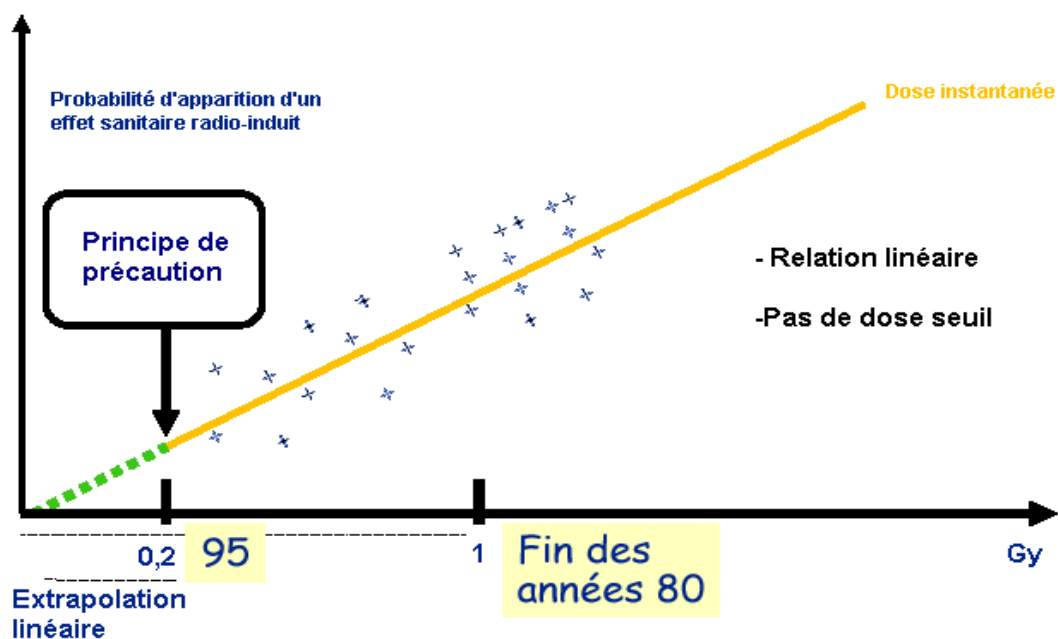


Figure 2 : modèle linéaire avec extrapolation aux faibles doses

Pour de tels modèles, l'extrapolation linéaire entraîne une surestimation du risque d'apparition d'un cancer, pour les faibles doses.

Les modèles les plus récents pour l'évaluation des relations dose-effet sont :

- BEIR VI (2006, Phase 2 Board on Radiation Effects Research ; Health Risks from Exposure to Low Levels of Ionizing Radiation);
- NCRP (2001 Evaluation of the linear nonthreshold dose-response model for ionizing radiation NCRP report);
- UNSCEAR (2000, Sources and effects of ionizing radiation, UNSCEAR 2000 report to the General Assembly with scientific annexes).

## IV. Analyse des différentes études

### A. Epidémiologie sans valeur d'exposition

(articles du CépiDc, IFR69, Inserm et CEPN [1,2])

Ici, les estimations sont fondées sur des observations provenant directement de registres (comme les causes de décès recensées par l'Inserm). Des enquêtes ou des sondages sont aussi utilisés lorsque les données ne sont pas enregistrées. Ces indicateurs sont des définitions et utilisent, selon les données disponibles, la mortalité ou le nombre de maladies dans une population (incidence).

L'article « Principaux indicateurs de mortalité (CépiDc, IFR69, Inserm) » nous donne les principaux indicateurs utilisés en épidémiologie :

### Effectif de décès

L'effectif de décès est l'indicateur de mortalité le plus simple. Les effectifs de décès peuvent être analysés toutes causes confondues ou répartis par cause, pour une ou plusieurs années, en fonction des variables enregistrées.

