Tumour volume distribution can yield information on tumour growth and tumour 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: Tumour volume distribution can yield information on tumour growth and tumour control

Abstract

Background: It is shown that tumour volume distributions, can yield information on two aspects of cancer research: tumour induction and tumour control.

Materials and methods: From the hypothesis that the intrinsic distribution of breast cancer volumes follows an exponential distribution, firstly the probability density function of tumour growth time was deduced via a mathematical transformation of the probability density functions of tumour volumes. In a second step, the distribution of tumour volumes was used to model the variation of the clonogenic cell number between patients in order to determine tumour control probabilities for radiotherapy patients.

Results: Distribution of lag times, i.e. the time from the appearance of the first fully malignant cell until a clinicaly observable cancer, can be used to deduce the probability of tumour induction as a function of patient age. The integration of the volume variation with a Poisson-TCP model results in a logistic function which explains population-averaged survival data of radiotherapy patients.

Conclusions: The inclusion of tumour volume distributions into the TCP formalism enables a direct link to be deduced between a cohort TCP model (logistic) and a TCP model for individual patients (Poisson). The TCP model can be applied to non-uniform tumour dose distributions.

Introduction

The estimation and analysis of tumour size[ǂ] and associated size distributions have mainly been applied to modelling their effects on cancer survival or tumour progression [1-8]. A major conclusion of published work is that tumour size has a significant impact on survival of patients treated with radiotherapy. In this work we focus on alternative applications of tumour size distributions. We will show that tumour-size distributions, or more precise tumour-volume distributions, can yield valuable information on two completely different areas of cancer research. Firstly we will show that the probability density distribution of detected tumour sizes is strictly related to the probability density function of tumour growth time. The tumour 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 tumour induction models this lag time is usually estimated as a fixed number. We believe that the use of the determined lag time distribution in tumour induction models has the potential to improve the quality of such models.

Secondly, the distribution of cancer volumes reflects the distribution of clonogenic cell numbers (assuming a constant clonogenic cell density) and thus can be used for modelling tumour control probability for cohorts of radiotherapy cancer patients. Currently no general analytical formulation of TCP exists which includes the population variation of tumor volumes. 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.

[ǂ] By tumour size we mean here the size of the tumour when the treatment starts.

Materials and Methods

Probability distributions of tumour volumes

We assume here that the intrinsic probability density function (pdf) for the distribution of tumour 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 $1/\lambda$ one can write:

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

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

With intrinsic we mean here that the exponential probability distribution describes the distribution of all tumour volumes in a cohort of subjects, independently of a potential diagnosis. E.g., if we would select a specific age group of women, all breast tumours, including also un-observed and microscopically small ones, would be distributed exponentially. However, the mean tumour volume is assumed to change with the age of the observed cohort.

Obviously, such an intrinsic probability density function can never be observed, as tumours are only detected after they reach a certain volume. For tumour diagnosis a certain minimal tumour size is required such that they are detectable with state-of-the–art diagnostic procedures. If it is assumed that for small volumes the detection probability of a tumour is increasing linearly with volume, the actually observed tumour volume distribution $g_0(V)$ 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

representing a characteristic volume such that $\left(1 - exp(-V/a)\right)$ describes the detection rate. In the limit of $a = 0$ all tumours are detected and the distribution function describes the intrinsic tumour-volume distribution. The value of a depends on the quality of the applied diagnostic procedures. As we will see later, a was decreased after mammography was employed for the detection of breast cancers.

For model testing, breast cancer volume distributions were selected, as they include usually a large number of patients and the data are available at different times, which allows the impact of improved diagnosis on the parameter a to be studied. 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 identical 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 tumour growth times

Many tumour induction models [12-16] assume that pre-malignant stem cells transform themselves with a certain probability, into fully malignant cells, and, after a certain lag time L, into a clinically observable cancer. Thus, L is the lag (time) between the first fully malignant cell and cancer. Usually the parameters of tumour induction models are determined by fitting tumour incidence data by assuming a fixed lag time usually estimated from the A-bomb survivor data [17 - 18] or from comparing tumour size at different time [19].

