Hierarchical modeling and Bayesian statistics for a better consideration of uncertainties when estimating radiation-related risks

Sophie Ancelet

Institute for Radiological Protection and Nuclear Safety (IRSN), France (sophie.ancelet@irsn.fr)

Workshop "Statistical methods for safety and decommissioning"
Avignon, 22nd November 2022
All exposed to ionizing radiations (IR)

Worldwide distribution of radiation exposure

- Cosmic: 13%
- Artificial: 16%
- Medical: 20%
- Soil: 42%
- Radon: 42%
- Food: 9%

Workers (France)
Average individual dose by field of activity (mSv)

- Research: 0.25 mSv/year
- Others: 0.33 mSv/year
- Medical: 0.34 mSv/year
- Nuclear: 1.17 mSv/year
- Industry: 1.38 mSv/year
- Aviation: 1.98 mSv/year

French population: average individual dose 4.5 mSv/year (IRSN, 2015)

Workers: average individual dose 0.72 mSv/year
A low and controlled exposure
Radiation dose and health effects

Relationship of radiation doses and health effects

- Burns, radiation sickness and death
- Clinically observable in individuals
- Increasing risk of cancer about 5% per Sv
- Statistically observable in populations (epidemiology)
- Biologically plausible
- Statistical limitations

- Very low
  - Natural background, occupational doses
- Low
  - Chemobyl child thyroid doses
- Moderate
- High
  - Chemobyl firemen

UNSCEAR, 2016
Current issues on stochastic effects

- Non-threshold linearity of the dose-response relationship for cancers: discrepancy between epidemiology and radiobiology
- Multi-exposure situations
- Taking into account the complexity of biological mechanisms
- Variability factors, individual susceptibility
- Tissue sensitivity, integration of new cancers
- Validity of the assessment of heritable effects, consideration of epigenetic mechanisms
International Radiological Protection system

**RESEARCH**
- Physics
- Dosimetry
- Radiochemistry
- Genetics
- Physiology
- Radiobiology
- Radiotoxicology
- Oncology
- Epidemiology
- Radioecology

**SYNTHESIS**
- United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR)

**RECOMMENDATIONS**
- International Commission on Radiological Protection
- Biological Effects Of Ionising Radiations (NAS, USA)

**RADIATION PROTECTION STANDARDS**
- International Atomic Energy Agency
- European Community EURATOM
- Country

1928
International Commission on Radiological Protection (ICRP)

- **Aim**
  - "To contribute to an appropriate level of protection for people and the environment from the adverse effects of radiation exposure, without unduly limiting the desirable human actions that may be associated with such exposure"
  - Avoiding deterministic effects and limiting stochastic effects

- **Management tool**
  - Strong simplification necessary for the practical application of radiation protection
  - Radiological detriment computed from weighted nominal risk coefficients of a given terminal event (e.g., death by cancer) on a given organ over the entire life
  - Nominal risk coefficients estimated from dose-response analysis

- **Some priority scientific issues**
  - Effects of prolonged exposures and low dose rates
  - Non-cancer effects and heritable effects, and contribution to radiological detriment
  - Mechanisms of low-dose effects and integration of these mechanisms into dose-response modeling

⇒ The assessment of the risk of stochastic effects in the current radiation protection system is mainly based on knowledge from epidemiological studies
**Some priority scientific issues**

- **Identification and estimation** of the effects of:
  - chronic or repeated exposures at **low doses** and characterization of the form of the dose-response relationship for cancer risk
  - exposure during childhood
  - non-cancer effects associated with exposure to low and moderate doses

**Main statistical aims**

- **Estimate** the magnitude of the association (and its uncertainty) between one (several) exposure(s) to ionizing radiations (IR) and a given disease
  - *(Probabilistic) modelling and statistical learning*

- **Identify** the existence of an association between one (several) exposure(s) to ionizing radiations (IR) and a given disease
  - **Statistical hypothesis testing/model selection**

- **Characterize** the shape of dose-response relationships
  - **Model selection/Model averaging**
Radiation epidemiology: an observational science

- First step: Build, validate and maintain **large databases over the long term**, in compliance with health data confidentiality constraints.

