Supplementary Material for

Risk of lung adenocarcinoma from smoking and radiation arises in different molecular pathways

Noemi Castelletti, Jan Christian Kaiser, Cristoforo Simonetto, Kyoji Furukawa, Helmut Küchenhoff and Georgios T. Stathopoulos

Imputation of smoking information in the LSS cohort

While data on radiation dose were virtually complete in this cohort, data on smoking histories were available for about 60% of the members, which were mainly obtained from a series of mail surveys (at most three surveys for each individual) conducted between 1965 and 1991. The questionnaire at each data collection included a question about the smoking status ('never', 'current' or 'past' categories). In addition, the age at starting smoking and the intensity (the average number of cigarettes smoked per day) were asked to ever-smokers and the age at quitting to past-smokers. For analyses, multiple survey sources of smoking data were combined and summarized as a relatively simple set of variables (start age, intensity, and quit age) to describe the smoking history for each individual having any information of smoking [2, 3]. In total, about 40% of the study subjects had no information on smoking habits due to non-response to all surveys, and the proportion of subjects with missing smoking data varied depending on sex, birth-year and radiation exposure [4]. In addition, the majority (about 60%) of the survey respondents had data from only one or two sources. In the earlier analyses [1, 2, 3, 5, 7], subjects with no smoking data were treated as having 'unknown' smoking status throughout their time at risk. For those with smoking data, the smoking status during the period up to the first survey response was treated as 'unknown', and the status at the last response was carried forward to the end of follow-up. With a concern on potential impact of the incompleteness in smoking data, Furukawa et al.(2014) [4] applied a common missing data approach of multiple imputation (e.g., [9]) to individual smoking histories in the LSS cohort and used the imputed data in analysis to evaluate the joint effects of radiation and smoking on the lung cancer incidence. In the current analysis we used these imputed data. A data set with a longer follow-up 1958-2009 could not be used since it lacked information on histological types, and smoking imputation was not performed with sufficient detail [1]. Supplementary Table S1 summarizes the LSS cohort data broken down by sex and smoking status. The cases in the whole cohort have been diagnosed at mean age 68.5 yr with mean radiation dose of 0.161 Gy and mean age at radiation exposure of 30.5 yr.

The descriptive model

The choice of the descriptive model is motivated by results of Egawa et al.(2012) [2]. Based on the Akaike Information Criterion (AIC) additive action of smoking ($S = S(packyr, smkdyr, smkdqyr, smkint)$) and radiation ($R = R(D)$) is slightly favored. This action leads to a total hazard function

$$h = h_0(1 + \rho(R) + \Psi(S)) \tag{1}$$

which applies the baseline hazard h_0 and the corresponding Excess Relative Risks (ERRs) $\rho(R)$ and $\Psi(S)$ according to

$$h_0 = \exp\left[\beta_1 + \beta_2 \cdot (city - 1) + \beta_3 \cdot agexp + \beta_4 \cdot \ln(age) + \beta_5 \cdot \ln(age)^2\right], \tag{2}$$

$$\rho(R) = \beta_6 \cdot D \cdot \exp\left[\beta_7 \cdot \ln(age)\right], \tag{3}$$

$$\Psi(S)_{f,m} = \beta_8 \cdot packyr \cdot \exp\left[\beta_9 \cdot smkdyr + \beta_{10} \cdot \ln(smkdqyr + 1) + \beta_{11f,m} \cdot smkint\right]. \tag{4}$$

The baseline hazard depends on city of residence (Hiroshima or Nagasaki), age at exposure and attained age. The radiation-related ERR $\rho(R)$ depends linearly on the lung dose and is modified by attained age.

The smoking-related ERR $\Psi(S)$ depends linearly on the cumulative smoking amount and is modified by years smoked, years since quit smoking, and smoking intensity. Only the last modifier was found to be sex-dependent.

The meaning and the values of the fitted parameters of the model are presented in Supplementary Tables S3 and S6.

