PONTIFICIA
UNIVERSIDAD
CATÓLICA
DE CHILE

INSTITUTE OF PHYSICS

A Critical Analysis of Shuryak's Predictive Radiocarcinogenesis Model

By:

NICOLAS MATTHIAS HEUMANN SCHRÖDER

Thesis submitted to the Faculty of Physics of the

Pontificia Universidad Católica de Chile

As a requirement to qualify for the degree of master's in medical physics

THESIS SUPERVISORS:

BEATRIZ SÁNCHEZ NIETO

IGNACIO ESPINOZA

January 8thth 2024

Santiago, Chile

TABLE OF CONTENTS

Summary

Objective: To critically analyze, simplify and implement a predictive radiocarcinogenesis model to estimate the risk of secondary cancer after RT which can effectively compare different radiotherapy treatment plans with the aim of having an additional element of information during the decision-making process for the best RT plan.

Methodology: A Python software was developed that was able to implement the model proposed by Shuryak et al. (2009). Simplifications and minor corrections were made which allowed for more compactness and more efficient run times. The model was then reparametrized with newer data from the *Surveillance, Epidemiology, and End Results Program* (SEER) database and several epidemiological studies using Bayesian Inference. Uncertainty propagation studies were then conducted to understand their propagation better. Finally, the model with its new parameters was applied to a selection of prostate plans to determine if it could construct a risk hierarchy.

Results: The model was successfully reparametrized with newer data. Although some parameters show significant deviation from Shuryak's original parameters, they are mostly on the same order of magnitude, and the differences arise likely due to differences in fitted data and the fitting process itself. Shuryak's model successfully built a risk hierarchy between prostate plans, although it deviated from the more simplistic linear non threshold BEIR VII model. It was also possible to simplify some complex mathematical equations, both in general and for particular cases, allowing for easier implementation and more efficient run times.

Conclusions: Shuryak's model was successfully reparametrized and implemented, showing potential to become clinically applicable. However, more comparisons between the model's result and epidemiological data must be made to evaluate its accuracy better, and more concise and complete second primary cancer studies must be used before the model is reliable enough for clinical decision-making.

Glossary

• **Absorbed dose:** Quantity used to determine the amount of ionizing radiation a material has absorbed. Measures energy deposited per unit mass and has units:

$$1\ Gray \equiv 1\ Gy = \frac{1\ Joule}{1\ kg}$$

• **Premalignant cell:** A cell that, although it is not yet fully malignant, has certain mutations that increase its likelihood of becoming malignant. They can arise due to DNA replication mistakes or when exposed to certain chemicals or ionizing radiation.

• **Mechanistic:** In the context of secondary cancer risk, a mechanistic model is one that tries to model the microscopic phenomena that give rise to the macroscopic manifestation of cancer.

• **Long-term process:** This is one that takes years to have noticeable effects. Example: The spontaneous apparition (i.e., not induced by an external agent) of premalignant cells and their posterior replication.

• **Short-term process:** This process happens fast enough to be considered instantaneous. For example, cellular death or mutations induced by ionizing radiation during radiotherapy.

• **Activation:** Process through which a normal cell becomes premalignant. If caused by radiation, we will call it "radiation-induced."

• **Inactivation by radiation:** Process in which a cell is considered to die. For healthy cells, this means the destruction of the cell itself. For premalignant cells, it suffices if it loses the ability to reproduce indefinitely.

• **Niche:** In this context, a niche is a microenvironment in a specific sector of the tissue in which it shares common traits with nearby cells.

• **Repopulation:** Process through which cells in a tissue damaged by radiation replicate to mend the damage to its condition prior to irradiation.

• **Survival:** A cell is said to survive radiotherapy if it retains the capacity to replicate afterward.

• **Relative risk:** Ratio of the risks for an event for the exposure group to the risks for the non-exposure group. Mathematically, it is calculated as $RR = \frac{risk\ of\ exposed\ population}{risk\ of\ non{-}exposed\ population}$.

• **Excess Relative Risk:** The proportional increase of risk of a population that is exposed to a specific treatment with respect to a population that has not been exposed to it. Mathematically, it is written as $ERR = RR - 1$.

• **Life Attributable Risk:** Additional cumulated probability of developing a specific cancer over the expected lifetime. It can be understood as the time integral of ERR.

• **Senescence:** Phenomenon during which a cell loses the ability to replicate and, thus, produce a lineage. Senescence is usually considered one of the primary forms of cellular death. For the purpose of radiotherapy, a senescent malignant or premalignant cell will be considered to be controlled.

• **Attained Age:** The age at which the second primary cancer manifests.

Chapter 1: Introduction

The carcinogenic potential of ionizing radiation, especially at low doses, has been firmly established [1], as was observed by studying atomic bomb survivors [2]. With RT being one of the most widely used cancer treatments worldwide, it is of utmost importance to be able to quantify the risk that the treatment designed to heal the patient ends up causing second primary cancer on a nearby previously healthy tissue. This danger is especially relevant upon consideration of the longer life expectancies of cancer patients that have come as a consequence of recent technological advancements and the improvement of screening programs. Pediatric patients are at the highest risk of suffering radiation-induced second primary cancer, given their higher radiosensitivity and the fact that their young age of exposure and long potential life span give ample opportunity for cancer to develop.

Second primary malignancies arise when part of the radiation from the treatment is absorbed by healthy tissue near the tumor, which is why they tend to appear on or close to the edge of the tumoral volume[3]. Figure 1 shows the frequency of apparition of second primary malignancies in pediatric patients after radiotherapy as a function of the distance to the edge of the treatment field.

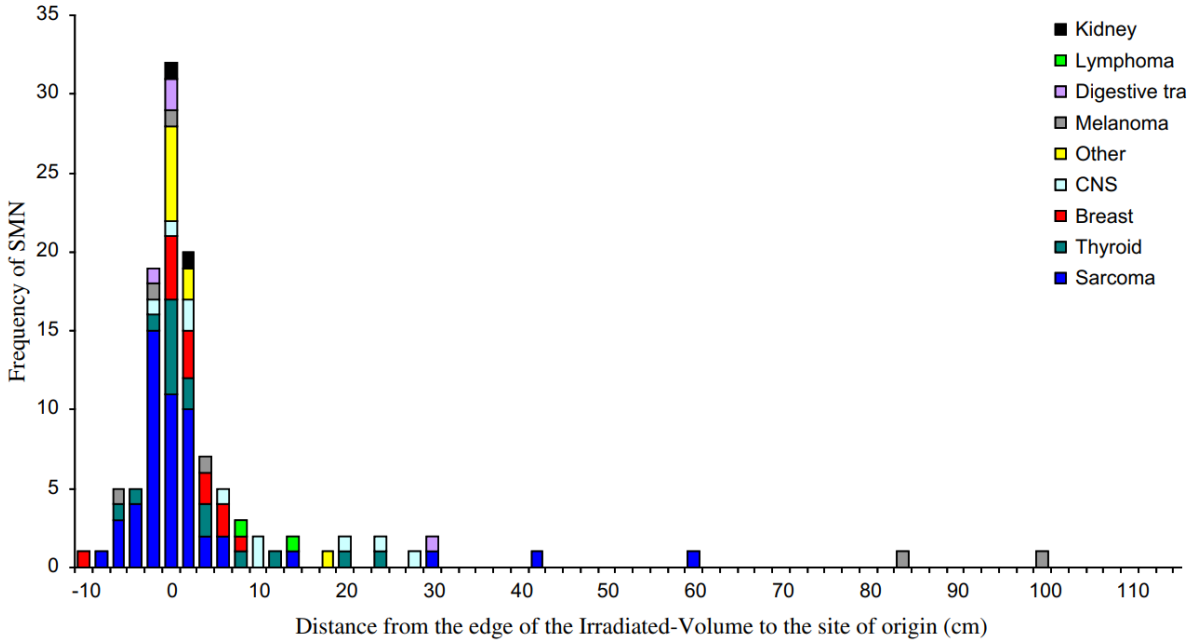


Figure 1. Distribution of frequencies of the distances from the edge of the irradiated volume to the site of secondary malignancy apparition (Second Malignant Neoplasm, SMN), among 115 pediatric patients. Source: Diallo et al.2009[4]

Although the small yet significant risk of radiocarcinogenesis has been intensively studied since the onset of radiation therapy, there are limited resources available for medical professionals when optimizing a patient's treatment. This scarcity is especially notable when comparing existing secondary cancer risk estimation models to those available for tumor control or normal tissue complication probabilities.

Ideally, medical staff should be able to estimate the secondary cancer risk for each available treatment plan to make the most informed decision. In practice, this would involve ranking several treatment options with equal probability for tumor control based on their likelihood of radiocarcinogenesis.

In principle, decisions of this nature could be based on available epidemiological data; observing a cohort of patients who have undergone a specific RT and tracking their development over a couple of decades could lead to a reasonably accurate risk estimation. However, unlike other fields of medicine, the prolonged latency periods before the appearance of second primary cancer and the relatively low likelihood of their occurrence require unrealistically long waiting periods and large cohorts. This is further complicated by the rapid development of cancer treatments, which basically ensures that by the time sufficient epidemiological data is available for a given treatment, that treatment will likely become obsolete. While it might be possible to extrapolate knowledge from an outdated treatment to a newer, albeit relatively similar one it is far from a reliable approach.

For that reason, there is a growing desire to develop a predictive mathematical model capable of estimating the second primary cancer risk of any given treatment, considering the radiobiological processes that occur inside the human body[4].

It is one of the motivations of the present work not only to study these models, and Shuryak's in particular, but also to critically analyze the latter to make the mathematical expressions more compact and more accessible to utilize while updating its parameters with newer epidemiological data for a more accurate clinical application.

Chapter 2: Objectives

2.1 General Objectives

The main objective of the present work is to analyze and implement a predictive mechanistic radiocarcinogenesis model to estimate the risk of second primary cancer after RT that can effectively compare two RT treatment plans.

2.2 Specific Objectives

a) Develop a Python code that can successfully implement the mechanistic model proposed by Shuryak and estimate the risk of second primary cancer after RT.

b) Reparametrize Shuryak's model using more recent epidemiological data.

c) Critically analyze the model's parameters and their impact on carcinogenesis.

d) Implement the model to compare a selection of alternative RT treatment plans and identify the one that minimizes the risk of developing a second primary cancer

e) Compare the resulting conclusions utilizing Shuryak's model to those obtained by using the BEIR VII and Schneider's model to determine the congruency of their results.

Chapter 3: Theoretical Framework

3.1 Premalignant cells

Cancerous lesions form as a consequence of a cell's genetic mutations, which can occur due to a genetic predisposition, environmental factors (such as exposure to asbestos or long-time smoking), and ionizing radiation. Whatever the cause, a cancerous cell is born once it accumulates a mutation and manages to evade the body's cancer-detection systems. Once a cell has accomplished this, we will call it a $premalignant$ cell. This is a cell that has begun accumulating mutations and thus has the potential to become malignant in the future but still behaves similarly enough to normal, healthy tissue. The premalignant cell will become fully malignant after going through each of the *Hallmarks of Cancer*, as described by Douglas Hanahan and Robert Weinberg [5].

3.2 DNA damage and repair

Understanding how radiation causes mutations or cell death requires understanding the structure of cell DNA and its response to ionizing radiation. There are two mechanisms through which the double-strand structure of DNA can be damaged: single and double-strand breaks. Single-strand breaks correspond to sub-lethal damage and can usually be repaired. However, there is a small probability of an error occurring during the reparation, which can transform the cell into a premalignant cell or cause it to fail one of the checkpoints during cell division, leading to its death. While double-strand breaks can also be repaired, they are more likely to lead to cell death. Since tumors repair much slower than healthy tissue, delivering the radiation across several fractions is beneficial to allow healthy tissue to repair, thus minimizing damage to it.

