



REVIEW ARTICLE

A compact review of Probability Models for Cancer

Christos P. Kitsos¹, Constantinos-Symeon Nisiotis²

¹University of West Attica,
Department of Informatics and
Computer Engineering, Egaleo,
Athens, GR

²University of West Attica,
Department of Informatics and
Computer Engineering, Egaleo,
Athens, GR

ABSTRACT

The target of this paper is to review the main Probability models that have been proposed to examine different problems in (experimental) carcinogenesis. The models have been grouped, classified and analysed, while their necessity was discussed. We were referred Data Analysis for Breast Cancer, which has been faced under different Mathematical lines of approach, as with fractals, information measures, among them.



OPEN ACCESS

PUBLISHED

30 October 2024

CITATION

Kitsos, CP., and Nisiotis, CS., 2024.
A compact review of Probability
Models for Cancer. Medical
Research Archives, [online] 12(10).
<https://doi.org/10.18103/mra.v12i10.5819>

COPYRIGHT

© 2024 European Society of
Medicine. This is an open- access
article distributed under the terms
of the Creative Commons
Attribution License, which permits
unrestricted use, distribution, and
reproduction in any medium,
provided the original author and
source are credited.

DOI

<https://doi.org/10.18103/mra.v12i10.5819>

ISSN

2375-1924

1. Introduction

The aim of this paper is to discuss and provide a compact and critical review for the appropriate Probability Models (Probability Model Oriented Data Analysis) concerning the statistical bioassays of the experimental carcinogenesis. Recently more Mathematical approaches have been developed to face the Cancer problem, the fractals being the most popular.¹ These methods are not stochastic, there is not a Probability distribution following the underlying Biological mechanism, but certainly deserve particular interest. Since the early 50's interest in the cancer risk assessment has led to various statistical lines of thought, due to different intrinsic biological background considered each time to tackle the problem and supported by the appropriate statistical analysis.

The biological insight (cell proliferation, mechanism of inhibition in mutagenesis and carcinogenesis) of a cancer risk assessment has not been discussed here – it is really too complicated, but we devote this paper to Statistical consideration to face the problem. Emphasis was given to the statistical models considered in the literature in the area of experimental carcinogenesis. Moreover, an optimal experimental design approach is discussed, for evaluating the percentiles of a distribution, facing the Low Dose problem, while the D – optimality, always with an aesthetic appeal in Biological – Toxicological problems, is chosen as an optimal design criterion. We applied this method for the typical One-Hit model.

There is a different line of thought facing the "Cancer Problem" for different Sciences: There is the Medical approach, Biological, Toxicological and even Mathematical. There is a completely mathematical approach to the Probability Models emerged to discuss the Cancer, which are unpleasant for the experimentalist,² however beyond the target of this paper. We try to bridge the different lines of facing the problem, around the Statistical way of supporting the investigation of the problem.^{3,4}

An example of Epidemiological analysis was discussed by Bernal et al.,⁵ Hayat et al.,⁶ Angelopoulou et al.,⁷

among others, while a Biologically based approach adopting the Statistical theory was followed by Coglianno et al.,⁸ a Medical approach by Chyczewski et al.,⁹ and, finally, as a Relative Risk development appeared in various cases and approaches by Edler and Kitsos,¹⁰ where a number of more than one thousand references concerning the Cancer Risk Assessment are listed. A large area of different research was covered by Cogliano et al.,⁸ covering both Biological and Statistical development, in their pioneering work.

The pioneering work of Moolgavkar and Venzon,¹¹ Moolgavkar and Knudson¹² and latter that of Luebeck and Moolgavkar^{13,14,15} addressed the model building problem towards to probability models, rather than to the statistical line of thought. Kopp-Schneider¹⁶ is working within the same probability theory spirit, while the problem known as the "low – dose extrapolation", is in the boundary with the curve fitting methodology. In such cases the limits of the confidence interval of the extrapolating value are used. We believe that the tolerance intervals are more appropriate, covering that the 95% of future observations will lie to the tolerance interval, with certain probability, Muller and Kitsos.¹⁷

There is a number of chemicals that have been identified as *initiators* for a cancer process, through experimentation, while others have been considered as *promoters*. Initiation and promotion, in animal experiments are referred as well as *early – stage* and *late – stage* respectively, from the epidemiological point of view. It has been underlined that a cancer bioassay should be considered as the one that proliferation might increase the frequency of mutations as a consequence of errors in replication or of the conversion of endogenous or exogenous DNA adducts into stable mutations. Moreover, a number of different animal experiments, especially those conducted in the skin of mice, have been used to distinguish the different levels of cancer risk, working under the Risk Analysis. Therefore, experimental carcinogenesis has an important role in examining cancer initiation and promotion.

The Risk Analysis problem does not accept a unique line of thought and development, but highly depends on the nature of the problem. Typical example is the food processing. It is clear in such cases, and not only, that the chemical hazards can be either naturally occurring (mycotoxins, pyrrolizidine, alkaloids etc) or added chemical hazards (pesticides, antibiotics, hormones, heavy metals etc), see Kitsos,¹⁸ for examples. The Cancer problem, , admits, so to speak, this data analysis line, and created "models" for the development of tumour, and epidemiological analysis for describing other characteristics, food being one of them, associating with the Relative Risk. This paper is focused to the former case of study and provides a classification of the developed models.

The optimal experimental design techniques, Kitsos,¹⁹ usually have as a primary goal to extract the maximum amount of unbiased information regarding the factors affecting the response of the experimentation, from a "small" number of observations and get the "best" possible estimators, with Statistical criteria. Risk Assessment in a Bioassay, National Research Council (NCR),²⁰ complex experimental design among many factors that influence the response are often regarded as a "nuisance", and the problem is getting worse in the neighbourhood of zero, as all the assumed models, appears to be "linear" in that small area, so extrapolation is dangerous for misleading results. We believe the tolerance intervals are more appropriate, as we have already mentioned.

The focus of interest lies in the qualification of the time to tumour, as it depends on various environmental factors, such as chemical substances or radiation. Since the early work of Armitage and Doll²¹ – also Armitage,^{22,23} Doll^{24,25} – Probability models have been used to describe the process of forming benign and malignant tumours. The cancer problem was eventually the most beloved and impressive statistical problem under consideration and Sir David Cox was providing a number of working examples, Cox,²⁶ Cox and Snell,²⁷ Kitsos,²⁸ with his hazardous function being a fundamental

tool for applications. Epidemiological studies were performed under Statistical cover and essential "parameters" were evaluating, even without the support of Statistical packages. In principle there are two main reasons for formulating Probability models of carcinogenesis, to:

- provide a framework for evaluating the consequences of the proposed mechanisms of carcinogenesis.
- help to determine allowable concentrations of known carcinogens in the environment, and to estimate the consequences of exceeding them.

This is necessary because animal experiments must be done at concentrations high enough to cause some of the animals to develop tumours, while environmental concentrations must be low enough to produce very few tumours in man. Thus, apart from the great difficulties due to interspecies differences, animal experiments cannot be used directly to study low concentrations. Therefore, beyond the Statistical analysis, covered widely with software now, some Probability model theory is needed, which is not always covered from popular software, to extrapolate the dose – response relationships downward from the high doses used in animal experiments to the low doses to be allowed in the environment.

In principle there are two different approaches in cancer modelling:

1. Those models that consider the whole organism as the modelling unit and describe the time to overt tumour in this unit.
2. Models that describe the formation of carcinomas at the level of the cell, since knowledge is accumulating about the cellular biological events leading to cancer.

