



## **Les relations dose-effet**

*Phase 3 :*

*Prise en compte de la contamination*

Document présenté au CEA

*Direction de la Protection et de la Sûreté Nucléaire*

(à l'attention de M. Thierry de Bruyne)

par la

Société de Calcul Mathématique S. A.

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

rédaction Thomas Vanaret

février 2008

## Résumé opérationnel

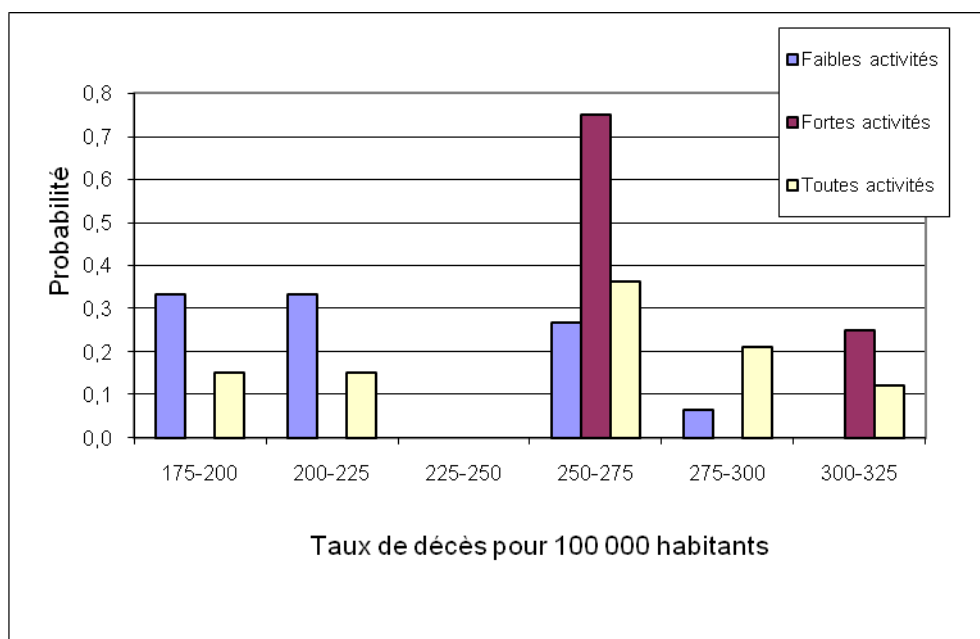
Dans le précédent rapport, nous avons montré comment utiliser les méthodes probabilistes pour comparer des régions entre elles, du point de vue du nombre de cancers. Dans ce rapport, nous montrons comment faire intervenir la contamination liée à la radioactivité.

Nous disposons seulement de données très fragmentaires : mesures de concentration en Césium 137, sur 5 régions (Aquitaine, Auvergne, Basse-Normandie, Ile de France, PACA), sur 12 années (1993 à 2004). Il s'agit de l'activité dans différentes espèces de bryophytes (mousses), mesurée par le réseau OPERA de l'IRSN. En parallèle, nous disposons des taux de cancers, par région et par an, déjà exploités dans notre rapport précédent. Nous montrons comment "combiner" ces informations pour obtenir la "loi conjointe" du couple exposition  $\times$  cancers et comment exploiter cette loi conjointe.

Notre conclusion est ici claire : les régions de plus forte exposition sont aussi celles où le taux de cancers est le plus élevé : pour les régions où l'activité est d'au moins 150 Bq/kg, le taux de cancers est toujours d'au moins 250 pour 100 000. Pour les régions où l'activité est entre 0 et 50 Bq/kg (faibles activités), deux sur trois sont au dessous de 225 décès pour 100 000.

Nous avons également pris en compte l'aspect temporel. Un cancer peut ne pas apparaître l'année de l'exposition ; nous avons repris nos calculs avec un taux de cancer cumulé sur cinq années suivant l'exposition. Les conclusions ne sont pas significativement différentes.

Sur le plan de la méthodologie probabiliste, on obtient ainsi des conclusions tout à fait claires : on voit sur le graphique ci-dessous que la loi de probabilité "fortes activités" (en rouge) est beaucoup plus "à droite" que la loi de probabilité "faibles activités" (en bleu).



