RadoNorm
Managing risks from radon and NORM

Deliverable 4.5

Final report on consideration of different sources of uncertainty in radon induced lung cancer risk inference

Work Package 4

This project has received funding from the Euratom research and training programme 2019-2020 under grant agreement No 900009.

Document information

Project Acronym	**RadoNorm**
Project Title	**Towards effective radiation protection based on improved scientific evidence and social considerations - focus on radon and NORM**
Project Type	**RIA**
EC grant agreement No.	**900009**
Project starting / end date	**1st September 2020 – 31 August 2025**
Work Package No.	**4**
Work Package Title	**Effects & Risks**
Deliverable No.	**4.5**
Deliverable Title	**Final report on consideration of different sources of uncertainty in radon induced lung cancer risk inference**
Lead Beneficiary	**ASNR**
Contractual Delivery Date	**31st August 2025**
Actual Delivery Date	**30th August 2025**
Type	**Report**
Dissemination level	**Public**
Authors	**Sophie Ancelet (ASNR), Sara Antignani (ISS), Francesco Bochicchio (ISS), Enora Clero (ASNR), Chantal Guihenneuc (UPC), Kossi Gnameho (UPC), Raja Ben Hajria (ASNR), Felix Heinzl (BfS), Sabine Hoffmann (ULM), Jan Christian Kaiser (BfS), Niccolò Loret (ISS), Estelle Rage (ASNR), David Richardson (University of California), Manuel Sommer (BfS), Ladislav Tomášek (SURO), Stefania Trentin (BfS)**

To be cited as:

Ancelet S., Antignani S., Bochicchio F., Clero E., Guihenneuc C., Gnameho K., Hajria R.B., Heinzl, F., Hoffmann S., Kaiser J.C., Loret N., Rage E., Richardson D., Sommer M., Tomášek L., Trentin S. (2025): Final report on consideration of different sources of uncertainty in radon induced lung cancer risk inference. Final version as of 31.08.2025 of deliverable D4.5 of project RadoNorm.

Disclaimer

Acknowledgement

This document is a deliverable of the RadoNorm project. This project has received funding from the Euratom research and training programme 2019-2020 under grant agreement No 900009.

Status of deliverable		
	By	Date
Delivered (Lead Beneficiary)	ASNR	29.08.2025
Verified (WP Leader)	UEF	30.08.2025
Reviewed (Reviewers)	WPL, PC	31.08.2025
Approved (PC)	PC	31.08.2025
Submitted to EC (PC)	PC	31.08.2025

Executive Summary

Task 4.6 of Work Package 4 is divided into three independent sub-tasks.

Subtask 4.6.1 aims to refine the estimation of the risk of death from lung cancer due to chronic exposure to radon, by considering both the French and the Czech cohort of uranium miners and accounting for complex patterns of radon exposure measurement error. An essential preliminary step was to develop models to account for exposure measurement error when estimating the risk of death from radon-related lung cancer in the Czech cohort of uranium miners. Twelve different hierarchical exposure-risk models, based on survival disease submodels, were proposed, fitted and compared from the last update of the Czech cohort of uranium miners. Monte-Carlo Markov Chain (MCMC) algorithms were implemented to fit these models. We proposed joint probabilistic models - with and without consideration of exposure measurement error - to estimate the risk of death from radon-related lung cancer in a multi-cohort of uranium miners - namely the Czech and the French cohorts - in order to boost the statistical power to estimate this risk. These joint models are based on the combination of the hierarchical models previously defined for each cohort. Interestingly, they assume a common risk parameter between both cohorts and cohort-specific parameters to define the baseline hazard risk and the radon exposure measurement error. An MCMC algorithm was implemented to fit these joint models without accounting for measurement error. We also performed simulations to study the impact of pooling cohorts on the estimated risk parameter. Our first results show that multi-cohort Bayesian modelling allows to refine our risk estimation in comparison to cohort per cohort estimation. These results still need to be confirmed by further analyses. An MCMC algorithm is still under development to account for exposure measurement error.

Subtask 4.6.2 aims to develop biologically-based models on lung carcinogenesis to improve causality assessment risk quantification for lung cancer caused by exposure to radon. Biologically-based models (BBRM) have been applied to study a cohort of 4309 male Sprague-Dawley rats exposed to radon at the Commissariat à l'Energie in France. Survival analysis shows an increase in the lung cancer risk with both increasing exposure and exposure rate. A similar analysis on the incidence of precancerous lesions also shows a significant increase with exposure and exposure rate. Linear, respectively ordinal, regression was performed on the size distribution of premalignant lesions, respectively on size class distribution of lung cancers. The linear regression evidence larger preneoplastic lesions in two of the higher exposed groups in comparison to the unexposed rats, while ordinal regression showed no significant difference in the cancer size classes distribution among exposure groups. To study the subset of the CEA dataset containing information on the size of precancerous lesions we have applied a hyperplasia growth model. This enables us to properly estimate biological parameters such as premalignant cells duplication rates. We further extended the hyperplasia growth model to lung cancer, so that we could integrate the information on lung cancer size. Last, on the subset of the dataset which contains information on both hyperplasia and lung cancer size we tested a model integrating both these aspects and compared it with a model using only lung cancer size on this subset. The combined model resulted more numerically robust, with almost equal goodness of fit as the tumour only model, but without the risk of overfitting due to the low number of lung cancer cases in this subset. Overall, the integration of subclinical information such as preneoplastic lesions or size of lung cancers resulted in better numerical stability and more robust fits, together with the possibility to estimate biological parameters otherwise not identifiable.

Subtask 4.6.3 aims to estimate the lifetime lung cancer risk associated with radon exposure, based on different populations, risk models, and exposure scenarios. The population groups considered include national data from Italy, Germany, and France, as well as data from a composite Euro-American population used in ICRP Publication 103. This approach allows for the consideration of variations in population characteristics that may affect radon-related risk. The risk models used in the analysis were

derived from both occupational miner studies (i.e., BEIR VI and PUMA) and residential studies (i.e., the European pooled analysis by Darby et al., 2006). The exposure scenarios allowed for the evaluation of constant lifetime exposure, increasing exposure over time, or decreasing exposure due, for example, to the implementation of mitigation measures in high-radon environments. The metric used to quantify the lifetime risk of dying from lung cancer caused by radon exposure is the Lifetime Excess Absolute Risk (LEAR). LEAR calculations were performed separately for males and females. Smoking data were also incorporated into the analysis to assess their modifying effect on radon-related lung cancer risk. LEAR estimates were computed separately for current smokers, ex-smokers, and never-smokers. Two types of radon–smoking interaction models were considered: multiplicative and sub-multiplicative. As general conclusions, LEAR values for France, Italy, and Germany are broadly similar for women across all models and exposure scenarios. For men, however, some differences emerge, likely reflecting variations in smoking habits and baseline lung cancer mortality. Among the models, BEIR VI and PUMA consistently yield higher LEAR estimates than the Darby model, which is expected given their derivation from miner cohorts. Regarding the impact of exposure scenarios, the results clearly show that both the timing of exposure and the choice of risk model significantly influence the estimated lifetime lung cancer risk. The assumed type of radon–smoking interaction appears to have a notable impact only in the case of the BEIR VI model.

Table of content

List of Figures

List of Tables

1. Subtask 4.6.1: A Bayesian hierarchical approach to account for radon exposure measurement error from a multi-cohort of uranium miners

1.1 Introduction

When estimating the risk of death from lung cancer due to radon exposure, one major source of uncertainty relates to radon exposure assessment, especially when no individual monitoring data are available. In occupational cohorts of uranium miners, for instance, changes in the methods of radon exposure assessment can lead to complex patterns of exposure measurement error (Tirmarche et al. 2009, Allodji *et al. 2012*, Hoffmann 2017). This is particularly the case when using a joint cohort that may be affected by multiple and heterogeneous exposure measurement error depending on both the cohort and the calendar period. Moreover, the use of a strategy of group-exposure assessment, for instance via job-exposure matrices or retrospective exposure reconstructions, as well as individual worker characteristics may lead to error components that are shared between or within workers, respectively (Hoffmann et al. 2018).

The uncertainties on radon exposure assessment are often neglected in practice when estimating lung cancer risk. Usually, only point estimates of radon exposures are plugged into exposure-risk models to estimate this risk without accounting for the related uncertainties (Kreuzer *et al.* 2008, Tomasek 2012, Rage *et al.* 2015, Kelly-Reif *et al*. 2023). This may lead to biased risk estimates, a distortion of the exposure-risk relationship and an under/overestimation of the uncertainty of risk estimates (Carroll 2005, Hoffmann 2017, Keogh *et al.* 2020) which may be problematic for scientific bodies and decision makers in charge of defining radiation protection standards.

IRSN and University Paris-Cité (UPC) have recently proposed Bayesian hierarchical models to account for complex patterns of measurement error in radon exposure in the analysis of lung cancer mortality in the French cohort of uranium miners (Hoffmann *et al.* 2017, Fendler *et al.* 2023). A simulation study based on exposure data of the French cohort of uranium miners has also been conducted to compare the effects of shared and unshared exposure uncertainty on risk estimation and on the shape of the exposure-response curve in survival models (Hoffmann et al. 2018). The current findings underline the importance of both a careful characterisation of all components of exposure measurement error in epidemiological studies and the realistic modelling of the identified error structures, in particular in cases where there are exposure periods that were characterized by shared exposure uncertainty.

More research is now needed to promote the use of the Bayesian hierarchical approach to account for multiple and heterogeneous sources of uncertainty in radiation epidemiology. Particularly, the Bayesian hierarchical approach must be further explored to estimate the corrected risk of death by lung cancer from multiple cohorts of uranium miners, simultaneously. Indeed, fitting a Bayesian hierarchical model to a joint cohort will allow benefiting from a larger statistical power to estimate the risk of death from radon-related lung cancer. Nevertheless, the main technical difficulties will be to account for heterogeneous types of exposure measurement error depending on the cohort and the calendar period and to be able to fit a hierarchical model to a large database within a reasonable time.

The aim of sub-task 4.6.1 is to refine the estimation of the risk of death from lung cancer due to chronic exposure to radon by considering the Czech and French cohort of uranium miners simultaneously and accounting for the radon exposure measurement error in both cohorts. An essential preliminary step (detailed in the following sub-section 1.3) was to develop joint probabilistic models - with and without consideration of exposure measurement error – to estimate the risk of death from radon-related lung cancer in the joint cohort of Czech and French uranium miners in order to boost the statistical power to estimate this risk. These joint models are based on the combination of a Bayesian hierarchical model previously proposed for the French cohort (Hoffmann *et al.* 2017, Fendler *et al.* 2023) and a Bayesian hierarchical model developed for the Czech cohort in this sub-task 4.6.1 of RadoNorm. One

methodological challenge was to implement an efficient MCMC algorithm to fit a complex joint hierarchical model under the Bayesian paradigm in order to estimate a joint and corrected risk of death from lung cancer from both cohorts of uranium miners.

The hierarchical approach provides a flexible modelling framework to describe complex exposure measurement error (Richardson and Gilks 1993). In this context, a hierarchical model combines two main conditionally independent probabilistic sub-models:

- a disease sub-model: it describes the probability of occurrence of a response variable (e.g., age at death from lung cancer) as a function of 'true' radon exposure.
- a measurement sub-model: it describes the relationship between 'true' exposures to radon and measured or estimated radon exposures, that are prone to uncertainty.

The Bayesian paradigm allows us to infer these often complex and high-dimensional models within a relevant framework (Gelman *et al.* 2014). This choice also makes it possible to account for prior knowledge about unknown parameters of the proposed model. This may facilitate their estimation in situations where the available data are poorly informative, which may be the case when we are interested in radon-related cancer risks.

1.2 The French and Czech cohort of uranium miners

The French and Czech cohorts of uranium miners are retrospective cohorts whose characteristics, sources of data and methods of data collection have been described previously (Laurier et al. 2004, Tomasek 2012, Rage et al. 2015, Fendler et al. 2024).

In this work, the French cohort of uranium miners includes 5,086 males who were employed as uranium miners for at least one year in the Atomic Energy Commission (CEA) and the General Company of Nuclear Fuel (COGEMA, which became AREVA in 2006) company between 1946 and 1990. The Czech cohort includes 9,978 men who worked as underground miners. It can be seen as two historical sub-cohorts named S and N. Miners in sub-cohort S worked mainly in the Jáchymov mines. Those in sub-cohort N worked main in the Příbram mines. The uranium miners in the S sub-cohort began working between 1948 and 1959. The inclusion criterion for this sub-cohort was to have worked for at least 4 years as a uranium miner in the Jáchymov mines. The uranium miners in the N sub-cohort began working between 1968 and 1974. The inclusion criterion for this sub-cohort was to have worked for at least one year as a uranium miner in the Příbram mines.

The inclusion period of uranium miners extends from 1946 to 1990 for the French cohort and from 1948 to 1974 for the Czech cohort. The follow-up began at the date of first employment (defined as one year before the cohort entry date except for the sub-cohort S) and ended at the earliest date among the following date of death, date of loss to follow-up and December 31st, 2014 for French miners and December 31st, 2012 for the Czech miners.

The main characteristics of the French and Czech cohorts of uranium miners data used in this work are summarized in Table 1. Compared with the French cohort of uranium miners, the average duration of employment and exposure to radon are approximately 0.4 and 0.6 times shorter in the Czech cohort, while the average cumulative exposure to radon is approximately twice as high. As for the percentage of deaths from lung cancer, it is slightly more than twice as high in the Czech cohort (11.8%) compared to the French cohort (5.3%).

Table 1. Main characteristics of the French cohort of uranium miners (1946–2014) and the Czech cohort of uranium miners (1946-2012).

	French cohort	**Czech cohort**
No. of miners, n	5086	9978
Vital status, n (%)		
Alive	2580 (50.7)	4016 (40.2)
Deceased	2464 (48.5)	5559 (55.7)
Deceased (lung cancer)	268 (5.3)	1176 (11.8)
Lost to follow-up	42 (0.8)	403 (4.1)
Age and duration in years, mean (range)		
Age at entry into study	28.8 (16.0-68.4)	27.7 (13.6-63.6)
Duration of follow-up	39.0 (0.1-67.1)	35.0 (0.94-66.5)
Duration of employment	17.4 (1.0-62.9)	7.8 (0.2-39)
Radon exposure		
Cumulative exposure in WLM, mean (range)	36.6 (0.003-960.1)	70.4 (0-869.8)
Duration of exposure in year, mean (range)	13.2 (1.0-41.0)	8.3 (0.3-39)
Age at first exposure	29.6 (14.2-63.0)	27.7 (13.6-63.6)

1.3 Accounting for radon measurement error in the joint cohort of French and Czech uranium miners.

We develop Bayesian models to quantify/estimate the risk of death from lung cancer potentially associated with chronic exposure to radon in a multi-cohort of uranium miners: the Czech and the French cohorts in order to boost the statistical power of estimating this risk. Based on a parametric survival model, we develop a Bayesian modelling approach assuming a common association between lung cancer mortality and radon exposure but with specific parameters (baseline hazard risk) for each country. In other words, each cohort has its own specific disease parameters which reflect the level of instantaneous risk function locally country per country. Inference by adaptive MCMC algorithms are implemented in high-dimensional setting.

1.3.1 The disease sub-model

Following Kleinbaum and Klein (2010) and Thiébaut and Bénichou (2004), we use the age as a time scale. Therefore, as the monitoring of each miner i begins at the time of its inclusion in the cohort, the random variable T_i is truncated on the left to the age (in days) of the miner at the time of his entry into the study. We denote this variable by r_i^0 subsequently. Given the instantaneous hazard risk function h_i,

the truncated probability density in r_i^0 and the actualized truncated survival function in r_i^0 of the miner i are respectively given by

$$\tilde{f}_i(t) = \frac{f_i(t)}{S_i(r_i^0)} \text{ and } \tilde{S}_i(t) = \frac{S_i(t)}{S_i(r_i^0)} = \exp(-\int_{r_i^0}^{t} h_i(s)ds)$$

where S_i denotes the survival function.

