



**HAL**  
open science

## Tumor growth modeling based on cell and tumor lifespans

Roukaya Keinj, Thierry Bastogne, Pierre Vallois

► **To cite this version:**

Roukaya Keinj, Thierry Bastogne, Pierre Vallois. Tumor growth modeling based on cell and tumor lifespans. *Journal of Theoretical Biology*, 2012, 312, pp.76-86. 10.1016/j.jtbi.2012.07.008. hal-00730315

**HAL Id: hal-00730315**

**<https://hal.science/hal-00730315v1>**

Submitted on 9 Sep 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Tumor growth modeling based on cell and tumor lifespans

R. Keinj<sup>1</sup>, T. Bastogne<sup>2,4,6</sup>, P. Vallois<sup>3,5,6</sup>

<sup>1</sup>Université Joseph Fourier, Laboratoire Jean Kuntzmann,  
Equipe Projet INRIA MOISE, B.P. 53, 38041 Grenoble Cedex 9, France

<sup>2</sup>Université de Lorraine, CRAN, UMR 7039,  
Campus Sciences, BP 70239, Vandoeuvre-lès-Nancy Cedex, 54506, France

<sup>3</sup>Université de Lorraine, IECN, UMR 7502,  
Campus Sciences, BP 70239, Vandoeuvre-lès-Nancy Cedex, 54506, France

<sup>4</sup>CNRS, CRAN, UMR 7039, France

<sup>5</sup>CNRS, IECN, UMR 7502, France

<sup>6</sup>INRIA, BIGS, France

September 9, 2012

## Abstract

This paper deals with the lifespan modeling of heterogenous tumors treated by radiotherapy. A bi-scale model describing the cell and tumor lifespans by random variables is proposed. First and second-order moments, as well as the cumulative distribution functions and confidence intervals are expressed for the two lifespans with respect to the model parameters. One interesting result is that the mean value of the tumor lifespan can be approached by a logarithmic function of the initial cancer cell number. Moreover, we show that TCP and NTCP, used in radiotherapy to evaluate, optimize and compare treatment plans, can be derived from the tumor lifespan and the surrounding healthy tissue respectively. Finally, we propose a ROC curve, entitled ECT (Efficiency-Complication Trade-off), suited to the selection by clinicians of the appropriate treatment planning.

**Keywords:** Markov chain; cancer cells; radiotherapy

## 1 Introduction

Cancer is a disease that affects millions of people worldwide. One of the common therapies used to treat cancer is external beam radiotherapy. The ionizations induced by radiation cause a variety of possible lesions in cells [3] and the most harmful damage are the lesions which affect the DNA structure [18, 10]. Probabilistic modeling is a helpful tool for describing these biological damage. For

instance, the tumor control probability (TCP) [19, 4, 8] and the normal tissue complication probability (NTCP) [14, 11] are used to characterize and evaluate radiotherapy treatment planning. Their mathematical expressions can be derived from different stochastic models of the tumor response such as the linear quadratic model [7, 19], cell population-dynamic models [17], mixed-effects behavioral models [1] and cell-cycle models [13]. The main drawback of those mathematical representations is their inability to handle biological heterogeneity. There are different types of heterogeneity [15] but we have chosen to focus on the tumor damage heterogeneity. A large majority of models suppose that the cell sensitivity to radiation is constant during the treatment and over the entire cell population. Meaning that a surviving cell is thought to be as viable as an unirradiated cell and that all cells are supposed to have the same survival probability. However, evidence suggests that a damaged cell partially loses its ability to resist. As stressed in [9, 6], the intratumor heterogeneity of cell phenotypes or damage is of direct clinical importance. Therefore, the clinical challenge is to define the suited treatment duration for each patient by accounting for variability of the therapeutic response.

In a previous study [12], we proposed a *multinomial* model of tumor response based on a discrete-time Markov chain. This model is derived from the *Target Theory* and assumes there exists a number of radio-sensitive sites within the cell, called targets. The cell death is finally caused by the deactivation of those targets by radiation particules. We showed in [12] that the *multinomial* model is a generalization of typical target models [2, 16], able to account for the heterogeneity of cell damage caused by the treatment. However, like the majority of models used to measure the tumor response to treatment, the *multinomial* model examines the number of surviving cells in the tumor and not the tumor lifespan.

In this paper, we firstly address the stochastic modeling of the tumor lifespan. We start by considering the lifespan of a single cancer cell that behaves as described in [12]. It is important to emphasize that the cell lifespan is the minimal random number of radiation dose fractions to be applied to kill the cell. We study this random time by calculating its mean, variance and cumulative distribution function. We then assume that a tumor is a group of independent cells. This allows us to define the lifespan of the tumor as the maximum of individual lifespans. In practice, the tumor lifespan is of clinical importance since it corresponds to the minimal number of radiation dose fractions (*i.e.* treatment duration) to be applied to completely eradicate the tumor. When the initial number  $n_0$  of cancer cells is not too large, we can explicitly calculate the mean, variance and the cumulative distribution function of the tumor lifespan. When  $n_0$  is large, the previous parameters are no longer calculable. However, we show that, under some assumptions, the mean lifespan of the tumor behaves as a logarithmic function of the initial number  $n_0$ . The second goal is to show that TCP and NTCP can be completely formulated with respect to the tumor and normal tissue lifespans. These expressions of TCP and NTCP are finally used to propose a ROC curve, called ECT (Efficiency-Complication Trade-off), suited to the determination of the appropriate treatment schedule.

This paper is structured as follows: in Section 2 we give a reminder of the individual cell behavior (cf [12]). In Section 3, we study the cancer cell lifespan  $T$  as a random variable, determining its mean, variance and confidence intervals and a related set of numerical results is given. We introduce the tumor lifespan  $L$  in Section 4, taking the cell proliferation into account and we study theoretically and numerically this random variable. Expressions for TCP and NTCP are given in Section 5. Before concluding, we also propose the ECT diagram.

Table 1: Main notations

Notations	
Not.	Definition
$k$	discrete time related to the $k^{th}$ dose fraction
$u_0$	magnitude of each dose fraction
$Z_k$	number of deactivated targets in the cell
$\Pi$	transition matrix associated with the Markov chain $(Z_k)$
$\mathbf{P}$	matrix associated with treatment effects
$\mathbf{R}$	matrix associated with cell repair process
$m$	number of targets in a cancer cell
$q$	probability to deactivate a target in a cancer cell
$r$	probability for an inactive target to be reactivated in a cancer cell
$T$	cancer cell lifespan
$F$	cumulative distribution function of $T$
$I$	$1 - \theta$ confidence interval of $T$
$n_0$	initial total number of cancer cells in the tumor
$L$	lifespan of the tumor
$G$	cumulative distribution function of $L$
$J$	$1 - \theta$ confidence interval of $L$
$\bar{m}$	number of targets in a normal cell
$\bar{q}$	probability to deactivate a target in a normal cell
$\bar{r}$	probability for an inactive target to be reactivated in a normal cell
$\bar{T}$	normal cell lifespan
$\bar{F}$	cumulative distribution function of $\bar{T}$
$\bar{n}_0$	initial total number of normal cells in the irradiated zone
$\bar{n}$	the complication threshold number of dead normal cells
$L_{\bar{n}, \bar{n}_0}$	lifespan of the normal tissue

## 2 Behavior of a single cell

The main notations used thereafter are presented in Table 1 and  $\log(\cdot)$  denotes the natural logarithm function. In [12], it has been proposed a multinomial model of tumor growth relying on the target and hit modeling paradigm and based on a discrete-time Markov chain. We keep the same modeling assumptions stated in [12]. Since they play a crucial role in our model, we should briefly recall them:

- a cell has  $m$  targets;
- each target may be made inactive after the application of a fraction dose  $u_0$  with a probability  $q$ . The relationship between  $q$  and  $u_0$  is given in (5);
- the cell death, due to radiation, happens when the  $m$  targets are deactivated;
- between two consecutive dose fractions (*i.e.* during 24h separating two dose fractions in a conventional daily fractionated radiation schedule), if the cell is still alive then an inactive target may repair with a probability  $r$ ;
- we suppose that cancer cells that compose the tumor all have the same phenotype;
- to reduce complexity of the model, cell cycle positions are not accounted for;
- we assume there is no delay effect between the radiation dose applied to the mother cell and the damage consequences on daughter cells.

Let  $Z_k$  the random variable denoting the number of deactivated targets in the cell at time  $k$ , *i.e.* after the  $k^{th}$  dose fraction.  $k = 0$  corresponds to the beginning of treatment. Moreover, we assume that a constant fraction dose (typically  $u_0 = 2Gy$ ) is applied every day. We suppose that  $(Z_k)$  is a discrete-time Markov chain, *i.e.* the cell state at time  $k + 1$  only depends on the current state at time  $k$ .

### 2.1 Probability distribution of $Z_k$

Let  $\Pi$  be the corresponding transition matrix of  $(Z_k)$ . We briefly define  $\Pi$ , interested readers can refer to [12] for details. The dynamics of  $(Z_k)$  takes the effects of dose fractions and repair mechanisms into account :

$$\Pi = \mathbf{P}\mathbf{R}, \quad (1)$$

where  $\mathbf{P}$  models the treatment effects and  $\mathbf{R}$  describes repair mechanisms and given as follows

$$\mathbf{P}(i, j) = \begin{cases} \binom{m-i}{j-i} q^{j-i} (1-q)^{m-j} & i \leq j \\ 0 & j < i. \end{cases} \quad (2)$$

$$\mathbf{R}(i, j) = \begin{cases} \binom{i}{j} r^{i-j} (1-r)^j & j \leq i < m \\ 0 & i < j. \end{cases} \quad (3)$$

When  $i = m$ ,  $\mathbf{R}(m, m) = 1$  and  $\mathbf{R}(m, j) = 0$  for  $0 \leq j < m$ .

If we assume that the cell is initially in state  $i_0$  ( $i_0$  inactive targets), *i.e.*  $P(Z_0 = i_0) = 1$ , then

$$P(Z_k = i) = \Pi^k(i_0, i) \quad i \in \{0, \dots, m\}. \quad (4)$$

One interesting feature of our model is the possible simulation of the state trajectory  $k \mapsto Z_k$  by a recursive approach given in A. Moreover, the following proposition describes the influence of the model parameters  $q$ ,  $r$  and the initial cell state  $i_0$ , on the cell death probability.