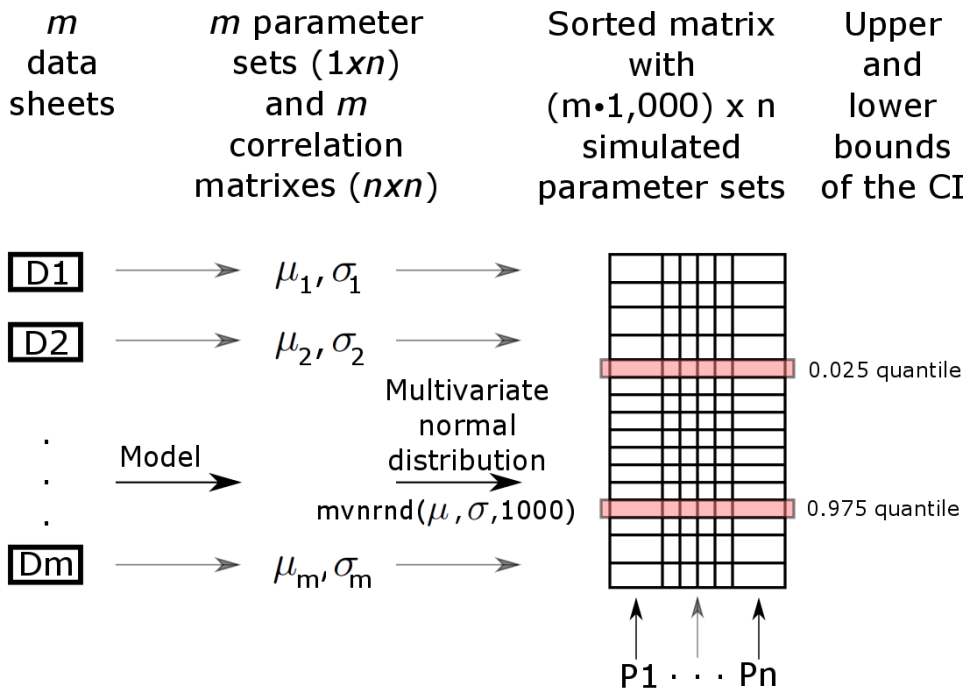
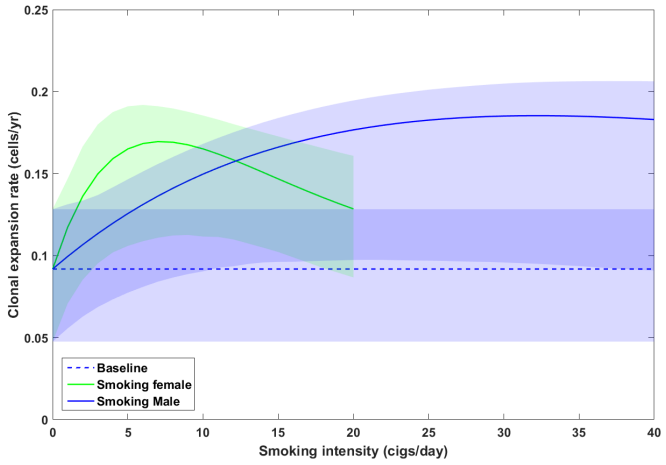


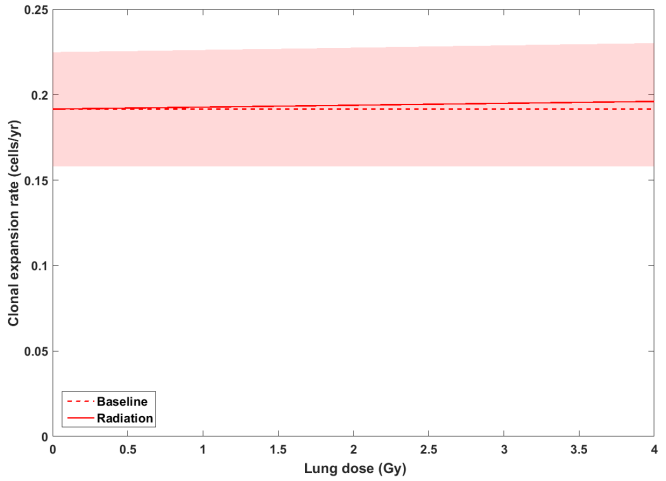
Figure S1: Flow chart for the calculation of confidence intervals for a multiple imputation overall point estimate. The gold standard for merging parameter estimates from imputed data sheets into one multiple imputation (MI) estimate is the so-called Rubin's rule [6, 8]. We extend Rubin's rule by including the parameter correlation matrix in the calculation of confidence intervals (CIs) for overall point estimates from MI.

The flow chart for the calculation is depicted in Figure S1. As in the Rubin's rule, from 50 data sheets 50 parameter sets of maximum likelihood estimates were fitted by Poisson regression and the respective correlation matrices were retained. The MI overall point estimates for model parameters and risk quantities of interest are then calculated as the means of 50 maximum likelihood estimates. To derive CIs let us consider only one dataset $D1$ with maximum likelihood estimate μ_1 and correlation matrix σ_1. From a multivariate normal distribution with inputs μ_1 and σ_1 1000 simulated parameter sets are simulated. These procedure is applied to all 50 dataset, ending up with $50 \cdot 1000 = 50000$ simulated parameter sets. The parameter sets are merged in a single matrix with 50000 rows and as many columns as the number of parameters contained in the model. A frequency distribution of 50000 entries for the risk quantities of interest such as EAR and ERR will yield the CIs as quantiles 0.025 and 0.975.

In this extension of the Rubin's rule, all components presented in the Rubin's rule itself are used. The within and the between imputation variances play both a role in the simulation of the parameter sets due to the correlation matrix. Since all parameters are hence merged into a single matrix and the quantiles are taken, number of imputed data sheets and variation in the imputation keep playing a role.