In order to include the exposure-risk dimension in the modelling, it is necessary to define survival models that could link survival data and the annual cumulative exposure to radon of each miner. Following Rage et al. 2015, we considered a lag of 5 years between a date of exposure to radon and a date of death from lung cancer which could be attributed to this exposure. For the sake of simplicity in the notations, we will note $X_i^{cum}(t)$ the cumulative exposure to radon (in WLM) to which miner i was exposed at age $t - 5365.25$ (in days).

As previously indicated, data were combined from two cohorts where we specify a single and common parameter β for the two cohorts but each cohort has its own specific disease parameters. In the pool of the two cohorts, for a single individual indexed by i, we define the instantaneous hazard risk function as:

$$h_i^{\chi}(t, \theta_{\chi}, \beta) = h_0^{\chi}(t, \theta_{\chi})\rho(\beta, X_i^{cum}(t)), t \geq 0 \text{ where:}$$

- β is the common shared unknown radiation-related risk coefficient
- the index $\chi \in \{C,F\}$ and C (resp. F) is linked to Czech (resp. French) country
- $X_i^{cum}(t)$: 5-year lagged cumulative exposure to radon of miner i at age t in the total cohort
- θ_{χ} a list of intra-cohort parameters for cohort χ.
- h_0^{χ} the baseline hazard specific to cohort χ.
- ρ is the radiation-related hazard ratio

Based on the independence assumption among individuals, the general likelihood resulting from the combined cohorts is given by

$$\mathcal{L}(y_{1:N}, \delta_{1:N} | \theta_C, \theta_F, \beta, X^{cum}) = \prod_{\chi \in \{C,F\}} \prod_{i=1}^{n_{\chi}} \left(h_0(y_i, \theta_{\chi}).\rho(\beta, X_i^{cum}(y_i))\right)^{\delta_i} . exp\left(-\int_{r_i^0}^{y_i} h_0(u, \theta_{\chi}).\rho(\beta, X_i^{cum}(u))du\right)$$

where $N = \sum_{\chi \in \{C,F\}} n_{\chi}$ and n_{χ} is the number of individuals in the cohort $\chi \in \{C,F\}$.

The cumulative radon exposure covariate depends on time. In this form, the integral in the previous equation is not tractable and therefore need to be approached numerically. In practice, the tracking time of each miner i defined by the time interval $[r_i^0; y_i]$ is divided into many intervals $r_i^j; r_i^{j+1}$ on which the cumulative exposure to radon of miner i is constant. The integral of the previous equation becomes then tractable. Concerning the baseline hazard risk, a piece-wise constant function is assumed over K age intervals denoted I_k with $k \in 1,...,K$:

$$h_0(t, \lambda_{1:K}^{\chi}) = \sum_{k=0}^{n_i-1} \lambda_k^{\chi}. 1_{\{I_k\}}(t), \quad \text{where} \quad (t, \lambda_k^{\chi}) \in \mathbb{R}^+ \times \mathbb{R}^+ \text{ and } \chi \in \{C,F\}.$$

$\lambda_{1:K}^{\chi}$ designates the set of unknown parameters related to the baseline instantaneous risk in the cohort χ. Depending on the work carried out in the French cohort of uranium miners (Clayton & Hills 2013), four

intervals are considered where baseline risk is assumed to be constant. These intervals of ages are [0,40); [40,55); [55,70) and [70,∞) years, used in the Czech and in the French cohorts.

Finally, ρ corresponds to the class of the survival model in excess of instantaneous risk (denoted EHR for Excess Hazard Ratio subsequently). It assumes the following form

$$\rho\left(\beta, X_i^{cum}(t)\right) = 1 + \beta X_i^{cum}(t)$$

Choice of prior distributions

As the instantaneous risk functions has to be positive for all $t \in [0,\infty)$, we impose the constraint $\beta \geq 0$.

Table 2 indicates the list of prior distributions considered for the parameters of interest. The choice of the prior distribution for the risk coefficient β is a non-informative normal distribution centred at 0, with a standard deviation of 1000 and truncated to the left at 0. For the positive parameters λ_k^{χ} ($k = 1, 2, 3, 4$; $\chi = C, F$) defining the piece-wise function h_0, two types of proposal distributions could be implemented: a flat distribution uniform on $[0,1]$ and an informative distribution gamma of parameters (1;1) useful when data are sparse.

Table 2. Prior distributions.

Country	Parameters	Prior
F - C	β	$\log N(0,1000^2)1_{\beta>0}$
C	λ_1^C	$\Gamma(1,1)$
C	λ_2^C	$\Gamma(1,1)$
C	λ_3^C	$U(0,1)$
C	λ_4^C	$U(0,1)$
F	λ_1^F	$\Gamma(1,1)$
F	λ_2^F	$\Gamma(1,1)$
F	λ_3^F	$U(0,1)$
F	λ_4^F	$U(0,1)$

1.3.2 The measurement error model

The measurement sub-model includes many parameters to be estimated with 15704 error components of shared measurements U_i^j for period j and miner i in Czech cohort only. This presents serious computational challenges in term of CPU time. Adding the French cohort of data rises the problem of limited computational resources. The estimated models therefore do not include measurement errors. Concerning simulation studies, one scenario includes simulated data sets with measurement errors. For this simulation, we choose measurement models specific for each country as follow:

For the Czech cohort,

$$M_1^C = \begin{cases} X_i(t) = Z_i(t).U_i^1 & \forall\, t \in [1946\,,1952] \\ X_i(t) = Z_i(t).U_i^2 & \forall\, t \in [1953\,,1959] \\ X_i(t) = Z_i(t).U_i^3 & \forall\, t \in [1960\,,1967] \\ X_i(t) = Z_i(t).U_i^4 & \forall\, t \in [1968\,,1999] \end{cases}$$

For the French cohort,

$$M_1^F = \begin{cases} X_i(t) = Z_i(t).U_i^1 & \forall\, t \in [1945\,,1955] \\ X_i(t) = Z_i(t).U_i^2 & \forall\, t \in [1956\,,1982] \\ X_i(t) = Z_i(t) & \forall\, t \in [1983\,,2007] \end{cases}$$

Where $X_i(t)$ is the "true" exposure and $Z_i(t)$ the estimated exposure of miner i at time t. The distribution of errors terms U_i^j are assumed to be LogNormal with fixed variances (Allodji et al. 2012).

1.3.3 Bayesian inference

Within Bayesian frameworks, MCMC algorithms enable us to simulate Markovian samples of the posterior distribution associated with proposed models. Diagnostics of convergence and mixing are carried out to ensure the quality of the posterior estimates. The posterior mean, median and the 95% credibility intervals denoted CI-95% are obtained for the risk coefficient β and the parameters of h_0. All parameters were updated with an adaptive MH sampler. The MCMC algorithms were based on several simulated Markov chains of $30{,}000$ iterations including 10,000 iterations for the adaptive phase, 5000 iterations for the burn-in and 15,000 iterations for the sample used to estimate the posterior.

1.3.4 Results

Results on real data set

The following results are based on the pooled cohort. For the Post-55 French cohort of 3723 miners, the mortality rate from lung cancer is 4.35 %, the cumulative radon from 0 to 960 WLM and the follow up period from 0 to 62 years. For the Czech cohort of 9978 miners, the mortality rate from lung cancer is 11.79 %, the cumulative radon exposures from 0 to 870 WLM and the follow up period from 0.94 to 66.5 years.

The following table 3 shows the results from the French, Czech and pooled cohort with common β in terms of posterior medians and CI-95% obtained for association β (for 100 WLM) and the parameters λ related to the baseline instantaneous risk.

Table 3. Mean of posterior medians and credible intervals.

		F Cohort		C Cohort		Pooled Cohort	
Parameters		Mean	CI-95%	Mean	CI-95%	Mean	CI-95%
β per 100 WLM		0.98	[0.58 ; 1.51]	3.33	[2.83 ; 3.92]	2.69	[2.27 ; 3.07]
λ_1^F	in 10^{-8}	4.66	[3.05 ; 6.72]			4.63	[3.01 ; 6.64]
λ_2^F	in 10^{-6}	1.34	[1.01 ; 1.72]			1.10	[0.84 ; 1.43]
λ_3^F	in 10^{-6}	5.70	[4.78 ; 6.73]			4.56	[3.86 ; 5.36]
λ_4^F	in 10^{-6}	9.40	[7.46 ; 11.75]			6.78	[5.46 ; 8.28]
λ_1^C	in 10^{-8}			8.48	[6.40 ; 11.07]	8.53	[6.30 ; 11.23]
λ_2^C	in 10^{-6}			2.25	[1.97 ; 2.57]	2.51	[2.22 ; 2.85]
λ_3^C	in 10^{-6}			6.04	[5.33 ; 6.82]	6.86	[6.15 ; 7.67]
λ_4^C	in 10^{-6}			7.61	[6.43 ; 8.98]	8.83	[7.58 ; 10.38]

When pooling the two cohorts, precisions of estimates on λ are often improved with narrower credible intervals. The common association β is estimated as a compromise between estimates obtained on each cohort. The value is closer to that obtained on the Czech cohort; this is probably due to the fact that the Czech cohort is more informative because has a higher death rate.

The results on pooled cohort confirm the existence of a statistically positive association between cumulative radon exposure and risk of death from lung cancer. DIC values under the model with null association and under the model with a positive association β for the post55-French cohort, for Czech cohort and for the pooled cohort are in Table 4.

Table 4. DIC values.

	DIC VALUES		
	$\beta \neq 0$	$\beta = 0$	Difference
Post-55 French cohort	4084	4112	28
Czech cohort	27759	28537	778
Pooled cohort	31848	32577	729

Positive difference of DIC values clearly confirms the existence of a positive association β.

Simulation studies

In this part, we perform simulations to study the impact of pooling cohorts on the estimated parameters above. We consider two first cases without measurement errors corresponding to different or equal numbers of individuals per cohort and to different or equal levels of association β per cohort. In a last case, data are simulated with measurement errors but estimated without in order to assess if a pooling approach improves estimates under a misspecified model. For all cases, the true considered values of the disease's parameters lambda (λ_1, λ_2, λ_3, λ_4) are the same for both cohorts and are respectively $8.11\ 10^{-8}$, $1.38\ 10^{-6}$, $5.46\ 10^{-6}$ and $9.16\ 10^{-6}$. As previously, results are presented cohort per cohort and on the pooled cohort.

Simulation 1

In the first simulation, we consider 20 simulated data sets of French and Czech Republic cohort. The parameter β is the same for the two cohorts equal to 0.05 and the number of individuals are equal to 2000 for each cohort. This ideal case with two similar cohorts aims to check algorithms and to assess impact on β when pooling two identical cohorts. The results presented in Table 5 show that the risk β is estimated more precisely as expected with the pooled cohort: All other parameters which are specific for each country are well estimated except λ_1 corresponding to ages with very few deaths.

Table 5. Mean of posterior medians and credible intervals of medians on 20 simulated French, Czech and pooled data sets with β_{FR} = 0.05, β_{CZ} = 0.05, n_{FR} = 2000 and n_{CZ} = 2000.

	F Cohort		C Cohort		Pooled Cohort	
Parameters	Mean	CI-95%	Mean	CI-95%	Mean	CI-95%
β per 100 WLM	4.84	[3.82 ; 5.81]	4.83	[4.33 ; 5.44]	4.80	[4.26 ; 5.31]
λ_1^F in 10^{-8}	4.79	[4.60 ; 5.22]			4.80	[4.61 ; 5.25]
λ_2^F in 10^{-6}	1.32	[1.08 ; 1.53]			1.33	[1.08 ; 1.54]
λ_3^F in 10^{-6}	5.22	[4.81 ; 5.66]			5.23	[4.82 ; 5.51]
λ_4^F in 10^{-6}	9.25	[8.18 ; 10.46]			9.27	[8.31 ; 10.31]
λ_1^C in 10^{-8}			4.86	[4.54 ; 5.46]	4.86	[4.56 ; 5.46]
λ_2^C in 10^{-6}			1.31	[1.13 ; 1.48]	1.32	[1.15 ; 1.47]
λ_3^C in 10^{-6}			5.29	[4.89 ; 5.66]	5.31	[4.70 ; 5.72]
λ_4^C in 10^{-6}			9.47	[8.71 ; 10.52]	9.52	[8.52 ; 10.84]

Simulation 2

In order to mimic the real data, the number of miners is more important in Czech cohort (2000 for French cohort and 5000 for Czech cohort) and the association β is higher in Czech cohort (0.05) than in French cohort (0.02).

Table 6 presents estimates of parameters. As expected, the association is estimated as a compromise between the two cohorts with a value closer to that obtained on Czech cohort due to the more important number of miners.

Table 6. Mean of posterior medians and credible intervals of medians on 50 simulated French, Czech and pooled data sets with β_{FR} = 0.02, β_{CZ} = 0.05, n_{FR} = 2000 and n_{CZ} = 5000.

	FR Cohort		CZ Cohort		Pooled Cohort	
Parameters	Mean	CI-95%	Mean	CI-95%	Mean	CI-95%
β per 100 WLM	2.00	[1.52 ; 2.65]	4.86	[4.29 ; 5.56]	3.91	[3.54 ; 4.47]
λ_1^F in 10^{-8}	4.76	[4.61 ; 5.09]			4.74	[4.60 ; 5.12]
λ_2^F in 10^{-6}	1.32	[1.14 ; 1.52]			1.18	[1.02 ; 1.36]
λ_3^F in 10^{-6}	5.26	[4.91 ; 5.72]			4.50	[4.15 ; 4.88]
λ_4^F in 10^{-6}	9.14	[7.84 ; 10.47]			7.01	[5.94 ; 8.19]
λ_1^C in 10^{-8}			5.02	[4.53 ; 5.66]	5.06	[4.54 ; 5.71]
λ_2^C in 10^{-6}			1.30	[1.16 ; 1.46]	1.46	[1.31 ; 1.65]
λ_3^C in 10^{-6}			5.23	[4.81 ; 5.63]	6.01	[5.47 ; 6.44]
λ_4^C in 10^{-6}			9.49	[8.37 ; 10.56]	11.25	[9.99 ; 12.47]

Table 7 gives coverage rate (expected to be 0.95), RMSE (Root-Mean-Square Error) and relative bias of parameters. Concerning the risk β, there is no true value for the pooled cohort anymore. As a benchmark, these criteria are calculated on a weighted mean of the different associations that is on $\frac{2}{7}0.02 + \frac{5}{7}0.05 \approx 0.041$. For β, results are similar between the pooled cohort and those obtained cohort per cohort with very low bias. For λ, results are similar and sometimes better with pooled cohort except for λ_4 with bad results for pooled cohort (under-estimated for λ_4^f and over-estimated for λ_4^c). An explanation could be that β is estimated on pooled cohort by a higher value compared to the real value in the French cohort and by a lower value compared to the real value in the Czech. Consequently, the baseline is under or over-estimated corresponding to a compensation phenomenon

Table 7. Coverage rate (0.95), RMSE and relative bias on 50 simulated French and Czech datasets with β_{FR} = 0.02, β_{CZ} = 0.05, n_{FR} = 2000 and n_{CZ} = 5000. (: for pooled cohort, compared with β = 2/7 0.02 + 5/7 0.05 ≈ 0.041).*

	FR Cohort			CZ Cohort			Pooled Cohort		
	Cov. rate	RMSE	Relative Bias	Cov. rate	RMSE	Relative Bias	Cov. rate	RMSE	Relative Bias
β per 100 WLM*	0.98	0.34	0.00	1.00	0.36	-0.03	0.94	0.35	-0.06
λ_1^F in 10^{-8}	0.00	3.36	-0.41				0.00	3.37	-0.42
λ_2^F in 10^{-6}	1.00	0.12	-0.05				0.92	0.22	-0.14
λ_3^F in 10^{-6}	1.00	0.28	-0.04				0.24	0.98	-0.18
λ_4^F in 10^{-6}	1.00	0.74	0.00				0.14	2.23	-0.23
λ_1^C in 10^{-8}				0.00	3.11	-0.38	0.00	3.07	-0.38
λ_2^C in 10^{-6}				1.00	0.12	-0.06	0.96	0.11	0.06
λ_3^C in 10^{-6}				0.98	0.33	-0.04	0.74	0.61	0.10
λ_4^C in 10^{-6}				1.00	0.67	0.04	0.10	2.18	0.23

Simulation 3: Simulation with error

In this part, data are simulated with measurement errors as presented in section 1.5.2. All parameters are estimated without taking into account of errors. Results are based on 50 simulated data sets. The parameter β is the same for the two cohorts fixed to 0.05 and there are 2000 miners in each cohort. We obtain the following Table 8 and Table 9.