The problems we shall describe in this paper are under the line of thought of Probability and Statistics. As far as the shape of the tumour concerns, based on tissue image analysis, can be studied under (multi)fractal analysis, Stehlik et al.,²⁹ describing

different tumour groups. In principle a fractal is a geometric shape containing detailed structure even at arbitrarily small scales, attracting interesting in Biology and Medicine, due to the computational improvements, for their strong Mathematical insight, Losa and Nonnemacher.³⁰ A particular study based on fractals for mammary cancer, Hermann et al.,³¹ provides evidence that the method can be adopted on case study, by case study as the probability models are devoted for particular cases. Fractal methods, although have developed a strong Mathematical background they haven't developed an "intimacy", there is no feedback with users and have not grasped the potential for failure, eventually the existent

$$P(d, t) \propto P(X \leq d|t) := F(d) \quad \text{or} \quad P(d, t) \propto P(T \leq t|d) := G(t) \quad (2.1)$$

The Probability models, which have been proposed to describe the process of carcinogenesis, differ by their realism and the description of the underlying mechanism. due the assumed Biological insight and their Mathematical tractability. In principle, the underlying mechanism of the process is not usually known. There are special guidelines for carcinogenic risk, see US EPA.^{32,33} We are looking for that model, which provides a "satisfactory" approximation to the true process, using the Kolmogorov Statistical Nonparametric test to verify the approximation, to provide evidence that the chosen model, among various rivals is the right one, Wosniok et al. (1998).³⁴ In principle when the distribution of cancer occurrence is requested, in dependence of dose level d , and time dependence t , in experimental carcinogenesis, and not only, it is needed to define the cancer occurrence for the underlying model: the common definition is to declare that cancer has irreversibly occurred, when the first cell has reached the final stage. The description of the involved parameters, the comparison of different models, those models considered for "low dose" cases, as well as the possible data analysis, concerning the Statistical inference, has been extensively discussed by Wosniok et al. (1998).³⁴

A pioneering paper on quantitative model of carcinogenesis was that of Iverson and Arley,³⁵ who

Relevant Risk per case. In section 2 the development of the Probability models, at the early stages are discussed.

2. Probability Models in Carcinogenesis

Let D be the set of random variables, of the dose level of a carcinogen that induces a tumour in an individual agent and let T be the time at which an individual develops a tumour. Then, it is assumed that, an individual can develop tumour, at dose level d , say, at a particular time t , with probability $P(x, t)$, which can be represented either restricted on time or on dose as

assumed that the transformed cells are subject to a pure linear birth process and a clone of transformed cells is detected, if it exceeds a certain threshold. It has been noted this model could be modified in two different ways:

- If we assume that only one step is needed for the transformation of a normal to a malignant cell.
- To assume that a tumour arises from a single transformed cell.

In the former case we could assume that a number of these transformed cells must accumulate for a tumour to arise, while in latter a number (how many?) of steps are needed for a transformed cell to arise from a normal cell.

The improvement was from Nordling³⁶ who assumed that k specific mutational events have to occur, for a normal cell to transform into a malignant cell and called this the *multi-hit theory*. Armitage and Doll²¹ modified the multi-hit theory: they assumed that a certain sequence of irreversible cell alterations has to be followed. Moreover, the quantitative implication of this approach was really masterly investigated. The approach was called *multi-stage theory*. It was not only accepted but, eventually, found widespread application as the biological plausibility was combined with the applicability to

data sets (either from experimental or epidemiological studies).

Now, let us consider the case of a constant, continuously applied dose at level d . Moreover, the transformation rate from each stage, i , say, to the next one, $i + 1$, is assumed to increase linearly with the dose. In mathematical terms this is equivalent to: the transformation rate from the stage $i - 1$ to the next stage i is assumed to be equal to $a_i + b_i d$,

$$F(d, t) = 1 - \exp[-(a_1 + b_1 d)(a_2 + b_2 d) \cdots (a_k + b_k d)t^k / k!]. \quad (2.2)$$

As the main assumption was that the transformation rate, from each stage, to the next one is linear model, then (2.2) can be written as

$$G(d) = 1 - \exp[-(\theta_0 + \theta_1 d + \cdots + \theta_k d^k)] \quad (2.3)$$

where θ_i , $i = 0, 1, \dots, k$ are defined through the coefficients of the linear transformations assumed between stages, i.e. $\vartheta_i = \theta_i(t) = \alpha_i + \beta_i t$. Notice the biological inside on the coefficients of model (2.3), and therefore the value of k is essential: describes that the susceptible cell can be transformed through k distinct stages in order to be a malignant one as the multistage model describes the phenomenon, and not just an extension to a non-linear mathematical model – this is a crucial point for the Statistician to clarify. Same line of thought exists for the coefficients $\vartheta_i = \theta_i(t)$, which are function of time, which is not

$$\lambda(t) = c(t - t_0)^{k-1}, \quad c > 0, \quad k \geq 1, \quad \text{or} \quad \log \lambda(t) = \log c + (k - 1) \log(t - t_0) \quad (2.4)$$

where t_0 is fixed for the growth of tumour and k is the number of stages.

The above hazard function is the basis of the Armitage-Doll model, which may be considered biologically inappropriate for very old persons. The explanation is based on the fact that the very old cells lose their propensity to divide, and, therefore,

$$f(t) = c(t - t_0) \exp\left[-\frac{c}{k}(t - t_0)^k\right]. \quad (2.5)$$

The target of low dose exposure is to estimate effects of low exposure level of agents, known already and to identify if there are hazardous to human health. The interspecies extrapolation problem has been discussed by Luebeck et al.,³⁸ among others and is based mainly to the fact that the "body weight", say B , is related to the physiological parameter of

where $a_i > 0$ and $b_i \geq 0$. The parameter a_i presents the *spontaneous transformation rate*, in the absence of dosing (i.e. $d = 0$). Suppose that a tumour will develop before time t if all k transformations have occurred in sequence: the commutative probability density function – that is the time t is involved, and therefore mentioned in (2.2) model. Eventually the model describing the k -th change has occurred, is then given by

coming into light. Points like this bring Medicine-Biology, close to Statistics.

It was noticed that for some cancers (e.g. lung cancer) the cancer incidence rate increases with age, as *log* incidence is linear related to *log* age. Therefore, a *log – log* linear model describes the mechanism, Mc Cullagh and Nelder.³⁶ In case that the time independence is assumed and the mutation rates are supposed to be small the hazard function, Cox,²⁶ Kitsos²⁸ can be written as

are more refractive to new transformations. Thus, the "plateau" at older ages may simply reflect a compensating mechanism. Notice that to the Armitage-Doll hazard function corresponds to the density function

interest h exponentially as $h = pB^q$, with p and q parameters (very often q is reported around the value of $q \approx 0.75$). Therefore, the experimentation is based on animals and the results are transferred to humans. So, the target to calculate the probability of the occurrence of a tumour during the individual's lifetime is exposed to an agent of dose d during

lifetime is replacing humans with animals. Moreover, the idea of "tolerance dose distribution" was introduced to provide a statistical link and generate the class of dose risk functions.

Consider a nutshell - a tumour occurs at dose level $x = d$ if the individuals' resistance is broken at x , then the excess tumour risk is given by the Probability "model"

$$F(d) = P(X \leq d) = P(\text{Tolerance} \leq d)$$

The above "model" is actually the unknown cumulative distribution function that is eventually modelled, in the sense that it is assumed: there exists a statistical model which approximates $F(d)$, which acts as a cumulative distribution function. Then the dose level "d" is linked with the binary response (success or failure)

$$Y_i = \begin{cases} 1 & \text{success with probability } F(d) \\ 0 & \text{failure with probability } 1 - F(d) \end{cases}$$

Technically speaking the researcher only knows that the parameter vector comes from a subspace of the real numbers, $\theta \in \Theta \subseteq \mathcal{R}^p$, and we try to estimate as well as possible, different to be considered models, for different studies, see also Hartley and Sielken,³⁹ as far as the safe dose concerns.

