Biometrical Letters
Vol. 47 (2010), No. 2, 133-146

Estimating the Relative Risk for Breast Cancer

Christos P. Kitsos

Department of Mathematics, Technological Educational Institute of Athens, Ag. Spyridonos & Palikaridi, 122 10 Egaleo, Athens, Greece, xkitsos@teiath.gr

SUMMARY

The aim of this paper is to discuss cases involving the Relative Risk for Breast Cancer. The Relative Risk is examined in either *in vitro* or *in vivo* studies. This research is based on a sample for women in Greece, among whom a number of variables were studied. It was found that full-term pregnancy and the menopause are statistically significant factors influencing the Relative Risk for breast cancer. A critical discussion and extension of the existing theoretical background are also provided.

Keywords: Relative Risk, Binary Regression, Logit model, Design, Fisher's Information.

1. Introduction

In principle, Risk Assessment is defined as the likelihood of adverse/unwanted responses to exposure to a restrictive agent. It is crucial to know the Cancer Bioassays ((Zapponni, 2002; Kitsos, 2005a), in order to analyze the collected data, as well as possible. Moreover when an optimal design approach is adopting the Relative Risk is reduced. The application of statistical analysis to medical problems goes back to Mantel and Haenszel (1959), while for an extended list of 1100 references concerning Cancer Risk Assessment, see Edler and Kitsos (2005). The "best" estimation leads to an optimal design (Kitsos, 2005b), with D-optimality being the most applicable to biological studies (Kitsos,1999). The particular statistical analysis needed for Biologically Based Models has been extensively discussed by Wosniok et. al. (1998). Research on Relative Risk for Cancer is based on:

(i) Previewing of the probability of a person's developing cancer, or

(ii) To assess the probability of cure, for patients.

In the case of prediction, there are two different lines of approach:

1. To identify the parameters affecting the probability of malignancy.

2. To estimate the likelihood of cancer appearing in the first place.

The former is realised through biochemical, genetic and clinical studies. The latter is considered through a statistical methodology applied to real populations.

Every parameter promoting tumorigenesis increases the risk associated with cancer's appearance or the probability of an unwanted development (poor prognosis) when cancer is present (Ratto *et al.*, 1998). These factors can be environmental, such as radiation and chemicals, or endogenous, mainly associated with the genetic profile of the patient.

In other cases a gene may affect a pathway or a cycle that in turn affects or regulates other biochemical aspects which can alter the hormonal profile of the subject. Whether the presence of that compound is the cause or the result of the malignancy is irrelevant. What matters is to use these molecules or their biochemical results as tools for prediction and assessment, or – more practically – for medical diagnosis. Many malignancies are still diagnosed when the metastatic process has already started, indicating a poor prognosis.

It has been pointed out that tumour markers (usually proteins associated with a malignancy) might be clinically usable in patients with cancer (Cheung 2000, Amaral-Mendes and Pluygers 1999). 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. A tumour marker may be used to define a particular disease entity, in which case it may be used for diagnosis, staging, or population screening.

Markers may also be used to detect the presence of occult metastatic disease, to monitor response to treatment, or to detect recurrent disease. Some of the markers related to breast cancer are widely monitored and the relevant tests

are routinely performed on patient samples, with the following being the most popular:

1. Estrogen Receptors (ER): Estrogen, one of the female sex hormones, often regulates the growth of breast cancer, and can therefore be useful for prognosis.
2. Progesterone Receptors (PR): Helps prediction of the response to hormonal therapy.
3. HER-2: This is a protein and/or gene amplification, both of which contribute to aggressive growth of cancer, while HER-2 overexpression occurs in approximately 25 percent of women with breast cancer.
4. p53: This is a tumour suppressor gene. Normally the *p53* protein, coded for by the *p53* gene, stops cells with DNA damage from multiplying until the DNA is repaired naturally or sends the defective cell into programmed cell death. When the *p53* gene becomes damaged or mutated, the protein becomes non-functional and loses its checkpoint control, allowing cancerous cells to replicate more readily.
5. S phase**:** When a cell has duplicated its genetic material and divided through the process of mitosis, it may become inactive or it can start another replicate cycle, beginning with the "S" or synthesis phase during which genetic material duplicates again. A higher than normal proportion of S phase is a measure of how actively a tumour is proliferating.