Graphique 1 : répartition des différentes lois

Cependant, nous insistons beaucoup sur ceci : l'objet de ce rapport n'est qu'une présentation méthodologique. Nous avons très peu de données, et rien ne dit qu'elles soient représentatives : elles ont pu être recueillies dans des zones où l'on savait par avance que le taux de cancers serait élevé. Pour pouvoir conclure, il nous faudrait les mesures des taux d'exposition pour toutes les régions. Nous les avons demandés à l'IRSN, mais nous ne les avons pas obtenus à ce jour.

Une autre remarque méthodologique importante est qu'une activité élevée n'est pas nécessairement la cause d'un taux de cancers élevé. Notre méthode met en évidence une "dépendance conditionnelle", qui, grossièrement, se traduit par la phrase : là où l'activité est élevée, il y a plus de cancers. Mais ce peut être dû à une cause externe, qui influe sur les deux (caractéristiques du sol, de l'atmosphère, etc.). Nous n'avons aucun moyen de le savoir.

## Sommaire

I.	Les données disponibles.....	5
A.	Mesures de d'activité .....	5
B.	Mesures des effets : taux de décès par tumeurs .....	6
II.	Loi conjointe Activité - Effet .....	7
A.	Table d'occurrences.....	7
B.	Loi conjointe.....	8
C.	Test d'indépendance .....	8
D.	Lois conditionnelles .....	9
III.	Prise en compte de l'aspect temporel.....	10
A.	Méthode.....	10
B.	Table d'occurrence et loi conjointe.....	10
C.	Lois conditionnelles .....	11
IV.	Fonctions de répartition.....	13

## I. Les données disponibles

### A. Mesures d'activité massique

On utilise les mesures de l'activité du Cs137 présent dans différentes espèces de bryophytes (mousse) effectuées par le réseau OPERA de l'IRSN.

On dispose de mesures pour 5 régions entre les années 1993 et 2004. Les mesures ne sont pas systématiques. Différentes espèces de bryophytes ont été analysées de manière irrégulière.

Parmi les 8 espèces de bryophytes, seules deux ont été analysées sur plusieurs années consécutives (au mieux 10 ans) dans 3 régions : « mousse terrestre » et « mousse terrestre non identifiée ». Nous regroupons arbitrairement ces deux espèces.

Mesures Cs137	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004
<b>Aquitaine</b>										54.8	67.0	
<b>Auvergne</b>	142.3	163.5	134.9	112.6	113.2	77.1	63.4	75.5				
<b>Basse-Normandie</b>			17.8	17.6	12.4					1.3	1.1	
<b>Ile de France</b>			4.0	7.2	4.9	3.8	2.8	2.8	2.9	2.9	4.8	4.4
<b>PACA</b>	192.6	290.2	247.5	108.6	77.3	89.6	131.1	104.5				

Tableau 1 : Données « activité » exploitées

Les mesures sont exprimées en Bq/kg de mousse. A titre de comparaison, la radioactivité naturelle d'un être humain est de l'ordre de 100 Bq/kg.

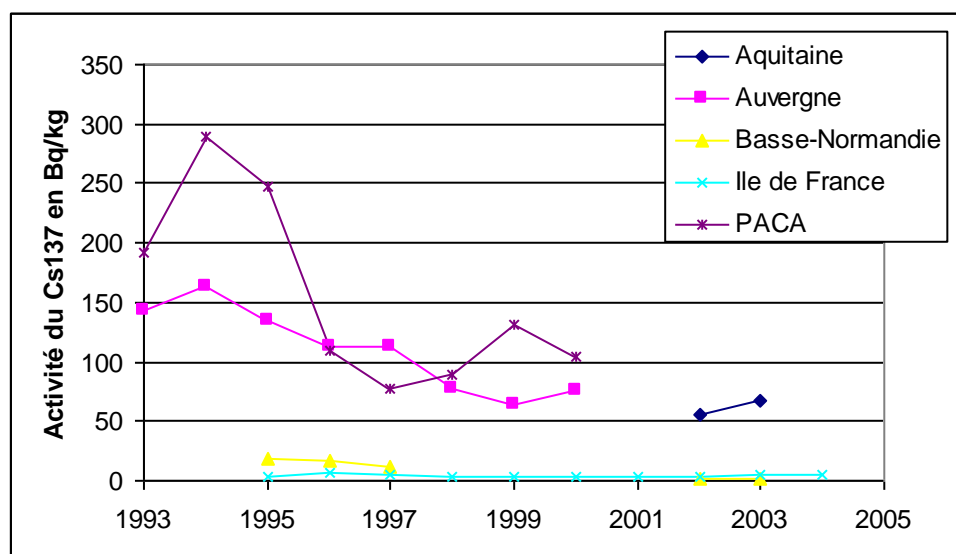


Figure 1 : Données « activité » exploitées

Les données d'activité exploitées pour cette étude sont en quantité relativement faible. Cet exemple permet donc essentiellement d'illustrer la méthode probabiliste permettant l'étude des relations activité – effet.

