

ESTIMATION D'UN RISQUE DE CANCER RADIO-INDUIT EN SITUATION DE CO-EXPOSITIONS PROFESSIONNELLES: CAS DES MINEURS D'URANIUM FRANÇAIS

Sophie Ancelet¹, Marion Belloni¹, Olivier Laurent¹, Chantal Guihenneuc²

¹ IRSN, PSE-SANTE/SESANE/LEPID, Fontenay-Aux-Roses, France; [Email: sophie.ancelet@irsn.fr](mailto:sophie.ancelet@irsn.fr)

² Université de Paris, UR 7537 BioSTM "Biostatistique, Traitement et Modélisation des données biologiques", Paris, France

Contexte

- Les **cancers**: pathologies **chroniques** et **multifactorielles** pour lesquelles le concept d'**exposome** est essentiel [Zabaleta, 2012]
 - Résultent de l'influence combinée de multiples facteurs de risque (i.e., génétiques, environnementaux, comportementaux) pouvant **agir simultanément et interagir**
 - Mieux appréhender comment ces situations de multi-expositions peuvent conduire au développement de cancers
 - Mieux comprendre l'étiologie des cancers afin d'aboutir à de meilleures stratégies de prévention

- En épidémiologie, importance de raisonner dans un **cadre d'exposition multifactorielle** pour estimer/prédire des risques de cancer au niveau individuel ou populationnel
 - Historiquement, focus sur l'effet d'**un seul facteur de risque**, en contrôlant sur facteurs de confusion potentiels
 - Très peu d'études ont visé à estimer:
 - Interaction entre exposition à un facteur de risque environnemental et autres facteurs de risque (ex: tabagisme et amiante [Klebe et al., 2019], tabagisme et radon [Leuraud et al., 2011])
 - Effets conjoints de l'exposition à plusieurs facteurs de risque (ex: particules ambiantes et ozone [Lin et al., 2019])

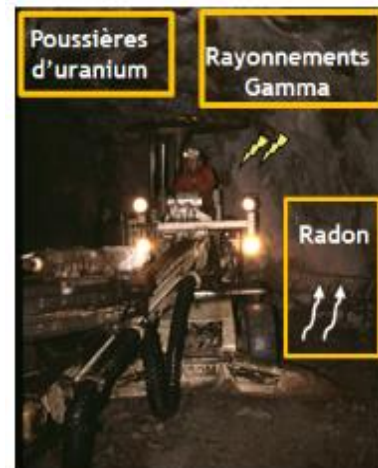


Effets sanitaires **d'expositions en mélange** très mal caractérisés

Qu'en est-il en épidémiologie des rayonnements ionisants (RIs)?

- Un **objectif clé** depuis des décennies : Estimation des risques de cancers radio-induits (+ incertitude)
- Nombreuses connaissances déjà accumulées concernant les **effets cancers** des expositions aux RIs
- Des **interrogations persistent** cependant:
 - Effets des doses faibles et modérées
 - Effets d'une exposition simultanée (co-expositions) à de multiples facteurs de risque radiologiques [NRC, 2006]
- Normes de radioprotection principalement basées sur un **cadre d'exposition mono-factoriel**
 - Risque faible, dans un cadre simplifié d'exposition radiologique supposée unique
=> Risque faible en situation de co-expositions à de multiples sources radiologiques ?
- Quelques **enjeux actuels** en épidémiologie des RIs:
 - Mieux estimer les effets sanitaires associés à des expositions conjointes prolongées à faibles doses de RIs et à d'autres facteurs de risque
 - Mieux appréhender les potentielles interactions entre expositions radiologiques et autres facteurs de risque
 - Mettre en perspective/mieux positionner les effets sanitaires de l'exposition aux RIs par rapport aux effets associés à d'autres facteurs de risque.

Cas d'étude



Des travailleurs multi-exposés

Estimation de l'association entre **des co-expositions radiologiques** à faibles doses et le risque de décès par cancer du poumon dans la **cohorte française des mineurs d'uranium**.



Pourquoi cette population d'étude?



Risque non estimé jusqu'à présent - Focus sur l'exposition au radon !