### Taux brut de mortalité

Le taux brut de mortalité est un indice pour décrire le taux de mortalité dans une population en tenant en compte l'effectif de cette population. Il est calculé en rapportant les décès enregistrés dans la population pendant une période donnée à l'effectif de cette population. Ce taux est généralement calculé pour 1 000, 10 000 ou 100 000 habitants.

$$\text{Taux brut de mortalité pour } 1000\text{hab} = \frac{\text{Nombre de décès pour une période donnée}}{\text{Effectif de la population}} 1000$$

La mortalité peut être étudiée en fonction de différentes variables comme l'âge, le sexe, la catégorie socio-professionnelle ....

### Standardisation directe

La standardisation directe est utilisée lorsque les variables sont différentes entre la population de référence et la population observée. Par exemple, si la population observée n'a pas la même structure d'âge que la population de référence, la standardisation directe permet de prendre en compte cette différence.

Bien entendu, la méthode de standardisation est sujette à caution : on fait une hypothèse (pas toujours explicite) quand on passe d'une population à une autre.

### Standardised Mortality Ratio (SMR)

Les études réalisées sans valeur d'exposition comparent les effets observés par rapport aux effets attendus. Le SMR (Standardised Mortality Ratio) se calcule comme suit :

$$SMR = \frac{\text{mortalité observé}}{\text{mortalité attendue (mortalité type)}} 100$$

Un SMR supérieur à 100 indique un excédent de mortalité dans la population observée par rapport à la population de référence.

### Taux d'incidence

Le taux d'incidence permet d'évaluer la progression d'une maladie au cours du temps. Il est défini comme étant le nombre de nouveaux cas d'une maladie dans une population choisie au cours d'une année.

Le tableau suivant montre l'évolution du cancer ; il est extrait de l'article « Epidémiologie du cancer thyroïdien » (Médecine Nucléaire-2003-vol.27)

Période	Homme	Femme
1978-1982	1,1 (0,9-1,3)	3,4 (3,1-3,8)
1983-1987	1,5 (1,2-1,7)	3,9 (3,5-4,3)
1988-1992	1,6 (1,4-1,9)	5,3 (4,9-5,7)
1993-1997	1,9 (1,6-2,1)	7,1 (6,6-7,6)

Tableau 1 : Evolution temporelle du taux d'incidence pour 100 000 habitants standardisé à la population mondiale (intervalle de confiance à 95 %)

La variation de l'incidence au cours du temps n'est généralement pas fiable, car les diagnostics s'améliorent au cours du temps : une augmentation de l'incidence sera observée car les diagnostics récents permettent une meilleure détection de la maladie. Les études de taux d'incidence sont donc à considérer avec prudence si les périodes d'observation sont anciennes.

Le CEPN (Centre d'Etude sur l'Evaluation de la Protection dans le Domaine Nucléaire) décrit d'autres indicateurs couramment utilisés dans « les principes généraux de la modélisation » :

#### Excès de risque absolu

L'excès de risque absolu tient compte à la fois de la taille et de la durée de suivi de cette population, il est défini comme suit :

$$ERA = \frac{nb \text{ effets observés} - nb \text{ effets attendus}}{nb \text{ de personnes} - \text{années}}$$

#### Risque relatif

Le risque relatif est le rapport de risque entre un groupe exposé et un groupe non-exposé. Les effets observés sont donc comparés aux effets attendus.

$$RR = \frac{nb \text{ effets observés}}{nb \text{ effets attendus}}$$

Si le risque relatif est égal à 2, cela signifie que pour une exposition considérée, on a observé un doublement du nombre d'effets attendus.

#### Excès de risque relatif

L'excès de risque relatif (ERR) est défini comme étant le risque relatif moins un. Un ERR positif indique donc un risque plus élevé chez les personnes exposées.

$$ERR = \frac{nb\text{ effets observés}}{nb\text{ effets attendus}} - 1 = \frac{nb\text{ effets observés} - nb\text{ effets attendus}}{nb\text{ effets attendus}}$$

Ces calculs, contrairement aux calculs provenant d'études « dose-effet » n'utilisent pas d'informations sur la dose reçue par les personnes.

