SORT 30 (2) July-December 2006, 193-204

Statistical models to study subtoxic concentrations for some standard mutagens in three colon cancer cell lines

Xavier Bardina[1], Laura Fernández[2], Elisabet Piñeiro[2],
Jordi Surrallés[2] & Antonia Velázquez[2]

Universitat Autònoma de Barcelona

Abstract

The aim of this work is to propose models to study the toxic effect of different concentrations of some standard mutagens in different colon cancer cell lines. We find estimates and, by means of an inverse regression problem, confidence intervals for the subtoxic concentration, that is the concentration that reduces by thirty percent the number of colonies obtained in the absence of mutagen.

MSC: 62J05

Keywords: Inverse regression problem, subtoxic concentration, confidence interval.

1 Introduction

Human populations are exposed to a variety of environmental agents, including biological, chemical and physical entities, that can injure the DNA and cause adverse health consequences, such as cancer. It is therefore extremely important to detect these mutagenic agents, unravel their mechanisms of action, and define what type of injury they produce. All this information is crucial to determine the genetic risk of exposed population.

[1] Departament de Matemàtiques, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain. bardina@mat.uab.cat

[2] Departament de Genètica, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain. lfernandez@einstein.uab.es, epineiro@einstein.uab.es, jordi.surralles@blues.uab.es, avh@cc.uab.es

Received: September 2006
Accepted: November 2006

Mutagenicity assays are specifically designed to detect DNA damaging agents and to analyze their biological effects. Most of these assays are performed *in vitro* with a variety of cell types and under controlled conditions of cell growth and viability. The range of the treatment concentrations is usually determined in a previous toxicity study. The chosen concentrations for the mutagenicity study must be subtoxic in order to ensure biological effects without extreme cell death. This is why a well performed toxicity assay is absolutely required as a previous routine in all mutagenicity assays.

Tandem repeated DNA sequences of few nucleotides, the so-called microsatellites, are known to be highly unstable in colon cancer cells defective in DNA mismatch repair (MMR). In addition, expansion of tandem repeated sequences have been causally related to a number of degenerative diseases including miotonic dystrophy, fragile X syndrome and Huntington's disease. The final aim of the present investigation was to determine whether microsatellite instability is inducible *in vitro* by a set of mutagens of different mode of action. To do so, subtoxic concentrations, i.e. those inducing a reduction of thirty percent in cell viability, had to be previously determined for the following standard mutagens: bleomycin (BLEO), N-methyl-N-nitrosourea (MNU), ethoposide (ETO), mitomycin C (MMC) and ethidium bromide (EtBr). The toxicity assay was carry out in three human fibroblast cell lines derived from colon tumours: the wild-type cell line SW480, and lines HCT116 and LoVo which are both defective in MMR.

The toxicity data were obtained by the colony forming efficiency method. About 200 cells from exponentially growing cell cultures were platted in triplicate on $25cm^2$ plates (falcons). After allowing for attachment to the plate for 24 hours, the medium was replaced with fresh medium containing the test chemical at different concentration for each replica. Cell lines were maintained in these conditions for 10 days, replacing the medium every 3 days. The plates were washed with phosphate buffer saline, fixed with methanol, and stained with Giemsa. Colonies with more than 50 growing cells were counted. A reduction in the number of colonies after treatment is interpreted to be a consequence of the chemical toxicity.

In this article we will propose different types of statistical models to study the effect of the concentration of the mutagens in three colon cancer cell lines. We have considered linear and exponential models and in each case we propose the model that approximates better the data. When we consider a regression linear model, we need to assume that the errors are additive and normally distributed. When the model considered is an exponential one, we assume that the errors are multiplicative and their logarithms are normally distributed. In all the models proposed, the analysis of the residuals does not give evidence that controvert this hypothesis. For each mutagen, cell line and concentration we have three values. Then, we will consider models with weights, where the weights are calculated by the inverse of the estimated variances.

For each mutagen and each cell line we will obtain an estimation and a confidence interval for the subtoxic concentration, that is the concentration that reduces a thirty percent the initial number of colonies.

This article is organized as follows. In Section 2 we provide the mathematical justification for the calculation of the confidence intervals used in this study. In Section 3 we present some models for the standard mutagens: bleomycin, N-methyl-N-nitosourea, ethoposide, mitomycin C and ethidium bromide respectively. We also provide for each mutagen and each cell line the estimate of the subtoxic concentration and a confidence interval for this concentration. In Section 4 the use of weighted models is justified. Finally, in the appendix the data used in the study are shown.