Table 8. Mean of posterior medians and credible intervals of medians on 50 simulated French, Czech and pooled data sets with measurement error (β_{FR} = 0.05, β_{CZ} = 0.05, n_{FR} = 2000 and n_{CZ} = 2000).

	FR Cohort		CZ Cohort		Pooled Cohort	
Parameters	Mean	CI-95%	Mean	CI-95%	Mean	CI-95%
β per 100 WLM	3.14	[2.48 ; 3.84]	4.34	[3.72 ; 5.01]	3.81	[3.35 ; 4.31]
λ_1^F in 10^{-8}	4.82	[4.61 ; 5.07]			4.81	[4.61 ; 5.09]
λ_2^F in 10^{-6}	1.34	[1.19 ; 1.58]			1.29	[1.15 ; 1.51]
λ_3^F in 10^{-6}	5.48	[4.92 ; 5.99]			5.20	[4.64 ; 5.86]
λ_4^F in 10^{-6}	9.43	[7.79 ; 10.82]			8.65	[7.47 ; 9.83]
λ_1^C in 10^{-8}			4.80	[4.56 ; 5.21]	4.80	[4.57 ; 5.19]
λ_2^C in 10^{-6}			1.32	[1.09 ; 1.48]	1.39	[1.16 ; 1.59]
λ_3^C in 10^{-6}			5.31	[4.94 ; 5.64]	5.67	[5.10 ; 6.14]
λ_4^C in 10^{-6}			9.47	[8.47 ; 10.42]	10.40	[9.29 ; 11.67]

Table 9. Coverage rate (0.95), RMSE and relative bias on 50 simulated French, Czech and pooled data sets with measurement error (β_{FR} = 0.05, β_{CZ} = 0.05, n_{FR} = 2000 and n_{CZ} = 2000).

	FR Cohort			CZ Cohort			Pooled Cohort		
	Cov. rate	RMSE	Relative Bias	Cov. rate	RMSE	Relative Bias	Cov. rate	RMSE	Relative Bias
β per 100 WLM	0.06	1.90	-0.37	0.86	0.73	-0.13	0.04	1.22	-0.24
λ_1^F in 10^{-8}	0.00	3.29	-0.41				0.00	3.30	-0.41
λ_2^F in 10^{-6}	1.00	0.11	-0.03				1.00	0.13	-0.06
λ_3^F in 10^{-6}	1.00	0.27	0.00				0.96	0.39	-0.05
λ_4^F in 10^{-6}	0.96	0.92	0.03				0.94	0.84	-0.06
λ_1^C in 10^{-8}				0.00	3.31	-0.41	0.00	3.31	-0.41
λ_2^C in 10^{-6}				1.00	0.11	-0.04	1.00	0.11	0.01
λ_3^C in 10^{-6}				1.00	0.25	-0.03	0.98	0.33	0.04
λ_4^C in 10^{-6}				1.00	0.70	0.03	0.78	1.40	0.13

As already studied, when measurement errors are not taken into account, the risk β is underestimated in particular for the French cohort which has a lower mortality rate than that for the Czech cohort. The common association is estimated as a compromise between estimates in each cohort. When pooling the two cohorts, results on β are clearly improved compared to that obtained on French cohort. We notice an improvement in terms of precision for all the parameters.

Summary: The results obtained on pooled cohort confirm the existence of a statistically significant positive association between exposure cumulative radon exposure and the risk of death from lung cancer in the French cohort and Czech cohort of uranium miners. The multi-cohort Bayesian modelling refines our estimation in comparison to cohort per cohort estimation. These results are confirmed in simulation studies.

2. Subtask 4.6.2: Biologically based models of lung carcinogenesis

2.1 Introduction

2.2 Data

We analysed two historical rat datasets: the PNNL dataset, consisting of 4776 male Winstar rats exposed to radon at the Pacific Northwest National Laboratory in the United States (Table 10), and the CEA dataset with 4309 male Sprague-Dawley rats exposed to radon at the Commissariat à l'Energie in France (Table 11). All rats were exposed in their adulthood with cumulative exposure ranging from 0 to 6000 Working-Level Months (WLM), and exposure rates from 0 to 600 Working Level (WL). Both datasets contained mortality data from lung cancer, with the PNNL dataset also containing histological classification of the lung cancer cases, the two most common types being adenocarcinoma (ADC) and squamous cell carcinoma (SQCC). We analysed these two subtypes separately for the PNNL dataset. The CEA dataset also reported a size classification of the lung cancer cases in terms of discrete size classes: T0 being no tumour, T1 tumour smaller than 2 mm, T2 between 2 and 5 mm, T3 between 5 and 10 mm, T4 larger than 10 mm. Moreover, the CEA dataset reported incidence of preneoplastic lesions and for a subset of rats also the size in mm of these lesions.

Table 10. Number of rats and lung cancer cases with histological classification and exposure parameters for the six main exposure groups of the PNNL dataset.

	Exposure (WLM)	**Exposure rate (WL)**	**Rats**	**Lung cancers**		
	Range	**Range**		**ADC**	**SQCC**	**All LC**
Unexposed	0	0	668	3	2	7
Low-light (Ll)	<100	< 100	1786	20	13	41
Medium-light (Ml)	100–1000	< 100	829	55	13	79
Medium-heavy (Mh)	100–1000	> 100	427	31	3	38
High-light (Hl)	> 1000	< 100	128	55	33	79
High-heavy (Hh)	> 1000	> 100	422	138	40	177

Table 11. Number of rats, hyperplasia and lung cancers with exposure parameters for the exposure groups of the CEA dataset.

Exposure group	Exp. rate (WL)	Exposure (WLM)	Rats	Hyperplasia	Lung cancer
HIS	0	0	785	28	5
RnC	0	0	120	13	0
RnSham	0	0	120	10	0
RnD1	33.2	105	251	70	17
RnPr	10.31	107	240	39	7
RnD3	10.12	100	246	38	10
RnD12	2.25	100	241	31	12
RnAg	26.62	100	120	14	2
RnD6	2.44	42	211	33	3
25HDR	2.53	25	497	44	11
50HDR	3.79	50	497	41	19
200/	34.89	200	218	42	28
500/	72.28	500	108	27	31
1000/	142.5	1000	115	37	36
3000/	293	3000	54	15	20
6000/	285	6000	53	8	8
25LDR	0.36	25	500	22	3

2.3 Descriptive survival analysis and size analysis of hyperplasia and lung cancers

For a first descriptive survival analysis of both datasets, we performed Cox regression on the mortality of lung cancer. To improve visualization, the rats in the PNNL dataset were grouped into six exposure classes according to cumulative exposure (Low, Medium and High) and exposure rate (light or heavy) instead of the original 22 experimental cohorts, while for the CEA dataset we kept the original 15 exposure groups, but joined together all unexposed rats. Table 10 *and* Table 11above summarize the number of lung cancer cases in each dataset according to exposure class.

From survival analysis on adenocarcinomas (ADC) and squamous cell carcinomas (SQCC) in the PNNL dataset emerged two different dose-response behaviour as depicted in Figure 1*Table 10*. We performed a Cox regression on the number of lung cancer cases separately for ADC and SQCC by exposure class. From Figure 1 can be observed that the mortality risk for ADC increases with both higher cumulative exposure and exposure rate, while classes with same cumulative exposure show comparable risk for SQCC independently on the exposure rate.

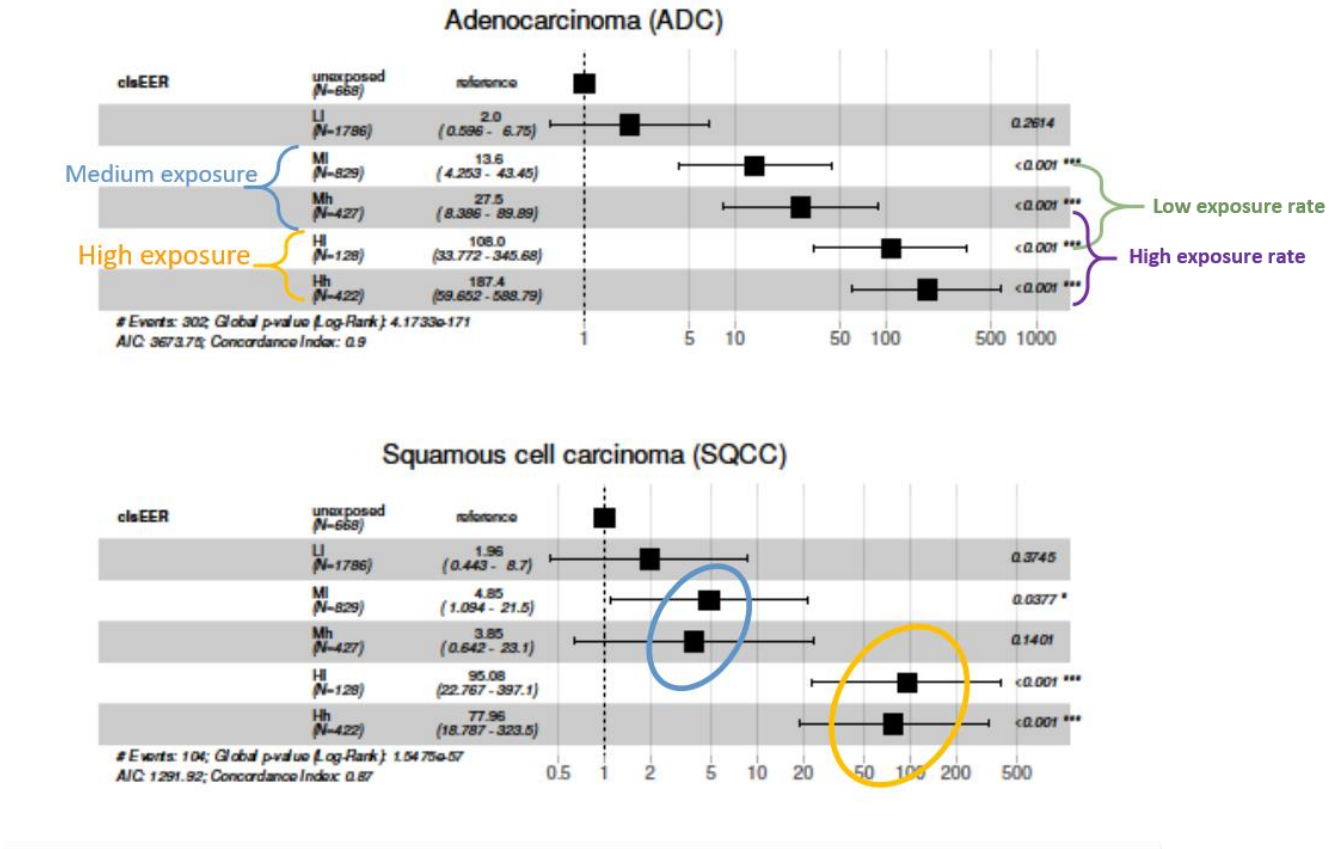

Figure 1. Survival analysis for ADC and SQCC by exposure class in the PNNL dataset.

Survival analysis for the CEA dataset shows an increase in the lung cancer risk with both increasing exposure and exposure rate, see Figure 2. A similar analysis on the incidence of precancerous lesions also shows a significant increase with exposure and exposure rate, see Figure 2. The class with lowest exposure and exposure rate 25LDR shows lower incidence of hyperplasia and lung cancer in comparison to the unexposed rats, with the former also statistically significant at 95% level.

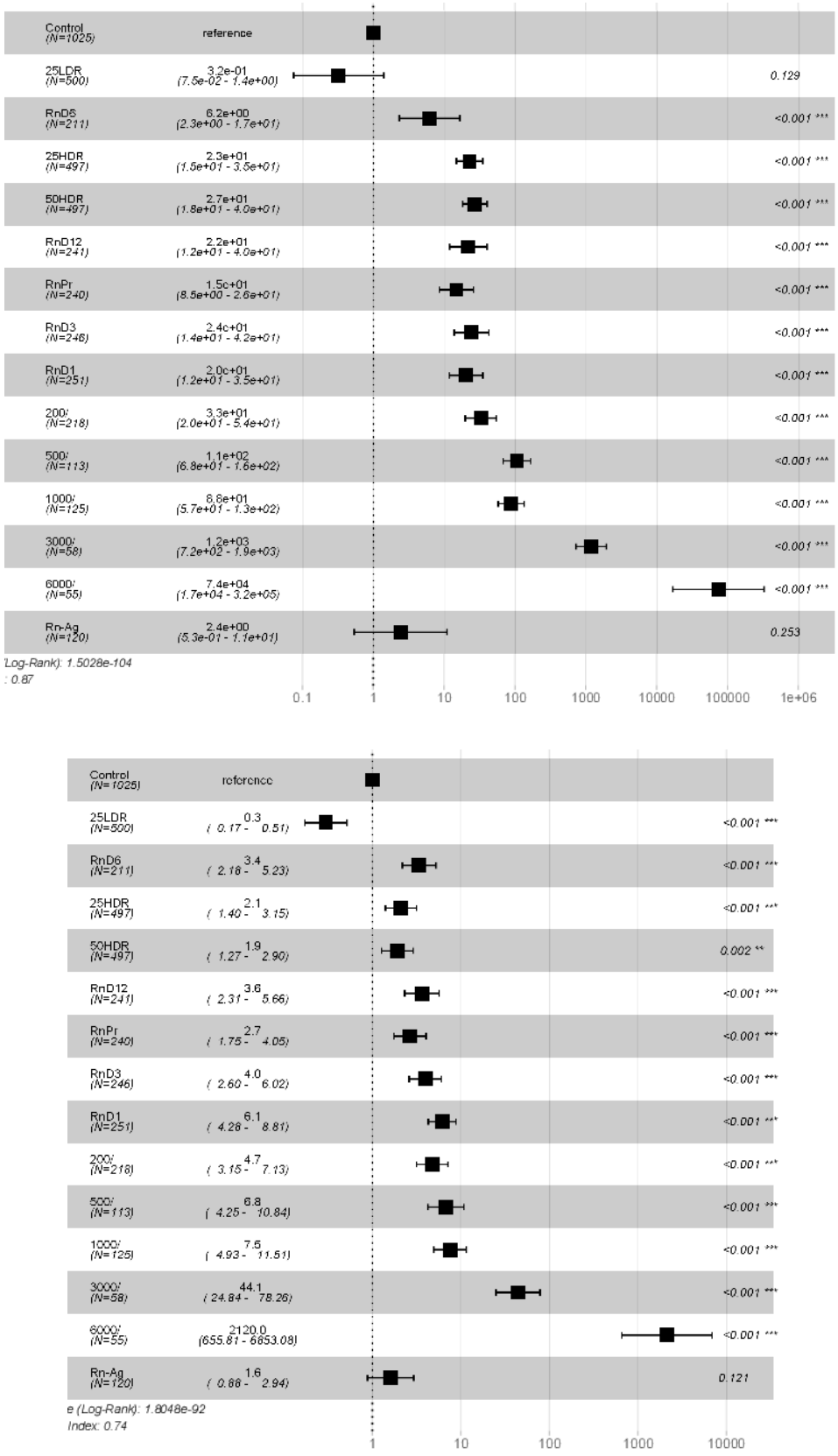

Figure 2. Hazard ratio on lung cancer mortality in the CEA dataset (top) and incidence of precancerous lesions (bottom).