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

Here an exponential tumour growth function is used converts one probability density function into another:

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

where b is the tumour growth factor and L the lag time. The volume $V_{\emptyset}$ is the initial volume before exponential growth starts. Here $V_{\emptyset} = 10^{-9}$ cm^3, i.e., the volume of one cell, was used. The transformation between the two pdfs is a strict mathematical procedure given by [20]:

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

Using equation (4) we get:

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

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

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

where equation 7 was normalized such that the integral over $f(L)$ from $L = 0$ to infinity is unity. We have chosen infinity as the upper integration limit instead of a realistic maximum tumour volume, as the probability for very large tumours is practically zero. E.g., assuming an average tumor volume of 40 cm^3 an integration of the exponential distribution up to a realistic upper limit of 500 cm^3 yields 0.999996 instead of 1.

For the visualization of a realistic lag time distribution we have used the tumour growth factor $b = 1.05$ y^{-1} for breast cancer taken from Talkington and Durrett [19].

Determination of TCP

Another interesting application of the exponential pdf of tumour volumes described by equation 2 is the application to models of tumour control probability (TCP) following radiotherapy. Most TCP models are based on a calculation of the Poisson probability of there being no surviving cells in a population of tumours after a fractionated treatment [21, 22]:

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

where ρ is the number of tumour clonogens per unit volume, V is the tumour volume and S is 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 tumours that have been similarly treated include different tumour volumes, different initial clonogen densities, different intrinsic radiosensitivities and different oxygenation status. However, these parameters cannot all be modeled analytically and as a consequence, they are often varied numerically (e.g. [23, 24]). An alternative to numerical calculations is the use of a logistic model to fit the observed tumour control (e.g. [9]). However, to our knowledge it has never been shown that this empirical approach is based on a mechanistic understanding of the involved processes.

By assuming that the tumour volumes are distributed according to equation 2, and assuming a constant clonogenic density we obtain:

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

The result of this integration yields an analytic expression for TCP in terms of a logistic model were survival S is modelled by an exponential function 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}, $\bar{V}$ = 10, 100 and 1000 cm^3, $\alpha = 0.36$ Gy^{-1}, $\beta = 0.036$ Gy^{-2}, D = 30…70 Gy and d_f = 2 Gy, where α and β 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 tumour sizes

In Figure 2 tumour 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. The data were fitted to equation 3. A shift of the maximum of the probability distribution of the newer data to smaller volumes can clearly be observed. This is reflected by a smaller value of a = 0.42 cm^3, hence better detection efficiency, when compared to older data (a = 1.84 cm^3). The average volume obtained from the fit of the exponential distribution to the two data sets is 20.7 cm^3. The fitted parameters with the corresponding errors are listed in table 1.

Determination of the probability distribution for tumour growth times

The expression for the pdf of breast tumour lag times from equation 7 is plotted in Figure 3. A typical tumour growth factor *b* for breast cancer and an average tumour volume of 20.7 cm^3 obtained from the comparison of the tumour size distribution with the epidemiological data was used. The average lag time for breast cancer was determined to be 22.1 years with a variance of 1.5 years. This value is consistent with the findings from the A-bomb survivors with a lag time between 10 and 30 years [17 - 18] and with studies comparing tumour size at different time. Talkington and Durrett [20] found for breast cancer a growth rate of 1.05/year for an exponential model. This translates into a lag time of 22 years for a tumor of volume 1.3 cm^3.