2 Mathematical justification of the confidence intervals used in this study

In this section we will explain the method used in order to obtain a confidence interval of the subtoxic concentration, that is the concentration that reduces by thirty percent the initial number of colonies.

Consider the linear regression model

$$y = \beta_0 + \beta_1 x,$$

where y is the number of colonies and x the concentration of the mutagen.

We will estimate the concentration that reduces to 70 percent the number of colonies for $x = 0$ (that is, the concentration x such that $y = 0.7\beta_0$) by

$$\hat{x} = \frac{-0.3 \cdot \hat{\beta}_0}{\hat{\beta}_1},$$

where $\hat{\beta}' = \left(\hat{\beta}_0, \hat{\beta}_1\right)$ is the usual estimate of (β_0, β_1) obtained by the least-squares method. That is,

$$\hat{\beta} = (X'X)^{-1} X'y,$$

where

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

and $(y_1, x_1), \cdots, (y_n, x_n)$ are the data of the number of colonies and concentrations, respectively.

If we consider the vector $\lambda' = (\lambda_1, \lambda_2)$, it is well known that we can obtain a $100(1-\alpha)$ % confidence region for $\lambda'\beta$ by using the following expression:

$$\lambda'\beta = \lambda'\hat{\beta} \pm t_{\alpha/2,n-2} S \sqrt{\lambda'\Lambda\lambda},$$

where $\Lambda = (X'X)^{-1}$, S is the estimate of the standard deviation of the errors and $t_{\alpha/2,n-2}$ is the critical point such that $P(|T| > t_{\alpha/2,n-2}) = \alpha$ where T is a Student's t distribution with $n-2$ degrees of freedom.

On the other hand, from the linear regression model, the concentration x that reduces to 70 percent the inicial number of colonies satisfies that

$$0.7\beta_0 = \beta_0 + \beta_1 x,$$

that is

$$0.3\beta_0 + x\beta_1 = 0$$

and we can write this expression as

$$\lambda'\beta = 0$$

with $\lambda' = (0.3, x)$.

So, in order to obtain a sort of $100(1-\alpha)$ % confidence interval for the concentration x that reduces a 30 % the inicial number of colonies, we can solve the following system:

$$\begin{cases} \lambda'\beta = \lambda'\hat{\beta} \pm t_{\alpha/2,n-2} S \sqrt{\lambda'\Lambda\lambda} \\ \lambda'\beta = 0 \end{cases} \tag{1}$$

whith $\lambda' = (0.3, x)$. That is, we have to find the intersections (if there exists) between the line $\lambda'\beta = 0$ and the curves given by $\lambda'\beta = \lambda'\hat{\beta} \pm t_{\alpha/2,n-2} S \sqrt{\lambda'\Lambda\lambda}$. This kind of problems are called inverse regression problems (see for example Draper and Smith, 1981).

When we use a linear model with weights, the matrix Λ now is given by

$$\Lambda = \left(X'V^{-1}X\right)^{-1}$$

and the estimate of β is

$$\hat{\beta} = \left(X'V^{-1}X\right)^{-1} X'V^{-1}y,$$

where

$$V^{-1} = \begin{pmatrix} w_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & w_n \end{pmatrix}$$

and $w_1, w_2, \dots, w_n$ are the weights of the data $y_1, y_2, \dots, y_n$, respectively (see for example

Montgomery, 1992, or Draper and Smith, 1981). Observe that in our examples the confidence interval are approximated because the weights are estimated from the data.

When we consider an exponential model, that is,

$$y = e^{\beta_0+\beta_1 x},$$

we assume that the residuals are multiplicative and that their logarithms are normally distributed. That is, we suppose that if we consider the transformation

$$\ln y = \beta_0 + \beta_1 x$$

we can proceed like in a linear model. But in this case, we have to estimate the concentration x such that $y = 0.7e^{\beta_0}$. Thus, we will estimate x by

$$\hat{x} = \frac{\ln 0.7}{\hat{\beta}_1}.$$

Then, we obtain the confidence interval solving the system:

$$\begin{cases} \lambda'\beta - \ln 0.7 = \lambda'\hat{\beta} - \ln 0.7 \pm t_{\alpha/2,n-2} S \sqrt{\lambda'\Lambda\lambda} \\ \lambda'\beta - \ln 0.7 = 0, \end{cases}$$

where $\lambda' = (0,\ x)$. We remark that this confidence interval has an exact confidence level of $100(1-\alpha)$ % only when the weights are perfectly known. If the weights are estimated from the sample then the confidence levels are approximated.

