# Tumor size distribution can yield valuable information on tumor growth and tumor control

Uwe Schneider[1,2*], Jürgen Besserer[1,2]

[1] Department of Physics, Science Faculty, University of Zürich, Zürich, Switzerland

[2] Radiotherapy Hirslanden, Witellikerstrasse 40, CH-8032 Zürich, Switzerland

* Corresponding author:
E-mail: uwe.schneider@uzh.ch

Short Title: Tumor size distribution can yield valuable information on tumor growth and tumor control

# Abstract

In this work it is shown that tumor volume distributions, can yield valuable information on two completely different topics of cancer research. From the hypothesis that the intrinsic distributions of breast cancer volumes follows an exponential distribution, first the probability density function of tumor growth time was deduced. The resulting distribution of lag times can be used in tumor induction models instead of a fixed lag time to deduce the probability of tumor induction as a function of patient age. In a second step, the distribution of cancer volumes was used to model the variation of the clonogenic cell number for tumor control probability calculations for radiotherapy cancer patients. The integration of the volume variation into a Poisson-TCP model resulted in a logistic function which fits population averaged survival data of radiotherapy patients. To our knowledge this is the first direct derivation of a logistic TCP model for cohorts of patients from a Poisson TCP-model for individuals.

# Introduction

The estimation and analysis of tumor size distributions is mainly focused on modelling the effect of tumor size on cancer survival or tumor progression [1-8]. In this work we focus on alternative applications of tumor size distributions. We will show, that tumor size distributions, or more precise tumor volume distributions, can yield valuable information on two completely different topics of cancer research. First we will show that the probability density distribution of detected tumor sizes is strictly related to the probability density function of tumor growth time. The tumor growth time, often called lag time, is the time from the appearance of the first fully malignant cell until development of a clinical observable cancer. In tumor induction models this lag time is usually estimated as a fixed number. We believe that the use of the determined lag time distribution in tumor induction models has the potential to improve the quality of such models.

Secondly, the distribution of cancer volumes can be used to calculate the variation of the clonogenic cell number for modelling tumor control probability for radiotherapy cancer patients. Currently no general analytical formulation of TCP exists which includes the population variation of TCP. Often a logistic function is used to model population averaged TCP [9], however, to our knowledge no strict derivation of a logistic TCP model exists. We will show that the probability density function of cancer sizes together with a Poisson based TCP model yields an analytic expression for TCP in terms of a logistic function.

# Materials and Methods

## Probability distributions of tumor sizes

We assume here that the intrinsic probability density function (pdf) for the distribution of tumor volumes follows an exponential distribution.

$$g_i(V) = \lambda \cdot \exp(-\lambda \cdot V) \tag{1}$$

As the mean value of an exponential distribution is $1/\lambda$ it consequently writes:

$$g_i(V) = \frac{1}{\bar{V}} \cdot \exp(-V/\bar{V}) \ , \tag{2}$$

where $\bar{V}$ is the average tumor volume.

With intrinsic we mean here that the exponential probability distribution describes the distribution of all tumor volumes in a cohort of subjects, independently of a potential diagnosis. Obviously such an intrinsic probability density function can never be observed, as tumors are only detected after they reach a certain volume. For tumor diagnosis a certain minimal tumor size is required such that side effects occur or that they are detectable with state-of-the–art diagnostic procedures. If it is assumed that for small volumes the probability that a tumor causes side effects and/or is imaged by a diagnostic procedure is increasing linearly with volume the actually observed tumor size distribution can be written as:

$$g_o(V) = \frac{\bar{V}+a}{\bar{V}^2} \cdot \left(1 - exp(-V/a)\right) \cdot \exp(-V/\bar{V}) \ , \tag{3}$$

where the term $1 - exp(-V/a)$ is for small volumes $V/a$. The normalization constant was chosen such that the integral over the distribution function $g_0$ is 1. In equation 3, $a$ is a parameter which characterizes the volume-specific detection efficiency of tumors. In the limit of $a = 0$ all tumors are detected and equation (3) becomes equation (2).

