

Modélisation hiérarchique bayésienne pour la prise en compte d'erreurs de mesure d'exposition complexes dans les études de cohorte. Application en épidémiologie des rayonnements ionisants.

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

¹Institut de Radioprotection et de Sécurité Nucléaire (IRSN), 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

Journées de STATistique de Rennes (JSTAR), INSA Rennes, 4 Avril 2019

1 Context & Objectives

2 Measurement error characteristics in occupational cohort studies

3 Simulation studies : the effects of different error structures on statistical inference

4 Application 1 : Accounting for radon exposure measurement error in the French cohort of uranium miners

5 Application 2 : Accounting for left-censored gamma exposure prone to measurement error in the French cohort of uranium miners

6 Conclusion

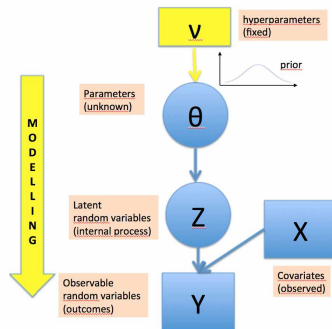
Exposure measurement error in epidemiological studies

- 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 occupational 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 error structures
 - ▶ Disjoint steps to estimate true exposure and risk parameters
 - ▶ Difficult to obtain confidence intervals
- The Bayesian hierarchical approach
 - ▶ Provides a natural way of combining exposure and parameter uncertainty in a coherent framework
 - ▶ Flexible approach to describe and account for complex measurement error :
 - Measurement error in time-varying exposures
 - Different types of measurement error simultaneously
 - Heteroscedastic measurement error variances
 - Uncertainty in measurement error variance parameters
 - ▶ 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

Bayesian hierarchical approach : Overview



- Latent random variables Z describe an **unobserved internal random process** \Rightarrow Out-of-control internal fluctuations (variability)
 - ▶ Ex : "true" exposure, organ-specific dose, inter-individual variability
- Complex models to infer...
- Under the Bayesian paradigm : one additional level of modelling : assigning a prior probability distribution to θ to describe epistemic uncertainty.

Objectives

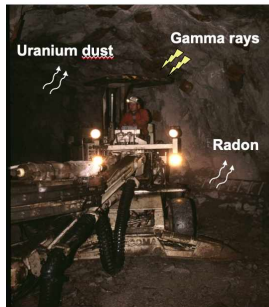
Promote the use of **Bayesian hierarchical approach** to account for exposure measurement error in risk estimates in epidemiology

- 1 Study the **impact** of different error structures on statistical inference on simulated data
- 2 Propose different **hierarchical models** to account for exposure uncertainty in risk estimates
- 3 Implement **Bayesian inference** for the proposed hierarchical models in the French cohort of uranium miners

Motivating case study : the French cohort of uranium miners (1/2)

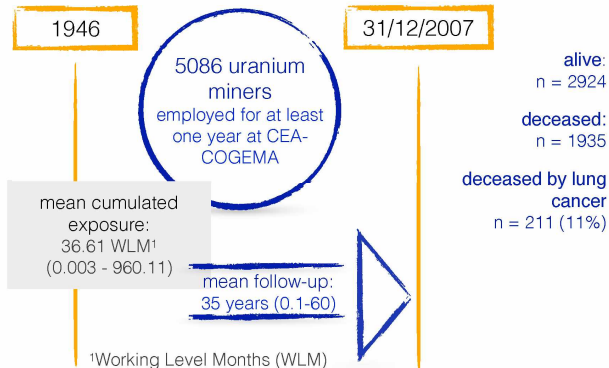
The French cohort of uranium miners allows to estimate the health effects of **chronic low-dose exposure** to :

- radon (Inhalation exposure)
- γ -rays (External exposure)
- uranium dust (Inhalation exposure)



- Radon is a radioactive gas which presents the **primary source of background radiation**
- Radon is the **second cause of lung cancer** [Samet and Eradze, 2000]
- General population is also exposed to γ -rays (natural and artificial sources)
- γ -rays **exposure** may increase the risk of lung and liver cancer [Vacquier et al., 2011] and the risk of childhood leukemia [Evrard et al., 2006]

Motivating case study : the French cohort of uranium miners (2/2)



Challenges

Survival data, individual and longitudinal exposure data prone to complex measurement error and potentially combined with a censoring process, weak signal in the data ...

1 Context & Objectives

2 Measurement error characteristics in occupational cohort studies

3 Simulation studies : the effects of different error structures on statistical inference

4 Application 1 : Accounting for radon exposure measurement error in the French cohort of uranium miners

5 Application 2 : Accounting for left-censored gamma exposure prone to measurement error in the French cohort of uranium miners

6 Conclusion

Different types of measurement error

Classical measurement error

$$Z_i(t) = X_i(t) \cdot U_i(t)$$

- $U_i(t) \perp X_i(t)$
- $Var(Z_i(t)) > Var(X_i(t))$

$Z_i(t)$: observed exposure

$X_i(t)$: true exposure

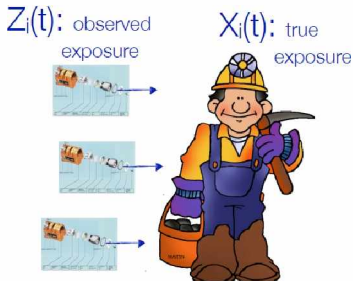


Different types of measurement error

Classical measurement error

$$Z_i(t) = X_i(t) \cdot U_i(t)$$

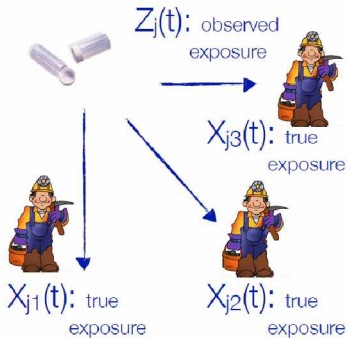
- $U_i(t) \perp X_i(t)$
- $Var(Z_i(t)) > Var(X_i(t))$



Berkson error

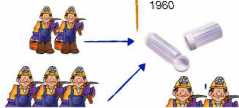
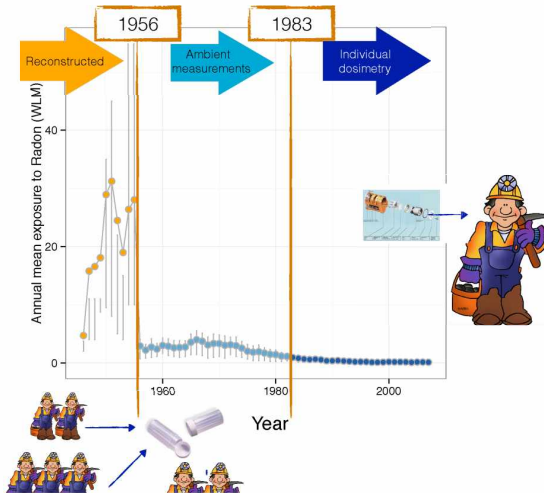
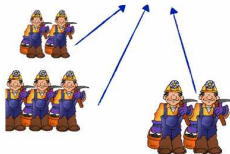
$$X_{ji}(t) = Z_j(t) \cdot U_{ji}(t)$$