When the proportion of "successes", as a response in a binary problem, is the proportion of experimental animals killed by various dose levels of a toxic substance, Finney,⁴⁰ called this experimental design *bioassay*.

When it is assumed that the cancer occurs when a portion of the tissue sustains a fixed number of "hits"; cancer is observed when the first such portion has sustained the required number of "hits", then the One-Hit model is considered. That is the carcinogenesis problem is formulated to quantify incidence data, then the *tolerance distribution models* are adopted. In this family of models, the parametric function is suggested as the distribution of the time.

The best known tolerance distribution is proposed by Finney,⁴⁰ in his early work, the *probit model* of the form

$$P(d) = \Phi(\mu + \sigma d),$$

with Φ being, as usual, the cumulative distribution function of the Normal distribution and μ and $\sigma > 0$ are location and scale parameters estimated from the data. Practically, the logarithm of dose is used that implies a *log* normal tolerance distribution.

The most commonly used parametric model for carcinogenesis is the Weibull model

$$P(T \leq t) = 1 - e^{-(\theta t)^\kappa}$$

With t being the time to create a tumour, and with hazard function.

$$\lambda(t) = \kappa \theta^\kappa t^{\kappa-1} \quad (2.6)$$

For the shape and scale parameters θ , $\kappa > 0$, respectively, it is assumed when the Weibull model can exhibit a dose-response relationship that is either sub-linear (shape parameter $\kappa > 1$) or supra-linear ($\kappa < 1$), and has a point of inflection at $x = ((\kappa - 1)/\kappa)^{1/\kappa}$.

The Weibull distribution is the fundamental distribution in Survival Analysis and is used in Reliability theory as a lifetime distribution. It is an extreme value distribution and is obtained as the distribution of the minimum of identical exponentially distributed random variables. Hence, if a "system" consists of independent components, each having identical exponential lifetimes, and if the system fails, whenever the first component reaches the end of its lifetime, then the lifetime of the system follows a Weibull distribution.

Due to its flexibility, the Weibull model is suited to describe incidence data as they arise in animal experiments and in epidemiological studies. It has been used for common parametrical analyses, e.g. comparison between experimental groups, which were treated with different doses of a carcinogenic substance.

The Weibull model has been extensively discussed for tumorigenic potency by Dewanji et al.,⁴¹ through the survival functions and the maximum likelihood estimators.

Thus, from the maximum likelihood function, from a censored sample the MLE of θ is

$$\theta^* = \left(\frac{d}{\sum t_i^\kappa} \right)^{1/\kappa}. \quad (2.7)$$

The second derivatives of the \log -likelihood l are

$$I_{\theta\theta} = -\frac{\kappa d}{\theta^2} - \kappa(\kappa - 1)\theta^{\kappa-2} \sum t_i^\kappa, \quad I_{\kappa\kappa} = -\frac{d}{\kappa^2} - \theta^\kappa \sum t_i^\kappa [\log(\theta t_i)]^2$$

$$I_{\theta\kappa} = \frac{d}{\theta} - \theta^{\kappa-1}(1 + \kappa \log \theta) \sum t_i^\kappa - \kappa\theta^{\kappa-1} \sum t_i^\kappa \log t_i,$$

Therefore, the 2×2 Fisher's information matrix, with diagonal elements $I_{\theta\theta}$, $I_{\kappa\kappa}$, can be evaluated and then the Variance-Covariance matrix is the inverse of Fisher's matrix.

Example 2.1: For a working example on the above, see Kitsos and Limakopoulou.⁴²

Example 2.2. Simulation Study for the One Hit Model.

The One Hit model was chosen for a particular study under the imposed frame work We face this

$$T(x; \theta) = P(y = 1|x, \theta) = \exp(-\theta x) = 1 - P(y = 0|u, \theta)$$

The corresponding log-likelihood function is

$$\ell(\theta) = -\sum x_i y_i - \sum (1 - y_i) [x_i \exp(-\theta x_i)] / [1 - \exp(-\theta x_i)]$$

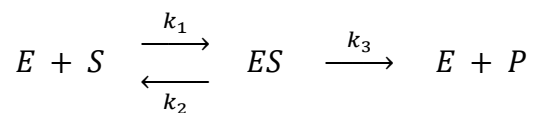
while the summation \sum runs from 1 to n . In probability terms the value $x_{opt} = 1.59/\theta$ corresponds to $p = 0.2$ – notice the dependence on the unknown parameter θ . For the binomial model with success probability $p = P(y = 1|x, \theta)$ it seems reasonable in practice to keep probability levels within the interval $[0.025, 0.975]$. Notice that the optimum design point depends on the unknown parameter, therefore a prior guess is needed. The unknown parameter is estimated sequentially.

3. On the Michaelis-Menten Model

A general theory for enzyme kinetics was firstly developed by Michaelis and Menten⁴⁴ in their pioneering work. This is discussed very briefly below: When an enzyme, say E , combines reversibly with a substrate, say S , to form an enzyme-substrate

problem as a sequential design, Kitsos⁴³ with equal and unequal batches. Practically the definition of the design space means "where the input variable" for example dose level, takes values, the range of the dose level, while the parameter space means where the parameter lies. Then the Probability model describing this assessment, depending on the unknown parameter we try to estimate is,

complex, say ES , which can be dissociate or proceed to the product, say P , the following scheme is assumed



with k_1, k_2, k_3 the associated rate constants. We let $K_M = \frac{k_2 + k_3}{k_1}$, the Michaelis-Menten (MM) constant, $V_{max} = k_3 C_{TOT}$, C_{TOT} = the total enzyme concentration. In principle K_M is the concentration at one half V_{max} , with V_{max} being the maximum metabolic rate constant. As far as the interspecies extrapolation concerns the MM constant is assumed to remain constant. Then a plot of the initial velocity of reaction u , against the concentration of substrate, C_S , will provide the MM rectangular hyperbola of the form

$$u = \frac{V_{max}C_S}{K_M + C_S} \quad (3.1)$$

Notice that in practice we only have readings of the form (3.1) with $f(x, \theta) = u$ being a non-linear function as above with the parameter vector to be $\theta = (V_{max}, K_M)$, $x = C_S$. The deterministic relation (3.1) is linked with the experimental error. And in practice only readings for the stochastic non-linear model of the form $y_i = u_i + e_i$, $i = 1, 2, \dots, n$ is obtained. That is, readings y_i are associated with noise, with mean zero and variance constant, with the Normality assumption valid when inference is needed. For the optimal design consideration and evaluation of the parameters in Michaelis-Menten Model, see the early work of Endrenyi and Chan,⁴⁵ while Currie⁴⁶ have discussed the heteroscedasticity

$$(\theta - \hat{\theta})^T I(\hat{\theta}, \xi)(\theta - \hat{\theta}) \leq Bps^2 F(a; p, n - p) \quad (3.3)$$

When $B = 1$, a linear approximation is considered. As usually s^2 is a suitable estimator of σ^2 i.e. $s^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2$.

There are two lines of thought to approach the MM model as an optimal experimental design:

1. From the biological point of view, which eventually ensures: enzymatic process plays an import role in practice. In cancer studies, the question is whether enzymatic induced interactions have an influence on the production of a carcinogen. This is strongly related with the estimation of V_{max} and K_M , with such attempts give more emphasis to practical problems, Currie,⁴⁶ Endrenyi and Chan,⁴⁵ Gilberg et al..⁴⁷
2. The statistical point of view includes Michaelis Menten within the class of non-linear models and uses the model within this theoretical framework, Kitsos⁵⁰ among others.

