



HAL
open science

Modélisation hiérarchique et apprentissage statistique bayésien pour la prise en compte d'erreurs de mesure d'exposition complexes en épidémiologie des rayonnements ionisants.

Sophie Ancelet, Sabine Hoffmann, Marion Belloni, Chantal Guihenneuc

► To cite this version:

Sophie Ancelet, Sabine Hoffmann, Marion Belloni, Chantal Guihenneuc. Modélisation hiérarchique et apprentissage statistique bayésien pour la prise en compte d'erreurs de mesure d'exposition complexes en épidémiologie des rayonnements ionisants.. WORKSHOP SCIDONI, WORKSHOP SCIDONI, Dec 2020, LYON, France. hal-03106523

HAL Id: hal-03106523

<https://hal.science/hal-03106523>

Submitted on 11 Jan 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

Modélisation hiérarchique et apprentissage statistique bayésien pour la prise en compte d'erreurs de mesure d'exposition en épidémiologie des rayonnements ionisants

Sophie Ancelet¹, Sabine Hoffmann^{1,2}, Marion Belloni¹, Chantal Guihenneuc³

¹Institut de Radioprotection et de Sûreté Nucléaire (IRSN), PSE-SANTE/SESANE/LEPID, Fontenay-Aux-Roses (sophie.ancelet@irsn.fr)

²Institute for Medical Information Processing, Biometry, and Epidemiology, University of Munich, Germany

³Université Paris Descartes, Faculté de Pharmacie, Paris

Atelier SCIDONI, 12 Novembre 2020

Context : Exposure measurement error in epidemiological studies

- Exposure measurement error is
 - ▶ ubiquitous
 - ▶ one of the most important source of uncertaintyin epidemiological studies
- If not or only poorly accounted for, exposure measurement error may cause :
 - ▶ bias in risk estimates
 - ▶ a distortion of the exposure-risk relationship
 - ▶ a loss in power
- In retrospective cohort studies :
 - ▶ complex patterns of exposure measurement error
 - ▶ attenuation of the exposure-risk relationship for high exposure values [HertzPicciotto (1993), Stayner (2003)]
 - Suggested to be explained by the effect of measurement error

Methods to account for measurement error

- Many methods for the correction of measurement error, like regression calibration and simulation extrapolation (SIMEX)
 - ▶ Lack of flexibility to account for complex measurement error
 - Measurement error in time-varying exposure variables
 - Combination of different types of measurement error
 - Heteroscedastic measurement error variances
 - Uncertainty in measurement error variance parameters
 - ▶ Potential lack of consistency in risk estimates in proportional hazards models [Bartlett and Keogh, 2016]
 - ▶ Disjoint steps to estimate true exposure and risk parameters
- Bayesian hierarchical models (also called multi-levels or random-effects models)
 - ▶ Provides a natural way of combining exposure and parameter uncertainty
 - ▶ Flexible approach to describe and account for complex measurement error
 - ▶ Allows for the joint estimation of true exposure and risk parameters
 - ▶ Allows to integrate external information through the specification of informative prior distributions to obtain more precise estimates

Aims of the work

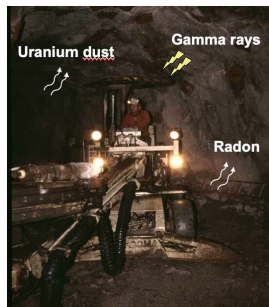
- 1 Promote the use of **Bayesian hierarchical approach** to account for **complex** exposure measurement error in risk estimates in radiation epidemiology
 - ▶ Improved considerations of **uncertainties**
 - ▶ (Potentially) improved risk estimates in a **weakly informative context**
- 2 Study the **impact** of different error structures on statistical inference
 - ▶ Simulation studies
 - ▶ Specific context of survival models with time-dependent covariates
- 3 Refine, if possible, the estimation of radiation-induced health risks at low doses
 - ▶ Réponses vis-à-vis de la stratégie scientifique de l'IRSN en radioprotection
 - Question 1 : Quels sont les effets d'une exposition à de faibles doses ?
 - Question 3 : Comment améliorer les concepts, méthodes et outils destinés à évaluer le risque suite aux expositions des travailleurs, de la population et des patients ?
- 4 Improve the **knowledge** about the shape of dose-response relationships
 - ▶ Réponses vis-à-vis des priorités scientifiques identifiées au niveau européen
 - Forme de la relation dose-réponse pour les cancers