3 Examples

Using the method described in Section 2 we will obtain now a confidence interval for the subtoxic concentration of the Bleomycin in the cell line LoVo. From the data given in the Appendix we have that

$$X = \begin{pmatrix} 1 & 0.0000 \\ 1 & 0.0000 \\ 1 & 0.0000 \\ 1 & 0.0001 \\ \vdots & \vdots \\ 1 & 0.0500 \end{pmatrix}, \quad y = \begin{pmatrix} 26.5 \\ 29.0 \\ 31.0 \\ 32.5 \\ \vdots \\ 6.5 \end{pmatrix}, \quad V^{-1} = \begin{pmatrix} 0.1967 & 0 & \dots & 0 \\ 0 & 0.1967 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0.2449 \end{pmatrix}.$$

Thus, with an easy computation we obtain that

$$\Lambda = \left(X'V^{-1}X\right)^{-1} = \begin{pmatrix} 1.039 & -22.082 \\ -22.082 & 1008.297 \end{pmatrix},$$

and the estimate of β is given by

$$\hat{\beta} = \left(X'V^{-1}X\right)^{-1} X'V^{-1}y = \begin{pmatrix} 27.205 \\ -440.677 \end{pmatrix}.$$

In this example the estimate of the standard deviation of the errors is $S = 1.083$ and the critical point of the Student's t is $t_{\frac{0.05}{2},11} = 2.201$.

Then in order to obtain the confidence interval explained in this section we have to solve the system (1). In this example we get

$$\begin{cases} 0.3\beta_0 + \beta_1 x = 8.161 - 440.677x \pm 2.383\sqrt{0.093 - 13.249x + 1008.297x^2} \\ 0.3\beta_0 + \beta_1 x = 0. \end{cases}$$

So, we have to find the roots of the following equation:

$$440.677x - 8.161 = \pm 2.383\sqrt{0.093 - 13.249x + 1008.297x^2},$$

that is, the roots of the quadratic equation

$$33182.804x^2 - 1253.193x + 11.634 = 0,$$

that are given by the values

$$\begin{aligned} x_1 &= 0.0164 \\ x_2 &= 0.0213. \end{aligned}$$

Then, (0.0164, 0.0213) is a confidence interval of approximately 95 % for the subtoxic concentration of the Bleomycin in the cell line Lovo.

3.1 Bleomycin (BLEO)

We have considered a regression line with weights for each cell line. Recall that y is the number of colonies and x the concentration of bleomycin. The models obtained using the SAS system are:

cell line	regression line
HCT116	$y = 67.597 - 1168.925\,x + \varepsilon$
LoVo	$y = 27.205 - 440.677\,x + \varepsilon$
SW480	$y = 53.488 - 566.464\,x + \varepsilon$

The regression coefficients R-square were 0.9285, 0.9372 and 0.9452, respectively. Finally, with these models we obtain the following estimates, and the following confidence intervals, for the subtoxic concentration in each cell line.

cell line	concentration estimate	95% confidence interval
HCT116	0.0173	(0.0152, 0.0203)
LoVo	0.0185	(0.0164, 0.0213)
SW480	0.0283	(0.0259, 0.0315)

3.2 N-methyl-N-nitrosourea (MNU)

We have considered a regression line with weights for the cell lines LoVo and SW480. In this case, the models obtained using the SAS system are the following:

cell line	regression line
LoVo	$y = 22.117 - 0.218\,x + \varepsilon$
SW480	$y = 67.826 - 0.663\,x + \varepsilon$

The regression coefficients R-square were 0.9330 and 0.9512, respectively. With these models we obtain the following estimates, and the following confidence intervals, for the subtoxic concentration in each cell line.

cell line	concentration estimate	95% confidence interval
LoVo	30.477	(29.299, 31.777)
SW480	30.708	(29.249, 32.344)