Population d'étude : la sous-cohorte post-55 des mineurs d'uranium français

- Mineurs d'uranium employés **au moins un an** en tant que mineurs d'uranium dans le groupe CEA-Cogema et embauchés après 1955 [Rage et al. 2015]

Mineurs, n	3 377
Âge à l'entrée dans l'étude, moyenne [min-max]	28.3 [16.9-57.7]
Durée de travail en années, moyenne [min-max]	16.7 [1.0-40.9]
Durée de suivi en années, moyenne [min-max]	32.8 [0.1-51.0]
Statut vital, n (%)	
Vivant	2 412 (71.4)
Âgé \geq 85 ans	74 (2.2)
Décès par cancer du poumon	94 (2.8)
Décès d'une autre cause	777 (23.0)
Perdu de vue	20 (0.6)

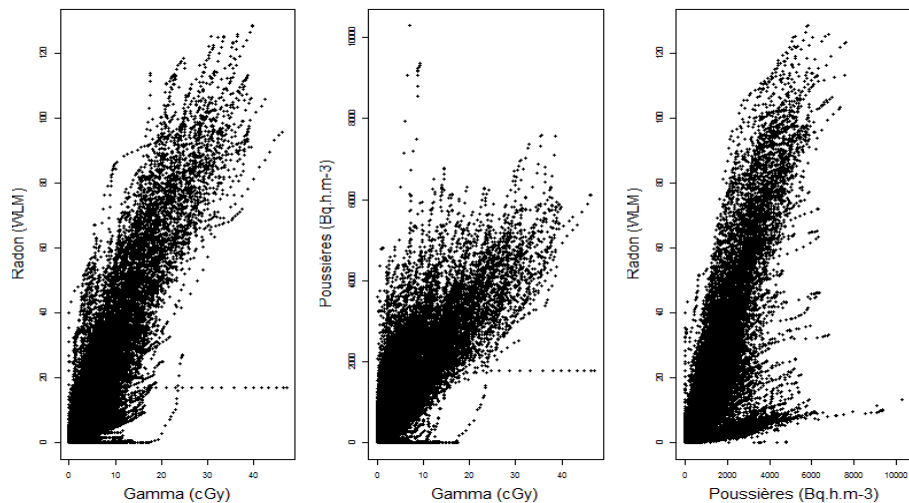


Une difficulté majeure en situation de co-expositions : la multicollinéarité

- En situation de co-expositions a de multiples stressseurs (ex: environnementaux), les covariables d'exposition sont souvent fortement corrélées entre elles => multicollinéarité
 - Exemple : Cas de travailleurs simultanément exposés à de multiples stressseurs chimiques, radiologiques et/ou physiques (e.g. chaleur) pendant leur activité professionnelle
- Appliquer un **modèle de régression multiple** dans lequel au moins 2 prédicteurs sont fortement corrélés
 - Estimations instables des coefficients de risque (grande variance!) donc non interprétables
 - Conclusions trompeuses concernant l'effet de chaque prédicteur sur la variable réponse d'intérêt



Exemple dans la cohorte post-55 des mineurs d'uranium français



Modèle*	$\hat{\beta}_{\text{gamma}}/100\text{mSv}$	$\hat{\beta}_{\text{radon}}/100\text{WLM}$	$\hat{\beta}_{\text{poussière}}/100\text{Bq.h.m-3}$
Univarié	0.78 [0.28;1.67]	2.7 [1.1;5.2]	3.34e-2 [1.07e-2;7.00e-2]
Trivarié	-0.00 [-0.39;1.17]	2.7 [-0.2;5.8]	-0.15e-2 [-1.66e-2;3.81e-2]

* Modèle de survie en excès de risque instantané

Comment pallier un problème de multicolinéarité?

ENVIRONMENT-WIDE ASSOCIATION APPROACH (EWAS)

- Estimation par régressions séparées de l'association entre l'événement d'intérêt et chaque exposition étudiée
- Approche exploratoire qui ne permet pas de comprendre la relation entre une combinaison d'expositions et la pathologie étudiée

SELECTION DE VARIABLES EN REGRESSION

- Permet d'identifier un petit sous-ensemble de variables d'exposition ayant un impact majeur sur le risque d'intérêt.
- Surtout utile quand le nombre de facteurs de risque est très important.

APPROCHES PAR REDUCTION DE DIMENSION

- Régression sur composantes principales, régression sPLS
- Interprétabilité
- Construction des axes et estimation du risque dans 2 étapes disjointes. Quid de la prise en compte des incertitudes sur la construction des axes ?