Determination of TCP

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

Discussion

In this work it was hypothesized that the intrinsic distribution of tumour volumes in a cohort is exponentially distributed. To account for the fact that the probability of the clinical detection of a tumour increases with tumour 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 different values of parameter *a*. The data set covering

diagnosed breast cancers from 1973 – 1998 results in a much smaller *a* (0.31) than the older data set (1977 - 1982; *a*=1.20). This can be explained by the clinical introduction of mammography in the late 1980s which resulted in the diagnosis of much smaller breast tumours than before [8]. Therefore, a smaller *a* is expected for the newer data-set as only mammography-diagnosed tumours are included in the newer data [11]. A limitation of the tumour 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 tumour volume distributions for various tumour sites is in preparation.

If exponential tumour growth is used as a transformation function the probability density function for tumour volumes can be transformed into the probability density function of tumour growth or lag times, respectively. The obtained lag time distribution function can be used in the development of mechanistic models of background cancer risk. Usually, such models combine the probability of occurrence of a malignant cell per individual per year with a constant lag time in order to estimate cancer induction [12-16]. Replacing a constant lag time by a lag time distribution may improve the quality of such models. We would like to note that the average lag time for breast cancer shown in Figure 3 depends strongly on the tumour growth factor *b* which we have taken from literature. The goal of this work was not to determine the absolute lag times for breast cancer, but to derive an expression for the distribution of lag times.

A limitation of this work is that tumour growth was simply modeled with an exponential model. An exponential model might be a satisfying approximation as long as the tumours 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 [19]. It might be interesting to apply the formalism of this work to other tumour growth models and to study its impact on the tumour size distribution.

It was shown that the assumed exponential variation of tumour volumes can be used to derive

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 tumour control data for radiotherapy patients without any bio-physical justification. In this work a logistic TCP model for cohorts of patients is explicitly derived from a Poisson TCP-model for individuals.

The methodology presented in this work enables in principle the cohort TCP model given by equation 9 to be fitted to survival data for groups of patients. However, radiotherapy patients who have been similarly treated include not only different tumour volumes, but also different clonogen densities, different intrinsic radiosensitivites and different oxygenation, none of which were considered in this work. We are currently working on an integration of most of these factors into our analytical model. Therefore, this work should be viewed as a first step towards the development of a TCP model that can be applied equally to individuals and to populations of patients.

References

1. Verschraegen C, Vinh-Hung V, Cserni G, Gordon R, Royce ME, Vlastos G, Tai P, Storme G. Modeling the effect of tumour 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 tumour 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. Tumour 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 tumour 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 tumour 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 Tumour 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 tumours. 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 tumour 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, tumour size, tumour 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. Tokunaga M, Norman JE Jr, Asano M, Tokuoka S, Ezaki H, Nishimori I, Tsuji Y. Malignant breast tumours among atomic bomb survivors, Hiroshima and Nagasaki, 1950-74. J Natl Cancer Inst. 1979 Jun;62(6):1347-59.

18. Tokunaga M, Land CE, Tokuoka S, Nishimori I, Soda M, Akiba S. Incidence of female breast cancer among atomic bomb survivors, 1950-1985. Radiat Res. 1994 May;138(2):209-23.

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

20. https://en.wikibooks.org/wiki/Probability/Transformation_of_Probability_Densities at 25.01.2021.

21. 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

22. Chapman JD, Nahum AE. RADIOTHERAPY TREATMENT PLANNING: Linear-Quadratic Radiobiology, CRC Press Taylor and Francis, Boca Raton 2015.

23. Chapman J.D., Nahum A. Radiotherapy Treatment Planning: Linear-Quadratic Radiobiology. CRC Press Boca Raton, Florida; first edition 21. April 2016.

24. Roberts SA, Hendry JH. A realistic closed-form radiobiological model of clinical tumour-

control data incorporating intertumour heterogeneity. Int J Radiat Oncol Biol Phys. 1998 Jun 1;41(3):689-99. doi: 10.1016/s0360-3016(98)00100-x.

