

# A Bayesian hierarchical approach to account for shared exposure measurement error in an occupational cohort

Sabine Hoffmann<sup>1,2</sup>, Chantal Guihenneuc<sup>3</sup>, Sophie Ancelet<sup>1</sup>

<sup>1</sup>Institut de Radioprotection et de Sûreté Nucléaire, Paris

<sup>2</sup>Ludwig-Maximilians University, Munich

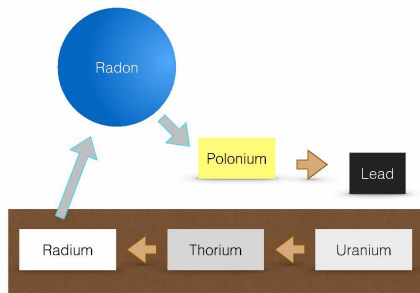
<sup>3</sup>Université Paris Descartes, Paris

ISCB 2019, Leuven

# Background

- Radon is a radioactive gas which presents the **primary source of background radiation**
- It is known to be the **second leading cause of lung cancer** [Samet and Eradze, 2000] responsible for about 2% of cancer deaths in Europe [Darby et al., 2005]

**Cohorts of uranium miners** present an important source of information on the association between radon and lung cancer



# Background

- Exposure measurement error may cause bias in risk estimates, a distortion in the exposure-risk relationship and a loss in power

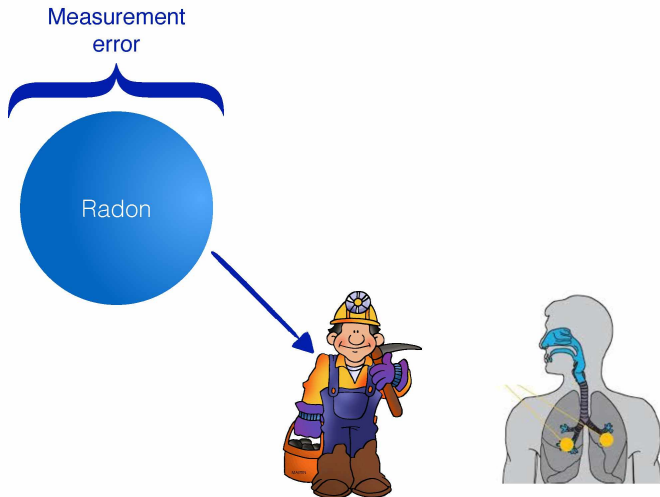
# Background

- Exposure measurement error may cause bias in risk estimates, a distortion in the exposure-risk relationship and a loss in power
- Classical methods to account for measurement error may lack the flexibility to account for complex measurement error characteristics

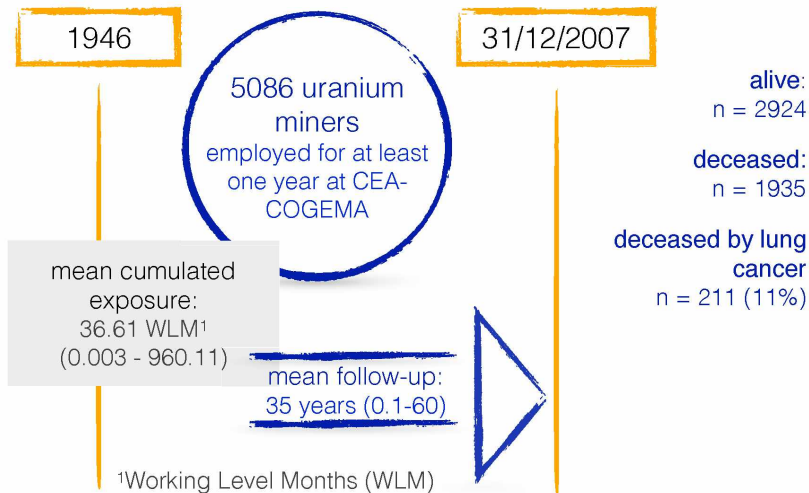
# Background

- Exposure measurement error may cause **bias in risk estimates**, a **distortion in the exposure-risk relationship** and a **loss in power**
- Classical methods to account for measurement error may lack the flexibility to account for **complex measurement error characteristics**
- The hierarchical Bayesian approach provides a flexible and coherent way to model
  - Measurement error in time-varying exposures
  - Heteroscedastic measurement error
  - Measurement error that is shared between or within subjects