### *B. Etude des valeurs d'exposition*

(article de l'IRSN, Radioactivité en Césium 137 du sol français [3])

#### Objectif

L'IRSN interprète les données OPERA (Observatoire Permanent de la Radioactivité de l'Environnement) qui rassemble 1 200 données d'activité en Cs137.

Les données OPERA interprétées par l'IRSN dans l'étude ont été acquises entre 1993 et 2004.

#### Données

Les résultats d'activités mesurées comprennent l'activité massique en Cs137, l'incertitude de mesure, la limite de détection si aucune activité n'est détectée, le lieu et la date du prélèvement.

Les données sont réparties sur 7 stations, qui dans l'ensemble se situent en dehors des zones reconnues comme les plus contaminées en Cs137 (5 stations correspondent à des zones de dépôt relativement faible et 2 stations correspondent à des zones potentiellement plus touchées par les retombées de Tchernobyl).

#### Méthodes

Parmi les valeurs acquises, certaines sont apparemment très élevées. L'étude de l'IRSN exclut ces valeurs, considérées comme aberrantes.

Pour déterminer si une valeur est aberrante ou non, l'étude utilise un test statistique, basé sur l'hypothèse que les échantillons mesurés suivent une loi normale. Cette hypothèse ne repose sur aucun fondement scientifique ; elle est purement factice.

On réalise ensuite un test de Student, test standard en statistiques, qui permet de comparer les moyennes issues de deux échantillons et de décider si elles sont différentes ou non.

Ce test permet de comparer les échantillons de deux stations différentes, ou bien d'une même station, pour deux saisons différentes.

Malheureusement, le test de Student n'est valable que si les deux échantillons suivent chacun une loi normale, ce qui n'est pas le cas ici : nous retombons sur le même présupposé idéologique.

Ce test est réalisé de la manière suivante :

1. On calcule les moyennes  $X_1$  et  $X_2$  de chaque échantillon
2. On calcule les écarts type  $S_1$  et  $S_2$  (il s'agit ici des écarts-types empiriques)
3. On calcule la quantité :

$$t = \frac{X_1 - X_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

où  $n_1$  et  $n_2$  représentent le nombre de données dans chacun des échantillons respectivement.

4. On cherche l'écart critique  $t_a$  dans la table de Student en fonction du nombre de degrés de liberté et en fonction du risque d'erreur.
5. Si  $t < t_a$ , on accepte l'hypothèse si  $t > t_a$ , on rejette l'hypothèse.

Ce test est très simple à mettre en œuvre et il est robuste à condition que les données soient suffisamment nombreuses, ce qui est le cas ici. Mais il requiert que les lois soient normales, ce qui n'a aucune raison d'être le cas.

Pour deux échantillons, on peut également mesurer la relation existant entre deux séries de données, en analysant les corrélations. Le coefficient de corrélation usuel est défini par :

$$\rho_{X_1, X_2} = \frac{\text{cov}(X_1, X_2)}{\sigma_{X_1} \cdot \sigma_{X_2}}$$

La corrélation est d'autant plus forte que  $\rho_{X_1, X_2}$  est proche de 1.

Il est important de mentionner que le coefficient de corrélation ne détecte que les corrélations linéaires : voir par exemple le livre de Bernard Beauzamy "Méthodes probabilistes pour l'étude des phénomènes réels" [8]. Son utilisation pour détecter des liens quelconques est donc impropre. Dans le cas général, pour mettre en évidence des corrélations non linéaires, il est préférable d'utiliser des notions probabilistes appropriées, comme par exemple la variance conditionnelle, voir [8].



## Résultats

L'activité massique moyenne ainsi que les écarts type (dispersion des résultats) sont obtenus au final en fonction du lieu et de la date des relevés

Tous les résultats pris au cours du temps permettent la comparaison entre la décroissance radioactive issue des mesures et la période de décroissance théorique.