Two more very crucial points are also mentioned namely: that this design is optimum for estimating the ratio V_{max}/K_M and that the variance σ^2 is not constant.

Endrenyi and Chan⁴⁵ discussed and obtained the D -optimal design points, without the minimization

in the Michaelis-Menden Model and Gilberg et al.⁴⁷ provides an extensive discussion, while for all the possible ways the experimentalist wishes to fit the model, see Toulas and Kitsos.⁴⁸ As far as the construction of the confidence intervals, here a special consideration is needed. An extra statistic is defined: the supremum value of Beale's measure of nonlinearity, Bates and Watts,⁴⁹ for models with two or more parameters, can be reduced to

$$B = 1 + \frac{n}{n-2} \frac{1}{\sqrt{F}} \quad (3.2)$$

with F being the F distribution with the appropriate degrees of freedom (df). Therefore, the approximate confidence region for the MM model is

of the information matrix. They worked out, empirically, different designs to calculate efficiencies and claim calculations for 100% or 95% efficient designs. The Optimal Design theory can be applied in experimental Carcinogenesis, Kitsos,^{51,19} while the Metabolizing Enzymes for Lung Cancer Risk Factors have been discussed by Risch et al..⁵²

It is clear that in enzyme kinetic studies, with constant variance the one design point might be at the highest possible concentration, and the second at the maximal feasible velocity. The main result, Kitsos⁵⁰ for the Michaelis - Menten Model is that the D -optimal design for the MM model as in (3.1) does not depend on V_{max} , the "linearly contributed" parameter. That practical means that only for one parameter you need prior information, and the D -optimal design for the extended Michaelis - Menten model of the form

$$u = \frac{V_{max}C_S}{K_M + C_S} + \theta_0 C_S$$

for estimating $\theta = (V_{max}, K_M, \theta_0)$ does not depend on V_{max} and θ_0 , which can be considered as the linear metabolic rate constant. Moreover, the D_S -optimal design for estimating θ_0 , depends only on K_M . This comes as a result that an extra linear term

does not influence the MM optimal experimental design, Kitsos,⁵⁰ while if $U \gg K_0$ (U is too large comparing K_0) the locally D -optimal design (the one which allocates half observations at two contributed optimal design points) allocates half points at U (the maximum value of the concentration of substrate) and K_0 (the initial guess for the MM constant). Moreover, it is Optimum value of $C_S \cong K_0$.

Notice that the D -optimal designs as above for estimating the parameter vector (V_{max}, K_M) are the D -optimal designs for estimating the ratio $\varphi = V_{max}/K_M$. This result is really helpful for the experimentalist, as to get a ratio estimates is a difficult task in Statistics.

This result is Statistically essential, as there is, in principle, a difficulty on ratio estimates.

The problem is always the initial guess. Therefore, we propose two steps of calculations:

A1: Devote a proportion p for your n observations at first stage i.e. allocate $\frac{np}{2}$ observations at U and K_0 , or at U and Optimum C_S . Get the estimates. Use these estimates to feed the next step A2.

A2: Perform sequentially $\frac{(1-p)n}{2}$ more runs OR a second static design with the estimates obtained at stage A1 (this case is known as *two stage design*).

The MMM is rather a Statistical model, than a Probability one, but still within the useful models facing the Cancer problem. More elegant, mathematically, models can be constructed and are discussed below.

4. Advanced Probability Models

The *mechanistic models* have been named so, because they are based on the presumed mechanism of the carcinogenesis and they form a particular class of models. The main and typical models of this class are the dose response models, used in Risk Assessment, based on the following main characteristics:

- There is no threshold dose below which the carcinogenic effect will not occur.
- Carcinogenic effects of chemicals are induced proportionally to dose (target tissue concentration) at low dose levels.
- A tumour originates from a single cell that has been damaged by one of the two reasons: either the parent compound or one of its metabolites.

The mechanistic class of models can be subdivided into those models that describe the process on the level of the organism or on the level of individual cells, see section 1. The following subclasses are referred.

The sub-class of *Global models*, includes those models that on the level of the whole organism, are closely related to the introduced Probability Models for Cancer, as also describe the time to detectable carcinoma. Typical example is, the *cumulative damage model* is motivated by reliability theory. This model considers that the environmental factors cause damage to a system, which although does not fail "immediately", eventually fails whenever the accumulating damage exceeds a threshold. The model adopts a Poisson process generating time points in which damage occurs.

The second sub-class of the class of Mechanistic Models is the *Cell Kinetic models*. They attempt to incorporate a number of biological theories, based on the line of thought that the process of carcinogenesis is on the cellular level. There has been a common understanding among biologists that the process of carcinogenesis involves several biological phenomena including mutations and replication of altered cells.

Cell kinetic models are subdivided by the method, which is used for their analysis, in Multistage Models and Cell Interaction Models, as follows.

The main class of cell kinetic models comprises *multistage models*, which describe the fate of single cells, but does not take into account interactions

between cells. In these models, mutations and cell divisions are described. These models stay analytically tractable because cells are assumed to act independently to each other. The variables of interest can be derived explicitly, and hence the usual statistical techniques can be used to apply these models to data. The pioneering work of Moolgavkar and Venzon¹¹ formulates a two - stage model with stochastic clonal expansion of both normal and intermediate cells and they introduced a mathematical technique to analyse a two - stage model with deterministic growth of normal cells and stochastic growth of intermediate cells. Moreover, Moolgavkar and Knudson¹² showed how to apply this latter model to data from epidemiological studies.

Working on incidence data from animal experiments and epidemiological studies, Kopp – Schneider,¹⁶ is providing the appropriate definitions and understanding for the underlying probabilistic mechanism model for applications.

The second sub-class of the cell kinetic models is the *Cell interaction models* incorporating both the geometrical structures of the tissue and communication between cells are too complicated for analytical results. These models aim to describe the behaviour (described by a number of simulations) of complex tissues in order to test biological hypotheses about the mechanism of carcinoma formation.

The *Generalised Multistage Models* (GMS) or *Cell Interaction Models* and the Moolgavkar-Venzon-Knudson Models (MVK) are based on the following assumptions:

- Carcinogenesis is a stochastic multistage process on cell level.
- Transition between stages is caused by an external carcinogen, but it may also occur spontaneously.

Cell death and division is important in MVK models. The normal, intermediate and malignant cells are depending on time. Intermediate cells arise from a

normal cell due to a Poisson process with known rate. A single intermediate cell may die with rate β , divide into two intermediate cells with rate α , or divide into one intermediate and one malignant cell with rate μ . Therefore, the process gives rise to the three steps:

- Initiation, promotion, progression which are very different from the biological point of view. If an agent increases the net cell proliferation $\alpha - \beta$, the cancer risk will also increase, Leubeck et al.,³⁸ Leubeck and Moolgavkar.¹⁵

These cell kinetic models were used to describe the time to tumour as a function of exposure to a carcinogenic agent. Two objectives guide this research. On one hand, the models can be used to investigate the mechanism of tumour formation by testing biological hypotheses that are incorporated into the models. On the other hand, they are used to describe the dose-response relationship for carcinogens.

