## Supplementary Information

# Biologically-based modeling of radiation risk and biomarker prevalence for papillary thyroid cancer in Japanese a-bomb survivors 1958 - 2005

by Jan Christian Kaiser, Munechika Misumi and Kyoji Furukawa

# Mechanistic model implementation

## MSC pathway for models M0 and M1

The concept for the mechanistic model of the MSC pathway (subscript $_{MSC}$) is depicted in the upper leg of Fig. 1. In the MSC pathway PTC occurs only sporadically according to the baseline hazard function of the stochastic TSCE model, which depends on attained age $a$ and sex $s$ with functional form

$$\begin{aligned} h_{0,MSC}(s,a) &= X_{MSC}\frac{\exp[(\gamma_{MSC}(s)+2q)a]-1}{\gamma_{MSC}(s)+q\,[\exp[(\gamma_{MSC}(s)+2q)a]+1]} \\ \text{with} \quad q &= \frac{1}{2}(\sqrt{\gamma_{MSC}(s)^2+4\delta_{MSC}}-\gamma_{MSC}(s)). \end{aligned} \tag{S1}$$

A mathematical derivation has been outlined earlier [1, 2].

The biological parameters $N_c$,$\nu_0$,$\nu_1$,$\alpha_{MSC}$ and $\beta_{MSC}$ in the upper leg of Fig. 1 are not directly identifiable from fits to incidence data [3]. They are mapped to a set of three identifiable baseline parameters

$$\begin{aligned} X_{MSC} &= N_c\nu_0\nu_1 \\ \gamma_{MSC} &= \alpha_{MSC}-\beta_{MSC}-\nu_1 \\ \delta_{MSC} &= \alpha_{MSC}\nu_1 \end{aligned} \tag{S2}$$

which belong to characteristic phases of cancer development [1, 2]. $X_{MSC}$ is termed Armitage-Doll parameter because the product of transition rates determines the incidence rate in the historic model of Armitage and Doll [4]. $N_c$ denotes the large but unknown number of thyrocytes assumed as susceptible stem cells on the pathway to cancer. The transition rates $\nu_0$ and $\nu_1$ appeal to the two mutational 'hits' necessary to arrive at fully developed PTC [5]. The complex carcinogenic processes after the creation of initiated clones are summarized by rate $\nu_1$.

In Figs. 1 and S1 the arrows curving downward denote the asymmetric division of an intermediate cell: one progenitor cells acquires an additional mutation and the second progenitor cells returns to the clone of intermediate cells. This division scheme constitutes a Poisson point process in contrast to an Armitage-Doll process where the second progenitor cell differentiates or dies. Although normal thyroid cells also divide according to Poisson processes the downward arrows are not shown. Here we assume a large

number of normal cells in homeostasis which remains effectively constant. The difference in cell balance between both processes is negligible for rare events. For a thorough discussion see e.g. Dewanji et al. [6].

The residual difference between symmetric cell division rate $\alpha_{MSC}$ and cell inactivation rate $\beta_{MSC}$ determines the net clonal expansion rate $\gamma_{MSC} \simeq \alpha_{MSC} - \beta_{MSC}$ since $\nu_1$ is usually some orders of magnitude smaller as $\gamma_{MSC}$. The clonal growth rate $\gamma_{MSC}$ is the only sex-dependent parameter in the MSC model. Values for sex $s$ are estimated according to

$$\gamma_{MSC}(s) = \gamma_{MSC} \, \exp\left(\pm p_{MSC}\right) \tag{S3}$$

with $+p_{MSC}$, $-p_{MSC}$ in females, males.

The stochasticity parameter $\delta_{MSC}$ determines the leveling of the hazard at old age. Leveling can be interpreted as a *frailty* effect of inter-personal variation. Persons burdened with large clones of intermediate cells are prone to acquire PTC at younger age and drop out of the group with persons under risk earlier. At old age the number of persons with large clones approaches a saturated equilibrium, which is reflected in the hazard [7, 8].

The complete hazard function

$$h_{MSC}(s, a, b) = \exp\left[b_{byr} \frac{(1935.6 - b)}{10}\right] \, h_{0,MSC}(s, a) \tag{S4}$$

in the MSC pathway is obtained by adjusting for birth year $b$ in the same way as for descriptive models. The MSC pathway consumes five adjustable parameters.