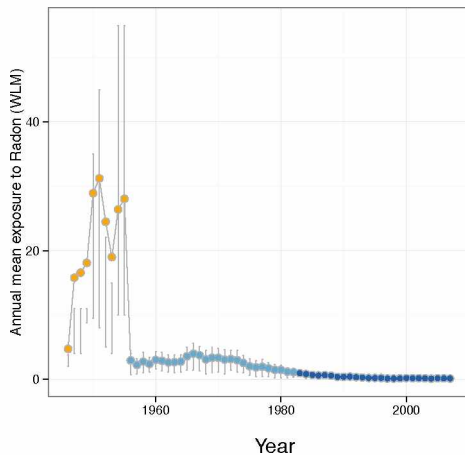
# Account for exposure measurement error in the association between radon exposure and lung cancer mortality



# The French Uranium Miners' Cohort

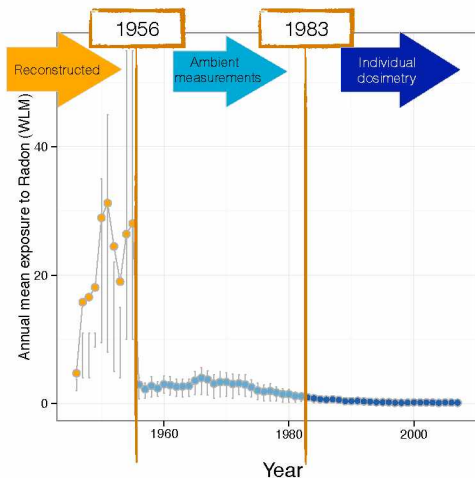


# Radon exposure in the French cohort of uranium miners



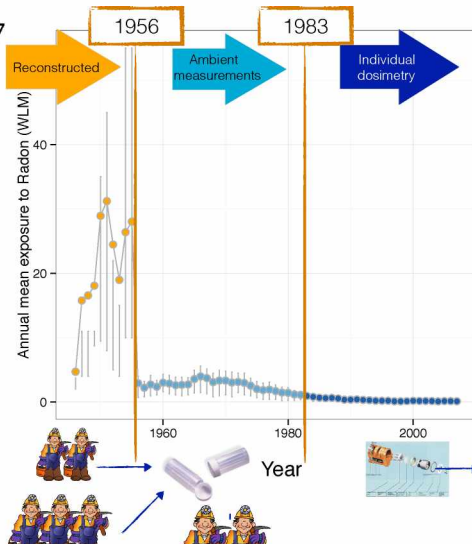
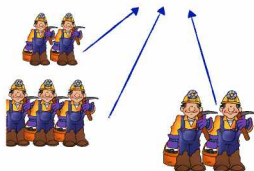


## Radon exposure in the French cohort of uranium miners



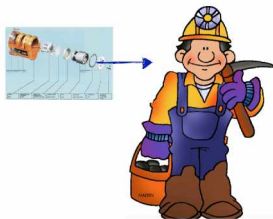
# Radon exposure in the French cohort of uranium miners

MINE	Henriette	Brugseaud	Margnac	Franç-Sagnes	Scarprière	Chapelle
De 1947 à 1950	10					
1951	10	2		2		
1952	10	2		2		
1953	10	5	5	5	5	5
1954	10	5	5	5	5	5
1955	10	5	5	5	5	5
1956	1					