3.3 State of the art of radiation induced risk models

There are two main kinds of cancer estimation risk models. First, there are empirical models based on the observation of cancer incidence in the years following RT. An example is the linear response model developed from observations of atomic bomb survivors [6], which concluded that irradiation in the dose ranges between 1-2 Gy correlates linearly with secondary cancer risk. However, these are statistical results designed to be applied to populations rather than individuals. Furthermore, doses delivered to organs during RT often surpass this range and thus require a different approach.

The second type, the mechanistic model, leverages our existing understanding of radiobiological processes to estimate the risk of second primary cancers mathematically. However, achieving a purely mechanistic model would require perfect knowledge of every intracellular and radiobiological process, rendering such an approach unrealistic. Nevertheless, a more viable option involves combining both approaches into a semi-mechanistic model, offering a substantial advantage. By incorporating radiobiological processes, such a model can strive to deliver personalized results by considering patient- and treatment-specific characteristics, such as age, number of fractions, and organ doses.

To construct a mathematical model, one must thus select the most relevant biological aspects and assign them a representative function whose parameters should consequently be fitted from epidemiological data.

Alexandru Dasu and Iuliana Toma-Dasu [7] provide an insightful overview of mechanistic and semi-mechanistic models, which I will briefly summarize here. It is important that the reader keeps in consideration the impact that the choice of parameters (and thus biological processes) has on the resulting risk curve.

One can construct a model to estimate second primary cancer risk after irradiation by considering the damage evolution of the DNA of irradiated cells, considering the possibility of both lethal and sub-lethal lesions, where the former results in cell killing and the latter in the potential for carcinogenesis (by initiating premalignant cells that can eventually become fully malignant and form a tumor). Cancer risk, as per such a model, becomes a competition between the initiation of malignant cells and their inactivation by radiation. As reported by the United Nations Scientific Committee on the Effects of Atomic Radiation (UNSCEAR)[8], it is possible to describe the risk as:

$$Risk = (\alpha_1 D + \beta_1 D^2)\exp[-(\alpha_2 D + \beta_2 D^2)] \qquad (1)$$

Here, D is the radiation dose and $\alpha_1, \beta_1, \alpha_2, \beta_2$ are parameters related to the induction of mutations and cell survival, respectively. It is interesting to note the shape of the curve generated by this model. The risk increases as the delivered dose remains in a predominantly sub-lethal range, but as it continues to increase, we observe a falloff in risk as the cell-killing begins to dominate.

Some authors [9][10] suggest no such decline in risk at higher doses, which leads to the formulation of the plateau model proposed by Robert Davis [11]:

$$Risk = \frac{cN}{a}\exp(1 - e^{-aD}) \qquad (2)$$

Where D is the radiation dose, c the probability per unit dose of the development of a second cancer from a precancer cell, N is the average number of initial precancer cells, and a is the probability of radiation killing mutated cells.

However, neither of these models has accounted for the known mechanism of interfraction cell repair, which plays a pivotal role in safeguarding healthy tissue during RT, which, unlike tumoral tissue, is known to repair. It is possible to modify the UNSCEAR model to account for interfraction repairs for n fractions of RT:

$$Risk = \left(\alpha_1 D + \frac{\beta_1 D^2}{n}\right)\exp\left[-\left(\alpha_2 D + \frac{\beta_2 D^2}{n}\right)\right. \qquad (3)$$

This expression also predicts a bell-shaped risk-dose relation, except it shows a slower falloff due to DNA repair. All four models are illustrated in Figure 1, where the different curves can be easily identified by their linear, bell, or plateau shape.

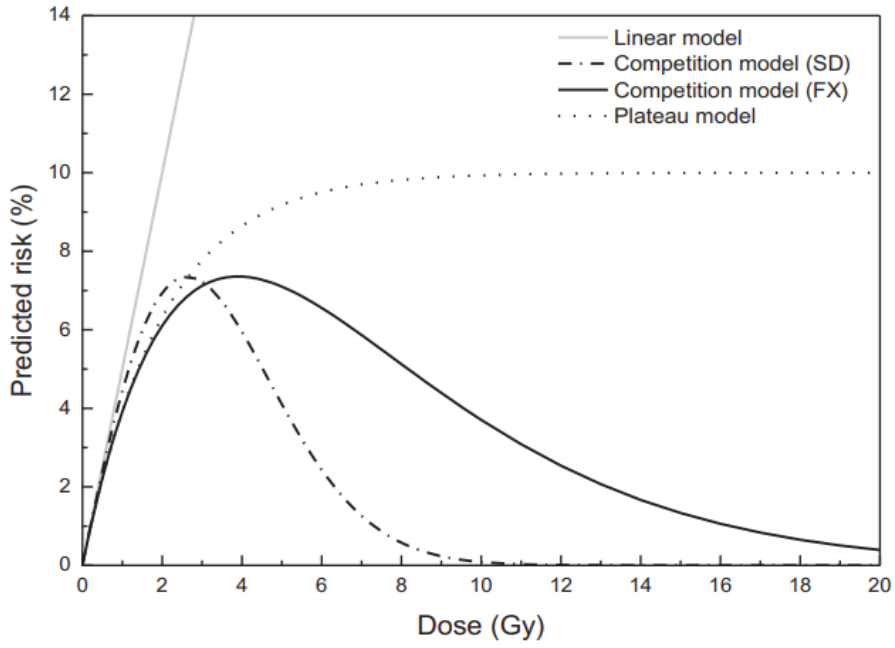


Figure 2. Illustration of the dose-risk dependence for risk models used in radiation therapy (Dasu et al 2017)[7]. Here, SD refers to the UNSCEAR report model (Equation 1), which does not consider dose fractionation. The plateau model corresponds to equation 2. FX refers to the modified UNSCEAR report that does consider it (Equation 3).

Incorporating interfraction repair makes the models more realistic, yet several critical aspects remain unaccounted for. Sach and Brenner [12] proposed a model assuming that radiation-induced cell killing triggered accelerated repopulation to restore the steady-state number of functional cells in the irradiated tissue.

$$ERR = N[\exp(\gamma D) - 1]B \qquad (4)$$

Here, N is the number of functional cells before irradiation, γ the probability of mutation per unit dose, and B is the proportionality factor linking the yield of mutated cells to the risk. The risk here is given as Excess Relative Risk (ERR), a very important quantity that was defined in the glossary.

Schneider [13] also proposed a risk model with similar considerations but allowing for the differentiation between carcinomas and sarcomas.

$$Risk_{carcinomas} = \alpha_1 \frac{\exp\left(-\alpha D - \frac{\beta D^2}{n}\right)}{\left(\alpha + \frac{\beta D}{n}\right)R} \left\{1 - 2R + R^2 \exp\left(\alpha D + \frac{\beta D^2}{n}\right) - (1-R)^2 \exp\left[-\frac{R}{1-R}\left(\alpha D + \frac{\beta D^2}{n}\right)\right]\right\} \qquad (5)$$

$$Risk_{sarcomas} = \alpha_1 \frac{\exp\left(-\alpha D - \frac{\beta D^2}{n}\right)}{\left(\alpha + \frac{\beta D}{n}\right)R} \left\{1 - 2R + R^2 \exp\left(\alpha D + \frac{\beta D^2}{n}\right) - (1-R)^2 \exp\left[-\frac{R}{1-R}\left(\alpha D + \frac{\beta D^2}{n}\right)\right] - R\left(\alpha D + \frac{\beta D^2}{n}\right)\right\} \qquad (6)$$

Here, α_1 is the linear risk coefficient, α and β are the LQ parameters describing cell survival, D is the total dose, n is the number of fractions, and R is a parameter characterising the ability of the tissue to repopulate.

In the BEIR VII report, "H*ealth risks from exposure to low levels of ionizing radiation"* [35], models are presented to estimate risk expressions. In particular, the expression to estimate ERR is:

$$ERR = \exp(\gamma e^*)\, a^{\eta} \qquad (7)$$

Where $e^* = e - 30$, with e age at exposure in years when $e < 30$ and equal to zero when $e > 30$. a is attained age in years, and γ and η are parameters that quantify the dependence of the ERR on e and a.

3.3 Shuryak's model

From the vast array of semi-mechanistic models that exist, the one proposed by Shuryak[14][15][16] considers the largest number of biological processes, thus possessing great potential. Yet this model has been mostly overlooked in the decade since its original publication.

Due to the proposed model's complexity, I will use Figure 2 from Part I of Shuryak's publication[14], which is a schematic representation of the proposed premalignant cell kinetics.

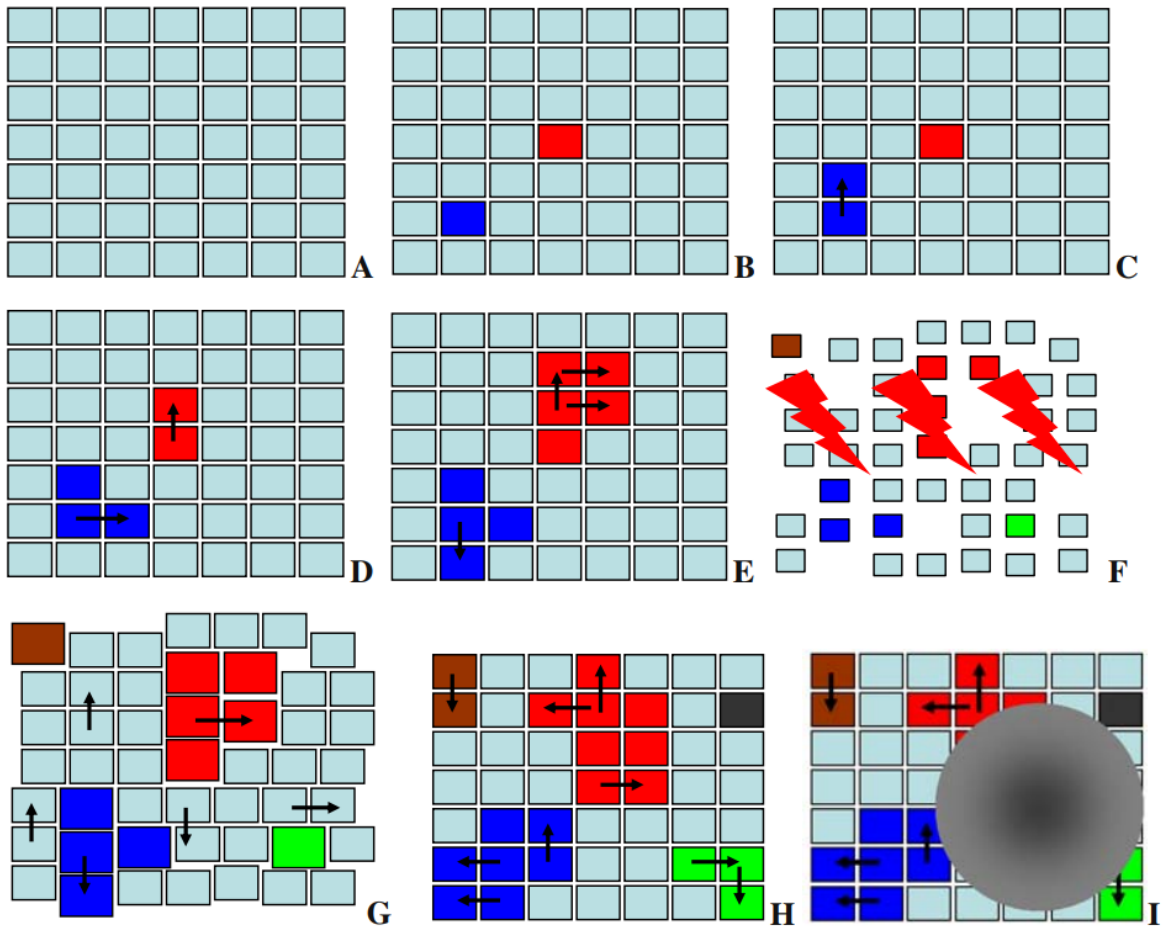


Figure 3. Schematic representation of the proposed model. Source: Shuryak et al 2009 [14]

The model assumes that both normal and premalignant cells inhabit specific tissue niches, which are collections of cells that share characteristics and lineage. The carrying capacity and number of these niches are assumed to be homeostatically regulated. It is also assumed that premalignant cells gradually lose their carcinogenic potential with age so that they have a progressively smaller ability to transform into a purely malignant cell.