- $U_i(t) \perp Z(t)$
- $Var(X_{ij}(t)) > Var(Z(t))$



Example : Radon exposure in the French cohort of uranium miners

MINE	
Henriette	
De 1947	10
à 1950	10
1951	10
1952	10
1953	10
1954	10
1955	10

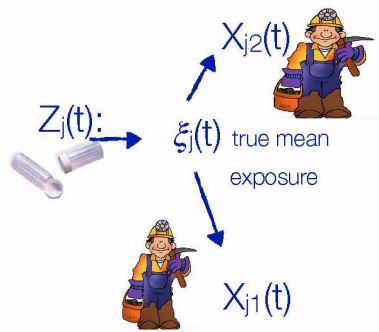


Period 3 - Individual dosimetry

Unshared
classical measurement error



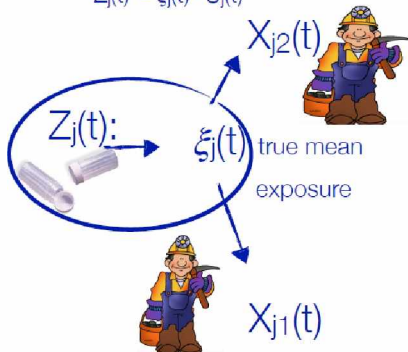
Period 2 - Prospective group exposure assessment



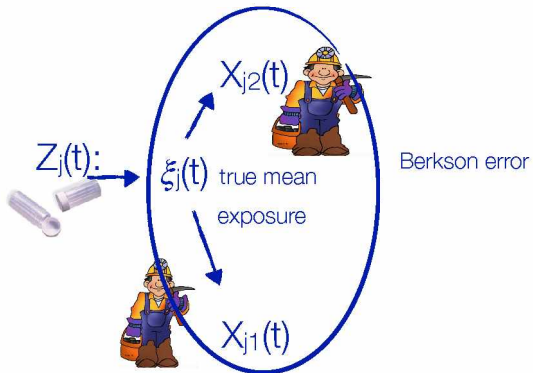
Period 2 - Prospective group exposure assessment

Classical measurement error
shared between miners

$$Z_j(t) = \xi_j(t) \cdot U_j(t)$$

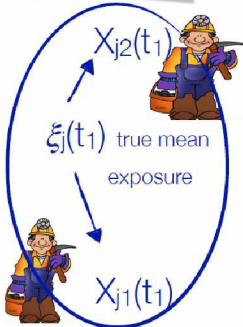


Period 2 - Prospective group exposure assessment

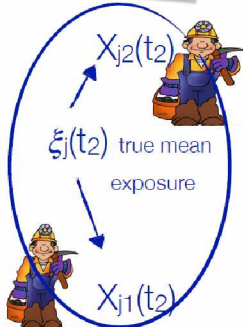


Period 2 - Prospective group exposure assessment

Timepoint 1

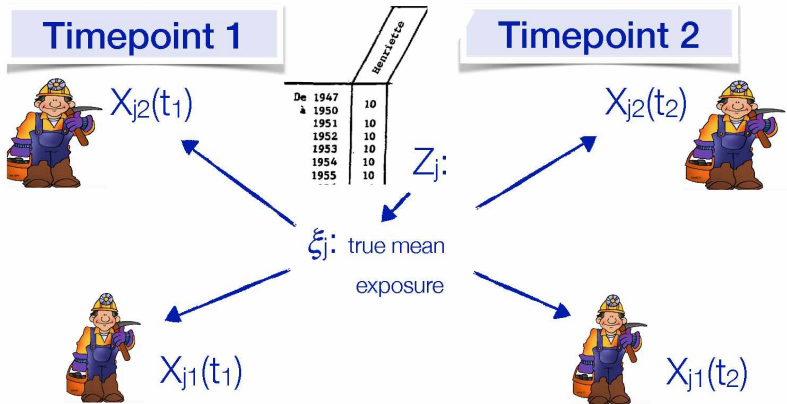


Timepoint 2

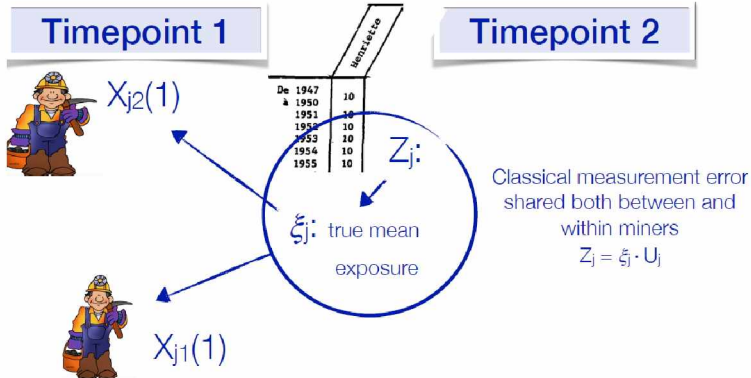


Berkson error shared within miners
 $X_{ij}(t) = \xi_j(t) \cdot U_i$

Period 1 - Retrospective group exposure assessment

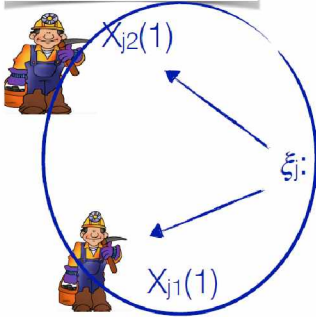


Period 1 - Retrospective group exposure assessment

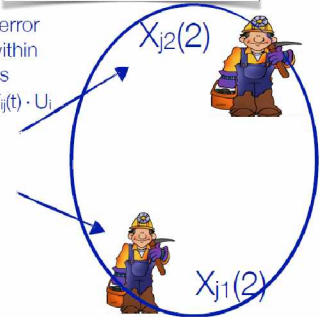


Period 1 - Retrospective group exposure assessment

Timepoint 1



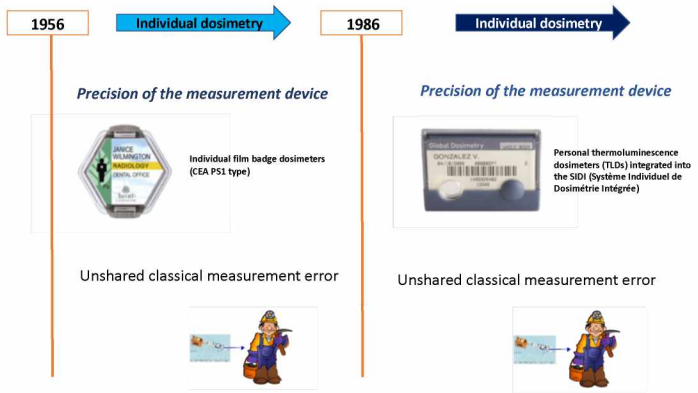
Timepoint 2



Berkson error shared within miners
 $X_{ij}(t) = \xi_j \cdot T_{ij}(t) \cdot U_i$