For model testing, breast cancer distributions were selected, as they include usually a high

number of patients and data are available at different years of diagnosis, which allows to study the impact of improved diagnosis on the parameter *a*. Here we use the data of Carter et al. [10] and Rosenberg et al. [11]. The study of Carter et al. [10] included 24,740 cases recorded in the Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute between 1977 and 1982. Rosenberg et al. [11] collected data from 72,367 cases from 1973 to 1998 also using the SEER data base. Equation 3 was fitted to the epidemiological data by using the least squares method. The fit was performed such that the average volume $\bar{V}$ was kept the same for both data-sets, assuming that the intrinsic volume distribution is equivalent for both data-sets. The parameter *a* was fitted separately, as the observation interval for the two data-sets was different.

## Determination of the probability distribution of tumor growth times

Many tumor induction models [12-16] assume that pre-malignant stem cells are transforming with a certain probability into a fully malignant cell, and, after a certain lag time into a clinically observable cancer. Usually the parameters of tumor induction models are determined by fitting tumor incidence data by assuming a fixed lag time usually taken from literature.

If a simple exponential tumor growth model is used to describe the size increase of the tumor during the lag time, the pdf which describes the tumor volume distribution $g(V)$ can be transformed into the pdf of the tumor growth time $f(L)$. Thus knowledge of the tumor volume distribution function can yield the shape of the tumor lag time distribution and vice versa. A schematic concept of the process is provided in Figure 1.