Motivating case study

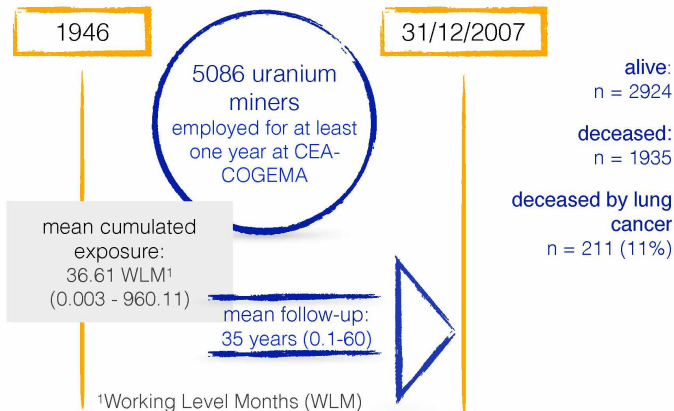
Lung cancer and chronic low-dose exposure to radon

- Radon is a radioactive gas which presents the primary source of background radiation
- Radon is the second cause of lung cancer (after tobacco) [Samet and Eradze, 2000]

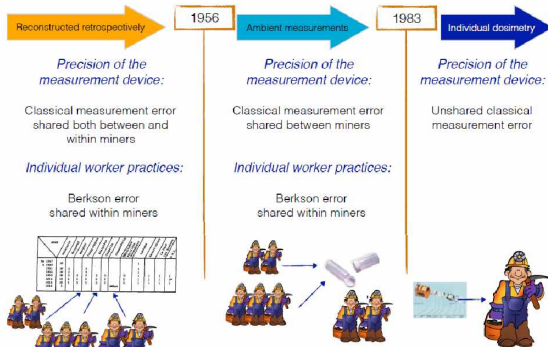
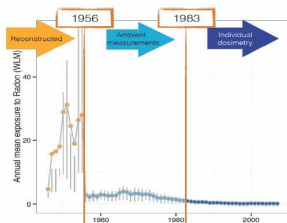
Thanks to an individual dosimetry collected over the entire career, the French cohort of uranium miners is a reference population for the study of the long-term health effects of chronic low-dose exposure to radon (Inhalation exposure)



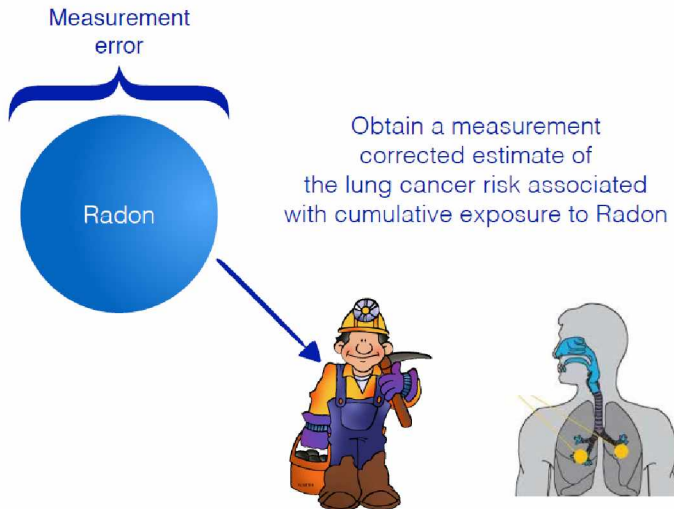
The French cohort of uranium miners



Radon measurement error characteristics in the cohort



Challenge



Accounting for measurement error through a hierarchical structure

In a hierarchical framework, one can account for exposure measurement error combining the following submodels :

- **Disease submodel**
- **Measurement submodel**
- **Exposure submodel**

⇒ Joint modelling of all sub-models using conditional independence assumptions

[Richardson and Gilks, 1993]

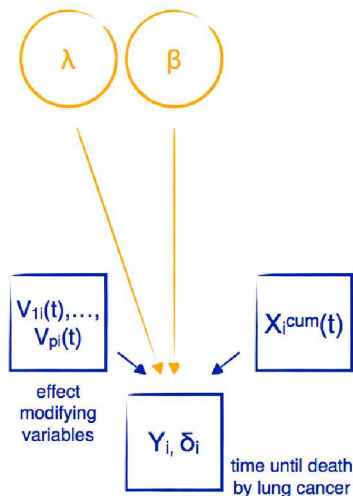
⇒ Joint estimation of all unknown quantities

Modelling failure times

- $Y_i = \min(T_i, C_i)$ is right-censored age at death by lung cancer
- $h_i(t) = h_0(t) (1 + \beta X_i^{cum}(t))$
- Suppose a piecewise constant model for $h_0(t)$ with $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)$