ξ_j : true mean exposure

Summary of gamma measurement error characteristics in the cohort



1 Context & Objectives

2 Measurement error characteristics in occupational cohort studies

3 Simulation studies : the effects of different error structures on statistical inference

4 Application 1 : Accounting for radon exposure measurement error in the French cohort of uranium miners

5 Application 2 : Accounting for left-censored gamma exposure prone to measurement error in the French cohort of uranium miners

6 Conclusion

The effects of different structures of measurement error

- In simulation studies that assume unshared heteroscedastic error, only a mild attenuation of the exposure-response relationship is observed [Stayner et al. 2003, Steenland et al. 2015]
- Error components that are shared between individuals or for several years of the same individual could have more impact on risk estimation than unshared error components [Kipnis et al. 2001, Simon et al. 2015]

- ① *Do shared and unshared error components have the same impact on risk estimation ?*
⇒ Simulation study 1
- ② *Can error structures which are typical for an occupational cohort study cause an attenuation of the exposure-risk relationship ?*
⇒ Simulation study 2

Design of the simulation study

- Use exposure data of the French cohort of uranium miners

Design of the simulation study

- Use exposure data of the French cohort of uranium miners
- Generate
 - 1 Measurement error assuming one period of exposure with measurement error that is unshared, shared between workers or within workers
 - 2 Measurement error assuming 3 periods of exposure with changes in the type of measurement error that reflect the exposure conditions of an occupational cohort

Design of the simulation study

- Use exposure data of the French cohort of uranium miners
- Generate
 - ① Measurement error assuming one period of exposure with measurement error that is unshared, shared between workers or within workers
 - ② Measurement error assuming 3 periods of exposure with changes in the type of measurement error that reflect the exposure conditions of an occupational cohort
- Compare the effects when data are generated according to the
 - ▶ Cox model : $h_i(t) = h_0(t) \cdot \exp(\beta X_i^{cum}(t))$
 - ▶ Excess Hazard Ratio model : $h_i(t) = h_0(t) \cdot (1 + \beta X_i^{cum}(t))$

where $h_0(t)$ is described by a piecewise-constant model

Design of the simulation study

- Generate
 - ① Measurement error assuming one period of exposure with measurement error that is unshared, shared between workers or within workers
 - ② Measurement error assuming 3 periods of exposure with changes in the type of measurement error that reflect the exposure conditions of an occupational cohort
- Compare the effects when data are generated according to the
 - ▶ Cox model : $h_i(t) = h_0(t) \cdot \exp(\beta X_i^{cum}(t))$
 - ▶ Excess Hazard Ratio model : $h_i(t) = h_0(t) \cdot (1 + \beta X_i^{cum}(t))$where $h_0(t)$ is described by a piecewise-constant model
- Generate failure times as a function of time-varying covariates using a method based on piecewise-exponential variables proposed by [Hendry, 2014]

Design of the simulation study

- Generate
 - 1 Measurement error assuming one period of exposure with measurement error that is unshared, shared between workers or within workers
 - 2 Measurement error assuming 3 periods of exposure with changes in the type of measurement error that reflect the exposure conditions of an occupational cohort

- Compare the effects when data are generated according to the

- ▶ Cox model : $h_i(t) = h_0(t) \cdot \exp(\beta X_i^{cum}(t))$
- ▶ Excess Hazard Ratio model : $h_i(t) = h_0(t) \cdot (1 + \beta X_i^{cum}(t))$

where $h_0(t)$ is described by a piecewise-constant model

- Generate failure times as a function of time-varying covariates using a method based on piecewise-exponential variables proposed by [Hendry, 2014]
- Conduct statistical inference without accounting for measurement error

Simulation study 1 : One period of exposure

Aim : Compare the impact of shared and unshared error on risk estimation Berkson error :

$$X_{ij}(t) = Z_j(t) \cdot U_{ij}(t)$$

Classical measurement error :

$$Z_{ij}(t) = X_{ij}(t) \cdot U_{ij}(t)$$

where :

- $U_{ij}(t) \perp\!\!\!\perp Z_j(t)$ for Berkson error and $U_{ij}(t) \perp\!\!\!\perp X_{ij}(t)$ for classical error
- $U_{ij}(t) = U_j(t)$ for error shared between workers,
- $U_{ij}(t) = U_i$ for error shared within workers
- $U_{ij}(t), U_j(t), U_i \sim \mathcal{LN}(-\frac{\sigma^2}{2}, \sigma^2) \Rightarrow$ No systematic error

Results : The impact of measurement error on risk estimation

Impact of measurement error for $\beta = 5$ (EHR), $\beta = 2$ (Cox) and $\sigma = 0.9$

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

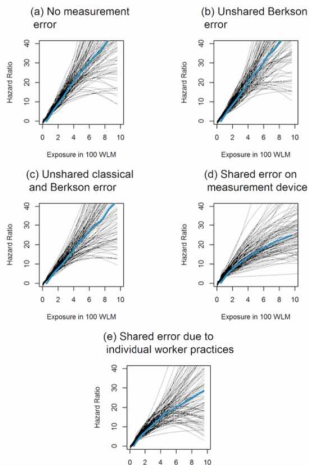
Simulation study 2 : Three periods of exposure

Aim : Compare the effects of shared and unshared components of error on the shape of the exposure-risk relationship

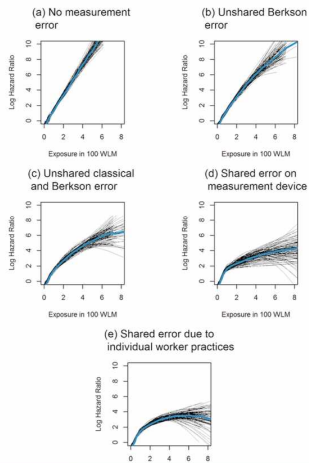
- Introduce different types of measurement error for the three exposure periods
- Study the shape of the exposure-response relationship via natural cubic splines