The most important risk factor associated with breast cancer is exposure to endogenous and exogenous oestrogen throughout the patient’s life. Many gene polymorphisms in the metabolism of breast cancer have been described as possible neoplasm etiological factors (Bugano et al.. 2008). In such cases, an accurate estimate of a woman's breast cancer risk is essential for optimal patient counselling and management.

Estrogens, as activators of cellular proliferation, have been related to breast cancer progression. In this paper, we examine this problem in a case-control study: the frequencies of genotype polymorphisms were determined for genes involved in catechol estrogen formation, via estrogen biosynthesis (*CYP17*) and/or inactivation (*COMT*) (Huang et. al. 1999), and their association with an elevated risk for breast cancer was studied in Greek women.

Regarding breast cancer risk assessment, susceptibility, and its relation to CYP17, MspA1 polymorphism, different opinions have been expressed (Feigelson et al., 2002). Moreover Huang et al. (1999) found a positive association between the breast cancer relative risk and the individual susceptibility genotypes. The association of the relative risk with the number of susceptibility genotypes was stronger in women with prolonged exposure, women with higher estrogen levels – implied by early menarche – and women with higher body mass index (Kobayashi and Kawakub, 1994). The collected data set for the breast cancer measures: full term pregnancies, menarche, menopause, COMT, CYP17, among other variables, see section 3.

In this paper, the logit model is adopted to estimate the relative risk. Therefore the second section is devoted to a compact consideration of the logistic regression.

2. The Logit Model and the Relative Risk

Consider a subject with attributes given by the input vector $\mathbf{X}=(X_1, X_2,\ldots, X_p)^T$. Then, risk analysis concentrates interest on the parameter p(x), i.e. the probability that this subject has a certain characteristic C, given that the input vector takes the real vector value x, i.e. **X=x**, and measures the odds ratio or the Relative Risk RR=p(**x**)/(1-p(**x**)). The logit model has been suggested since the pioneering paper of Berkson (1955). That is, the log odds have been assumed linear i.e.

$$\log\{p(\mathbf{x})/(1-p(\mathbf{x})\} = \mathbf{x}^T\boldsymbol{\beta} \qquad (2.1)$$

with $\boldsymbol{\beta}$ being an appropriate vector of regression parameters, and $\mathbf{x}^T\boldsymbol{\beta} = \beta_0+\beta_1x_1 +\ldots.+\beta_px_p,$ (Hosmer and Lemeshow, 1989). Then, from (2.1), considering the cumulative distribution function of the logistic distribution function, say $\text{Logit}(z) = e^z/(1+e^z)$, Rao and Toutenburg (1999), we can evaluate that $p(\mathbf{x}) = \text{Logit}^{-1}(\mathbf{x}^T\boldsymbol{\beta})$, see also Kitsos (2006).

A very similar approach was adopted by Bliss (1935) in his pioneering paper on Risk Analysis, where he considered the cumulative distribution function of the standard normal distribution φ(t), say

$$\text{Normal}(z) = \int_{-\infty}^{z} \varphi(t)dt \text{ , and let } p(x) = \text{Normal}^{-1}(\mathbf{x}^T\boldsymbol{\beta}).$$

This method is known as probit analysis, and it has been pointed out that, for values of p = p(x) within [0.2, 0.8], both methods are very close. Various extensions have been discussed (Aranda-Ordaz, 1981, Taylor, 1988, among others). In principle a general binary regression model is considered as $F^1(p(\mathbf{x})) = \mathbf{x}^T\boldsymbol{\beta}$ equivalent to $p(\mathbf{x}) = F(\mathbf{x}^T\boldsymbol{\beta})$, for a given continuous cumulative distribution function F.

If we assume that with the covariate x_i, i = 1,2,…,p are associated n_i persons, then the MLE of **β**, say **b**, provides linear predictors x, through the estimated probabilities π_i, i = 1,2,…,p. Then we can evaluate the estimated expected information matrix $\mathbf{I}(\xi,\mathbf{b})$ which, for the design matrix $\mathbf{D}=(x_{ij})$ i = 1,2…,p, j = 1,2,…,n, is equal to $\mathbf{I}(\xi,\mathbf{b}) = \mathbf{D}^T\mathbf{W}\mathbf{D}$, with $\mathbf{W} = \text{diag}\,(w_{ii})$, $w_{ii} = n_i\pi_i(1-\pi_i)$, i = 1,2,…,p with ξ being a design measure; see McCullagh and Nedler (1989) for details. Thanks to Fisher's information the variances can be estimated, and thanks to the Rao-Blackwell theorem, bounds for the variances are considered, and approximated confidence intervals can be obtained.