### B. Mesures des effets : taux de décès par tumeurs

On utilise les taux bruts de décès pour 100 000 habitants causés par des tumeurs. Ces données proviennent du CépiDC (Centre d'épidémiologie sur les causes médicales de décès). Nous les avons déjà utilisés dans notre précédent rapport.

La période d'observation est comprise entre 1979 et 2005 ; seule la période 1993 – 2005 peut être confrontée aux données d'activité.

Le tableau suivant indique les taux bruts de décès pour les 5 régions précitées :

Taux bruts	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
<b>Aquitaine</b>	282.0	274.8	279.5	276.0	277.7	276.6	277.8	276.0	281.2	280.1	286.5	273.8	276.7
<b>Auvergne</b>	297.4	313.1	295.6	295.8	294.8	307.1	311.0	309.6	315.3	318.8	311.6	305.6	312.1
<b>Basse-Normandie</b>	261.9	258.7	266.9	269.4	273.7	276.1	266.9	269.6	272.6	271.5	281.2	276.1	278.4
<b>Ile de France</b>	208.9	204.0	204.6	206.8	203.0	203.3	202.3	198.6	199.3	198.0	199.7	196.0	198.9
<b>PACA</b>	272.5	272.9	267.8	272.9	268.6	266.4	271.7	272.6	268.5	275.7	272.4	269.8	271.2

Tableau 2 : Données « effets » - taux bruts de décès par tumeur

Le graphique suivant représente ces mêmes données :

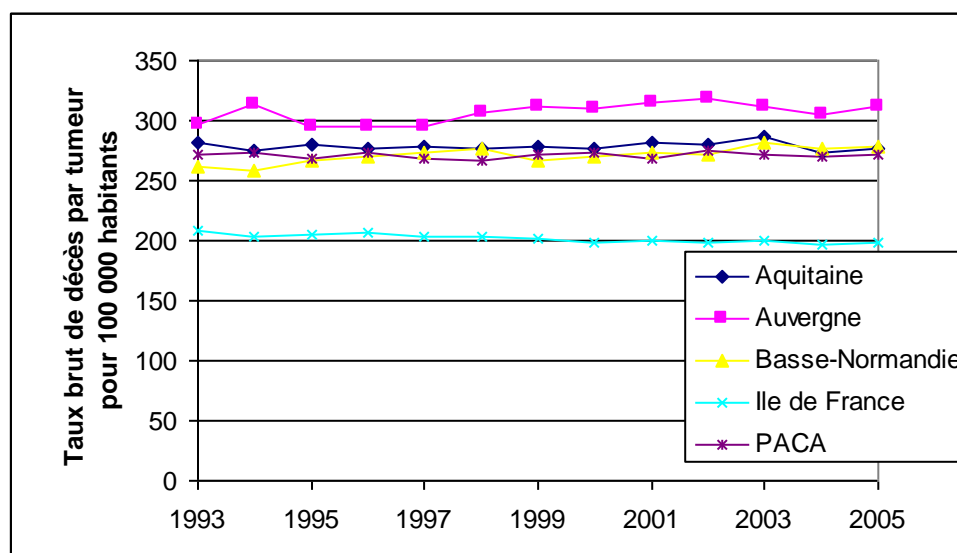


Figure 2 : Données « effets » - taux bruts de décès par tumeur

## II. Loi conjointe Activité - Effet

### A. Table d'occurrences

D'une manière générale, on désignera par « activité » l'activité du Cs137 dans les bryophytes (en Bq/kg) et par « effet » le taux brut de décès pour 100 000 habitants pour cause de tumeur.