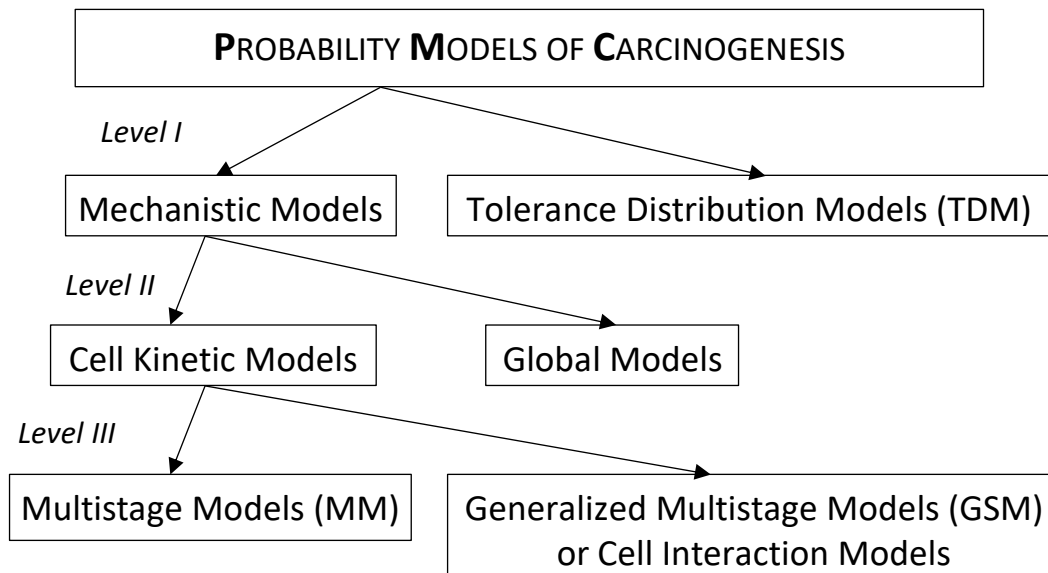


Figure 4.1: A classification of the stochastic models of carcinogenesis.

The compact presentation and discrimination in Figure 4.1 has been taken place at first level according to the intention, at second level according to biological detail and at third level according to the desired method, and is based on the above discussion.

5. Theoretical Implications in Practice

There is a need for calculating *low level percentiles*, Kitsos,⁵³ working for low dose extrapolation problems. For practical reasons the calculation should be simple such that it can be adopted easily by the experimentalist, while at the same time, this

$$F(L_p) = 1 - \exp(-(\theta L_p)^\kappa) \Rightarrow L_p = \left(-\frac{1}{\theta^\kappa} \ln(1-p)\right)^{1/\kappa} \quad (5.1)$$

This result can be generalized. It has been proved, Kitsos,⁵⁵ that within the class of Multistage Models there exists an iterative scheme which converges to the percentile L_p . Such a need is essential when p is very small and can always be evaluated sequentially, rather than working with a static design, as the estimated percentile is re-estimated.⁵⁶

Another point of interest, with a strong theoretical background, is *mixtures*. Humans are exposed to a large number of chemicals from a variety of different sources, either sequentially or static/concurrently. We would like to clarify that the mixtures of chemicals are ubiquitous:

derivation has to be optimal in the statistical sense. One of the targets of this paper has been to refer on the calculation of the low – dose percentiles by adopting the sequential principle of design, Hu.⁵⁴ The sequential approach can be faced easier through the stochastic approximation scheme. This is discussed below in comparison with a static design approach: i.e., perform once an experiment.

Example 5.1 Recall the One Hit model $F(x) = 1 - \exp(-\theta x), \theta > 0, x \geq 0$.

According to the definition of the percentile point it is

1. in the *air* the animals and humans are breathing
2. the *food* that all species are eating
3. the *water* that all species and especially humans are drinking.

Therefore, an extended analysis to the involved chemical mixtures is needed to provide a further analysis on Cancer Risk Assessment, Kitsos and Edler,⁵⁷ who worked on the different Statistical models concerning the mixtures and the corresponding Geometrical presentation, to clarify the situation to the experimenter.

A simple chemical mixture consists of a composition of chemicals. In practice this composition consists of no more than ten chemicals. The qualitatively and quantitatively synthesis is supposed to be known, in any case, that means that even if it is not known has to be investigated and analysed. The various combination of chemicals can eventually affect on either *different target organs or the same target organ*.

The above two groups can be combined with the two assumptions on the action of these chemicals. This action is consisted of chemicals that can be either with *different mode of action or with the same mode of action*.

The main terms in a biological mixture bioassay, when a mixture experimental approach is adopted for the biological/toxicological problem, under investigation, has been developed by Hodgson and Levi,⁵⁸ Kitsos and Edler⁵⁷:

- *Synergism*: Both involve toxicity greater than would be expected from the toxicities of the compounds administrated separately, by in the case of synergism one compound has little or no intrinsic toxicity administered alone, where in the case of
- *Potentiation*: both compounds have appreciable toxicity when administered alone.

As a toxicological interaction, the National Research Council (NCR),²⁰ defines a circumstance in which exposure to two or more chemicals results in a qualitatively or quantitatively altered biological response relative to that predicted from the actions of a single chemical.

$$\log\{p(x)/(1 - p(x))\} = x^T = \beta_0 + \beta_1x_1 + \dots + \beta_px_p,$$

Hosmer and Lemeshow,⁶¹ Rao and Toutenburg,⁶² we can evaluate that

$$p(x) = \text{Logit}^{-1}(x^T\beta)$$

which remains invariant under affine transformations,⁶⁶ while for a development of the Generalized Linear Models see Collet,⁶⁷ Cox and Snell.²⁷ A very similar

In principle the investigator is interested to get more information and they are usually constructed in such a way, that they extract as much information about the process as is possible, following a certain criterion. An "optimal experimental design" is that design which estimates the involved parameters "as well as possible", with the most popular criterion being to have the estimators the smallest variance, i.e., the smallest possible error – this is the *D*-optimal criterion.

There are different types of optimal experimental design for different targets. The main ones are Kitsos and Edler⁵⁷:

- ❖ 2^k Factorial Design, 2^{k-q} fractional factorial, Rotatable, Simplex,
- ❖ Optimal designs (e.g., *D*-optimal etc.) and other.

Another theoretical issue, which it is not considered, unfortunately, is the *effect of Covariates to prognosis*, Prentice and Kalbfleisch,⁵⁹ Kitsos,⁶⁰ Begg and Legakos,⁶¹ among others. The adoption of the covariate idea to the development of tumour, needs a careful study and it is beyond the target of this paper.

In this paper, the *logit model* is adopted to estimate the relative risk, through the relative risk. The logit model has been suggested, since the pioneering paper of Berkson.⁶² That is the "log odds" have been assumed linear in terms that considering a subject with attributes, given by the input vector $X = (X_1, \dots, X_p)^T$, Breslow and Day,⁶³ Cox and Snell,²⁷ then

approach had been adopted by Bliss⁶⁸ in his pioneering paper of Risk Analysis, were counting the number of dead insects after being exposed to CS_2 .

Usually, interest is focused on the Relative Risk (RR), with $RR = \exp(\beta_1)$ and we test the null hypothesis $H_0: RR = 1$ vs the alternative $H_1: RR \neq 1$. The Chi-

square test, Lemeshow and Hosmer⁶⁹ among others, is very useful, when a "large" number of observations are considered, while the hazard function has been also proved useful to the experimenter, especially when the time is part of the whole study. Although the Normal distribution has been generalized and the hazard function for the Generalized Normal has been evaluated, Toulas and Kitsos,⁷⁰ the researcher still needs the classic hazard function, as it is easier to be evaluated and there are not Mathematics associated with it.

