CREATIVE MATH. & INF.
19 (2010), No. 2, 199 - 208

Online version available at http://creative-mathematics.ubm.ro/
Print Edition: ISSN 1584 - 286X Online Edition: ISSN 1843 - 441X

Methods to evaluate the relationship between survival times

N. TODOR, G. SĂPLĂCAN and M. RĂDULESCU

ABSTRACT.

For two groups of patients selected by the value of a prognostic factor from a larger set, the common practice is to evaluate the difference by the logrank test with some variants. Now if we have a set of indexed rules that link the survival times of the two groups it is natural to choose the rules which minimize the logrank test. To find this minimum is a difficult task in the general case because the functions are not analytical ones. Our strategy is to transform the observations of one group by a set of predefined indexed rules to identify at least one rule that minimize the log rank test. If T is survival time for one group, let say basic group, we solved the problem for the sets of rules as $\{aT|a\ real\ value\}$, $\{a+T|a\ real\ value\}$and $\{a+bT|a,b\ real\ value\}$. Mathematical foundations for an algorithm and a generalization for $\{ag(T)|a\ real\ value\}$, $\{a+g(T)|a\ real\ value\}$ and $\{a+bg(T)|a,b\ real\ value\}$ with $g(.)$ an increasing function are presented. For breast cancer, some examples solved by Mathematica programs are presented.

1. Introduction

In the clinical studies implying survival analysis there are two modalities to report survival [1, 2, 3, 4, 5]. The first one supposes a report of the survival at an predefined fixed time. The values from Kaplan-Meier curves [1, 3, 4, 5, 7] or variations of that are revealed if the object of the study has a mature history and the literature has enough information about this fixed time when it exists. For example in the breast cancer studies on stages I-II which has generally a slow evolution the unwritten recommendation is to report survival at more then 10 years. Anyway the studies reporting survival at less then 10 years are generally not accepted [4, 5]. Of course all these recommendation have to be connected with the numbers of patients at risk and the follow-up of the groups of patients. If the object of the study is new enough and there is not yet a structured literature a good choice is to report at the median follow-up [2]. This attitude in reporting give us a fast image of the results when we intend to compare many groups of patients. On the other side to report survival at a fixed point could be unrealistic if the survival functions has large deviations from the exponential model. The second attitude, more elaborated uses hazard ratio as a global estimation of the differences between two groups [2]. Unfortunately this works well only on exponential survival times. In the present paper we extend this view in the case when the groups of patients suppose more complex links. We think at links where one can say that the new treatment or the favorable factor shifts, makes longer or do both shifts and makes longer the survival times. Some generalizations are suggested. Our solution finds a minimum of logrank statistics for a set of links transforming the observations of one group by the supposed link. We present the theoretical justification of the algorithm and Mathematica programs. All this were used for cases of operable breast cancers recruited at Oncology Institute "Ion Chiricuţă" from Cluj-Napoca in 1995 and 1996.

2. Notations and previous results

Let index a set of $p > 2$ patients by $1, 2, \ldots, p$ and define a selection criteria by which the patients are dichotomized. Selection criteria could be any condition or variable that could be of medical interest. For example we could think at stage, sex, treatment, etc. Our interest is to show a relationship between the group of patients which verifies a selection criteria and the group which does not verify. For simplicity, in this paper, we denote by A and B the set of indexes associated with these two groups. We are expecting a link between the survival times of group A and B. If we denote by T the survival time for A group then for group B we are expecting one of the following forms:

bT**:** with b a real value
$a + T$**:** with a a real value
$a + bT$**:** with a, b real values.

Or with $g(.)$ an increasing function

$bg(T)$**:** with b a real value
$a + g(T)$**:** with a a real value
$a + bg(T)$**:** with a, b real values.

The observation for a patient is a pair of numbers (δ, t) where t is survival time and δ the censor variable with $\delta = 0$ if the patient is alive at time t and $\delta = 1$ if the death was noticed at time t. For groups A and B, the observations are denoted by $\{(\delta_i, t_i)|i \in A\}$, $\{(\delta_j, t_j)|j \in B\}$ and from the times from both groups let be $\tau = 0 < \tau_1 < \cdots < \tau_n$ the distinct death times. For $i = 0, 1, \ldots, n$ let denote by n_{A_i}, n_{B_i} and $n_i = n_{A_i} + n_{B_i}$ the numbers of patients exposed to

Received: 09.12.2009; In revised form: 16.06.2010; Accepted: 15.08.2010
2000 *Mathematics Subject Classification.* 62N01, 62N02, 62N05.
Key words and phrases. *Relation between survival times, prognostic factors, breast cancer.*

the risk of death just before time τ_i for group A and B that is the numbers of patients alive just before τ_i; by d_{A_i}, d_{B_i} the numbers of deaths exactly at time τ_i for group A, B with $d_i = d_{A_i} + d_{B_i}$.