La première étape à suivre pour la mise en place d'une méthode probabiliste est la construction d'une table d'occurrences (voir [1]). Les deux indicateurs "activité" et "effet" sont répartis en intervalles, et on compte le nombre d'occurrences conjointes :

Activité \ Effet	175-200	200-225	225-250	250-275	275-300	300-325	Total
<b>0-50</b>	5	5	0	4	1	0	15
<b>50-100</b>	0	0	0	2	2	3	7
<b>100-150</b>	0	0	0	3	4	0	7
<b>≥150</b>	0	0	0	3	0	1	4
<b>Total</b>	5	5	0	12	7	4	33

Tableau 3 : Table d'occurrences

Dans chaque case de ce tableau, on indique le nombre d'observations pour lesquelles l'activité est dans l'intervalle indiqué à gauche et l'effet est dans l'intervalle indiqué en haut. Une observation correspond à un couple de valeur d'activité et d'effet d'une région et d'une année. Les cas où l'on ne dispose que de l'une des deux observations ne sont pas pris en compte lors de la construction de la table d'occurrences.

Ainsi, pour notre exemple, il y a 5 couples (activité, effet) tels que l'activité de Cs137 est comprise entre 0 et 50 Bq/kg et le taux brut de mortalité est compris entre 175 et 200 pour 100 000 habitants.

A ce stade, on peut déjà comparer les situations de faible activité (entre 0 et 50 Bq/kg) et de forte activité (plus de 150 Bq/kg) en calculant moyenne et écart type de ces deux situations :

Activité de Cs137 en Bq/kg	Taux de décès moyen	Ecart type relatif
0-50	225	15%
≥ 150	281.6	6.5%

Tableau 4 : Comparaisons entre l'effet des fortes activités et celui des faibles activités

Le taux de décès moyen des situations de forte activité est nettement supérieur aux situations de faible activité. De plus, l'écart type relatif (écart type divisé par le moyenne) est plus faible, ce qui renforce cette conclusion.

### B. Loi conjointe

On obtient ensuite la loi conjointe Activité – Effet en divisant chaque valeur de la table d'occurrence par le nombre d'observations, soit 33 :

Activité \ Effet	175-200	200-225	225-250	250-275	275-300	300-325	Total
<b>0-50</b>	0.15	0.15	0.00	0.12	0.03	0.00	0.45
<b>50-100</b>	0.00	0.00	0.00	0.06	0.06	0.09	0.21
<b>100-150</b>	0.00	0.00	0.00	0.09	0.12	0.00	0.21
<b>≥150</b>	0.00	0.00	0.00	0.09	0.00	0.03	0.12
<b>Total</b>	0.15	0.15	0.00	0.36	0.21	0.12	1.00

Tableau 5 : Loi conjointe Activité – Effet

La dernière ligne et la dernière colonne correspondent aux lois marginales. Elles indiquent respectivement la probabilité de chaque intervalle de mortalité quelle que soit l'activité et la probabilité de chaque intervalle d'activité de Cs137 quel que soit l'effet. Nous allons voir leur utilisation plus bas.

### C. Test d'indépendance

En probabilités, deux variables aléatoires  $X$  et  $Y$  sont indépendantes lorsque la probabilité d'observer simultanément  $X = a$  et  $Y = b$  est égale au produit de la probabilité d'observer  $X = a$  (quelle que soit la valeur de  $Y$ ) et de la probabilité d'observer  $Y = b$  (quelle que soit la valeur de  $X$ ) :

$$P(X = a \text{ et } Y = b) = P(X = a) \times P(Y = b)$$

Cette définition se décline également lorsque l'on dispose de probabilités sur des intervalles seulement :

$$P(X \in A \text{ et } Y \in B) = P(X \in A) \times P(Y \in B)$$

Appliqué à notre sujet, si l'on considère activité et mortalité comme des variables aléatoires, on peut donc tester leur indépendance. La loi conjointe nous donne les probabilités du type  $P(X \in A \text{ et } Y \in B)$  et les lois marginales correspondent à  $P(X \in A)$  et  $P(Y \in B)$ .