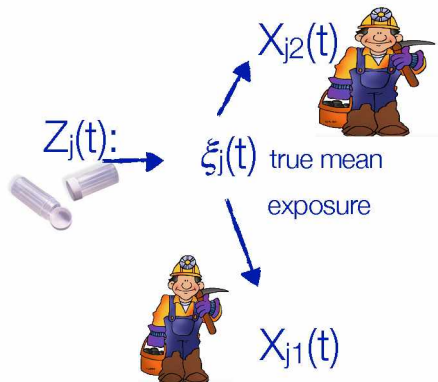


# Period 3 - Individual dosimetry

Unshared  
classical measurement error

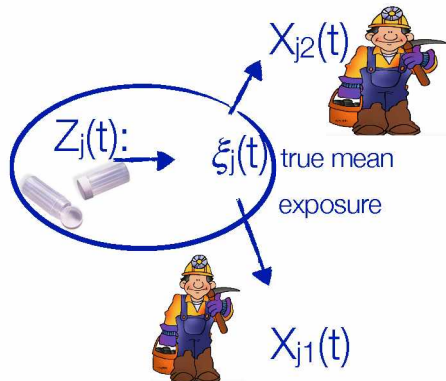


## Period 2 - Prospective group exposure assessment

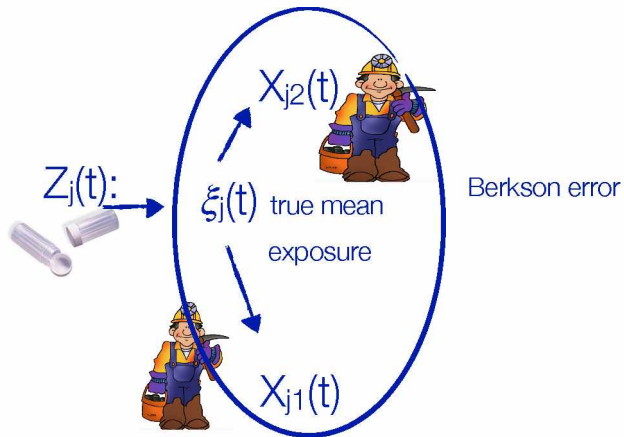


## Period 2 - Prospective group exposure assessment

Classical measurement error  
shared between miners

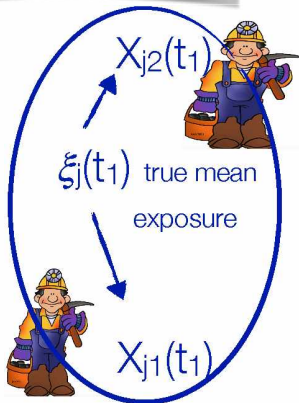


## Period 2 - Prospective group exposure assessment



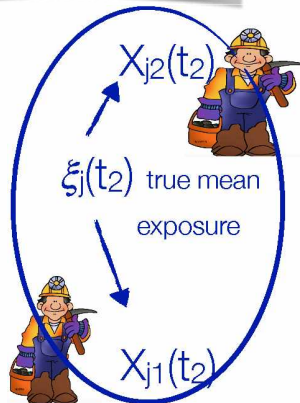
# Period 2 - Prospective group exposure assessment

## Timepoint 1

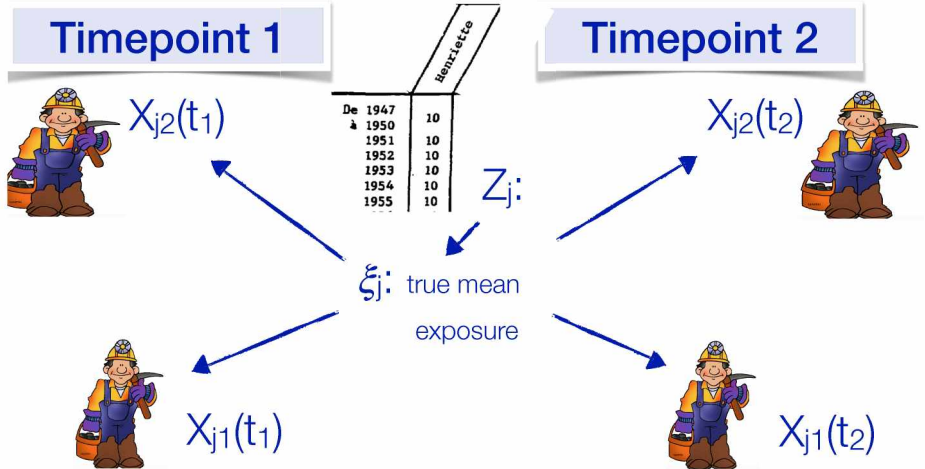


Berkson error  
shared within  
miners

## Timepoint 2



# Period 1 - Retrospective group exposure assessment





# Period 1 - Retrospective group exposure assessment

## Timepoint 1


 $X_{j2}(t_1)$ 

 $X_{j1}(t_1)$ 


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

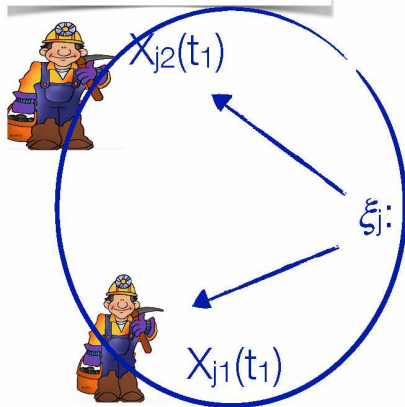
 $Z_j:$ 
 $\xi_j:$  true mean  
exposure

## Timepoint 2

Classical measurement error  
shared both between and  
within miners

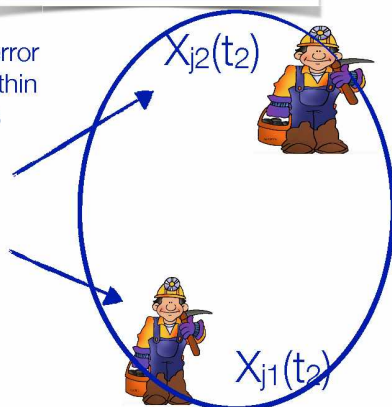
# Period 1 - Retrospective group exposure assessment