The logrank test [2, 3, 6, 7] to evaluate the difference between the two groups is based on statistic $\chi^2 = \frac{\left[\sum\limits_{i=1}^{n}(d_{B_i} - d_i \frac{n_{B_i}}{n_i})\right]^2}{\sum\limits_{i=1}^{n} d_i \frac{n_i - d_i}{n_i - 1} \frac{n_{A_i} n_{B_i}}{n_i^2}}$ where the terms of sums are counted only if $n_{A_i} > 0$ and $n_{B_i} > 0$.

3. Basic results

Let say for example that the group B contains the treated patients. We supposed that the effect of treatment is improved by one of the relationships from above. If we apply the transformation to the observations of untreated patients and compute the X^2 statistic, the difference we expect to be not significant. We take one by one the relationship: bT, $a + T$, $a + bT$ with a, b real values. In the case when the survival time for group B is supposed to be linked with the survival time of A by bT, we replace the observations for the group $A : \{t_i | i \in A\}$ by $\{bt_i | i \in A\}$. In this way χ^2 from the above definition is a function of b. We seek the b which minimize $\chi^2\,(b)$. The following theorem is the key of the algorithm showed in the next sections.

Theorem 3.1. *If the set* $\left\{\frac{t_j}{t_i} | i \in A, j \in B\right\}$ *has* $\nu > 1$ *distinct values denoted by* $b_1 < b_2 < \cdots < b_\nu$ *then* $\chi^2(b)$ *is constant on the intervals*

$$(0, b_1); (b_1, b_2); \ldots; (b_{\nu-1}, b_\nu); (b_\nu, \infty).$$

Proof. Let be b so that none of the points of the set $\{t_i \,| i \in A\}$ is transformed by bT in a point from the set $\{t_j \,| j \in B\}$. If we modify continuously b to left and right on the real axis so that none of the points $\{t_i \,| i \in A\}$ is transformed in a point from $\{t_j \,| j \in B\}$ then the death times preserve same order and n_{A_i}, n_{B_i}, d_{A_i}, d_{B_i} have same values for $i = 1, \ldots, n$. We have to identify the values b which give us a superposition between the sets $\{at_i \,| i \in A\}$ and $\{t_j \,| j \in B\}$ that is equivalently with the solutions of the equations $bt_i = t_j$ for $i \in A$ and $j \in B$. So we have to take the distinct values for $\frac{t_j}{t_i}$ with $i \in A$ and $j \in B$ which is equivalent with the statement of the theorem. □

Remark 3.1. $b_1 = \frac{\min_{j \in B} t_j}{\max_{i \in A} t_i}$;$b_\nu = \frac{\max_{j \in B} t_j}{\min_{i \in A} t_i}$.

Remark 3.2. In order to find the minimum value of $\chi^2(b)$ we can compute $\chi^2(b)$ on the points $\left\{\frac{b_1}{2}, b_1, \frac{b_1 + b_2}{2}, \ldots, \frac{b_{\nu-1} + b_\nu}{2}, b_\nu, b\right.$
$\left.1\right\}$.

Remark 3.3. The confidence interval for b is the set $\left\{b \,|\chi^2(b) < \chi_1^2\right\}$ where χ_1^2 is the corresponding value of the χ^2 distribution with zero degree of freedom. For example for the chosen level of significance of 0.05 the confidence interval is the set $\left\{b \,|\chi^2(b) < 3.84\right\}$. The confidence interval is an union of intervals $(0, b_1); (b_1, b_2); \ldots; (b_{\nu-1}, b_\nu), (b_\nu, \infty)$ where the extremities of these intervals might be considered.