Le tableau suivant donne, pour chaque intervalle d'activité et de taux de décès, la valeur de l'écart  $P(X \in A) \times P(Y \in B) - P(X \in A \text{ et } Y \in B)$  :

Activité \ Effet	175-200	200-225	225-250	250-275	275-300	300-325
<b>0-50</b>	-0.08	-0.08	0.00	0.04	0.07	0.06
<b>50-100</b>	0.03	0.03	0.00	0.02	-0.02	-0.07
<b>100-150</b>	0.03	0.03	0.00	-0.01	-0.08	0.03
<b>≥150</b>	0.02	0.02	0.00	-0.05	0.03	-0.02

Tableau 6 : Test d'indépendance



Les écarts sont faibles, mais ne sont pas négligeable devant les probabilités qui figurent dans la loi conjointe. On peut donc déjà affirmer qu'il n'y a pas indépendance entre l'activité et le taux de décès d'après les données étudiées. Mais cette approche dépend du choix qui a été fait pour les intervalles ; de plus, elle ne renseigne pas de manière quantitative : on se contente de conclure qu'il n'y a pas indépendance.

Pour obtenir des résultats plus précis, nous allons construire les lois conditionnelles : quelle est la probabilité du nombre de cancers, sachant que l'activité est forte, ou sachant que l'activité est faible ? La comparaison de ces lois, au contraire, est très frappante.

#### D. Lois conditionnelles

Nous construisons les différentes lois conditionnelles : probabilité d'être dans chaque intervalle de taux de décès sachant que l'activité est faible ou forte.

Le graphique suivant présente ainsi la probabilité de chaque intervalle de taux de décès pour une activité de Cs137 comprise entre 0 et 50 Bq/kg (en bleu), pour une activité supérieure à 150 Bq/kg et la loi marginale pour les effets (en jaune) :

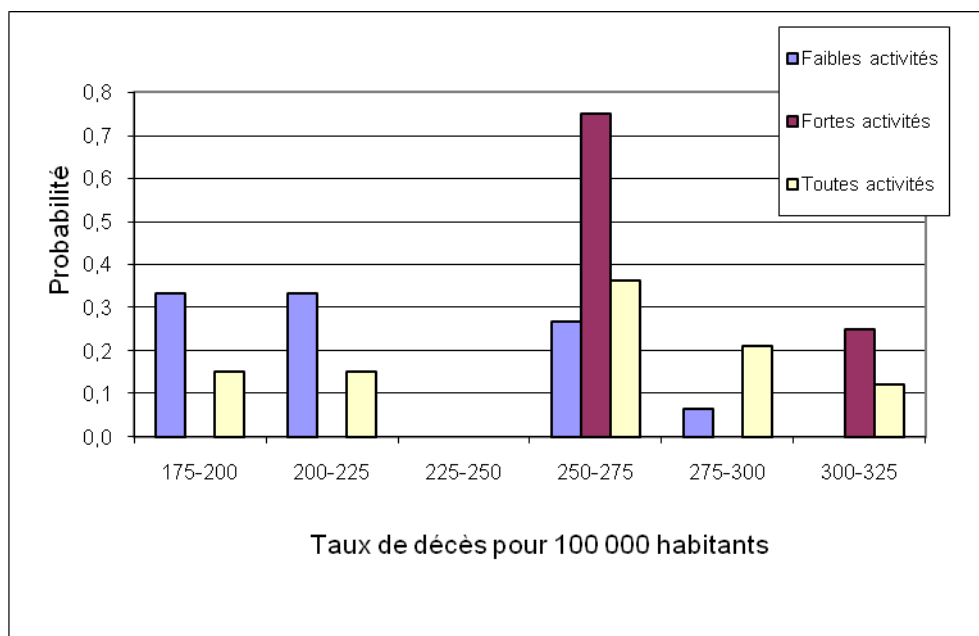


Figure 3 : Lois de probabilité pour différentes situations

Les deux premiers intervalles de taux de décès sont clairement plus probables que la moyenne pour les situations de faible activité. Ils sont en outre de probabilité nulle pour les situations de forte activité.

A l'inverse, deux des trois intervalles des taux de décès les plus forts sont surreprésentés pour les situations de forte activité. L'intervalle 275 – 300 est de probabilité nulle pour les fortes activités, mais sa probabilité pour les faibles activités (environ 0.07) reste inférieure à la probabilité toutes activités confondues (environ 0.21).

### III. Prise en compte de l'aspect temporel

#### A. Méthode

En réalité, l'apparition d'un cancer peut survenir plusieurs années après l'exposition à la radioactivité. Les effets peuvent également s'étaler dans le temps.