## Timepoint 1



Berkson error  
shared within  
miners

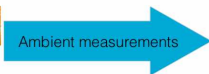
## Timepoint 2



# Summary of measurement error characteristics in the cohort



1956



1983



*Precision of the measurement device:*

Classical measurement error shared both between and within miners

*Individual worker practices:*

Berkson error shared within miners

Year	Measurement Method									
	1956-1957	1958	1959	1960	1961	1962	1963	1964	1965	1966
1956	100									
1957	100									
1958	100	10								
1959	100	20	10							
1960	100	30	20	10						
1961	100	40	30	20	10					
1962	100	50	40	30	20	10				
1963	100	60	50	40	30	20	10			
1964	100	70	60	50	40	30	20	10		
1965	100	80	70	60	50	40	30	20	10	
1966	100	90	80	70	60	50	40	30	20	10



*Precision of the measurement device:*

Classical measurement error shared between miners

*Individual worker practices:*

Berkson error shared within miners

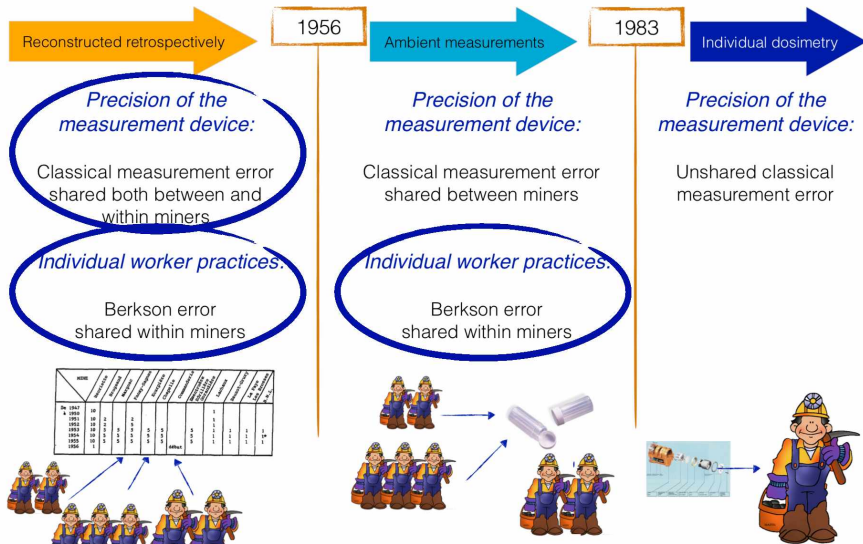


*Precision of the measurement device:*

Unshared classical measurement error



# Summary of measurement error characteristics in the cohort



# Modelling classical measurement error

In a hierarchical Bayesian approach, we can account for **classical measurement error** using the following conditional independence models ([Richardson and Gilks, 1993]):

- **disease model:**  $[Y_i | \mathbf{X}_i, \beta]$
- **measurement model:**  
 $[Z_i | \mathbf{X}_i, \sigma]$  for classical error
- **exposure model:**  $[\mathbf{X}_i | \mathbf{V}_i, \pi]$  for classical error

# Modelling Berkson measurement error

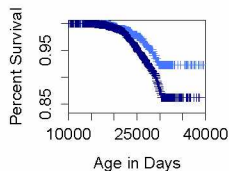
In a hierarchical Bayesian approach, we can account for **Berkson measurement error** using the following conditional independence models ([Richardson and Gilks, 1993]):

- **disease model:**  $[Y_i | \mathbf{X}_i, \beta]$
- **measurement model:**  
 $[\mathbf{X}_i | \mathbf{Z}_i, \sigma]$  for Berkson error

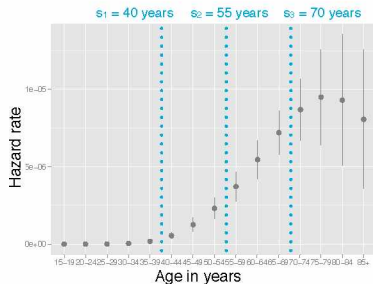
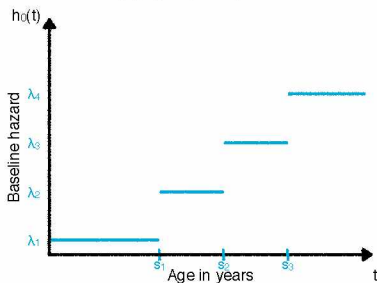
# The disease model: Modelling failure times

- $Y_i = \min(T_i, C_i)$  ,  $\delta_i = \mathbb{1}_{[T_i < C_i]}$

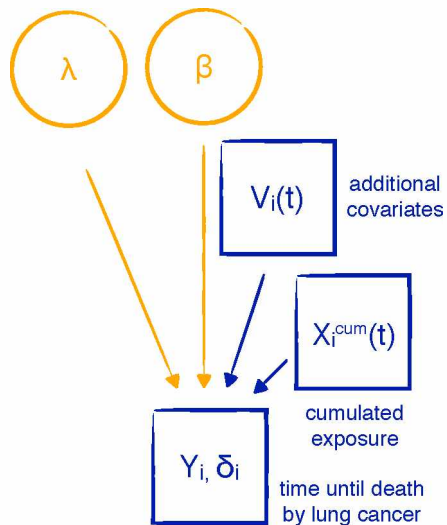
$$h_i(t) = h_0(t) \left( 1 + \beta \sum_{q=1}^Q X_{iq}(t) t_j^{-5} \right)$$