Remark 3.4. Switching the two set of patients the solution b obtained initially becomes $\frac{1}{b}$ omitting that the solutions are in fact intervals

When the survival times T of the group A is linked to group A by relationship $a + T$, the problem makes sense only if $a + t_i > 0$ for any $i \in A$ that is $a > -t_i$ for any $i \in A$ or $a > -\min_{i \in A} t_i$. Now χ^2 is a function of a and in order to find the minimum of $\chi^2(a)$ we have the following result.

Theorem 3.2. *If the set* $\{t_j - t_i \,| i \in A, j \in B, t_j - t_i > -\min_{i \in A} t_i\}$ *has* $\nu > 1$ *distinct values denoted by* $a_1 < a_2 < \cdots < a_\nu$ *then* $\chi^2(a)$ *is constant on the intervals* $(-\min_{i \in A} t_i, a_1)\,; (a_1, a_2)\,; \ldots; (a_{\nu-1}, a_\nu)\,; (a_\nu, \infty)$.

Proof. The proof is similar with that from the Theorem 3.1 but the equations $bt_i = t_j$ for $i \in A$ and $j \in B$ have to be replaced by equations $a + t_i = t_j$ for $i \in A$ and $j \in B$ and add the constraints $a + t_i > 0$ for any $i \in A$. □

Remark 3.5. $a_\nu = \max_{j \in B} t_j - \min_{i \in A} t_i$.

Remark 3.6. In order to find the minimum of $\chi^2(a)$ we compute $\chi^2(a)$ for $a =$

$$\frac{-\min_{i \in A} t_i + a_1}{2}, a_1, \frac{a_1 + a_2}{2}, \ldots, \frac{a_{\nu-1} + a_\nu}{2}, a_\nu, a_\nu + 1.$$

Remark 3.7. The confidence interval is similar with that derived from Remark 3.3.

Remark 3.8. If in Theorem 3.1 we can switch the groups of patients for the model $a + T$ this is not true and we must solve two problems: group A against group B and group B against group A.

We consider now a fixed real value b, $b \neq 0$. We have the following statement.

Theorem 3.3. *For a fixed real number b, if the set*

$$\left\{ t_j - bt_i \middle| i \in A, j \in B, t_j - bt_i > \max_{i \in A} (-bt_i) \right\}$$

has $\nu > 1$ distinct values denoted by $a_1 < a_2 < \cdots < a_\nu$ then $\chi^2(a)$ is constant on the intervals

$$\left(\max_{i \in A} (-bt_i), a_1 \right); (a_1, a_2); \ldots; (a_{\nu-1}, a_\nu); (a_\nu, \infty).$$

Proof. The proof is almost identical with that for Theorem 3.2. □

Remark 3.9. $a_\nu = \max_{j \in B} t_j - \min_{i \in A} bt_i$.

Remark 3.10. *In order to find the minimum of $\chi^2(a)$ we compute $\chi^2(a)$ for $a =$*

$$\frac{\max_{i \in A}(-bt_i) + a_1}{2}, a_1, \frac{a_1 + a_2}{2}, \ldots, \frac{a_{\nu-1} + a_\nu}{2}, a_\nu, a_\nu + 1.$$

Remark 3.11. For $b = 1$ we obtain again Theorem 3.2.

Remark 3.12. The confidence interval is similar with that derived from Remark 3.3.

For the general relationship $a + bT$ with a, b real values we describe an algorithm by which we find the minimum of $\chi^2(a, b)$. The algorithm is based on the following theorem.

Theorem 3.4. *Let $\nu_b > 1$ distinct values of the set*

$$\left\{ \frac{t_{j_2} - t_{j_1}}{t_{i_2} - t_{i_1}} | i_1, i_2 \in A, i_1 < i_2; j_1, j_2 \in B, j_1 < j_2 \right\}$$

be denoted by $b_1 < b_2 < \cdots < b_{\nu_b}$. Applying the results of Theorem 3.3 for

$$b = \frac{b_1}{2}, b_1, \frac{b_1 + b_2}{2}, \ldots, \frac{b_{\nu_b - 1} + b_{\nu_b}}{2}, b_{\nu_b}, b_{\nu_b} + 1$$

we obtain all values of $\chi^2(a, b)$.

Proof. We are interested only with the points of the set $\{a + bt_i \,| i \in A\}$ which are identical with some points from the set $\{t_j \,| j \in B\}$. This can be expressed as the set of equations $a + bt_i = t_j$ for $i \in A$ and $j \in B$. These equations determine in the plane (b, a) a set of polygonal zones on which interior $\chi^2(a, b)$ is constant because the order of death times and the corresponding numbers of exposed and death patients are constant in the computation. The vertexes of these polygons are computed solving a system formed with a pair of equations $a + bt_i = t_j$.