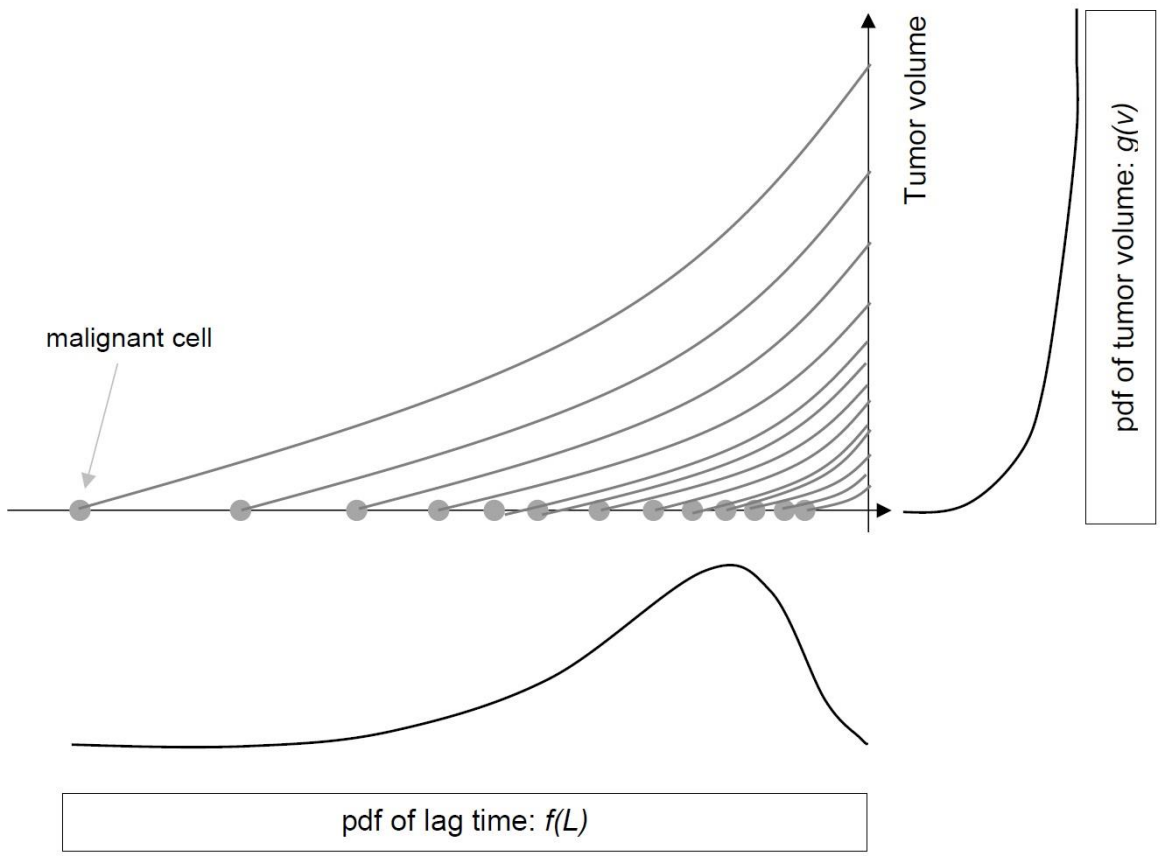


***Figure 1.*** *A schematic concept of the relation between the tumor volume pdf and the tumor growth time pdf. The horizontal axis shows tumor induction as a function of age. The occurrence of malignant cells is displayed with gray circles. Each cell grows exponentially (shown by the lines) and after a certain lag time a clinical observable tumor is grown. The clinical observable tumors are distributed according to plot shown on the vertical axis (equation 2), the corresponding lag times distributions are shown on the horizontal axis.*

Here an exponential tumor growth function is used as the transformation function to convert one probability density function into another:

$$V = V_0 \cdot exp(b \cdot L) \ , \tag{4}$$

where $b$ is the tumor growth factor and $L$ the time which is needed to develop a clinical

observable cancer from one fully malignant cell. The volume $V_0$ is the initial volume before exponential growth starts. Here $V_0 = 10^{-9}$ cm$^3$, i.e., the volume of one cell, was used. The transformation between the two pdfs is a strict mathematical procedure given by:

$$g\big(V(L)\big) = f(L) \cdot \left|\frac{dL}{dV}\right| \tag{5}$$

Using equation (4) we get:

$$L = \frac{\ln(V/V_0)}{b} \quad \text{and} \quad \frac{dL}{dV} = \frac{1}{b \cdot V}\ , \tag{6}$$

with $g(V)$ from equation (2), $f(L)$ can be easily determined using equations 5 and 6:

$$f(L) = \frac{V_0 \cdot b}{\bar{V} \cdot \exp(-\frac{V_0}{\bar{V}})} \cdot exp(b \cdot L) \cdot exp\left(-\frac{V_0}{\bar{V}} \cdot exp(b \cdot L)\right), \tag{7}$$

where equation 7 was normalized such that the integral over the lag time from 0 to infinity is one.

For the visualization of a realistic lag time distribution we have used the tumor growth factor $b$ = 1.05 for breast cancer taken from Talkington and Durrett [17].

## Determination of TCP

Another interesting application of the exponential pdf of tumor volumes described by equation 2 is the application to models of tumor control probability (TCP) after radiotherapy of cancer. Most TCP models are based on a calculation of the Poisson probability of there being no surviving cells in a population of like tumors after a fractionated treatment is given [18, 19]:

$$TCP = exp(-\rho \cdot V \cdot S) \tag{8}$$

where $\rho$ is the density of tumor clonogens per volume, $V$ the tumor volume and $S$ the surviving fraction of cells after radiotherapy treatment. If the TCP curves are generated with single values of intrinsic radiosensitivity and clonogenic cell number, the experimentally obtained data cannot be fitted, as the TCP-curves show a very steep gradient. This discrepancy is

usually explained by the fact that TCP curves derived from populations of human tumors that have been similarly treated include different tumor volumes, different initial clonogen densities, different intrinsic radiosensitivites and different oxygenation status. Unfortunately all these variations cannot be modeled analytically (at least in a general view) and as a consequence, equation 8 is usually calculated numerically by modelling the above mentioned variations. Instead of numerical calculations, very often the observed tumor control is fitted to a logistic model of TCP (e.g. [9]). However, to our knowledge it was never shown, that this empirical approach is based on a mechanistic understanding of the involved processes.