Linear, respectively ordinal, regression was performed on the size distribution of premalignant lesions, respectively on size class distribution of lung cancers. The linear regression evidence larger lesions in two of the higher exposed groups in comparison to the unexposed rats, while ordinal regression showed no significant difference in the cancer size classes distribution among exposure groups.

2.4 Lung cancer risk model and timeline of carcinogenesis for radon-exposed rats

To analyse the PNNL dataset we applied a biologically-based model based on the multistage-clonal expansion model (MSCE). The core concept of this model is the description of the carcinogenic process as a sequence of stochastic cell-based processes (stages) which gradually turn normal stem cells into premalignant and cancer stem cells. Hyperplasia and tumour growth stages are represented by clonal expansion, and each stage is started by genetic events. Our conceptual model extends the model of Heidenreich et al. 1999, which was already used to analyse this dataset, by adding the cancer growth stage, which was previously simplified by inserting a fixed lag time in the model. Moreover, we consider the stochastic nature of cancer detection, as discussed by Little et al. 2024, introducing a detection probability of lung cancer proportional to its size. The effect of radon exposure is integrated in the model in two ways. The number of new precancerous lesions per time unit is increased during exposure linearly with the exposure rate, an effect known as initiation and motivated by the higher number of lung cancer cases in higher exposed rats, while the net growth rate of precancerous lesions can also be increased proportionally with the exposure rate, a phenomenon known as promotion, which is confirmed in our later analysis of the size distribution of precancerous lesions in the CEA dataset.

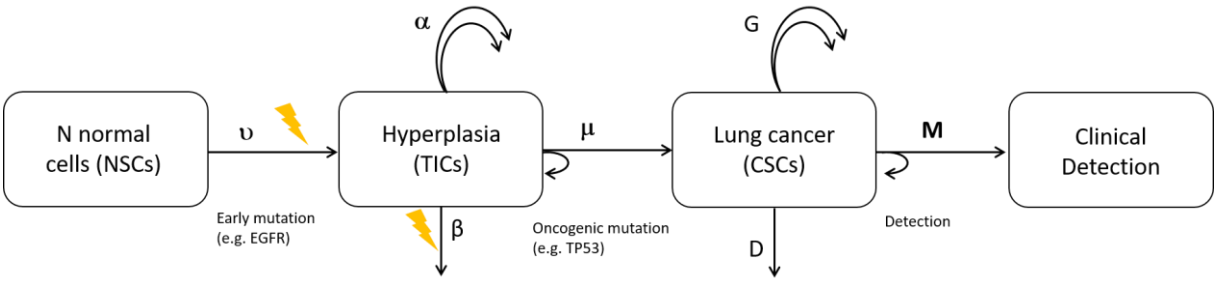

Figure 3. Schematic depiction of the biologically based risk model.

Motivated by the results of the survival analysis for ADC and SQCC the model was also applied separately on these two histological subtypes. The survival curves predicted by the model fit the mortality data very accurately, as shown in Figure 4. In particular, the best model for SQCC shows no significant radon induced initiation, coherent with the survival analysis showing only cumulative exposure increasing the lung cancer risk for SQCC and no role of the exposure rate.

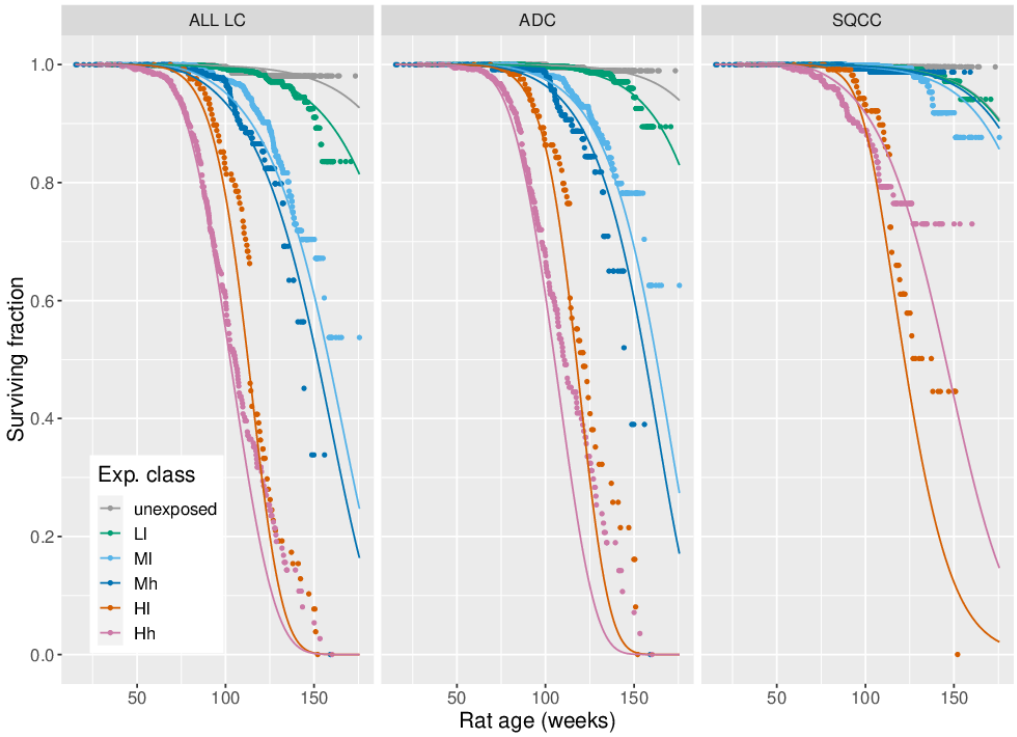

Figure 4. Comparison of survival curves predicted by the model with lung cancer cases in the dataset.

Based on this model we compared two approaches to the description of the time development of lung cancer. Luebeck et al. 2013 already established time quantities representing the mean life span of a tumour called latency time, respectively of a premalignant lesion, called sojourn time. These time quantities pertain to already developed objects and are therefore *retrospective*. We propose a new, *prospective* approach dividing the time between birth and cancer detection into three phases:

- an initiation period, where only normal stem cells populate the tissue, which is then ended by the appearance of the first premalignant cell,
- a clone formation period, where only premalignant cells are present and hyperplasia can grow until the first cancer stem cell appears and
- last a tumour formation period where malignant clones grow until detection.

Formation periods have to be understood in the light of dynamic cancer or hyperplasia evolution with stochastic growth, regression or even extinction of (pre-)malignant lesions. This is a main difference to the retrospective time quantities of Luebeck et al. 2013, which pertain to fully developed objects. The prospective timeline of lung cancer is given by the sum of these three quantities which are computed by means of nested survival functions and are therefore additive, that is their sum gives the duration of the whole carcinogenic process. Figure 5 shows how these three phases, in the figure depicted in terms of the underlying survival functions, are compressed and accelerated by higher radon exposure. This is a further difference with latency and sojourn time: the former is not influenced by radon exposure, the latter only by radon-induced promotion.

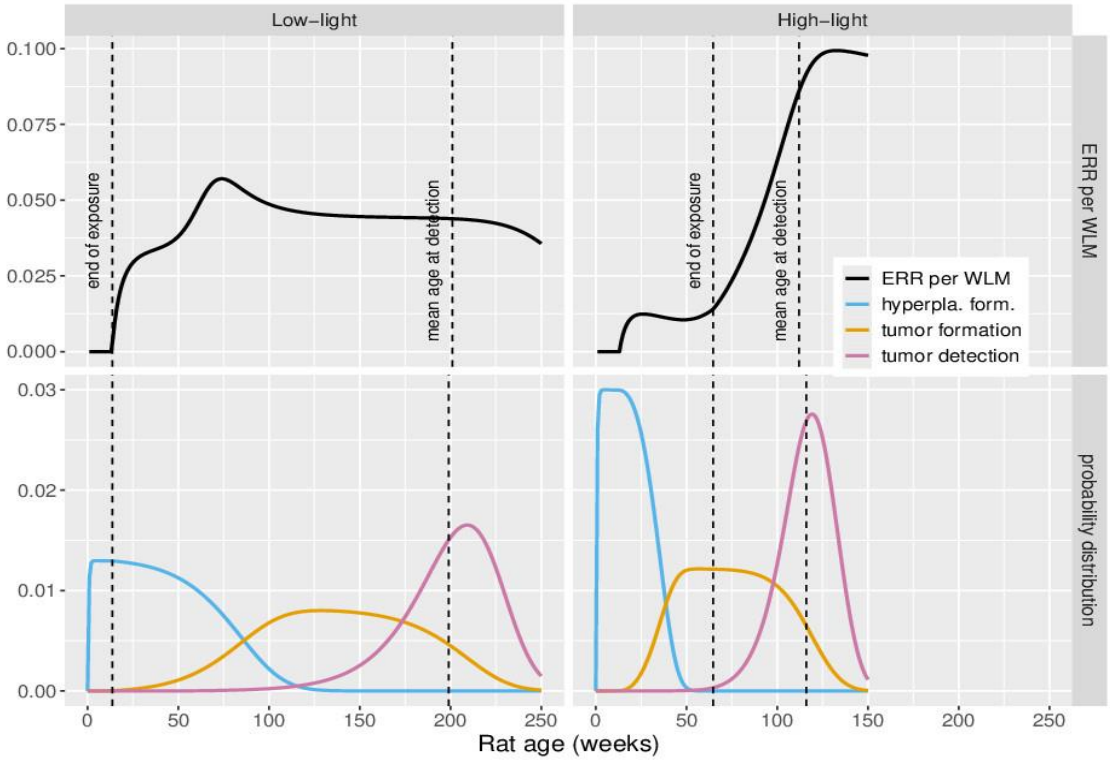

Figure 5. Comparison of excess relative risk per WLM and timeline of carcinogenesis for ADC and two exposure scenarios.

2.5 Hyperplasia and cancer growth model

To study the subset of the CEA dataset containing information on the size of precancerous lesions we have developed a hyperplasia growth model based on the work by Simonetto et al. 2023 for human colorectal adenoma. The model is depicted in Figure 6 below.

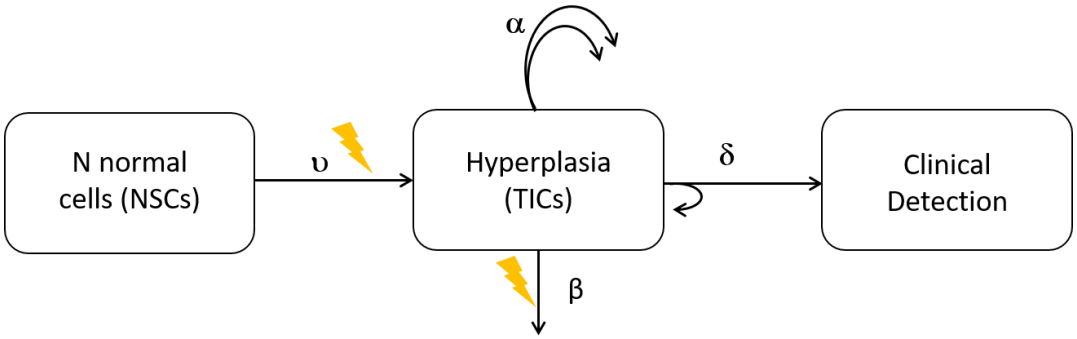

Figure 6. Hyperplasia growth model.

In comparison to the original model of Simonetto et al. 2023 we have added the effect of radon exposure on the age dependent parameters as explained for the PNNL subset, together with a stochastic detection rate proportional to hyperplasia size. The model uses the not only binary information on hyperplasia incidence but also size in mm of the lesions. This enables us to estimate biological parameters such as premalignant cells duplication rates, which were not identifiable in the model used above for the PNNL dataset. Several variants of this model have been tested, accounting for different effects of radon exposure, hyperplasia growth size and number of mutations necessary for hyperplasia

growth. The best model in terms of goodness of fit implements both initiation and promotion effect of radon exposure and a two-dimensional hyperplasia growth. In particular, the predicted initiation effect is significant only for exposure rates higher than 25 WL which is coherent with the observation from the Cox regression that low exposed rats no significantly higher risk for hyperplasia than the unexposed rats. The model can predict the number of hyperplasia cases by age and exposure classes together with the probability for a precancerous lesion to regress or extinct instead of growing, see Figure 7 and Figure 8.

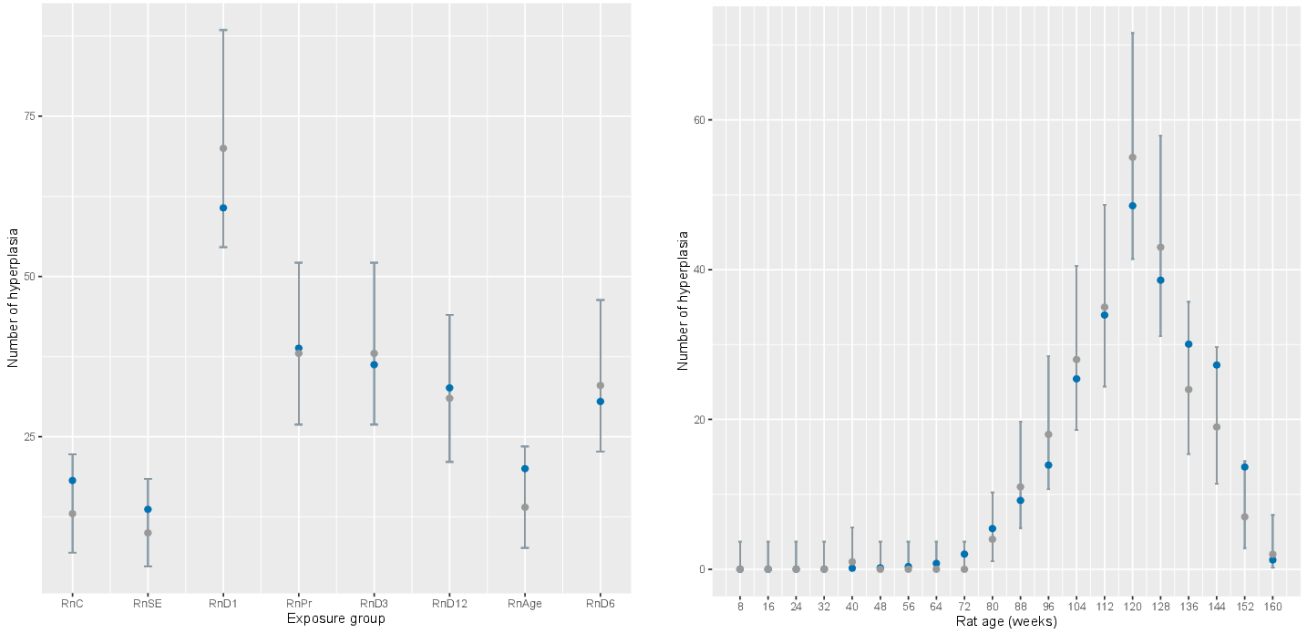

Figure 7. Comparison of predicted (blue) and cases (gray) in the dataset by age and exposure classes.