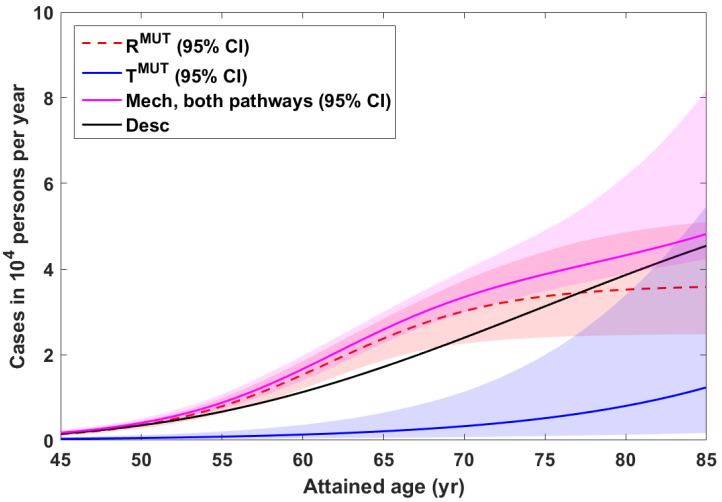


(A) T^{MUT} pathway

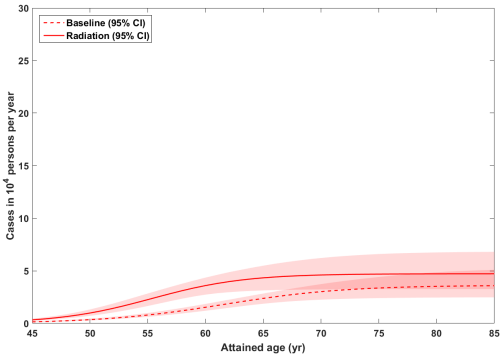


(B) R^{MUT} pathway

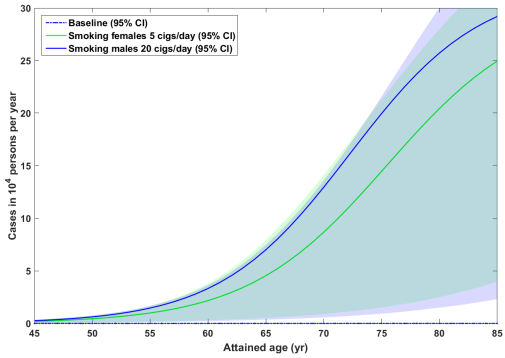
Figure S2: Clonal expansion rates for the two pathways T^{MUT} and R^{MUT} in M_3^{LADC}. (A) In the T^{MUT} pathway smoking intensity *smkint* linearly enhances the clonal expansion rate $\gamma_T(smkint) = \gamma_{T0}\,[1 + g_{Sm,f}\,smkint\,\exp(-\kappa_{m,f}\,smkint)]$ with an attenuated effect for high smoking intensity. The implausibly strong attenuation of the clonal expansion rate for females smoking more the 10 cigs/day is possibly caused by a reporting bias. (B) In the R^{MUT} pathway a radiation dose D linearly enhances the clonal expansion rate $\gamma_R(D) = \gamma_{R0}\,[1 + g_R\,D\,H(a - a_e)]$. The Heaviside function $H(a - a_e)$ equals one if attained age a exceeds age at exposure a_e and zero otherwise. It ensures that clonal expansion remains permanently elevated after exposure.



(A) Baseline hazards

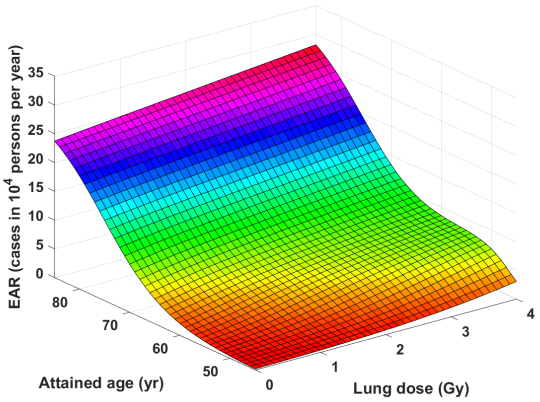


(B) Baseline hazard and hazard from radiation exposure in the R^{MUT} pathway

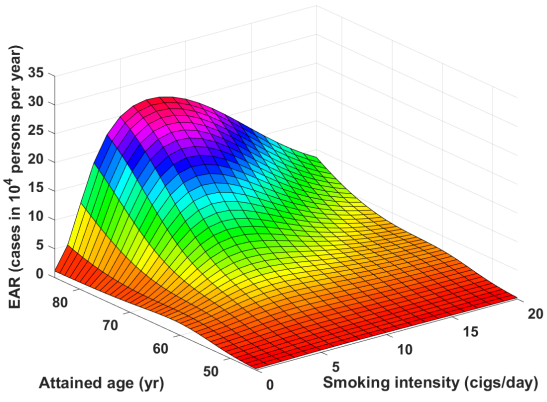


(C) Baseline hazard and hazard from smoking in the T^{MUT} pathway

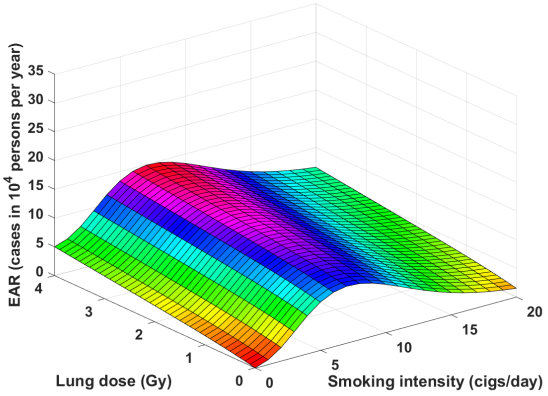
Figure S3: (A) Baseline hazards in pathways R^{MUT} and T^{MUT} for radiation-induced lung adenocarcinoma in the Japanese life span study cohort. To eliminate the influence for city of residence, person-year weighted city means are used. For comparison with the baseline hazard from a descriptive radio-epidemiological model (Desc) the total baseline hazard (as the sum of pathway-specific hazards) from the preferred mechanistic model M_3^{LADC} is shown. (B) Baseline hazard and hazard from radiation exposure in the R^{MUT} pathway for a person exposed at 30 yr to a lung dose of 1 Gy (C) Baseline hazard and hazard from smoking in the T^{MUT} pathway for lifelong smokers starting at age 20 yr with smoking intensity 20 cigs/day (male) and 5 cigs/day (female).