## C2C pathway for model M1

The design of the C2C model (subscript $_{C2C}$) is guided by the observation that the CLIP2 marker has been detected in radiation-induced PTC at young age below 20 yr but also in older patients where it occurred *sporadically* [9]. Hence, biological evidence demands that a mechanistic model should be able to generate both sporadic and radiation-induced PTCs carrying the CLIP2 marker. To achieve the desired outcome several versions of the TSCE model with different radiation targets have been tested.

The blueprint for the preferred mechanistic model of the C2C pathway is shown in the lower leg of Fig. 1. It can be mathematically translated into a TSCE model with radiation action on clonal expansion. The identifiable baseline parameters

$$
\begin{aligned}
X_{C2C} &= N_b \mu_0 \mu_1 \\
\gamma_{C2C} &= \alpha_{C2C} - \beta_{C2C} - \mu_1 \\
\delta_{C2C} &= \alpha_{C2C} \mu_1 .
\end{aligned}
\tag{S5}
$$

have the same meaning as the parameters of the MSC pathway given in Eq. (S2).

Takano [10] explained decreasing radiation risk with increasing age at exposure by inactivation of thyroblasts as targets of radiation during adolescence. Consequently, waning radio-sensitivity is modeled by multiplying the logistic attenuation function for age at exposure $e$

$$
f_b(e) = \frac{1}{2} \left[ 1 + \tanh\left( \frac{e_{cen} - e}{b_{slp}} \right) \right] \tag{S6}
$$

to the total number of thyroblasts $N_b$.

To obtain a value of one at $e = 0$ the attenuation function is rescaled according to $f_b(e) \rightarrow f_b(e)/f_b(0)$. With increasing $e$, $f_b(e)$ approaches zero with a characteristic downward step of slope $1/b_{slp}$ around an inflection point at $e_{cen}$. The total cell number $N_b$ is not known at birth, and $f_b(e)$ describes the relative reduction of thyroblasts.

Radiation action is modeled by enhancing the baseline parameters of either initiation $\mu_0$ or promotion $\gamma_{C2C}$ linearly with dose. The preferred radiation target is determined by goodness-of-fit and biological plausibility. Enhancement from acute exposure is allowed to endure for a very short period of one week or to remain permanently until end of follow-up. The dose response is sex-dependent. No closed analytical form exists for the hazard function $h_{C2C}(s, a, e, D)$ of the C2C pathway. For piecewise constant parameters the hazard is obtained by solving a backward recursion relation for the TSCE model [11].

For the Chernobyl accident Heidenreich et al. [12] estimated a minimum lag period of three years after exposure for the onset of radiation risk. A functional form for the lag period cannot be derived from PTC incidence data of the LSS since follow-up started in 1958 about thirteen years after the bombings.

In the preferred model M1 radiation increases the clonal expansion rate of initiated cells

$$
\gamma_{C2C}(s, D) = \gamma_{C2C} \left[ 1 + \Theta(a - e) g_d \, D \, \exp(\pm p_{C2C}) \right] . \tag{S7}
$$

The Heaviside function $\Theta(a-e)$ equals one if attained age $a$ exceeds age at exposure $e$ and zero otherwise. Hence, $\gamma_{C2C}(s,D)$ is permanently enhanced after exposure by a constant value. The dose effect is modified by sex $s$ with parameter $+p_{C2C}$, $-p_{C2C}$ for females, males. We also tested an extension of model M1 which additionally applies radiation action according to Eq. (S7) in the MSC pathway. With this extension a complete confusion matrix can be generated to assess sensitivity and specificity of a potential radiation marker.

The hazard function $h_{C2C}(s,a,e,D)$ in the C2C pathway applies seven adjustable parameters.

## Rich mechanistic model M1

The total hazard function for rich mechanistic model M1 of Fig. 1

$$h_{M1}(s,a,e,b,D,AHS) = \exp(b_{AHS})\,[h_{MSC}(s,a,b) + h_{C2C}(s,a,e,D)] \tag{S8}$$

is adjusted by a factor of $\exp(b_{AHS})$ for AHS participation and uses thirteen fit parameters.

## Sparse mechanistic model M0

The blueprint for the sparse mechanistic model M0 is shown in Fig. S1. For the MSC pathway it uses the hazard function defined in Eq. (S4). In the lower leg of Fig. S1 a pathway for purely radiation-induced carcinogenesis (RIC) is shown which does not produce sporadic PTC.