Example 5.2. Some of the markers related to the breast cancer are widely monitored and the relevant tests are routinely performed on patient samples, with the following being the most popular: Estrogen Receptors (ER), Progesterone Receptors (PR), HER-2, p53, S phase. Many gene polymorphisms in the metabolism of breast cancer have been described as possible neoplasm etiologic factors, Bugano et al..⁷¹ Regarding the breast cancer risk assessment, susceptibility, and its relation to CYP17, MspA1 polymorphism, different opinions have been expressed, Feigelson et al.,⁷² while Huang et al.⁷³ found a positive association between the breast cancer relative risk and the individual susceptibility genotypes. See Kitsos⁷⁴ for the details of this study, while the method was extended for the Generalized Normal distribution and Information criteria providing an appropriate software, Kitsos and Toulas.⁷⁵

Therefore, a study, was performed referring to 98 breast cancer patients and 125 healthy controls where they were compared considering the age at menarche, age at menopause, the number of full-term pregnancies and the CYP17, COMT genotypes. The frequency of CYP17 A1/A1 genotype was compared to A1/A2 and A2/A2, whereas the frequency of COMT G/G genotype was compared to G/A and A/A. The result was that the full pregnancy with Relative Risk, $RR = 1.42$ and Menopause with $RR = 1.04$ ($p < 0.05$) influence the final RR .

Interpretation of the coefficients of the statistically significant variables, provide evidence, that on the basis of this study, when the age at menopause

increases one year the probability of breast cancer increases 4%.

Moreover, women with full time pregnancy have 42% less probability for breast cancer than the other women. In that point it is useful to remark that from the interpretation of the coefficient of the variable age of menarche from the full model we have that when the age of menarche increases one year then the probability of cancer decreases 5%. There is a strong interest on the subject from the statistical point of view, see Duffy,⁷⁶ Prentice and Gloeckler,⁷⁷ among others.

6. Conclusion

It has mentioned that tumour markers (usually proteins associated with a malignancy) might be clinically usable in patients with cancer, Cheung et al.,⁷⁸ Amaral-Mendes and Plugers.⁷⁹ A tumour marker can be detected in a solid tumour, in circulating tumour cells in peripheral blood, in lymph nodes, in bone marrow or in other body fluids. As the problem of Carcinogenesis is so important a number of models have been already reviewed, Vineis et al.,⁸⁰ while others provide new ones, Gatenby and Vincent,⁸¹ as it is expected due to the evolution.

The aim of this paper has been to offer Statistical methods and discussion either for experimental carcinogenesis problems or for real life Cancer problems. The former needs extrapolation to humans or interspecies extrapolation, Travis et al.,⁸² while the latter is a real problem with the most references in the Science, Edler and Kitsos,¹⁰ Crump et al..⁸³ Different accidents, the main one being the Chernobyl, Baverstock and Williams,⁸⁴ among others, provided less information than it was expected, due to the problems to collect the appropriate data. As the Risk Assessment of Cancer has studied in detail from the medical, Ladenson,⁸⁵ Kawai et al.,⁸⁶ Hayat, et al.,⁶ toxicological studies, Bowman et al.,⁸⁷ and biological points of view, Arden et al.⁸⁸ the main objective of this work has been to provide an insight into this problem from the Statistical point of view proposing a sequential bioassay for the estimation

of low – dose exposure, i.e. low – dose percentiles. The optimal experiment design has been adopted and the sequential principle has been considered. Static designs, where all the observations are used once have the disadvantage that a costly experiment might be performed and the acquired estimator might be far from the "true" value. The effect of covariates in Cancer problems, it is always a useful idea to proceed, Petersen,⁸⁹ Kitsos.⁶⁰

The Probability Models it is not related to epidemiological studies, Ferlay et al.,⁹⁰ Horn-Ross et al.,⁹¹ Kafadar and Tukey,⁹² among others. These studies can be helpful in constructing statistical parameters, and being helpful in Risk Analysis studies. We strongly encourage the development and use of such models trying to explain the underlying mechanics through Statistical modeling. It is better to approach the situation with an error than to have no model to describe the phenomenon. In the development we attempted, mainly based on our research work, we believe that Luebeck et al.,³⁸ Montie and Meyers⁹³ are among those who tackled the real problem, while the majority of Statistical work offer ideas for facing the problem, as it happens to Risk Analysis which oscillates between practice and theory, Kitsos.¹⁸ But still we believe that a simple geometrical figure clarifies the situation, providing food for investigating the Statistics behind. It is a real need the Statistical coverage, usually through an appropriate, and assumed correct, model. There are a number of theoretical techniques, trying to support the research on this kind of Bioassays.

The affine Geometry approach for the invariance of the logistic model, Kitsos,⁵⁶ can be useful on defining the appropriate transportation from animals to humans (usually it is assumed the transformation: the body weight to the power around the value of 0.74), while Hermann et al.³¹ provided more Mathematics to study the surface of a tumour. Statistics is the right hand of Sciences, so we believe it can proved herself useful facing Cancer problems, communicating with Medicine and Biology. The advice "Keep It Simple", Kitsos,^{28,18} is depending on

the definition of "simple", which is a function of time. What is "simple" today was not 50 years ago, with typical examples being the conic sections (of Apollonius), which were waiting for centuries Kepler to adopt them. Now ellipse, parabola etc. are simple questions at high schools. Therefore, it needs determination, and capable communication skills, to adopt the Probability models of today to study, in a team work, the evolution of the cancer problems.

Conflict of Interest:

None

Acknowledgments:

CPK would like to thank the CCMS/NATO pilot study for the generous grand, on carcinogenesis, for the time period 1990-2008 and the participants for the nice discussions. The comments of the referees are very much appreciated and help us to promote the final form of this paper.

References:

1. Baish WJ, Jain KR. Fractals and Cancer. *Cancer Research*, 2000; 60:3683-3688.
2. Tan WY. *Stochastic Models of Carcinogenesis*. Marcel-Dekker, N.Y.;1991
3. Wosniok W, Kitsos C, Watanabe K. Statistical issues in the Application of Multistage and Biologically Based Models. In: Cogliano V, Luebeck G, Zapponi G. eds. *Respectives on Biologically Based Cancer Risk Assessment*. NATO-Challenges of Modern Society. 1998; Vol. 23:243-272.
4. Zapponi, G. A. Carcinogenetic Risk Assessment: Some Points of Interest for a Discussion. In: Chyczewski L, Niklinski J, Plugers E. eds. *Endocrine Disruptors and Carcinogenic Risk Assessment*. IOS press. 2002; 15-27.
5. Bernal M, Chalikias MS, Kitsos CP. Analysing data set on Thyroid Cancer. In: *2d International Conference on Cancer Risk Assessment (ICCRA2, e-proceedings)*. 2007; Santorini, 25-27 May 2007.
6. Hayat MJ, Howlader N, Reichman ME, Edwards BK. Cancer statistics, trends, and multiple primary cancer analyses from the Surveillance, Epidemiology, and End Results (SEER). *Program Oncologist*. 2007; 12:20-37.
7. Angelopoulou R, Bala M, Lavranos G, Chalikias M, Kitsos C, Baka S, Kittas C. Evaluation of immunohistochemical markers of germ cells' proliferation. In the developing rat testis: A comparative study. *Tissue and Cell*. 2008; 40(1):43-50
8. Cogliano VJ, Luebeck EG, Zapponi G. Perspectives on Biologically-Based Cancer Risk Assessment eds. NATO Challenges of Modern Society. Kluwer Academic/Plenum Pub. 1999; Vol 23.
9. Chyczewski L, Niklinski J, Pluygers E. Endocrine Disruptors and Carcinogenic Risk Assessment. NATO Science Series I. IOS Press. Amsterdam; 2002; Vol 340.
10. Edler L, Kitsos, CP. *Recent Advances in Qualitative Methods in Cancer and Human Health Risk Assessment*. Editors, Wiley, UK; 2005.
11. Moolgavkar SH, Venzon D. Two-Event Modls for Carcinogenesis: Incidence Cures for Childhood and Adult Tumors. *Mathem. Biosciences*. 1979; 47:55-77.
12. Moolgavkar SH, Knudson A. Mutation and cancer: A model for human carcinogenesis. *Journal of the National Cancer Institute*. 1981; 66:1037-1052.
13. Luebeck GE, Moolgavkar SH. Two-Event Model for Carcinogenesis: Biological, Mathematical and Statistical Considerations. *Risk Analysis*. 1989; 10:323-341.
14. Luebeck GE, Moolgavkar SH. Stochastic Description of Initiation and Promotion in Experimental Carcinogenesis. *Ann. Ist. Super.Sanita*. 1991; 27(4):575-580.
15. Luebeck GE, Moolgavkar SH. Multistage Carcinogenesis: Population-Based Model for Colon Cancer. *Journal of National Institute*. 1992; 610-618.
16. Kopp-Schneider A. Carcinogenesis models for risk assessment. *Stat. Methods in Medical Research*. 1997; 6:317-340.
17. Muller CH, Kitsos CP. Optimal Design Criteria Based on Tolerance Regions. In: Di Bucchianno A, Lauter H, Wynn H, eds. *MODA7 – Advances in Model-Oriented Design and Analysis*. Physica-Verlag. 2004; 107-115.
18. Kitsos CP. Risk Analysis in Practice and Theory. In: Kitsos CP, Oliveira TA, Pierri F, Restaino MR, eds. *Statistical Modelling and Risk Analysis*. Springer. 2023; 107-118.
19. Kitsos CP. Optimal Design for Bioassays in Carcinogenesis. In: Edler L, Kitsos CP, eds. *Quantitative Methods for Cancer and Human Health Risk Assessment*. Wiley, England. 2005; 267-279.
20. National Research Council (NCR). *Principles of Toxicological Interactions Associated with Multiple Chemical Exposures*. National Academic Press. Washington, DC. 1980.
21. Armitage P, Doll R. The Age Distribution of Cancer and a Multi-Stage Theory of Carcinogenesis. *Brit. J. Cancer*. 1954; 8:1-12.

22. Armitage P. The Assessment of Low Dose Carcinogenicity. *Biometrics*. 1982; 28 (sup.):119-129.
23. Armitage P. Multistage Models of Carcinogenesis. *Environmental Health Perspectives*. 1985; 63:195-201.
24. Doll R. The Age Distribution of Cancer: Implications for Models of Carcinogenesis. *JRSS*. 1971; A134:133-166.
25. Doll R. An Epidemiological Perspective on the Biology of Cancer. *Cancer Res*. 1978; 38:3573-3583.
26. Cox DR. Regression Models and Life Tables (with discussion). *JRSS*. 1972; B, 74:187-220.
27. Cox DR, Snell EJ. *Analysis of binary data*. Second Edition, Chapman and Hall. 1989.
28. Kitsos CP. Sir David Cox: A wise and noble statistician (1924–2022). *Eur. Math. Soc. Mag*. 2022; 27 – 32.
29. Stehlik M, Hermann P, Giebel S, Schenk JP. Multifractal Analysis on Cancer Risk. In: Oliveira TA, Kitsos CP, Oliveira A, Grilo L, eds. *Recent Studies on Risk Analysis and Statistical Modeling*. Springer. 2018; 17-34.
30. Losa GA, Nonnenmacher TF. *Fractals in biology and medicine*. Springer. 2005.
31. Hermann P, Piza S, Ruderstorfer S, Spreitzer S, Stehlik M. Fractal Case Study for Mammary Cancer: Analysis of Interobserver Variability. In: Kitsos CP, Oliveira TA, Rigas A, Gulati S, eds. *Theory and Practice of Risk Assessment*. Springer. 2015; 21-36.
32. US EPA. *Guidelines for exposure assessment*. Federal Register. 1992; 51:33992-34003.
33. US EPA. *Guidelines for Carcinogenic Risk Assessment*. EPA, Washington, DC. 1999.
34. Wosniok W, Kitsos CP, Watanabe K. Statistical issues in the Application of Multistage and Biologically Based Models. In: Cogliano V, Luebeck G, Zapponi G. *Respectives on Biologically Based Cancer Risk Assessment*. NATO-Challenges of Modern Society. 1998; Vol. 23:243-272.
35. Iverson S, Arley N. On the mechanism of experimental carcinogenesis. 1950; 27:773-803.
36. Nordling CO. A new theory of the cancer inducing mechanism. *British J. of Ca*. 1953; 7:78-72.
37. McCullagh P, Nelder JA. *Generalized Linear Models*. Chapman and Hall, London. 1989.
38. Luebeck EG, Watanabe K, Travis C. Biologically based models of carcinogenesis. In: Cogliano VJ, Luebeck EG, Zapponi GA, eds. *Prospectives on Biologically Based Cancer Risk Assessment*. Kluwer Academic/Plenum Publishers. New York. 1999; 205-241.
39. Hartley HO, Sielken RL. Estimation of 'Safe Dose' in Carcinogenic Experiments. *Biometrics*. 1977; 33:1-30.
40. Finney DJ. *Probit Analysis, 3d ed*. Cambridge Un. Press. 1971.
41. Dewanji A, Venzon DJ, Moolgavkar SH. A stochastic two-stage model for cancer risk assessment: II. The number and size of premalignant clones. *Risk Analysis*. 1989; 9:179-187.
42. Kitsos CP, Limakopoulou A. Optimal Risk Assessment for Estimating the VSD in Experimental Carcinogenesis. In: A. Hasman et al., eds. *Medical Infobank for Europe*. IOS Press. 2000; 763 – 766.
43. Kitsos CP. *Cancer Bioassays: A Statistical Approach*. Lampert. 2012.
44. Michaelis L, Menten ML. Kinetics for Intertase action. *Biochemische Zeiturg*. 1913; 49:333-369.
45. Endrenyi L, Chan FY. Optimal Design of Experiments for the Estimation of Precise Hyperbolic Kinetic and Binding Parameters. *J. Theor. Biol*. 1981; 90:241-263.
46. Currie DJ. Estimating Michaelis-Menten Parameters: Bias, Variance and Experimental design. *Biometrics*. 1982; 38:907-919.
47. Gilberg F, Urfer W, Elder L. Heteroscedastic Nonlinear Regression Models with Random Effects and their Application to Enzyme Kinetic Data. *Biometrical Journal*. 1999; 41:543-557.
48. Toulas TL, Kitsos CP. Fitting the Michaelis-Menten Model. *Journal of Comp. and Appl. Math*. 2016; 296:303-319.