Is the occurrence of the event different in **individuals exposed to IR compared to unexposed or less exposed individuals**?
Many sources of uncertainty

- **Exposures** (measurement/estimation, left-censored value due to detection limits, missing data, ...)
- **Organ dose** estimation
- **Right-censored survival data** (competing risks...)
- **Baseline risk for rare diseases** (but not only)
- Cause of death
- Multifactorial diseases (e.g., cancer)
- Confounding factors
- **Shape of the dose-response/exposure-risk model**
- **Individual variability**
- ...
Uncertainty on radiological exposure values (predictor variables) is:

- ubiquitous
- one of the most important source of input uncertainty in epidemiological studies
Exposure uncertainty

**Classical measurement error**

\[ Z_i(t) = X_i(t) \cdot U_i(t) \]
- \( U_i(t) \perp X_i(t) \)
- \( \text{Var}(Z_i(t)) > \text{Var}(X_i(t)) \)

**Berkson error**

\[ X_{ji}(t) = Z_j(t) \cdot U_{ji}(t) \]
- \( U_i(t) \perp Z(t) \)
- \( \text{Var}(X_{ji}(t)) > \text{Var}(Z(t)) \)
Exposure uncertainty

- In retrospective cohort studies:
  - complex patterns of exposure measurement error
  - attenuation of the exposure-risk relationship for high exposure values [Stayner (2003)]
    - Measurement error?

- If not accounted for, exposure uncertainty may cause [Carroll et al. (2006)]:
  - bias in risk estimates
  - misleading conclusions about the effect of these exposures on the disease risk
  - a distortion of the exposure-risk relationship

⇒ It is important to account for exposure uncertainty in risk estimation [ICRP103 (2007); UNSCEAR (2012)]
Standard methods to account for exposure uncertainty in risk estimates

- Exposure measurement error $\Rightarrow$ Frequentist functional methods: regression calibration and simulation extrapolation \([\text{Carroll et al. (2006)}; \text{Keogh et al. (2020)}]\)
  - Lack of flexibility to account for complex measurement error on time-varying exposures
    - Mixture of different types of measurement error
    - Heteroscedastic measurement error
  - Disjoint steps to estimate "true" exposure and risk parameters
  - Applicability restricted to cases where a validation sample is available to estimate the expected value of true exposure given observed exposure or the true size of the error
  - Potential lack of consistency in risk estimates in proportional hazards models \([\text{Bartlett and Keogh, 2016}]\)
Dose uncertainty

- The health effects of IR are associated with **radiation dose** rather than with radiation exposure [Preston et al. (2013); Birchall and Marsh (2005)].
- The values of radiation dose do not only depend on the exposure to radioactive material, but also on the exposure conditions.
- The calculation of radiation doses involves further uncertainties.
Dose uncertainty

Measurement error

Uncertainties

Radon

breathing characteristics
activity median diameter
unattached fraction

Intake

Time-activity curves in the source regions

Number of nuclear transformations

Nuclear data, phantoms and particle transport

Tissue absorbed dose (mGy)

Radiation weighting factors $w_R$

Tissue equivalent dose (mSv)

Tissue weighting factors $w_T$

Effective dose (mSv)

ICRP66

Human Respiratory Tract Model (HRTM)

Extrathoracic

ET_1

Environment

Extrathoracic

LN

BB

bb

GI tract

CT

ET_2

LN

Thoracic

Al

Blood

Initial

$(1-f_s)s_3$

Bound

(f_s)s_3

IRSN

INSTITUT DE RADIOPROTECTION ET DE SÉCURITÉ NUCLEAIRES
Dose uncertainty

NCICT: National Cancer Institute dosimetry system for CT

=> Radiation transport Monte Carlo simulation within ICRP reference pediatric and adult computational anatomic phantoms
Dose uncertainty

- The input parameters of dosimetric models are uncertain ⇒ The estimation of radiation doses is uncertain when estimating the health effects of radiation exposure
- If not accounted for, dose uncertainty may cause:
  ▶ bias in risk estimates
  ▶ misleading conclusions about the effect of these exposures on the disease risk
  ▶ a distortion of the dose-response relationship
- However, they are most often neglected in epidemiological studies!
- NB: The dosimetric models are black box for epidemiologists/statisticians (but dose calculations from these models are fast)

⇒ It is important to account for dose uncertainty in risk estimation [ICRP103 (2007); UNSCEAR (2012)]
Standard methods to account for dose uncertainty in risk estimates

- **Step 1:** Simulate plausible dose values using 2-dimensional Monte-Carlo algorithm [Simon et al. (2015)]