Table 1. Fitted parameters of the for tumour volume distributions with one standard deviation

Parameter		Model data from	
		Carter et al. [10]	Rosenberg et al. [11]
detection efficiency	a [cm^3]	1.84±0.48	0.42±0.22
average volume	$\bar{V}$ [cm^3]	20.7±2.2	20.7±1.2

Figure legends

Figure 1. A schematic concept of the relation between the tumour volume pdf and the tumour growth time pdf. The horizontal axis shows tumour 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 tumour is grown. The clinical observable tumours are distributed according to plot shown on the vertical axis (equation 2), the corresponding lag times distributions are shown on the horizontal axis.

Figure 2. Probability density distribution of breast cancer induction. The diamonds and squares show the frequency distribution of the 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.

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

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

Figure 1

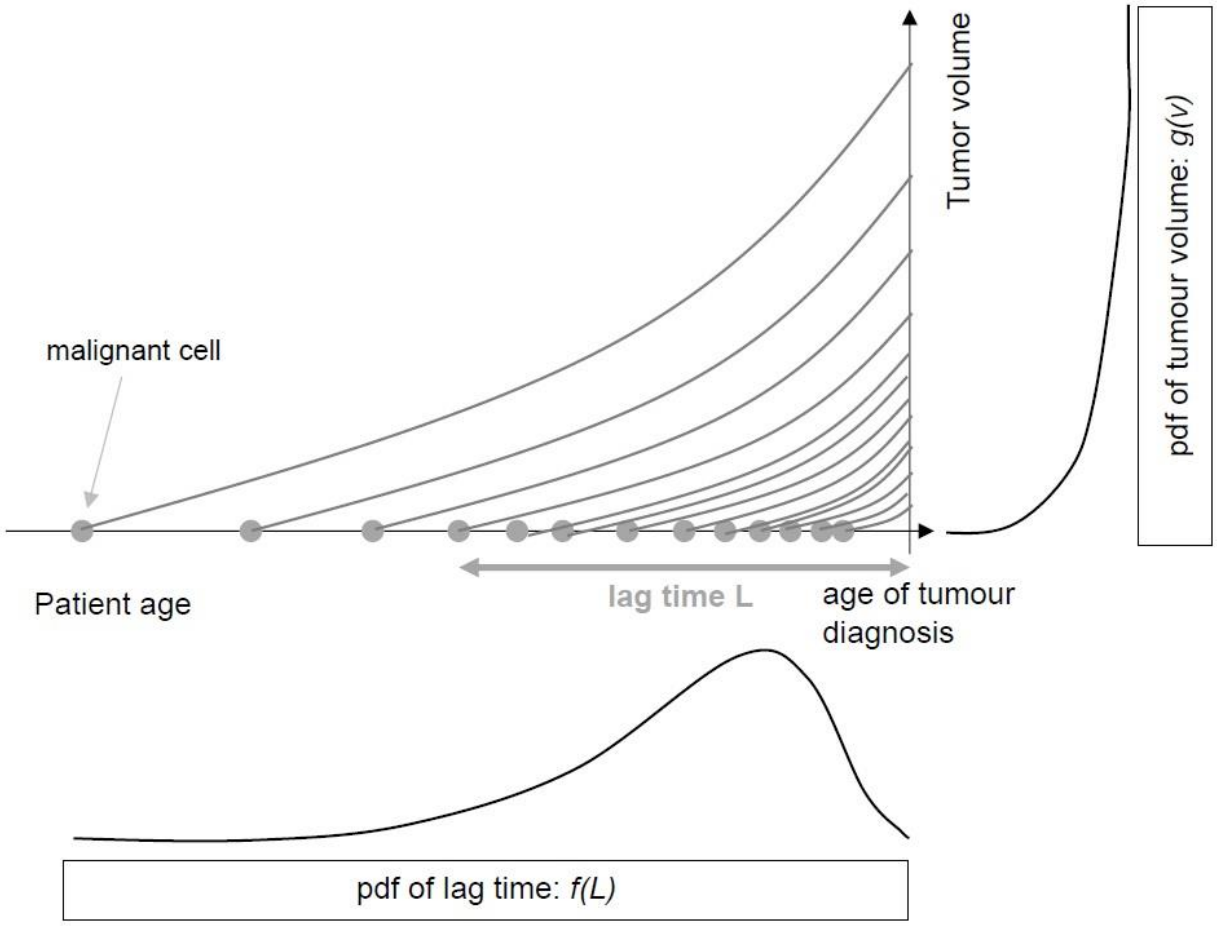

Tumor volume
pdf of tumour volume: g(v)
malignant cell
Patient age
lag time L
age of tumour diagnosis
pdf of lag time: f(L)