49. Bates DM, Watts DG. *Nonlinear-Regression Analysis and its Applications*. Oliver and Boyd, Edinburgh. 1998.
50. Kitsos CP. Design aspects for the Michaelis – Menten Model. *Biometrical Letters*. 2001; 38:53-66.
51. Kitsos CP. The Cancer Risk Assessment as an Experimental Design. In: Chyczewski L, Niklinski J, Pluygers E, eds. *Endocrine Disruptors and Carcinogenic Risk Assessment*. IOS Press. 2002; 37:329-337.
52. Risch A, Dally H, Edler L. Genetic Polymorphisms in Metabolising Enzymes as Lung Cancer Risk Factors. In: Edler L, Kitsos CP, eds. *Recent Advances in Quantitative Methods in Cancer and Human Risk Assessment*. Wiley, UK. 2005.
53. Kitsos CP. Optimal Designs for Estimating the Percentiles of the Risk in Multistage Models in Carcinogenesis. *Biometrical Journal*. 1999; 41, No 1:33-43.
54. Hu I. On sequential designs in nonlinear problems. *Biometrika*. 1998; 85:496-503.
55. Kitsos CP. Optimal Designs for Percentiles at Multistage Models in Carcinogenesis. *Biometrical Journal*. 1997; 41, No 1:33-43.
56. Kitsos CP. Sequential Approaches for Ca Tolerance Models. *Biometrie und Medizinische Informatik Greifswalder Seminarberichte*. 2011; Heft 18:87-98.
57. Kitsos CP, Edler L. Cancer Risk assessment mixtures. In: Edler L, Kitsos CP, eds. *Quantitative Methods for Cancer and Human Health Risk Assessment*. Wiley, England. 2005; 283-298.
58. Hodgson E, Levi EP. *A textbook of modern toxicology*. Elsevier, New York. 1987.
59. Prentice RL, Kalbfleisch JD. Hazard Rate Models with Covariates. *Biometrics*. 1979; 25-39.
60. Kitsos CP. The Role of Covariates in Experimental Carcinogenesis. *Biometrical Letters*. 1998; Vol. 35, No 2:95-106.
61. Begg MD, Legakos S. Loss in efficiency caused by omitting covariates and misspecifying exposure in logistic regression models. *JASA*. 1993; 88:166-170.
62. Berkson J. Maximum Likelihood and Minimum X² estimates of the logistic function. *JASA*. 1955; 50:130-162.
63. Breslow NE, Day NE. *Statistical Methods on Cancer-Research*. IARC. Lyon, France. 1980; No 32.
64. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. John Wiley. 1989.
65. Rao CR, Toutenburg H. *Linear Models, 2nd ed*. Springer-Verlag, N. Y. 1999.
66. Kitsos CP. On the Logit Methods for Ca Problems. In: Vonta F, ed. *Statistical Methods for Biomedical and Technical Systems*. Limassol, Cyprus. 2006; 335-340
67. Collet D. *Modelling Binary Data*. Chapman and Hall/CRC. 1999.
68. Bliss CI. The calculation of the dosage mortality curve. *Annals of Applied Biology*. 1935; 22:134.
69. Lemeshow S, Hosmer DW. The use of goodness-of fit statistics in the development of logistic regression models. *American Journal of Epidemiology*. 1982; 115:92-106.
70. Toulas TL, Kitsos CP. Hazard Rate and Future Lifetime for the Generalized Normal Distribution. In: Oliveira TA, Kitsos CP, Oliveira A, Grilo L, eds. *Recent Studies on Risk Analysis and Statistical Modeling*. Springer. 2018; 165-180.
71. Bugano DD, Conforti-Froes N, Yamaguchi NH, Baracat EC. Genetic polymorphisms, the metabolism of oestrogens and breast cancer: A review. *European Journal Gynaecological Oncology*. 2008; 29(4):313-20.
72. Feigelson SH, McKean-Cowdlin R, Henderson EB. Concerning the CYP MspA1 and breast cancer risk: a meta-analysis. *Mutagenesis*. 2002; 17:445-446.
73. Huang CS, Chern HD, Chang KJ, Cheng CW, Hsu SM, Shen CY. Breast Cancer Risk Associated with Genotype Polymorphism of the

- Oestrogen-metabolizing CYP17, CYP1A1 and COMT: A Multigenic Study on Cancer Susceptibility. *Cancer Research*. 1999; 59:4870-4875.
74. Kitsos CP. Estimating the Relative Risk for the Breast Cancer. *Biometrical Letters*. 2010; Vol 47(2):133-146.
75. Kitsos CP, Toulas TL. Generalized Information Criteria for the Best Logit Model. In: Kitsos CP, Oliveira TA, Rigas A, Gulati S, eds. *Theory and Practice of Risk Assessment*. Springer. 2015; 3-20.
76. Duffy MJ. CA 15-3 and related mucins as circulating markers in breast cancer. *Ann Clin Biochem*. 1999; 36:579-86.
77. Prentice RL, Gloeckler LA. Regression analysis of grouped survival data with application to breast cancer data. *Biometrics*. 1978; 34:56-67.
78. Cheung K, Graves CRL, Robertson JFR. Tumour marker measurements in the diagnosis and monitoring of breast cancer. *Cancer Treat Review*. 2000; 26:91-102.
79. Amaral-Mendes JJ, Pluygers E. Use of Biochemical and Molecular Biomarkers for Cancer Risk Assessment in Humans. In: Cogliano V, Luebeck G, Zapponi G. *Respectives on Biologically Based Cancer Risk Assessment*. NATO-Challenges of Modern Society. 1999; Vol. 23:81-152.
80. Vineis P, Schatzkin A, Potter JD. Models of carcinogenesis: an overview. *Carcinogenesis*. 2010; 31(10):1703-1709.
81. Gatenby AR, Vincent LT. An Evolutionary Model of Carcinogenesis. *Cancer Research*. 2003; 63:6212- 6220.
82. Travis CC, White RK, Ward RC. Interspecies Extrapolation of Pharmacokinetics. *J. Theor. Biol*. 1990; 42:285-304.
83. Crump KS, Guess HA, Deal KL. Confidence Intervals and Tests of Hypotheses Concerning Dose Response Relations Inferred from Animal Carcinogenicity Data. *Biometrics*. 1977; 33:437-451.
84. Baverstock K, Williams D. The Chernobyl accident 20 years on: an assessment of the health consequences and the international response. *Environ Health Perspect*. 2006; 114(9):1312-1317.
85. Ladenson PW. Optimal laboratory testing for diagnosis and monitoring of thyroid nodules, goiter, and thyroid cancer. *Clinical chemistry*. 1996; 42(1):183-187.
86. Kawai R, Mathew D, Tanaka C, Rowland M. Physiologically based pharmacokinetics of cyclosporine A: extension to tissue distribution kinetics in rats and scale-up to human. *Journal of Pharmacology and Experimental Therapeutics*. 1998; 287(2):457-468.
87. Bowman D, Chen JJ, George EO. Estimating Variance Function in Developmental Toxicity Studies. *Biometrics*. 1995; 51:1523-1528.
88. Arden KC, Anderson MJ, Finckenstein FG, Czekay S, Cavenee WK. Detection of the t(2:13) chromosomal translocation in alveolar rhabdomyosarcomah using the reverse transcriptase polymerase chain reaction. *Genes Chrom Cancer*. 1996; 16:254-260.
89. Petersen T. Fitting Parametric Survival Models with Time-Dependent Covariates. *Appl. Statist*. 1986; 35:281-288.
90. Ferlay J, Autier P, Boniol M, Heanue M, Colombet M, Boyle P. Estimates of the cancer incidence and mortality in Europe in 2006. *Annals of Oncology*. 2007; 18:581-592.
91. Horn-Ross PL, Morris JS, Lee M. Iodine and thyroid cancer risk among women in a multiethnic population: the Bay Area Thyroid Cancer Study. *Cancer Epidemiol Biomarkers Prev*. 2001; 10:979-85.
92. Kafadar K, Tukey JW. U.S. Cancer Death Rates: A Simple Adjustment for Urbanization. *International Statistical Review*. 1993; 61:257-281.
93. Montie JE, Meyers SE. Defining the ideal tumor marker for prostate cancer. *Urol Clin North Am*. 1997; 24:247-252.