Results : Effects of error structures on the exposure-risk relationship

EHR with $\beta = 5$



Cox with $\beta = 2$



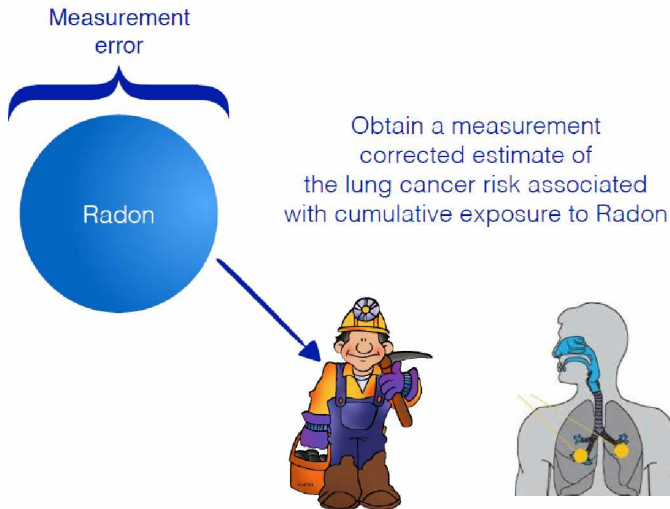
Summary of results on simulated data

- Simulation study 1 :
 - ▶ More bias in risk estimation caused by measurement error shared within workers than shared between workers or unshared
- Simulation study 2 :
 - ▶ The attenuation of the exposure-risk relationship observed in many occupational cohorts could be caused by components of shared exposure uncertainty that occur in a retrospective exposure estimation

⇒ Important to account for measurement error in occupational cohort studies and to distinguish shared and unshared error components

- 1 Context & Objectives
- 2 Measurement error characteristics in occupational cohort studies
- 3 Simulation studies : the effects of different error structures on statistical inference
- 4 Application 1 : Accounting for radon exposure measurement error in the French cohort of uranium miners**
- 5 Application 2 : Accounting for left-censored gamma exposure prone to measurement error in the French cohort of uranium miners
- 6 Conclusion

Aim



Accounting for exposure measurement error through a hierarchical model

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

- Disease model
- Measurement model
- Exposure model

- ⇒ Joint modelling of all sub-models using conditional independence assumptions
[Richardson and Gilks, 1993]
- ⇒ Joint estimation of all unknown quantities

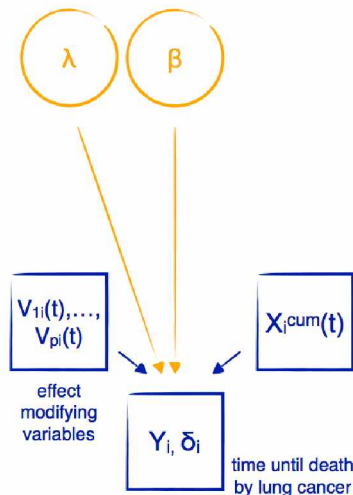
The disease model

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 and exposure model : \mathcal{M}_1

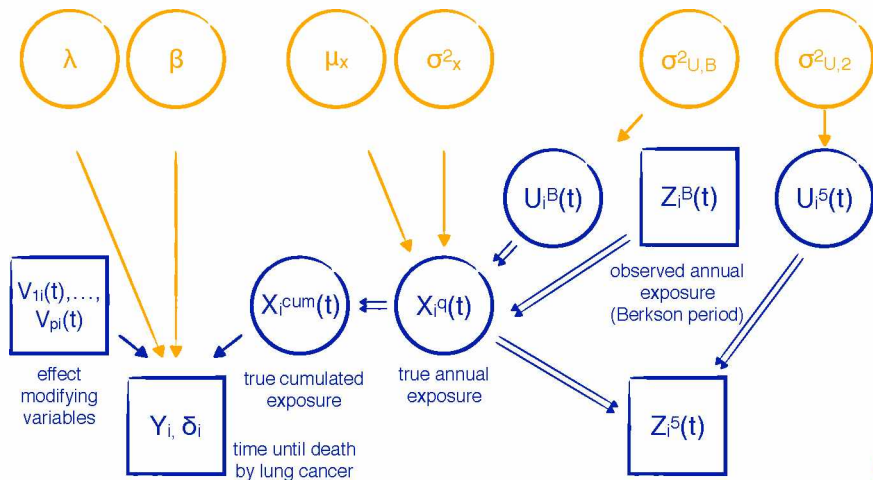
The measurement model

- Berkson error 1946 -1982 :
 $X_i(t) = Z_i(t) \cdot U_i(t)$
 where $\mathbb{E}(U_i(t)|Z_i(t)) = 1$
- Classical error after 1983 :
 $Z_i(t) = X_i(t) \cdot U_i(t)$
 where $\mathbb{E}(U_i(t)|X_i(t)) = 1$
- $U_i(t) \sim \mathcal{LN}\left(-\frac{\sigma_{p_i(t)}^2}{2}, \sigma_{p_i(t)}^2\right)$ [Darby, 1998]
- Assume different error variances $\sigma_{p_i(t)}^2$ for the periods p 1946-1955, 1956-1974, 1975-1977, 1978-1982 and 1983-2007

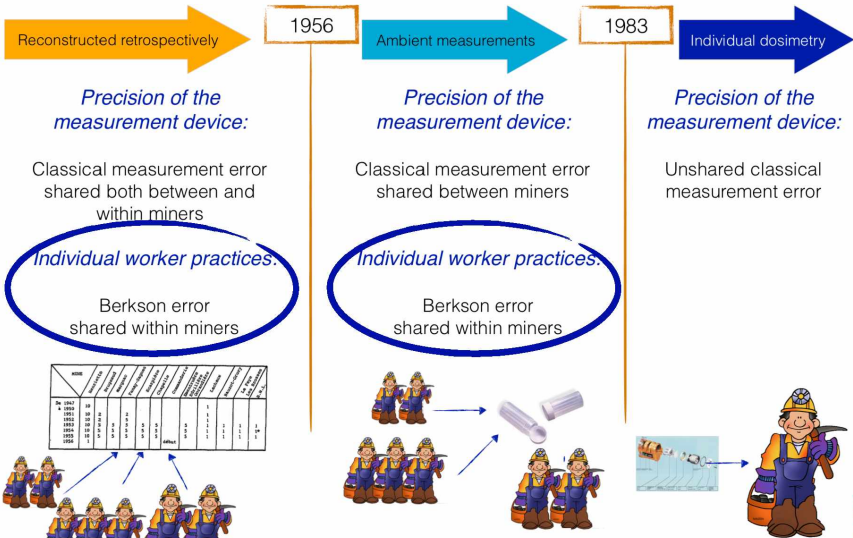
The exposure model for classical measurement error

- $X_{iq} \sim \mathcal{LN}(\mu_x, \sigma_x^2)$ [Lubin et al. 1995, Heid et al. 2002]
- Assume four distinct exposure periods in the classical measurement error period (1983-1985, 1986-1988, 1989-1992, 1992-2001)

Accounting for unshared measurement error



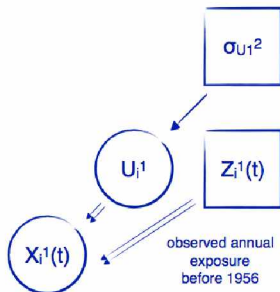
Measurement model \mathcal{M}_2



Measurement model \mathcal{M}_2

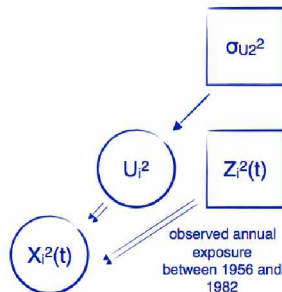
First period 1946 -1955:

- Berkson error shared within miners:
 - $X_i^1(t) = Z_i^1(t) \cdot U_i^1$
 - $U_i^1 \sim \mathcal{LN}\left(-\frac{\sigma_{U1}^2}{2}, \sigma_{U1}^2\right)$

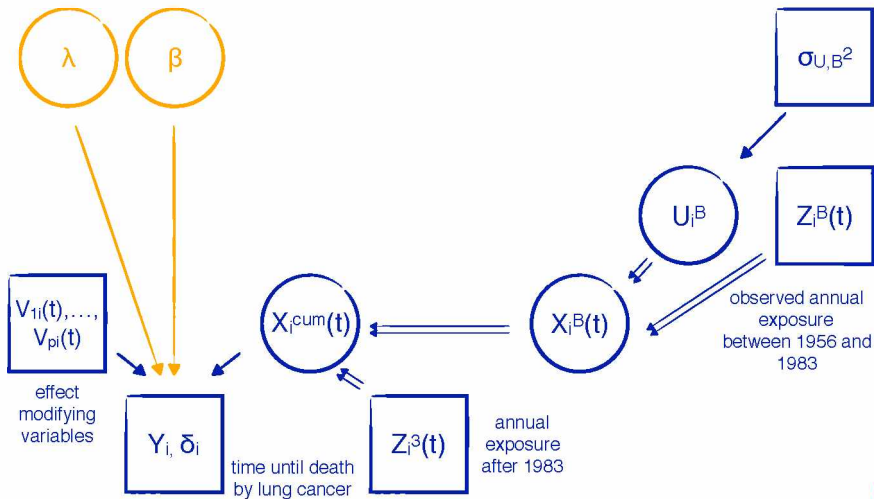


Second period 1956 -1982:

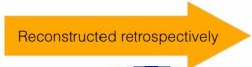
- Berkson error shared within miners:
 - $X_i^2(t) = Z_i^2(t) \cdot U_i^2$
 - $U_i^2 \sim \mathcal{LN}\left(-\frac{\sigma_{U2}^2}{2}, \sigma_{U2}^2\right)$



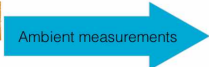
Accounting for measurement model \mathcal{M}_2



Measurement model \mathcal{M}_3



1956



1983



Precision of the measurement device:
Classical measurement error shared both between and within miners

Individual worker practices:
Berkson error shared within miners

Precision of the measurement device:
Classical measurement error shared between miners

Individual worker practices:
Berkson error shared within miners

Precision of the measurement device:
Unshared classical measurement error

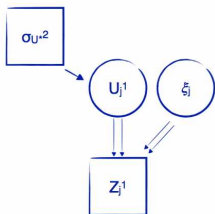
Year	Country	Measurement Type	Device	Accuracy	Resolution	Dynamic Range	Stability	Temperature Range	Humidity Range	Pressure Range	Power Source	Size	Weight	Cost
1950s	USA	Ambient	Geiger-Mueller	±10%	0.1	0.1 - 10	±5%	0 - 40°C	0 - 100%	1013 - 1015 hPa	Battery	10 cm	1 kg	\$50
1960s	USA	Ambient	Geiger-Mueller	±10%	0.1	0.1 - 10	±5%	0 - 40°C	0 - 100%	1013 - 1015 hPa	Battery	10 cm	1 kg	\$50
1970s	USA	Ambient	Geiger-Mueller	±10%	0.1	0.1 - 10	±5%	0 - 40°C	0 - 100%	1013 - 1015 hPa	Battery	10 cm	1 kg	\$50
1980s	USA	Individual	Geiger-Mueller	±10%	0.1	0.1 - 10	±5%	0 - 40°C	0 - 100%	1013 - 1015 hPa	Battery	5 cm	0.5 kg	\$100
1990s	USA	Individual	Geiger-Mueller	±10%	0.1	0.1 - 10	±5%	0 - 40°C	0 - 100%	1013 - 1015 hPa	Battery	5 cm	0.5 kg	\$100
2000s	USA	Individual	Geiger-Mueller	±10%	0.1	0.1 - 10	±5%	0 - 40°C	0 - 100%	1013 - 1015 hPa	Battery	5 cm	0.5 kg	\$100
2010s	USA	Individual	Geiger-Mueller	±10%	0.1	0.1 - 10	±5%	0 - 40°C	0 - 100%	1013 - 1015 hPa	Battery	5 cm	0.5 kg	\$100
2020s	USA	Individual	Geiger-Mueller	±10%	0.1	0.1 - 10	±5%	0 - 40°C	0 - 100%	1013 - 1015 hPa	Battery	5 cm	0.5 kg	\$100



Measurement model \mathcal{M}_3

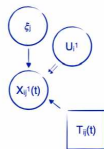
First period 1946 -1955:

- Classical measurement error shared both within and between miners:
 - $Z_j^1 = \xi_j \cdot U_j^1$
 - $U_j^1 \sim \mathcal{LN}\left(-\frac{\sigma_{U^*}^2}{2}, \sigma_{U^*}^2\right)$

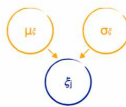


- Berkson error shared within miners:

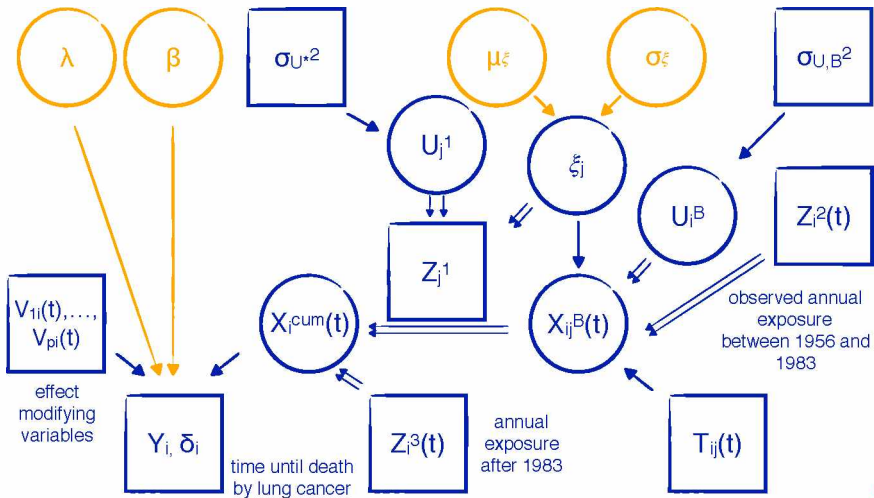
- $X_{ij}^1(t) = \xi_j \cdot T_{ij}(t) \cdot U_i^1$
- $U_i^1 \sim \mathcal{LN}\left(-\frac{\sigma_{U1}^2}{2}, \sigma_{U1}^2\right)$