Let be $i_1, i_2 \in A$ with $i_1 < i_2$ and $j_1, j_2 \in B$ with $j_1 < j_2$. The equations $a + bt_{i_1} = t_{j_1}$ and $a + bt_{i_2} = t_{j_2}$ give us the point

$$(b, a) = \left(\frac{t_{j_2} - t_{j_1}}{t_{i_2} - t_{i_1}}, \frac{t_{i_2} t_{j_1} - t_{i_1} t_{j_2}}{t_{i_2} - t_{i_1}} \right).$$

We consider now the set of points

$$\left\{ \left(\frac{t_{j_2} - t_{j_1}}{t_{i_2} - t_{i_1}}, \frac{t_{j_2} t_{j_1} - t_{i_1} t_{j_2}}{t_{i_2} - t_{i_1}} \right) \right\}$$

with $i_1, i_2 \in A, i_1 < i_2; j_1, j_2 \in B, j_1 < j_2$ and let $\nu_b > 1$ be the distinct values of the first coordinate in ascending order $b_1 < b_2 < \cdots < b_{\nu_b}$. Let fix two consecutive values b_k, b_{k+1}. The vertical band between b_k and b_{k+1} contains adjacent zones where $\chi^2(a, b)$ is constant. Each zone has at most an upper and a lower neighbour zone.

For a $b' \in (b_k, b_{k+1})$ we denote by $a_1 < a_2 < \cdots < a_{\nu_a}$ the coordinate a where a vertical line in b' cut the lines $a + b't_i = t_j$ for $i \in A$ and $j \in B$. As in Theorem 3.3 the values of $\chi^2(a, b)$ are constant for all values of a from the set

$$\left\{ \left(\max_{i \in A} (-b't_i), a_1 \right); (a_1, a_2); \ldots; (a_{\nu_a - 1}, a_{\nu_a}) \right\}.$$

Let observe that the values of $\chi^2(a, b')$ are same for any value of b' in the interval (b_k, b_{k+1}). We adopt same strategy as before so that we compute $\chi^2(a, b')$ for

$$a = \frac{\max_{i \in A}(-b't_i) + a_1}{2}, a_1, \frac{a_1 + a_2}{2}, \ldots, \frac{a_{\nu_a - 1} + a_{\nu_a}}{2}, a_{\nu_a}, a_{\nu_a} + 1.$$

For simplicity, in the general case we take for b' similar values as in Theorem 3.1 that is

$$\frac{b_1}{2}, b_1, \frac{b_1 + b_2}{2}, \ldots, \frac{b_{\nu_b - 1} + b_{\nu_b}}{2}, b_{\nu_b}, b_{\nu_b} + 1$$