If it is assumed that the tumor volumes which are treated with radiotherapy are distributed according to equation 2, it is possible to model the variation of tumor volumes and thus clonogen number, assuming a constant clonogenic density. When integrating TCP over a variation of tumor volumes we get:

$$TCP = \int_0^\infty exp(-\rho \cdot V \cdot S) \cdot \frac{1}{\bar{V}} \cdot \exp\left(-\frac{V}{\bar{V}}\right) \cdot dV = \frac{1}{1+\rho \cdot \bar{V} \cdot S} \tag{9}$$

The result of this integration yields an analytic expression for TCP in terms of a logistic model.

For the visualization of TCP curves the surviving fraction $S$ was modeled using the linear quadratic model:

$$S = \exp\left(-D \cdot \left(\alpha + \beta \cdot d_f\right)\right) \tag{10}$$

Some typical parameters have been used: $\rho = 10^8$ cm$^{-3}$, $\alpha = 0.36$ Gy$^{-1}$, $\beta = 0.036$ Gy$^{-2}$, $D$ = 30…70 Gy and $d_f$ = 2 Gy, where $\alpha$ and $\beta$ are the parameters of the LQ-model, $D$ is the total radiotherapy dose and $d_f$ is the dose per fraction.

# Results

## Probability distributions of tumor sizes

In Figure 2 tumor volume distributions of breast cancer are shown. The data points marked by diamonds represent the volume distribution from the Carter data including cases until 1982, whereas the squares belong to the Rosenberg data obtained between 1973 and 1998. The statistical error is of the order of the size of the symbols and was therefore not plotted. Clearly a shift of the maximum of the probability distribution of the newer data to smaller volumes can be observed. This is reflected by a smaller value of $a = 0.42cm^3$ when compared to older data ($a = 1.84\ cm^3$). The fitted average volume of the two data sets is 20.7 $cm^3$. It should be noted that the average volume $\bar{V}$ corresponds to the intrinsic tumor volume distribution and not of the observed one. The fitted parameters are listed in table 1.

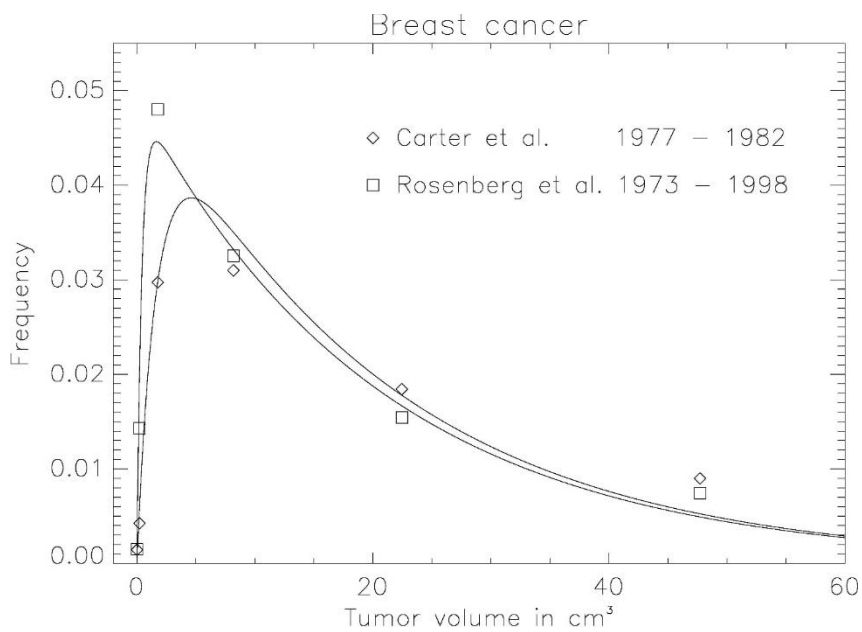


***Figure 2.*** *Probability density distribution of breast cancer induction. The diamonds and squares show observed data from Carter et al. [10] and Rosenberg et al. [11], respectively. The lines represent the fit of equation 3. The fitted parameters a and* $\bar{V}$ *are listed in table 1.*