- The exposure model:
 - $\xi_j \sim \mathcal{LN}(\mu_\xi, \sigma_\xi)$

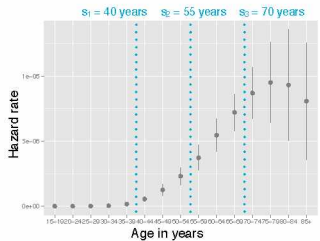


Accounting for measurement model \mathcal{M}_3



Prior distributions

- $[\beta]$: $\beta \sim \mathcal{N}(0, 10^4)$
truncated to guarantee $h_i > 0$
- $[\lambda]$: $\lambda_j \sim \mathcal{G}(\alpha_{0j}, \lambda_{0j})$ for each component j , $j = 1, \dots, 4$
based on the lung cancer mortality in the general French male population between 1968 and 2005



- $[\mu_X]$: $\mu_X \sim \mathcal{N}(-1.44, 10.24)$
- $[\sigma_X^2]$: $\sigma_X^2 \sim IG(1.75, 0.88)$
- $[\sigma_U]$: Normal distributions with small variances, centered around "guess estimates" : $\hat{\sigma}_{U,B_1}=0.94$, $\hat{\sigma}_{U,B_2}=0.47$; $\hat{\sigma}_{U,B_3}=0.42$; $\hat{\sigma}_{U,B_4}=0.33$; $\hat{\sigma}_{U,C_5}=0.10$

Bayesian inference - Model \mathcal{M}_3

Target: Joint posterior distribution of $\theta = (\beta, \lambda, \mathbf{X}, \xi, \mu_\xi, \sigma_\xi)$

$$[\theta|y, Z] \propto [\beta][\lambda] \prod_{i=1}^n \left[y_i \mid \sum_{q=1}^{Q_i} X_{iq}, \beta, \lambda \right]$$

$$\cdot \prod_{i=1}^n \prod_{q=1}^{Q_2} [X_{iq}^2 | Z_{iq}^2, \sigma_{U2}^2]$$

$$\cdot \prod_{i=1}^n \prod_{q=1}^{Q_1} [X_{iq}^1 | \xi_j, T_{ij}(t), \sigma_{U1}^2]$$

$$\cdot [\mu_\xi][\sigma_\xi] \cdot [Z_j | \xi_j, \sigma_{U^*}^2] [\xi_j | \mu_\xi, \sigma_\xi]$$

- Complex
- No analytical solution
- High-dimensional

⇒ Adaptive Metropolis-Within-Gibbs algorithm developed and tested in Python

Main difficulty : high dimensional latent variables

- About 6,000 (Model \mathcal{M}_3) to 49,000 (Model \mathcal{M}_1) latent exposure variables to update at each iteration of the MCMC algorithm!
- Idea 1 \Rightarrow **Block updating** of X after defining 239 homogeneous groups of miners (hierarchical clustering) based on information on :
 - ▶ Mine location
 - ▶ Type of mine
 - ▶ Job type
- Idea 2 \Rightarrow Update $\log(X)$ instead of X in order to respect the constraint $X > 0$ and to improve convergence

Results when accounting for unshared measurement error (\mathcal{M}_1)

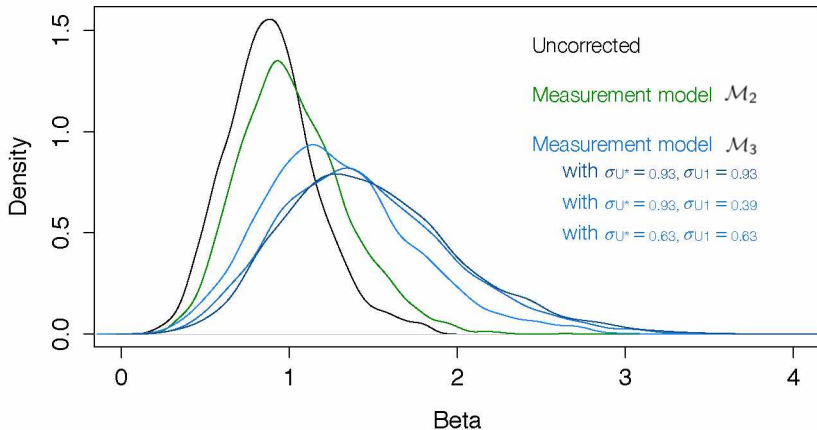
Model	Uncorrected		Corrected	
	EHR per 100 WLM	DIC	EHR per 100 WLM	DIC
Linear	0.87 [0.42;1.50]	5435.08	0.90 [0.51;1.41]	5433.30
Period of exposure		5428.12		5423.59
until 1955	0.31 [-0.01;0.80]		0.31 [0.02;0.79]	
after 1955	1.94 [1.15;2.96]		2.06 [1.34;3.00]	
Piecewise linear		5429.31		5422.50
<50 WLM	2.57 [1.18;4.57]		2.12 [1.07;3.37]	
≥50 WLM	0.30 [-0.15;0.98]		0.34 [0.02;0.98]	

Results when accounting for shared measurement error (\mathcal{M}_2 & \mathcal{M}_3)

Model	Uncorrected Full cohort EHR per 100 WLM	Measurement model \mathcal{M}_2 Full cohort EHR per 100 WLM	Measurement model \mathcal{M}_3 Full cohort EHR per 100 WLM	Uncorrected Post-55 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]
\geq 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]

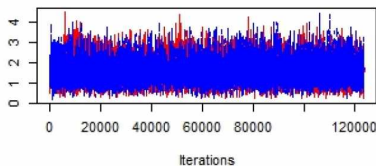
Results : comparing the posterior distributions for β

Posterior distribution for β

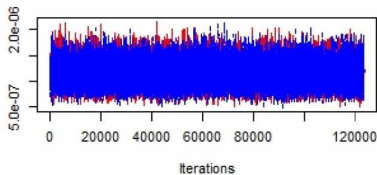


Convergence diagnostics : Trace plots

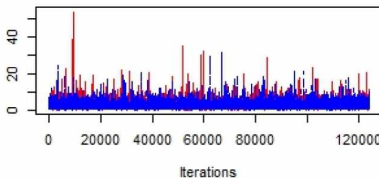
Trace of beta1



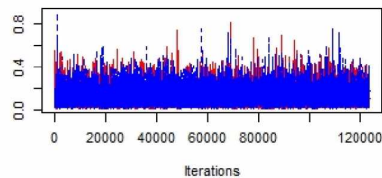
Trace of lambda2



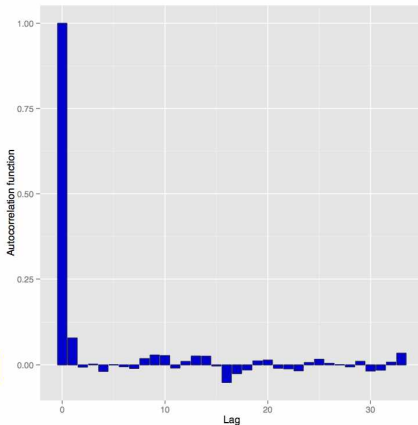
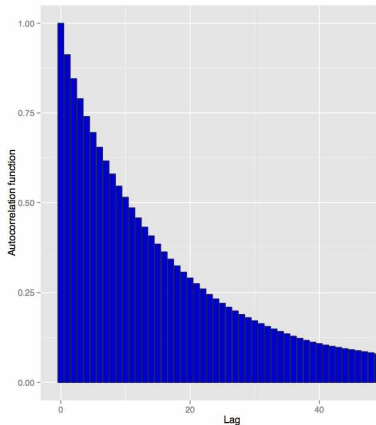
Trace of sigma_ksi