ALGORITHMES DE MACHINE LEARNING

- k-means, forêts aléatoires
- Approches efficaces et pertinentes quand le nombre de variables est très important mais ne permet pas d'estimer les risques d'intérêt

APPROCHES PAR CLASSIFICATION DE DONNEES D'EXPOSITIONS CORRELEES

- Analyses en classes latentes: variable-réponse non prise en compte dans la classification
- **Modèles de mélange par régression sur profils d'exposition** [Molitor et al. 2010] : permet d'estimer l'effet sanitaire d'une combinaison d'expositions



Pas d'utilisation en épidémiologie des RIs

Qu'est-ce qu'un modèle de mélange par régression sur profils d'exposition ?

- Ce modèle probabiliste permet:
 - de définir des **groupes** d'individus ayant des **profils d'exposition proches**
 - d'**estimer**, pour chaque groupe, le **risque** associé à la pathologie d'intérêt en supposant que tous les individus appartenant au même groupe ont le même risque
- Ce modèle peut être vu comme la combinaison de 3 sous-modèles probabilistes:
 - Un **sous-modèle de maladie** : il décrit mathématiquement le lien entre l'âge au décès par cancer du poumon (événement d'intérêt) et le risque de pathologie d'intérêt qui lui-même dépend du risque de base et du groupe d'appartenance
 - Le **sous-modèle d'exposition** : il décrit les caractéristiques d'exposition de chaque groupe pour chaque facteur de risque (prédicteur)
 - Le **sous-modèle d'attribution** : il définit la probabilité d'appartenance d'un mineur à chaque groupe

Le sous-modèle de maladie

Variables réponse

- Age (en jours) au décès par cancer du poumon d'un mineur i - noté T_i (Événement d'intérêt)
 - Tronqué à gauche à l'âge à l'entrée dans la cohorte
 - Censuré à droite (décès autre cause, 31 décembre 2007, perdu de vue ou 85 ans)
- Age (en jours) à la censure - noté C_i
- Ce qui est observé : (Y_i, δ_i) avec $Y_i = \min(T_i, C_i)$ et l'indicateur binaire $\delta_i = 1$ si décès par cancer du poumon ($\delta_i = 0$ sinon).

Modèle de survie en excès de risque instantané (EHR)

- Risque instantané de décès par cancer du poumon du mineur i à l'âge t

$$h_i(t) = h_0(t) \cdot (1 + \beta_{C_i})$$

- $h_0(t)$: Risque instantané de base supposé constant par morceaux sur 4 intervalles d'âge (en années):
<40 ans, [40;55], [55;70], >70 => $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)$
- C_i : Groupe d'appartenance du mineur i (inconnu)
- β_c : Excès de risque instantané de décès par cancer du poumon du groupe de mineurs c



Deux individus appartenant au même groupe c ont le même risque β_c de décès par cancer du poumon

Le sous-modèle d'exposition

■ Le profil d'exposition $Z_i=(Z_{i,1},\dots,Z_{i,p})$ d'un mineur i où $P=7$ = nombre de facteurs de risque (prédicteurs)

■ Prédicteurs à valeurs continues positives:

- Exposition cumulée au radon (lag 5 ans) du mineur i
- Exposition cumulée aux rayonnements gamma (lag 5 ans) du mineur i
- Exposition cumulée aux poussières d'uranium (lag 5 ans) du mineur i
- Age (en années) à la première exposition du mineur i

▪ Sous-modèle => $Z_i^{Cont,q} | C_i = c \sim \text{LogNormal}(\mu_c^q, \sigma_c^q)$

■ Prédicteurs catégoriels:

- Type de poste le plus souvent occupé par le mineur i (proxy conditions d'exposition)
 - 5 catégories: foreur avant mécanisation, foreur après mécanisation, autre mineur de fond avant mécanisation, autre mineur de fond après mécanisation; mineur de surface
- Mine dans laquelle le mineur i a le plus travaillé
 - 2 catégories: Hérault (roche sédimentaire) ou autre (roche granitique)
- Durée d'exposition en années => 4 catégories : [1;5], [6;12], [13;18], 19 ans et +

▪ Sous-modèle => $Z_i^{Cat,q} | C_i = c \sim \text{Multinomial}(p_c^q)$

Le sous-modèle d'attribution

- Le nombre de groupes de mineurs est **supposé inconnu et donc estimé**. Seul le nombre maximal C_{\max} de groupes est fixé.

$$C_i \stackrel{iid}{\sim} \text{Multinomial}(\psi_1, \psi_2, \dots, \psi_{C_{\max}})$$