**Table 1. Fitted parameters of the for tumor volume distributions**

| Parameter | | Model data from | |
|---|---|---|---|
| | | Carter et al. [10] | Rosenberg et al. [11] |
| detection efficiency | $a$ [cm$^3$] | 1.84 | 0.42 |
| average volume | $\bar{V}$ [cm$^3$] | 20.7 | 20.7 |

## Determination of the probability distribution for tumor growth times

The expression for the pdf of breast tumor lag times from equation 7 is plotted in Figure 3. A typical tumor growth factor $b$ for breast cancer and an average tumor volume of 20.7 cm$^3$ obtained from the comparison of the tumor size distribution with the epidemiological data (last section) was used. The average lag time for breast cancer was determined to be 22.1 years with a variance of 1.5 years.

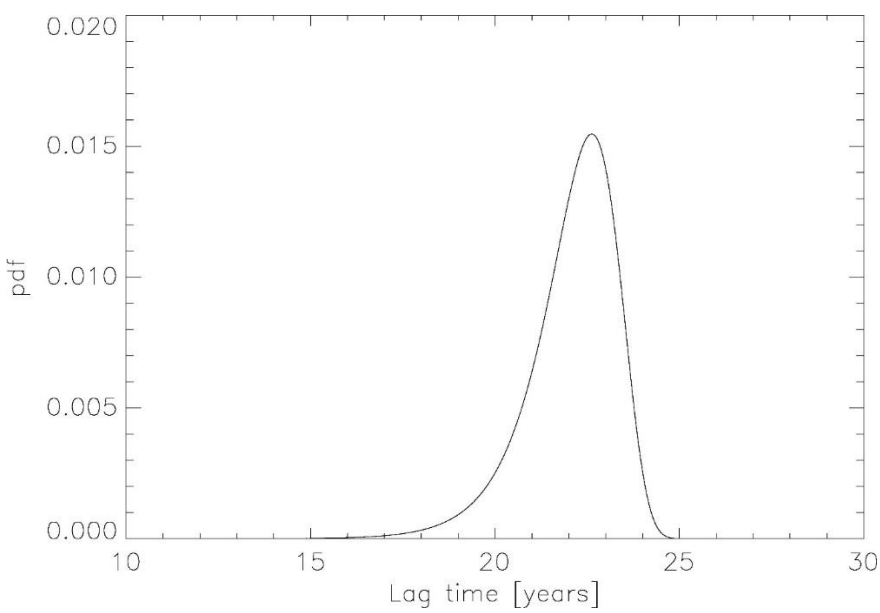


***Figure 3.*** *Plot of the pdf of breast tumor lag times from equation 7. A typical tumor growth factor b = 1.05 for breast cancer and an average tumor volume of 20.7 cm$^3$ was used.*

## Determination of TCP

In Fig. 4 the TCP model of equation 8 is plotted as the dotted line for three different average tumor volumes 10, 100 and 1000 $cm^3$, respectively. The solid lines represent the TCP calculations including an exponential distribution of tumor volumes from equation 9. Clearly, a broadening of the TCP curve is visible.