The mathematical derivation of the hazard for the RIC pathway is fully described in [9] and briefly summarized below. Radiation exposure increases the first transition rate

$$\mu_0(a) = \begin{cases} \mu_0 & : \quad a < e \\ \mu_0 + \mu_r & : \quad e \le a < e + \Delta t_e \\ \mu_0 & : \quad a \ge e + \Delta t_e \end{cases} \tag{S9}$$

by $\mu_r$ for a short period $\Delta t_e$ (i.e. one week) after age at exposure $e$. Before and after exposure the first transition rate $\mu_0$ remains constant.

Integration of the corresponding master equation yields the radiation-induced hazard (supplement of [9], their Eq. (S5)).

$$h_{RIC} = X_{RIC}\, a + \Theta(a-e)\, X_{RIC}\, \frac{\mu_r}{\mu_0} \Delta t_e. \tag{S10}$$

$X_{RIC} = N_b\mu_0\mu_1$ is the Armitage-Doll parameter for radiation-induced PTC where $N_b$ denotes the number of thyroblasts susceptible to RIC. The Heaviside function $\Theta(a - e)$ equals one if attained age $a$ exceeds age at exposure $e$ and zero otherwise. The first term $X_{RIC}\, a$ of Eq. (S10) is some orders of magnitude smaller than the second term and can thus be neglected. The radiation response

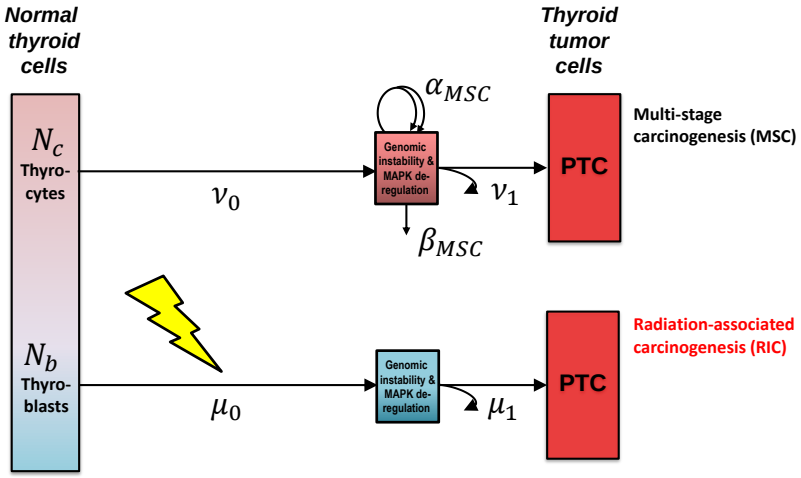


Figure S1: Conceptual model M0 for the development of sporadic PTC from multi-stage carcinogenesis (MSC) (upper path) and PTC from radiation-induced carcinogenesis (RIC) (lower path), PTC development starts from either $N_c$ thyrocytes in the MSC pathway or $N_b$ thyroblasts in the RIC pathway (left box), small central boxes in MSC and RIC pathways represent clones of intermediate cells with genomic instability and deregulated MAPK signaling; arrows between boxes indicate Poisson point processes for asymmetrically dividing cells using Greek letters $\nu, \mu$ as transition rates; $\alpha, \beta$ denote rates of symmetric cell division and inactivation in pre-neoplastic clones, radiation exposure (jagged yellow arrow) targets the first transition rate $\mu_0$ in the RIC pathway.

$$X_{RIC}\,\frac{\mu_r}{\mu_0}\,\Delta t_e = x_d\,\exp(\pm p_d)\,D \tag{S11}$$

increases linearly with sex-dependence $+p_d$, $-p_d$ for females, males and thyroid dose $D$. In the present parametrization of the mechanistic RIC model $x_d$ is expressed as an EAR per 10,000 PY.

After multiplying the scaled attenuation function $f_b(e)/f_b(0)$ of Eq. (S6) the mechanistic radiation hazard acquires the final form

$$h_{RIC}(s,e,D) = \Theta(a-e)\,x_d\,\exp(\pm p_d)\,D\,f_b(e)/(f_b(0) \tag{S12}$$

with four adjustable parameters $x_d$, $p_d$, $b_{slp}$ and $e_{cen}$.