**Proposition 2.1** *Suppose that  $m$  is fixed, the maps*

1.  $i \mapsto \Pi^k(i, m)$  and  $q \mapsto \Pi^k(i_0, m)$  are increasing;
2.  $r \mapsto \Pi^k(i_0, m)$  is decreasing.

The proof is based on coupling and identities (33) and (33) in A.

## 2.2 Deactivation parameter and radiation dose

The target deactivation probability  $q$  is a parameter associated with the sensitivity of cell damage to the magnitude of dose fractions  $u_0$ , which is a parameter of the radiation treatment schedule. To connect these two parameters, we use the linear quadratic (LQ) model with parameters  $\alpha$  and  $\beta$  (cf [5]). According to the latter, after one dose fraction  $u_0$  a cell dies with the probability  $q_c = 1 - e^{-\alpha u_0 - \beta u_0^2}$ . However, it is clear that in our model,  $q_c = \mathbf{P}(0, m) = q^m$ . Thus, we obtain

$$q = \left(1 - e^{-\alpha u_0 - \beta u_0^2}\right)^{\frac{1}{m}}. \quad (5)$$

Note that, we proved in [12] that the *multi-target-single-hit* model may be regarded as a specific case of the multinomial model when the repair parameter  $r$  is equal to 1. Similarly, if we fix the target number parameter  $m = 1$  and the deactivation target parameter  $q$  equals (5), then we recover the linear quadratic model.

## 2.3 Cell proliferation modeling

Let us now describe the behavior of a given cancer cell, initially in state 0, taking the reproduction phase into account. We suppose that the tumor is composed of only one cell phenotype to have a simple mathematical model, which allows us to explicit calculations. Once the  $k^{th}$  dose fraction is applied, the cell may be damaged, thus repair mechanisms and proliferation follow until the  $k + 1^{th}$  dose fraction.

We suppose that a cell may reproduce only if it is in state 0 and still in this state after the application of a dose fraction, which occurs with a probability  $\mathbf{P}(0, 0) = (1 - q)^m$ . Moreover, such a cell either divides and gives birth to two daughter cells in state 0 with probability  $\mu$  or it remains unchanged with probability  $1 - \mu$ .

In the case of cell division, it is convenient for further developments to consider the mother cell still alive and a new one (in state 0) is artificially added.

### 3 Lifespan of a cancer cell

The lifespan of a tumor can be expressed in terms of lifespan of independent cells (see Section 4). This leads us to consider the lifespan of a single cell. We do not take cell proliferation into account.

The cell lifespan is the first (random) instant  $T$  of treatment time when the cell is dead :

$$T = \inf\{i \geq 1; Z_i = m\} \quad (6)$$

where  $(Z_k)$  is the Markov chain introduced in Section 2.

We would like to emphasize this new viewpoint of cell mortality modeling. Indeed, in the usual hit models and the previous multinomial model,  $k$  is fixed and only random variables at time  $k$  are considered, for instance, the number of deactivated targets  $Z_k$ . The knowledge of  $T$  allows us to recover information when time  $k$  for instance  $\{T \leq k\}$  is equal to  $\{Z_k = m\}$ . Moreover  $T$  reveals the dynamical aspect of the cell lifespan, so  $\{T = k\} = \{Z_1 \neq m, \dots, Z_{k-1} \neq m, Z_k = m\}$  and therefore involves the sequence  $Z_1, \dots, Z_k$  instead of  $Z_k$  only.

Note that  $T$  is of direct therapeutic importance since it corresponds to the minimal number of radiation dose fractions to be applied to kill the cancer cell. However,  $T$  is a random variable and cannot be studied as a real parameter. Nevertheless, we are able to measure the randomness of  $T$  determining its mean, variance, distribution function and confidence intervals. These real numbers and function permit in particular the understanding of how  $T$  fluctuates. In the case of numerical values of  $q$ ,  $r$  and  $m$ , the above quantities can be determined explicitly (see Section 3.2).

#### 3.1 Expected value, variance and confidence interval of $T$

**Proposition 3.1** *If the cell is initially in state  $Z_0 = i_0 \in \{0, \dots, m - 1\}$ , the random variable  $T$  of the cancer cell lifespan verifies the following properties :*

1. *The cumulative distribution function  $F$  of  $T$  is related to the transition matrix  $\Pi$  by*

$$F(k) = P(T \leq k) = \Pi^k(i_0, m). \quad (7)$$

2. *For any  $k \in \mathbb{N}^*$  we have*

$$q^m(1 - q)^{k-1} \leq P(T = k) \leq q(1 - q^m)^{k-1}. \quad (8)$$

*Specially when  $m = 1$ ,  $T$  has a geometric distribution with parameter  $q$ .*

3. For any  $n \in \mathbb{N}^*$  the expected value  $E(T^n)$  is finite and

$$E(T) = \sum_{k \geq 1} k(\Pi^k(i_0, m) - \Pi^{k-1}(i_0, m)). \quad (9)$$

4. The variance  $V(T)$  is finite and given by

$$V(T) = \sum_{k \geq 1} k^2(\Pi^k(i_0, m) - \Pi^{k-1}(i_0, m)) - (E(T))^2. \quad (10)$$

The proof is given in B.

### Confidence interval of $T$

We are interested in the confidence interval  $I$  defined as

$$P(T \in I) \geq 1 - \theta, \quad (11)$$

where  $1 - \theta$  is the confidence level. We restrict ourselves to interval  $I$  of the type  $I = ]\iota_1, \iota_2]$ .  $F$  being the cumulative distribution function of  $T$ , then the positive real numbers  $\iota_1$  and  $\iota_2$  are determined by the conditions :

$$\iota_1 = \max\{k; F(k) \leq \beta_1\} \quad \text{and} \quad \iota_2 = \min\{k; 1 - F(k) \leq \beta_2\}, \quad (12)$$

where  $\beta_1 + \beta_2 = \theta$ . For instance,  $\theta = 0.05$ ,  $\beta_1 = \beta_2 = \frac{\theta}{2} = 0.025$ . We observe thereafter that  $E(T)$  belongs to  $I$ .

### 3.2 Numerical analysis

Once the parameter values of  $q$ ,  $r$  and  $m$  are given, one can determine numerically  $\Pi$  (via (1), (2) and (3)) and thus  $\Pi^k$  for  $k \in \mathbb{N}^*$ . We have implemented the cumulative distribution function  $F$ , the expectation  $E(T)$  and the variance  $V(T)$ , using (7), (9) and (10), into the computing environment Matlab and we have fixed  $m = 3$ . From  $F$ , we obtain the bounds of the confidence interval  $I = ]\iota_1, \iota_2]$  through (12)(with  $\beta_1 = \beta_2 = \frac{\theta}{2} = 0.025$ ).

We present here a set of numerical results to outline the influence of the parameters  $q$ ,  $r$  and the initial state  $Z_0 = i_0$ , on  $F$  and  $I$ . We fix two parameters and we vary the third one.

For the fixed values  $q = 0.4$  and  $r = 0.3$ , Figures 1-(a) and (b) show  $F$  when  $Z_0 = 0$  and  $Z_0 = 2$  respectively. We observe that the size of the confidence interval  $I$  slightly decreases when  $Z_0$  is larger.

Secondly, we study the influence of the parameter  $q$  on  $I$  by fixing  $r = 0.3$  and  $Z_0 = 0$ . Figures 2-(a) and (b) show  $F$  when  $q = 0.2$  and  $q = 0.6$  respectively. We observe now that increasing  $q$  causes a drastic reduction of  $I$ .



Fixed parameters	Varied parameter	$E(T)$	$V(T)$	$I$	Figure
$q = 0.4$	$Z_0 = 0$	5.77	18.77	[1, 17]	2-(a)
$r = 0.3$	$Z_0 = 2$	3.64	14.98	[1, 15]	2-(b)
$r = 0.3$	$q = 0.2$	20.23	322.79	[2, 68]	3-(a)
$Z_0 = 0$	$q = 0.6$	2.90	3.51	[1, 8]	3-(b)
$q = 0.6$	$r = 0.2$	2.75	2.77	[1, 7]	4-(a)
$Z_0 = 0$	$r = 0.6$	3.48	6.96	[1, 11]	4-(b)

Table 2: Influence of the multinomial parameters  $q$ ,  $r$  and  $i_0$  on the confidence interval  $I$  of  $T$  with  $m = 3$  and  $\beta_1 = \beta_2 = 0.025$ .

Finally, we fix  $q = 0.6$  and  $Z_0 = 0$  to show the influence of the parameter  $r$ . Figures 3-(a) and (b) show  $F$  when  $r = 0.2$  and  $r = 0.6$  respectively. In this case, the width of  $I$  increases with  $r$ .

Table 2 presents the numerical results corresponding to the above figures. The width of  $I$  is reduced when the initial state of the cell is fragile and when the treatment is more efficient. Conversely, the interval  $I$  becomes larger when the cellular repair mechanisms increases. Figures 4-(a) and (b) complete this numerical analysis by showing the influence of  $q$  and  $r$  on  $E(T)$ .

It is an important aspect of our approach that we can observe and quantify numerically the monotonicity of the confidence interval  $I$ , especially when either the efficiency of the treatment or the repair mechanisms are changed.

## 4 Lifespan of a tumor

Our aim is to measure the tumor growth through its lifespan. We consider a tumor initially composed of  $n_0$  cells that have an independent behaviour and are all in state 0. We consider cell division as explained in Section 2.3 and recall the convention that the mother cell is still alive and a new one (in state 0) is artificially added. So at time  $k$ ,  $n_k$  new cells in state 0 are "added". This key property allows us to define the lifespan  $L$  of the whole tumor as the lifespan of  $n_0$  "mother" cells combined with the lifespan of the  $n_1$  additional cells at time  $k = 1$  and so on.

As application of our theoretical results, two regimes are distinguished. For  $n_0$  "small", calculations of mean time, variance and other properties of  $L$  are numerically possible. For large  $n_0$ , the risk of numerical miscalculations becomes significant. However, we can estimate how  $L$  behaves when the initial cancer number  $n_0$  is large and we show that the expectation of  $L$  roughly grows as a logarithmic function of  $n_0$ .