(A) Females with smoking intensity 5 cigs/day

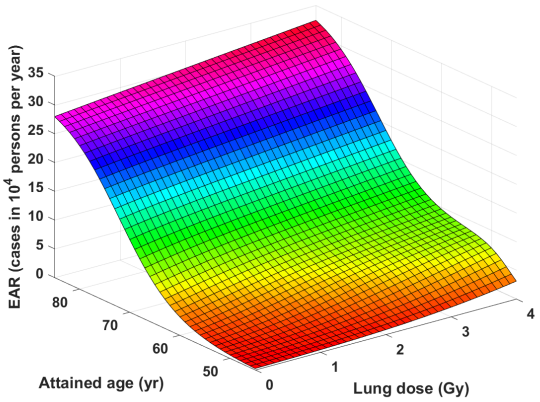


(B) Smoking females with lung dose 1 Gy

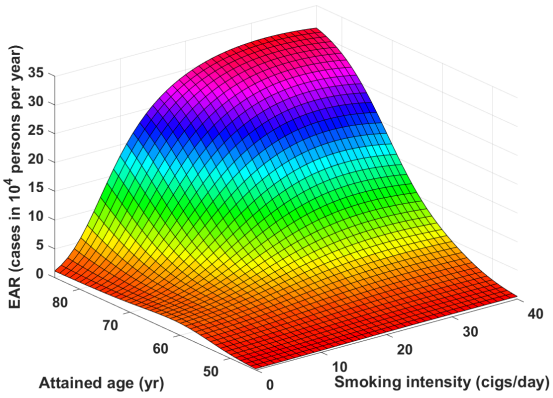


(C) Smoking females at attained age 70 yr exposed to radiation at age 30 yr

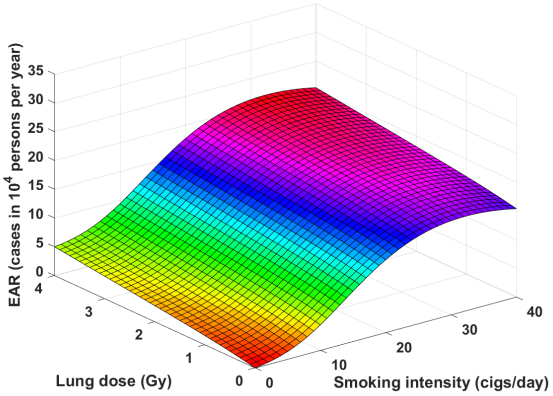
Figure S4: Bivariate Excess Absolute Rates (EARs as cases in 10,000 persons per year) for smoking-induced and radiation-induced lung adenocarcinoma in the Japanese life span study cohort from the preferred mechanistic model M_3^{LADC} for lifelong female smokers starting at age 20 yr and exposed to radiation at age 30 yr. To eliminate the influence for city of residence, person-year weighted city means are used. (A) Dependence on attained age and lung dose for comparison with Figure 5A. (B) Dependence on attained age and smoking intensity. For a lung dose of 1 Gy comparison with Figure 4A reveals that the radiation effect on the EAR is negligible. (C) Additive effect of radiation and smoking at attained age 70 yr.



(A) Males with smoking intensity 20 cigs/day

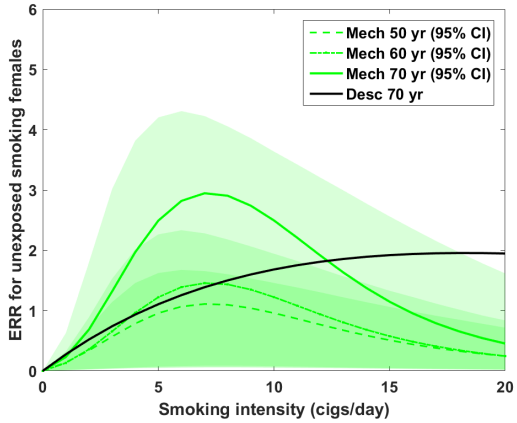


(B) Smoking males with lung dose 1 Gy

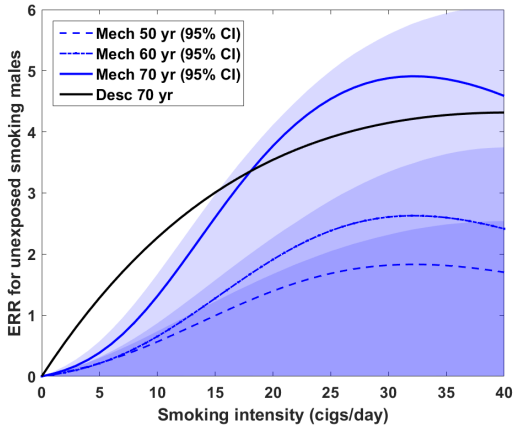


(C) Smoking males at attained age 70 yr exposed to radiation at age 30 yr

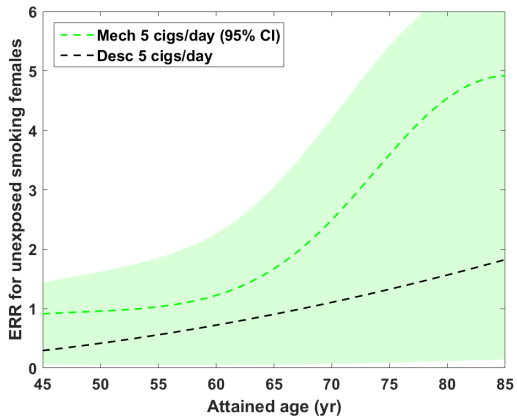
Figure S5: Bivariate Excess Absolute Rates (EARs as cases in 10,000 persons per year) for smoking-induced and radiation-induced for lung adenocarcinoma in the Japanese life span study cohort from the preferred mechanistic model M_3^{LADC} for lifelong male smokers starting at age 20 yr and exposed to radiation at age 30 yr. To eliminate the influence for city of residence, person-year weighted city means are used. (A) Dependence on attained age and lung dose for comparison with Figure 5A. (B) Dependence on attained age and smoking intensity. For a lung dose of 1 Gy comparison with Figure 4B reveals that the radiation effect on the EAR is negligible especially for heavy smokers. (C) Additive effect of radiation and smoking at attained age 70 yr.