The sparse mechanistic model M0 consists of two independent molecular pathways based on the corresponding hazards of Eqns. (S4) and (S12). The total hazard function is adjusted by AHS participation according to

$$h_{M0}(s,a,e,b,D,AHS) = \exp(b_{AHS})\,[h_{MSC}(s,a,b) + h_{RIC}(s,e,D)]\,. \tag{S13}$$

M0 tries to detect imprints of sporadic and radiation-induced pathways in the PTC incidence data and relies on the assumption of a perfect biomarker in the RIC pathway. In total M0 applies ten adjustable parameters, three parameters less than model M1.

## Supplementary Tables

Table S1: Summary of PTC incidence data from 1958-2005 for the Life Span Study (LSS), the Adult Health Study (AHS) as its clinical subset and the LSS without AHS.

| | females | males | both sexes |
|---|---|---|---|
| Life Span Study (LSS) | | | |
| subjects | 62511 | 42890 | 105401 |
| PTC cases | 248 | 44 | 292 |
| person years (PY) $\times 10^6$ | 1.881 | 1.124 | 3.005 |
| PY-weighted mean age at exposure (yr) | 23.5 | 20.4 | 22.2 |
| case-weighted mean age at exposure (yr) | 21.0 | 18.4 | 20.6 |
| PY-weighted mean thyroid dose (Gy) | 0.101 | 0.106 | 0.103 |
| case-weighted mean thyroid dose (Gy) | 0.289 | 0.255 | 0.284 |
| case-weighted mean attained age (yr) | 57.9 | 55.8 | 57.6 |
| case-weighted mean time since exposure (yr) | 36.9 | 37.4 | 37.0 |
| Adult Health Study (AHS) | | | |
| subjects | 10319 | 6357 | 16676 |
| PTC cases | 72 | 12 | 84 |
| person years (PY) $\times 10^6$ | 0.264 | 0.143 | 0.407 |
| PY-weighted mean age at exposure (yr) | 25.9 | 25.1 | 25.7 |
| case-weighted mean age at exposure (yr) | 21.3 | 14.1 | 20.3 |
| PY-weighted mean thyroid dose (Gy) | 0.349 | 0.391 | 0.364 |
| case-weighted mean thyroid dose (Gy) | 0.757 | 0.750 | 0.756 |
| case-weighted mean attained age (yr) | 55.8 | 46.7 | 54.6 |
| case-weighted mean time since exposure (yr) | 34.6 | 32.8 | 34.4 |
| LSS without AHS | | | |
| subjects | 52192 | 36533 | 88725 |
| PTC cases | 176 | 32 | 208 |
| person years (PY) $\times 10^6$ | 1.617 | 0.981 | 2.598 |
| PY-weighted mean age at exposure (yr) | 23.1 | 19.3 | 21.7 |
| case-weighted mean age at exposure (yr) | 20.9 | 20.0 | 20.8 |
| PY-weighted mean thyroid dose (Gy) | 0.061 | 0.064 | 0.062 |
| case-weighted mean thyroid dose (Gy) | 0.097 | 0.070 | 0.093 |
| case-weighted mean attained age (yr) | 58.7 | 59.2 | 58.8 |
| case-weighted mean time since exposure (yr) | 37.8 | 39.1 | 38.0 |

Table S2: Prediction of mechanistic models M0 and M1 for number of PTC cases, case-weighted arithmetic mean (standard deviation) for age at exposure, attained age, time since exposure and case-weighted arithmetic mean (geometric mean) for thyroid dose pertaining to the total LSS cohort, and broken down by molecular pathways MSC and RIC for model M0, and by molecular pathways MSC and C2C for model M1, in the C2C pathway PTC cases occur both sporadically (sC2C) and radiation-induced (rC2C), in the first row the crude estimates for the total LSS cohort are given for comparison with model results.