Effect modifying variables

- Test the linearity of the exposure-risk relationship via piecewise linear models



Measurement submodel \mathcal{M}_2

- $Z_i^B(t)$: estimated/observed exposure of individual i for year t of the calendar period B
 $X_i^B(t)$: "true" exposure of individual i for year t of the calendar period B
 U_i^B : error component for individual i and the period B (individual worker practices)

First calendar period 1946-1955 :

- **Berkson error shared within individuals :**

- ▶ $X_i^1(t) = Z_i^1(t) \cdot U_i^1$
- ▶ $U_i^1 \sim \mathcal{LN}\left(-\frac{\sigma_{U_1}^2}{2}, \sigma_{U_1}^2\right)$ [Darby, 1998]
- ▶ $U_i^1 \perp\!\!\!\perp Z_i^1(t)$ and $U_i^1 \perp\!\!\!\perp Y_i$
- ▶ $\mathbb{E}(X_i^1(t)|Z_i^1(t)) = Z_i^1(t)$
- ▶ $\sigma_{U_1}^2 = 0.86$ [Allodji et al., 2012]

Second calendar period 1956-1982 :

- **Berkson error shared within individuals :**

- ▶ $X_i^2(t) = Z_i^2(t) \cdot U_i^2$
- ▶ $U_i^2 \sim \mathcal{LN}\left(-\frac{\sigma_{U_2}^2}{2}, \sigma_{U_2}^2\right)$
- ▶ $U_i^2 \perp\!\!\!\perp Z_i^2(t)$ and $U_i^2 \perp\!\!\!\perp Y_i$
- ▶ $\mathbb{E}(X_i^2(t)|Z_i^2(t)) = Z_i^2(t)$
- ▶ $\sigma_{U_2}^2 = 0.15$ [Allodji et al., 2012]

Third calendar period 1983-2007 : Classical measurement error **neglected** (magnitude of error much smaller than in earlier periods)

Measurement submodel \mathcal{M}_3 + Exposure submodel

=> Refined measurement submodel only for the first calendar period 1946-1955

$X_{ij}^1(t)$: "true" exposure of individual i for year t , given that he worked in mine j

Z_j^1 : estimated mean exposure for one month of work in mine j (1946-1955)

$T_{ij}(t)$: Time in months spent by individual i in mine j during year t

ξ_j : "true" mean exposure for one month of work in mine j (1946-1955)

U_i^1 : error component for individual i over the period 1946-1955

U_j^1 : error component for all miners working in mine j over the period 1946-1955

- **Berkson error shared within individuals :**

- ▶ $X_{ij}^1(t) = \xi_j \cdot T_{ij}(t) \cdot U_i^1$
- ▶ $U_i^1 \sim \mathcal{LN}\left(-\frac{\sigma_{U_1}^2}{2}, \sigma_{U_1}^2\right)$
- ▶ $\mathbb{E}(X_{ij}^1(t) | \xi_j, T_{ij}(t)) = \xi_j \cdot T_{ij}(t)$

- **Classical measurement error shared both within and between miners**

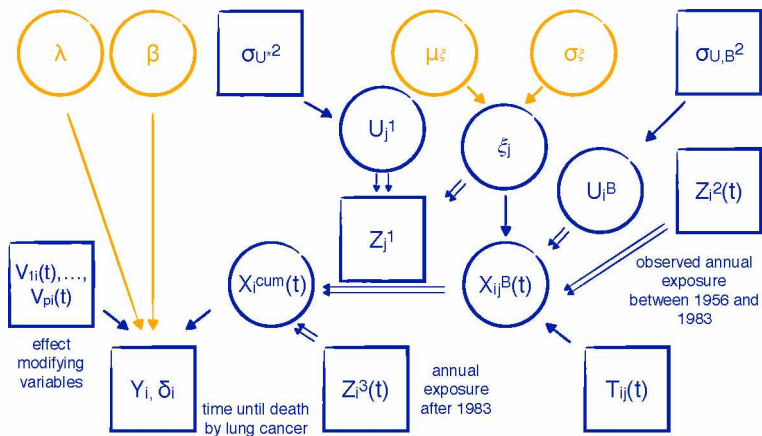
- ▶ $Z_j^1 = \xi_j \cdot U_j^1$
- ▶ $U_j^1 \sim \mathcal{LN}\left(-\frac{\sigma_{U_*}^2}{2}, \sigma_{U_*}^2\right)$
- ▶ $U_j^1 \perp \xi_j$

- **The exposure model**

- ▶ $\xi_j \sim \mathcal{LN}(\mu_\xi, \sigma_\xi)$

$\sigma_{U_1}^2$ and $\sigma_{U_*}^2$ fixed according to [Allodji et al., 2012] + sensitivity analysis