Pour tenir compte de cet aspect temporel, on remplace le taux de décès chaque année  $i$  par le taux maximum entre l'année  $i$  et l'année  $i+5$ , et ce pour chaque région.

Ceci est une manière de prendre en compte l'aspect temporel ; d'autres solutions sont possibles. La durée d'observation est un premier paramètre. On pourrait également considérer non pas le maximum, mais la somme glissante sur 5 ans. Enfin, on pourrait aussi considérer l'activité cumulée sur 5 ans.

#### B. Table d'occurrence et loi conjointe

On établit de la même manière que précédemment la table des occurrences puis la loi conjointe à partir des données « corrigées » pour prendre en compte 5 années pour les effets et non une seule :

Activité \ Effet	175-200	200-225	225-250	250-275	275-300	300-325	Total
0-50	5	5	0	0	5	0	15
50-100	0	0	0	1	3	3	7
100-150	0	0	0	1	2	4	7
≥150	0	0	0	3	0	1	4
<b>Total</b>	5	5	0	5	10	8	33

Tableau 7 : Table d'occurrence - aspect temporel sur 5 ans

Activité \ Effet	175-200	200-225	225-250	250-275	275-300	300-325	Total
0-50	0.15	0.15	0.00	0.00	0.15	0.00	0.45
50-100	0.00	0.00	0.00	0.03	0.09	0.09	0.21
100-150	0.00	0.00	0.00	0.03	0.06	0.12	0.21
≥150	0.00	0.00	0.00	0.09	0.00	0.03	0.12
<b>Total</b>	0.15	0.15	0.00	0.15	0.30	0.24	1.00

Tableau 8 : Loi conjointe Activité – Effet - aspect temporel sur 5 ans

La prise en compte de l'aspect temporel n'a pas d'impact pour les trois premiers intervalles de taux de décès, quelle que soit l'activité. On observe en revanche un glissement vers le haut pour les intervalles au delà de 250 décès pour 100 000 habitants.

La comparaison des situations de faible activité et de forte activité montre là aussi un taux de décès moyen nettement supérieur en présence d'une forte activité :

Activité de Cs137 en Bq/kg	Taux de décès moyen	Ecart type relatif
0-50	227.4	15.8%
≥ 150	282.9	6.1%

Tableau 9 : comparaison activité faible - activité forte

La représentation graphique des lois marginales pour les effets avec et sans prise en compte de l'aspect temporel montre l'augmentation de la probabilité des deux derniers intervalles de taux de décès :

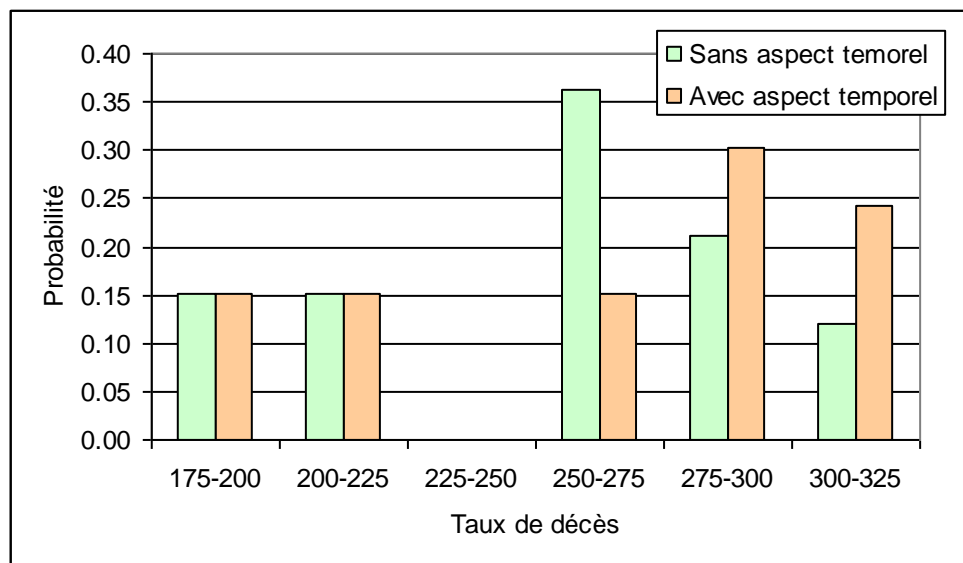


Figure 4 : Lois marginales avec et sans prise en compte de l'aspect temporel

### C. Lois conditionnelles

Le graphique suivant présente ainsi la probabilité de chaque intervalle de taux de décès pour les situations de faible activité (entre 0 et 50 Bq/kg – en bleu), de forte activité (plus de 150Bq/kg – en rouge) et la loi marginale pour les effets (en jaune) :