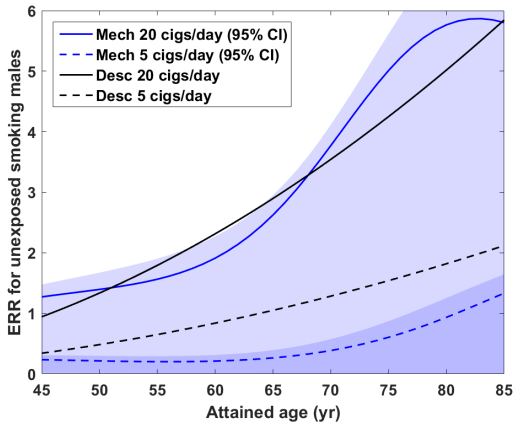
(A) Lifelong female smokers starting at age 20 yr



(B) Lifelong male smokers starting at age 20 yr

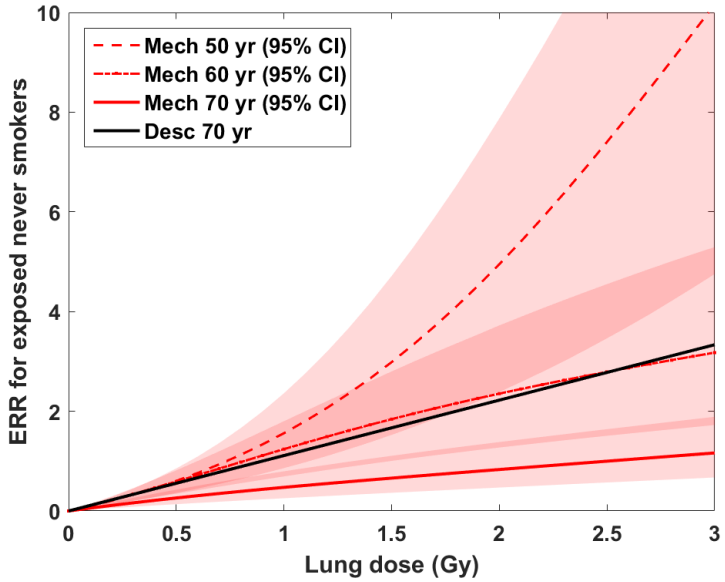


(C) Lifelong female smokers of 5 cigs/day starting at age 20 yr

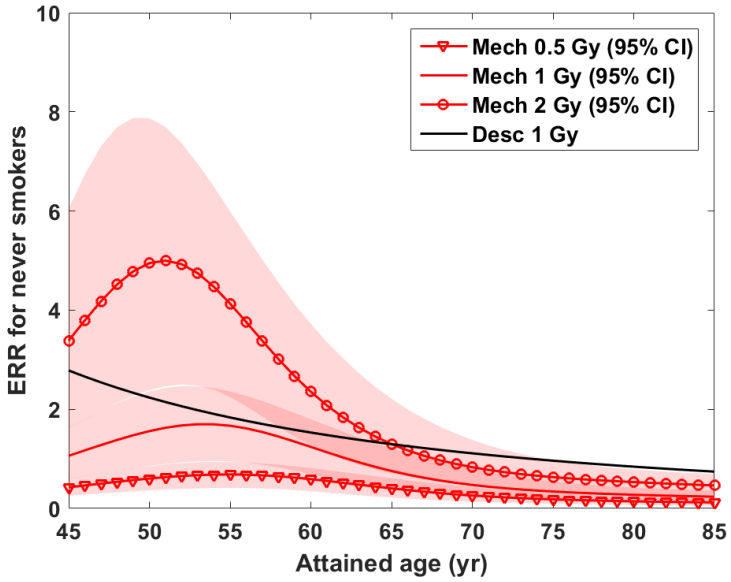


(D) Lifelong male smokers of 20 cigs/day starting at age 20 yr

Figure S6: Excess Relative Risk (ERR) from the preferred mechanistic model M_3^{LADC} (Mech) for smoking-induced lung adenocarcinoma in the Japanese life span study cohort for lifelong smokers starting at age 20 yr. Smoking only affects the T^{MUT} pathway with a markedly different response in both sexes but independent of radiation exposure. Panels (A) and (B) depict the ERR for attained ages of 50, 60 and 70 yr which is determined by the sex-dependent linear-exponential response of the clonal expansion rate in the T^{MUT} pathway. The implausibly strong attenuation of the ERR for females smoking more the 10 cigs/day is possibly caused by a reporting bias. Panels (C) and (D) depict the ERR over age for 5 cigs/day (males and females) and 20 cigs/day (males only). Female smokers of 5 cigs/day and males smokers of 20 cigs/day possess about the same risk. ERR estimates from a descriptive radio-epidemiological risk model (Desc) are shown for comparison.