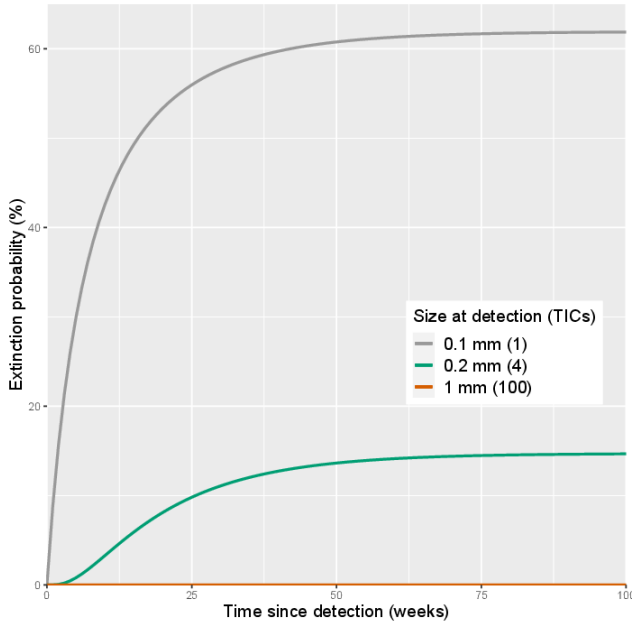

Figure 8. Extinction probability for hyperplasia of 0.1, 0.2 and 1 mm as function of the time since detection.

We further extended the hyperplasia growth model to lung cancer, so that we can integrate the information on lung cancer size. Schematically, the growth model looks like Figure 3 above with the subtle difference that the focus is on single tumours and their size rather than just on binary survival information. The role of radon in the carcinogenic process is the same as in the previous two models and we also keep stochastic cancer detection rates. The information on cancer size allows to estimate cancer cells duplication rate, which was not identifiable in the model used for the PNNL dataset. We tested several variants of this model, both in terms of expected radon effect and number of growth dimensions for lung cancer, with best model in term of goodness of fit a 3D model where radon induces both initiation and promotion of precancerous lesions. The resulting model could predict the number of cases by age and exposure group as well as the size of lung cancers, see Figure 9.

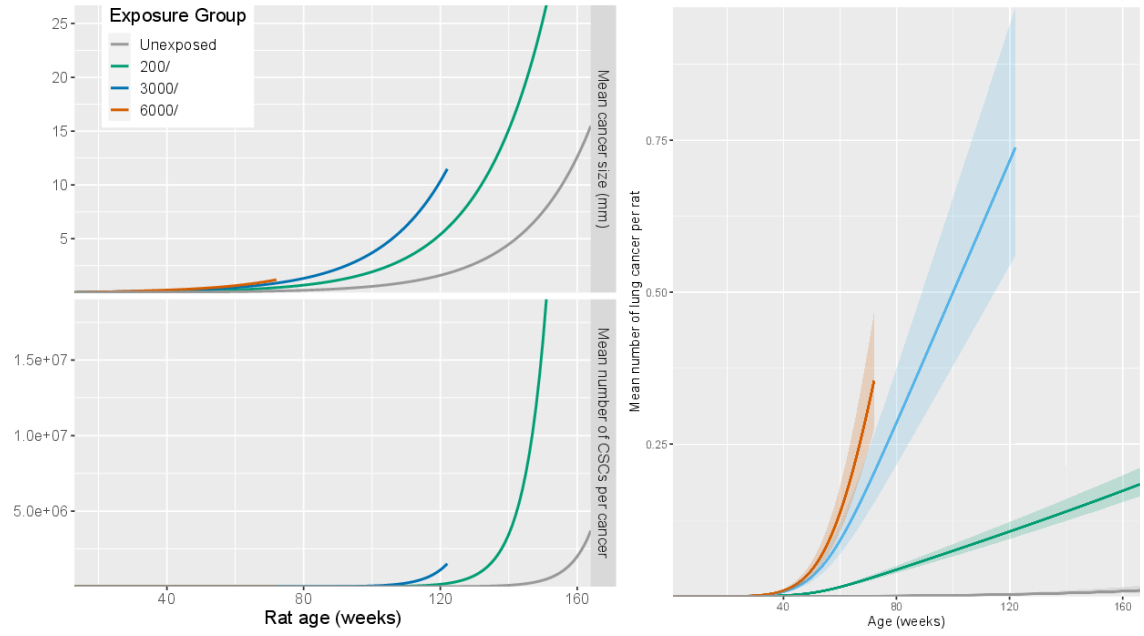

Figure 9. Expected lung cancer size and number of lung cancers per rat in three exposure groups.

Last, on the subset of the CEA dataset which contains information on both hyperplasia and lung cancer size, we tested a model integrating both these aspects and compared it with a model using only lung cancer size on this subset. The combined model resulted more numerically robust, with almost equal goodness of fit as the tumour only model, but without the risk of overfitting due to the low number of lung cancer cases in this subset.

3. Subtask 4.6.3: Analysis of lung cancer lifetime risk estimates

3.1 Introduction

As part of WP4 on health effects and risk assessment, task 4.6 is dedicated to the quantification of various sources of uncertainty in the inference of risks associated with radon, particularly in the estimation of lifetime lung cancer risk related to radon and smoking.

Lifetime lung cancer risk reflects the probability of dying from this disease during a lifetime. The lifetime excess absolute risk (LEAR) due to radon exposure can be defined as the difference between the lifetime risk of developing lung cancer for an individual exposed to radon and the lifetime risk for an individual not exposed. The lifetime lung cancer risk associated with radon is estimated based on

survival rates and reference rates for lung cancer in the general population, the exposure-risk relationship between radon and lung cancer, and radon exposure for an individual. In this work, LEAR calculations allow the comparison of risk models obtained from different occupational and residential studies with different characteristics while using the same radon exposure scenarios and baseline mortality rates. Of course, every population should be considered exposed to radon, and baseline death rates can only be inferred basing on the average excess relative risk (ERR) expected from the different models, given the weighted average indoor radon concentration (IRC) in each country.

This approach will be implemented based on selected and up-to-date European and continental data from the period preceding the COVID-19 pandemic (in order to limit any further confounding factor possibly affecting our datasets). In addition, we examine the impact of mitigation measures on lifetime lung cancer risk, particularly in homes with high radon concentrations or in situations where radon levels suddenly increase, such as after moving to a different residence.

Secondly, smoking, the major risk factor for lung cancer, is incorporated into the analysis to assess its influence on estimates of lifetime lung cancer risk associated with radon exposure. Different types of smoking–radon interactions are investigated in the present analyses.

3.2 LEAR calculation

The following section explains how a LEAR is calculated on the basis of an excess relative risk (ERR) model. The LEARS are differentiated according to gender and smoking group. With $ERR_S := ERR_{Rn|S}$ we obtain for the LEAR until age a with radon exposure $Rn = \rho$, smoking group S and sex g:

$$\begin{aligned} LEAR_{Sg}(\rho, a) &= \int_0^a \mu_{E,Sg}^{LC}(\rho, t) \cdot S_g(t)\, dt - \int_0^a \mu_{0,Sg}^{LC}(t) \cdot S_g(t)\, dt = \\ &= \int_0^a \mu_{0,Sg}^{LC}(t)\big(1 + ERR_S(\rho, t)\big) \cdot S_g(t)\, dt - \int_0^a \mu_{0,Sg}^{LC}(t) \cdot S_g(t)\, dt \\ &= \int_0^a \mu_{0,Sg}^{LC}(t) \cdot ERR_S(\rho, t) \cdot S_g(t)\, dt \approx \sum_{t=0}^{a} \mu_{0,Sg}^{LC}(t) \cdot ERR_S(\rho, t) \cdot \tilde{S}_g(t) \\ &= \sum_{t=0}^{a} \frac{\mu_{E,Sg}^{LC}(\bar{\rho}, t)}{1 + ERR_S(\bar{\rho}, t)} \cdot ERR_S(\rho, t) \cdot \tilde{S}_g(t) \end{aligned}$$

with

$$\tilde{S}_g(t) = \exp\left\{-\sum_{u=0}^{t-1} \mu_g(u)\right\}$$

Notation:

$\mu_{E,Sg}^{LC}(\rho, t)$: model-based lung cancer mortality rate at age t in smoking group S for sex g after long-term average radon concentration ρ

$\mu_{E,Sg}^{LC}(\bar{\rho}, t)$: observed lung cancer mortality rate at age t in smoking group S for sex g in a population with assumed average radon concentration $\bar{\rho}$

$\mu_{0,Sg}^{LC}(t)$: lung cancer mortality rate at age t in smoking group S for sex g without radon exposure extrapolated from $\mu_{E,Sg}^{LC}(\bar{\rho}, t)$

$S_g(t)$: Survival function for age t for sex g in the population

$\mu_g(t)$: mortality rate (all causes) at age t for sex g in the population

$a = 95$: Maximum lifetime

t: continuous age (in the integral) or discrete age (in the sum)

Note that we assume:

$\bar{\rho}$ does not depend on smoking group S

$S_g(t)$ and $\mu_g(t)$ do not depend on smoking group S

The above LEAR formula is used in the analyses in section 3.4. In the analyses in Section 3.3, no distinction is made between smoking groups, so that the index *f* for the smoking group can be omitted.

3.3 Lifetime lung cancer risk due to radon

3.3.1 Studied populations

Country-level data

French data: We use datasets provided by INSEE (The National Institute of Statistics and Economic Studies) and the French National Cancer Institute (INCa). The data includes all causes and lung cancer mortality rates, disaggregated by sex (male and female) and age group. Figure 10 presents the average mortality rates for lung cancer (left) and all causes (right) in France between 2015 and 2018, across different age groups for males and females. Additionally, we note that the national population-weighted average indoor radon concentration in France is approximately 70 Bq/m³.

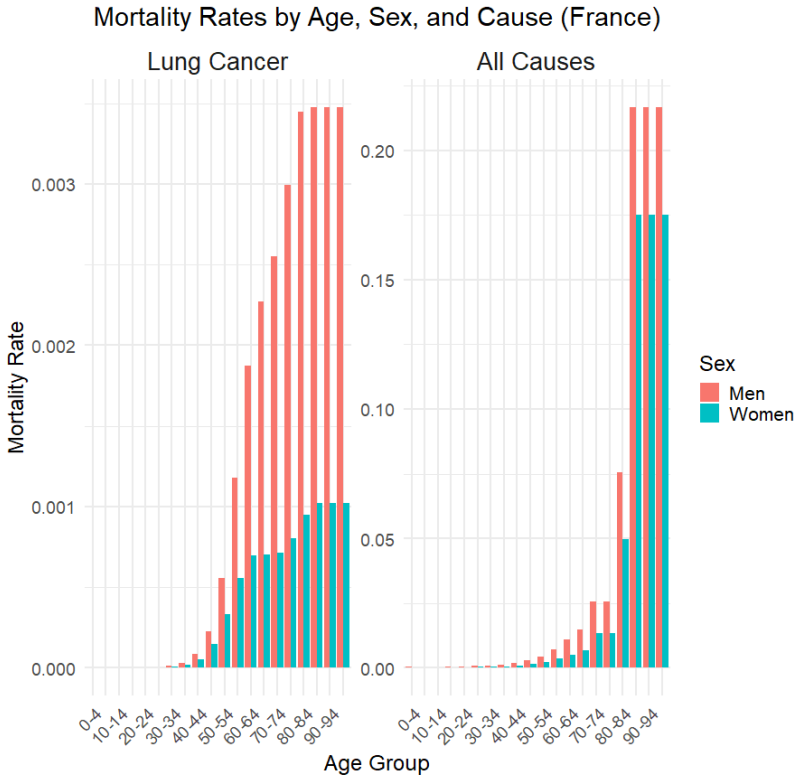

Figure 10. Average lung cancer death rates (on the left) and all-cause death rates (on the right) in France between 2015 and 2018 for males and females.

Italian data: In order to calculate lifetime lung cancer (LC) risk due to radon we used demographic data provided by the Italian Institute for Statistics (ISTAT), and average indoor radon concentration weighted on the population of 79 Bq/m³, obtained basing on the Second Italian Radon National Survey[1].

The general death rate $\mu_g(t)$, and the lung cancer-specific death rate $\mu_g^{LC}(\overline{\rho},t)$, subdivided by 5-years age groups and gender, were calculated using the Italian population distribution $P_g(t)$, the yearly overall deaths in Italy $M_g(t)$, and the yearly LC deaths $M_g^{LC}(t)$ as:

$$\mu_g(t) = M_g(t)/P_g(t)\ ,\ \mu_g^{LC}(\overline{\rho},t) = M_g^{LC}(\overline{\rho},t)/P_g(t)\ ,$$

where the index g specifies the gender (Male or Female), LC stands for Lung Cancer, the variable t specifies the age-group, and $\overline{\rho}$ is the radon concentration level to which the Italian population is on average exposed, estimated at 79 Bq/m³. The overall and the Lung Cancer-specific death rates were calculated for each of the years between 2015-2019, in order to obtain a 5-year average value. The Italian Institute of Statistics, moreover, provides also data about age- and gender-specific smoke prevalence.

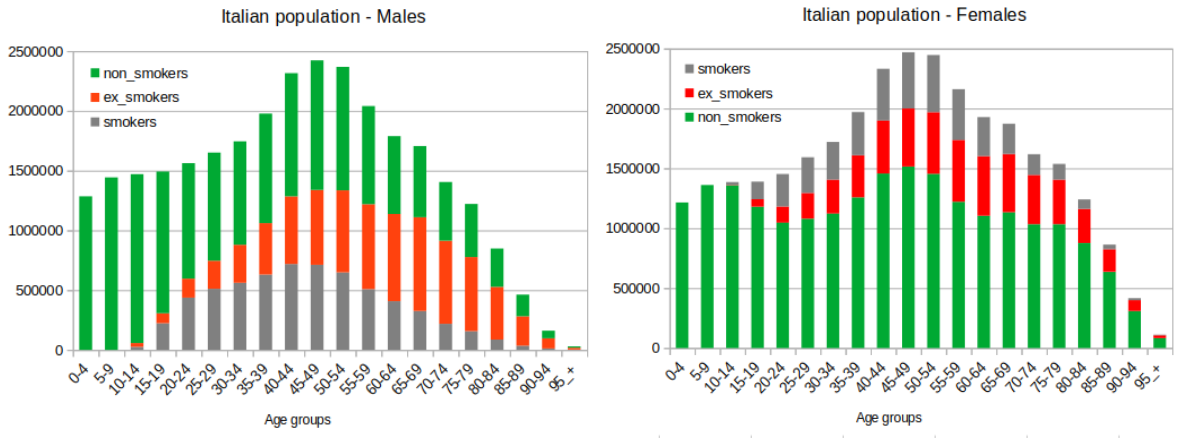

Figure 11. Histograms of Italian population (Males on the left and Females on the right), divided by age and smoking status.

Summarizing, the average 2015-2019 Italian population consists of around 60 million people, 29 million males and 31 million females, the average (weighted) age is 43 years old for men and 46 for women, while the mode for both groups is between 45-49 years old.

In terms of smoking prevalence, the Italian population (see Figure 11) is composed of 22% smokers for males and 14% for women, while ex-smokers are respectively 31% and 16%.