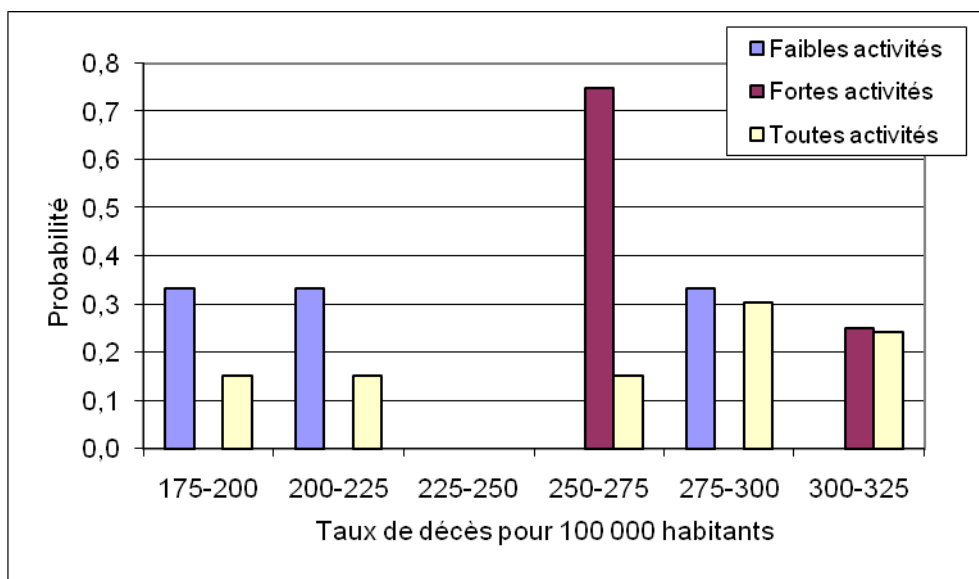


Figure 5 : Lois de probabilité pour différentes situations

Ici aussi, les deux premiers intervalles de taux de décès sont plus probables que la moyenne pour les situations de faible activité et sont de probabilité nulle pour les situations de forte activité.

En ce qui concerne les taux de décès supérieurs à 250 pour 100 000 habitants, il est plus difficile de conclure. L'intervalle de taux de décès 250-275 est 5 fois plus probable en situation de forte activité, mais l'intervalle supérieur est plus probable en situation de faible activité.

## IV. Fonctions de répartition

Nous traçons maintenant la fonction de répartition (plus précisément son complémentaire, défini par  $P\{X > x\} = 1 - P\{X \leq x\}$ ) des deux lois précédentes (avec et sans l'aspect temporel). La fonction de répartition permet une visualisation plus facile.

La voici pour la loi de probabilité sans tenir compte de l'aspect temporel :