## 4.1 Lifespan of the $n_0$ "mother" cells

For each single cell  $j$ , among the  $n_0$  ones, we associate the corresponding lifespan:

$$T^{(j)} = \inf\{i \geq 1; Z_i^{(j)} = m\}. \quad (13)$$

$L_{n_0}$  is the time when all the  $n_0$  cells are killed and can be expressed with respect to the variables  $\{T^{(j)}\}_{j \in \{1, \dots, n_0\}}$  as

$$L_{n_0} = \max\{T^{(1)}, \dots, T^{(n_0)}\}. \quad (14)$$

Under the assumption that all cells are initially in state 0 and behave independently with the same dynamics, the random variables  $T^{(1)}, \dots, T^{(n_0)}$  are *i.i.d.* and have the same distribution as  $T$  (cf (7)). Therefore, the cumulative distribution function of  $L_{n_0}$  is

$$P(L_{n_0} \leq k) = P(T^{(1)} \leq k, \dots, T^{(n_0)} \leq k) = (\Pi^k(0, m))^{n_0}. \quad (15)$$

The expected value of  $L_{n_0}$  is then expressed as

$$\begin{aligned} E(L_{n_0}) &= \sum_{k \geq 1} k (P(L_{n_0} \leq k) - P(L_{n_0} \leq k - 1)). \\ &= \sum_{k \geq 1} k ((\Pi^k(0, m))^{n_0} - (\Pi^{k-1}(0, m))^{n_0}). \end{aligned} \quad (16)$$

The next proposition gives the asymptotic behavior of  $E(L_{n_0})$  when  $n_0$  is large.

**Proposition 4.1** *For fixed parameters  $m, q$  and  $r$ , the expected value  $E(L_{n_0})$  verifies*

$$\frac{\log(n_0)}{-\log(1-q)} + \frac{3q^{m-1} - q^{1-m}}{-4\log(1-q)} \leq E(L_{n_0}) \leq \frac{\log(n_0)}{-\log(1-q^m)} - \frac{2q^{1-m}}{(1-q^m)\log(1-q^m)}, \quad (17)$$

for all  $n_0 \geq \frac{2}{q^{m-1}(1-q^m)}$ .

The proof is given in C. This proposition will be generalized thereafter to the expected value of the tumor lifespan (see Section 4.3).

**Remark 4.2** *1) Note that  $q \mapsto \frac{1}{-\log(1-q)}$  and  $q \mapsto \frac{1}{-\log(1-q^m)}$  are increasing. Therefore (17) intuitively implies that  $E(L_{n_0})$  is also decreasing, which is a reasonable result since the trend of the tumor lifespan is to decrease when the treatment becomes more efficient.*

*2) The terms  $\frac{-1}{\log(1-q)}$  and  $\frac{-1}{\log(1-q^m)}$  which appear in (17) are probably not optimal.*

## 4.2 Proliferation and tumor growth modeling

According to the cell division rules introduced in Section 2.3, after the application of the first dose fraction on the  $n_0$  initial cells, a number  $S_1$  of new cells in

state 0 is produced and  $S_1 \sim \mathcal{B}(n_0, \mu(1-q)^m)$ . Our assumption of this branching process consists of replacing the random number  $S_1$  by the integer part  $n_1$  of its mean,

$$n_1 = \lfloor E(S_1) \rfloor = \lfloor n_0 \mu(1-q)^m \rfloor. \quad (18)$$

Thus, we consider that the treatment effects, repair mechanisms and cell proliferation in the initial  $n_0$  cells are equivalent to the treatment and repair effects on the  $n_0$  "mother" cells and addition of  $n_1$  new cells in state 0.

Next, we consider cell division associated with the second dose fraction. Obviously, duplication concerns the  $n_1$  new cells and the  $X_1$  cells that stay in state 0, among the  $n_0$  initial ones, after the first dose fraction. It is clear that  $X_1 \sim \mathcal{B}(n_0, \Pi(0,0))$ . Reasoning as above,  $S_2$  new cells appear, and conditionally on  $X_1 = x_1$ ,  $S_2 \sim \mathcal{B}(x_1 + n_1, \mu(1-q)^m)$ . Consequently

$$E(S_2) = \mu(1-q)^m (n_1 + E(X_1)) = \mu(1-q)^m (n_1 + n_0 \Pi(0,0)). \quad (19)$$

Then we replace random new cells  $S_2$  by  $n_2 = \lfloor E(S_2) \rfloor$ .

This analysis can be repeated at each dose fraction  $j$  adding a number  $n_j$  of new cells in state 0 resulting from cell reproduction. The sequence  $n_j$  is then given recursively by

$$n_j = \left\lfloor \mu(1-q)^m \sum_{l=0}^{j-1} \Pi^{j-1-l}(0,0) n_l \right\rfloor \quad j \geq 1. \quad (20)$$

### 4.3 Lifespan of the whole tumor

We define the lifespan  $L$  of the whole tumor as

$$L = \max\{L_{n_0}, 1 + L_{n_1}, 2 + L_{n_2}, \dots\}, \quad (21)$$

where  $L_{n_j}$   $j \geq 1$  is the lifespan of the  $n_j$  additional cells resulting from cell division at time  $j$ ,

$$P(L_{n_j} \leq k) = (\Pi^k(0, m))^{n_j}. \quad (22)$$

**Proposition 4.3** *The cumulative distribution function  $G$  of  $L$  is given by*

$$G(k) = P(L \leq k) = \prod_{j=0}^{k-1} (\Pi^{k-j}(0, m))^{n_j} \quad k \geq 1. \quad (23)$$

Therefore, the expected value is

$$E(L) = \sum_{k \geq 1} k(G(k) - G(k-1)). \quad (24)$$

**Proof 4.1** *Since all additional cell groups are independent and  $L$  is given by (21), then according to (15) we obtain*

$$P(L \leq k) = \prod_{j=0}^{k-1} P(L_{n_j} \leq k-j) = \prod_{j=0}^{k-1} (\Pi^{k-j}(0, m))^{n_j} \quad k \geq 1.$$

Let us observe that (23) allows numerical determination of  $G(k)$  for any given value of  $k$ . However, (24) needs the knowledge of a large number of  $G(k)$ . In particular, this implies to control the behavior of the sequence  $(n_j)_{j \geq 1}$ . Note that  $(n_j)$  cannot stop and the lifespan  $L$  defined by (21) would be infinite. For this reason, it is important to determine conditions under which  $n_j = 0$ , for large  $j$ .

**Proposition 4.4** *In the case*

$$\mu(1-q)^m \sum_{k \geq 0} \Pi^k(0,0) < 1, \quad (25)$$

there exists  $j_0 \in \mathbb{N}^*$  such as

$$n_j = 0 \quad \forall j > j_0. \quad (26)$$

Moreover, if there exists  $\lambda_0 > 1$  such as

$$\lambda_0 \mu(1-q)^m \sum_{k \geq 0} \lambda_0^k \Pi^k(0,0) < 1, \quad (27)$$

then (26) holds with  $j_0 = \left\lceil \frac{\log(n_0)}{\log(\lambda_0)} + C \right\rceil$ , where  $C$  is a constant.

The proof is given in D.

Note that (25) holds either if  $\mu$  is small or if  $q$  is close to 1. The first condition means that the proliferation level is low while the second one reveals the treatment efficiency. It is intuitively clear that one of the two conditions contributes to stopping new offsprings of cancer cells. Proposition 4.4 shows that the expected number  $j_0$  of dose fractions that stops the birth of additional cells is at most a logarithmic function of  $n_0$ . This crucial result allows us to generalize Proposition 4.1 as follows.

**Proposition 4.5** *Suppose that (27) holds. Then, for fixed parameters  $m, q$  and  $r$ , there exists  $a > 0, b \in \mathbb{R}$  such as for large  $n_0$*

$$\frac{\log(n_0)}{-\log(1-q)} + \frac{3q^{m-1} - q^{1-m}}{-4\log(1-q)} \leq E(L) \leq a \log(n_0) + b. \quad (28)$$

The proof is given in E.

We deduce from Proposition 4.5 that the expected lifespan of the whole tumor is bound lower and upper by two logarithmic functions of the initial cancer cell number  $n_0$ .

#### 4.4 Numerical analysis

Numerical miscalculations occur during Matlab computations when  $n_0$  is very large. This numerical problem currently limits the application of our model to *in vitro* culture and *in vivo* applications in which the region of interest can be

decomposed into sub-volumes, e.g. voxels, whose size is compatible with the computation limits of the proposed model.

Let  $J$  the confidence interval of  $L$  of the type  $J = ]\kappa_1, \kappa_2]$  such as

$$\kappa_1 = \max\{k; G(k) \leq 0.025\} \quad \text{and} \quad \kappa_2 = \min\{k; 1 - G(k) \leq 0.025\}. \quad (29)$$

We have implemented the cumulative distribution function  $G$  and the expectation  $E(L)$  (cf (23) and (24)) in Matlab and fixed the parameters  $m = 3$ ,  $q = 0.5$ ,  $r = 0.1$ . Figures 5-(a) and (b) show the mean  $E(L)$  and the confidence interval  $J$  as functions of the initial cancer cell number  $n_0$  and the cell division probability  $\mu$  respectively. Dotted lines correspond to the 95% confidence interval  $J$  and solid lines represent  $E(L)$ . We observe that  $E(L)$  increases with  $n_0$  while the width of the confidence interval remains almost constant. Moreover,  $E(L)$  increases with the proliferation rate of cells.

Proposition 4.4 shows that the expected number  $j_0$  of dose fractions that stops the birth of additional cells is at most a logarithmic function of  $n_0$ . For this, we have implemented the sequence  $(n_j)$  of additional cells (cf (20)) and we retrieve  $j_0$  when it exists, *i.e.* in the extinction case corresponding to (25). Figures 6-(a) and (b) show an example for the variation of  $j_0$  in terms of  $\log(n_0)$  ( $n_0 \leq 10^5$ ) and the deactivation parameter  $q$  respectively, with  $m = 3$ ,  $r = 0.5$  and  $\mu = 1$ . We observe in Figure 6-(a) that  $j_0$  is close to being an increasing linear function of  $\log(n_0)$ . As expected, Figure 6-(b) shows that  $j_0$  decreases when the treatment becomes more efficient.