- **Step 2:**
  - **Plug-in** of dose point estimates (i.e., empirical mean, median or other quantiles derived from the simulated dose distributions) in dose-response models
  - **Monte-Carlo Maximum Likelihood** [Stayner et al. (2007)]: Estimate the risk coefficient \( \beta \) and its uncertainty by maximizing the estimated average likelihood from a grid of fixed values for \( \beta \)

- Asymptotical confidence intervals
Model uncertainty

- In radiation epidemiology, different radiation-related risk models may fit **similarly well** to a given dataset.
- Usual practice ignores such a model uncertainty by selecting **a single model**  
  $\Rightarrow$ Some excess risks may be wrongly declared as significant or non-significant.

Model uncertainty $\rightarrow$ Uncertainty by ignorance/Epistemic uncertainty
Example: Modelling the radiation-related leukaemia excess risk

<table>
<thead>
<tr>
<th>ERR models</th>
<th>Form of $\text{ERR}_{\theta, i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNSCEAR (2006)</td>
<td>$(\alpha d_i + \beta d_i^2)\exp(\kappa \log(a_i/55))$</td>
</tr>
<tr>
<td>Qexp</td>
<td>$\beta d_i^2 \exp(\gamma d_i)\exp(\kappa \log(a_i/55))$</td>
</tr>
<tr>
<td>Sigmoid</td>
<td>$\frac{A}{\exp(B) + (\frac{1}{d_i})\exp(C)}\exp(\kappa \log(a_i/55))$</td>
</tr>
<tr>
<td>Spline</td>
<td>$[\alpha_1 d_i + \alpha_2 (d_i - d_k)1_{d_i \geq d_k}]\exp(\kappa \log(a_i/55))$</td>
</tr>
<tr>
<td>Little (2008)</td>
<td>$(\alpha d_i + \beta d_i^2)\exp(\kappa_1 \log(a_i/55) + \kappa_2 \log(e_i/25))$</td>
</tr>
<tr>
<td>Littleexp (2008)</td>
<td>$(\alpha d_i + \beta d_i^2)\exp(\gamma d_i)\exp(\kappa_1 \log(a_i/55) + \kappa_2 \log(e_i/25))$</td>
</tr>
<tr>
<td>BEIRVII (2006)</td>
<td>$\beta(s_{i+1})(d_i + \theta d_i^2)\exp(\gamma e'_i + \delta \log(t_i/25) + \phi e'_i \log(t_i/25))$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EAR models</th>
<th>Form of $\text{EAR}_{\theta, i}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNSCEAR (2006)</td>
<td>$(\alpha d_i + \beta d_i^2)\exp(\kappa_1 s_i + \kappa_2 \log(t_i/25))$</td>
</tr>
<tr>
<td>Littleexp (2008)</td>
<td>$(\alpha d_i + \beta d_i^2)\exp(\gamma d_i)\exp(\kappa_1 s_i + \kappa_2 \log(t_i/25))$</td>
</tr>
<tr>
<td>BEIRVII (2006)</td>
<td>$\beta(s_{i+1})(d_i + \theta d_i^2)\exp(\gamma e'_i + \phi e'_i \log(t_i/25))$</td>
</tr>
<tr>
<td>Schneider (2009)</td>
<td>$(1 + \alpha \tilde{s}_i)(\beta d_i + \delta d_i^2)\exp(\gamma_1 (e_i - 41) + \gamma_2 \log(a_i/60))$</td>
</tr>
<tr>
<td>Schneiderexp (2009)</td>
<td>$(1 + \alpha \tilde{s}_i)(\beta d_i + \delta d_i^2)\exp(\gamma d_i)\exp(\gamma_1 (e_i - 41) + \gamma_2 \log(a_i/60))$</td>
</tr>
<tr>
<td>Preston (2004)</td>
<td>$\beta(s_{i+1})(\alpha d_i + \delta d_i^2)\exp(\gamma \text{ecat}_i + \tau \text{ecat}_i \log(t_i/25))$</td>
</tr>
</tbody>
</table>
A specific statistical challenge: to deal with weakly informative data

FRENCH COHORT OF URANIUM MINERS
5086 URANIUM MINERS (31/12/2014)

Mean follow-up: 39 years
Mean duration of exposure to radon: 13 years
268 deaths by lung cancer
30 deaths by kidney cancer
Alive: 2580 miners (50.7%)
Mean cumulative total absorbed lung dose among exposed miners (post-55 cohort): 133.9 mGy