(A) Never smokers exposed at age 30 yr



(B) Never smokers exposed at age 30 yr

Figure S7: Excess Relative Risk (ERR) from the preferred mechanistic model M_3^{LADC} (Mech) for lung adenocarcinoma in the Japanese life span study cohort for never smokers exposed to radiation at 30 yr. Radiation only affects the R^{MUT} pathway independent of sex and smoking status. (A) For attained ages 50, 60 and 70 yr the ERR responds *non-linearly* to doses in the range 0-4 Gy. However, on the biological level radiation action is modeled by a *linear* increase of the clonal expansion rate in the R^{MUT} pathway which lasts for life. (B) For lung doses of 0.5, 1 and 2 Gy the ERR from the preferred mechanistic model peaks at decreasing age with increasing value. The ERR estimate at 1 Gy from a descriptive radio-epidemiological model (Desc) is shown for comparison.

	Women			Men		
Smoking status	**Never**	**Past**	**Current**	**Never**	**Past**	**Current**
Cases (% of 636)	234 (37)	19 (3)	68 (21)	23 (4)	72 (11)	220 (35)
Age at diagnosis (years)	68.1	72.0	70.3	71.1	69.9	67.7
Age at begin smoking (years)	-	32.8	34.4	-	21.6	21.9
Smoking duration (years)	-	29.2	36.0	-	34.5	45.7
Cumulative smoking amount (pack-years)	-	14.5	16.8	-	32.4	43.3
Smoking intensity (cigarettes/day)	-	9.1	9.3	-	19.2	19.4
Age at radiation exposure (years)	29.6	32.4	34.8	30.4	27.6	31.1
Radiation dose (Gy)	0.190	0.076	0.229	0.012	0.081	0.157

Table S1: Summary of mean values for age and exposure-related co-variables of the Japanese life span study cohort data broken down by smoking status. For smoking related co-variables the measmns are taken over 50 data sets with imputed smoking information.

Imputed data set no.	**AIC of candidate models**			
	One-path TSCE models		*Two-paths models*	
	rad and smk. in promotion	rad and smk. in promotion, smk. in initiation	$TSCE_R$-$3SCE_T$	$TSCE_R$-$TSCE_T$
2	5138.1	5124.7	5100.9	**5099.6**
9	5099.0	5090.7	5061.4	**5060.9**
11	5156.2	5139.9	5106.3	**5106.1**
18	5106.4	5096.7	**5057.3**	5057.8
23	5011.7	5002.7	**4961.7**	4962.7
28	5172.0	5158.5	**5123.5**	5124.1
39	5024.6	5014.6	**4973.4**	4974.2
43	5149.3	5137.3	5109.5	**5108.7**
45	5102.8	5091.8	**5047.7**	5048.6
50	5106.7	5093.2	5063.1	**5062.2**
Cumulative AIC	51066.8	50950.1	**50604.8**	**50604.9**

Table S2: A large number of multi-stage models has been tested. To speed up the selection process, model were ïňĄtted to 10 (out of 50) randomly chosen LSS cohort data sets with imputed smoking information. Model selection was based on goodness-of-ïňĄt measured by the cumulative AIC = deviance + × ÃŮ no. of parameters for the 10 data sets. Only the main results with statistically significant parameter estimates on a 95% confidence level are shown. We started with one-path two stage clonal expansion (TSCE) models applying action of smoking and radiation in promotion only, and for smoking acting additionally in initiation. Initiating radiation action did not improve fits of one-path TSCE models. Two-path models were tested in pairs of TSCE and three stage clonal expansion (3SCE) models. For the R^{MUT} pathway (subscript $_R$) only a TSCE model survived the test phase. For the T^{MUT} pathway (subscript $_T$) a TSCE model and 3SCE model yielded the same AIC, when paired with the TSCE model for the R^{MUT} pathway. Compared to the TSCE model the 3SCE model contains an additional mutational stage before clonal expansion but has the same number of parameters. The TSCE model was chosen for the T^{MUT} pathway because it required substantially less computation time, nevertheless the 3SCE model is also biologically plausible and cannot be excluded. The impact of this choice on the results is negligible. The preferred $TSCE_R$-$TSCE_T$ model is termed in M_3^{LADC} in the main paper.

Variable	**Unit**	**Meaning**
city	-	Hiroshima (1) or Nagasaki (2)
agexp	-	(age at exposure - 30 yr)/10 yr
age	-	attained age/70 yr
D	Gy	lung dose
packyr	packs/day × yr	cumulative amount of cigarette packs (packs smoked per day × years smoked)/50 yr
smkdyr	-	years smoked/50 yr
smkdqyr	-	years since quit smoking/50 yr
smkint	cigs/day	smoking intensity (cigs smoked per day)

Table S3: Explanatory variables for descriptive and mechanistic models. In the baseline hazard age at exposure is equivalent to birth year (birth year = 1945.7 - age at exposure). A pack contains 20 cigarettes. In both descriptive and mechanistic models the only sex-dependent parameters, $\beta_{11f,m}$ and $\gamma_{Sf,m}$ and $\kappa_{f,m}$, respectively, are related to smoking intensity. For the other parameters the sex-difference was found to be not statistically significant based on likelihood ratio tests on the 95% level.