### Case of large $n_0$

For *in vivo* studies with a very large number  $n_0$ , we deduce from (28) that the expected value  $E(L)$  is bound by a logarithmic function of  $n_0$ . This property of Proposition 4.5 is also satisfied when  $n_0 \leq 10^5$ , as corroborated by numerical results presented in Figure 7. We observe that, for  $m = 3$ ,  $q = 0.6$ ,  $\mu = 1$  and  $r = 0.4$ ,  $E(L)$  is approximately a linear function of  $\log(n_0)$ .

## 5 Optimization of treatment schedule

The main task in radiotherapy is to choose the treatment schedule able to eliminate the tumor without causing significant damage on the adjacent normal tissue. The proposed protocol is split up into three steps:

1. firstly, provided that relevant clinical data are available, the model parameters are estimated for each patient and each tumor type;
2. secondly, Tumor Control and Normal Tissue Complication Probabilities (TCP & NTCP) are computed from the model parameter estimates and for a daily fractionated radiation schedule (see Section 5.1 below);

3. thirdly, an Efficiency-Complication Trade-off diagram proposes to the clinician three possible protocols: (1) minimizing complications with an acceptable level of therapeutic efficiency, (2) privileging efficiency *vs* complication or (3) an optimal compromise (see Section 5.2 below).

The ECT diagram (Efficiency Complication Trade-off) is inspired from the ROC (Receiver Operating Characteristic) / DET (Detection Error Trade-off) curves used in biometrics and is derived from the computation of TCP & NTCP values.

### 5.1 Tumor Control and Normal Tissue Complication Probabilities

The Tumor Control Probability (TCP) at time  $k$  is the probability to destroy the tumor at this time *i.e.* the lifespan of the whole tumor is lower than  $k$ ,

$$TCP_k = P(L \leq k) = \prod_{j=0}^{k-1} (\Pi^{k-j}(0, m))^{n_j} \quad k \geq 1. \quad (30)$$

Normal Tissue Complication Probability, noted NTCP, is defined as the probability that a complication appears in the adjacent normal tissue. We consider a normal tissue initially containing  $\bar{n}_0$  cells and a complication occurs when the number of damaged normal cells is larger than a given threshold number  $\bar{n}$ . We assume that normal cells behave similarly to cancer cells, the difference comes from the different values of the model parameters  $\bar{m}$ ,  $\bar{q}$  and  $\bar{r}$  (cf Table 1).

We define  $L_{\bar{n}, \bar{n}_0}$  as the lifespan of the entire normal tissue. Note that, normal cells divide in an orderly way and only when it is needed. For this, we do not consider normal cell proliferation which is replaced by the  $(\bar{n} - 1)$  reserve capacity in the normal tissue.

**Proposition 5.1** *The NTCP at time  $k$  is given by*

$$NTCP_k = P(L_{\bar{n}, \bar{n}_0} \leq k) = \sum_{i=\bar{n}}^{\bar{n}_0} \binom{\bar{n}_0}{i} (\bar{F}(k))^i (1 - \bar{F}(k))^{\bar{n}_0 - i}, \quad (31)$$

where  $\bar{F}$  is the cumulative distribution function of the lifespan of a normal cell similar to  $F$  (cf (7)). We deduce the expected value,

$$E(L_{\bar{n}, \bar{n}_0}) = \sum_{k \geq 1} (P(L_{\bar{n}, \bar{n}_0} \leq k) - P(L_{\bar{n}, \bar{n}_0} \leq k - 1)). \quad (32)$$

This expression of  $NTCP_k$  was given in [12]. Numerical analysis for  $TCP_k$  and  $NTCP_k$  are presented in the next section, and a diagram that summarizes the possible choices of the appropriate treatment schedule is also proposed.

## 5.2 Determination of a suited number of dose fractions

The ECT diagram allows us to compute both NCTP ( $x$ -axis) and the treatment non-efficiency probability (1-TCP) ( $y$ -axis) for different values of  $k$  (number of treatment days). In this diagram, the ideal therapeutic goal is the origin point  $O(0, 0)$  since it corresponds to a 100% efficient treatment and a null risk of complication. Note that  $k \rightarrow TCP_k$  and  $k \rightarrow NTCP_k$  are increasing as they are the cumulative distribution functions of  $L$  and  $L_{\bar{n}, \bar{n}_0}$  respectively (cf (23) and (31)).

Figure 8 shows the ECT curve for  $m = \bar{m} = 3$ ,  $q = 0.7$ ,  $\bar{q} = 0.6$ ,  $r = 0.3$ ,  $\bar{r} = 0.5$ ,  $n_0 = 10000$ , and  $\bar{n} = \bar{n}_0 = 1000$  ( $\bar{n}$  and  $\bar{n}_0$  are taken relatively small to avoid numerical complication). For simplicity, we do not consider cell proliferation ( $\mu = 0$ ) since we get the same analysis.

We propose three strategies to select a suitable treatment duration:

- a first choice consists in limiting the probability of complication on normal tissues with a threshold fixed in this example at 0.05 (dotted vertical line). Beyond this level, the risk of complication is not acceptable. Subsequently, the most appropriate number of dose fractions to be applied is given by  $k = 13$  dose fractions with a treatment non-efficiency probability estimated at  $1 - TCP \approx 0.3$ .
- A second choice is to give priority to the therapeutic efficiency by using a non-efficiency threshold fixed in this example to 0.05 (dotted horizontal line). This choice leads to choose at least  $k = 16$  dose fractions with a risk of complication estimated at  $NTCP \approx 0.3$ .
- A third possibility is a trade-off choice which consists of selecting the "nearest" point of the curve to the ideal point  $O(0, 0)$ . In this case, the choice leads to select  $k = 14$  dose fractions with a non-efficiency probability lower than 0.2 and a complication probability  $NTCP \approx 0.05$ .

## 6 Conclusion

We have proposed an original approach that expresses the probability distribution of the cancer and normal cell lifespans with respect to the number of dose fractions in radiotherapy. Conversely to previous models that examine the number of surviving cells in the treated population at fixed time instants, our modeling approach better reveals the dynamics of the tumor response. It relies on a discrete-time Markov chain. This model accounts for (i) the heterogeneity of cell damage, (ii) the repair mechanisms that occur between dose fractions and (iii) the mean cell proliferation. The main contribution is the introduction of new response variables: the tumor and normal tissue lifespan. We have shown that the average lifespan of a treated tumor may be approached by a logarithmic function of the initial cancer cell number. Moreover, we have

provided original formulations of the tumor control probability and normal tissue complication probability, and we have proposed a new curve called ECT (Efficiency-Complication Trade-off), which allows us to control the benefit-risk ratio of the treatment. This synthetic representation summarizes both efficiency and complication of the treatment and point out several possibilities of choice for the radiotherapist : treatment efficiency, priority to safety of normal tissue, or a trade-off between them. The current challenge is to assess the relevance of the proposed model in preclinical studies focused on brain tumors. Ongoing works is aimed at testing the ability of the model to predict *in vitro* responses based on U87 cells.

## References

- [1] T. Bastogne, A. Samson, R. Keinj, P. Vallois, S. Wantz-Mézières, S. Pinel, D. Bechet, and M. Barberi-Heyob. Phenomenological modeling of tumor diameter growth based on a mixed effects model. *Journal of Theoretical Biology*, 262:544–552, 2010.
- [2] J D Chapman. Target theory revisited: Why physicists are essential for radiobiology research. *Clinical Oncology*, 19:S12, 2007.
- [3] S B Curtis. Lethal and potentially lethal lesions induced by radiation — A unified repair model. *Radiat. Res*, 106:252–279, 1986.
- [4] A Dawson and T Hillen. Derivation of the tumour control probability (TCP) from a cell cycle model. *Computational and Mathematical Methods in Medicine*, 7:121–141, 2006.
- [5] Brener DJ and EJ Hall. The origin and basis of the linear quadratic model. *Int J Radiat Oncol Biol Phys*, 23:252, 1992.
- [6] R. Durrett, J. Foo, K. Leder, J. Mayberry, and F. Michor. Intratumor heterogeneity in evolutionary models of tumor progression. *Genetics*, 188:461–477, 2011.
- [7] J Fowler. The linear-quadratic formula and progress in fractionated radiotherapy. *Br. J Radiol*, 62:679–694, 1989.
- [8] H A Gay and A Niemierko. A free program for calculating EUD-based NTCP and TCP in external beam radiotherapy. *Physica Medica*, 23:115–125, 2007.
- [9] P. B. Gupta, C. M. Fillmore, G. Jiang, S. D. Shapira, K. Tao, C. Kuperwasser, and E. S. Lander. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell*, 146:633–644, August 19 2011.
- [10] J H Hoeijmakers. Genome maintenance mechanisms for preventing cancer. *Nature*, 411:366–374, 2001.



- [11] P Kallman, A Agren, and A Brahme. Tumour and normal tissue responses to fractionated non-uniform dose delivery. *Int. J. Radiat. Biol.*, 62:249–262, 1992.
- [12] R Keinj, T Bastogne, and P Vallois. Multinomial model-based formulations of TCP and NTCP for radiotherapy treatment planning. *Journal of Theoretical Biology*, 279:55–62, 2011.
- [13] N F Kirkby, Burnet N G, and D B F Faraday. Mathematical modelling of the response of tumour cells to radiotherapy. *Nuclear Instruments and Methods in Physics Research. Section B: Beam Interactions with Materials and Atoms*, 188:210–215, 2002.
- [14] J T Lyman. Complication probability as assessed from dose volume histograms. *Radiat. Res.*, 104:513–519, 1985.
- [15] S. Michelson and J. T. Leith. *A survey of Models for Tumor-Immune System Dynamics*, chapter Tumor Heterogeneity and Growth Control, pages 295–333. Birkhäuser, 1997.
- [16] E C Pollard, W R Guild, F Hutchinson, and R B Setlow. The direct action of ionizing radiation on enzymes and antigen. *Progress in Biophysics*, 5:72–108, 1955.
- [17] R K Sachs, L R Hlatky, and P Hahnfeldt. Simple ODE models of tumor growth and anti-angiogenic or radiation treatment. *Mathematical and Computer Modelling*, 33:1297–1305, 2001.
- [18] C Wyman and R Kanaar. DNA double-strand break repair: all’s well that ends well. *Annu. Rev. Genet.*, 40:363–383, 2006.
- [19] M Zaider and G N Minerbo. Tumour control probability: a formulation applicable to any temporal protocol of dose delivery. *Phys. Med. Biol.*, 45:279–293, 2000.