In Figure 2, each color represents a different niche, with its carrying capacity represented by the size of the square. Subfigure A represents healthy tissue at the moment of birth, which assumes there are no premalignant cells present at this point, a reasonable assumption for most cancers. Subfigure B represents the phenomenon of spontaneous premalignant cell initiation, which occurs as a consequence of cell replication errors or environmental factors. The newly formed premalignant cell niches are represented in red and blue. It is assumed that these premalignant cell niches have the capacity to reproduce and invade adjacent niches, which is depicted in C, D, and E. It is also possible for the premalignant niche to divide into two daughter niches, as is the case with colon crypt fission [17][18][19], and the new colonies may have different mutations than the original one. However, the model does not distinguish between these processes, as both yield the same outcome. These processes occur over the course of decades, constituting a long-term process (Subfigures A to E).

With these assumptions, one can mathematically model the number of premalignant cell niches as proportional to:

$$A_{background} = \frac{a}{b}\left(e^{bT_x} - 1\right)e^{-cT_x^2} \qquad (8)$$

Here, a is a parameter representing the spontaneous initiation rate of premalignant cell niches, b is proportional to their replication rate, c is an age-dependent premalignant cell senescence parameter, while T_x is the age of the patient.

At some point, the patient is subjected to radiotherapy in a tissue nearby. As a consequence, the healthy tissue we are currently observing is irradiated, and we will observe both the effects of inactivation and initiation. Niches become smaller in size as they lose a fraction of their cells. Some may be wiped out entirely. At the same time, there exists a possibility of radiation-induced initiation, which is represented by the newly formed green and dark red premalignant niches.

This short-term process is considered instantaneous compared to the decades it takes premalignant cells to form and replicate. The three main aspects of the short-term process are initiation, inactivation, and repopulation. The model allows for custom fractionization (i.e., it provides freedom to choose the spacing, number, and dose per fraction), and considers cell repopulation between fractions. There are three key mathematical quantities that will represent each one of the aforementioned mechanisms. $Sf(Z,D)$ will be the probability that at least a single premalignant cell in a niche with a carrying capacity of Z will survive a dose D in n fractions:

$$Sf(Z,D) = 1 - [1 - F(0)]^Z \qquad (9)$$

$F(k)$ is the probability that a premalignant cell present before irradiation will produce a lineage that survives all dose fractions, and is given by:

$$F(k) = \left(1 + \sum_{j=k+1}^{K} \left[\frac{S^{(k-j)}(1-S)}{\prod_{i=k}^{j-1} Sn^{-}(i)}\right]\right)^{-1} \qquad (10)$$

Here, $n^{-}(i+1)$ is the number of premalignant cell niches present right before the $(i+1)th$ dose fraction. This quantity can be found recursively using the formula provided in the errata to Shuryak's publication[16]:

$$n^{-}(i+1) = \frac{\nu \mathrm{Sn}^{-}(i)}{\{Sn^{-}(i) + [\nu - Sn^{-}(i)] \exp[-\lambda(\Delta \mathrm{T}(i))]\}} \qquad (11)$$

Here, ν is the number of cells present before irradiation, λ is the maximum net cell repopulation rate, and $\Delta T(i)$ is the time interval in days between the ith and $(i+1)th$ fractions, while $S = e^{-\alpha d - \beta d^2}$ is the standard LQ survival fraction.

The second relevant quantity is $ISf(D)$, which is the net probability outcome of inactivation, initiation and repopulation during radiation exposure, and is given by:

$$ISf(D) = \sum_{k=1}^{K} I(k)F(k) = \sum_{k=1}^{K} \frac{dSn^{-}(k)}{\nu} F(k) \qquad (12)$$

$ISf(D)$ can be understood as the probability of a niche being initiated in a given dose fraction k times the probability of that niche surviving all remaining fractions. The final contribution of premalignant cells initiated during radiotherapy can be determined by assuming that it is Poisson distributed, with an average of $aXISf(D)$.

The last term to consider is related to the process in which promoted or initiated cell niches increase their carrying capacity Z after radiotherapy, as shown in subfigure G of Figure 2. This initial excess is assumed to be linearly dependent on the radiation dose, with the coefficient being parameter Y. In the months following radiotherapy (H), it is assumed that these niches return to their normal carrying capacities. Both extinction of niches and shrinkage are combined into a single adjustable parameter δ, which allows us to find the average normalized number of premalignant cells per surviving niche at any given time T_y after exposure:

$$C_{rad}(T_y) = \frac{1 + Y \cdot D}{\left[1 + Y \cdot D\left(1 - e^{-\delta T_y}\right)\right]} \qquad (13)$$

Finally, the model considers that premalignant cell niches can continue to be spontaneously initiated and replicate in the years following radiotherapy.

With this, the number of premalignant cells in a patient treated with radiotherapy at age T_x at a time T_y after treatment is given by:

$$A_{rad} = \frac{a}{b}\left(\frac{1 + Y \cdot D}{\left[1 + Y \cdot D\left(1 - e^{-\delta T_y}\right)\right]}\left(\left(e^{bT_x} - 1\right)Sf(Z, D) + b \cdot X \cdot ISf(D)\right)e^{bT_y} + \left(e^{bT_y} - 1\right)\right)e^{-c\left(T_x + T_y\right)^2} \qquad (14)$$

It is important to note that the age of the patient is written here as the sum of the age at RT T_x and the time since radiotherapy T_y.

To be able to quantitatively estimate risk, Shuryak compares this number to the number of premalignant cells present in a healthy individual of the same age who has not been exposed to radiotherapy (equation 1):

$$ERR = \frac{A_{rad}}{A_{background}} - 1 = \frac{\frac{1 + YD}{\left[1 + YD\left(1 - e^{-\delta T_y}\right)\right]}\left(\left(e^{bT_x} - 1\right)Sf(Z, D) + b \cdot X \cdot ISf(D)\right)e^{bT_y} + \left(e^{bT_y} - 1\right)}{e^{b\left(T_x + T_y\right)} - 1} \qquad (15)$$

The parameters a and c have canceled out and are thus not required for the estimation of ERR. Parameters for selected organs are given by Shuryak in the second part of his publication after fitting his

expressions to epidemiological data. The parameters α and β, and λ are taken from literature [20][21][22][23].

Thus, to estimate the ERR after a particular RT, one must know the dose absorbed by each relevant organ, the number of fractions, the patient's age, and the parameters for each organ. A summary of each parameter and its meaning can be found in Table 1.

Parameter	Units	Interpretation
a	$time^{-2}$	Spontaneous stem cell initiation and transformation
b	$time^{-1}$	Pre-malignant niche replication
c	$time^{-2}$	Age-dependent pre-malignant cell senescence
X	$time/dose$	Radiation-induced initiation
Y	$dose^{-1}$	Radiation-induced promotion
Z	cells/niche	Carrying capacity for pre-malignant cells per niche
δ	$time^{-1}$	Homeostatic regulation of pre-malignant cell number per niche
α, β	$dose^{-1}, dose^{-2}$	Stem cell inactivation by radiation
λ	$time^{-1}$	Maximum net stem cell proliferation (repopulation) rate

Table 1. Summary of model parameters defined in Shuryak's original publication [14].

Although Shuryak's model is by far the one that considers the largest number of radiobiological processes and thus is most likely to represent reality accurately, it has seen very little use since its publication, likely due to the difficulty of applying it. This has been a big motivator behind the desire to analyze, simplify, and reparametrize the model and make it more accessible.

Chapter 4: Methods

4.1 Reparametrization

A Python script was written to execute the model using the parameters found in his original publication[15] and reproduce the graphs presented there. After confidence in the correct functioning of the model was established, it was necessary to reparametrize it to update it with newer epidemiological data. Given that the parameter b appears both in the background cancer incidence as well as in the secondary cancer model, it is convenient to do a two-step fitting process since there are more available data for background cancer than for second primary cancer. This ensures the best possible fit for b before fitting the rest of the parameters to the much less certain second primary cancer studies.

Background cancer incidence data was obtained from the *Surveillance, Epidemiology, and End Results Program (SEER)[24]* software (SEERStat). Incidence rates were obtained for the years 2015 to 2020 for ages 2 to 87 at 5-year intervals. The background risk was fitted to data from the SEER database. The sex of the subjects was chosen to fit the sex of the epidemiological studies that would be used for the second phase. It is important to note that parameters *a* and *c* cancel out from the final expression for ERR, which is why the discussion will only concern *b.*

Once ranges for b were determined, Bayesian inference was used to find the best fit for selected organs to epidemiological data found in the literature. It is important to note that these studies had to include information about the subject's age at exposure, attained age, and dose to organ, as well as have the risk reported as either relative risk, excess relative risk, or a quantity that could be reliably converted to these. Some of these quantities, especially dose to organ and the relevant ages were usually reported as averages of the studied cohorts, thus obscuring some of the data. Bayesian inference was chosen due to its ability to include the uncertainty ranges in b and the ERR values found in the literature.

Organs were chosen based on the availability of the required data. In the end, parameters for the lung, stomach, colon, rectum, pancreas, bladder, breast, central nervous system, and thyroid were chosen and reparametrized, as they were in Shuryak's original publication. Additionally, parameters for the esophagus were also calculated. Preston's work on the second primary cancer incidences following the atomic bomb [6] was used in the low-dose range for every organ. The reported ERR/Gy ranges reported by Preston were used in the linear, low dose range, in which 5 sample points were chosen for each case, assuming that $ERR = \frac{ERR}{Gy} \times Dose$. A literature review was conducted to find additional studies to the ones used by

Shuryak. All studies used to fit the model parameters can be found in Table 2. Two additional studies were found to be suitable for the fitting process.

To fit the data, Bayesian inference was used in which priors were given uniform distribution within the biologically plausible range. The likelihood, or observed data, included their respective uncertainties. It is important to note that these uncertainties varied widely and, in some cases, were large enough to obscure a meaningful conclusion. To visualize the propagation of these uncertainties, each parameter's distribution was assumed to be normal, and 10.000 synthetic data sets were generated, from which the ERR for a given dose was computed 500 times by randomly drawing parameter samples from the generated distributions. This process was done for all the original nine organs in Shuryak's publication (stomach, lung, colon, rectum, pancreas, bladder, breast, central nervous system, and thyroid), as well as the esophagus, which was additionally parametrized during the present work.