| total | PTC cases | age at exposure (yr) | attained age (yr) | time since exposure (yr) | thyroid dose (Gy) |
|---|---|---|---|---|---|
| crude | 292 | 20.6 (14.4) | 57.6 (14.9) | 37.0 (13.8) | 0.284 (0.125) |
| model M0 | 292.0 | 20.6 (13.8) | 57.4 (15.3) | 36.7 (13.3) | 0.304 (0.126) |
| model M1 | 292.0 | 20.6 (13.9) | 57.3 (15.2) | 36.7 (13.3) | 0.309 (0.129) |
| model M0 pathway | | | | | |
| [a]MSC | 247.6 | 23.4 (14.0) | 60.7 (13.4) | 37.3 (13.2) | 0.132 (0.043) |
| [b]RIC | 44.3 | 10.4 (6.7) | 44.2 (14.9) | 33.7 (13.5) | 1.27 (0.99) |
| model M1 pathway | | | | | |
| [a]MSC | 230.7 | 23.4 (14.0) | 60.7 (13.4) | 37.3 (13.1) | 0.131 (0.043) |
| C2C | 61.3 | 10.3 (6.8) | 44.4 (14.9) | 34.1 (13.5) | 0.979 (0.675) |
| [a]sC2C | 17.2 | 9.3 (6.5) | 44.4 (14.7) | 35.1 (13.6) | 0.124 (0.038) |
| [b]rC2C | 44.0 | 10.8 (6.8) | 44.5 (15.0) | 33.7 (13.5) | 1.31 (1.04) |

[a] sporadic
[b] radiation-induced

## Supplementary Figures

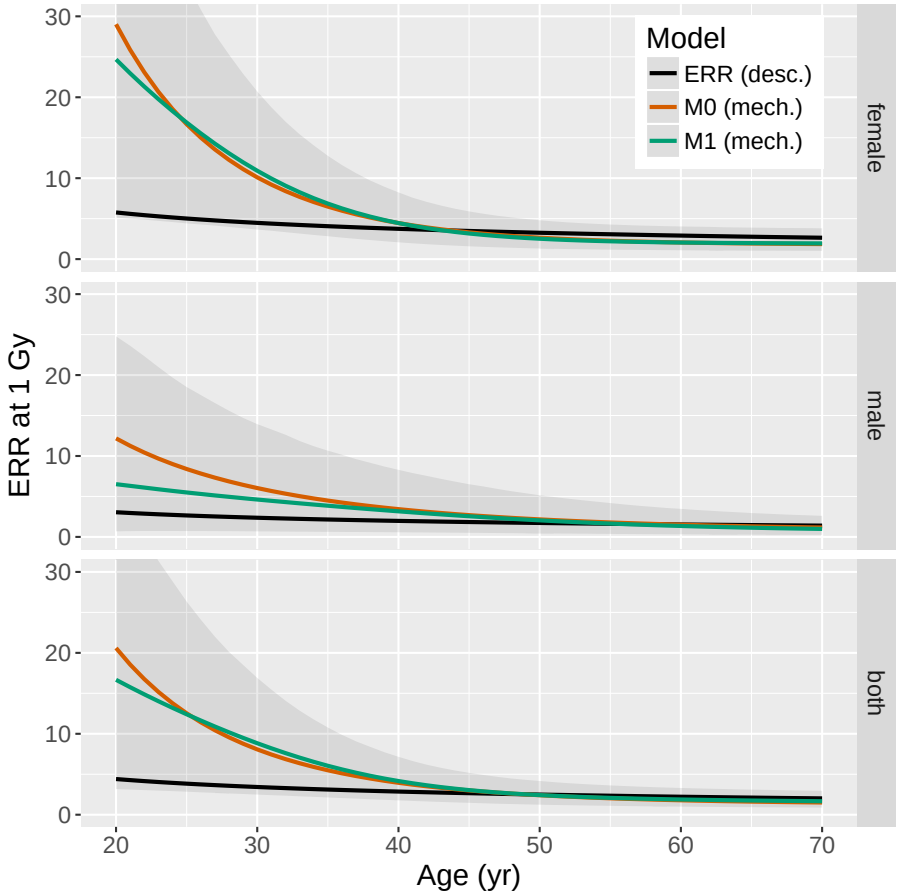


Figure S2: ERR at 1 Gy and age at exposure 8 yr for females, males and averaged for both sexes from the descriptive ERR model and mechanistic models M0 and M1, shaded 95% CI from model M1.