## A Algorithmic expression of $Z_k$

The following recursive scheme, taking into account first the treatment effect and second the repair mechanisms, gives a realization of  $(Z_k)$ :

$$\left\{ \begin{array}{l} Z_0 = i_0 \\ Z_{k+\frac{1}{2}} = Z_k + \sum_{j=1}^{m-Z_k} \xi_{k+\frac{1}{2},j} \\ Z_{k+1} = \begin{cases} \sum_{j=1}^{Z_{k+\frac{1}{2}}} \xi_{k+1,j} & \text{if } Z_{k+\frac{1}{2}} < m \\ m & \text{if } Z_{k+\frac{1}{2}} = m, \end{cases} \end{array} \right. \quad (33)$$

where  $(\xi_{k+\frac{1}{2},j})$  and  $(\xi_{k+1,j})$  are independent and Bernoulli distributed,  $P(\xi_{k+\frac{1}{2},j} = 1) = q$  and  $P(\xi_{k+1,j} = 1) = 1 - r$ . Note that  $Z_{k+\frac{1}{2}}$  is the state of the cell after

application of the  $(k + 1)$ -th dose and just before the repair phase.

Moreover, we give an equivalent expression:

$$\begin{cases} Z_0 &= i_0 \\ Z_{k+\frac{1}{2}} &= Z_k + \sum_{j=1}^{m-Z_k} 1_{\{U_{k+\frac{1}{2},j} \leq q\}} \\ Z_{k+1} &= \begin{cases} \sum_{j=1}^{Z_{k+\frac{1}{2}}} 1_{\{U_{k+1,j} \leq 1-r\}} & \text{if } Z_{k+\frac{1}{2}} < m \\ m & \text{if } Z_{k+\frac{1}{2}} = m, \end{cases} \end{cases} \quad (34)$$

where  $(U_{k+\frac{1}{2},j})$  and  $(U_{k+1,j})$  are independent and uniformly distributed over  $[0, 1]$ .

## B Proof of Proposition 3.1

1. Using the definition of  $T$  (cf (6)) and the crucial fact that  $m$  is an absorbant state we have

$$P(T \leq k) = P(Z_k = m) \quad \forall k \in \mathbb{N}^*. \quad (35)$$

According to (4) and the assumption that the cell is initially in state  $i_0 < m$ , we obtain

$$F(k) = P(T \leq k) = \Pi^k(i_0, m).$$

2. Let us introduce

$$\lambda_1 = \max_{i \leq m-1} \left\{ \sum_{j \leq m-1} \Pi(i, j) \right\} = \max_{i \leq m-1} \{1 - \Pi(i, m)\} \quad (36)$$

$$\lambda_2 = 1 - \max_{i \leq m-1} \{\Pi(i, m)\}. \quad (37)$$

Consequently,

$$\lambda_1 = 1 - \min_{i \leq m-1} \{\Pi(i, m)\} \quad (38)$$

and for all  $i \leq m - 1$  we have

$$1 - \lambda_1 \leq \Pi(i, m) \leq 1 - \lambda_2. \quad (39)$$

Note that  $\lambda_1 < 1$  and  $\lambda_2 < 1$ . Firstly, we prove that

$$P(T = k) \leq (1 - \lambda_2) \lambda_1^{k-1} \quad \forall k \leq 1, \quad (40)$$

only when  $k = 3$  (the proof for any  $k \in \mathbb{N}^*$  is analogue):

$$\begin{aligned}
P(T = 3) &= P(Z_1 \neq m, Z_2 \neq m, Z_3 = m) \\
&= \sum_{i_1, i_2 \leq m-1} P(Z_1 = i_1 \mid Z_0 = i_0) P(Z_2 = i_2 \mid Z_1 = i_1) P(Z_3 = m \mid Z_2 = i_2) \\
&= \sum_{i_1, i_2 \leq m-1} \Pi(i_0, i_1) \Pi(i_1, i_2) \Pi(i_2, m) \\
&\leq (1 - \lambda_2) \left( \sum_{i_1 \leq m-1} \left\{ \sum_{i_2 \leq m-1} \Pi(i_1, i_2) \right\} \Pi(i_0, i_1) \right) \\
&\leq (1 - \lambda_2) \lambda_1 \left( \sum_{i_1 \leq m-1} \Pi(i_0, i_1) \right) \leq (1 - \lambda_2) \lambda_1^2.
\end{aligned}$$

Similarly, using (39) and

$$\sum_{j \leq m-1} \Pi(i_0, j) = 1 - \Pi(i_0, m) \geq \lambda_2, \quad (41)$$

we deduce

$$(1 - \lambda_1)(\lambda_2)^{k-1} \leq P(T = k). \quad (42)$$

Note that,  $\lambda_1$  and  $\lambda_2$  are related to the parameters  $m$  and  $q$ . According to (2), (3) and (1) we have

$$\Pi(i, m) = \sum_{j=0}^m \mathbf{P}(i, j) \mathbf{R}(j, m) = \mathbf{P}(i, m) = q^{m-i} \quad (i < m).$$

Thus, we obtain

$$\lambda_1 = 1 - q^m \quad \text{and} \quad \lambda_2 = 1 - q. \quad (43)$$

3. According to (8) we deduce that the mean value  $E(T^n)$  is finite for any  $n \geq 1$ :

$$E(T^n) = \sum_{k \geq 1} k^n P(T = k) \leq q \sum_{k \geq 1} k^n (1 - q^m)^{k-1} < \infty.$$

Now, we express  $E(T)$  in terms of the cumulative distribution function  $F$ . Thus,

$$\begin{aligned}
E(T) &= \sum_{k \geq 1} k (P(T < k+1) - P(T < k)) \\
&= \sum_{k \geq 1} k (F(k) - F(k-1)). \\
&= \sum_{k \geq 1} k (\Pi^k(i_0, m) - \Pi^{k-1}(i_0, m)).
\end{aligned}$$

4. Recall that  $E(T^2) < \infty$ , then the variance of  $T$  is finite and

$$\begin{aligned} V(T) &= E(T^2) - (E(T))^2 \\ &= \sum_{k \geq 1} k^2 (P(T < k+1) - P(T < k)) - (E(T))^2 \\ &= \sum_{k \geq 1} k^2 (\Pi^k(i_0, m) - \Pi^{k-1}(i_0, m)) - (E(T))^2. \end{aligned}$$

## C Proof of Proposition 4.1

1) First, we show that

$$-\frac{\log(n_0)}{\log(1-q)} + \frac{3q^{m-1} - q^{1-m}}{-4\log(1-q)} \leq E(L_{n_0}). \quad (44)$$

As we have  $1 - F(k) = \sum_{i \geq k+1} P(T = i)$ , then (8) implies

$$q^m \sum_{i \geq k+1} (1-q)^{i-1} \leq 1 - F(k) \leq q \sum_{i \geq k+1} (1-q^m)^{i-1}.$$

This leads to

$$q^{m-1}(1-q)^k \leq 1 - F(k) \leq \frac{1}{q^{m-1}}(1-q^m)^k$$

and

$$1 - \frac{1}{q^{m-1}}(1-q^m)^k \leq F(k) \leq 1 - q^{m-1}(1-q)^k. \quad (45)$$

On the other hand, set  $U_{n_0}$  as

$$U_{n_0} := L_{n_0} + \frac{\log(n_0)}{\log(1-q)}. \quad (46)$$

From (15) we get

$$\begin{aligned} P(U_{n_0} \leq x) &= P\left(L_{n_0} \leq x - \frac{\log(n_0)}{\log(1-q)}\right) = P\left(L_{n_0} \leq \left\lfloor x - \frac{\log(n_0)}{\log(1-q)} \right\rfloor\right) \\ &= \left(F\left(\left\lfloor x - \frac{\log(n_0)}{\log(1-q)} \right\rfloor\right)\right)^{n_0} \end{aligned} \quad (47)$$

for any  $x \geq \frac{\log(n_0)}{\log(1-q)}$ , and  $[a]$  denotes the integer part of  $a$ . Using (45) we obtain

$$F\left(\left\lfloor x - \frac{\log(n_0)}{\log(1-q)} \right\rfloor\right) \leq 1 - q^{m-1}(1-q)^{\left\lfloor x - \frac{\log(n_0)}{\log(1-q)} \right\rfloor}.$$

Since we have

$$\left\lfloor x - \frac{\log(n_0)}{\log(1-q)} \right\rfloor \leq x - \frac{\log(n_0)}{\log(1-q)} \quad \text{and} \quad 1-q < 1,$$

then

$$1 - q^{m-1}(1-q)^{\left\lfloor x - \frac{\log(n_0)}{\log(1-q)} \right\rfloor} \leq 1 - q^{m-1}(1-q)^{x - \frac{\log(n_0)}{\log(1-q)}} = 1 - \frac{q^{m-1}(1-q)^x}{n_0}.$$

Combining previous inequalities, we deduce :

$$P(U_{n_0} \leq x) \leq \left(1 - \frac{q^{m-1}(1-q)^x}{n_0}\right)^{n_0} = e^{n_0 \log\left(1 - \frac{q^{m-1}(1-q)^x}{n_0}\right)}, \quad x \geq \frac{\log(n_0)}{\log(1-q)}.$$