Trace of ksi_FOR



Convergence diagnostics : Intra-Chain autocorrelation



Summary of measurement error corrected results

- When accounting for a combination of shared Berkson and classical measurement error
 - ▶ We observe a **marked increase** in the risk coefficient estimated in the linear EHR model
 - ▶ There is 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

1 Context & Objectives

2 Measurement error characteristics in occupational cohort studies

3 Simulation studies : the effects of different error structures on statistical inference

4 Application 1 : Accounting for radon exposure measurement error in the French cohort of uranium miners

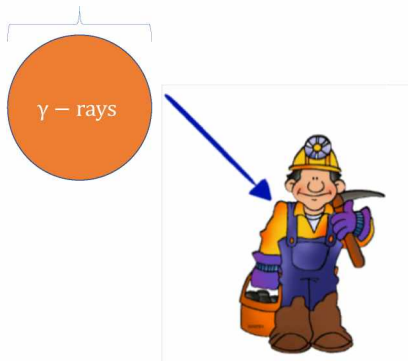
5 Application 2 : Accounting for left-censored gamma exposure prone to measurement error in the French cohort of uranium miners

6 Conclusion

Aim

Measurement uncertainty:

- Measurement error
- Left-censored exposure
- Missing exposure



Obtain a « corrected » estimate of the lung cancer risk associated with cumulative exposure to γ - rays

Classification of zeros and missing γ -rays exposure data in the post-55 French cohort of uranium miners

γ -rays exposure recorded	Co-exposure measurement	Miner's job position	γ -rays exposure "most likely" measured	Classification	Proportion of γ -rays exposure values (in %)
0	Radon=0 and LLR=0	Regular	= 0	True zero	2.75%
	Radon>0 or LLR>0	Regular	<DL**	(False zero) Left-censored	4.85%
NA*	Radon=0 and LLR=0	Regular	=0	Zero	0.50%
	Radon=NA and LLR=0				
	Radon=0 and LLR=NA				
	Radon>0 or LLR>0	Regular	<DL	Left-censored	0.45%
	Radon=NA and LLR=NA	Regular	<DL	Left-censored	1.41%
Expatriate		>0	>0	0.30%	
	Other	=0	Zero	0.28%	

* missing exposure value

** denotes the detection limit of the dosimeter: DL(t)=2.2 mGy for t between 1956 and 1985, and

DL(t)=0.55 mGy for t between 1986 and 2007

The hierarchical model (1/2)

The disease sub-model

The same survival model (Excess Hazard Ratio) as Application 1 !

The measurement sub-model

- $X_i(t)$: True γ -rays exposure of miner i at time t
- $Z_i(t)$: Strictly positive measured or "most likely" measured γ -rays (Left censored !)
- Outcome of interest : $(W_i(t), \delta_{it}^W)$ where :

$$W_i(t) = \max(Z_i(t), DL(t))$$

$$\delta_{it}^W = 1 \quad \text{if} \quad Z_i(t) \geq DL(t) \quad (\text{0 otherwise})$$

- Classical error :

$$Z_i(t) = X_i(t) \cdot U_i(t)$$

$$U_i(t) \sim^j \mathcal{LN}\left(-\frac{\sigma_{U,q(t)}^2}{2}, \sigma_{U,q(t)}^2\right)$$

with $\sigma_{U,1}^2 = 0.245$ (1956-1985); $\sigma_{U,2}^2 = 0.16$ (1986-2007) [Allodji et al., 2011]

The hierarchical model (2/2)

Three considered exposure sub-models

$$\log(X_i(t)) \sim \mathcal{N}(\mu_{x,i}(t), \tau_x) \quad \forall i \forall t$$

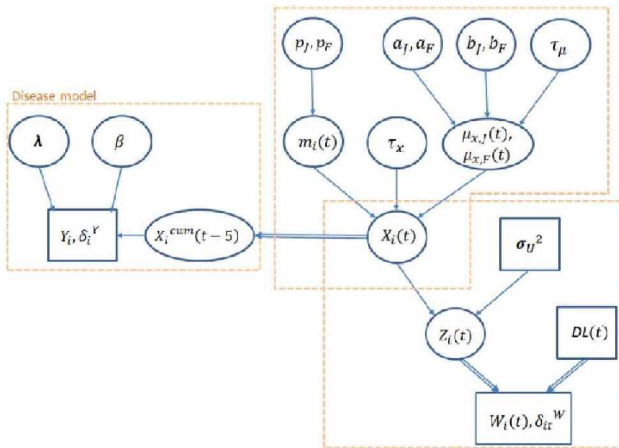
with :

$$\mathcal{M}_1 : \mu_{x,i}(t) = \mu_{x,p(t)} \text{ with } p(t) = \{1, 2, 3, 4, 5\}$$

$$\mathcal{M}_2 : \mu_{x,i}(t) \sim \mathcal{N}(a.f(t) + b, \tau_\mu) \text{ with } f(t) = t - 1956$$

$$\mathcal{M}_3 : \mu_{x,i}(t) \sim \mathcal{N}(a_{m_i(t)}.f(t) + b_{m_i(t)}, \tau_\mu) \text{ with } m_i(t) = \{J, F\}$$

The Directed Acyclic Graph (Exposure sub-model = \mathcal{M}_3)



Bayesian inference

- Implementation in Python 3.4 of an adaptive Metropolis-Within-Gibbs algorithm
 - ▶ Variance of proposal distributions : calibrated to target an acceptance rate of 40% for single parameters and 20% for vectors of latent variables
 - ▶ 3 Markov chains with different initial values
 - ▶ Adaptive phase : 100 cycles based on 100 iterations
 - ▶ Burn-in phase : 10,000 iterations, additional iterations : 55,000, thin step = 30
 - ▶ Posterior samples : 5,500 values.
- Difficulty : high dimensional latent variables \Rightarrow Long computation time

Impact of the exposure sub-model on lung cancer risk estimates

Table 4 Posterior medians and 95% CI of the EHR (per 100mGy) for lung cancer mortality in the post-55 sub-cohort, assuming 3 different exposure sub-models. Watanabe-Akaike Information Criterion (WAIC) for the three exposure sub-models.