For the log-likelihood ratio statistic $\log L$, the deviance, say Dev, is defined as $Dev = 2\log L$, and it can be proved that $Dev \sim \chi^2_{n-p}$. Therefore the expected value of Dev has to be close to n-p, when a good fit of the data by the model occurs (Breslow and Day, 1980).

Alternatively to the MLE, Firth (1993) introduced a modification of the score function so as to reduce the small sample bias of MLE. He imposed the so-called "penalty" term on the typical likelihood function; the "penalized likelihood function" is defined as

$$\mathcal{L}^* = \mathcal{L}(\beta\,;\,\text{x})\,|\text{I}(\xi,\beta)|^{1/2} \qquad (2.2)$$

where **β** is the parameters involved, x is the data, ξ is the design measure. The log-likelihood of (2.2) is equal to

$$l^* = l + \frac{1}{2}\ \ln|\mathrm{I}(\xi,\beta)| \tag{2.3}$$

and the modified score function U is reduced to

$$\mathrm{U}^* = \mathrm{U} + \ \frac{1}{2}\mathrm{tr}\left\{\mathrm{I}(\xi,\beta)^{-1}T\right\} \text{ with } T = \left\{T_{\mathrm{ij}}\right\} = \left\{\frac{\partial \mathrm{I}(\xi,\beta)_{\mathrm{ij}}}{\partial \theta_{\mathrm{j}}}\right\} \tag{2.4}$$

Now, the regression coefficients β of the proposed logistic model (2.1) quantify the relationship of the independent variables to the dependent variable, involving the parameter as the Relative Risk (RR) defined already. A typical χ^2 test is applied for the null hypothesis H_0: RR=1 vs. H_1: RR≠1 is identical to a test of the equality of the two proportions' having or not having the characteristic C, i.e. H_0: $P_0 = P_1$ vs. H_1: $P_0 \neq P_1$.

We mention that $\mathrm{RR} = \exp(\beta_1) \cong \exp(\hat{\beta}_1)$ can be biased due to population heterogeneity caused by confounding factors associated with the response, for the simple logit model and not only for that model. Recall (2.1). To decide for a quadratic term in (2.1), namely $\mathrm{logit}[P(Y=1)] = m_0 + m_1 x + m_2 x^2$, is quite different. If the term m_2 is essentially different from zero, the question is how "robust" is the linear model. We shall apply this consideration and we shall not restrict ourselves to linear model approximation, as occurs in all applications. Moreover it has been extensively discussed by Fornius (2008) that when g(π), the link function, is of the form

$$\mathrm{g}(\pi) = \beta_0 + \beta_1(\mathrm{x} - \mu)^2 . \tag{2.5}$$

for the parameter vector $\theta = (\beta_0, \beta_1, \mu)$ a 3-point D-optimal design is:

$$\xi_3 = \begin{Bmatrix} -\mathrm{x} & 0 & \mathrm{x} \\ 1/3 & 1/3 & 1/3 \end{Bmatrix}, \text{ with x = 1.957 or 1.238,}$$

while a 4-point D-optimal design can be also obtained; Fornius (2008).

For the Logit model and link function defined by the logit transformation, the D-optimal design allocates half observations (i.e. ξ = 1/2) at the optimal design points ±1.5434, when the vector of coefficients is (0, 1). For the optimal design approaches in Ca Bioassays, see Kitsos (2002, 2005a, b). The sequential designs for the logit model have been also discussed by Fornius (2008, Chapter 5). The stochastic approximation plays an important role in the development of a sequential design, usually related to D-optimality (Kitsos, 1999) in applications. Fornius (2008a) creates a sequential **c**-optimal design and compares it with the D-optimal designs for the logit model. Different **c**-optimal designs can be produced with different $\boldsymbol{\beta} = (\beta_0,\beta_1)$ and "direction ray" **c**. For $\boldsymbol{\beta}$ = (1, 1), **c** = (1,3) and the explanatory variable within [-3,3] the **c**-optimal design is a two point design at -3 and 1.4164, with the corresponding weights equal to 0.1826 and 0.8174 respectively. It was pointed out that the values of the input vector (0, 1) are essential to construct an optimal design, therefore for the Logit model, the "canonical form" is introduced. This is based on the fact that **c**-optimality remains invariant when the data undergoes a “linear” transformation . Now we state and prove the following Theorem 1.