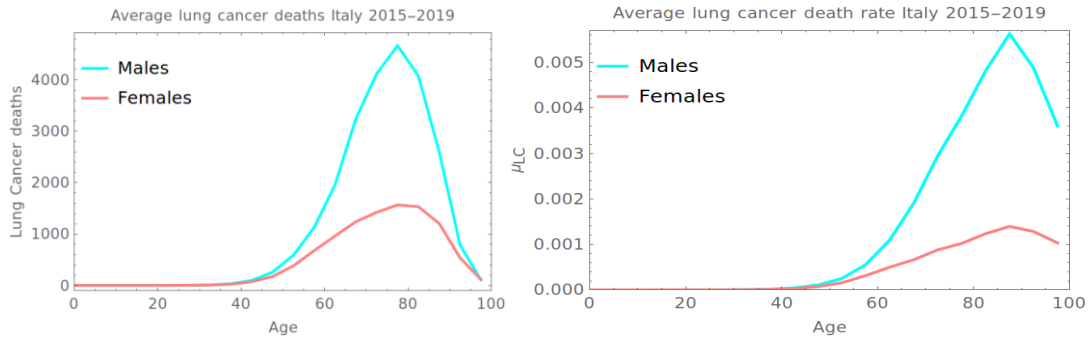

Figure 12. Average Lung Cancer deaths (on the left and death rates (on the right) in Italy between 2015 and 2019 for male (cyan) and female (light red) subgroups.

The average number of Lung Cancer deaths between 2015-2019, according to ISTAT, is 33.572, 23.665 males and 9.907 women. The peak of LC deaths for both groups is between 75 and 79 years old (see Figure 12); on the other hand, the maximum value for LC death rates distribution $\mu_g^{LC}(\overline{\rho},t)$ in both cases is around 85 years old.

German data: We use the mortality rates for all cancer and lung cancer, separated by sex (male and female), age class (0, 1–14, 15–19, ..., 85–89, 90+ years) and calendar year. This data was provided by the Federal Health Monitoring of Germany (GBE 2023). The mortality rates for the years 2015-2019 were averaged for the further calculations in order to equalize annual fluctuations. 63 Bq/m³ is used as the average radon concentration in living spaces in Germany (Petermann et al. 2024).

Continental data

ICRP data: Data from Tables A.4.12 and A.4.13 of the Publication 103 of ICRP (International Commission on Radiological Protection) were used to estimate the lifetime risk of lung cancer resulting from radon exposure. These tables provide age- and sex-specific baseline cancer mortality rates for various organs, including lung, based on a Euro-American population (see Figure 13). For radon exposure in the American population, we considered an average indoor radon concentration of 46 Bq/m³, as estimated by Marcinowski et al. (1994). The U.S. population is approximately 340 million, while Europe has about 450 million people exposed to an average concentration of 63 Bq/m³, an approximate value derived as the weighted average of national radon averages reported in the UNSCEAR (2006) report 5 for the European countries that conducted a national survey. Using population weights of roughly 60% for Europe and 40% for the U.S., the resulting weighted average indoor radon concentration across these regions would be approximately 56 Bq/m³.

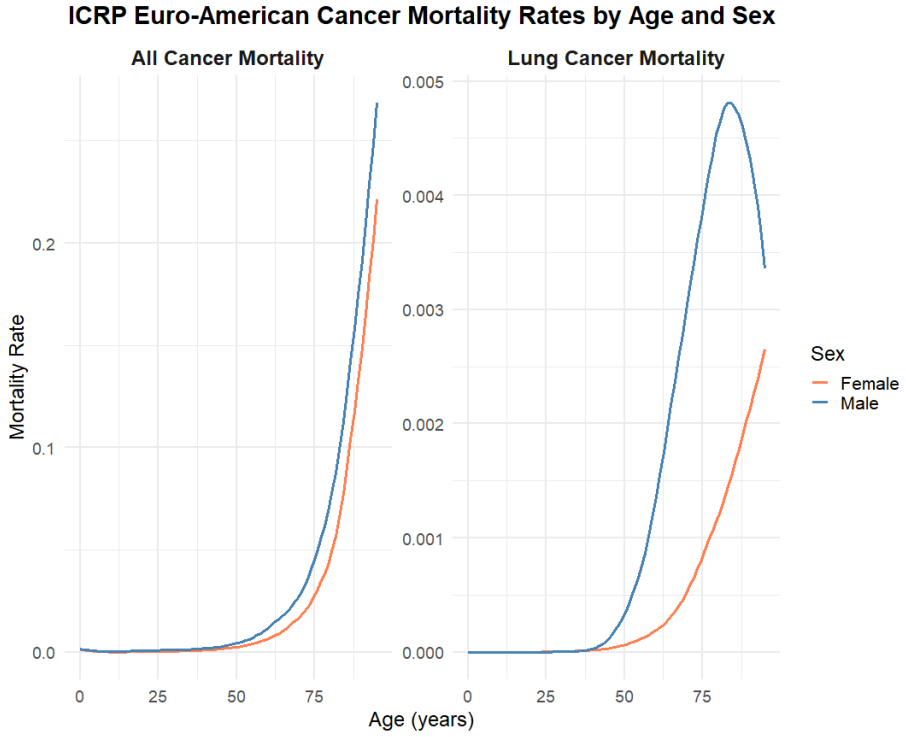

Figure 13. Average lung cancer death rates (on the left) and all-cause death rates (on the right) from ICRP data for males and females, expressed per person-years.

3.3.2 Selection of risk models for radon-associated lung cancer

Several radon risk models have been developed based on epidemiological studies of miners occupationally exposed to radon and individuals exposed to radon in residential settings. We present here three exposure-risk models selected for this work.

Darby model from the joint European study on residential radon

The joint analysis of 13 case-control studies of residential radon and lung cancer in nine European countries (Darby et al. 2006) focused on the mean radon concentration in homes occupied throughout a period of 5 to 34 years. The relation between residential radon exposure and lung cancer was defined as a simple linear model:

$$ERR\,(t,\bar{\rho}) = \beta(t)\cdot\bar{\rho}$$

- t is the attained age
- $\beta(t)$ is the slope of the exposure-risk relationship (β = 0.16/ (100 Bq/m³)) after 30 years (35 minus some 5 years lag)

$$\beta(t) = \begin{cases} 0, & t < 5 \\ \dfrac{(t-5)}{30}\cdot 0.16, & 5 \le t < 35 \\ 0.16, & t \ge 35 \end{cases}$$

- $\overline{\rho}$ average Radon concentration in homes within the previous 35 years, expressed in Bq/m³

BEIR VI model from studies on Uranium miners

The BEIR-VI excess relative risk (ERR) is presented as:

$$ERR(\omega,t,er) = \beta\big(\theta_{[5-14]}\omega_{[5-14]} + \theta_{[15-24]}\omega_{[15-24]} + \theta_{[25+]}\omega_{[25+]}\big)\phi_t\cdot\gamma_{er}$$

The used model is titled “BEIR-VI Age-Concentration Model” (NRC 1999). Here, $\beta = 0.0768$ represents the ERR/WLM in the reference category, since $\theta_{[5-14]} = 1$ by definition. Parameters $\theta_{[15-24]} = 0.78$ and $\theta_{[25+]} = 0.51$ quantify the effect modification by time since exposure with cumulative radon exposure $\omega_{[5-14]}$, $\omega_{[15-24]}$, $\omega_{[25+]}$in the windows 5-14, 15-24 or 25+ years before attained age t. ϕ_t and γ_{er} denote the effect-modifying factors of attained age and exposure rate. See NRC (1999) for more details.

The Pooled Uranium Miners Cohort Study: the PUMA model

The PUMA study (Pooling Uranium Miners Analysis) includes data from seven cohorts across Canada, the Czech Republic, France, Germany, and the USA. PUMA full cohort (Kelly-Reif et al. 2023) includes all eligible miners from these seven cohorts. However, the PUMA 1960+ sub-cohort (Richardson et al. 2022) only focuses on miners hired in 1960 or later. The lung cancer Excess Relative Risk (ERR) in the both cohorts were estimated using models with the BEIR VI model structure as above but with varying categories of time since exposure and the effect-modifying variables. See Kelly-Reif et al. (2023) and Richardson et al. (2022) for more details.

3.3.3 Characterization of radon exposure scenarios

To better understand the various sources of uncertainties in the relationship between radon exposure and lung cancer lifetime risk, we consider three distinct exposure patterns. These scenarios reflect typical real-life situations where radon levels may remain constant, decrease due to mitigation efforts, or increase due to changes.

a. Constant radon exposure

We begin by examining the predicted lung cancer risks associated with constant radon exposure levels of 50 Bq/m³ and 300 Bq/m³.

b. Decreasing radon exposure

People living in residences with high radon levels require remedial action. Therefore, it is important to investigate the impact of reductions in radon exposure. Specifically, we examined scenarios of risk mitigation for an individual who have been exposed the first half of his/her life to a radon concentration of 300 Bq/m³ in a home up to the age of 45, then reduced to 50 Bq/m³ after mitigation.

c. Increasing radon exposure

This scenario reflects a rise in radon exposure, for example, due to a change in residence. The radon concentration increases from 50 Bq/m³ in the initial home to 300 Bq/m³ in the new home. The housing change is assumed to occur when the individual is 45 years old.

3.3.4 Results of LEAR calculation based on national and continental data for radon-only exposure scenarios

Using the average lung cancer mortality rates, Tables 12 to 14 present lifetime radon-induced lung cancer risk estimates (in %) by sex, based on different risk models and national datasets (French, German, and Italian). The estimates are provided for four exposure scenarios: constant exposure from age 0 to 94, exposure reduction at age 45 (from 300 to 50 Bq/m³), and exposure increase at age 45 (from 50 to 300 Bq/m³). Risk calculations were performed using the following risk models: Darby, BEIR VI, and two versions of the PUMA model and the results are presented by sex (males and females). Table 4 presents lifetime radon-induced lung cancer risk estimates (in %) for males and females, based on same models using ICRP reference data and the same exposure scenarios. This dataset allows comparison with national datasets.

Results of LEAR calculation based on French, German, and Italian data

The LEAR results (Tables 12 to 14) demonstrate consistent patterns across the French, German, and Italian datasets. Males exhibit higher lifetime radon-induced lung cancer risks than females. For example, under constant exposure to 300 Bq/m³ using the BEIR VI model, the estimated LEAR for French males is 5.31%, compared to 2.24% for females. Among models, PUMA_Sub produces the highest LEAR values, such as 6.77% for French males, while the Darby model gives the lowest, at 2.59%. Reducing radon exposure at age 45 (from 300 to 50 Bq/m³) lowers the lifetime risk, though the decrease is moderate: using German data and BEIR VI, male risk drops from 4.60% to 3.24%.

Table 12. Lifetime excess absolute risk (%) estimates by sex according to different models using French data.

Risk model	Constant Exposure from age 0-94		Reduction of Exposure at age 45	Increase in Exposure at age 45
	100 Bq/ m³	300 Bq/m³	300 Bq/m³ to 50 Bq/m³	50 Bq/m³ to 300 Bq/m³
Males				
Darby	0.86	2.59	0.46	2.54
BEIR VI	1.77	5.31	3.92	2.33
PUMA_Full	1.66	5.00	3.30	2.58
PUMA_Sub	2.25	6.77	5.40	2.59
Female				
Darby	0.34	1.03	0.19	1.01
BEIR VI	0.74	2.24	1.71	0.92
PUMA_Full	0.68	2.06	1.41	1.01
PUMA_Sub	0.94	2.84	2.31	1.04

Table 13. Lifetime excess absolute risk (%) estimates by sex according to different models using German data.

Risk model	Constant Exposure from age 0-94		Reduction of Exposure at age 45	Increase in Exposure at age 45
	100 Bq/m³	300 Bq/m³	300 Bq/m³ to 50 Bq/m³	50 Bq/m³ to 300 Bq/m³
Males				
Darby	0.80	2.41	0.93	1.88
BEIR VI	1.53	4.60	3.24	2.12
PUMA_Full	1.55	4.66	2.91	2.52
PUMA_Sub	2.03	6.09	4.66	2.44
Females				
Darby	0.44	1.32	0.55	1.00
BEIR VI	0.89	2.68	1.93	1.20
PUMA_Full	0.86	2.59	1.66	1.36
PUMA_Sub	1.16	3.47	2.70	1.35

Table 14. Lifetime excess absolute risk (%) estimates by sex according to different models using Italian data.

Risk model	Constant Exposure from age 0-94		Reduction of Exposure at age 45	Increase in Exposure at age 45
	100 Bq/m³	300 Bq/m³	300 Bq/m³ to 50 Bq/m³	50 Bq/m³ to 300 Bq/m³
Males				
Darby	0.99	2.96	0.51	2.95
BEIR VI	1.61	4.84	2.49	3.16
PUMA_Full	1.87	5.60	3.74	2.79
PUMA_Sub	2.24	6.73	3.59	4.27
Females				
Darby	0.37	1.11	0.19	1.10
BEIR VI	0.66	1.97	1.04	1.26
PUMA_Full	0.71	2.14	1.45	1.05
PUMA_Sub	0.89	2.67	1.46	1.65

Results of LEAR calculation based on continental data (ICRP data)

The LEAR results based on ICRP data follow similar trends observed in the national datasets, with higher risk estimates for males than females (Table 15). For example, under constant exposure to 300 Bq/m³, male lifetime risk ranges from 2.68% (Darby) to 6.93% (PUMA_Sub), closely aligning with the highest values seen in the French (6.77%) and Italian (6.73%) data, and slightly higher than German estimates (6.09%). Female lifetime risk reaches up to 1.93% (PUMA_Sub), aligning with the upper estimates from the French data (2.84%) and Italian data (2.67%), and slightly lower than the German maximum (3.47%). Modifying exposure scenarios has a noticeable impact: reducing the exposure at age 45 (from 300 to 50 Bq/m³) leads to a moderate decrease in risk, while increasing the exposure at age 45 (from 50 to 300 Bq/m³) results a lower lifetime risk than constant high exposure. This pattern reflects the importance of mitigation measures.

Table 15. Lifetime excess absolute risk (%) estimates for males and females according to different models using ICRP data.

Risk model	Constant Exposure from age 0-94		Reduction of Exposure at age 45	Increase in Exposure at age 45
	100 Bq/ m³	300 Bq/m³	300 Bq/m³ to 50 Bq/m³	50 Bq/m³ to 300 Bq/m³
Males				
Darby	0.89	2.68	0.46	2.66
BEIR VI	1.77	5.32	3.83	2.44
PUMA Full	1.74	5.22	3.34	2.80
PUMA_Sub	2.31	6.93	5.44	2.75
Females				
Darby	0.27	0.81	0.14	0.80
BEIR VI	0.43	1.29	0.89	0.63
PUMA_Full	0.53	1.60	0.96	0.93
PUMA_Sub	0.64	1.93	1.43	0.84

3.4 Lifetime lung cancer risk due to radon and smoking

3.4.1 Studied populations

Country-level data

Italian data: In order subdivide by smoking category the Italian lung cancer (LC) data, already split by gender and age-group, we can make use of the different LC Relative Risks (with respect to non-smokers) related to each group available in Thun et al. (1997) and Pesch et al. (2012) studies. The analysis of the American CPS-II study (Thun et al. 1997) provided age-specific RRs by gender and smoking status for the US population between 1982 and 1997. On the other hand, the more recent Pesch study, based on a pooled analysis predominantly of European studies, provides estimated smoking-related relative risks (RRs), but does not provide insights into age-specific RR values. Therefore, regarding the Italian population, we have a reliable estimate of the average RR for each smoking category. In fact, the Pesch study is consistent with previous results from a pooling of European studies only (Simonato et al. 2001; see Table 5) but we do not have information on the age-specific distribution of RR. For this reason, we derived it based on data from the Thun study, which was conducted on the U.S. population.

Table 16. Summary of Relative Risks (the value in bold) at 95% confidence level for each smoking category (with respect to non-smokers) in the major Lung Cancer studies in the USA and Europe.