- Describe  $h_0(t)$  by a piecewise constant hazard model



## The disease model





# Measurement and exposure models

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:

- $X_{ij}^1(t) = \xi_j \cdot T_{ij}(t) \cdot U_i^1$

- $U_i^1 \sim \mathcal{LN}\left(-\frac{\sigma_{U_1}^2}{2}, \sigma_{U_1}^2\right)$

- The exposure model:

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

# Measurement and exposure models

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:

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

- The exposure model:

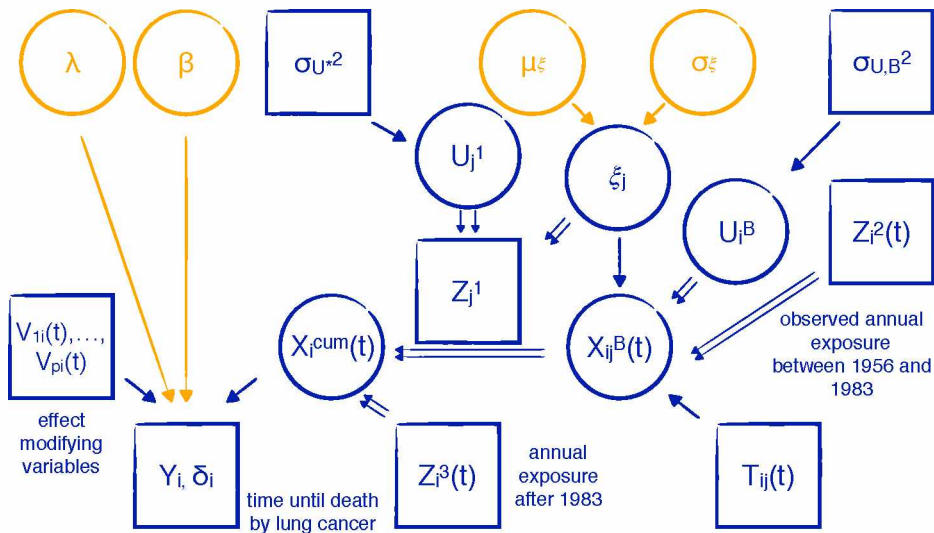
- $\xi_j \sim \mathcal{LN}(\mu_\xi, \sigma_\xi)$

Second period 1946 - 1955:

- Berkson error shared within miners:

- $X_{ij}^2(t) = Z_j^2(t) \cdot U_i^2$
- $U_i^2 \sim \mathcal{LN}\left(-\frac{\sigma_{U_2}^2}{2}, \sigma_{U_2}^2\right)$

## Accounting for shared classical and Berkson error



# Bayesian inference

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

$$\begin{aligned}
 [\theta|y, Z] &\propto [\beta][\lambda] \prod_{i=1}^n \left[ y_i \mid \sum_{q=1}^{Q_i} X_{iq}, \beta, \lambda \right] \\
 &\quad \cdot \prod_{i=1}^n \prod_{q=1}^{Q_2} [X_{iq}^2 | Z_{iq}^2, \sigma_{U2}^2] \\
 &\quad \cdot \prod_{i=1}^n \prod_{q=1}^{Q_1} [X_{iq}^1 | \xi_j, T_{ij}(t), \sigma_{U1}^2] \\
 &\quad [\mu_\xi][\sigma_\xi] \prod_{j=1}^J [Z_j | \xi_j, \sigma_{U^*}^2] [\xi_j | \mu_\xi, \sigma_\xi]
 \end{aligned}$$

# Bayesian inference

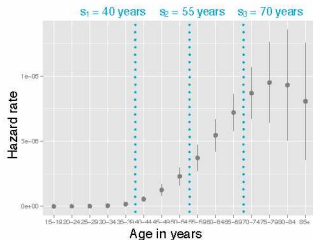
Target: Joint posterior distribution of  $\theta = (\beta, \mathbf{X}, \lambda, \sigma_\epsilon, \mu_x, \sigma_x)$

$$\begin{aligned}
 [\theta|y, Z] &\propto [\beta][\lambda] \prod_{i=1}^n \left[ y_i \mid \sum_{q=1}^{Q_i} X_{iq}, \beta, \lambda \right] \\
 &\quad \cdot \prod_{i=1}^n \prod_{q=1}^{Q_2} [X_{iq}^2 | Z_{iq}^2, \sigma_{U2}^2] \\
 &\quad \cdot \prod_{i=1}^n \prod_{q=1}^{Q_1} [X_{iq}^1 | \xi_j, T_{ij}(t), \sigma_{U1}^2] \\
 &\quad [\mu_\xi][\sigma_\xi] \prod_{j=1}^J [Z_j | \xi_j, \sigma_{U^*}^2] [\xi_j | \mu_\xi, \sigma_\xi]
 \end{aligned}$$