Theorem 1. The set of the transformations

$$\Im = \left\{ G = \begin{pmatrix} 1 & 0 \\ \beta_0 & \beta_1 \end{pmatrix}, \beta_0, \beta_1 \in \mathsf{R} \right\} \tag{2.6}$$

forms an affine transformation group, under matrix multiplication, which preserves the c-optimal information measure.

Proof: Considering another element of the group $\Im$ as

$H = \begin{pmatrix} 1 & 0 \\ a_0 & a_1 \end{pmatrix}$, $a_0, a_1 \in \mathsf{R}$ then it can be proved easily that

$$HG = \begin{pmatrix} 1 & 0 \\ \gamma_0 & \gamma_1 \end{pmatrix} \in \Im,\ \gamma_0, \gamma_1 \in \mathsf{R}, \text{ with } \gamma_0 = \alpha_0 + a_1\beta_0 \in \mathsf{R},\quad \gamma_1 = a_1\beta_1 \in \mathsf{R}.$$

Similarly if $K \in \Im$ then it can be proved that (HG) K = H (GK)$\in \Im$. The unit transformation $I \in \Im$ is the identity matrix, the inverse transformation $G\text{-}1 \in \Im$ is the inverse matrix of G, and GI = IG = G$\in \Im$. We introduce the

transformation of the design space U to W of the form: w = G u$\in$ W with G from the group $\Im$. Take c-optimality as the optimal design criterion function Φ, then $\Phi(M_u) = c^t M_u^{-1} c$, with $M_u = M_u(\beta,\xi)$ the average per observation information matrix in U space, and Mw in W space. Then it is easy to see that $c^T M_u^{-1} c = c_w^T M_w^{-1} c_w$ with $c_w = G c_u, G \in \Im$.

This theoretical result practically means that we can work as follows: perform the experiment with the "easiest" scale and position parameters. To perform the experiment at the optimal design points as in **c**-optimality, prior knowledge of β_0,β_1 is needed. Then transform the results with an element from the group of affine transformations, and we still have an optimal design. With the group of affine transformations the experimenter can move to a different "orbit", adopting the $\mathbf{Gc}^T$ "direction ray", when he has evaluated only one set of experiments. The discussion was based on the fact that the evaluation of Relative Risk is mainly based on the logit model, and this discussion is applied in section 3 below.

Now let us consider the general case, where the dichotomous response variable Y denotes whether (Y=1) or not (Y=0) the characteristic under investigation is linked with the k regression variables $X=(X_1, X_2, \ldots, X_k)$ via the logit equation:

$$P(Y=1) = \frac{\exp\left\{\beta_0 + \sum_{k=1}^{K} \beta_k X_k\right\}}{1+\exp\left\{\beta_0 + \sum_{k=1}^{K} \beta_k X_k\right\}} \tag{2.7}$$

This is equivalent to Logit $\Pr(Y=1|\,\mathbf{X}) = \beta_0 + \sum_{k=1}^{K} \beta_k X_k$

With this formulation we have the benefit that the relative risk (RR) for individuals having two different sets $\mathbf{X}'$ and $\mathbf{X}$ of risk variables is

$$RR = \frac{P(X')[1-P(X)]}{P(X)[1-P(X')]} = \exp\left\{\sum_{i=1}^{K} \beta_i (X_i' - X_i)\right\} \tag{2.8}$$

It is essential that the RR of the k regressors (RR_k) influences the RR of the k+1 regressors, RR_{k+1} as in relation (12) below, due to the following:

Theorem 2. Let the relative risk, as above in (2.8), be $RR_K = \sum^{K} \beta_i \delta_i$, $\delta_i = X_i' - X_i$. Then if a variable is added, the relative risk of the k+1 variables equals the k-variable relative risk times the new variable's relative risk, i.e.

$$RR_{k+1} = RR_k \, r_{k+1} \tag{2.9}$$

This theoretical result informs the experimenter that the k input variables continue to participate with a "total" relative risk RR_k either in a k-variable model or in a (k+1)-variable model.