3.3 Ethoposide (ETO)

We have considered a regression line with weights for each cell line. The models obtained using the SAS system are:

cell line	regression line
HCT116	$y = 130.075 - 860.173\,x + \varepsilon$
LoVo	$y = 43.509 - 233.079\,x + \varepsilon$
SW480	$y = 104.453 - 423.569\,x + \varepsilon$

The regression coefficients R-square were 0.9032, 0.7189 and 0.8282, respectively. With these models we obtain the following estimates, and the following confidence intervals, for the subtoxic concentration in each cell line.

cell line	concentration estimate	95% confidence interval
HCT116	0.0454	(0.0380, 0.0563)
LoVo	0.0560	(0.0432, 0.0838)
SW480	0.0740	(0.0593, 0.0996)

3.4 Mitomycin C (MMC)

For this data sets we have fitted a regression line with weights for each cell line. The models obtained using the SAS system are:

cell line	regression line
HCT116	$y = 85.321 - 6334.859\,x + \varepsilon$
LoVo	$y = 13.989 - 714.984\,x + \varepsilon$
SW480	$y = 48.554 - 4174.451\,x + \varepsilon$

The regression coefficients R-square were 0.9253, 0.5364 and 0.9314, respectively. With these models we obtain the following estimates, and the following confidence intervals, for the subtoxic concentration in each cell line.

cell line	concentration estimate	95% confidence interval
HCT116	0.0040	(0.0036, 0.0046)
LoVo	0.0059	(0.0040, 0.0115)
SW480	0.0035	(0.0032, 0.0039)

3.5 Ethidium bromide (EtBr)

In this situation we have considered a exponential model with weights for the cell lines HCT116 and SW480, i.e., a regression linear model with weights for the logarithms of the data. The models obtained using the SAS system are the following:

cell line	model
HCT116	$y = \exp(4.477 - 19.474\,x) \times \varepsilon$
SW480	$y = \exp(3.722 - 24.330\,x) \times \varepsilon$

The regression coefficients R-square were 0.8584 and 0.8601, respectively. With these models we obtain the following estimates, and the following confidence intervals, for the subtoxic concentration in each cell line.

cell line	concentration estimate	95% confidence interval
HCT116	0.0183	(0.0147 , 0.0242)
SW480	0.0147	(0.0118, 0.0193)

4 Conclusions

Our exploration of the data sets (see the Appendix) demonstrates the non-homogeneity of the variances of the errors of the corresponding models, and consequently the non-adequateness of the classical least squares method to estimate the parameters. For instance, for the Ethoposide line HCT116, the variances corresponding to each concentration are 2.583, 288.250, 64.333, 376.333 and 82.583. The Levene test, that can be performed with the SPSS statistical package, rejects the equaliy of variances with $p = .029$.

Then a possible option would be to transform the dependent variable by using a suitable stabilizing variance transformation. However the usual transformations (powers, logarithms, etc.) are employed when certain specific patterns are observed between the sampling variances and their respective means. This is not the situation as can be shown in the example above by plotting the variances against their corresponding means.

Another option (those considered in this report) is to use the weighted linear regression models. These models can be implemented by using any of the more usual statistical packages such as SAS or SPSS. The ideal setting is when the weight (the inverse of the variance) for each observation is perfectly known. There are real examples where it happens (see Draper and Smith, 1981) but this is not our case here – for the data sets studied in this report the weights are estimated.

For the linear regression model we have expressed the value of the variable y (number of colonies) corresponding to the subtoxic concentration x as a linear combination of the coefficients of the regression line. The same has happened with the exponential model and the variable $\ln y$. This is the key point that has allowed us to find the confidence intervals for the subtoxic concentrations. That is, in a linear regression model (respectively, in an exponential model), this method can be applied to find confidence intervals for the value of the variable x for which the variable y (respectively, the variable $\ln y$) can be expressed as a linear combination of the coefficients of the regression line.

Appendix

In this Appendix the data used in this work are provided. The methodology used in order to obtain these results has been explained in Section 1. We express also, between parenthesis, the weight of each datum. Recall from the theory of linear models that the weights are given by the inverse of the estimated variances of the three data obtained for each concentration. Observe that some data corresponding to the number of colonies are not integer numbers. The reason is that the number of colonies was counted twice, and we have used their average.