FRENCH CT COHORT
100560 CHILDREN (31/12/2016)

Mean age at entry in the cohort (1st scanner): 3.4 years
Mean follow-up: 9.5 years
Mean cumulative brain dose: 24 mGy
Mean cumulative red bone marrow dose: 9 mGy
75 central nervous system tumors
39 leukaemia
41 lymphoma

Approximated statistical power at level = 0.05 of the following hypothesis test:

\[ H_0: \text{exp}(\beta) = 1 \text{ vs } H_1: \text{exp}(\beta) \neq 0 \]

where \( \text{exp}(\beta) \) is the hazard ratio (for 1 mSv) of death by radiation-induced solid cancer from a cohort of nuclear workers (Cox model)

<table>
<thead>
<tr>
<th>( \text{exp}(\beta) )</th>
<th>1.0001</th>
<th>1.0005</th>
<th>1.0009</th>
<th>1.001</th>
<th>1.003</th>
<th>1.005</th>
<th>1.007</th>
<th>1.009</th>
<th>1.015</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_{2003} )</td>
<td>8%</td>
<td>8%</td>
<td>15%</td>
<td>15%</td>
<td>56%</td>
<td>89%</td>
<td>99%</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>( P_{2014} )</td>
<td>5.85%</td>
<td>10.22%</td>
<td>21.76%</td>
<td>23.2%</td>
<td>97%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>
- **Aim**

  - **Promote the use of hierarchical (also called multilevel models) modeling and Bayesian statistical methods** when estimating radiation-related risks at low doses

  - **Why hierarchical modeling?**
    - Flexible modelling approach to describe and simultaneously account for several and heterogeneous sources of uncertainty
    - Benefit of "borrowing strength" in the inference of multiple groups of data

  - **Why Bayesian statistics?**
    - Allows for the joint inference of all unknown quantities (e.g., true exposure/dose and risk parameters) when fitting complex models like hierarchical models
    - Allows to integrate external information through the specification of informative priors or transfer/sequential learning
    - Credible intervals (i.e., for risk estimates) are easily obtained as by-product of Bayesian inference (without asymptotic assumption!)
    - ...
What is a hierarchical (probabilistic) model?

- Main idea: **Think conditionally** to build complexity!

- Combination of conditionally independent submodels
- Each sub-model describes one source of uncertainty
- Many latent layers can be combined
Case study 1: 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 annual radiological exposures collected over the entire career, the French cohort of uranium miners is a reference population to study the long-term health effects of chronic low-dose exposure to radon (Inhalation exposure) and define radon exposure thresholds.

Obtain a measurement corrected estimation of lung cancer mortality risk as well as its associated uncertainty.
Building a Bayesian hierarchical model

Work in collaboration with Julie Fendler (IRSN), Chantal Guihenneuc (Univ. Paris Cité), Sabine Hoffmann (Univ. Ludwig Maximilians)

- Two (or three) conditionally independent submodels [Richardson & Gilks (1993)]
  - **Disease submodel**: it describes the relation between the "true" unknown exposures/doses and the disease outcome
  - **Measurement submodel**: it describes the relation between the observed and the "true" unknown exposures
  - **Exposure submodel**: it describes the probability distribution of the "true" exposures

- **Specific context**:
  - Heteroscedastic measurement error components
  - Time-varying exposure covariates
  - Right-censored survival data (outcome variable)
  - Weak signal in the data (low dose/exposure and low radiation-related risks)

⇒ New models are required
The disease submodel $M_0$ (1/2)

Let’s consider one event of interest (e.g., death by lung cancer)

- Disease outcomes: $(Y_i, \delta_i)$ with $Y_i = \min(T_i, C_i)$, $T_i$ the age at the time of event for individual $i = \{1, \ldots, N\}$, $C_i$ the age at censorship and $\delta_i$ the non-censoring indicator

- Modelling the hazard rate of event for individual $i$ at time $t \in [0, +\infty[$

$$h_i(t; \beta, \theta) = h_0(t; \theta)\rho(\beta; X_i^{\text{cum}}(t))$$

- $X_i^{\text{cum}}(t)$: 5-year lagged cumulative exposure to radon of individual $i$ at time $t$
- $h_0(t; \theta)$: Baseline hazard rate at time $t$ (i.e., for any unexposed individual)
- $\rho(\beta; X_i^{\text{cum}}(t))$: Radiation-related hazard ratio (HR)
- $\beta$: Unknown risk coefficient