and we obtain all values for $\chi^2(a, b)$. We have proved so the theorem. □

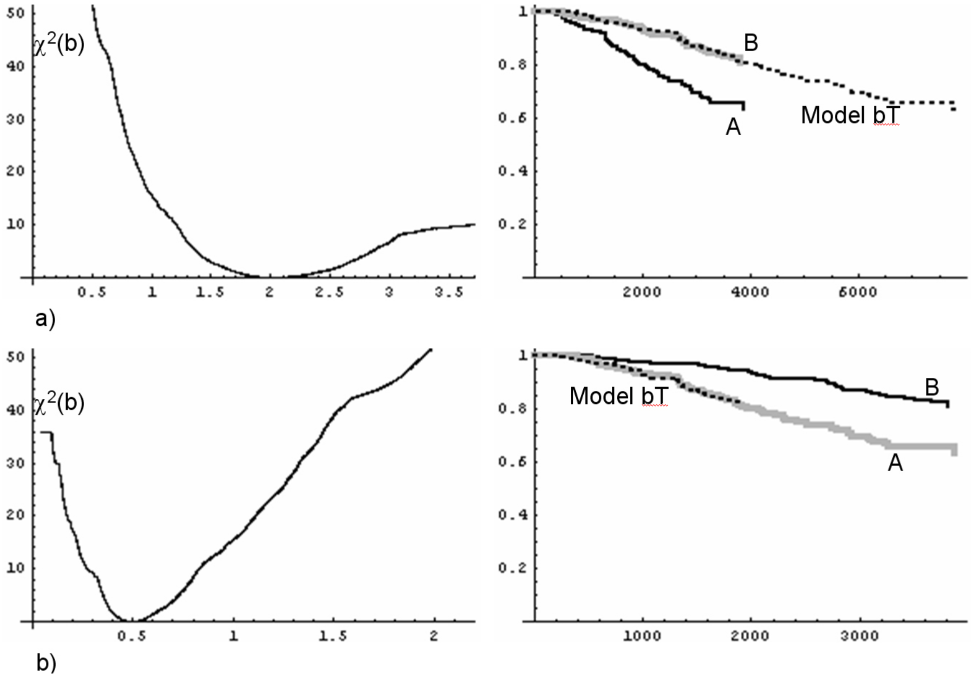


FIGURE 1. Model bT, graphic of $\chi^2(b)$ and survival curves for a) A versus B ($b = 2.0097$) and b) B versus A ($b = 0.497424$).

Remark 3.13. The algorithm emerging from Theorem 3.4 can be improved if we add the constraints $a + bt_i > 0$ for $i \in A$. Also we are not interested with the zone where by transformation $a + bt$ the set $\{t_i \,|i \in A\}$ is placed at left or right of the set $\{t_j \,|j \in B\}$ because there $\chi^2(a, b)$ is maximum. That can be written $a + bt_i < \min_{j\in B} t_j$ for $i \in A$ and $a + bt_i > \max_{j\in B} t_j$ for $i \in A$.

Remark 3.14. All proofs of the stated theorems are maintained if we consider instead of the set $\{t_i \,|i \in A\}$ the set $\{g(t_i) \,|i \in A\}$ where $g(.)$is an increasing positive function.

Remark 3.15. The confidence interval is almost similar with that derived for Theorems 3.1 and 3.2 but now we have the set $\left\{(a, b) \,|\chi^2(a, b) < \chi^2_1\right\}$ with χ^2 distribution with one degree of freedom. Evidently the *confidence interval* in this case is an union of zones as defined in Theorem 3.4.