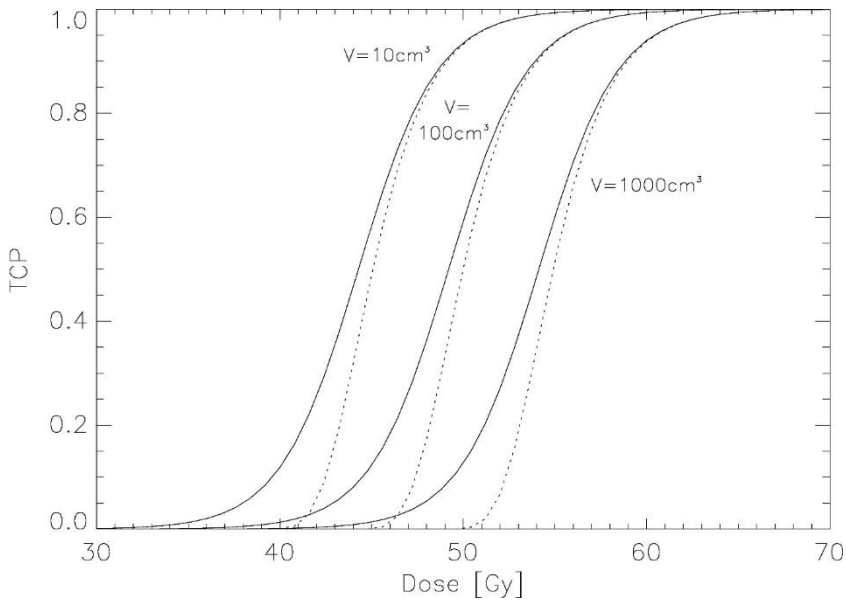


***Figure 4.*** *Plot of the TCP model of equation 8 for individual patients as the dotted line for three different average tumor volumes 10, 100 and 1000 cm3, respectively. The solid lines shows the TCP model for a volume averaged patient cohort according to equation 9.*

# Discussion

In this work it was hypothesized that the intrinsic distribution of tumor volumes in a cohort is exponentially distributed. To account for the fact that the probability for the clinical detection of a tumor increases with tumor volume the exponential distribution was modified with a

sigmoidal function. The fit of this distribution function to breast volume data from literature covering two different time intervals of breast cancer diagnosis results in two different locations of the distribution maximum and consequently in a different parameter *a*. The data set covering diagnosed breast cancers from 1973 – 1998 results in much smaller *a* (0.31) than the older data set (1977 - 1982; *a*=1.20). This can be explained with the clinical introduction of mammography in the late 1980's which resulted in the diagnosis of much smaller breast tumors than before [8]. Therefore a smaller *a* is expected for the newer data-set as the mammography diagnosed tumors are only included in the newer data [11]. A limitation of the tumor volume model is that we considered in this work only breast cancer data. However, the intent of this work was a proof of principle and detailed analysis of tumor volume distributions for various tumor sites is in preparation.

If exponential tumor growth is used as a transformation function the probability density function for tumor volumes can be transformed into the probability density function of tumor growth or lag times, respectively. The obtained lag time distribution function can be used in the development of mechanistic models of background cancer risk, which basically model the occurrence of a malignant cell per individual per year [12-16]. Usually an approximated constant lag time is used in such models. It can be hypothesized, that replacement of a constant lag time by a lag time distribution could improve the quality of such models.

It was also shown that the assumed exponential variation of tumor volumes can be used to deduce an analytical representation for the population averaged TCP. The Poisson based TCP model for individuals was integrated by using the volume distribution as a weighting function. This resulted in the logistic function as a population averaged TCP model. The logistic function was used in the past to fit successfully observed tumor control of radiotherapy patients without giving a bio-physical justification for its use.

A limitation of the deduction of the TCP model is that we considered here only tumor size variations. However radiotherapy patients who have been similarly treated include not only different tumor volumes, but also different initial clonogen numbers, different intrinsic radiosensitivites, different oxygenation status and different growth fractions, all of which were not considered in this work.

Another limitation is that tumor growth was simply modeled with an exponential model. An exponential model might be a satisfying approximation as long as the tumors are small. However, it is known that the exponential model is not universally appropriate and many other models including the Gompertz model have been proposed [17]. It might be interesting to apply the formalism of this work in the future to other tumor growth models.

# References


1. Verschraegen C, Vinh-Hung V, Cserni G, Gordon R, Royce ME, Vlastos G, Tai P, Storme G. Modeling the effect of tumor size in early breast cancer. Ann Surg. 2005 Feb;241(2):309-18. doi: 10.1097/01.sla.0000150245.45558.a9.

2. P Ghosh D, Banerjee M, Biswas P. Inference for constrained estimation of tumor size distributions. Biometrics. 2008 Dec;64(4):1009-17. doi:10.1111/j.1541-0420.2008.01001.x. Epub 2008 Mar 27.