⇒ Adaptive Metropolis-Within-Gibbs algorithm developed and tested in Python to sample from the joint posterior distribution

# Prior distributions

- $[\beta]$ :  $\beta \sim \mathcal{N}(0, 10^4)$
- $[\lambda]$ :  $\lambda_j \sim \mathcal{G}(\alpha_{0j}, \lambda_{0j})$  for each component  $j$ ,  $j = 1, \dots, 4$



# Prior distributions

- $[\beta]$ :  $\beta \sim \mathcal{N}(0, 10^4)$
- $[\lambda]$ :  $\lambda_j \sim \mathcal{G}(\alpha_{0j}, \lambda_{0j})$  for each component  $j$ ,  $j = 1, \dots, 4$
- $[\mu_\xi]$ :  $\mu_\xi \sim \mathcal{N}(-0.8, 5.76)$

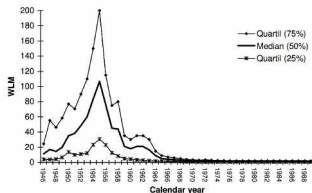


Figure 6.3-1: Quartiles (25%, 50%, 75%) of annual exposure to radon in WLM among exposed cohort members (n=50,773)

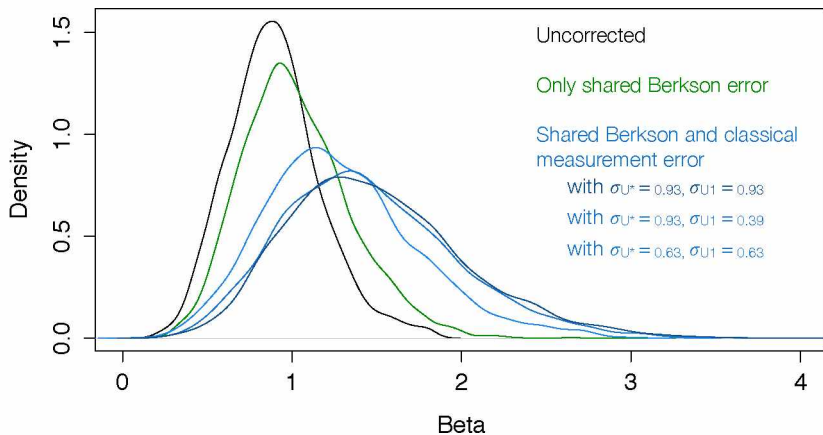
- $[\sigma_\xi^2]$ :  $\sigma_\xi^2 \sim IG(0.01, 0.01)$
- $[\sigma_U]$ : No validation sample available, fix values at estimates based on previous work [Allodji et al., 2012]

## Results

Model	Uncorrected EHR	EHR corrected for unshared ME	EHR corrected for shared Berkson error	EHR corrected for shared Berkson and classical ME
Risk estimate (per 100 WLM)	0.87	0.90	0.99	1.44
95% Credible interval	[0.49;1.36]	[0.51;1.41]	[0.48;1.73]	[0.66;2.69]



## Results

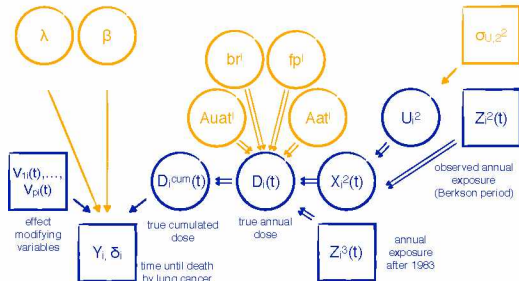
Posterior distribution for  $\beta$ 

# Discussion





- When assuming a measurement model that realistically describes the exposure uncertainty in the cohort, we find a substantial increase in the risk estimate for radon

# Discussion

- When assuming a measurement model that realistically describes the exposure uncertainty in the cohort, we find a substantial increase in the risk estimate for radon
- Outlook:
  - Assess the influence of model misspecification
  - Compare the proposed methodology with classical methods for measurement error correction
  - Account for additional sources of dose uncertainty



Thank you for your attention

-  Allodji, R. S., Leuraud, K., Bernhard, S., Henry, S., Bénichou, J., and Laurier, D. (2012).  
Assessment of uncertainty associated with measuring exposure to radon and decay products in the french uranium miners cohort.  
*Journal of Radiological Protection*, 32(1):85–100.
-  Darby, S., Hill, D., Auvinen, A., Barros-Dios, J., Baysson, H., Bochicchio, F., Deo, H., Falk, R., Forastiere, F., Hakama, M., et al. (2005).  
Radon in homes and risk of lung cancer: collaborative analysis of individual data from 13 european case-control studies.  
*Bmj*, 330(7485):223.
-  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.
-  Richardson, S. and Gilks, W. R. (1993).