Study	Main Author	Organs Reported	Journal, Volume, Year
Radiation Dose and Second Cancer Risk in Patients Treated for Cancer of the Cervix*	John D. Boice	Bladder, Stomach, Pancreas	Radiation Research, Vol. 116, 1988
Second Cancers Among 40,576 Testicular Cancer Patients: Focus on Long-term Survivors	LB Travis	Esophagus, Colon, Bladder, Pancreas, Stomach, Lung	Journal of the National Cancer Institute, Vol. 97, 2005
Risk of Second Malignant Neoplasms Among Long-term Survivors of Testicular Cancer	LB Travis	Stomach, Rectum, Colon, Bladder, Pancreas	Journal of the National Cancer Institute, Vol. 89, 1997
Roles of Radiation Dose, Chemotherapy, and Hormonal Factors in Breast Cancer Following Hodgkin's Disease	FE Van Leeuwen	Breast	Journal of the National Cancer Institute, Vol. 95, 2003
Breast Cancer Following RT and Chemotherapy Among Young Women with HD	LB Travis	Breast	Journal of the American Medical Association, 290,2003
Lung Cancer Following Chemotherapy and Radiotherapy for Hodgkin's Disease	LB Travis	Lung	Journal of the National Cancer Institute, Vol. 94, 2002
Lung Cancer After Treatment for Hodgkin's Disease: Focus on Radiation Effects	ES Gilbert	Lung	Radiation Research, Vol. 159, 2003
Thyroid Cancer in Childhood Cancer Survivors: A Detailed Evaluation of Radiation Dose Response and Its Modifiers	CM Ronckers	Thyroid	Radiation Research, Vol. 166, 2006
New Primary Neoplasms of the Central Nervous System in Survivors of Childhood Cancer: A Report from the Childhood Cancer Survivor Study	JP Neglia	CNS	Journal of the National Cancer Institute, Vol. 98, 2006
Second Malignant Neoplasms in Testicular Cancer Survivors*	C. Fung	Stomach, Rectum, Pancreas, Bladder	Journal of the National Comprehensive Cancer Network, Vol. 10, 2012

Table 2. Selected studies and the organs they were used for.[9][25-33][37] Studies marked with $*$ were found during the present work and not originally included in Shuryak's original publication.

4.2 Uncertainty Propagation

To better understand some of the uncertainties obscured by the averaging done by the studies, several plots were made varying the age at exposure and attained age.

With confidence in understanding the model's uses and limitations, it was relevant to test whether parameter uncertainties hinder the main objective: to be able to distinguish between two otherwise identical plans and select the one with the lowest second cancer probability.

For this purpose, a comparison between 36 prostate plans, as published by Sanchez-Nieto[34] was chosen. Plans of over 6 MV were eliminated in order to avoid dose contributions from neutrons. The present model was then implemented by picking one hundred random parameter combinations from within their possible range and estimating the resulting ERR distribution for each plan.

These results were compared to those obtained using the BEIR VII mechanistic model [35] to verify whether both models agree on the ranking of the plans in terms of most to least risk, which would be the deciding criteria when making a clinical decision. The results were also compared to those reported in the paper obtained using Schneider's model. It is essential to consider that Schneider's model reports the Life Attributable Risk (LAR) rather than ERR, and comparing both quantities is not straightforward. For that reason, the aim is to compare the hierarchy of the plans, which is, after all, the relevant result that one wishes to obtain in order to allow the health professionals to choose the plan with the least risk.

Chapter 5: Results

5.1 Improvements to Shuryak's model

The implementation of Shuryak's model gave rise to a series of challenges. First was the quantity ν, defined as the number of normal cells present before irradiation. Although defined in his publication and explicitly used in the model in equations 11 and 12, there was no mention of the values of ν for the different organs. It was noticed after testing with arbitrary values that the risk did not depend, in fact, on the particular choice of ν. This can be shown mathematically by assuming that the initial condition of the recursive equation 11 is

$$n^{-}(0) = \nu$$

Note here that the number of fractions is counted such that 0 is the first fraction. This parameter appears only in two places, and it is easy to show that it cancels out. Beginning with equation 11, it is important to note that we care only about the ratio of surviving cells from two successive fractions (equation 10):

$$\frac{n^{-}(1)}{n^{-}(0)} = \frac{\nu S}{\{Sn^{-}(0) + [\nu - Sn^{-}(0)]\exp[-\lambda(\Delta T(i))]\}} = \frac{\nu S}{\{S\nu + [\nu - S\nu]\exp[-\lambda(\Delta T(i))]\}}$$

One can factor out a ν from the denominator and rearrange the expression such as:

$$\frac{n^{-}(1)}{n^{-}(0)} = \frac{S}{\{S + [1 - S]\exp[-\lambda(\Delta T(i))]\}} = \gamma$$

The result of this fraction is thus independent of ν, and the result holds for posterior fractions:

$$\frac{n^{-}(2)}{n^{-}(1)} = \frac{\nu S}{\{Sn^{-}(1) + [\nu - Sn^{-}(1)]\exp[-\lambda(\Delta T(i))]\}}$$

$$= \frac{\nu S}{\{Sn^{-}(0)\gamma + [\nu - Sn^{-}(0)\gamma]\exp[-\lambda(\Delta T(i))]\}}$$

$$= \frac{\nu S}{\{S\nu\gamma + [\nu - S\nu\gamma]\exp[-\lambda(\Delta T(i))]\}} = \frac{S}{\{S\gamma + [1 - S\gamma]\exp[-\lambda(\Delta T(i))]\}}$$

Again, we find that the result is independent of the parameter ν. It remains to be shown that the ν in $ISf(D)$ (equation 12) also cancels out:

$$ISf(D) = \sum_{k=1}^{K} \frac{dSn^{-}(k)}{\nu} F(k)$$

$$ISf(D) = \sum_{k=1}^{K} \frac{dS\left[\frac{\nu \mathrm{Sn}^{-}(k-1)}{\{Sn^{-}(k-1) \ + \ [\nu - \ Sn^{-}(k-1)] \exp[-\lambda(\Delta \mathrm{T(i)})]\}}\right]}{\nu} F(k)$$

As before, the result is clear once you examine the cases for $k = 1$:

$$ISf(D)_{k=1} = \frac{dS\left[\frac{\nu \mathrm{Sn}^{-}(0)}{\{Sn^{-}(0) \ + \ [\nu - \ Sn^{-}(0)] \exp[-\lambda(\Delta \mathrm{T(i)})]\}}\right]}{\nu} F(k)$$

$$= \frac{dS\left[\frac{\nu^2 S}{\{\nu \ + \ [\nu - \ S\nu)] \exp[-\lambda(\Delta \mathrm{T(i)})]\}}\right]}{\nu} F(k) = dS\left[\frac{\nu S}{\{\nu \ + \ [\nu - \ S\nu)] \exp[-\lambda(\Delta \mathrm{T(i)})]\}}\right] F(k)$$

$$ISf(D)_{k=1} = dS\left[\frac{S}{\{1 \ + \ [1 - \ S)] \exp[-\lambda(\Delta \mathrm{T(i)})]\}}\right] F(k)$$

Thus, the parameter ν has been completely canceled out from every equation it originally appeared in. What this translates to is that the expressions for surviving cells $n^{-}(k)$ are normalized so that they refer not to the number but rather to the fraction of surviving cells, thus giving an initial condition for the recursive formula that $n^{-}(0) = 1$.

5.2 A compact expression for F(k)

Another important simplification can be done on equation 10, assuming that all fractions deliver the same dose. We can expand the product inside of $F(k)$:

$$\prod_{i=k}^{j-1} \frac{n^{-}(i+1)}{Sn^{-}(i)} = \frac{n^{-}(k+1)}{Sn^{-}(k)} \times \frac{n^{-}(k+2)}{Sn^{-}(k+1)} \times \frac{n^{-}(k+3)}{Sn^{-}(k+2)} \times \ldots \times \frac{n^{-}(j-1)}{Sn^{-}(j-2)} \times \frac{n^{-}(j)}{Sn^{-}(j-1)}$$

Evidently, this is a telescopic product which simplifies down to:

$$\prod_{i=k}^{j-1} \frac{n^{-}(i+1)}{Sn^{-}(i)} = \frac{n^{-}(j)}{S^{j-1-k}\, n^{-}(k)}$$

Now, plugging this back into equation 10:

$$F(k) = \left(1 + \sum_{j=k+1}^{K} \left[\frac{S^{(k-j)}(1-S)}{S^{j-1-k}}\frac{n^-(j)}{n^-(k)}\right]\right)^{-1} = \left(1 + \sum_{j=k+1}^{K} \left[\frac{S^{(k-j+j-1-k)}(1-S)n^-(k)}{n^-(j)}\right]\right)^{-1}$$

$$F(k) = \left(1 + (S^{-1}-1)\sum_{j=k+1}^{K} \left[\frac{n^-(k)}{n^-(j)}\right]\right)^{-1}$$

This expression is much simpler to compute than 3 and is completely equivalent as long as every fraction delivers an equal dose.

5.3 Best fit parameters

The best-fit parameters related to the background risk, a b and c, are shown in Table 3, while the generated plots are shown in Figure 3. The resulting parameters are mostly the same as the ones originally proposed by Shuryak. The differences arise due to the differences in cancer indices between the available data around the release of the original publication and the data available for recent years. If instead of fitting cancer incidence rates from 2015 to 2020, one fits to data before 2007, Shuryak's parameters and graphs can be flawlessly reproduced.

Cancer	$a \times 10^6\ (years^{-2})$	$b\ (years^{-1})$	$c \times 10^3\ (years^{-2})$	k
Stomach	1.466 (0.857, 2.076)	0.174 (0.160, 0.188)	0.793 (0.857, 2.076)	0.000 (0, 0)
Lung	0.005 (0.002, 0.007)	0.411 (0.395, 0.427)	2.421 (0.002, 0.007)	0.000 (0, 0)
Colon	25.046 (14.83, 35.26)	0.123 (0.109, 0.138)	0.509 (14.83, 35.26)	0.000 (0, 0)
Rectum	13.094 (5.55, 20.63)	0.150 (0.129, 0.171)	0.912 (5.55, 20.63)	0.000 (0, 0)
Pancreas	0.009 (0.002, 0.016)	0.362 (0.340, 0.383)	2.144 (0.002, 0.016)	0.000 (0, 0)
Bladder	0.004 (0.002, 0.005)	0.393 (0.371, 0.395)	2.110 (2.035, 2.185)	0.000 (0, 0)
Breast	74.316 (38.519, 110.113)	0.188 (0.170, 0.206)	1.247 (1.111, 1.383)	0.000 (0, 0)
CNS (modified)	0.651 (0.253, 1.450)	0.219 (0.196, 0.242)	1.380 (1.215, 1.544)	1044.657 (128.058, 1961.260)
CNS	2.072 (0, 4.669)	0.191 (0.148, 0.234)	1.189 (233.845, 1.497)	0.000 (0, 0)
Thyroid	156.650 (125.68, 187.61)	0.096 (0.085, 0.107)	0.975 (0.747, 0.934)	0.000 (0, 0)

Table 3. Best-fit parameter values for background incidence of all analyzed cancer types. An additional parameter k is shown here and will be explained shortly. The 95% confidence intervals appear in parenthesis next to their corresponding mean values.