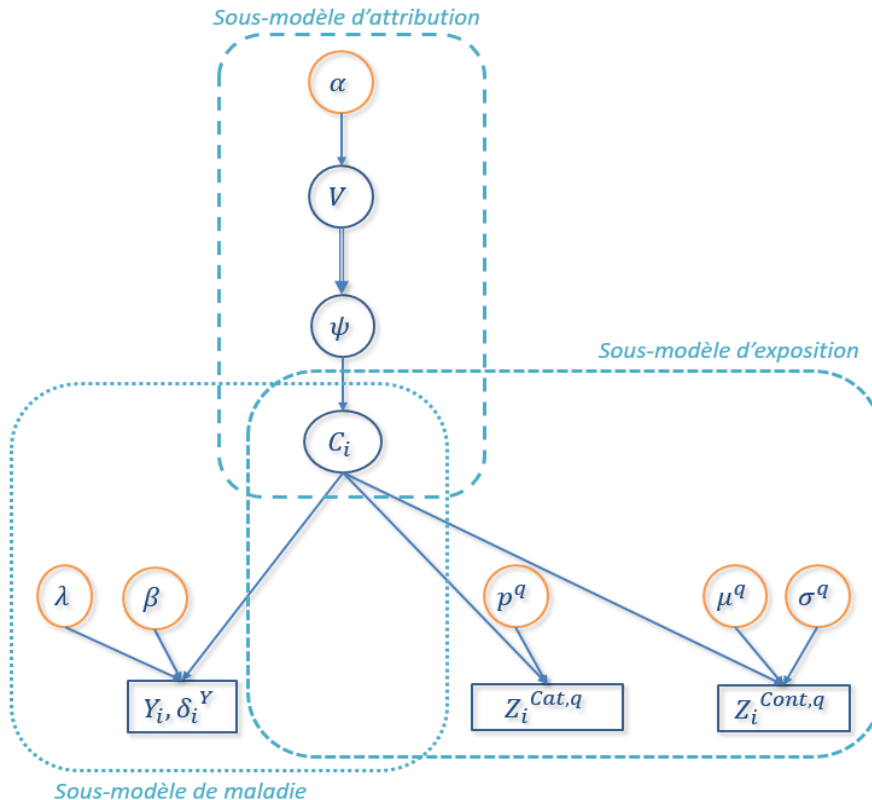
- Le vecteur des probabilités $\psi = (\psi_1, \dots, \psi_{C_{\max}})$ d'attribution à chaque groupe suit un **processus de Dirichlet tronqué** comme suit:

$$\begin{cases} V_c | \alpha \sim \text{Beta}(1, \alpha), c = 1, \dots, C_{\max} - 1 \\ \psi_c = V_c (1 - \sum_{k=1}^{c-1} \psi_k), c = 1, \dots, C_{\max} - 1 \\ \psi_{C_{\max}} = 1 - \sum_{k=1}^{C_{\max}} \psi_k \end{cases}$$



Plus le paramètre inconnu α est grand, plus le nombre de groupes non vides est grand

Représentation graphique du modèle et inférence statistique bayésienne



- Apprentissage statistique bayésien du modèle à l'aide d'un algorithme Monte-Carlo par Chaînes de Markov
- Choix de lois *a priori* sur les paramètres inconnus du modèle
 - Peu de signal dans les données => lois *a priori* informatives pour certains paramètres

Résultats : nombre de groupes non-vides estimé

- Nombre de groupes non-vides estimé en fonction de la valeur initiale du paramètre α

Valeur initiale de α	0.5	1.5	2.5	3.5	4.5	5.5	6.5	7.5	8.5	9.5
Nombre de groupes	5	6	8	8	7	7	7	8	7	8

- Loi *a posteriori* multimodale
 - Pour chaque partition identifiée, pas de problème de convergence identifié de l'algorithme MCMC
 - Problème de convergence local de l'algorithme MCMC implémenté
- Une approche possible : Estimation à nombre de groupes non-vides fixé à 5 puis 6 puis 7 puis 8 puis sélection du « meilleur » modèle