4. EXAMPLES AND PRACTICAL CONSIDERATIONS

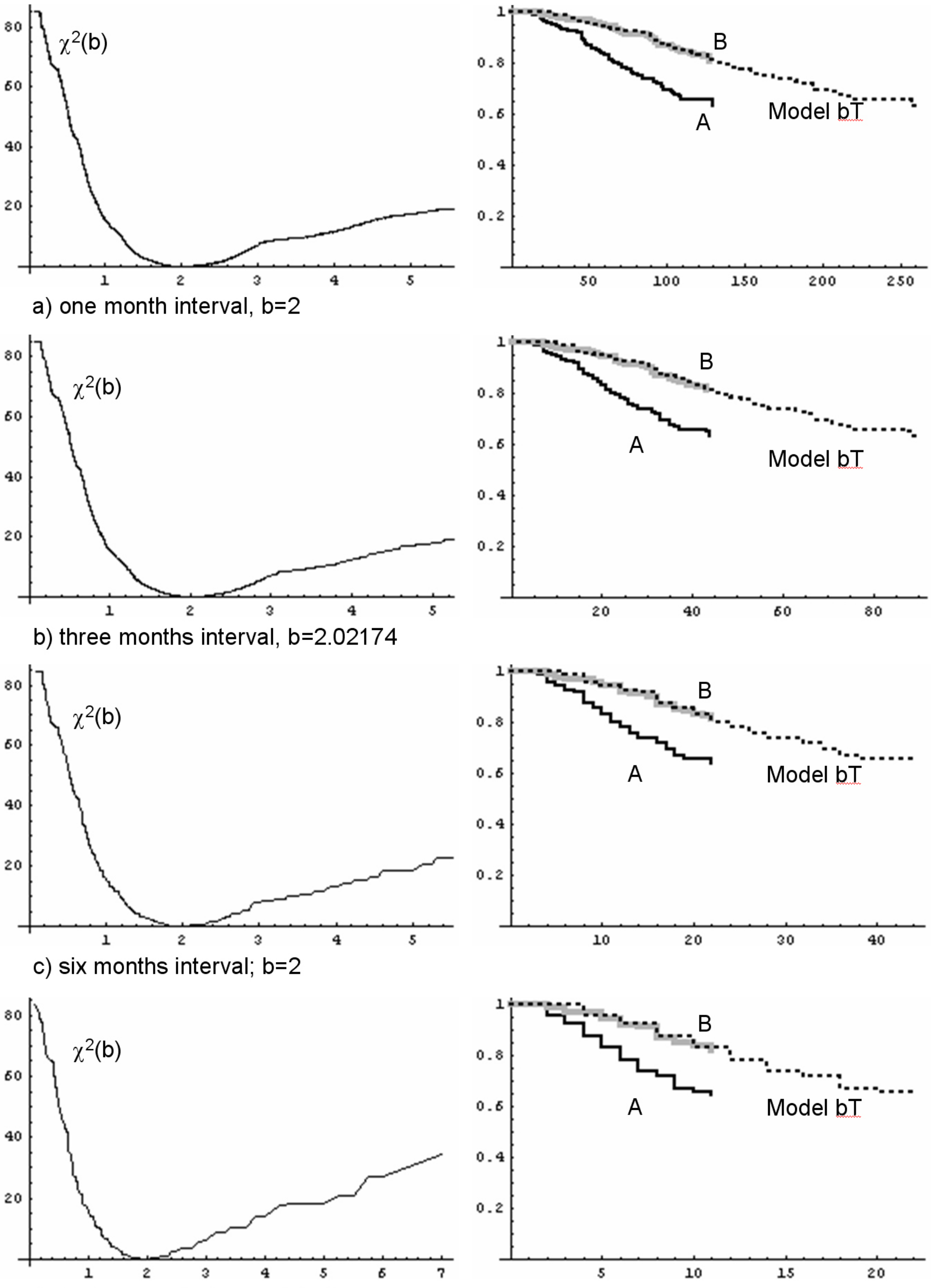
The idea of the present algorithm with the results from below has occurred in 2005 when one started to analyze the data from Oncology Institute concerning the operable breast cancer. In [8] we can see detailed medical conclusions of this report. The algorithms derived from the presented theorems are simple to program. At each iterations we compute the values of $\chi^2(b)$, $\chi^2(a)$, or $\chi^2(a, b)$ as the theorems ask. We have chosen Mathematica [9] as programming media for his graphic and mathematical capabilities. The authors can send you all the programs they used on these examples. The examples are related to data from [8] where the two groups are coming from our patients which after surgery were investigated for positive nodes. Group A contains 175 patients with positive nodes and group B 193 patients with negative nodes. For the model bT treated in Theorem 3.1 we obtain $b = 2.0097$ with $\chi^2(b) = 3.42169 \times 10^{-9}$. We can say that the survival times for the patients with negative nodes are two times longer. The graphic of $\chi^2(b)$ and the survival curves are presented in Figure 1a where black line is for group A, gray line for group B and dotted line for the result of transformation of group A.

If we switch the groups of patients the results are presented in Figure 1b. The value of $b = 0.497424$ is very close of $\dfrac{1}{2.0097}$. We can say that the survival times for the group with positive nodes is half of those with negative nodes.

When we take the observation times in days we need 56273 iterations. Naturally we ask ourselves if we alter the solution when we take an actuarial approximation for example on one month, 3 months, 6 months or 12 months with the objective to lower the number of iterations. Using the Mathematica program of the authors to minimize $\chi^2(b)$ we obtain for b the values: 2., 2.02174, 2, 2. The Figure 2 represents the graphical results similar to Figure 1a but for length of intervals of one month, 3, 6 and 12 months.

One can say that the solutions did not change much but the number of iterations did from 56273 iterations when the observation times are in days at 7075 for one month, at 1943 for 3 months interval length, at 563 for 6 months interval length and at 159 for 12 months interval length.

For the model $a + T$ with the observations times in days we have 12331 iterations with solution $a = 1856$ days that is 61.9 months; for observations on interval of length one month 797 iterations with solution $a = 62$ months; for