Assumption: Non-informative censoring

Contribution to the likelihood of individual $i$ for the disease submodel

$$\left[ (y_i, \delta_i) | \beta, \lambda \right] \propto h_i(y_i; \beta, \theta)^{\delta_i} S_i(y_i; \beta, \theta) \text{ where } S_i(y_i; \beta, \theta) = \exp \left( - \int_0^{+\infty} h_i(u; \beta, \theta) du \right)$$
The disease submodel $\mathcal{M}_0$ (2/2)

- Modeling the baseline hazard function:
  - $h_0(t; \lambda) = \sum_{k=1}^{K} \lambda_k 1_{t \in I_k}$ with $\lambda_k > 0$
  - $h_0(t; \alpha, \xi) = \xi t^{\alpha-1}$ with $\xi > 0$ (scale parameter) and $\alpha > 0$ (shape parameter)

- Modeling the hazard ratio function:
  - $\rho(\beta; X_{i}^{\text{cum}}(t)) = \exp(\beta X_{i}^{\text{cum}}(t)) \Rightarrow$ Cox Model
  - $\rho(\beta; X_{i}^{\text{cum}}(t)) = 1 + \beta X_{i}^{\text{cum}}(t) \Rightarrow$ Excess Hazard Ratio (EHR) model
  - Constraint: $\beta > -\frac{1}{X_{i}^{\text{cum}}(t)}$
Estimation of annual radon exposure in the French cohort of uranium miners

- De 1947 à 1950: 10
- 1951: 10
- 1952: 10
- 1953: 10
- 1954: 10
- 1955: 10

Reconstructed
Ambient measurements
Individual dosimetry

Annual mean exposure to Radon (WLM)

Year

1960 1980 2000
Measurement submodel $M_1$

For an individual $i$ working in mine $m$ at time $t$:

Berkson error components only

\[
\begin{aligned}
X_{im}^1(t) &= Z_m^1(t) \cdot T_{im}(t) \cdot U_i^1(t) & \text{period 1 : 1945-1955} \\
X_{im}^2(t) &= Z_m^2(t) \cdot T_{im}(t) \cdot U_i^2(t) & \text{period 2 : 1956-1982} \\
X_{im}^3(t) &= Z_{im}^3(t) & \text{Hoffmann et al., 2017} \\
\end{aligned}
\]

with $Z_m^1(t) \perp U_i^1(t) \ \forall i, m, t$ and $U_i^k = (U_i^k(t_1), \ldots, U_i^k(t_{ik}))^T \sim \mathcal{L}\mathcal{N}(\frac{-\sigma_k^2}{2} 1_{t_{ik}}, \sigma_k^2 \Gamma_{t_{ik}})$

\[
E[U_i^k] = 1_{t_{ik}} \ \forall k \in \{1, 2\}
\]

\[
1_{t_{ik}} = (1,\ldots,1)^T \text{ et } \Gamma_{t_{ik}} = \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho & \cdots & \rho & 1 \end{bmatrix}, \rho \in [0;1[.
\]

- $U_i^k$ : Shared Berkson error (individual worker practices)
- Fixed magnitudes of Berkson error : $\sigma_1 = 0.93$, $\sigma_2 = 0.39$ [Allodji et al., 2012]
Measurement submodel $\mathcal{M}_2$ : A mixture of Berkson and classical error for period 1

$$
\begin{align*}
&\bullet Z_m^1(t) = \zeta^1_m(t) \cdot U_m(t) \\
&\quad \text{if } Z_m^1(t) \text{ is known, period } 1: 1945-1955 \\
&\bullet X_{im}^1(t) = \zeta^1_m(t) \cdot T_{im}(t) \cdot U_i^1(t) \\
&\quad \text{true mean exposure} \\
&\bullet X_{im}^1(t) = Z_m^1(t) \cdot T_{im}(t) \cdot U_i^1(t) \\
&\quad \text{Berkson error} \\
&\bullet X_{im}^2(t) = Z_m^2(t) \cdot T_{im}(t) \cdot U_i^2(t) \\
&\quad \text{period } 2: 1956-1982 \\
&\bullet X_{im}^3(t) = Z_m^3(t) \\
&\quad \text{period } 3: \text{post } 1983
\end{align*}
$$