Using Taylor's formula, we get

$$-u - \frac{u^2}{2(1-u)^2} \leq \log(1-u) < -u \quad 0 \leq u < 1. \quad (48)$$

Consequently,

$$P(U_{n_0} \leq x) \leq e^{-q^{m-1}(1-q)^x}, \quad x \geq \frac{\log(n_0)}{\log(1-q)} \quad (49)$$

and

$$P(U_{n_0} > x) \geq 1 - e^{-q^{m-1}(1-q)^x}.$$

Starting with

$$e^{-x} \leq 1 - x + \frac{x^2}{2} \quad x \geq 0,$$

we deduce that

$$P(U_{n_0} > x) \geq q^{m-1}(1-q)^x - \frac{1}{2}q^{2(m-1)}(1-q)^{2x}, \quad x \geq 0. \quad (50)$$

If  $X$  is an integrable real random variable on  $\mathbb{R}$ , it is easy to prove :

$$E(X) = \int_0^\infty P(X > y)dy - \int_{-\infty}^0 P(X \leq y)dy. \quad (51)$$

In particular

$$E(X) \leq \int_0^\infty P(X > y)dy, \quad (52)$$

and if  $X > -c$  where  $c > 0$ , then

$$\int_{-\infty}^0 P(X \leq y)dy = \int_{-c}^0 P(X \leq y)dy. \quad (53)$$

Applying (51) and (53) for  $X = U_{n_0}$  and  $c = -\frac{\log(n_0)}{\log(1-q)}$ , we get

$$E(U_{n_0}) = \int_0^\infty P(U_{n_0} > x)dx - \int_{-c}^0 P(U_{n_0} \leq x)dx. \quad (54)$$

Then using (50) we obtain

$$\int_0^\infty P(U_{n_0} > x) dx \geq -\frac{q^{m-1}}{4 \log(1-q)} (4 - q^{m-1}) \geq \frac{-3q^{m-1}}{4 \log(1-q)}, \quad (q^{m-1} \leq 1). \quad (55)$$

On the other hand, using (49) and the fact  $e^x \geq 1 + x, x \geq 0$ , we obtain

$$\int_{-c}^0 P(U_{n_0} \leq x) dx \leq \int_0^\infty e^{-q^{m-1}(1-q)^{-x}} dx \leq \frac{-1}{q^{m-1} \log(1-q)}. \quad (56)$$

Consequently, (54), (55) and (56) give

$$E(U_{n_0}) \geq \frac{1}{-4 \log(1-q)} (3q^{m-1} - q^{1-m}),$$

and finally

$$E(L_{n_0}) \geq -\frac{\log(n_0)}{\log(1-q)} + \frac{1}{-4 \log(1-q)} (3q^{m-1} - q^{1-m}).$$

2) Similarly, to prove that  $E(L_{n_0})$  is upper bound by a logarithmic function of  $n_0$ . In this case, we introduce

$$U'_{n_0} := L_{n_0} + \frac{\log(n_0)}{\log(1-q^m)}. \quad (57)$$

Let  $x \geq 0$ . (45) and  $\lfloor a \rfloor \geq a - 1$ , imply

$$\begin{aligned} F\left(\left\lfloor x - \frac{\log(n_0)}{\log(1-q^m)} \right\rfloor\right) &\geq 1 - \frac{1}{q^{m-1}} e^{\log(1-q^m) \left\lfloor x - \frac{\log(n_0)}{\log(1-q^m)} \right\rfloor} \\ &\geq 1 - \frac{1}{q^{m-1}} (1-q^m)^{x-1} \frac{1}{n_0}. \end{aligned} \quad (58)$$

Consequently,

$$P(U'_{n_0} \leq x) \geq e^{n_0 \log\left(1 - \frac{(1-q^m)^{x-1}}{n_0 q^{m-1}}\right)}, \quad (59)$$

and

$$P(U'_{n_0} > x) \leq 1 - e^{n_0 \log\left(1 - \frac{(1-q^m)^{x-1}}{n_0 q^{m-1}}\right)} \leq -n_0 \log\left(1 - \frac{(1-q^m)^{x-1}}{n_0 q^{m-1}}\right), \quad (60)$$

since  $1 - e^{-x} \leq x, \forall x \geq 0$ .

We have the following inequalities

$$\begin{aligned} \frac{(1-q^m)^{x-1}}{n_0 q^{m-1}} &\leq \frac{1}{n_0 q^{m-1} (1-q^m)} \leq \frac{1}{2}; \quad n_0 \geq \frac{2}{q^{m-1} (1-q^m)} \\ \frac{u^2}{(1-u)^2} &\leq 2u, \quad 0 < u < \frac{1}{2}. \end{aligned}$$

Thus using (48) and (60), we get

$$P(U'_{n_0} > x) \leq \frac{2(1-q^m)^{x-1}}{q^{m-1}}.$$

Using (52) with  $X = U'_{n_0}$  gives

$$E(U'_{n_0}) \leq \frac{-2q^{1-m}}{(1-q^m)\log(1-q^m)},$$

and finally

$$E(L_{n_0}) \leq \frac{\log(n_0)}{-\log(1-q^m)} - \frac{2q^{1-m}}{(1-q^m)\log(1-q^m)}.$$

## D Proof of Proposition 4.4

In order to study the sequence  $(n_j)$  defined by (20), it is convenient to introduce the sequence  $(\eta_j)$ :

$$\begin{aligned} \eta_0 &:= n_0 \\ \eta_{j+1} &:= \mu(1-q)^m \sum_{l=0}^j \Pi^{j-l}(0,0)\eta_l, \quad j \geq 0. \end{aligned} \quad (61)$$

a) The following proposition gives the sufficient conditions ensuring the convergence of the sequence  $(\eta_j)$ . We then deduce the convergence of the sequence  $(n_j)$  in step b).

**Proposition D.1** *If  $\mu(1-q)^m \sum_{k \geq 0} \Pi^k(0,0) < 1$ , then  $\lim_{j \rightarrow +\infty} \eta_j = 0$ . Moreover, if it exists  $\lambda_0 > 1$  such as*

$$\lambda_0 \mu(1-q)^m \sum_{k \geq 0} \lambda_0^k \Pi^k(0,0) < 1, \quad (62)$$

then

$$\eta_j < 1 \quad \forall j > j_0, \quad (63)$$

where  $j_0 = \left\lfloor \frac{\log(n_0)}{\log(\lambda_0)} + C \right\rfloor$  and  $C$  is a constant depending only on  $\mu, m, q, r$  and  $\lambda_0$ .

**Proof D.1** *Let*

$$S(\lambda) := \sum_{k \geq 0} \lambda^k \Pi^k(0,0), \quad \lambda > 0. \quad (64)$$

*Recall that the Markov chain  $(Z_k)$  defined in Section 2 is transient. Since  $m$  is an absorbent state, then  $S(1) < \infty$ .*

*For  $\lambda > 0$ , we multiply  $\eta_{j+1}$  by  $\lambda^{j+1}$  in (61) and we sum over  $j \in \{0, \dots, k\}$ ,*

$$\sum_{j=1}^{k+1} \lambda^j \eta_j = \lambda \mu(1-q)^m \left( \sum_{j=0}^k \sum_{l=0}^j \lambda^j \Pi^{j-l}(0,0)\eta_l \right). \quad (65)$$

For fixed  $l$ , we set  $i = j - l$  and we obtain

$$\sum_{j=1}^{k+1} \lambda^j \eta_j = \lambda \mu (1-q)^m \sum_{l=0}^k \left( \lambda^l \eta_l \sum_{i=0}^{k-l} \lambda^i \Pi^i(0,0) \right). \quad (66)$$

However,

$$\sum_{i=0}^{k-l} \lambda^i \Pi^i(0,0) \leq \sum_{i \geq 0} \lambda^i \Pi^i(0,0). \quad (67)$$

Since  $\lambda$ ,  $\mu(1-q)^m$  and  $\eta_l$  are positive, then

$$\sum_{j=1}^{k+1} \lambda^j \eta_j \leq \lambda \mu (1-q)^m S(\lambda) \sum_{l=0}^k \lambda^l \eta_l \leq \lambda \mu (1-q)^m S(\lambda) \sum_{l=0}^{k+1} \lambda^l \eta_l. \quad (68)$$

Consequently, if  $S(\lambda) < \infty$ , we get

$$(1 - \lambda \mu (1-q)^m S(\lambda)) \sum_{j=1}^{k+1} \lambda^j \eta_j \leq \lambda \mu (1-q)^m S(\lambda) n_0. \quad (69)$$

Assume that

$$\lambda \mu (1-q)^m S(\lambda) < 1. \quad (70)$$

Thus, taking  $k \rightarrow \infty$  in (69) leads to

$$H(\lambda) := \sum_{j \geq 0} \lambda^j \eta_j \leq \frac{\lambda \mu (1-q)^m S(\lambda)}{1 - \lambda \mu (1-q)^m S(\lambda)} n_0 < \infty. \quad (71)$$

If  $\mu(1-q)^m S(1) < 1$ , then the condition (70) is verified for  $\lambda = 1$ . Therefore,  $H(1) < \infty$ ,

$$\sum_{j \geq 0} \eta_j < \infty.$$

This implies that  $\lim_{j \rightarrow +\infty} \eta_j = 0$ .

Let us consider the case  $\lambda_0 \mu (1-q)^m S(\lambda_0) < 1$  where  $\lambda_0 > 1$ . Since (70) is verified with  $\lambda = \lambda_0$ , then

$$\sum_{j \geq 0} \lambda_0^j \eta_j \leq C_0 n_0, \quad (72)$$

where  $C_0$  is a constant depending only on  $\mu$ ,  $m$ ,  $q$ ,  $r$  and  $\lambda_0$ . Therefore,

$$\lambda_0^j \eta_j < C_0 n_0, \quad \forall j \geq 0. \quad (73)$$

Consequently, (63) is verified with  $j_0 = \left\lfloor \frac{\log(n_0)}{\log(\lambda_0)} + C \right\rfloor$  and  $C = \frac{\log(C_0)}{\log(\lambda_0)}$ . b) We prove now the Proposition 4.4. Firstly, we prove by induction on  $j$  that

$$n_j \leq \eta_j \quad \forall j \geq 0. \quad (74)$$



Indeed, if  $j = 0$ ,  $\eta_0 = n_0$  by definition. Suppose that

$$n_k \leq \eta_k \quad \forall k \leq j.$$

Then

$$\mu(1-q)^m \sum_{k=0}^j \Pi^{j-k}(0,0)n_k \leq \mu(1-q)^m \sum_{k=0}^j \Pi^{j-k}(0,0)\eta_k.$$

Since  $\lfloor a \rfloor \leq a$ , according to (20) and (61) we deduce that

$$\begin{aligned} n_{j+1} &= \left\lfloor \mu(1-q)^m \sum_{k=0}^j \Pi^{j-k}(0,0)n_k \right\rfloor \leq \mu(1-q)^m \sum_{k=0}^j \Pi^{j-k}(0,0)n_k \\ &\leq \mu(1-q)^m \sum_{k=0}^j \Pi^{j-k}(0,0)\eta_k \\ &\leq \eta_{j+1}. \end{aligned}$$

Let us consider the case  $\mu(1-q)^m \sum_{k \geq 0} \Pi^k(0,0) < 1$ . According to the Proposition D.1,  $\lim_{j \rightarrow +\infty} \eta_j = 0$ . However,  $(n_j)$  is a sequence of integer numbers, then we deduce from (74) that there exists  $j_0$  such as  $n_j = 0 \quad \forall j \geq j_0$ .