3. Shaevitch D, Taghipour S, Miller AB, Montgomery N, Harvey B. Tumor size distribution of invasive breast cancers and the sensitivity of screening methods in the Canadian National Breast Screening Study. J Cancer Res Ther. 2017 Jul- Sep;13(3):562-569. doi: 10.4103/0973-

1482.174539. PMID: 28862227.

4. Kimmel M, Flehinger BJ. Nonparametric estimation of the size-metastasis relationship in solid cancers. Biometrics. 1991 Sep;47(3):987-1004. PMID: 1742451.

5. Xu JL, Prorok PC. Nonparametric estimation of solid cancer size at metastasis and probability of presenting with metastasis at detection. Biometrics. 1997 Jun;53(2):579-91. PMID: 9192453.

6. Xu JL, Prorok PC. Estimating a distribution function of the tumor size at metastasis. Biometrics. 1998 Sep;54(3):859-64. PMID: 9750239.

7. Ball D, Mitchell A, Giroux D, Rami-Porta R; IASLC Staging Committee and Participating Institutions. Effect of tumor size on prognosis in patients treated with radical radiotherapy or chemoradiotherapy for non-small cell lung cancer. An analysis of the staging project database of the International Association for the Study of Lung Cancer. J Thorac Oncol. 2013 Mar;8(3):315-21. doi: 10.1097/JTO.0b013e31827dc74d. PMID: 23334092.

8. Welch HG, Prorok PC, O'Malley AJ, Kramer BS. Breast-Cancer Tumor Size, Overdiagnosis, and Mammography Screening Effectiveness. N Engl J Med. 2016 Oct 13;375(15):1438-1447. doi: 10.1056/NEJMoa1600249.

9. Okunieff P, Morgan D, Niemierko A, Suit HD. Radiation dose-response of human tumors. Int J Radiat Oncol Biol Phys. 1995 Jul 15;32(4):1227-37. doi:10.1016/0360-3016(94)00475-z.

10. Carter CL, Allen C, Henson DE. Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases. Cancer. 1989 Jan 1;63(1):181-7. doi: 10.1002/1097-0142(19890101)

11. Rosenberg J, Chia YL, Plevritis S. The effect of age, race, tumor size, tumor grade, and disease stage on invasive ductal breast cancer survival in the U.S. SEER database. Breast Cancer Res Treat. 2005 Jan;89(1):47-54. doi: 10.1007/s10549-004-1470-1.

12. Armitage P, Doll R. The age distribution of cancer and a multi-stage theory of carcinogenesis. Br J Cancer. 1954 Mar;8(1):1-12. doi: 10.1038/bjc.1954.1.

13. Moolgavkar SH. The multistage theory of carcinogenesis and the age distribution of cancer in man. J Natl Cancer Inst. 1978 Jul;61(1):49-52. doi: 10.1093/jnci/61.1.49.

14. Moolgavkar S. Multistage models for carcinogenesis. J Natl Cancer Inst. 1980 Aug;65(2):215-6.

15. Moolgavkar SH, Knudson AG Jr. Mutation and cancer: a model for human carcinogenesis. J Natl Cancer Inst. 1981 Jun;66(6):1037-52. doi: 10.1093/jnci/66.6.1037.

16. Shuryak I, Hahnfeldt P, Hlatky L, Sachs RK, Brenner DJ. A new view of radiation-induced cancer: integrating short- and long-term processes. Part I: approach. Radiat Environ Biophys. 2009 Aug;48(3):263-74. doi: 10.1007/s00411-009-0230-3. Epub 2009 Jun 18. Erratum in: Radiat Environ Biophys. 2011 Nov;50(4):607-8.

17. Talkington A, Durrett R. Estimating tumor growth rates in vivo. Bull Math Biol. 2015 Oct; 77(10): 1934–1954

18. Webb S, Nahum AE. A model for calculating tumour control probability in radiotherapy including the effects of inhomogeneous distributions of dose and clonogenic cell density. Phys Med Biol. 1993; 38: 653

19. Chapman JD, Nahum AE. RADIOTHERAPY TREATMENT PLANNING: Linear-Quadratic

Radiobiology, CRC Press Taylor and Francis, Boca Raton 2015.