Parameter estimates of the preferred mechanistic model M_3^{LADC}			
		R^{MUT} pathway	T^{MUT} pathway
Name	Unit	Mean value (95% CI)	
c_{city}		0.23 (0.11, 0.35)	
c_{agexp}		-0.24 (-0.29, -0.18)	
X_R, X_T	10^{-9} cells/yr^2	0.48 (0.11, 2.26),	4.64 (1.20, 18.9)
γ_{R0}, γ_{T0}	cells/yr	0.19 (0.16, 0.22),	0.092 (0.048, 0.128)
g_R	10^{-2} Gy^{-1}	0.58 (0.39, 0.77)	-
g_{Sf}	day/cigs	-	0.32 (0.057, 0.68)
g_{Sm}	day/cigs	-	0.086 (0.013, 0.180)
κ_f	day/cigs	-	0.14 (0.21, 0.078)
κ_m	day/cigs	-	0.031 (0.042, 0.021)
δ	10^{-7} cells/yr^2	2.73 (0.92, 8.06)	
Cumulative AIC		253720	

Table S4: Parameter estimates (95% CI) for the preferred mechanistic model M_3^{LADC} with 12 parameters which consists of two TSCE models pertaining to pathways R^{MUT} (subscript $_R$) and T^{MUT} (subscript $_T$). Parameter definitions correspond to Figure 3A. Central estimates are given as mean values from 50 imputed data sets with 95% CI simulated from multi-variate normal uncertainty distributions conditioned on the parameter correlation matrix. Parameters $X_R = N\nu_R\mu_R$ and $X_T = N\nu_T\mu_T$ (where N is the number of healthy cells) are modified by city and age at exposure with the same functional form as in the descriptive model. In the R^{MUT} pathway the radiation-dependent clonal expansion rate γ_R is given by $\gamma_R(D) = \alpha_R - \beta_R(D) - \mu_R = \gamma_{R0}\,[1 + g_R\,D\,H(a - a_e)]$. The Heaviside function $H(a - a_e)$ equals one if attained age a exceeds age at exposure a_e and zero otherwise. Thus, the clonal expansion rate $\gamma_R(D)$ remains enhanced for life after being linearly increased by an acute radiation dose D. The smoking-dependent expansion rate $\gamma_T(smkint) = \alpha_T - \beta_T(smkint) - \mu_T = \gamma_{T0}\,[1 + g_{Sf,m}\,smkint\,\exp(-\kappa_{f,m}\,smkint)]$ increases linearly for low smoking intensity $smkint$. For high smoking intensity exponential attenuation sets in. The smoking parameters $g_{Sm,f}$ and $\kappa_{m,f}$ depend strongly on sex. The leveling parameter $\delta = \alpha_R\mu_R = \alpha_T\mu_T$ determines the leveling of the hazard at old ages and is the same in both pathways. The cumulative AIC from 50 imputed data sets is 370 points lower compared to the descriptive model (corresponding to 7.4 point per data set).

Backward recursion algorithm for the hazard function in a single pathway	
$h(t) = \sum_{i=1}^{k} \frac{X_i}{\delta_i} \frac{1}{f_i(s_{i-1})} \frac{\partial}{\partial t} f_i(s_{i-1})$	
$\delta_i = \delta$	
$X_i = X_{R,T} \cdot \mathrm{e}^{c_{agexp}\ agexp + c_{city}(city-1)}$ (3SCE model: $X_{R,T} \to X'_{R,T}\ s_i$)	
$A_i = -\frac{1}{2}(\gamma_i + \sqrt{\gamma_i^2 + 4\delta_i\theta_i})$	
$B_i = \frac{1}{2}(-\gamma_i + \sqrt{\gamma_i^2 + 4\delta_i\theta_i})$	
$\theta_i = \frac{\mu_i}{\mu_0}$	
$\gamma_i = \gamma_{R0}(1 + g_R\, D\, H(s_i - a_e))$ or $\gamma_{T0}(1 + g_{Sf,m}\ smkint_i\ \mathrm{e}^{-\kappa_{f,m}\ smkint_i})$	
$f_i(s_{i-1}) = B_i \mathrm{e}^{A_i(s_{i-1}-t)} - A_i \mathrm{e}^{B_i(s_{i-1}-t)}$	first step $i = k$
$f_i(s_{i-1}) = (B_i - w_i(s_i))\mathrm{e}^{A_i(s_{i-1}-s_i)} + (w_i(s_i) - A_i)\mathrm{e}^{B_i(s_{i-1}-s_i)}$	all steps $i \neq k$
$\frac{\partial}{\partial t} f_i(s_{i-1}) = A_i B_i (\mathrm{e}^{B_i(s_{i-1}-t)} - \mathrm{e}^{A_i(s_{i-1}-t)})$	first step $i = k$
$\frac{\partial}{\partial t} f_i(s_{i-1}) = -\frac{\partial}{\partial t} w_i(s_i)(\mathrm{e}^{A_i(s_{i-1}-s_i)} - \mathrm{e}^{B_i(s_{i-1}-s_i)})$	all steps $i \neq k$
$w_i(s_i) = 0$	first step $i = k$
$\frac{\partial}{\partial t} w_i(s_i) = \delta_i \theta_i$	first step $i = k$
$w_{i-1}(s_{i-1}) = \frac{\delta_{i-1}}{\delta_i} w_i(s_{i-1})$	all steps $i \neq k$
$\frac{\partial}{\partial t} w_{i-1}(s_{i-1}) = \frac{\delta_{i-1}}{\delta_i} \frac{\partial}{\partial t} w_i(s_{i-1})$	all steps $i \neq k$
$w_i(s_{i-1}) = \frac{A_i B_i (\mathrm{e}^{A_i(s_{i-1}-t)} - \mathrm{e}^{B_i(s_{i-1}-t)})}{B_i \mathrm{e}^{A_i(s_{i-1}-t)} - A_i \mathrm{e}^{B_i(s_{i-1}-t)}}$	first step $i = k$
$w_i(s_{i-1}) = \frac{A_i B_i (\mathrm{e}^{A_i(s_{i-1}-s_i)} - \mathrm{e}^{B_i(s_{i-1}-s_i)}) - w_i(s_i)(A_i \mathrm{e}^{A_i(s_{i-1}-s_i)} - B_i \mathrm{e}^{B_i(s_{i-1}-s_i)})}{(B_i - w_i(s_i))\mathrm{e}^{A_i(s_{i-1}-s_i)} + (w_i(s_i) - A_i)\mathrm{e}^{B_i(s_{i-1}-s_i)}}$	all steps $i \neq k$
$\frac{\partial}{\partial t} w_i(s_{i-1}) = \frac{A_i B_i (A_i - B_i)^2 \mathrm{e}^{A_i(s_{i-1}-t)} \mathrm{e}^{B_i(s_{i-1}-t)})}{(B_i \mathrm{e}^{A_i(s_{i-1}-t)} - A_i \mathrm{e}^{B_i(s_{i-1}-t)})^2}$	first step $i = k$
$\frac{\partial}{\partial t} w_i(s_{i-1}) = \frac{\partial}{\partial t} w_i(s_i) \frac{(A_i - B_i)^2 \mathrm{e}^{A_i(s_{i-1}-s_i)} \mathrm{e}^{B_i(s_{i-1}-s_i)}}{[(B_i - w_i(s_i))\mathrm{e}^{A_i(s_{i-1}-s_i)} + (w_i(s_i) - A_i)\mathrm{e}^{B_i(s_{i-1}-s_i)}]^2}$	all steps $i \neq k$