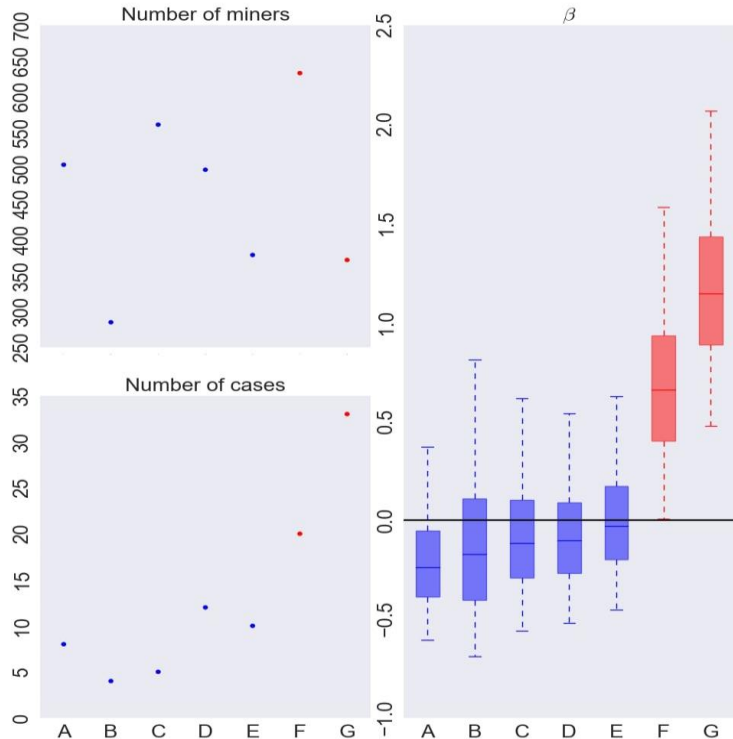
Nombre de groupes non-vides	DIC	WAIC
5	146 345	110 872
6	136 714	108 773
7	118 602	107 004
8	104 566	105 704

Meilleur modèle = Celui qui minimise le DIC (Deviance Information Criterion) ou le WAIC (Widely-Applicable Information Criterion)



Présentation des résultats du modèle à 8 groupes non-vides

Résultats: Caractérisation des groupes – Partition en 8 groupes



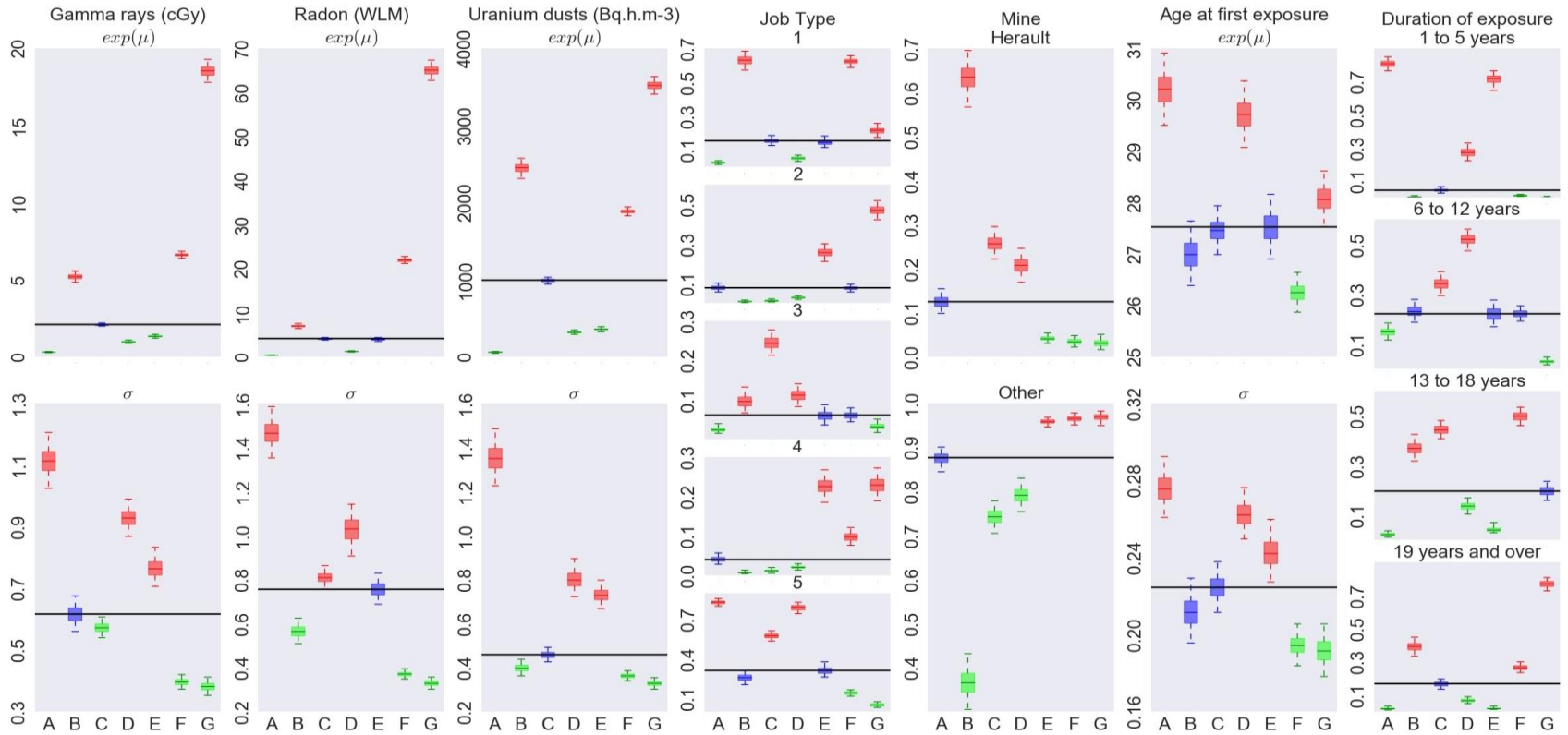
Médianes *a posteriori* et intervalles de crédibilité à 95%