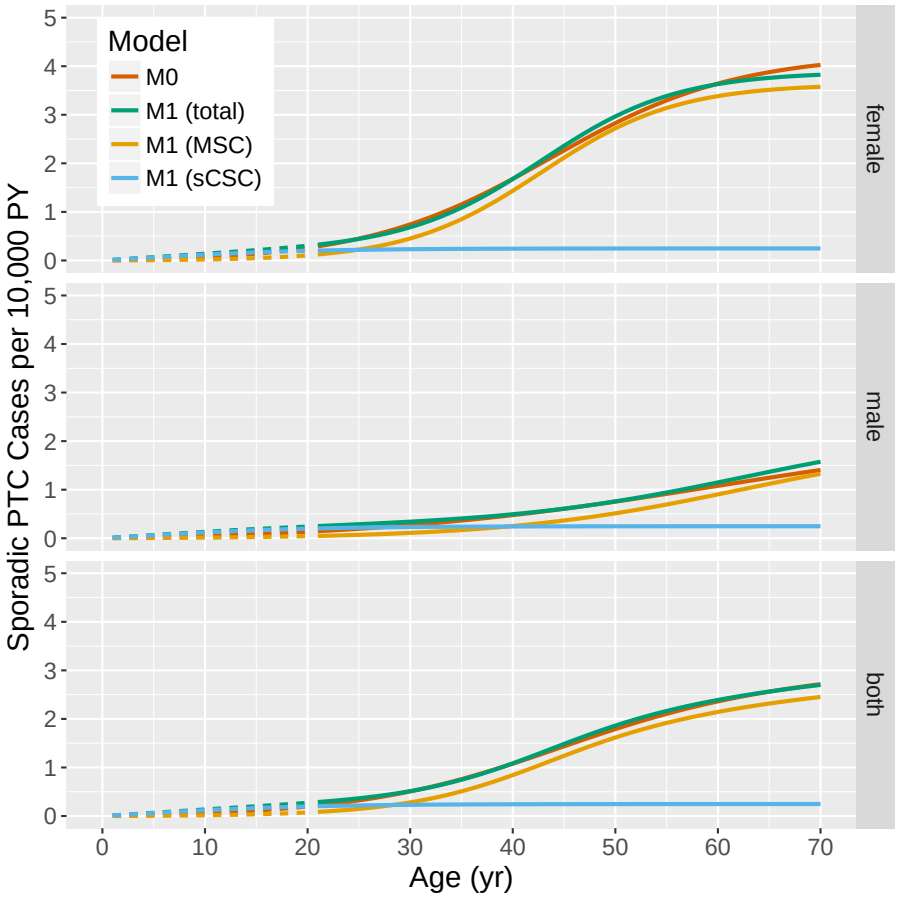


Figure S3: Incidence of sporadic PTC cases for females, males and averaged for both sexes from mechanistic models M0 and M1 for AHS participants born in 1937, total baseline of model M1 is the sum of two contributions M1 (MSC) and M1 (sC2C), dashed lines for age $< 21$ yr before 1958.