### *C. Etude sur les cohortes de travailleurs*

(article du British Medical Journal, Risk of cancer after low doses of ionising radiation: retrospective cohort study in 15 countries, 2005 [4])

## Objectif

Parmi les études sur des travailleurs dans le domaine nucléaire, l'étude réalisée par le CIRC (Centre International de Recherche sur le Cancer) prend en compte plus de 400 000 personnes réparties dans 15 pays. Elle permet d'étudier le risque relatif à faible dose des radiations ionisantes. Cette étude présente les estimations de risque de mortalité pour tous les cancers, et pour les leucémies, à l'exclusion des leucémies lymphoïdes chroniques.

## Données

L'étude sur les cohortes de travailleurs prend en compte les enregistrements dosimétriques de chaque individu ainsi que les données relatives à la santé (cause et date de décès).

## Méthodes

Pour savoir si les travailleurs affectés par des doses de rayonnements les plus élevées courent un risque de cancer plus élevé, des modèles statistiques ont été utilisés. Ces modèles prennent en compte différents facteurs liés au cancer, comme l'âge, le sexe, la période et le statut socio-économique.

La mortalité observée de la cohorte est comparée à la mortalité nationale en calculant le SMR (rapport mortalité observée sur mortalité type).

Le modèle du risque relatif est basé sur un modèle linéaire avec régression de Poisson de forme  $1 + \beta Z$ , où  $Z$  est la dose cumulée en Sievert et  $\beta$  l'excès de risque relatif par Sv. Une dépendance de type  $1 + \beta Z$  est un cas particulier de modèle linéaire (avec terme constant égal à 1). Il s'agit ici d'un modèle linéaire sans seuil

Le mot "régression" est d'usage courant en statistiques, mais il n'a aucun sens particulier. On veut simplement dire que la variable  $Z$  suit une loi de Poisson, c'est-à-dire que sa loi est de la forme :

$$P(Z = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

où  $\lambda$  est un paramètre, et on vérifie que  $E(Z) = \text{var}(Z) = \lambda$ . Il s'agit donc d'une loi très simple à utiliser.

### Résultats

Les résultats de l'étude des travailleurs dans le domaine du nucléaire sont comparés aux résultats de l'étude Hiroshima Nagasaki. Mais, selon l'Autorité de Sûreté Nucléaire (dans la présentation « Grandeurs et unités en radioprotection »), une même valeur de dose peut être reçue quasi instantanément, ou bien de façon continue ou fractionnée. Les effets biologiques produits seront alors différents.

Seules les moyennes des doses reçues sont données dans les différentes synthèses de cette étude.

Le nombre de cancers en fonction de la dose reçue est connu mais n'est pas dévoilé dans l'article. Sur l'histogramme ci-dessous, la répartition des travailleurs en fonction de la dose est clairement établie.