Moreover, if it exists  $\lambda_0 > 1$  such as (27) is verified, Proposition D.1 and (74) imply (26) with  $j_0 = \left\lceil \frac{\log(n_0)}{\log(\lambda_0)} + C \right\rceil$ , and  $C$  is a constant depending only on  $\mu$ ,  $m$ ,  $q$ ,  $r$  and  $\lambda_0$ .

**Remark D.2** Note that, in the case where  $\mu(1-q)^m S(1) > 1$ , we can prove that  $(\eta_j)$  is not bound. Moreover, let  $\gamma$  be the unique real number in  $]0, 1[$  such that  $\gamma \mu(1-q)^m S(\gamma) = 1$ . It can be proved that if  $n_0 > \frac{\gamma}{1-\gamma}$  then  $(n_j)$  is not bound.

## E Proof of Proposition 4.5

According to Proposition 4.4 and under (27),  $n_j = 0$  for any  $j \geq j_0$  where

$$j_0 := \left\lceil \frac{\log(n_0)}{\log(\lambda_0)} + C \right\rceil, \quad C \in \mathbb{R}.$$

Therefore, the lifespan of the whole tumor  $L$  reduces to

$$L = \max\{L_{n_0}, 1 + L_{n_1}, \dots, j_0 + L_{n_{j_0}}\}. \quad (75)$$

First, it is clear that  $L_{n_0} \leq L$ . Consequently, (17) implies

$$\frac{\log(n_0)}{-\log(1-q)} + \frac{3q^{m-1} - q^{1-m}}{-4\log(1-q)} \leq E(L). \quad (76)$$

On the other hand, for any  $0 \leq j \leq j_0$ , we have

$$j + L_{n_j} \leq j_0 + L_{n_j}.$$

Therefore,

$$L \leq j_0 + \max\{L_{n_0}, L_{n_1}, \dots, L_{n_{j_0}}\}. \quad (77)$$

Since all cells are independent, then

$$\max\{L_{n_0}, L_{n_1}, \dots, L_{n_{j_0}}\} \stackrel{(d)}{=} L_{N_{j_0}}, \quad (78)$$

where  $N_{j_0} = n_0 + \dots + n_{j_0}$ . According to (73), (74) and  $\lambda_0 > 1$ , we have

$$n_j \leq C_0 n_0 \quad \forall j \in \{0, \dots, j_0\}. \quad (79)$$

Then,

$$N_{j_0} \leq C_0 n_0 (j_0 + 1). \quad (80)$$

Thus, combining (77), (78), (80) and Proposition 4.1, we get

$$\begin{aligned} E(L) &\leq j_0 + E(L_{N_{j_0}}) \leq j_0 + \frac{\log(N_{j_0})}{-\log(1 - q^m)} + a_1 \\ &\leq a \log(n_0) + b, \end{aligned} \quad (81)$$

where  $a > 0$  and  $b$  are two constants.

## F Figures

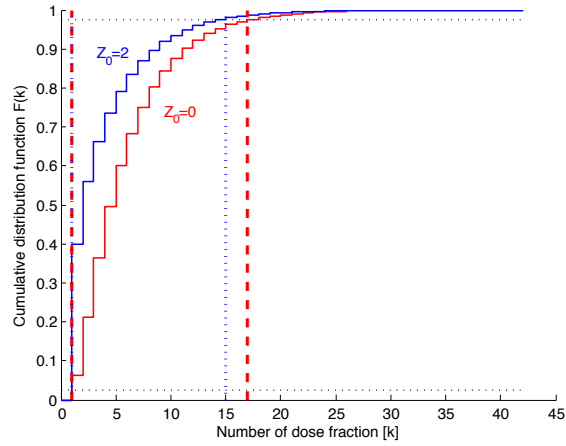


Figure 1: Influence of the initial state value  $i_0$  on the cumulative distribution function  $F$  and on the confidence interval  $I$  of  $T$ . The cumulative distribution functions  $F$  for  $Z_0 = 0$  and  $Z_0 = 2$  are described in red and blue lines respectively while dotted lines represent the bounds of  $I$  at the 5% level ( $\beta_1 = \beta_2 = 0.025$ ), with  $m = 3$ ,  $q = 0.4$  and  $r = 0.3$ .

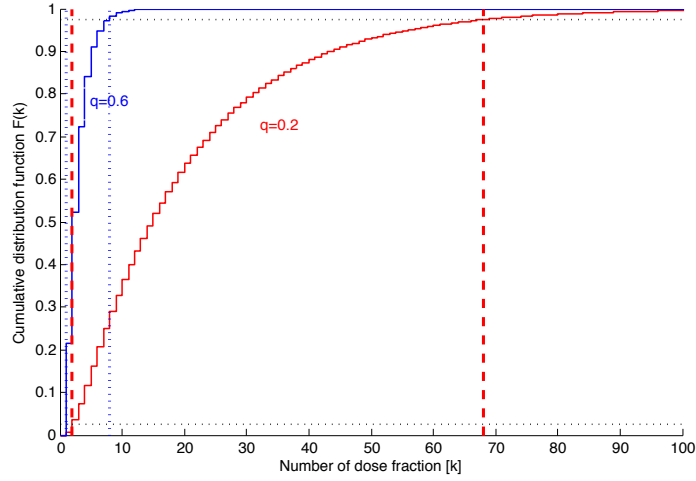


Figure 2: Influence of the deactivation parameter  $q$  on the cumulative distribution function  $F$  and on the confidence interval  $I$  of  $T$ . The cumulative distribution functions  $F$  for  $q = 0.2$  and  $q = 0.6$  are described in red and blue lines respectively while dotted lines represent the bounds of  $I$  at the 5% level ( $\beta_1 = \beta_2 = 0.025$ ), with  $m = 3$ ,  $r = 0.3$  and  $Z_0 = 0$ .

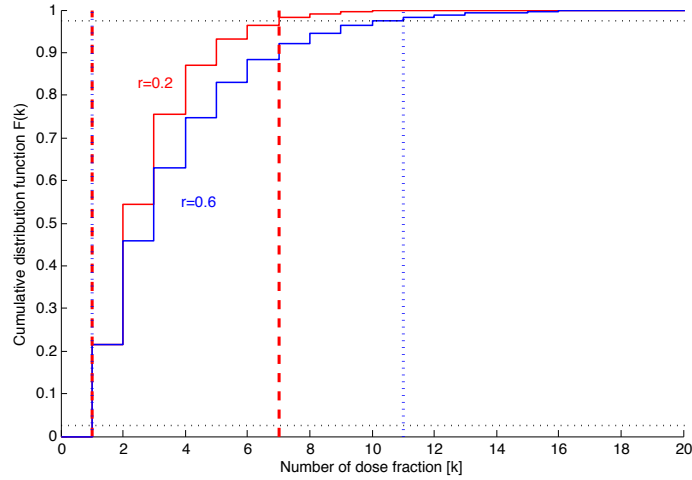
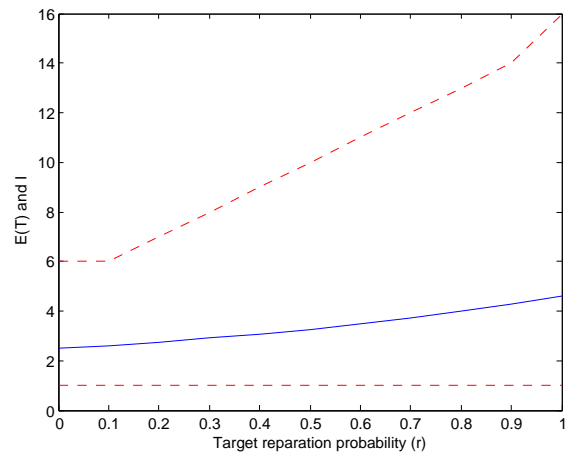
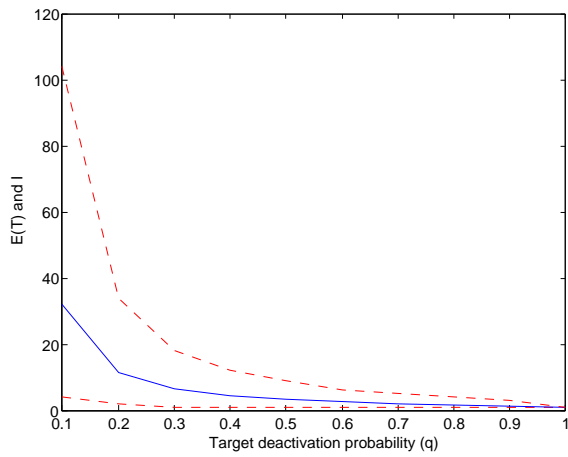


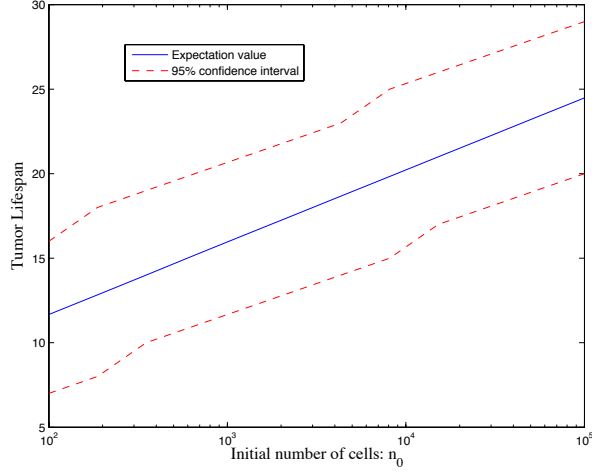
Figure 3: Influence of the repair parameter  $r$  on the cumulative distribution function  $F$  and on the confidence interval  $I$  of  $T$ . The cumulative distribution functions  $F$  for  $r = 0.2$  and  $r = 0.6$  are described in red and blue lines respectively while dotted lines represent the bounds of  $I$  at the 5% level ( $\beta_1 = \beta_2 = 0.025$ ), with  $m = 3$ ,  $q = 0.6$  and  $Z_0 = 0$ .