Figure 2

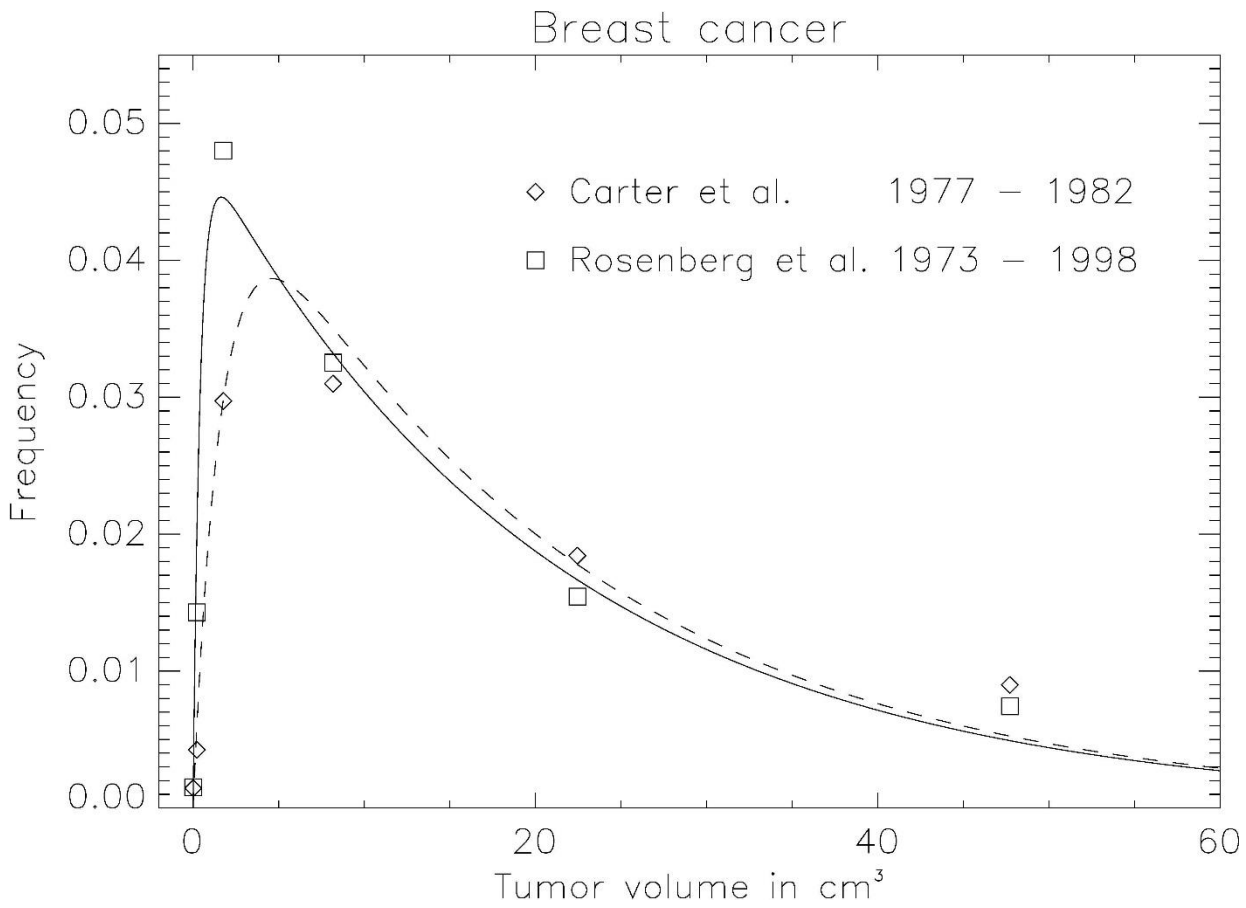

Breast cancer
Carter et al. 1977 – 1982
Rosenberg et al. 1973 – 1998
Frequency
0.05
0.04
0.03
0.02
0.01
0.00
0
20
40
60
Tumor volume in cm³