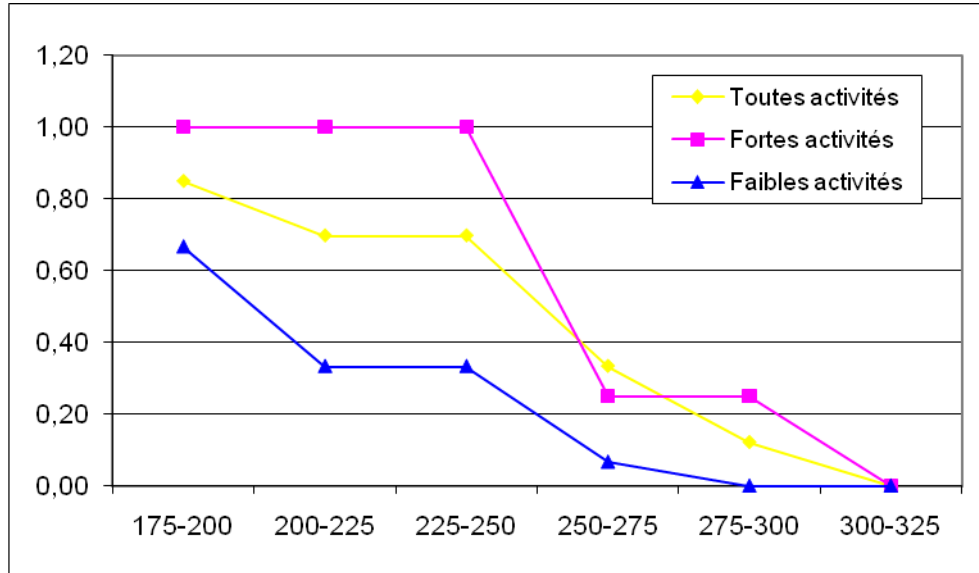


Figure 6 : Complémentaire de la fonction de répartition de la loi de probabilité sans tenir compte de l'aspect temporel

La voici pour la loi de probabilité tenant compte de l'aspect temporel :

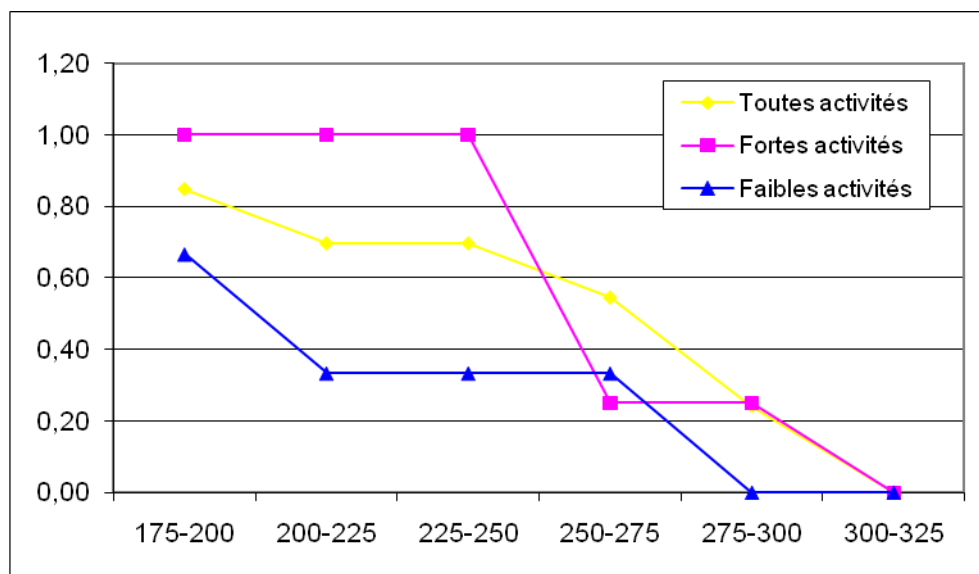


Figure 7 : Complémentaire de la fonction de répartition de la loi de probabilité, en tenant compte de l'aspect temporel

Dans les deux cas, les probabilités sont plus faibles pour les faibles activités. Cependant, lorsque l'on tient compte de l'aspect temporel, ce risque est plus fort pour le taux de décès supérieur ou égal à 250.

Pour les fortes activités, le risque est maximal jusqu'au taux de décès supérieur ou égal à 250. Il est toujours supérieur à celui des faibles activités lorsque l'on ne tient pas compte de l'aspect temporel.

### **Remarque**

On observe à chaque fois un artefact dans le cas des fortes activités, pour l'intervalle 250-275. Il provient du fait que la probabilité de cet intervalle de décès pour 100 000 habitants, en situation de forte activité, est cinq fois plus grande lorsque l'on tient compte de l'aspect temporel et deux fois plus grande lorsque cet aspect temporel est mis de côté. De plus, lorsque l'on s'intéresse au maximum des cancers sur 5 ans, on s'aperçoit que la probabilité est nulle pour les faibles activités (la fonction de répartition reste donc constante) dans l'intervalle de décès 250-275. Ainsi, la fonction de répartition complémentaire des faibles activités dépasse ponctuellement celle des fortes activités.

Toutes activités confondues, le risque décroît quasi-linéairement. Rappelons que ces conclusions ne se rapportent qu'aux valeurs de l'étude et qu'ils ne serait pas légitime de les généraliser.

### **Référence**

- [1] Bernard Beauzamy : Méthodes probabilistes pour l'étude des phénomènes réels. SCM SA, 2004.