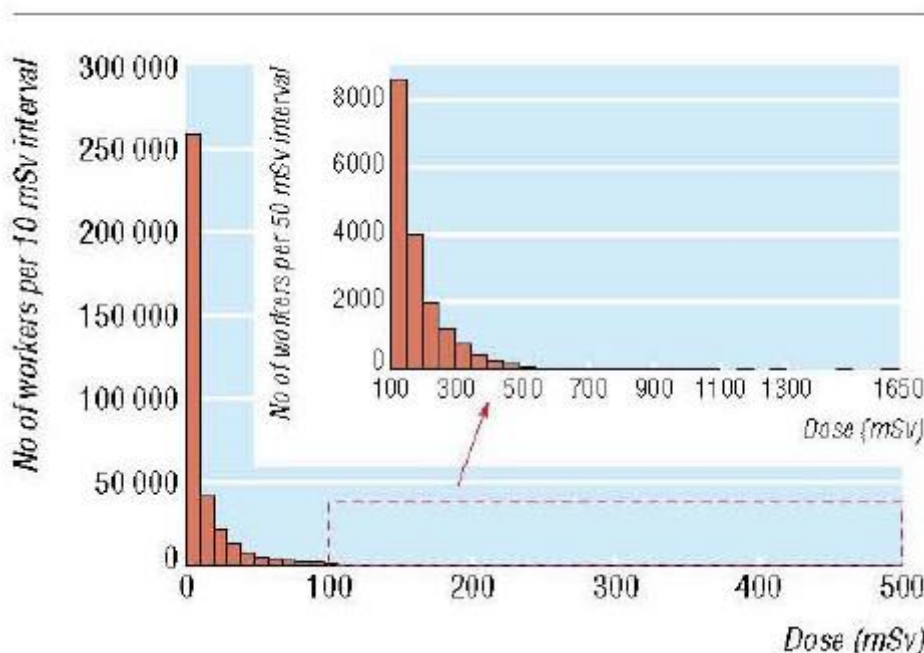


Figure 3 : répartition des travailleurs en fonction de la dose reçue

Plusieurs facteurs sont pris en compte dans l'étude, mais l'information sur le tabagisme des travailleurs manque. Or, le tabagisme est un facteur important car il est fortement lié avec le risque de certains types de cancers.

## *D. Quantification du risque lié à l'exposition au radon*

(article de l'IRSN, Quantification of lung cancer risk after low radon exposure and low exposure rate: synthesis from epidemiological and experimental data, 2003 [5])

### Objectif

Cette étude ressemble à celle des travailleurs dans le domaine nucléaire, car le but, ici, est de quantifier une relation dose-effet.

### Données

Les données prises en compte dans l'analyse sont : l'âge, l'âge d'exposition, le temps d'exposition, la cause de décès selon ICD (International Classification of Diseases), la moyenne annuelle d'exposition au radon ainsi qu'une information sur le tabagisme (âge au commencement, âge d'arrêt, nombre de cigarettes par jour) .

Le suivi des travailleurs est important pour garder le maximum de données pour l'étude statistique : le pourcentage de mineurs qui ont été perdus de vue (c'est-à-dire pour lesquels l'information sur le mineur n'est plus disponible) pendant l'étude est de moins de 3 %. Par contre les informations sur les causes de mortalité sont insuffisantes sur certaines cohortes : 13 % des décès sont de cause inconnue dans la cohorte allemande.

La perte d'information est préjudiciable pour l'analyse statistique : elle conduit à des incertitudes trop importantes. Il est important de s'en souvenir, et de ne pas "faire comme si" toutes les données étaient précises.

### Méthode

Comme dans l'étude sur les travailleurs dans le domaine nucléaire, la régression de Poisson est utilisée pour définir les modèles.

Le risque relatif obtenu varie fortement en fonction du modèle utilisé. La figure ci-dessous montre trois modèles différents estimant le risque en fonction de l'exposition.