- $U_m(t) \sim i.i.d. \mathcal{LN}(\frac{-\sigma^2_2}{2}, \sigma_2^2)$ $\forall t$
- Fixed magnitudes of errors : $\sigma_* = 0.41$ and $\sigma_1 = 0.84$ [Allodji et al., 2012]

Exposure submodel

$$
\zeta^1_m(t) \sim i.i.d. \mathcal{LN}(\mu_\zeta, \sigma_\zeta^2)
$$
Directed Acyclic Graph
Prior distributions and fixed parameters

- $[\beta]$: $\beta \sim \mathcal{N}(0, 10^6)$ left-sided truncated at 0 to guarantee $h_i > 0$
- $[\alpha]$: $\alpha \sim \mathcal{G}(0.01, 0.01)$
- $[\xi]$: $\xi \sim \mathcal{G}(1, 1)$
- $[\lambda]$: $\lambda_j \sim \mathcal{G}(\alpha_{0j}, \lambda_{0j})$ for each component $j$, of $\lambda$, $j = 1, \ldots, 4$ based on the lung cancer mortality in the general French male population between 1968 and 2005 or $\lambda_j \sim \text{Unif}(0, 1)\forall j$
- $[\mu_\zeta]$: $\mu_\zeta \sim \mathcal{N}(0, 100)$
- $[\tau_\zeta] = \left[\frac{1}{\sigma_\zeta^2}\right]$: $\tau_\zeta \sim \mathcal{G}(0.001, 0.001)$

+ Prior sensitivity analysis for $\alpha$ and $\xi$

- No validation sample to estimate the expected value of true exposure given observed/estimated exposure or the true magnitude/variance of the error components $\Rightarrow \sigma_1, \sigma_2, \sigma_\star$ [Allodji et al., 2012] and $\rho$ must be fixed...

+ Impact of these choices on risk estimates must be evaluated
Bayesian inference

- Complex joint posterior distribution $\theta = (\beta, \alpha, \xi, \mu_\zeta, \tau_\zeta, \zeta, U)$
- More than 198,000 pseudo-observations
- More than 40,000 unknown quantities to estimate
  $\Rightarrow$ High dimensional posterior distribution

- Adaptive Metropolis-Within-Gibbs algorithm developed in Python 3.4 + cluster HPC
  - (Left-sided truncated) Gaussian random Walk Metropolis-Hastings for $\beta$ and $\alpha$
  - Multiplicative random walk Metropolis-Hastings for $\zeta$, $U^k_i$ and $U_m$
  - Gibbs sampling for $\xi$, $\mu_\zeta$, $\tau_\zeta$
  - Block updating of shared Berkson error component $U_i$ after defining 239 homogeneous groups of miners (hierarchical clustering) based on information on mine location, type of min, job type

- Reparametrizations to improve mixing of the chains (e.g. $\xi$ parameter)
- Targeted acceptance rate: About 40% for single parameters and 20% for vectors
- Running time: 5 days for 2 Markov chains, 10,000 iterations for the adaptive phase + 60,000 iterations including 20,000 iterations for the burn-in phase ($\Rightarrow$ Effective Sample Size >4000)
### Application on the French cohort of uranium miners