Model	EHR per 100 mGy	WAIC
M ₁	0.90 [0.33; 2.00]	2461.359
M ₂	0.91 [0.34; 1.93]	2461.253
M ₃	0.90 [0.34; 1.89]	2459.912

Comparing the annual expected true γ -rays exposure (at log scale) in underground an open-pit mines

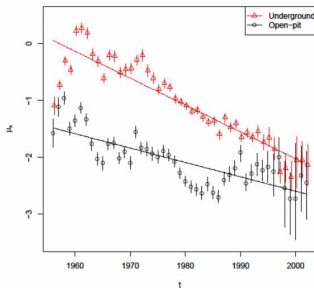
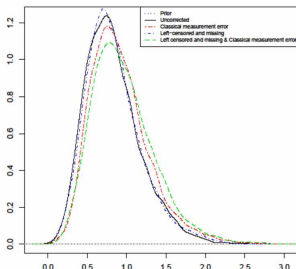


Figure 3 Annual expected true log-transformed γ -rays exposure in underground mines (i.e., $\mu_{x,U}(t)$) and open-pit mines (i.e., $\mu_{x,O}(t)$) in the post-55 sub-cohort of French uranium miners. The circles and triangles represent the posterior medians and the segments indicate the 95% credible intervals.

Impact of exposure measurement uncertainty on lung cancer risk estimates

Posterior medians and 95% credible intervals (95%CI) of the EHR (per 100mGy) for lung cancer mortality in the post-55 sub-cohort

Model	Uncorrected EHR per 100mGy	Full model EHR per 100 mGy	Classical error only EHR per 100mGy	Censoring only EHR per 100 mGy
Linear	0.78 [0.28;1.64]	0.90 [0.34, 1.89]	0.86 [0.34, 1.85]	0.79 [0.28 ; 1.72]



1 Context & Objectives

2 Measurement error characteristics in occupational cohort studies

3 Simulation studies : the effects of different error structures on statistical inference

4 Application 1 : Accounting for radon exposure measurement error in the French cohort of uranium miners

5 Application 2 : Accounting for left-censored gamma exposure prone to measurement error in the French cohort of uranium miners

6 Conclusion

Conclusion

- 1 The Bayesian hierarchical approach allows to account for complex structures of exposure measurement uncertainty (i.e., measurement error, left-censored, missing exposure) in risk estimation
- 2 Results on simulated data
 - ▶ Measurement error shared within miners can cause more bias in risk estimation than unshared measurement error (for the considered survival disease models !)
 - ▶ Error structures in an occupational cohort study can lead to an attenuation of the exposure-risk relationship
- 3 In the French cohort of uranium miners
 - ▶ Marked increase (65% !) in the lung cancer risk coefficient (posterior median) when accounting for shared Berkson and classical measurement error on radon exposure
 - ▶ Slight increase (15%) in the lung cancer risk coefficient (posterior median) when accounting for classical measurement error on gamma exposure
 - ▶ Wider 95% CI when accounting for exposure uncertainty but still do not include 0
 - ▶ No deleterious impact on lung cancer risk estimates when replacing all left-censored and missing γ -rays exposures by zero BUT only small proportion (about 7%) of these exposures in the post-55 sub-cohort

Limitations & Perspectives

- It was not possible to adjust for tobacco consumption, exposure to diesel exhaust, arsenic, asbestos and silica quartz : No or only partial information available
- The estimated risk coefficients which are corrected for measurement error are very sensitive to the chosen values for the measurement error variance parameters
- Our adaptive Metropolis-Within-Gibbs algorithm is quite inefficient when exploring high-dimensional posterior distributions \Rightarrow Hamiltonian Monte-Carlo ?
- Robustness to measurement and/or exposure model misspecification ? \Rightarrow To check !

References

- 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.
- Hoffmann S, Guihenneuc C, Ancelet S. (2018) *A cautionary comment on the generation of Berkson error in epidemiological studies*. Rad Environ Biophys 2018; 57(2) :189-193
- 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.
- Hendry D.J. (2014) *Data generation for the Cox proportional hazards model with time-dependent covariates : a method for medical researchers*. Statistics in Medicine; 33 :436-454

Thank you for your attention !

Design of the simulation study : Generate failure times with time-varying covariates

Adapted algorithm from the paper by Hendry (2014)

- Choose $h_0(t) = \frac{\partial[g^{-1}(t)]}{\partial t}$ with $g(0) = 0, g(t) \nearrow$ and $g^{-1}(t)$ differentiable
- Define a time partition $S = \{s_0, s_1, \dots, s_J\}$ with $s_0 = 0$ such as $Z_i(t) = Z_{ij} \forall t \in [s_{j-1}, s_j]$
- Choose a parameter value β
- For $i \in \{1, \dots, n\}$
 - ▶ Calculate $\{\gamma_{ij}\}_{j=1}^J = \{\exp(Z_{ij}\beta)\}_{j=1}^J$
 - ▶ Generate W_i according to a truncated piecewise exponential distribution with rates $\gamma_{i1}, \dots, \gamma_{iJ}$ on the partition $\{g^{-1}(s_1), \dots, g^{-1}(s_J)\}$
 - ▶ Calculate the failure time of interest $T_i = g(W_i)$
 - ▶ Generate the censoring time C_i (e.g., $C_i \sim Unif[a, b]$)
 - ▶ Derive $Y_i = \min(T_i, C_i)$ and the binary indicator δ_i^Y

Regression calibration

- Basic idea: Replace \mathbf{X} by the regression of \mathbf{X} on (\mathbf{Z}, \mathbf{W}) where \mathbf{W} are predictors measured without error
- Algorithm:
 - Using replication, validation or instrumental data, estimate the regression of \mathbf{X} on (\mathbf{Z}, \mathbf{W})
 - Replace the unobserved \mathbf{X} by its estimate and run a standard analysis to obtain parameter estimates
 - Adjust the resulting standard errors to account for exposure uncertainty using either the bootstrap or a sandwich method

Effect of calendar period on baseline hazard

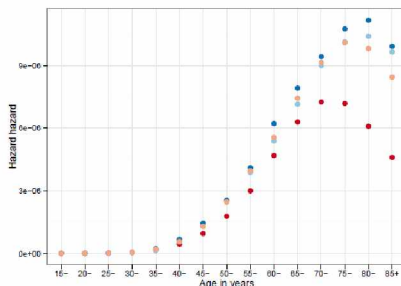


Figure: Hazard of lung cancer mortality in French males for the following periods: 1968-1977 (red), 1978-1987 (orange), 1988-1997 (blue), 1998-2005 (lightblue)

Resulting priors:

$$[\lambda_1] \sim \mathcal{G}(23.6, 4.90 \cdot 10^8), \quad [\lambda_2] \sim \mathcal{G}(35.5, 2.58 \cdot 10^7),$$

$$[\lambda_3] \sim \mathcal{G}(88.1, 1.61 \cdot 10^7), \quad [\lambda_4] \sim \mathcal{G}(29.7, 3.25 \cdot 10^6),$$