Proof: Indeed:

$$RR_{K+1} = \exp\left\{\sum_{i=1}^{K+1} \beta_i \delta_i\right\} = \exp\left\{\sum_{i=1}^{K} \beta_i \delta_i + \beta_{K+1}\delta_{K+1}\right\} =$$

$$\exp\left\{\sum_{i=1}^{K} \beta_i \delta_i\right\} \exp\left\{\beta_{K+1}\delta_{K+1}\right\} = RR_K \cdot r_{K+1},$$

where the definition of r_{K+1} is obvious. So the relative risk of the incoming variable is

$$r_{K+1} = \frac{RR_{K+1}}{RR_K}, \text{ q.e.d.}$$

3. Analyzing collected breast cancer data sets

The study concerns 98 breast cancer patients and 125 healthy controls, compared considering the age at menarche, age at menopause, number of full-term pregnancies and the CYP17, COMT genotypes. The frequency of the CYP17 A1/A1 genotype was compared to A1/A2 and A2/A2, while the frequency of the COMT G/G genotype was compared to G/A and A/A.

The logit model was used in order to compare the patients with the controls. The final model was chosen with a stepwise regression. In all cases the best

model is chosen with log-likelihood, and the goodness of fit is examined (Lemeshow and Hosmer,1982). In Table 1 the Relative Risks of the full model are given. Using stepwise regression the final model was estimated (Table 2), with two input variables, statistically significant (at a significance level of 0.05), namely full-term pregnancy and menopause.

The X^2 test was used in order to examine whether each of the genotypes CYP17 and COMT are related to Cancer Risk. Only COMT seems to be related to cancer at a significance level of 0.10 (value of P = 0.096). Moreover the two genotypes are related to Ca when the age of menarche is lower than the average value of 12.5 years and the age of menopause is over 48.8 years, while CYP17 is not (Kitsos et al., 2007).

Table 1. Relative Risks of the full model

Variable	RR	Std. Error	Z	p-value
Full term preg.	1.45	0.17	3.09	0.002
CYP17	0.98	0.19	-0.06	0.95
COMT	1.10	0.22	0.48	0.63
Menarche	0.95	0.08	-0.56	0.57
Menopause	1.07	0.03	2.18	0.029

Table 2 Relative Risks of the final model

Variable	RR	Std. Error	Z	p-value
Full term preg.	1.42	0.12	3.09	0.003
Menopause	1.04	0.03	2.14	0.028

Practical interpretation of the coefficients of the statistically significant variables provides evidence that when the age at menopause increases by one year, the probability of breast cancer increases by 4%. Moreover women with full-term pregnancy have a 42% lower probability of contracting breast cancer than other women. It is useful to notice that, from the interpretation of the coefficient of the variable age of menarche in the full model, it can be considered that when the age of menarche increases by one year, the probability of Ca decreases by 5%. Therefore there is evidence that for the data set under

consideration, when the age of menarche increases the relative risk for Ca is increased.

Table 3. Analyzing COMT restricted to menopause at over 48.8 years

COMT	Cases	Controls	Total
G/G	39	20	59
G/A	41	46	87
A/A	19	12	31
Total	99	78	177

For Table 3 we performed a X^2 test, relevant P-value = 0.062, which provides practical evidence that the genotype COMT is related to breast cancer ($0.05 < P < 0.1$). Finally when the age of menarche is lower than 12.5 years, both CYP17 and COMT are not related to cancer. The contribution of the time for which a woman is menstruating, with a Logit model, provided a parameter vector (-0.9925, -0.1542) for the (time, constant) with standard error 0.48, 0.27, and RRtime =0.3707. The contribution of the time a woman first menstruates (sample average age about 13.5 years old) was also studied with RRstart = 4.0615, when the coefficients were (-1.4016, -.7885) and corresponding standard errors 0.42 and 0.24 respectively. Table 4 shows the contribution of both variables to the model formed by them.

Table 4. Multivariate Logit Model Analysis

Variables	Coefficient b	se(b)	Sig at	RR
Start	-1.2543	0.4462	0.0049	3.5052
Time	-0.7829	0.4293	0.0688	0.3571
Constant	-0.4061	0.3487	0.2442	

When the square logit model was adopted and analysed, for various variables, the parameter vector for (const, menarche, $(\text{menarche})^2$) = (-3.782, .576, -.023) with the vector of RRs =(.023, .977, 1.779). The square term was not statistically significant, and this provides evidence that the linear logit was well assumed.