Table S5: Recursion equations for the hazard $h(t)$ at age t of the TSCE model and the 3SCE model with piecewise constant parameters in k age intervals.

Parameter estimates of the descriptive model		
Name	Meaning	Mean (95% CI)
β_1	baseline	1.07 (0.95, 1.18)
β_2	baseline, city	0.23 (0.11, 0.35)
β_3	baseline, age at exposure	-0.26 (-0.32, -0.21)
β_4	baseline, attained age	4.19 (3.47, 4.90)
β_5	baseline, attained age (squared)	-4.58 (-6.08, -3.08)
β_6	radiation, linear resp. (Gy^{-1})	1.11 (0.62, 1.60)
β_7	radiation, attained age	-2.08 (-3.92, -0.23)
β_8	smoking, linear resp. (day $\times$ yr/packs)	5.82 (3.38, 8.57)
β_9	smoking, years smoked	0.91 (-0.18, 2.06)
β_{10}	smoking, years since quitting	-0.33 (-0.63, -0.08)
β_{11f}	smoking, smoking intensity females	-0.055 (-0.094, -0.018)
β_{11m}	smoking, smoking intensity males	-0.025 (-0.045, -0.005)
Cumulative. AIC	254090	

Table S6: Parameter estimates for the descriptive model with 12 parameters. Central estimates are given as means from 50 imputed data sets with 95% CI simulated from multi-variate normal uncertainty distributions conditioned with the parameter correlation matrix.

References

[1] E. K. Cahoon, D. L. Preston, D. A. Pierce, E. Grant, A. V. Brenner, K. Mabuchi, M. Utada, and K. Ozasa. Lung, laryngeal and other respiratory cancer incidence among japanese atomic bomb survivors: An updated analysis from 1958 through 2009. Radiat Res, 187(5):538–548, 2017.

[2] H. Egawa, K. Furukawa, D. Preston, S. Funamoto, S. Yonehara, T. Matsuo, S. Tokuoka, A. Suyama, K. Ozasa, K. Kodama, and K. Mabuchi. Radiation and smoking effects on lung cancer incidence by histological types among atomic bomb survivors. Radiat Res, 178(3):191–201, 2012.

[3] K. Furukawa, D. L. Preston, S. Lonn, S. Funamoto, S. Yonehara, T. Matsuo, H. Egawa, S. Tokuoka, K. Ozasa, F. Kasagi, K. Kodama, and K. Mabuchi. Radiation and smoking effects on lung cancer incidence among atomic bomb survivors. Radiat Res, 174(1):72–82, 2010.

[4] K. Furukawa, D. L. Preston, M. Misumi, and H. M. Cullings. Handling incomplete smoking history data in survival analysis. Stat Methods Med Res, 2014.

[5] E. J. Grant, A. Brenner, H. Sugiyama, R. Sakata, A. Sadakane, M. Utada, E. K. Cahoon, C. M. Milder, M. Soda, H. M. Cullings, D. L. Preston, K. Mabuchi, and K. Ozasa. Solid cancer incidence among the life span study of atomic bomb survivors: 1958-2009. Radiat Res, 187(5):513–537, 2017.

[6] A. Marshall, D. G. Altman, R. L. Holder, and P. Royston. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. BioMediacal Central Medical Research Methodology, 2009.

[7] D. A. Pierce, G. B. Sharp, and K. Mabuchi. Joint effects of radiation and smoking on lung cancer risk among atomic bomb survivors. Radiat Res, 159(10):511–20, 2003.

[8] D.B. Rubin. Multiple Imputation for Nonresponse in Surveys New York. 2004.

[9] J. A. Sterne, I. R. White, J. B. Carlin, M. Spratt, P. Royston, M. G. Kenward, A. M. Wood, and J. R. Carpenter. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ, 338:b2393, 2009.