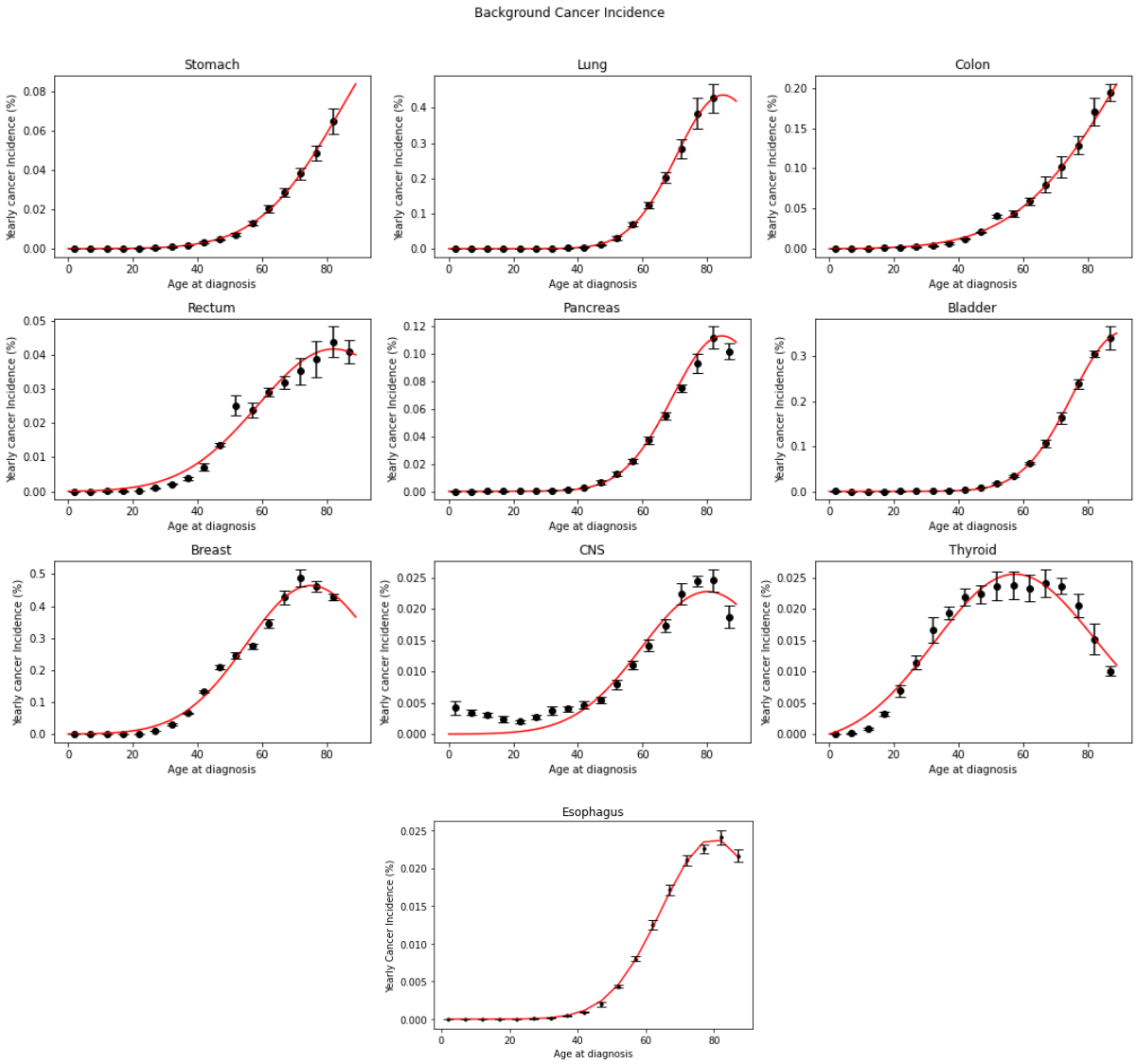


Figure 4. Best fit model predictions for US Background incidence for each cancer type from the SEER database.

The parameters for radiocarcinogenesis were fitted using both low doses from survivors of the atomic bomb and higher doses from second primary cancer studies after RT. A Bayesian Inference code was used in which priors were selected based on biologically plausible ranges. The results can be found in Table 4.

Cancer	X (Years Gy^{-1})	Y (Gy^{-1})	$\log_{10}$ Z	$\delta \times 10^3$ ($years^{-1}$)	$\alpha(Gy^{-1})$, β (Gy^{-2})	$\lambda(day^{-1})$
Stomach	5.000 (0,10)	0.243 (0.177,0.300)	2.614 (2.41,2.75)	4.172 (0,9.86)	0.25,0.025	0.05
Lung	5.000 (0,10)	0.360 (0.222,0.498)	9.698 (0,10)	15.835 (5.9,25.8)	0.25,0.025	0.05
Colon	56.829 (0.7,113)	2.229 (0.65,3.80)	4.724 (0,5)	33.058 (25.4,40.66)	0.25,0.025	0.1
Rectum	5.000 (0,10)	1.669 (0,5.7)	11.704 (0,12)	60.822 (7.644,114)	0.25,0.025	0.1
Pancreas	5.041 (0,10)	0.412 (0.224,0.600)	5.972 (5.92,6.02)	12.462 (7.126,17.799)	0.25,0.025	0.1
Bladder	5.000 (0,10)	0.241 (0.0016,0.48)	9.707 (0,10E10)	27.000 (12.644,42.283)	0.25,0.025	0.05
mama	50.000 (0,100)	0.793 (0.359,1.227)	6.748 (5.592,7.034)	3.942 (0,11.656)	0.25,0.05	0.05
CNS	1.777 (0,4.023)	0.722 (0.385,1.058)	6.701 (0,7)	8.025 (0,35.23)	0.25,0.025	0.1
Thyroid	13.840 (8.54,19.14)	0.163 (0.024,0.30)	5.649 (5.436,5.792)	1.904 (0,4.02)	0.3,0.03	0.05
Esophagus	5.000 (0,10)	1.186 (0.429,1.943)	7.700 (0,8.03)	44.348 (29.176,59.520)	0.25,0.025	0.05

Table 4. Best-fit parameters and their uncertainty ranges for selected organs. Note that only X, Y, Z, and δ were fitted. α, β, and λ were taken from the literature [20][21][22][23]. Note that the model appears to be almost entirely independent of the value of X, which consistently reflected the entirety of the biologically plausible range.

5.4 Adjustments to the CNS

In Figure 4, it is evident that all organs fit the data well for all ages except for the CNS. This is an issue that was initially raised by Shuryak. The underestimation that occurs for subjects under age forty is likely due to the model's assumption that there are no premalignant cells at birth. Yet CNS cancers are notorious for their genetic origin and thus require a modification of the model of background incidence. If we assume that an individual is born with a k number of premalignant cells, Equation 8 can be replaced by

$$A_{bac} = \frac{a}{b}\left(e^{bT_x} - 1 + k\right)e^{-c\left(T_x+T_y\right)^2} \qquad (9)$$

The plot of background cancer incidence of CNS using this new equation is shown in Figure 5.

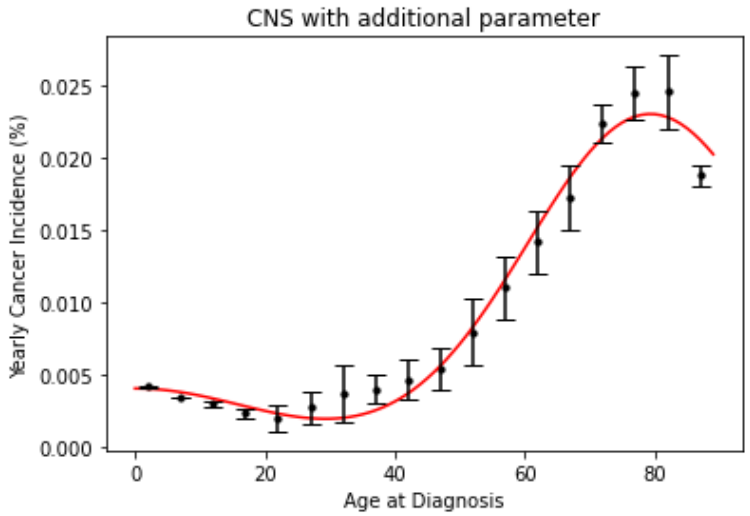


Figure 5. Best-fit model predictions utilizing the additional parameter k. When using this parameter, the values for a, b, and c should be replaced by 0.651×10^{-6},0.219 and 1.380×10^{-3}, respectively.

The model now fits the cancer rates at early ages more precisely, but it must be kept in mind that the modification to the background incidence model should only be used for cancers with high genetic causes, such as the central nervous system.

5.5 Fitted parameters and their uncertainties

After fitting, Montecarlo simulations of the parameter distributions were performed using random sampling. The plots can be found in Figure 5, where darker colors indicate a higher likelihood of occurrence.

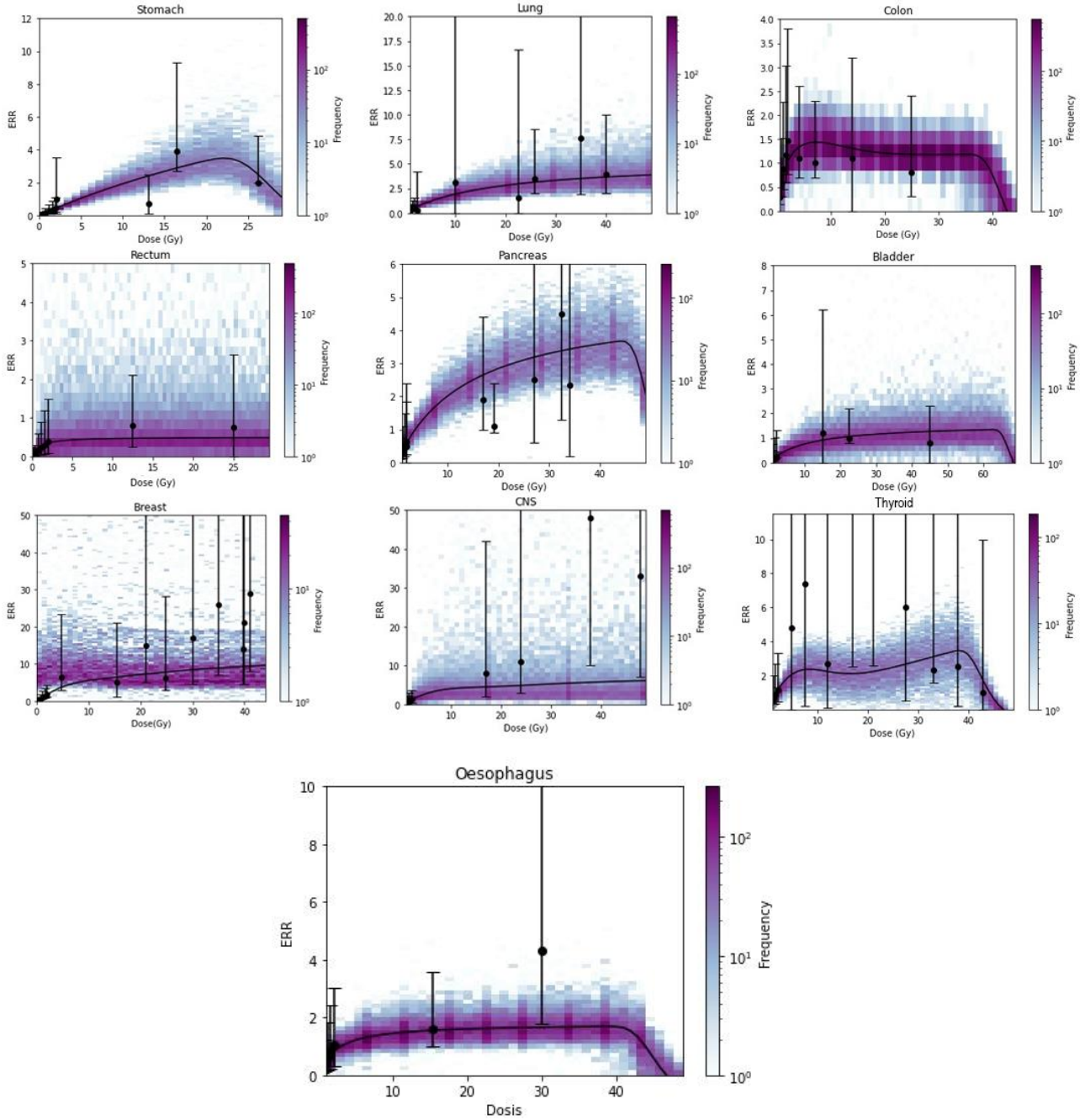
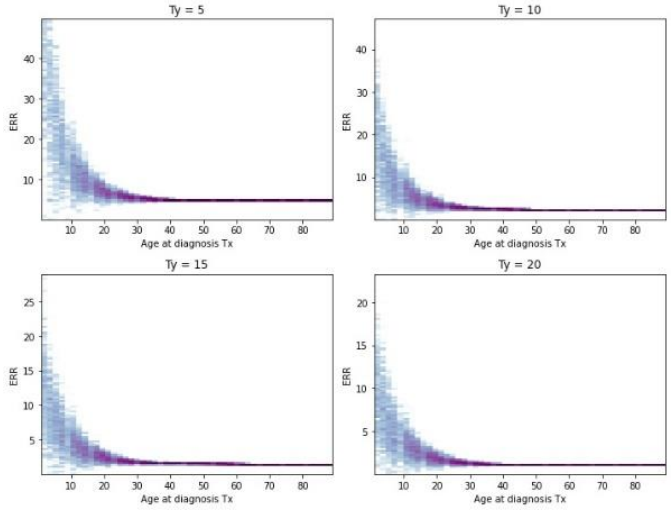


Figure 6. Results of the Montecarlo simulation for the ERR values for the selected organs. The black dots and lines correspond to the epidemiological data and their respective reported uncertainties.