Bleomycin

observation	concentration	HCT116	LoVo	SW480
1	0.0000	59.0 (0.0159)	26.5 (0.1967)	54.5 (0.0612)
2	0.0000	56.0 (0.0159)	29.0 (0.1967)	46.5 (0.0612)
3	0.0000	71.0 (0.0159)	31.0 (0.1967)	51.5 (0.0612)
4	0.0001	64.0 (0.0273)	32.5 (0.0221)	65.5 (0.0263)
5	0.0001	61.5 (0.0273)	23.0 (0.0221)	62.0 (0.0263)
6	0.0001	73.0 (0.0273)	*	53.5 (0.0263)
7	0.0050	58.0 (0.0058)	24.5 (0.1071)	55.0 (0.0569)
8	0.0050	76.5 (0.0058)	18.5 (0.1071)	48.0 (0.0569)
9	0.0050	*	22.5 (0.1071)	47.5 (0.0569)
10	0.0100	79.0 (0.0106)	19.0 (0.0556)	41.0 (0.0065)
11	0.0100	66.0 (0.0106)	25.0 (0.0556)	58.5 (0.0065)
12	0.0100	60.0 (0.0106)	*	*
13	0.0500	11.5 (0.0473)	6.5 (0.2449)	23.5 (0.4285)
14	0.0500	5.0 (0.0473)	3.0 (0.2449)	25.5 (0.4285)
15	0.0500	*	6.5 (0.2449)	26.5 (0.4285)

N-methyl-N-nitrosourea

observation	concentration	LoVo	SW480
1	0	34.5 (0.0317)	73.5 (0.0698)
2	0	27.0 (0.0317)	74.5 (0.0698)
3	0	23.5 (0.0317)	67.5 (0.0698)
4	1	23.0 (0.0323)	53.0 (0.0811)
5	1	34.0 (0.0323)	60.0 (0.0811)
6	1	30.0 (0.0323)	56.0 (0.0811)
7	5	20.5 (0.1791)	114.0 (0.0018)
8	5	19.5 (0.1791)	69.0 (0.0018)
9	5	16.0 (0.1791)	81.0 (0.0018)
10	10	15.5 (0.0968)	91.0 (0.0200)
11	10	21.5 (0.0968)	78.5 (0.0200)
12	10	16.5 (0.0968)	79.0 (0.0200)
13	25	16.5 (0.0424)	99.5 (0.0029)
14	25	26.0 (0.0424)	95.5 (0.0029)
15	25	19.5 (0.0424)	65.5 (0.0029)
16	50	22.0 (0.0141)	45.5 (0.0293)
17	50	20.0 (0.0141)	41.5 (0.0293)
18	50	6.5 (0.0141)	34.0 (0.0293)
19	100	1.0 (3.0000)	0.0 (0.6316)
20	100	0.0 (3.0000)	1.5 (0.6316)
21	100	0.0 (3.0000)	2.5 (0.6316)

Ethoposide

observation	concentration	HCT116	LoVo	SW480
1	0.0000	129.0 (0.3871)	81.0 (0.0012)	140.5 (0.0167)
2	0.0000	129.5 (0.3871)	28.0 (0.0012)	112.5 (0.0167)
3	0.0000	132.0 (0.3871)	35.0 (0.0012)	120.0 (0.0167)
4	0.0006	136.0 (0.0035)	39.0 (0.0033)	104.0 (0.1071)
5	0.0006	132.5 (0.0035)	70.0 (0.0033)	100.0 (0.1071)
6	0.0006	105.0 (0.0035)	40.5 (0.0033)	106.0 (0.1071)
7	0.0060	131.5 (0.0155)	38.5 (0.1081)	86.0 (0.0017)
8	0.0060	122.5 (0.0155)	43.5 (0.1081)	86.5 (0.0017)
9	0.0060	115.5 (0.0155)	44.0 (0.1081)	128.0 (0.0017)
10	0.0300	96.5 (0.0027)	33.0 (0.0116)	65.5 (0.0060)
11	0.0300	131.5 (0.0027)	43.5 (0.0116)	82.0 (0.0060)
12	0.0300	99.5 (0.0027)	25.0 (0.0116)	91.0 (0.0060)
13	0.0600	81.0 (0.0121)	27.5 (0.1165)	75.0 (0.0603)
14	0.0600	68.0 (0.0121)	28.5 (0.1165)	82.5 (0.0603)
15	0.0600	85.5 (0.0121)	33.0 (0.1165)	81.5 (0.0603)