A bayesian approach to measurement error problems in epidemiology using conditional independence models.

*American Journal of Epidemiology*, 138(6):430–442.



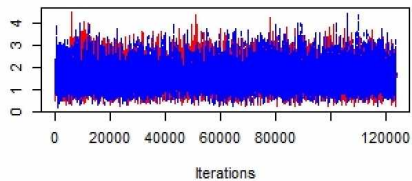
Samet, J. M. and Eradze, G. R. (2000).

Radon and lung cancer risk: taking stock at the millenium.

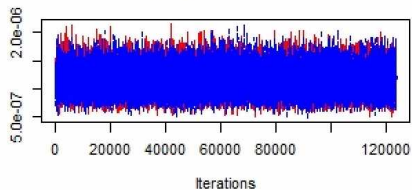
*Environmental health perspectives*, 108(Suppl 4):635–641.

# Convergence diagnostics: Trace plots

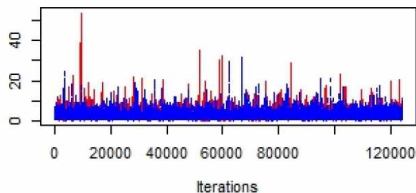
Trace of beta1



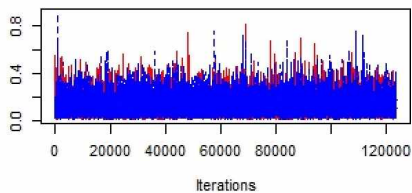
Trace of lambda2



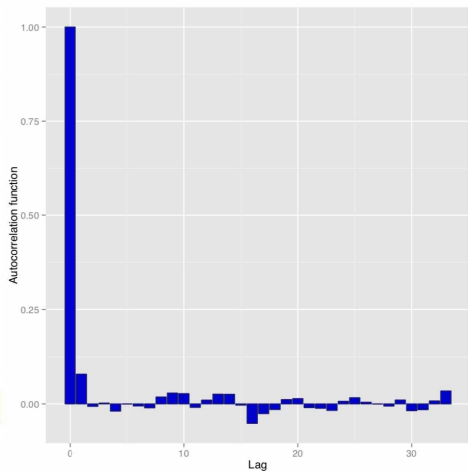
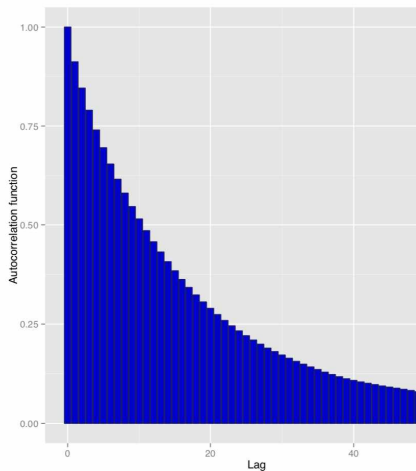
Trace of sigma\_ksi



Trace of ksi\_FOR



# Convergence diagnostics: Autocorrelations for beta





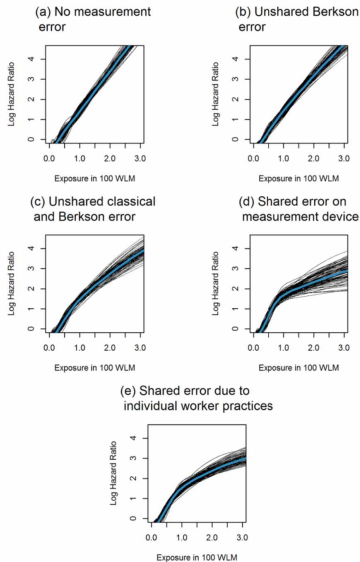
# Effects of shared and unshared measurement error

Table 1. Average posterior median ( $\hat{\beta}$ ), overall 95% credible intervals (CI<sub>95%</sub>), relative bias and coverage rate for 100 data sets generated according to the Cox model  $\mathcal{D}_2$ , a measurement model among  $\mathcal{M}_0$  to  $\mathcal{M}_8$  and a true risk coefficient of  $\beta = 2$  per 100 WLM.