Figure 3

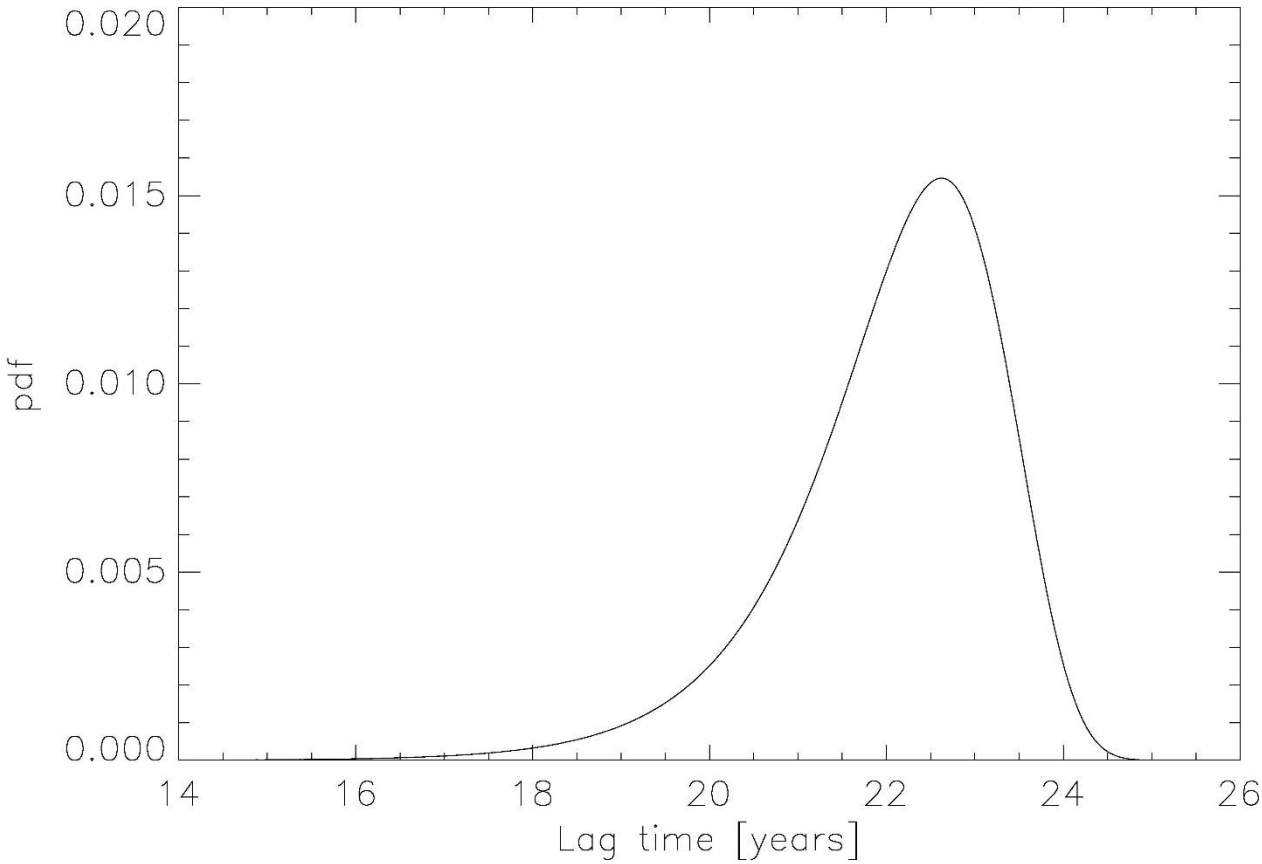
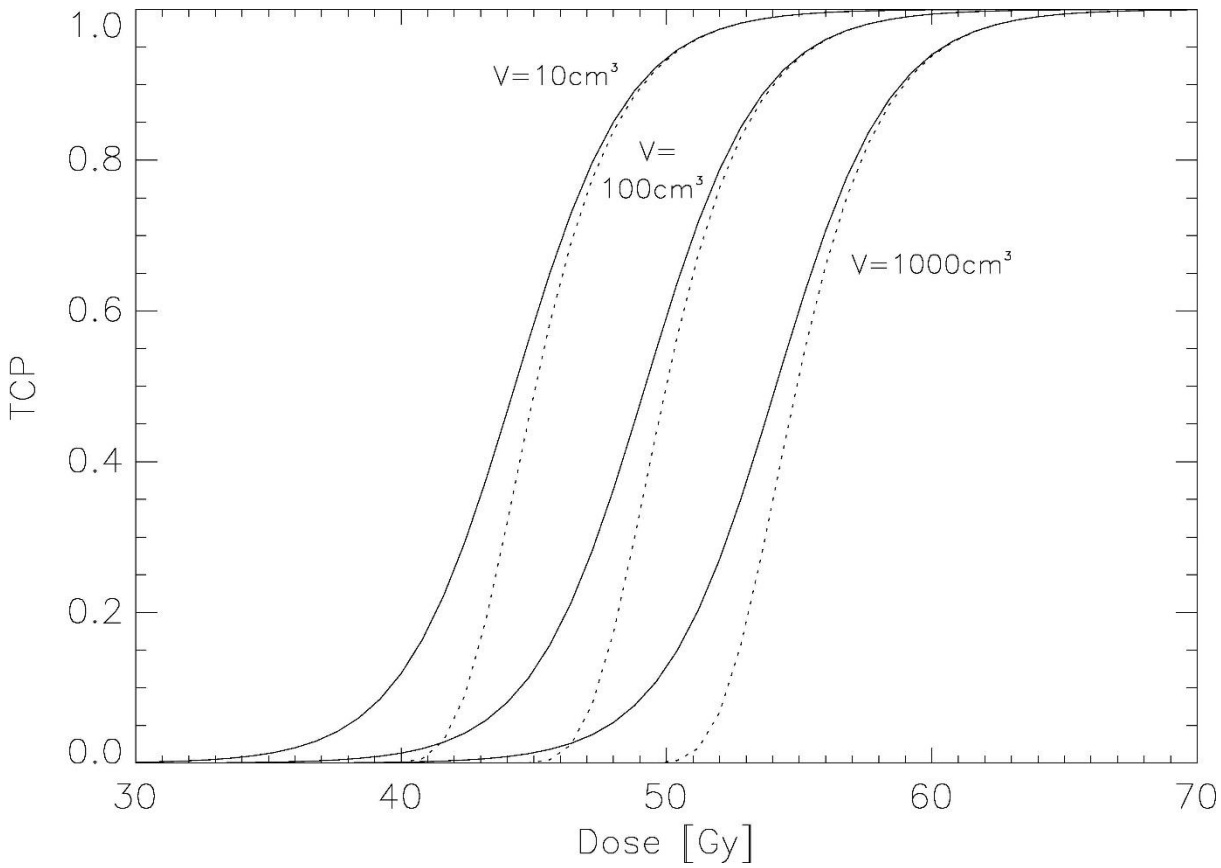


Figure 4


1.0
0.8
0.6
0.4
0.2
0.0
TCP
30
40
50
60
70
Dose [Gy]
V=10cm³
V=
100cm³
V=1000cm³