5.6 Analyzing X

When fitting, it became apparent that the model seemed to have only a loose dependence on the parameter X. To better understand this, another curve was plotted, this time fixing all parameters except for X, the values of which were randomly drawn from the range shown in Table 4. The resulting plots are shown in Figure 7.

Colon



Bladder

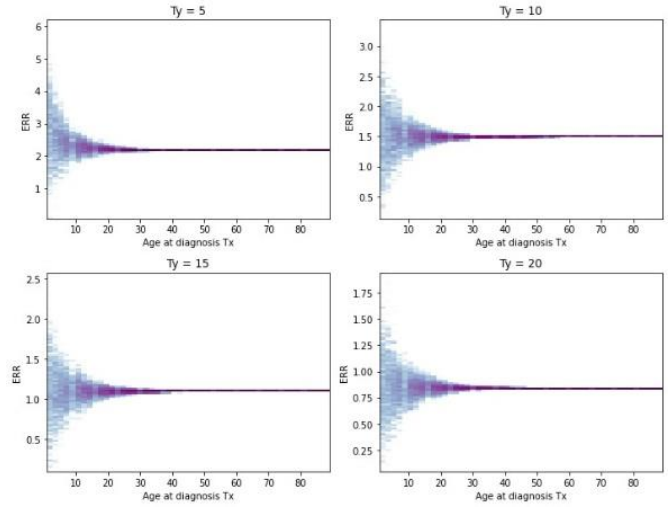


Figure 7. On the Left, ERR for Colon as a function of age at diagnosis (Tx) for four different values of time since diagnosis (Ty). On the right, ERR for Bladder under the same conditions. Keep in mind that graphs have different scales on the ERR axis. It is interesting to note that changes in X only have a noticeable impact if the age at diagnosis is lower than forty. Beyond that point, the X can take almost any value, and the model thus becomes independent of it.

While colon cancer at early ages is a relatively rare occurrence, one must consider that the model works with *relative* risk, so apparently large percentage increases often translate into a small number of additional cases. The behavior of the model as a function of parameter X remains consistent with other studied organs.

The justification of why X can be neglected at older ages is the fact that as T_x grows, there can be several orders of magnitude of difference between the quantities being added on the left side of equation 14 ($\left(e^{bT_x}-1\right)Sf(Z,D)$ and $+b\cdot X\cdot ISf(D)$). In other words, there exists such a large number of premalignant cells that have been accumulating during the individual's life that the few initiated ones become negligible. When the individual is young, however, since there are fewer premalignant cells present, initiation plays a much more important role. So, in cases when $T_x > 40\ years$, we can replace equation 6 with the much simpler:

$$ERR_{T_x>40} = \frac{\frac{1+YD}{\left[1+YD\left(1-e^{-\delta T_y}\right)\right]}\left(\left(e^{bT_x}-1\right)Sf(Z,D)\right)e^{bT_y}+\left(e^{bT_y}-1\right)}{e^{b\left(T_x+T_y\right)}-1} \quad (16)$$

5.5 The impact of uncertainties in Tx and Ty

All of the selected studies indicated the average attained age and average latency time before the apparition of second primary cancer. This obscures information that is relevant for a model that heavily depends on the time evolution of premalignant cells. To better understand the impact of averaging the ages of the cohort, ERR was plotted against different combinations of age at diagnosis (T_x) and time until apparition T_y, in Figures 8 and 9.

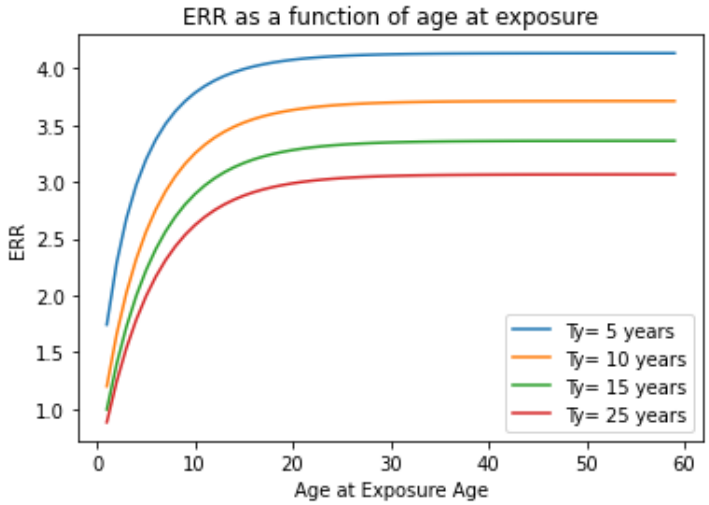


Figure 8. ERR for the stomach as a function of age at exposure for different values of Ty. This phenomenon is noticeable for all organs. The stomach was chosen for no particular reason. Similarly, an absorbed dose of 20 Gy at 17 fractions was chosen arbitrarily to illustrate the impact of the uncertainties in T_x and T_y.

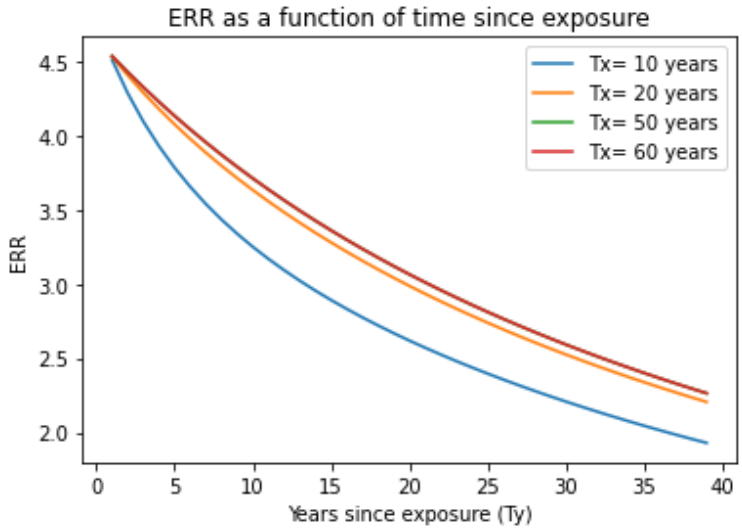


Figure 9. ERR for the stomach as a function of time since exposure for different values of Tx. This phenomenon is noticeable for all organs. The stomach was chosen for no particular reason. Similarly, an absorbed dose of 20 Gy at 17 fractions was chosen arbitrarily to illustrate the impact of the uncertainties in T_x and T_y.

Both plots clearly show significant variations in ERR for even relatively small age differences, which clearly indicates the impact of an uncertainty source that cannot be easily accounted for with the current method.

5.8 Results of the prostate plans

The resulting ERR estimation for the chosen prostate plans can be seen in Table 5.

Plan	Lung		Stomach		Colon		Rectum		Bladder		Thyroid		Total	
	ERR	Std	ERR	Std	ERR	Std	ERR	Std	ERR	Std	ERR	Std	ERR	Std
PRA6	0.0169	0.0015	0.0180	0.0023	0.1552	0.0053	0.0031	0.0008	0.3758	0.0518	0.0856	0.0063	0.6546	0.0525
SIMi6	0.0148	0.0005	0.0177	0.0022	0.1614	0.0050	0.0039	0.0008	0.3802	0.0450	0.0727	0.0051	0.6507	0.0456
EIMi6	0.0160	0.0011	0.0174	0.0032	0.1572	0.0051	0.0030	0.0005	0.3799	0.0367	0.0762	0.0055	0.6497	0.0377
SIMD6	0.0130	0.0016	0.0164	0.0026	0.1562	0.0063	0.0038	0.0007	0.3846	0.0432	0.0641	0.0021	0.6381	0.0438
VIMi6	0.0131	0.0011	0.0148	0.0029	0.1563	0.0073	0.0033	0.0008	0.3860	0.0636	0.0624	0.0033	0.6359	0.0642
EIMd6	0.0150	0.0017	0.0174	0.0028	0.1607	0.0062	0.0028	0.0007	0.3600	0.0580	0.0726	0.0043	0.6284	0.0586
VVM6	0.0124	0.0009	0.0133	0.0017	0.1559	0.0050	0.0036	0.0008	0.3764	0.0204	0.0587	0.0030	0.6203	0.0213
S3D6	0.0122	0.0007	0.0135	0.0018	0.1552	0.0037	0.0051	0.0014	0.3809	0.0598	0.0524	0.0030	0.6193	0.0601
V3D6	0.0119	0.0002	0.0133	0.0026	0.1559	0.0051	0.0042	0.0013	0.3794	0.0412	0.0527	0.0033	0.6173	0.0418
E3D6	0.0123	0.0008	0.0138	0.0018	0.1540	0.0055	0.0043	0.0016	0.3568	0.0300	0.0550	0.0032	0.5961	0.0308
VSB6	0.0027	0.0002	0.0034	0.0005	0.0953	0.0031	0.0009	0.0003	0.2638	0.0228	0.0119	0.0006	0.3780	0.0230

Table 5. ERR and standard deviation for the generated distributions obtained using the new parameters for Shuryak's model. The plan labels can be interpreted as follows: the first letter indicates the manufacturer of the linear accelerator (V for Varian, S for Siemens, E for Elekta, P for Philips). The middle letters correspond to the technique used (SB for SBRT, 3D for 3D-CRT, IMi for Inverse IMRT, IMd for direct IMRT, VM for VMAT, RA for rapid arc).

It can be seen by comparing the total ERR and their standard deviation that a hierarchy between plans can clearly be established. However, the model must still be tested against an alternative one. Tables 6 presents a comparison between the results using Shuryak's model and those found using the BEIR VII model. Results from the original study [34] were found using a combination of Schneider's and the BEIR model. The corresponding risks were transcribed into Table 7.