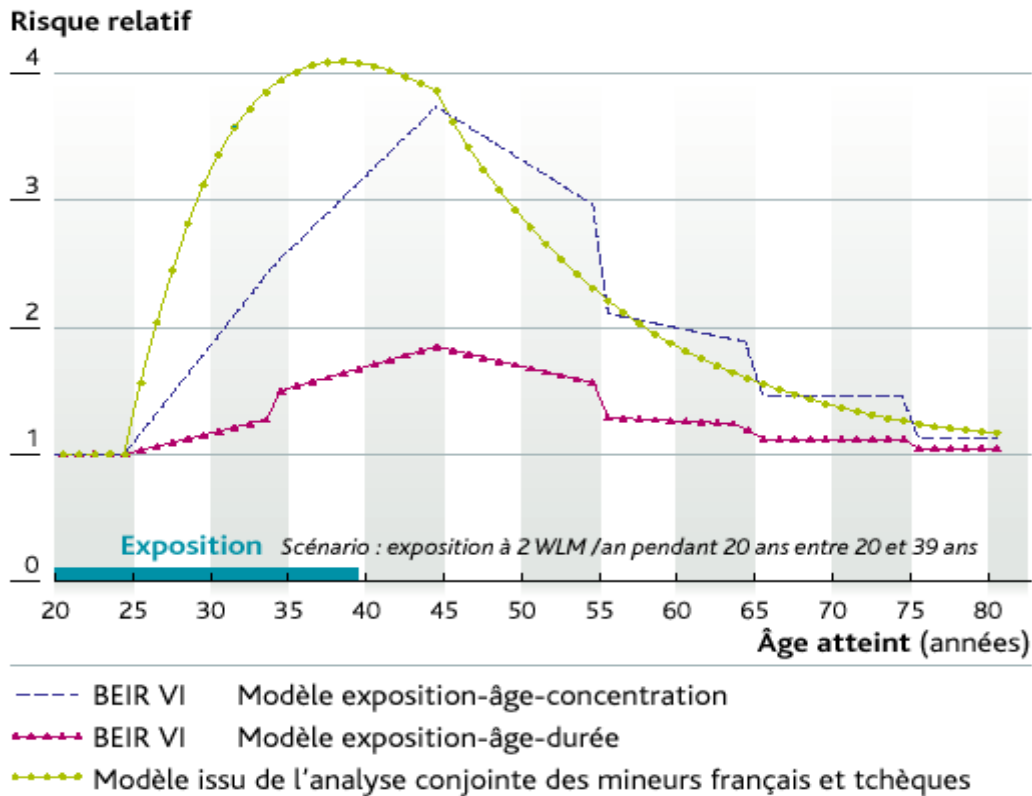


Figure 4 : risque relatif estimé par différents modèles en fonction de l'âge des travailleurs

Cette étude présente l'avantage de prendre en compte le tabagisme (mais pas pour toutes les cohortes).

D'autres modèles ont été réalisés pour estimer le risque relatif en prenant en compte l'exposition de radon ainsi que le facteur de tabagisme.

Les modèles de risque relatif (RR) sont les suivants :

$$RR = 1 + b \text{ WLM}$$

$$RR = r_s (1 + b \text{ WLM})$$

où :

- Le coefficient  $b$  est l'excès de risque lié à l'exposition ;
- $\text{WLM}$  (Working Level Month) est l'unité d'exposition au radon utilisée dans les mines.

Le deuxième modèle intègre l'excès de risque lié au tabac ( $r_s$ ). Mais les modèles utilisés dans l'étude sont nombreux et les résultats donnés divergent. Comme on le voit dans la figure 4, les facteurs pris en compte font varier fortement le risque relatif.

Dans ces études, le tabagisme n'est pas correctement quantifié. Sous réserve de vérification, la durée est connue mais pas l'« intensité », i.e. le nombre de cigarettes fumées

chaque jour. C'est l'inverse dans l'étude en milieu résidentiel. Pour certaines cohortes (en particulier la cohorte de travailleurs français), il n'y a aucune donnée sur le tabac.

## Résultats

La relation dose-effet est établie à partir des données d'exposition au radon et du risque de cancer des mineurs exposés à des faibles doses.

### *E. Etude sur les zones à forte radioactivité naturelle*

(IPSN, Etudes épidémiologiques dans des zones à haut niveau de radioactivité naturelle, 2000 [6]). Il s'agit, pour ces études particulières, d'expositions avec de très faibles débits de dose.

## Objectif

Comparaison des taux de mortalité ou de nombre de cancers entre des zones où la radioactivité naturelle est élevée à des zones « contrôles ».

## Données

L'étude sur les populations exposées à un haut niveau de radioactivité naturelle repose sur un indicateur moyen d'exposition et un indicateur de santé. La plupart des études écologiques sont réalisés à très large échelle (au niveau d'un pays entier par exemple).

Les données sont donc : la mortalité de la population étudiée, le nombre de cancers observés, la dose moyenne pour l'ensemble de la population.

## Méthodes

Il n'y a pas de modèle car les doses ne sont pas individuelles. Une dose moyenne est estimée pour l'ensemble de la population appartenant à une zone de forte activité naturelle.

De nombreuses campagnes de mesures sont réalisées à travers plusieurs pays (Chine, Inde, Brésil, Japon...) afin d'estimer les doses associées aux expositions.

Les études comparent la mortalité entre des zones « contrôles » et les zones de radioactivité naturelle élevée (entraînant pour la population locale une dose efficace individuelle supérieure à 5 mSv/an).

Les calculs ne sont pas détaillés. Ils ne sont pas basés sur des modèles, mais reposent sur des comparaisons de taux entre plusieurs groupes de population au sein desquels une dose moyenne a été estimée.