4. Discussion

In principle, Risk Assessment is viewed as an empirical process to determine the probability that an adverse health effect will be associated with exposure to a chemical. In this paper we were investigating not a chemical, but the main characteristics for females considering the frequency of certain genotypes correlated with the level of estrogens in blood as well as the duration of the exposure (menarche-menopause). Although nowadays it is easy to do with the existing software, which is becoming cheaper and cheaper, this is still a rather complicated task, which can be carried out with a given model among various rival models. The odds ratio models provide a solution in evaluating the Relative Risk, and the discussed (affine group of) transformation might be helpful in designing and then performing an experiment.

When a chemical is present, the object is to estimate the relative risk as a function of the dose (genotype) of the chemical that reaches the target organ (breast). In this paper the risk for breast cancer was evaluated as a function of full-term pregnancies and menopause.

Acknowledgements

I would like to thank the CCMS/NATO pilot study for the generous grant, under Dr Giovanni Zapponi's supervision (who continuously encouraged me) concerning experimental carcinogenesis, as well as all the members of the research group who made corrections to my attempts. The comments of the referee are very much appreciated.

REFERENCES

Amaral-Mendes J.J., Pluygers E. (1999): Use of Biochemical and Moleular Biomarkers for Cancer Risk Assessment in Humans. In: Respectives on Biologically Based Cancer Risk Assessment, by Cogliano V., Luebeck G., Zapponi G., NATO-Challenges of Modern Society, 23: 81–152.

Aranda-Ordaz R.D. (1981): On two families of transformations to additivity for binary response data. Biometrika 68: 357–364.

Berkson, J. (1955): Maximum Likelihood and Minimum X^2 estimates of the logistic function. Journal of American Stat. Association 50: 130–162.

Bliss C.I. (1935): The calculation of the dosage mortality curve. Annals of Applied Biology 22: 134.

Breslow N.E., Day N.E. (1980): Statistical methods in Cancer Research vol. 1: The analysis of case controls studies. Lyon: IARC.

Bugano D.D., Conforti-Froes N., Yamaguchi N.H., Baracat E.C. (2008): Genetic polymorphisms, the metabolism of estrogens and breast cancer: a review. European Journal Gynaecol Oncology 29 (4): 313–20.

Cheung K., Graves C.R.L., Robertson J.F.R. (2000): Tumour marker measurements in the diagnosis and monitoring of breast cancer. Cancer Treat Review 26: 91–102.

Edler L., Kitsos C.P. (2005): Recent Advances in Qualitative Methods in Cancer and Human Health Risk Assessment. Editors, Wiley, UK.

Firth D. (1993): Bias Reduction of Maximum Likelihood Estimates. Biometrika 20: 272–319.

Feigelson S.H., McKean-Cowdlin R., Henderson E.B. (2002): Concerning the CYP17 MspA1 polymorphism and breast cancer risk: a meta-analysis. Mutagenesis 17: 445–446.

Fornius E.F. (2008): Optimal Design of Experiments for the Quadratic Logistic Model. PhD thesis, Stockholm University, Sweden.

Fornius E.F. (2008a): Sequential designs for Binary Data with the Response to Maximize the Probability of Response. Communications in Statistics – Simulation and Computation 37: 6.

Gancberg D., Lespagnard L., Rouas G., Paesmans M., Piccart M., Di Leo A., Nogaret J.M., Hertens D., Verhest A., Larsimont D. (2000): Sensitivity of HER-2/neu antibodies in archival tissue samples of invasive breast carcinomas: correlation with oncogene amplification in 160 cases. American Journal of Clinical Pathology 113: 675–682.

Hosmer D.W., Lemeshow S. (1989): Applied Logistic Regression. John Wiley.

Huang C-S., Chern H-D., Chang K-J., Cheng C-W., Hsu S-M., Shen C-Y. (1999): Breast Cancer Risk Associated with Genotype Polymorphism of the Estrogen - metabolizing CYP17, CYP1A1 and COMT: A Multigenic Study on Cancer Susceptibility. Cancer Research 59: 4870–4875.