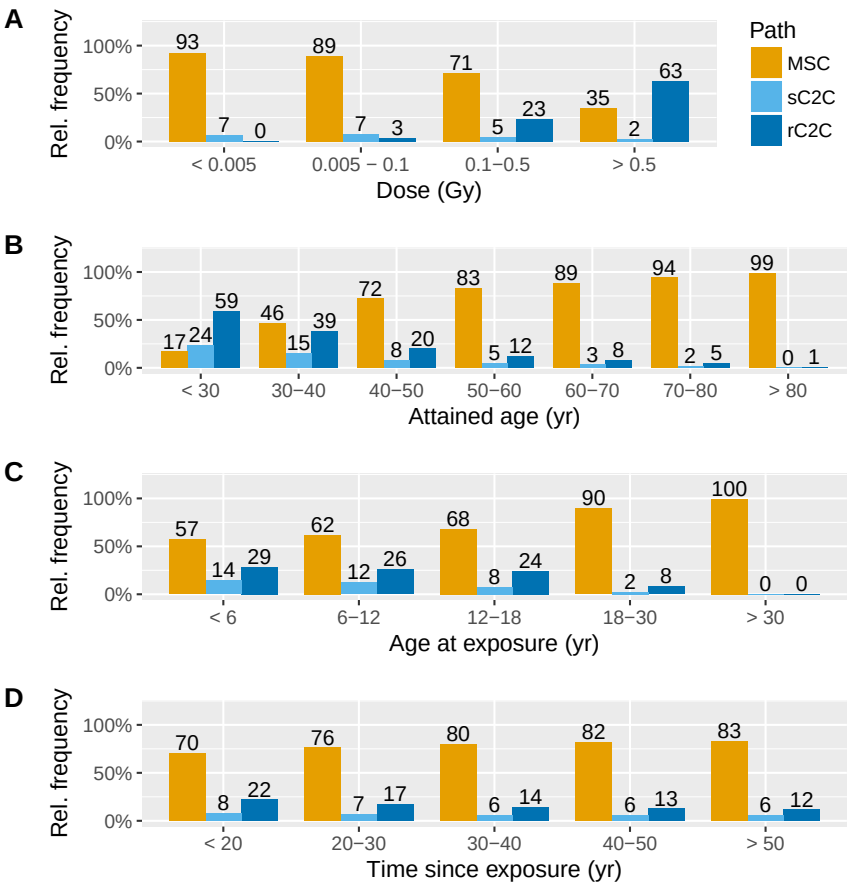


Figure S4: Relative frequency of PTC cases predicted by mechanistic model M1 in pathways MSC and C2C for categories of **A** thyroid dose, **B** attained age, **C** age at exposure and **D** time since exposure; in the MSC pathway PTC cases develop only sporadically, in the C2C pathway PTC cases develop either sporadically (sC2C) or induced by radiation (rC2C).

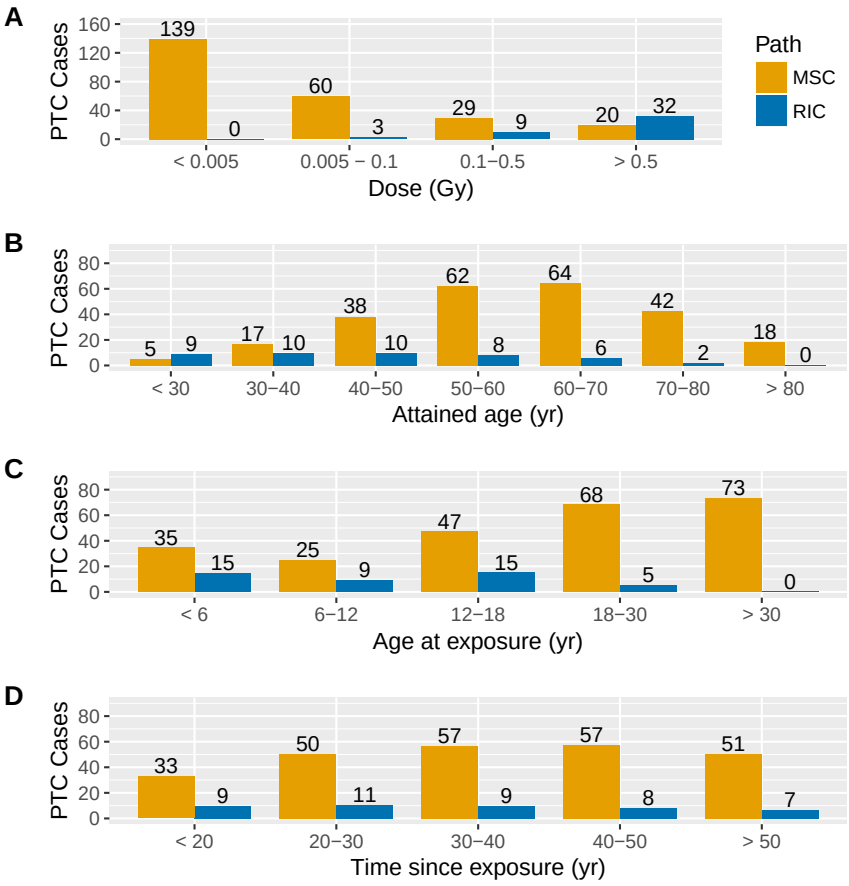


Figure S5: Number of PTC cases predicted by mechanistic model M0 in pathways MSC and RIC for categories of **A** thyroid dose, **B** attained age, **C** age at exposure and **D** time since exposure; in the MSC pathway PTC cases develop only sporadically, in the RIC pathway PTC cases are exclusively radiation-induced.