8 groupes identifiés dont:

- 2 groupes (F et G) à risques significativement plus élevés
- 5 groupes à risques non significativement différents de 0
- 1 groupe de référence (non représenté): les mineurs non-exposés

	Gamma Radon Poussière	Type de poste	Hérault	Âge 1 ^{ère} exposition	Durée
Groupe G ($\beta \approx 1.2$)	+++	Avant mécanisation	Non	Âgé	≥ 19 ans
Groupe F ($\beta \approx 0.6$)	+	Foreur après mécanisation	Non	Jeune	≥ 13 ans

Des résultats très riches !



Conclusion

- Première application en épidémiologie des rayonnements ionisants des modèles de mélange par régression sur profils d'exposition pour estimer l'impact d'une combinaison d'expositions radiologiques sur un risque de cancer
 - Extension de ces modèles au cadre des modèles de survie en EHR classiquement utilisés dans ce domaine
 - Résultats riches et nouveaux concernant l'analyse de l'association potentielle entre le risque de décès par cancer du poumon et des profils d'exposition spécifiques des mineurs d'uranium français
 - Possibilité d'utiliser ces modèles pour faire de la prédiction individualisée de risque en situation de co-expositions



Une approche prometteuse pour la recherche sur l'exposome en épidémiologie

Limites et perspectives

- Pas de prise en compte du **statut tabagique** :
 - Données seulement disponibles pour 4,2% des mineurs dans la cohorte post-55 + aucun prédicteur du statut tabagique disponible => Imputation non envisageable!
 - Néanmoins, selon de précédentes études :
 - le tabagisme n'apparaît pas comme un facteur de confusion dans les études de cohortes professionnelles de mineurs d'uranium [Richardson et al., 2014; Keil et al., 2015]
 - une association positive significative entre exposition au radon et mortalité par cancer du poumon a été observée chez les mineurs fumeurs et non-fumeurs [Kreuzer et al., 2017; Tomasek, 2002]
 - l'effet de l'exposition au radon sur le risque de décès par cancer du poumon persiste après ajustement sur le statut tabagique - études cas-témoins nichées [Leuraud et al., 2007; Leuraud et al., 2011]
- Loi *a posteriori* possiblement **multimodale** => **Manque de signal** dans les données?
 - Amélioration de l'algorithme MCMC pour éviter la convergence vers des modes locaux
- Pas de prise en compte de la **dimension temporelle** et des **erreurs de mesure** sur les expositions
- Pas d'interprétation aisée des **effets de synergie** ou **d'antagonisme** entre expositions conjointes
- Quelles performances des modèles de mélange par régression sur profils d'exposition dans des cas d'étude impliquant un **nombre plus élevé de facteurs de risque** (i.e., >7) ?

Si vous voulez en savoir plus :

Belloni M, Laurent O, Guihenneuc C, Ancelet S. (2020) *Bayesian Profile Regression to Deal With Multiple Highly Correlated Exposures and a Censored Survival Outcome. First Application in Ionizing Radiation Epidemiology*. *Frontiers Public Health*.

Merci de votre attention!