Directed Acyclic Graph



Bayesian inference

- Informative prior distributions for the λ parameters defining the baseline hazard function $h_0(t)$
 - ▶ External information : lung cancer mortality in the general French male population between 1968 and 2005
- Complex joint posterior distribution $\theta = (\beta, \lambda, \mu_\xi, \sigma_\xi, \xi, \mathbf{U})$
=> Adaptive Metropolis-Within-Gibbs algorithm developed and tested in Python
- About 6,000 (Model \mathcal{M}_3) latent exposure variables to update at each iteration of the MCMC algorithm
=> High-dimensional posterior distribution
- Idea 1 => Block updating of U_i after defining 239 homogeneous groups of miners (hierarchical clustering) based on information on :
 - ▶ Mine location
 - ▶ Type of mine
 - ▶ Job type
- Idea 2 => Update $\log(X)$ instead of X in order to respect the constraint $X > 0$ and to improve convergence

Simulation study : Do shared and unshared errors have the same impact on risk estimation when assuming survival disease models ?

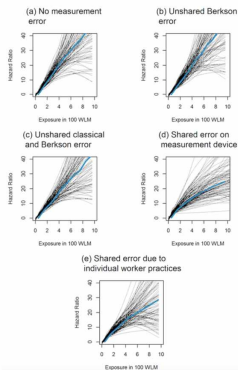
- Simulation of survival times and exposure measurement errors when assuming $\beta = 5$ (Excess Hazard Ratio (EHR)), $\beta = 2$ (Cox) and $\sigma = 0.9$
- Radon exposure data of the French cohort of uranium miners
- Bayesian inference of an EHR or a Cox model **without** accounting for error

Type of error		EHR			Cox		
		$\hat{\beta}$	Relative bias	Coverage rate	$\hat{\beta}$	Relative bias	Coverage rate
unshared	Berkson	4.65	-0.07	0.91	1.25	-0.38	0.00
	classical	4.34	-0.13	0.78	0.83	-0.59	0.00
between	Berkson	4.69	-0.06	0.93	1.25	-0.38	0.00
	classical	4.44	-0.11	0.85	0.80	-0.60	0.00
within	Berkson	3.98	-0.20	0.73	0.76	-0.62	0.00
	classical	3.03	-0.39	0.13	0.39	-0.81	0.00
none		4.90	-0.02	0.96	1.96	-0.02	0.95

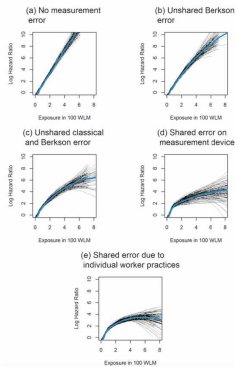
Simulation study : Can error structures that occur in retrospective studies cause an attenuation of the exposure-risk relationship ?

- Bayesian inference of natural cubic splines without accounting for exposure measurement error to study the shape of the exposure-response relationship

EHR with $\beta = 5$



Cox with $\beta = 2$



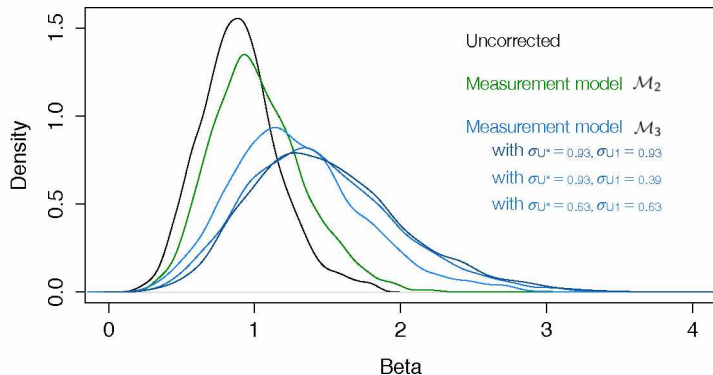
Application to the French cohort of uranium miners

Posterior medians and 95% Credible Intervals for the risk coefficient of death by lung cancer due to radon exposure

Model	Uncorrected Full cohort	Measurement model \mathcal{M}_2 Full cohort	Measurement model \mathcal{M}_3 Full cohort	Uncorrected Post-55
	EHR per 100 WLM	EHR per 100 WLM	EHR per 100 WLM	EHR per 100 WLM
Linear	0.87 [0.42;1.50]	0.99 [0.49;1.71]	1.44 [0.66;2.69]	2.75 [1.18; 5.32]
Piecewise linear				
<50 WLM	2.57 [1.18;4.57]	3.05 [1.10;3.54]	2.70 [0.95;5.11]	3.20 [0.80;7.03]
≥50 WLM	0.30 [-0.15;0.98]	0.44 [-0.13;1.11]	1.00 [-0.04;2.95]	2.82 [0.19;8.14]