Plan	Total ERR	Total Std
PRA6 (S)	0.6546	0.0525
SIMi6 (S)	0.6507	0.0456
EIMi6 (S)	0.6497	0.0377
SIMD6 (S)	0.6381	0.0438
VIMi6 (S)	0.6359	0.0642
EIMd6 (S)	0.6284	0.0586
VVM6 (S)	0.6203	0.0213
S3D6 (S)	0.6193	0.0601
V3D6 (S)	0.6173	0.0418
E3D6 (S)	0.5961	0.0308
VSB6 (S)	0.3780	0.0230
PRA6	0.2764	0.0819
EIMi6	0.2527	0.0954
SIMi6	0.2436	0.0686
VVM6	0.2387	0.0778
EIMd6	0.2332	0.0705
E3D6	0.2295	0.0627
SIMD6	0.2259	0.0625
V3D6	0.2259	0.0753
VIMi6	0.2238	0.0709
S3D6	0.2199	0.0785
VSB6	0.0839	0.0291

Table 6: Comparison between the results found using Shuryak's model with the new parameters and BEIR VII. Here, (B) stands for BEIR VII and (S) for Shuryak. Here, plans are ordered from highest to lowest total ERR.

Plan	Total Risk
PRA6 (S)	0.6545
SIMi6 (S)	0.6506
EIMi6 (S)	0.6497
SIMD6 (S)	0.6380
VIMi6 (S)	0.6358
EIMd6 (S)	0.6284
VVM6 (S)	0.6203
S3D6 (S)	0.6193
V3D6 (S)	0.6173
E3D6 (S)	0.5961
VSB6 (S)	0.3780
E3D6 (Sc)	1.49
EIMi6 (Sc)	1.49
S3D6 (Sc)	1.44
SIMD6 (Sc)	1.41
VRIMi6 (Sc)	1.4
EIMd6 (Sc)	1.39
VR3D6 (Sc)	1.39
SIMi6 (Sc)	1.37
PRA6 (Sc)	1.35
VRVM6 (Sc)	1.35
VSB6 (Sc)	1.09

Table 7: Comparison between the results found using Shuryak's model and those reported by Sanchez-Nieto et al. The latter risk is reported as LAR [34].

It is clear that Shuryak's model tends to underestimate the risk at very low doses while overestimating it at higher doses, with respect to the BEIR VII model.

While there are clear agreements between the models on the lowest risk plans, determining a hierarchy on the upper end is difficult, as the differences between ERR are smaller than the uncertainties. Differences could also be explained by the fact that the Shuryak model for low doses was parametrized using data from the atomic bomb, which not only has large uncertainties but also corresponds to a different ethnic group, which is known to have different cancer rates. This can be seen by examining the difference in cancer rates in the SEER database and the Radiation Effects Research Foundation (RERF) [36].

Chapter 6: Discussion

Several important contributions were accomplished during the present work. Shuryak's model was rewritten in a more compact and accessible way by eliminating the parameter v and rewriting $F(k)$. New parameters were calculated using newer data from epidemiological studies, and a new organ, the oesophagus, was added entirely. It was possible to obtain a better fit for the CNS to the cancer incidence data found on the SEER database by introducing a new parameter k. The fitting process was improved by using a Bayesian inference fitting method from Python's pymc database. It is interesting to note that, for most cancers, variations of the value of the parameter X only marginally affected the results. As briefly mentioned, this is a consequence of the fact that the first term in the bracket in equation 6 can be several orders of magnitude larger than the second one for sufficiently large T_x. This implies that the predominating factor that leads to second primary cancers might not be the initiation of premalignant cells during RT but rather the process of hyperproliferation of the surviving premalignant cells.

This issue may be related to the fact that equation 8 assumes exponential growth of the premalignant cell niches. However, cancerous growth cannot really be assumed to be exponential in the long term, as competition with adjacent niches and homeostatic regulatory mechanisms impose a limit on tumoral growth. It is possible that a limited growth function might be better suited for the modeling of growth and replication of premalignant cells and would thus avoid the order of magnitude difference that renders initiation a negligible factor in this model.

The age dependence of the parameter X might be an indicator of a future need to allow for age dependence in the fitting process, further increasing its complexity but also its accuracy. This change would make sense, however, as it is known that younger people are more radiosensitive than older people. Regardless, the result obtained in equation 16 should be used for all patients over the age of 40 (which is the vast majority of RT patients), as it significantly simplifies the expression and reduces computation time, which is even more impactful if implemented in a treatment planning system that requires large numbers of computations.

Despite the fact that Shuryak's model was clearly capable of distinguishing between different prostate plans, no clear agreement existed between different risk estimation models. Further study is required not only to obtain better-adjusted parameters by reduced uncertainty in the data fed to Shuryak's model but also to evaluate its reliability and accuracy.

In conclusion, Shuryak's model was successfully implemented and reparametrized, with a new organ being added altogether. Several simplifications were made, which will allow for easier access to the model. However, it was also made clear the impact that uncertainties have on a model that considers such a large number of parameters, and thus, before being clinically viable, Shuryak's model will require epidemiological studies that contain all the information that it needs with the lowest uncertainty possible.

Appendix: Python code

This is a sample code used for fitting the rectum. All other organs were similarly fitted, with their respective data.

```python
import pymc as pm
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.colors
import arviz as az
# Fixed values
b_original=0.217
alpha = 0.25
beta = 0.025
l = 0.1

#ranges for low dose ERR based on Preston et al
lower_range=0.04
avg_range=0.19
higher_range=0.56
#age considered
Tx=35
Ty=18

dosises = np.array([0.4, 0.8, 1.2, 1.6, 2, 12.5,25])


ERR_lower = [lower_range * 0.4, lower_range * 0.8, lower_range * 1.2, lower_range * 1.6, lower_range * 2,0.25,0]
ERRs = [avg_range * 0.4, avg_range * 0.8, avg_range * 1.2, avg_range * 1.6, avg_range * 2,0.8,0.75]
ERR_higher =np.array([higher_range * 0.4, higher_range * 0.8, higher_range * 1.2, higher_range * 1.6, higher_range * 2.0,1.3,1.9])
Error_lower = [ERRs[i]-ERR_lower[i]for i in range(len(ERRs))]
sigmas=[(ERR_higher[i]-ERR_lower[i])/4 for i in range(len(ERR_lower))]
def nmenos(i,num,S):
    if i==num:
        return S*nmenos(i-1,num,S)
    elif i==0:
        return 1
    else:
        if (i-1)%5==0 and i!=1:
            T=3
        else:
            T=1
        return(1/(1+(1/(S*nmenos(i-1,num,S))-1)*np.exp(-l*T)))

def F(k,num,S):
    if k==num:
        return 1
    else:
        suma=1
        for j in range(k+1, num+1):
            sumando=nmenos(k,num,S)*(1-S)/nmenos(j,num,S)
            suma+=sumando
        return(1/suma)

def Sf(num,S,Z):
    return(1-(1-F(0,num,S))**Z)
```

```
def Isf(num,d,S):
    suma=0
    for i in range(1,num+1):
        suma+=d*S*nmenos(i,num,S)*F(i,num,S)
    return suma


# Define your model using PyMC3
with pm.Model() as model:
    # Defining prior distributions for parameters
    X = pm.Uniform('X', lower=0, upper=10)
    Y = pm.Uniform('Y', lower=0, upper=10)
    Z = pm.Uniform('Z', lower=0, upper=10**12)
    delta = pm.Uniform('delta', lower=0, upper=3e-1)


    def ERR_pymc():
        #considering that different plans can have different fractionation
        for value in dosises:
            if value in [0.4, 0.8, 1.2, 1.6, 2]:
                num = 1
                break
            elif value == 12.5:
                num = 15
                break
            elif value == 25:
                num = 30
                break
        d=dosises/num
        S=np.exp(-alpha*d-beta*d**2)
        Q1=(1+Y*dosises)/(1+Y*dosises*(1-np.exp(-delta*Ty)))
        Q2=((np.exp(b_original*Tx)-1)*Sf(num,S,Z)+b_original*X*Isf(num,d,S))*np.exp(b_original*Ty)
        Q3=np.exp(b_original*Ty)-1
        Q4=np.exp(b_original*(Tx+Ty))-1
        ERR=(Q1*Q2+Q3)/Q4-1
        return(ERR)

    # Defining the likelihood
    mu = ERR_pymc()
    likelihood = pm.Normal('likelihood', mu=ERR_pymc(), sigma=sigmas, observed=ERRs)


with model:
    trace = pm.sample(5000, tune=1000, chains=2, cores=1)
hdi_result = az.hdi(trace, hdi_prob=0.95)


X_sigma=trace.posterior['X'].std().item()
X_mu=trace.posterior['X'].mean().item()

```

```
Y_mu=trace.posterior['Y'].mean().item()
Y_sigma=trace.posterior['Y'].std().item()

Z_mu=trace.posterior['Z'].mean().item()
Z_sigma=trace.posterior['Z'].std().item()

delta_mu=trace.posterior['delta'].mean().item()
delta_sigma=trace.posterior['delta'].std().item()
print("Rectum Parameters")
print("X_sigma =", X_sigma)
print("X_mu =", X_mu)

print("Y_sigma =", Y_sigma)
print("Y_mu =",Y_mu)
print("Z_sigma =", Z_sigma)
print("Z_mu =", Z_mu)
print("delta_sigma =", delta_sigma)
print("delta_mu =", delta_mu)
def ERR_plot(dosis,X,Y,Z,delta,num):
    lista=[]
    for i in dosis:
        d=i/num


        S=np.exp(-alpha*d-beta*d**2)
        Q1=(1+Y*i)/(1+Y*i*(1-np.exp(-delta*Ty)))
        Q2=((np.exp(b_original*Tx)-1)*Sf(num,S,Z)+b_original*X*Isf(num,d,S))*np.exp(b_original*Ty)
        Q3=np.exp(b_original*Ty)-1
        Q4=np.exp(b_original*(Tx+Ty))-1
        lista.append((Q1*Q2+Q3)/Q4-1)

    return(lista)
def ERR_montecarlo(dosis,X,Y,Z,delta,num):
    d=dosis/num
    S=np.exp(-alpha*d-beta*d**2)
    Q1=(1+Y*dosis)/(1+Y*dosis*(1-np.exp(-delta*Ty)))
    Q2=((np.exp(b_original*Tx)-1)*Sf(num,S,Z)+b_original*X*Isf(num,d,S))*np.exp(b_original*Ty)
    Q3=np.exp(b_original*Ty)-1
    Q4=np.exp(b_original*(Tx+Ty))-1
    return((Q1*Q2+Q3)/Q4-1)




#creating random distributions for each parameter
distX=np.random.normal(X_mu,X_sigma,10000)
distY=np.random.normal(Y_mu,Y_sigma,10000)
distZ=np.random.normal(Z_mu,Z_sigma,10000)
distdelta=np.random.normal(delta_mu,delta_sigma,10000)
```

```
resultados=[]
x_values_montecarlo=[]
for j in range(60):

    for i in range(500):

        X_chosen=np.random.choice(distX)
        Y_chosen=np.random.choice(distY)
        Z_chosen=np.random.choice(distZ)
        delta_chosen=np.random.choice(distdelta)
        resultados_temp=(ERR_montecarlo(j/2,X_chosen,Y_chosen,Z_chosen,delta_chosen,17))
        if resultados_temp>0 and resultados_temp <=50:
            resultados.append(resultados_temp)
            x_values_montecarlo.append(j/2)


x_values=[i/2 for i in range(60)]


fitted_ERRs_posterior = ERR_plot(x_values,trace.posterior['X'].mean(), trace.posterior['Y'].mean(), trace.posterior['Z'].mean(), trace.posterior['delta'].mean(),17)

plt.scatter(x_values_montecarlo,resultados,s=1)
plt.plot(x_values, fitted_ERRs_posterior,c='black')

plt.errorbar(dosises, ERRs, yerr=[Error_lower,ERR_higher], fmt='o', capsize=5,c="black")
plt.ylabel('ERR')
plt.xlabel('Dosis (Gy)')
plt.title('Recto')
plt.ylim([0,5])
plt.show()

log_norm = matplotlib.colors.LogNorm()
plt.hist2d(x_values_montecarlo, resultados, bins=(60, 200), cmap=plt.cm.BuPu, norm=log_norm)
plt.errorbar(dosises, ERRs, yerr=[Error_lower,ERR_higher], fmt='o', capsize=5,c="black")
plt.plot(x_values, fitted_ERRs_posterior,c='black')
plt.colorbar(label='Frequency')
plt.xlabel('Dose')
plt.title('Rectum')
plt.ylabel('ERR')
plt.ylim([0,5])
plt.show()



idata = az.convert_to_inference_data(trace)

# Plot the posteriors using ArviZ
az.plot_posterior(idata, var_names=['X', 'Y', 'Z', 'delta'], hdi_prob=0.95)


plt.show()
```

References

[1] N. Committee to Assess Health Risks from Exposure to Low Levels of Ionizing Radiation, Beir Vii, vol. 18, no. 3. 2006.