Mitomycin C

observation	concentration	HCT116	LoVo	SW480
1	0.0000	78.5 (0.0204)	18.0 (0.0902)	*
2	0.0000	92.0 (0.0204)	12.5 (0.0902)	*
3	0.0000	88.5 (0.0204)	18.5 (0.0902)	*
4	0.0005	86.0 (0.0227)	20.0 (0.1081)	60.0 (0.0142)
5	0.0005	74.5 (0.0227)	19.5 (0.1081)	59.0 (0.0142)
6	0.0005	74.5 (0.0227)	25.0 (0.1081)	45.0 (0.0142)
7	0.0010	73.0 (0.0553)	12.5 (4.0000)	39.0 (0.0663)
8	0.0010	74.5 (0.0553)	13.0 (4.0000)	46.5 (0.0663)
9	0.0010	81.0 (0.0553)	13.5 (4.0000)	41.0 (0.0663)
10	0.0025	72.0 (0.0769)	15.0 (0.0902)	36.0 (0.0074)
11	0.0025	67.0 (0.0769)	9.0 (0.0902)	50.5 (0.0074)
12	0.0025	74.0 (0.0769)	9.5 (0.0902)	27.5 (0.0074)
13	0.0050	63.0 (0.0526)	16.0 (0.0383)	31.0 (0.0902)
14	0.0050	55.0 (0.0526)	14.5 (0.0383)	29.0 (0.0902)
15	0.0050	56.0 (0.0526)	6.5 (0.0383)	24.5 (0.0902)
16	0.0100	13.0 (0.0356)	8.0 (0.5000)	5.0 (0.0902)
17	0.0100	20.5 (0.0356)	6.0 (0.5000)	4.5 (0.0902)
18	0.0100	*	*	10.5 (0.0902)

Ethidium bromide

observation	conc.	HCT116	LN HCT116	SW480	LN SW480
1	0.000	105.5	4.66 (14.92)	23.5	3.16 (12.29)
2	0.000	85.0	4.44 (14.92)	39.0	3.66 (12.29)
3	0.000	63.0	4.14 (14.92)	38.0	3.64 (12.29)
4	0.001	100.0	4.61 (3.78)	53.0	3.97 (87.81)
5	0.001	49.0	3.89 (3.78)	46.0	3.83 (87.81)
6	0.001	133.0	4.89 (3.78)	43.0	3.76 (87.81)
7	0.010	82.5	4.41 (27.33)	29.0	3.37 (114.39)
8	0.010	102.5	4.63 (27.33)	27.0	3.30 (114.39)
9	0.010	70.0	4.24 (27.33)	32.5	3.48 (114.39)
10	0.050	26.0	3.26 (57.23)	9.0	2.20 (12.03)
11	0.050	33.5	3.51 (57.23)	15.5	2.74 (12.03)
12	0.050	31.5	3.45 (57.23)	10.0	2.30 (12.03)
13	0.100	15.0	2.71 (7.47)	5.0	1.61 (6.47)
14	0.100	13.0	2.56 (7.47)	3.0	1.10 (6.47)
15	0.100	26.0	3.26 (7.47)	6.5	1.87 (6.47)

Acknowledgements

The experimental data shown in this report has obtained thanks to a research grant by the Fundació Marató TV3. We would like to thank Professors Federic Utzet and Pere Puig for their helpful comments and suggestions during the preparation of the report.

References

Boyer, J. C., Umar, A., Risinger, J. I., Lipford, J. R., Kane, M., Ying, S., Barret, J. C., Kolodner, R. D. and Kunkel, T. A. (1995). *Microsatellite instability, mismatch repair deficency, and genetic defects in human cancer cell lines.* Cancer research, 55, 6063-6070.

Draper, N. and Smith, H. (1981). *Applied Regression Analysis* (second edition). New York: John Wiley.

Montgomery, DC. and Peck, EA. (1992). *Introduction to Linear Regression Analysis.* New York: John Wiley.

Shuterland, G. I. and Richards, R. I. (1995). Simple tandem repeats and human genetic disease. Proceedings of the National Academy of Sciences of the United States of America, 92, 3636-3641.