Sensitivity to the error variance parameters

Posterior distribution for β

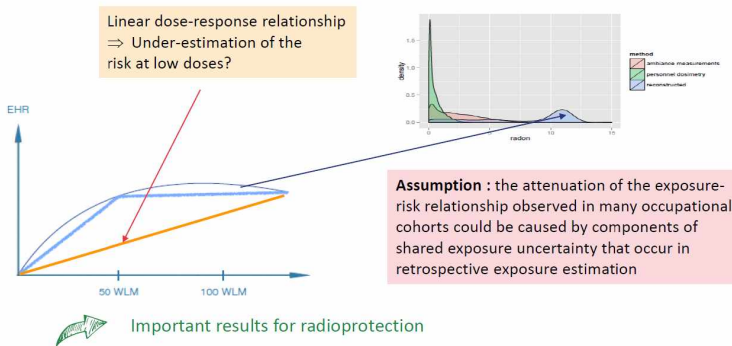


Conclusions (1/2)

- First application of the Bayesian hierarchical approach to account for exposure measurement error in radiation epidemiology
- We also extended the proposed hierarchical models to account for left-censored and missing exposure data prone to exposure measurement error
 - ▶ Belloni M, Guihenneuc C, Rage E, Ancelet S. (2020) *A Bayesian hierarchical approach to account for left-censored and missing exposure data prone to classical measurement error when analyzing lung cancer mortality due to γ -rays in the French cohort of uranium miners*. Radiat Environ Biophys; 59 :423-437.
- Application to the French cohort of uranium miners
 - ▶ When accounting for a combination of shared Berkson and classical measurement error
 - We observed a marked increase in the risk coefficient estimated in the linear EHR model
 - There was a less pronounced attenuation of the exposure-risk relationship
 - ▶ When accounting for unshared measurement error, we do not observe a substantial difference between corrected and uncorrected risk estimates (results not shown)
 - Hoffmann S, Rage E, Laurier D, Laroche P, Guihenneuc C, Ancelet S. (2017) *Accounting for Berkson and classical measurement error in radon exposure assessment using a Bayesian structural approach in the analysis of lung cancer mortality in the French cohort of uranium miners*. Radiat Res; 187(2) :196-209.

Conclusions (2/2)

- Results on simulated data
 - ▶ Measurement error shared within individuals can cause more bias in risk estimation than unshared measurement error or shared measurement error between individuals when assuming survival disease sub-models (**New result**)



- ▶ Hoffmann S, Laurier D, Rage E, Guihenneuc C, Ancelet S. (2018) *Shared and unshared measurement error in occupational cohort studies and their effects on statistical inference in proportional hazards models*. Plos One ; 13(2) :e0190792.

Limitations & Perspectives

- Corrected risk coefficients sensitive to measurement error variance parameters
- Adaptive Metropolis-Within-Gibbs algorithm => Very time-consuming to explore high-dimensional posterior distributions
 - ▶ Hamiltonian Monte-Carlo? Variational Bayes? Other Bayesian learning approaches?
 - ▶ Future application to a joint cohort of about 70000 uranium miners (French, Czech, German) in the subtask 6.4.1 of the RadoNorm European project
- Robustness to measurement and/or exposure model misspecification? => To check!
- Bayesian model selection for high-dimensional hierarchical models?

- **Atelier Statistique SFdS : "Apprentissage statistique des modèles à variables latentes"**

- ▶ En distanciel les **7, 8, 9, 10, 11 décembre 2020 (9h-13h30)**
- ▶ Animé par Estelle KUHN (INRAE), Maud DELATTRE (INRAE) et Sophie DONNET (INRAE)
- ▶ Inscription payante (tarifs réduits) avant le 30 novembre 2020 !

SFdS

- **Journée AppliBUGS**

- ▶ Séminaire scientifique sur les techniques et (surtout) l'application des méthodes statistiques bayésiennes dans divers domaines
- ▶ En distanciel les **17 et 18 décembre 2020 (9h-12h30)**
- ▶ Programme de la journée disponible sur "<http://genome.jouy.inra.fr/applibugs/applibugs.rencontres.html>"
- ▶ Séminaire gratuit mais inscription obligatoire



Pour plus d'infos, me contacter par email à sophie.ancelet@irsn.fr !