## Résultats

Les études réalisées ne mettent pas en évidence une modification des indicateurs de l'état de santé : aucune augmentation de la fréquence des cancers n'est montrée.

La plupart de ces études étant faites à une échelle très large, elles ne sont pas focalisées sur les zones avec un fort niveau d'exposition. De plus, comme les études ne reposent pas sur des situations individuelles, l'état de santé dans le temps pour chaque personne avec une estimation individuelle de la dose n'est pas disponible.

### *F. Etude du cancer en Europe lié à l'accident de Tchernobyl*

(article du Centre International de Recherche sur le Cancer : Le fardeau de cancer lié à Tchernobyl [7])

#### Objectif

L'étude a été réalisée dans le cadre d'un groupe de travail du CIRC, composé d'experts internationaux.

Dans cette étude, plusieurs approches ont été testées : la prédiction du nombre de cas de cancers et du nombre de décès par cancer à l'aide des modèles du risque radio-induit. Les taux d'incidence ou les taux de mortalité par cancer sont également étudiés.

#### Données

Le travail réalisé par le CIRC a consisté à :

- Mettre à jour la distribution des doses en Europe à l'aide de nouveaux modèles dosimétriques ;
- Examiner les tendances de l'incidence et de la mortalité par cancer en Europe dans le temps et par niveau de dose de rayonnement ;
- Estimer le nombre de cas de cancer qui peuvent être liés à l'exposition aux rayonnements au cours des 80 années suivant l'accident (jusqu'en 2065), en appliquant les modèles les plus récents (le modèle BEIR VII, de l'Académie des Sciences des Etats-Unis).

#### Méthodes

Le modèle utilisé est linéaire et sans seuil. Les prédictions calculées sont basées sur des modèles élaborés par le comité BEIR VII mais ne sont pas détaillées dans l'article.

Ce modèle est critiqué dans d'autres études car il surestime le risque de cancer pour les personnes exposées à de faibles doses, comme nous l'avons dit plus haut. L'article ne mentionne que les résultats finaux, les calculs ne sont pas visibles.

### Résultats

La prédiction du nombre de cancers en Europe, consécutifs à Tchernobyl, en fonction de la dose cumulée moyenne n'est pas justifiable : la dose moyenne cumulée de 0,5 mSv conduit à une prédiction d'un excès de 16 000 décès par cancer, alors que la dose naturelle accumulée est plus élevée.

Les données épidémiologiques actuelles, en ce qui concerne les faibles doses d'exposition, ne permettent pas d'affirmer l'existence d'un effet cancérigène proportionnel à la dose. Même si des modifications dans l'organisme sont détectables (stimulation des mécanismes de défense dans l'organisme), l'augmentation de cancer ne peut pas être confirmée.

### **Références**

- [1] Principaux indicateurs de mortalité (CépiDc, IFR69, Inserm).
- [2] Les principes généraux de la modélisation (Centre d'étude sur l'évaluation de la protection dans le domaine nucléaire).
- [3] Radioactivité en  $^{137}\text{Cs}$  de l'environnement terrestre français. Interprétation des données OPERA acquises de 1993 à 2004 (IRSN, S. Roussel-Debet, O. Masson, G. Sa-laun).
- [4] Risk of cancer after low doses of ionising radiation: retrospective cohort study in 15 countries (British Medical Journal, E Cardis, 2005).
- [5] Quantification of lung cancer risk after low radon exposure and low exposure rate: synthesis from epidemiological and experimental data (IRSN, M Tirmarche, 2003).
- [6] Etudes épidémiologiques dans des zones à haut niveau de radioactivité naturelle (IPSN, D Laurier, JM Martin, Ph Hubert, 2000).
- [7] Le fardeau de cancer lié à Tchernobyl (Centre International de Recherche sur le cancer, E Cardis, 2006).
- [8] Bernard Beauzamy : Méthodes probabilistes pour l'étude des phénomènes réels. Editions de la SCM, 2004.