Kitsos C.P. (1999): Optimal Designs for Estimating the Percentiles of the Risk in Multistage Models in Carcinogenesis. Biometrical Journal 41(1): 33–43.

Kitsos C.P. (2002): The Ca Risk Assessment as an Optimal Experimental Design. In: Endocrine Disrupters and Carcinogenic Risk Assessment, by L. Chyczewski et. al. (Eds): 329–337.

Kitsos C.P. (2005a): The Ca Risk Assessment as a Bioassay. In: 55th Session of the International Statistical Institute, 5–12 April 2005, Sydney.

Kitsos C.P. (2005b): Optimal Design for Bioassays in Carcinogenesis. In: Quantitative Methods for Cancer and Human Health Risk Assessment, by Lutz Edler and Christos Kitsos (Eds), Wiley, England: 267–279.

Kitsos C.P. (2006): On the Logit Methods for Ca Problems. In: Statistical Methods for Biomedical and Technical Systems, Limassol, Cyprus: 335–340.

Kitsos C.P., Chalikias M.S., Voutsinas G.P. (2007): A quantitative approach to Brest Cancer Data. In: 2d International Conference on Cancer Risk Assessment (ICCRA2), Santorini, 25–27 May 2007, e- proceedings.

Kobayashi T., Kawakub T. (1994): Prospective investigation of tumour markers and risk assessment in early cancer screening. Cancer 73(7): 1946–1953.

Lemeshow S., Hosmer D.W. (1982): The use of goodness-of fit statistics in the development of logistic regression models. American Journal of Epidemiology, 115: 92–106.

Mantel N., Haenszel W. (1959): Statistical aspects of the analysis of data from retrospective studies of disease. J. of the National Cancer Institute 22: 719–748.

Marty M., Cognetti F., Maraninchi D. (2005): Efficacy and safety of trastuzumab combined with docetaxel in patients with HER2-positive metastatic breast cancer given as first-line treatment: results of a randomized phase II trial. Journal of Clinical Oncology 23: 4265–4274.

McCullagh P., Nedler J.A. (1984): Generalized Linear Models. 2nd ed. Chapman and Hall, London.

Mitrunen K., Hirvonen A. (2003): Molecular Epidemiology of sporadic breast cancer. The role of polymorphic genes involved in estrogens biosynthesis and metabolism. Mutation Research 544: 9–41.

Nicolini A., Carpi A. (2000): Postoperative follow-up of breast cancer patients: overview of progress in the use of tumour markers. Tumour Biology 21: 235–48.

Rao C.R., Toutenburg H. (1999): Linear Models. 2nd ed. Springer-Verlag, New York.

Ratto C., Sofo L., Ippoliti M., Merico M., Doglietto G.B., Crucitti F. (1998): Prognostic factors in colorectal cancer. Literature review for clinical application. Dis. Colon. Rectum 41: 1033–1049.

Ridolfi R.L., Jamehdor M.R., Arber J.M. (2000): HER-2/ neu testing in breast carcinoma: a combined immunohistochemical and fluorescence in situ approach. Mod. Pathology 13: 866–873.

Risch A., Dally H., Edler L. (2005): Genetic Polymorphisms in Metabolising Enzymes as Lung Cancer Risk Factors. In: Recent Advances in Quantitative Methods in Cancer and Human Risk Assessment, by L. Edler and C. Kitsos (Eds). Wiley, UK.

Ross J.S., Fleisher J.A. (1999): HER-2/neu (c-erb-B2) gene and protein in breast cancer. American Journal of Clinical Pathology 112 (Suppl 1): 53–67.

Schwab M. (1998): Amplification of oncogenes in human cancer cells. Bioassays 20: 473–479.

Taylor J.M.G. (1988): The cost of generalizing logistic regression. Journal of American Statistical Association 83: 1078–1083.

Wosniok W., Kitsos C., Watanabe K. (1998): Statistical issues in the Application of Multistage and Biologically Based Models. In: Respectives on Biologically Based Cancer Risk Assessment, by Cogliano V., Luebeck G., Zapponi G. NATO-Challenges of Modern Society, Vol. 23: 243–272.

Zapponni G.A. (2002): Carcinogenetic Risk Assessment: Some Points of Interest for a Discussion. In: Endocrine Disrupters and Carcinogenic Risk Assessment, by Chyczewski L., Niklinski J., Plugers E. (Eds). IOS press: 15–27.