**Impact of the correlation parameter \( \rho \) on Bayesian inference**

<table>
<thead>
<tr>
<th>EHR</th>
<th>HR* ( 100_{\text{WLM}} )</th>
<th>IC** 95%</th>
<th>WAIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho = 0 )</td>
<td>2.14</td>
<td>1.65;2.81</td>
<td>6860</td>
</tr>
<tr>
<td>( \rho = 0.2 )</td>
<td>2.16</td>
<td>1.66;2.86</td>
<td>6860</td>
</tr>
<tr>
<td>( \rho = 0.4 )</td>
<td>2.21</td>
<td>1.70;2.94</td>
<td><strong>6858</strong></td>
</tr>
<tr>
<td>( \rho = 0.6 )</td>
<td>2.23</td>
<td>1.68;2.99</td>
<td>6859</td>
</tr>
<tr>
<td>( \rho = 0.8 )</td>
<td>2.17</td>
<td>1.66;2.89</td>
<td>6862</td>
</tr>
<tr>
<td>( \rho = 0.99 )</td>
<td>2.13</td>
<td>1.63;2.80</td>
<td>6863</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cox</th>
<th>HR* ( 100_{\text{WLM}} )</th>
<th>IC** 95%</th>
<th>WAIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho = 0 )</td>
<td>1.32</td>
<td>1.18;1.48</td>
<td>6971</td>
</tr>
<tr>
<td>( \rho = 0.2 )</td>
<td>1.33</td>
<td>1.18;1.50</td>
<td>6870</td>
</tr>
<tr>
<td>( \rho = 0.4 )</td>
<td>1.33</td>
<td>1.18;1.52</td>
<td><strong>6870</strong></td>
</tr>
<tr>
<td>( \rho = 0.6 )</td>
<td>1.35</td>
<td>1.18;1.54</td>
<td>6872</td>
</tr>
<tr>
<td>( \rho = 0.8 )</td>
<td>1.35</td>
<td>1.20;1.55</td>
<td>6875</td>
</tr>
<tr>
<td>( \rho = 0.99 )</td>
<td>1.19</td>
<td>1.08;1.35</td>
<td>6886</td>
</tr>
</tbody>
</table>

Results provided by the disease submodel combined with the \( M_1 \) measurement error submodel

*HR : *Posterior median* of the hazard ratio for 100 Working Level Months of death by lung cancer (i.e., \( 1 + \beta \times 100 \)), **IC : 95% credible interval
Uncorrected and measurement corrected estimation of lung cancer mortality risk and associated uncertainty

<table>
<thead>
<tr>
<th></th>
<th>HR* 100 WLM</th>
<th>IC** 95%</th>
<th>WAIC***</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline model (β = 0)</td>
<td></td>
<td></td>
<td>6895</td>
</tr>
<tr>
<td>$\mathcal{M}_0$</td>
<td>2.06</td>
<td>1.60;2.70</td>
<td>6861</td>
</tr>
<tr>
<td>$\mathcal{M}_1$</td>
<td>2.21</td>
<td>1.70;2.94</td>
<td>6858</td>
</tr>
<tr>
<td>$\mathcal{M}_2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_1 = 0.84$; $\sigma_* = 0.41$</td>
<td>2.38</td>
<td>1.78;3.26</td>
<td>6855</td>
</tr>
<tr>
<td>$\sigma_1 = 0.84$; $\sigma_* = 0.82$</td>
<td>2.50</td>
<td>1.81;3.46</td>
<td>6854</td>
</tr>
<tr>
<td>$\sigma_1 = 0.63$; $\sigma_* = 0.31$</td>
<td>2.32</td>
<td>1.73;3.13</td>
<td>6857</td>
</tr>
</tbody>
</table>

Results provided by the EHR disease submodel $\mathcal{M}_0$ (i.e., without accounting for exposure measurement error) and the measurement submodels $\mathcal{M}_1(\rho = 0.4)$ and $\mathcal{M}_2(\rho = 0.4)$ combined with $\mathcal{M}_0$

Some posterior probability distributions:
Impact of exposure measurement error on the instantaneous excess risk $\beta$ ($\rho = 0.4, \sigma_1 = 0.84, \sigma_* = 0.41$)

Posterior density of the excess risk coefficient $\beta$ (per 100 WLM) of death by lung cancer in the French cohort of uranium miners

Posterior density of the excess risk coefficient $\beta$ (per 100 WLM) of death by lung cancer in the French cohort of uranium miners
Application on the French cohort of uranium miners
Prior sensitivity

Influence of the prior density assigned on the baseline risk $\lambda$ on the posterior density of the risk coefficient $\beta$ in the French cohort of uranium miners.

Influence of the prior density (dotted lines) on the posterior density (solid line) when estimating the baseline risk of leukaemia between 55 and 65 years in the French cohort of uranium miners.

Lung cancer

Leukaemia
Adaptive Metropolis-Within-Gibbs algorithm are time-consuming to explore high-dimensional posterior distributions ⇒ Which alternative bayesian learning algorithm?
  ▶ Work under progress to implement a Metropolis-adjusted Langevin sampler and compare it to our current adaptive Metropolis-Hastings sampler ⇒ First promising results with about 40% reduction in calculation time for an equivalent ESS when updating unknown parameters $\alpha$ and $\beta$.

Robustness of our models to measurement and/or exposure model mispecification? ⇒ Simulation studies under progress...