[2] ICRP International Commission on Radiological Protection, "Annals of the ICRP Annals of the ICRP Annals of ICRP," ICRP Publ. 92, Ann. ICRP 28, no. 0, pp. 1–337, 2003.

[3] I. Diallo et al., "Frequency Distribution of Second Solid Cancer Locations in Relation to the Irradiated Volume Among 115 Patients Treated for Childhood Cancer," Int. J. Radiat. Oncol. Biol. Phys., vol. 74, no. 3, pp. 876–883, 2009, doi: 10.1016/j.ijrobp.2009.01.040.

[4] Schneider U. Modeling the risk of secondary malignancies after radiotherapy. Genes (Basel). 2011 Nov 29;2(4):1033-49. doi: 10.3390/genes2041033. PMID: 24710304; PMCID: PMC3927608.

[5] D. Hanahan and R. A. Weinberg, "Hallmarks of cancer: The next generation," Cell. 2011, doi: 10.1016/j.cell.2011.02.013.

[6] Preston DL, Ron E, Tokuoka S, Funamoto S, Nishi N, Soda M, Mabuchi K, Kodama K. Solid cancer incidence in atomic bomb survivors: 1958-1998. Radiat Res. 2007 Jul;168(1):1-64. doi: 10.1667/RR0763.1. PMID: 17722996.

[7] Dasu A, Toma-Dasu I. Models for the risk of secondary cancers from radiation therapy. Phys Med. 2017 Oct;42:232-238. doi: 10.1016/j.ejmp.2017.02.015. Epub 2017 Feb 24. PMID: 28242139.

[8] Unscear, U., 2000. Report to the General Assembly, with scientific annexes. United Nations.

[9] Boice JD Jr et al, Radiation dose and second cancer risk in patients treated for cancer of the cervix. Radiat Res. 1988 Oct;116(1):3-55. PMID: 3186929.

[10] B Tucker MA, Jones PH, Boice JD Jr, Robison LL, Stone BJ, Stovall M, Jenkin RD, Lubin JH, Baum ES, Siegel SE, et al. Therapeutic radiation at a young age is linked to secondary thyroid cancer. The Late Effects Study Group. Cancer Res. 1991 Jun 1;51(11):2885-8. PMID: 1851664.

[11] Robert H Davis, Production and killing of second cancer precursor cells in radiation therapy: in regard to Hall and Wuu (Int J Radiat Oncol Biol Phys 2003;56:83–88),International Journal of Radiation Oncology*Biology*Physics,

[12] Sachs RK, Brenner DJ. Solid tumor risks after high doses of ionizing radiation. Proc Natl Acad Sci U S A. 2005 Sep 13;102(37):13040-5. doi: 10.1073/pnas.0506648102. Epub 2005 Sep 6. PMID: 16150705; PMCID: PMC1199000.

[13] Schneider, U. (2009), Mechanistic model of radiation-induced cancer after fractionated radiotherapy using the linear-quadratic formula. Med. Phys., 36: 1138-1143. https://doi.org/10.1118/1.3089792

[14] I. Shuryak, P. Hahnfeldt, L. Hlatky, R. K. Sachs, and D. J. Brenner, "A new view of radiation-induced cancer: Integrating short- and long-term processes. Part I: Approach," *Radiat. Environ. Biophys.*, vol. 48, no. 3, pp. 263–274, 2009, doi: 10.1007/s00411-009-0230-3.

[15] I. Shuryak, P. Hahnfeldt, L. Hlatky, R. K. Sachs, and D. J. Brenner, "A new view of radiation-induced cancer: Integrating short- and long-term processes. Part II: Second cancer risk estimation," *Radiat. Environ. Biophys.*, vol. 48, no. 3, pp. 275–286, 2009, doi: 10.1007/s00411-009-0231-2.

[16] I. Shuryak, P. Hahnfeldt, L. Hlatky, R. K. Sachs, and D. J. Brenner, "Erratum to: A new view of radiation-induced cancer: integrating short- and long-term processes. Parts I and II," *Radiat. Environ. Biophys.*, 2011, doi: 10.1007/s00411-011-0378-5.

[17] Greaves LC et al, Preston SL, Tadrous PJ, Taylor RW, Barron MJ, Oukrif D, Leedham SJ, Deheragoda M, Sasieni P, Novelli MR, Jankowski JA, Turnbull DM, Wright NA, McDonald SA (2006) Mitochondrial DNA mutations are established in human colonic stem cells, and mutated clones expand by crypt fission. Proc Natl Acad Sci USA 103:714–719

[18] Johnston MD, Edwards CM, Bodmer WF, Maini PK, Chapman SJ (2007) Mathematical modeling of cell population dynamics in the colonic crypt and in colorectal cancer. Proc Natl Acad Sci USA 104:4008–401

[19] Edwards CM, Chapman SJ (2007) Biomechanical modelling of colorectal crypt budding and fission. Bull Math Biol 69:1927– 1942

[20] Bentzen SM, Ruifrok AC, Thames HD (1996) Repair capacity and kinetics for human mucosa and epithelial tumors in the head and neck: clinical data on the effect of changing the time interval between multiple fractions per day in radiotherapy. Radiother Oncol 38:89–101

[21] Challeton C, Branea F, Schlumberger M, Gaillard N, de Vathaire F, Badie C, Antonini P, Parmentier C (1997) Characterization and radiosensitivity at high or low dose rate of four cell lines derived from human thyroid tumors. Int J Radiat Oncol Biol Phys 37:163–169

[22] Thames HD, Bentzen SM, Turesson I, Overgaard M, van den Bogaert W (1989) Fractionation parameters for human tissues and tumors. Int J Radiat Biol 56:701–710

[23] Thames HD, Bentzen SM, Turesson I, Overgaard M, Van den Bogaert W (1990) Time-dose factors in radiotherapy: a review of the human data. Radiother Oncol 19:219–235

[24] "SEER Incidence Data, 1975-2023." https://seer.cancer.gov/data/.

[25] Travis LB, Gospodarowicz M, Curtis RE, Clarke EA, Andersson M, Glimelius B, Joensuu T, Lynch CF, van Leeuwen FE, Holowaty E, Storm H, Glimelius I, Pukkala E, Stovall M, Fraumeni JF Jr, Boice JD Jr, Gilbert E (2002) Lung cancer following chemotherapy and radiotherapy for Hodgkin's disease. J Natl Cancer Inst 94:182–192

[26] Travis LB, Hill DA, Dores GM, Gospodarowicz M, van Leeuwen FE, Holowaty E, Glimelius B, Andersson M, Wiklund T, Lynch CF, Van't Veer MB, Glimelius I, Storm H, Pukkala E, Stovall M, Curtis R, Boice JD Jr, Gilbert E (2003) Breast cancer following radiotherapy and chemotherapy among young women with Hodgkin disease. JAMA 290:465–475

[27] Travis LB, Fossa SD, Schonfeld SJ, McMaster ML, Lynch CF, Storm H, Hall P, Holowaty E, Andersen A, Pukkala E, Andersson M, Kaijser M, Gospodarowicz M, Joensuu T, Cohen RJ, Boice JD Jr, Dores GM, Gilbert ES (2005) Second cancers among 40, 576 testicular cancer patients: focus on long-term survivors. J Natl Cancer Inst 97:1354–1365

[28] van Leeuwen FE, Klokman WJ, Stovall M, Dahler EC, van't Veer MB, Noordijk EM, Crommelin MA, Aleman BM, Broeks A, Gospodarowicz M, Travis LB, Russell NS (2003) Roles of radiation dose, chemotherapy, and hormonal factors in breast cancer following Hodgkin's disease. J Natl Cancer Inst 95:971– 980

[29] Travis LB, Curtis RE, Boice JD Jr, Platz CE, Hankey BF, Fraumeni JF Jr (1996) Second malignant neoplasms among long-term survivors of ovarian cancer. Cancer Res 56:1564–1570

[30] Travis LB, Curtis RE, Storm H, Hall P, Holowaty E, Van Leeuwen FE, Kohler BA, Pukkala E, Lynch CF, Andersson M, Bergfeldt K, Clarke EA, Wiklund T, Stoter G, Gospodarowicz M, Sturgeon J,

Fraumeni JF Jr, Boice JD Jr (1997) Risk of second malignant neoplasms among long-term survivors of testicular cancer. J Natl Cancer Inst 89:1429–143

[31] Gilbert ES, Stovall M, Gospodarowicz M, Van Leeuwen FE, Andersson M, Glimelius B, Joensuu T, Lynch CF, Curtis RE, Holowaty E, Storm H, Pukkala E, van't Veer MB, Fraumeni JF, Boice JD Jr, Charke EA, Travis LB (2003) Lung cancer after treatment for Hodgkin's disease: focus on radiation effects. Radiat Res 159:161–173

[32] Neglia JP, Robison LL, Stovall M, Liu Y, Packer RJ, Hammond S, Yasui Y, Kasper CE, Mertens AC, Donaldson SS, Meadows AT, Inskip PD (2006) New primary neoplasms of the central nervous system in survivors of childhood cancer: a report from the Childhood Cancer Survivor Study. J Natl Cancer Inst 98:1528– 1537

[33] Ronckers CM, Sigurdson AJ, Stovall M, Smith SA, Mertens AC, Liu Y, Hammond S, Land CE, Neglia JP, Donaldson SS, Meadows AT, Sklar CA, Robison LL, Inskip PD (2006) Thyroid cancer in childhood cancer survivors: a detailed evaluation of radiation dose response and its modifiers. Radiat Res 166:618–628

[34] Sánchez-Nieto B, Romero-Expósito M, Terrón JA, Irazola L, García Hernández MT, Mateos JC, Roselló J, Planes D, Paiusco M, Sánchez-Doblado F. External photon radiation treatment for prostate cancer: Uncomplicated and cancer-free control probability assessment of 36 plans. Phys Med. 2019 Oct;66:88-96. doi: 10.1016/j.ejmp.2019.09.076. Epub 2019 Sep 29. PMID: 31574423.

[35] BEIR, Health Risks from Exposure to Low Levels of Ionizing Radiation: BEIR VII, Phase 2 (National Academy of Science, Washington, DC, 2006).

[36] Radiation Effects Research Foundation https://www.rerf.or.jp/en/

[37] Fung C, Fossa SD, Beard CJ, Travis LB. Second malignant neoplasms in testicular cancer survivors. J Natl Compr Canc Netw. 2012 Apr;10(4):545-56. doi: 10.6004/jnccn.2012.0052. PMID: 22491051.