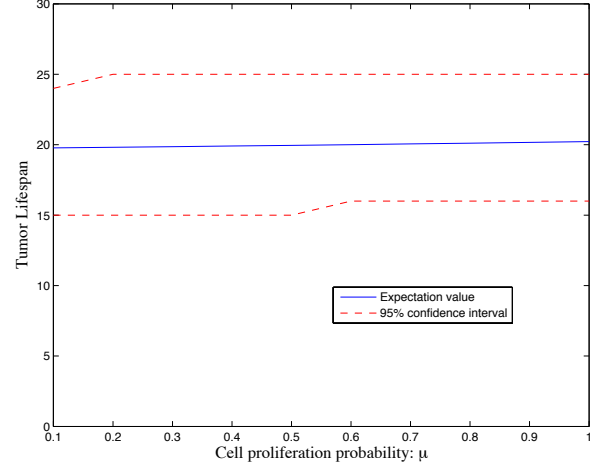
(a)  $E(T)$  and  $I$  as functions of  $q$  with  $r = 0.1$ .

(b)  $E(T)$  and  $I$  as functions of  $r$  with  $q = 0.6$ .

Figure 4: Variations of  $E(T)$  and the confidence interval  $I$  in terms of  $q$  and  $r$ , where  $m = 3$  and  $Z_0 = 0$ . Dotted lines correspond to the 95% confidence interval and solid lines represent  $E(T)$ .

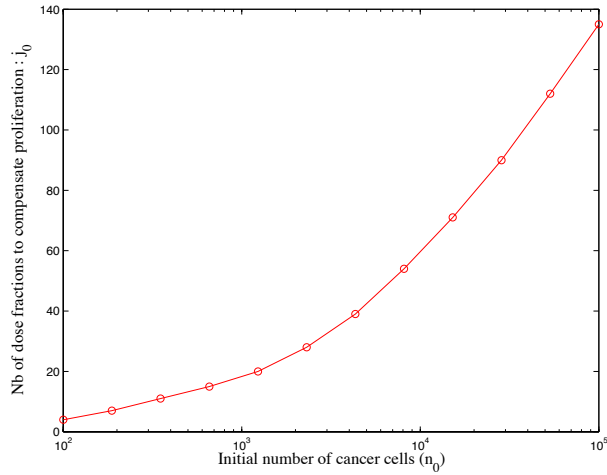


(a)  $E(L)$  and  $J$  as functions of  $n_0$  with  $\mu = 1$ .

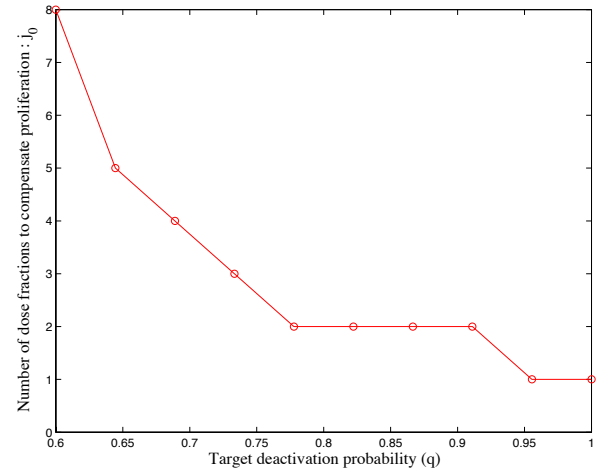


(b)  $E(L)$  and  $J$  as functions of  $\mu$  with  $n_0 = 1000$ .

Figure 5: Variations of  $E(L)$  and the confidence interval  $J$  in terms of  $n_0$  and  $\mu$ , where  $m = 3$  and  $q = 0.5$  and  $r = 0.1$ . Dotted lines correspond to the 95% confidence interval and solid lines represent  $E(L)$ .



(a)  $j_0$  as function of  $\log(n_0)$ , with  $q = 0.4$ .



(b)  $j_0$  as function of  $q$ , with  $n_0 = 10^4$ .

Figure 6: Variation of the expected number  $j_0$  that stops the birth of additional cells, as function of  $\log(n_0)$  (semi-log plot) and the deactivation parameter  $q$ , with  $m = 3$ ,  $r = 0.5$  and  $\mu = 1$ .

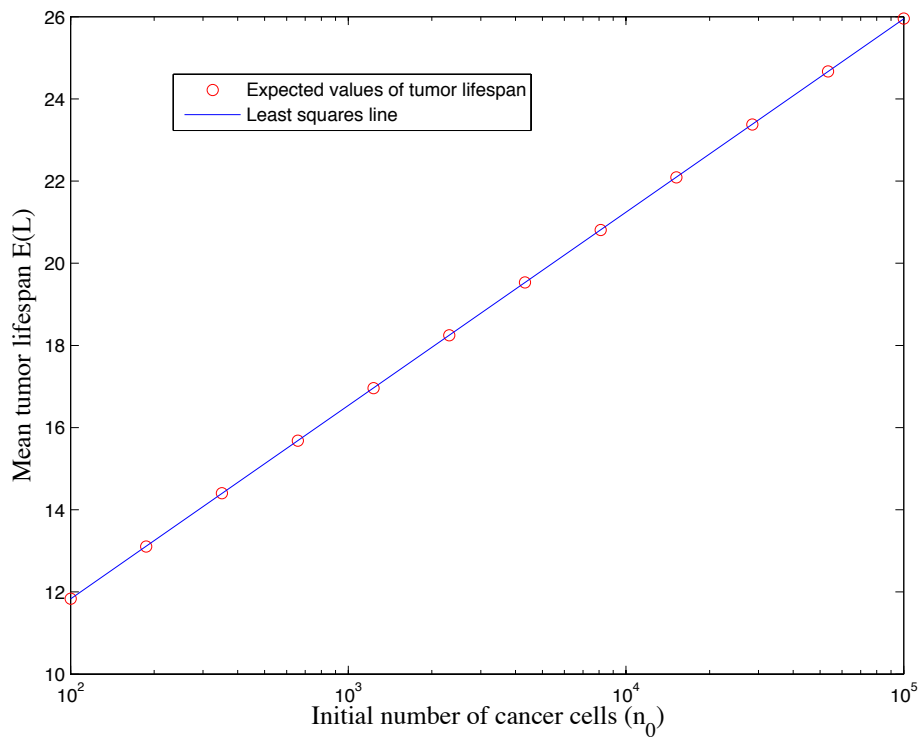


Figure 7: The expected value  $E(L)$  as function of  $\log(n_0)$  (semi-log plot), where  $m = 3$ ,  $q = 0.6$ ,  $\mu = 1$  and  $r = 0.4$ .

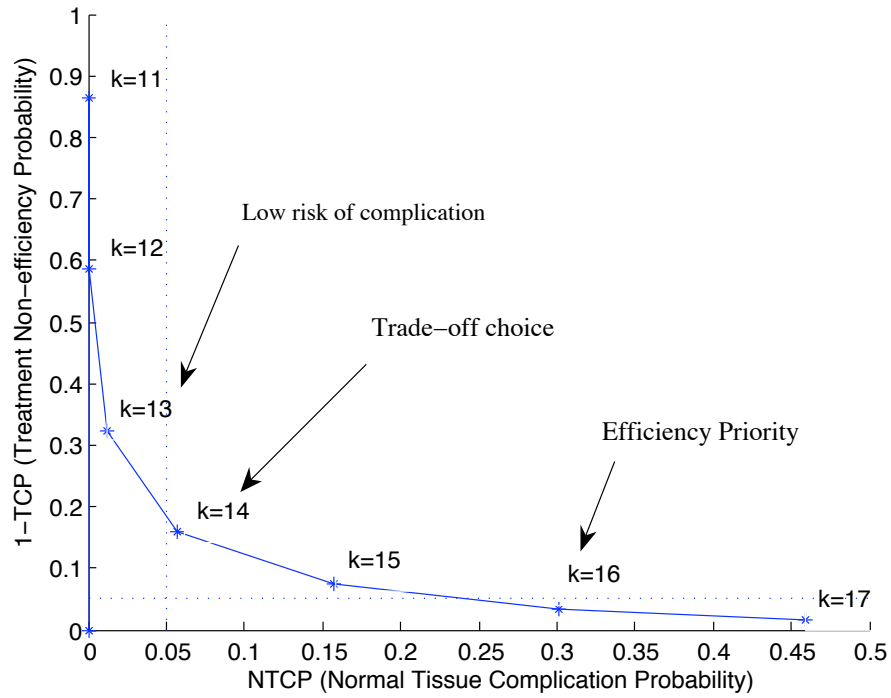


Figure 8: ECT diagram (Efficiency Complication Trade-off) for  $m = \bar{m} = 3$ ,  $q = 0.7$ ,  $\bar{q} = 0.6$ ,  $r = 0.3$ ,  $\bar{r} = 0.5$ ,  $n_0 = 10000$ , and  $\bar{n} = \bar{n}_0 = 1000$ . At  $k = 13$  (low risk of complication): TCP=0.68 and NTCP=0.011, at  $k = 14$  (trade-off choice): TCP=0.84 and NTCP=0.057 and for  $k = 16$  (efficiency priority): TCP=0.968 and NTCP=0.307.