	Thun (1997)	Simonato (2001)	Pesch (2012)
Smoking category & gender	**RR** (95% CI)	**RR** (95% CI)	**RR** (95% CI)
Smokers Males	**20.3** (16.4 – 25.1)	**23.9** (19.7 – 29.0)	**23.6** (20.4 – 27.2)
Smokers Females	11.**9** (10.1 – 13.7)	**8.7** (7.4 – 10.3)	**7.8** (6.8 – 9.0)
Ex-smokers Males	n.a.	**7.5** (6.2 – 9.1)	**7.5** (6.5 – 8.7)
Ex-smokers Females	n.a.	**2.0** (1.6 – 2.4)	**2.8** (2.4 – 3.3)

In order to obtain a reasonable starting point for reconstructing the age-specific LC death rates for each smoking category in Italy, we might assume the age pattern of distribution to be the same of CPS-II (Thun) study, and rescale each value according to the European mean RR values:

$$RR_{g,h}^{Ita}(t) = \frac{<RR_{g,S}^{Pesch}>}{<RR_{g,h}^{Thun}>} RR_{g,h}^{Thun}(t) ,$$

where the index g indicates the gender (M, F) and S the smoking group (ns, cs, exs). Any missing value was linearly interpolated. The overall Relative Risks age-distribution for each smoking category reconstructed for Italy is shown in Figure 14.

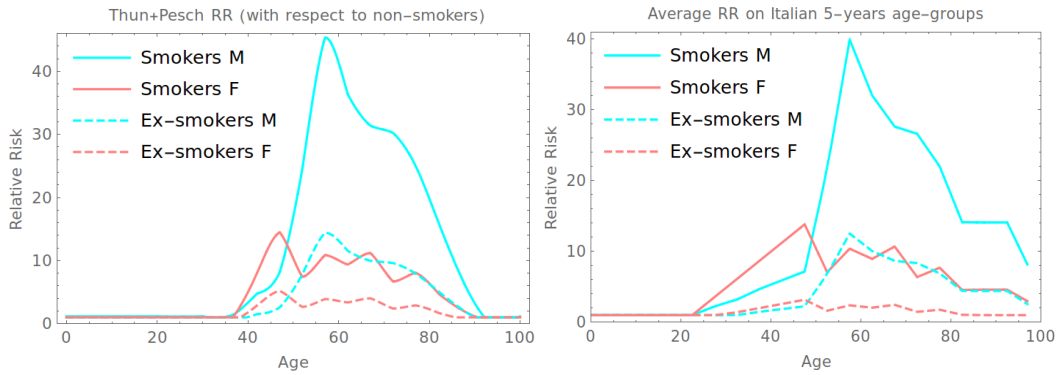

Figure 14. Relative Risk distributions by smoking status, according to Thun 1997 CPS-II analysis, rescaled using the average RR values in Europe, available in Pesch 2012 study. On the right the actual 5-years averages used in the calculations on the Italian population.

We now have the information and data necessary to reconstruct the trend in the risk of death from Lung Cancer by age-group for each smoking category (a similar approach was already discussed in Bochicchio et al., 2013). In general, we can re-express the total number of Lung Cancer deaths by age group $M_{g,S}(t)$ will be the sum of deaths within the single smoking categories, expressed as:

$$M_{g,h}(t) = \mu^{LC}_{g,ns}(t)\, Pop_{g,ns} + \mu^{LC}_{g,exs}(t)\, Pop_{g,exs} + \mu^{LC}_{g,s}(t)\, Pop_{g,cs}\ ,$$

where $\mu^{LC}_{g,S}(t)$ represent the absolute risk values by smoking category S and attained age t (clearly the same formula will be used for both for males and females $g = M, F$), and $Pop_{g,h}$ is the population in the age group. Of course, we still do not know the Absolute Risk values for each smoking category, however we know that relying on the non-smokers baseline, we have that $\mu^{LC}_{g,S}(t) = \mu^{LC}_{g,ns}(t)\, RR_{g,S}(t)$, then:

$$M_{g,S}(t) = \mu^{LC}_{g,ns}(t)\left(Pop_{g,ns} + RR_{g,exs}(t)\, Pop_{g,exs} + RR_{g,cs}(t)\, Pop_{g,cs}\right)$$

It will be now sufficient to overturn the previous equation to obtain the Baseline Risks for non-smokers:

$$\mu^{LC}_{g,ns}(t) = \frac{\mu^{LC}_{g}(t)\, Pop_{g}}{\left(Pop_{g,ns} + RR_{g,exs}(t)\, Pop_{g,exs} + RR_{g,cs}(t) Pop_{g,cs}\right)}$$

We are now able to split the total LC deaths and death rates by smoking status as shown in Figure 15.

The graphics in Figure 15 show the lung cancer death rates vs. attained age, for the different smoking categories in Italy between 2015-2019, according to the procedure here discussed. Below the inferred distribution of lung cancer deaths by gender, age and smoking categories.

The LEAR values for Italy specific by smoking status will be calculated using these estimates for the Lung Cancer death rates split by smoking category.

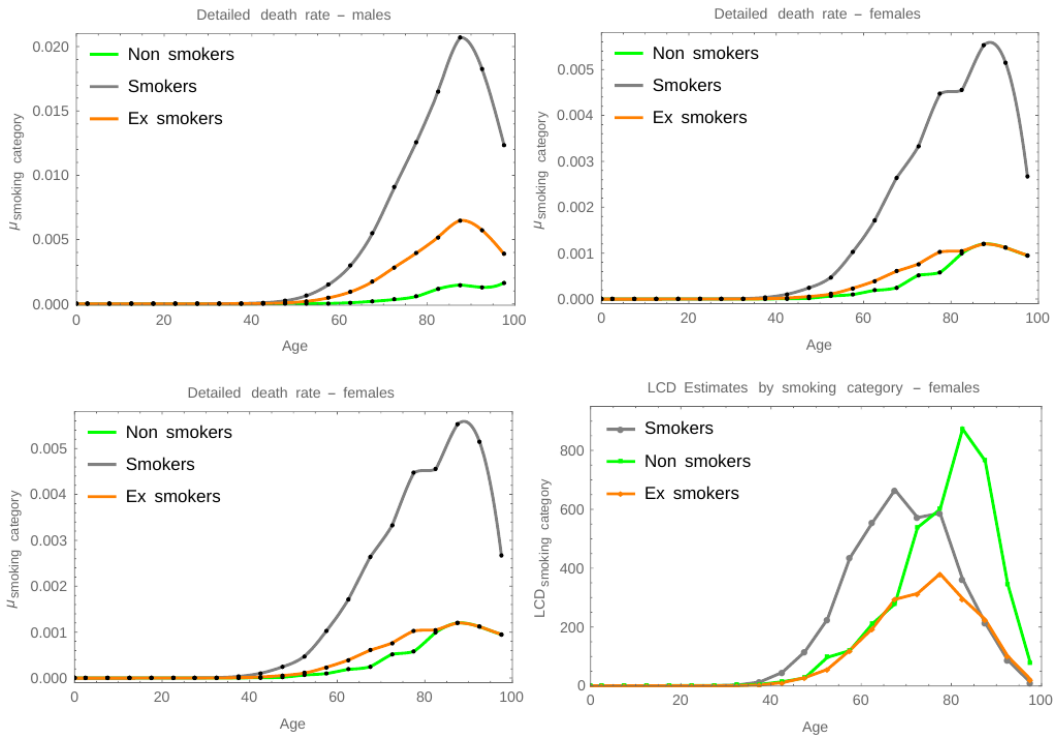

Figure 15. The above graphics show the lung cancer death rates vs. attained age, for the different smoking categories in Italy between 2015-2019, according to the procedure here discussed. Below the inferred distribution of lung cancer deaths by gender, age and smoking categories.

3.4.2 Risk models with interaction radon-smoking

Consider an individual who has been exposed to an average indoor radon concentration, denoted by Rn, and who belongs to a specific smoking category S (i.e., current smoker, ex-smoker and never smoker). The relative risk (RR) of lung cancer for an individual in smoking group S and exposed to a radon concentration Rn is defined as the ratio of their absolute lung cancer incidence rate, to the baseline rate, the rate observed in a non-smoker with no radon exposure. To describe the combined effects of smoking and radon exposure on lung cancer risk, a general linear model is adopted for the relative risk:

$$RR_{S,Rn} = 1 + ERR_S + ERR_{Rn} + ERR_{S\cdot Rn},$$

Here, ERR_S is the excess relative risk of lung cancer due to smoking, ERR_{Rn} is the excess relative risk due to radon exposure, and $ERR_{S\cdot Rn}$ is the excess relative risk due to the interaction of radon and smoking.

$$RR_{Rn|S} = \frac{1+ERR_S+ERR_{Rn}+ERR_{S\cdot Rn}}{1+ERR_S},$$

The $ERR_{Rn|S} = RR_{Rn|S} - 1$ indicates the excess radon-induced lung cancer risk in the respective smoker group S.

In the case of the additive model, it is assumed that there is no interaction between smoking and radon exposure, which mean that $ERR_{S\cdot Rn} = 0$. As a result, the relative risk of lung cancer for an individual in smoking group S and exposed to radon concentration Rn simplifies to:

$$RR_{S,Rn} = 1 + ERR_S + ERR_{Rn}.$$

Based on this formulation, the excess radon-related lung cancer risk within a specific smoking group S can be expressed conditionally as:

$$ERR_{Rn|S} = \frac{ERR_{Rn}}{1+ERR_S}.$$

This expression highlights how the apparent impact of radon is modulated by the baseline excess risk associated with smoking.

In the case of the multiplicative model, it is assumed that the interaction between smoking and radon exposure is multiplicative, such that $ERR_{S\cdot Rn} = ERR_s \cdot ERR_{Rn}$. This leads to the following expression for the relative risk of lung cancer:

$$RR_{S,Rn} = RR_S \cdot RR_{Rn} = (1 + ERR_S)(1 + ERR_{Rn}) = 1 + ERR_S + ERR_{Rn} + ERR_S \cdot ERR_{Rn}.$$

Under this model, the excess radon-related lung cancer risk within smoking group S simplifies to: $ERR_{Rn|S} = \frac{ERR_{Rn}+ERR_S\cdot ERR_{Rn}}{1+ERR_S} = ERR_{Rn}$.This result indicates that, in the multiplicative model, the excess relative risk due to radon is independent of smoking status.

Instead of an additive or a multiplicative model, a geometric mixture model can be applied as a mix of both models:

$$RR_{S,Rn} = [(1 + ERR_S)(1 + ERR_{Rn})]^{\theta} \cdot [1 + ERR_S + ERR_{Rn}]^{1-\theta}.$$

This model allows interpolation between the additive model ($\theta = 0$) and the multiplicative model ($\theta = 1$), with values of $0 < \theta < 1$ corresponding to a sub-multiplicative model. Within this framework, the excess relative risk due to radon in smoking group S is defined as:

$$ERR_{Rn|S} = \frac{[(1+ERR_S)(1+ERR_{Rn})]^{\theta}\cdot[1+ERR_S+ERR_{Rn}]^{1-\theta}}{1+ERR_S} - 1.$$

With the three types of radon-smoking interaction models presented, we can rely on findings from the literature that provide parameter estimates suitable for applying the Darby model and the BEIR VI models. These models are based on multiplicative and sub-multiplicative interaction structures.

Radon-smoking interaction using Darby model

If we assume a multiplicative interaction model, then the excess relative risk (ERR) from radon exposure is independent of smoking status. This means that:

$$ERR_S(\rho, t) = ERR(\rho, t) = \rho \cdot \beta(t)$$

Where

$$\beta(t) = \begin{cases} 0, & t < 5 \\ \frac{(t-5)}{30} \cdot 0.16, & 5 \le t < 35 \\ 0.16, & t \ge 35 \end{cases}$$

Alternatively, Darby et al. (2006) reported a suggestion of a possible sub-multiplicative interaction between radon exposure and smoking. The $ERR_f(\rho, t)$ per $100\ \mathrm{Bq/m^3}$were estimated and can therefore be used directly.

$$ERR_S(\rho, t) = \rho \cdot \beta_S(t)$$

$$\beta_S(t) = \begin{cases} 0, & t < 5 \\ \frac{(t-5)}{30} \cdot \beta_S, & 5 \le t < 35. \\ \beta_S, & t \ge 35 \end{cases}$$

The following parameter estimates were obtained: $\beta_{cs} = 0.10$ for current smokers, $\beta_{es} = 0.22$ for ex-smokers and $\beta_{ns} = 0.20$ for non-smokers.

Radon-smoking interaction using BEIR VI model

If we assume a multiplicative interaction model, then the excess relative risk (ERR) in BEIR VI model is independent of smoking status. This means that:

$$ERR_S(\rho,t) = ERR(\rho,t) = \beta \cdot \left(\theta_{[5-14]}\omega_{[5-14]} + \theta_{[15-24]}\omega_{[15-24]} + \theta_{[25-34]}\omega_{[25-34]}\right) \cdot \phi_t \cdot \gamma_\omega.$$

By applying a sub-multiplicative (Brand et al. 2005) interaction, the BEIR VI report (1999) pointed out that the underground miner studies used to estimate radon risks included very few never-smokers. To adjust for this, the report introduced correction factors, called δ_S parameters, that change the excess relative risk (ERR) depending on smoking status. These factors are $\delta_{cs} = 0.92$ for current smokers, $\delta_{ns} = 1.94$ for non-smokers, and 1.00 for the general population. The adjusted risk is calculated by multiplying the original $ERR(\rho,t)$ by the appropriate δ value:

$$ERR_S(\rho,t) = ERR(\rho,t) \cdot \delta_S.$$

3.4.3 Characterization of radon-smoking interaction exposure scenarios

The same three exposure scenarios (a, b, and c) studied in section 3.2 are considered in this section, now incorporating smoking habits as an additional factor influencing lung cancer risk. Three smoking status groups were considered: never smokers (people who have neither smoked regularly nor occasionally), ex-smokers (former regular and occasional smokers), and current smokers (regular and occasional smokers). Such stratification makes it possible to examine more precisely the combined effects of radon and smoking.

3.4.4 Results of LEAR calculations based on Italian data with radon-smoking interaction exposure scenarios

Tables 6 to 8 present the estimated lifetime excess absolute risk of radon-induced lung cancer for males and females, stratified by smoking categories (never smokers, ex-smokers, and current smokers). The risk values are expressed as percentages and were computed under four exposure scenarios: constant exposure from age 0 to 94 with different radon levels (100 and 300 Bq/m³), reduction of exposure at age 45 (from 300 to 50 Bq/m³), increase of exposure at age 45 (from 50 to 300 Bq/m³). To simplify comparison of results, calculations were performed on the Italian data using two risk models (Darby and BEIR VI), each with two radon-smoking interaction types: multiplicative and sub-multiplicative. These tables highlight the strong modifying effect of smoking status on LEAR calculations. Under constant exposure to 300 Bq/m³, for example, Tables 17 to 19 clearly show how smoking status strongly influences lifetime excess absolute risk. Among never-smokers, male LEAR ranges from 0.51% (Darby, multiplicative) to 1.38% (BEIR VI, sub-multiplicative), while female estimates range from 0.68% to 1.97%. In contrast, ex-smokers have higher risks, with male values between 3.04% and 4.33%, and females between 1.04% and 1.68%, depending on the model. The highest lifetime risks occur among current smokers, where male LEAR reaches up to 14.65% (BEIR VI, multiplicative) and female risk up to 7.25%. These differences illustrate the cumulative effect of smoking and radon exposure, for example, male LEAR values increase more than ten times, going from 1.38% in never-smokers to 14.65% in current-smokers.

Table 17. Lifetime excess absolute risk (%) estimates by sex, based on different radon exposure scenarios and smoking interaction models, for never-smokers, using Italian data.