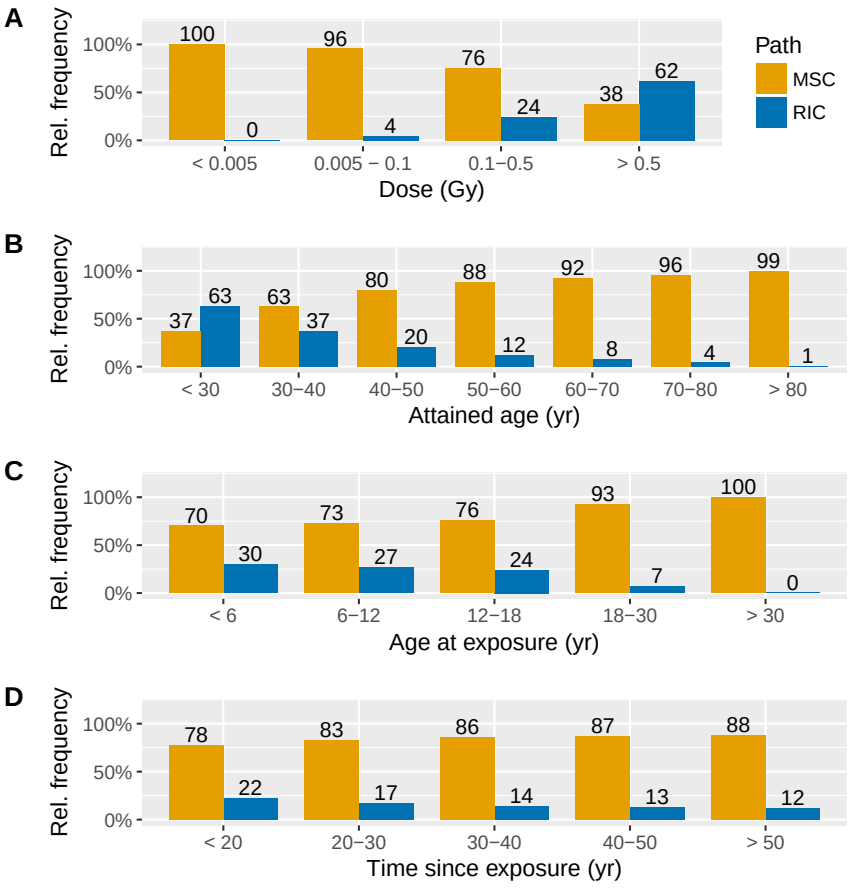


Figure S6: Relative frequency of PTC cases predicted by mechanistic model M0 in pathways MSC and RIC for categories of **A** thyroid dose, **B** attained age, **C** age at exposure and **D** time since exposure; in the MSC pathway PTC cases developed only sporadically, in the RIC pathway PTC cases are exclusively radiation-induced.

## References


1. Heidenreich, WF. On the parameters of the clonal expansion model. Radiat Environ Biophys 1996;35:127–9.
2. Meza, R, Jeon, J, Moolgavkar, SH, and Luebeck, EG. Age-specific incidence of cancer: Phases, transitions, and biological implications. Proc Natl Acad Sci U S A 2008;105:16284–9.
3. Little, MP, Heidenreich, WF, and Li, G. Parameter identifiability and redundancy: theoretical considerations. PLoS One 2010;5:e8915.
4. Armitage, P and Doll, R. The Two-Stage Theory of Carcinogenesis in Relation to the Age Distribution of Human Cancers. Br. J. Cancer 1957;11:161–169.
5. Knudson, AG. Mutation and cancer: statistical study of retinoma blastoma. Proc. Nat. Acad. Sci. USA 1971;68:820–823.
6. Dewanji, A, Jeon, J, Meza, R, and Luebeck, EG. Number and size distribution of colorectal adenomas under the multistage clonal expansion model of cancer. PLoS Comput Biol 2011;7:e1002213.
7. Aalen, OO, Valberg, M, Grotmol, T, and Tretli, S. Understanding variation in disease risk: the elusive concept of frailty. Int J Epidemiol 2015;44:1408–21.
8. Moolgavkar, SH. Commentary: Frailty and heterogeneity in epidemiological studies. Int J Epidemiol 2015;44:1425–6.
9. Kaiser, JC, Meckbach, R, Eidemüller, M, et al. Integration of a radiation biomarker into modeling of thyroid carcinogenesis and post-Chernobyl risk assessment. Carcinogenesis 2016;37:1152–1160.
10. Takano, T. Natural history of thyroid cancer [Review]. Endocr. J. 2017;64:237–244.
11. Heidenreich, WF, Luebeck, EG, and Moolgavkar, SH. Some properties of the hazard function of the two-mutation clonal expansion model. Risk Analysis 1997;17:391–399.
12. Heidenreich, WF, Kenigsberg, J, Jacob, P, et al. Time trends of thyroid cancer incidence in Belarus after the Chernobyl accident. Radiat. Res. 1999;151:617–25.