Model	Type of sharing	Type of error	Error variance	$\hat{\beta}$	CI <sub>95%</sub>	Relative bias	Coverage rate
$\mathcal{M}_1$	unshared	Berkson	0.1	1.81	[1.64; 1.99]	-0.10	0.10
			0.8	1.25	[0.97; 1.49]	-0.38	0.00
$\mathcal{M}_2$		classical	0.1	1.75	[1.55; 1.93]	-0.13	0.02
			0.8	0.83	[0.45; 1.21]	-0.59	0.00
$\mathcal{M}_3$	between	Berkson	0.1	1.82	[1.62; 2.01]	-0.09	0.22
			0.8	1.25	[1.03; 1.47]	-0.38	0.00
$\mathcal{M}_4$		classical	0.1	1.75	[1.53; 1.94]	-0.13	0.05
			0.8	0.80	[0.44; 1.16]	-0.60	0.00
$\mathcal{M}_5$	within	Berkson	0.1	1.45	[1.12; 1.69]	-0.28	0.00
			0.8	0.76	[0.54; 0.97]	-0.62	0.00
$\mathcal{M}_6$		classical	0.1	1.33	[1.04; 1.58]	-0.34	0.00
			0.8	0.39	[0.17; 0.62]	-0.81	0.00
$\mathcal{M}_7$	both	Berkson	0.1	1.46	[1.11; 1.78]	-0.27	0.00
			0.8	0.77	[0.54; 1.00]	-0.62	0.00
$\mathcal{M}_8$		classical	0.1	1.42	[1.04; 1.72]	-0.29	0.00
			0.8	0.49	[0.13; 0.86]	-0.76	0.00
$\mathcal{M}_0$	none	none	0	1.96	[1.80; 2.13]	-0.02	0.95

<https://doi.org/10.1371/journal.pone.0190792.t001>

# Effects of shared and unshared measurement error



## Validation of the algorithm: Simulation study

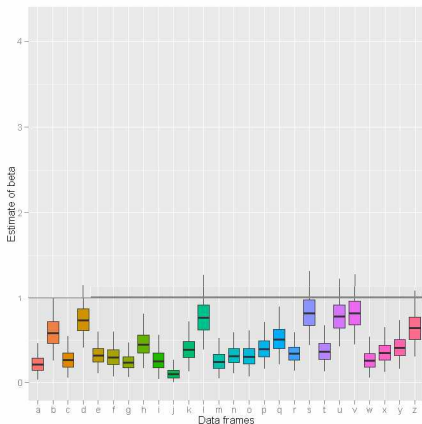
- $n = 1000$  miners
- Adapt a method based on piecewise exponential variables proposed by [Hendry, 2014] to generate survival times depending on a **continuous time-varying exposure**
- $\beta = 1$
- $\sigma_{\epsilon_1}^2 = 0.8$  and  $\sigma_{\epsilon_2}^2 = 0.5$
- Random censoring (exponential distribution)
- Flat prior distributions
- $\sigma_{\epsilon p} \sim \mathcal{N}(\hat{\sigma}_{\epsilon p}, 0.02)$  for  $p \in \{1, 2\}$

# Classical error: $\beta = 1$

## Uncorrected risk estimate:

$$\beta_{median} = 0.42 [0.17, 0.76]$$

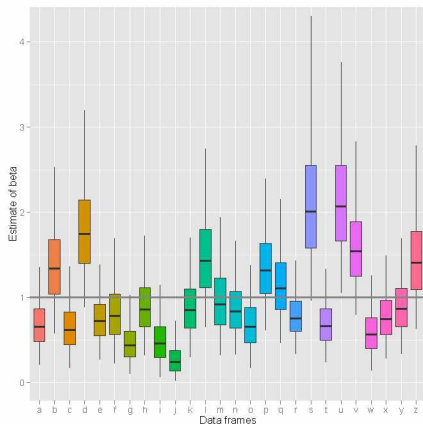
with cover probability 0.20



## Corrected risk estimate:

$$\beta_{median} = 0.995 [0.415, 1.921]$$

with cover probability 0.95

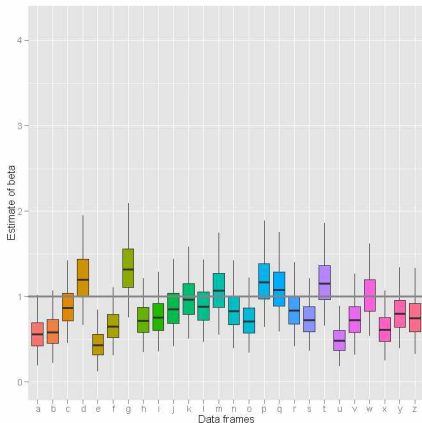


Berkson error:  $\beta = 1$

Uncorrected risk estimate:

$\beta_{median} = 0.83$  [0.42, 1.40]

with cover probability 0.90



Corrected risk estimate:

$\beta_{median} = 1.066$  [0.499, 1.966]

with cover probability 0.94