Risk model	Rn*Smoking Interaction	Constant Exposure from age 0-94		Reduction of Exposure at age 45	Increase in Exposure at age 45
		100 Bq/m³	300 Bq/m³	300 Bq/m³ to 50 Bq/m³	50 Bq/m³ to 300 Bq/m³
Males					
Darby	Multipl.	0.17	0.51	0.09	0.50
Darby	Sub-multipl.	0.21	0.64	0.11	0.63
BEIR VI	Multipl.	0.24	0.71	0.36	0.47
BEIR VI	Sub-multipl.	0.46	1.38	0.70	0.91
Females					
Darby	Multipl.	0.23	0.68	0.12	0.67
Darby	Sub-multipl.	0.28	0.85	0.14	0.84
BEIR VI	Multipl.	0.34	1.01	0.51	0.67
BEIR VI	Sub-multipl.	0.66	1.97	1.00	1.30

Table 18. Lifetime excess absolute risk (%) estimates by sex, based on different radon exposure scenarios and smoking interaction models, for ex-smokers, using Italian data.

Risk model		Constant Exposure from age 0-94		Reduction of Exposure at age 45	Increase in Exposure at age 45
	Rn*Smoking Interaction	100 Bq/m³	300 Bq/m³	300 Bq/m³ to 50 Bq/m³	50 Bq/m³ to 300 Bq/m³
Males					
Darby	Multipl.	1.01	3.04	0.53	3.02
Darby	Sub-multipl.	1.37	4.12	0.70	4.11
BEIR VI	Multipl.	1.52	3.18	2.09	2.42
BEIR VI	Sub-multipl.	2.09	4.33	2.83	3.35
Females					
Darby	Multipl.	0.34	1.04	0.20	1.01
Darby	Sub-multipl.	0.45	1.36	0.24	1.34
BEIR VI	Multipl.	0.61	1.31	0.91	0.91
BEIR VI	Sub-multipl.	0.80	1.68	1.15	1.22

Table 19. Lifetime excess absolute risk (%) estimates by sex, based on different radon exposure scenarios and smoking interaction models, for current-smokers, using Italian data.

Risk model	Rn*Smoking Interaction	Constant Exposure from age 0-94		Reduction of Exposure at age 45	Increase of Exposure at age 45
		100 Bq/m³	300 Bq/m³	300 Bq/m³ to 50 Bq/m³	50 Bq/m³ to 300 Bq/m³
Males					
Darby	Multipl.	3.17	9.52	1.61	9.50
Darby	Sub-multipl.	1.98	5.95	1.01	5.94
BEIR VI	Multipl.	4.88	14.65	7.41	9.68
BEIR VI	Sub-multipl.	4.49	13.48	6.82	8.90
Females					
Darby	Multipl.	1.42	4.26	0.74	4.23
Darby	Sub-multipl.	0.89	2.66	0.46	2.64
BEIR VI	Multipl.	2.42	7.25	3.78	4.68
BEIR VI	Sub-multipl.	2.22	6.67	3.48	4.30

4. Discussion

Across all models and exposure scenarios, the LEAR values for France, Italy, and Germany are very similar for women. For men, however, the LEAR values in Germany tend to be lower than in France and Italy, reflecting differences in smoking habits and baseline lung cancer mortality. Results based on ICRP data (continental data in Euro-American population) are close to those derived from national datasets. The continental estimates are slightly higher overall particularly for males and for the PUMA Sub model.

Comparison of the three models shows that the BEIR VI and PUMA models consistently produce higher LEAR estimates than the Darby model. This is expected given their derivation from miner cohorts. The PUMA-sub model, in particular, is sensitive to long-term exposure, which increases the resulting LEAR values. In contrast, the Darby model is derived from a simpler linear approach, with a smaller risk coefficient, which produces moderate risk estimates across various scenarios.

When applying the same model to different countries (Italy, France, and Germany), results exhibit expected variability due to demographic and behavioral differences. These include variations in the age distribution of lung cancer mortality, smoking groups by age and gender, and radon exposure levels (population-weighted average indoor radon concentration is approximately equal to 79 Bq/m³ in Italy, 70 Bq/m³ in France and 63 Bq/m³ in Germany). Although the differences are minor, a consistent trend is that LEAR estimates based on continental data are slightly higher than those derived from the Italian, German, and French datasets, while remaining overall comparable.

Under constant exposure to 300 Bq/m³ scenarios, all models show significantly higher risks compared to 100 Bq/m³, with PUMA-sub model generally yields the highest risk estimates due to its sensitivity to long-term exposure. In a scenario where exposure is reduced from 300 Bq/m³ to 50 Bq/m³ at age 45, all models confirm the benefit of mitigation, showing substantial risk reductions compared to a constant exposure to 300 Bq/m³ Scenario. However, the BEIR VI model shows no reduction in lifetime risk compared to a constant lifetime exposure of 100 Bq/m³. This is because the model attributes a very high risk up to ~50 years (then it decreases), i.e., exactly when exposure is at its highest (300 Bq/m³) in this scenario. A similar trend is observed with the PUMA model. In contrast, the Darby model, being linear, recognizes a benefit in mitigation, even if it occurs at age 45. In the scenario where exposure is

increased from 50 Bq/m³ to 300 Bq/m³ at age 45, with the BEIR VI model, the lifetime risk is lower than under a scenario of constant exposure to 300 Bq/m³: the reasons are the same as above, *i.e.* the BEIR VI model estimates peak risk before age 50, and in this scenario, exposure before age 45 is only 50 Bq/m³. With the Darby model, the risk is approximately the same for both a constant lifetime exposure of Bq/m³ and the scenario where exposure increases from 50 to 300 Bq/m³ at age 45, because lung cancer mortality before age 45 is very low. Across all scenarios, women show lower lifetime lung cancer risk values than men, mainly because lung cancer rates are in general lower in women. However, by using the same risk coefficients for males and females across the different models, implicitly assumes that the way radon exposure affects risk over time is similar for both sexes. Overall, the results highlight that the timing of exposure and the choice of models have a notable impact on lifetime lung cancer risk calculation.

Smoking is the main risk factor for lung cancer, and the assessment of its impact on the lifetime risk of lung cancer associated to radon is a major issue. Under a multiplicative interaction assumption, both BEIR VI and Darby models estimate that for smokers, the lifetime risk of lung cancer due to radon exposure for current smokers is approximately 20 times higher than for never-smokers, regardless of the exposure level (100 or 300 Bq/m³), as expected, considering the average relative risks identified in the smoking surveys. This trend also applies to women, where the increased lifetime risk for smokers is between about 6 times and 7 times higher than for never-smokers, depending on the model (Darby and BEIR VI, respectively). In the multiplicative model, LEAR values for former smokers are clearly lower because of the lower initial mortality, which is closer to that of non-smokers than to that of smokers. This brings to an overall LEAR increase by a factor of 4 to 6 in men, and a lifetime risk increase by 50% to 80% in women with respect to non-smokers with the same exposure to radon (100 or 300 Bq/m³).

Smoking acts as a major confounding factor when estimating the specific contribution of radon to the occurrence of lung cancer. Moreover, since smoking significantly increases the risk of numerous other fatal diseases, such as heart attacks and aneurysms, the overall probability of observing lung cancer in this population may decrease due to earlier competing causes of death. This is likely why, in models that assume a sub-multiplicative interaction between radon and smoking, the risk coefficients for smokers turn out to be about half those for non-smokers in both the Darby and BEIR VI models. This leads to a corresponding reduction in the lifetime risk of lung cancer due to radon in smokers. Even so, the LEAR for male smokers remains about 9 to 10 times higher than for non-smokers, and for female smokers, about 3 to 3.5 times higher consistently across both models. For ex-smokers, the LEAR calculations require the definition of a specific smoking habit scenario in which a person begins smoking at age 20 and quits at age 50. This means that from 0 to 20 (when lung cancer death rate is insignificant anyway) we are dealing with a non-smoker with some interaction coefficient. At age 21 he/she turns into a smoker and the coefficient halves. At age 51 then the Darby and BEIR-VI models behave differently: with the Darby model, a person who stops smoking will see an increase in lung cancer risk (almost double), while with the BEIR-VI model, smokers and ex-smokers are both categorized as “ever smokers”, with no change in the smoking-radon interaction coefficient. With the Darby model, no significant difference is observed between the multiplicative and the sub-multiplicative interaction since, in this scenario, the specific risk values balance each other out; on the other hand, with the BEIR-VI model, we observe a decrease of LEAR of around 40-50% for both men and women compared to the multiplicative case.

In perspective, several areas can be further explored and analysed to improve the estimation and the interpretation of LEAR. First, the uncertainties in LEAR estimates should be more explicitly addressed. Incorporating confidence intervals around LEAR values would help assess whether differences across models, exposure scenarios, or subpopulations are statistically significant. Second, a further investigation on methods for estimating age-specific relative risks by smoking status would improve the precision of calculations and enhance model comparability. Lastly, comparisons between countries should also be done with different radon–smoking interaction model applied (e.g. additive, multiplicative

and sub-multiplicative). Recent findings from RadoNorm subtask 4.1 on the interaction between radon and smoking, based on updated data from uranium miner cohorts and residential studies, should be used in the lifetime risk calculation and the results obtained should be compared to those presented in this chapter. This comparison will be the focus of a future publication.

5. References

Allodji, R.S. et al. Assessment of uncertainty associated with measuring exposure to radon and decay products in the French uranium miners cohort. Journal of radiological protection 32, 85-100 (2012).

Bochicchio, F., et al. Quantitative evaluation of the lung cancer deaths attributable to residential radon: A simple method and results for all the 21 Italian Regions. Radiat Meas 50, 121–126 (2013).

Brand, K. P. et al. Residential radon in Canada: An uncertainty analysis of population and individual lung cancer risk. Risk Analysis 25, 253–269 (2005).

Carroll, R.J. Measurement error in Epidemiologic Studies. Encyclopedia of Biostatistics, (2005).

Clayton D, & Hills M. (2013). Statistical models in epidemiology. OUP Oxford.

Darby, S. et al. Residential radon and lung cancer - Detailed results of a collaborative analysis of individual data on 7148 persons with lung cancer and 14 208 persons without lung cancer from 13 epidemiologic studies in Europe. Scandinavian Journal of Work, Environment and Health, Supplement 32, 1–84 (2006).

Fendler J. et al. Bayesian identification and estimation of radon-related increased hazard rates of cancer death in the updated French cohort of uranium miners (1946-2014). Int Arch Occup Environ Health. 97(9): 941-958 (2024)

Fendler J. et al. Bayesian hierarchical approach to account for radon exposure measurement error when estimating the risk of death by several cancers in an occupational cohort of uranium miners, Proceedings of the Inaugural Meeting of the International Society for Radiation Epidemiology and Dosimetry (ISORED) (2023).

GBE. Deaths, mortality figures (from 1998). Robert Koch Institute and the Federal Statistical Office. www.gbe-bund.de. Accessed 29 Dec 2023. (2023).

Gelman A. et al. Bayesian Data Analysis, Third Edition. chap.7 p.172-174. Chapman & Hall; CRC texts in statistical science (2014).

Heidenreich, W. F., et al. Time Trends of Thyroid Cancer Incidence in Belarus after the Chernobyl Accident. Radiation Research, vol. 151, no. 5, 1999, pp. 617–25. JSTOR, https://doi.org/10.2307/3580038. Accessed 29 Aug. 2025.

Hoffmann, S. et al. Accounting for Berkson and classical measurement error in radon exposure 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 (2017).

Hoffmann S. et al. Shared and unshared exposure measurement error in occupational cohort studies and their effects on statistical inference in proportional hazards models. PloS one, 13(2) (2018).

Hoffmann, S. Approche hiérarchique bayésienne pour la prise en compte d'erreurs de mesure d'exposition chronique et à faible doses aux rayonnements ionisants dans l'estimation du risque de cancers radio-induits: Application à une cohorte de mineurs d'uranium. PhD manuscript (written in English), University Paris Sud (2017).

Kelly-Reif, K. et al. Radon and lung cancer in the pooled uranium miners analysis (PUMA): highly exposed early miners and all miners. Occup Environ Med 80, 385–391 (2023).

Keogh, R.H. et al. “STRATOS guidance document on measurement error and misclassification of variables in observational epidemiology: Part 1—Basic theory and simple methods of adjustment”. Statistics in medicine 39(16), 2197-2231 (2020).

Kleinbaum, D.G. and Klein, M. Survival analysis. T. 3. Springer (2010).

Kreuzer, M. et al., Radon and risk of extrapulmonary cancers: results of the german uranium miners’ cohort study, 1960-2003. British journal of cancer, 99(11):1946—1953 (2008)

Laurier, D. et al. An update of cancer mortality among the French cohort of uranium miners: extended follow-up and new source of data for causes of death. Eur J Epidemiol 19(2):139–46 (2004).

Little, M. et al. A Historical Survey of Key Epidemiological Studies of Ionizing Radiation Exposure. Radiation Research, 2024, 202 (2), pp.432-487. ff10.1667/RADE-24-00021.1ff. ffhal-04852031f

Luebeck, EG. et al. Impact of Tumor Progression on Cancer Incidence Curves. Cancer Research, 73(3), 1086-96 (2013).

Marcinowski, F. et al. National and regional distributions of airborne radon concentrations in U.S. homes. Health Phys 66, 699–706 (1994).

NRC. National Research Council - Committee on Health Risks of Radon. Health Effects of Exposure to Radon: BEIR VI. (National Academy Press, Washington DC) (1999).

Pesch, B. et al. Cigarette smoking and lung cancer - Relative risk estimates for the major histological types from a pooled analysis of case - Control studies. Int J Cancer 131, 1210–1219 (2012).

Petermann, E. et al. Development of a High-Resolution Indoor Radon Map Using a New Machine Learning-Based Probabilistic Model and German Radon Survey Data. Environ Health Perspect 132, 97009 (2024).

Rage, E. et al. Mortality analyses in the updated French cohort of uranium miners (1946–2007). International archives of occupational and environmental health 88(6), 717-730 (2015)

Richardson, S. and Gilks, WR. A Bayesian approach to measurement error problems in epidemiology using conditional independence models. Am J Epidemiol, 138(6) :430-42 (1993)

Richardson, D. B. et al. Lung Cancer and Radon: Pooled Analysis of Uranium Miners Hired in 1960 or Later. Environ Health Perspect 130, (2022).

Simonato, L. et al. Lung cancer and cigarette smoking in Europe: An update of risk estimates and an assessment of inter-country heterogeneity. Int J Cancer 91, 876–887 (2001).

Simonetto, C. et al. Shape-specific characterization of colorectal adenoma growth and transition to cancer with stochastic cell-based models. PLoS Comput Biol. 2023 Jan 23;19(1):e1010831. doi: 10.1371/journal.pcbi.1010831. PMID: 36689547; PMCID: PMC9894544.

Thiébaut, A.C.M. and Benichou, J. Choice of time-scale in cox’s model analysis of epidemiologic cohort data: a simulation study. Statistics in Medicine, 23 (2004).

Thun, M.J. et al. Age and the Exposure-Response Relationships Between Cigarette Smoking and Premature Death in Cancer Prevention Study II. Changes in cigarette-related disease risks and their implications for prevention and control. Smoking and Tobacco Control Monograph No. 8 383–413 (1997).

Tirmarche M. et al. α-risk final Scientific Report. Quantification of cancer and non-cancer risks associated with multiple chronic radiation exposures: Epidemiological studies, organ dose calculation and risk assessment. ALPHA-RISK Final Activity Report-Publishable 19122009_V1 2dl (2009).

Tomasek L. Lung cancer mortality among czech uranium miners-60 years since exposure. Journal of radiological protection. Official journal of the Society for Radiological Protection, 32(3) :301—314, (2012)

UNSCEAR. Effects of Ionizing Radiation. Vol. II, Annex E. (United Nations, New York) (2008).