a) one month interval, b=2

b) three months interval, b=2.02174

c) six months interval; b=2

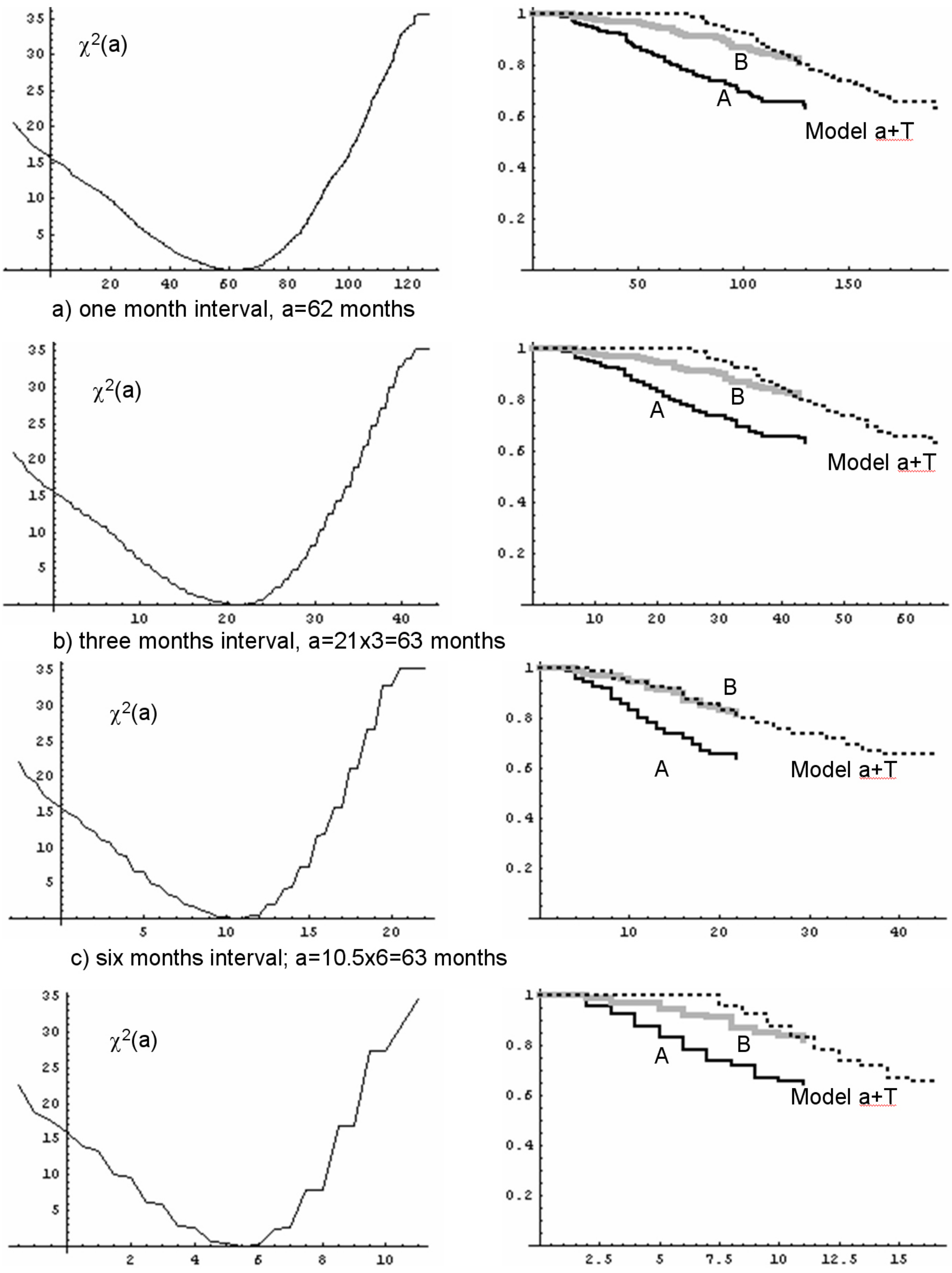
d) twelve months interval, b=2

FIGURE 2. Model bT, graphic of $\chi^2(b)$ and survival curves with actuarial approximation for one, three, six and twelve months.

observations on intervals of length 3 months 167 iterations with solution $a = 21 \times 3 = 63$ months; for observations on intervals of length 6 months 85 iterations with $a = 10.5 \times 6 = 63$ months; for observations on intervals of length 12 months 43 iterations with $a = 5.5 \times 12 = 66$ months. Figure 3 presents the graphical results for $\chi^2(a)$ and the corresponding survival curves.

We have also for this model a powerful stability near 62 months that is near 5 years and the number of iterations was lowered. We can conclude that the patients with negative node status has an extra 5 years in survival time. Other three important factors for breast cancer are listed in the table 1. We took as base the most favorable factor and we estimate the gain or loss in survival. For example stage IIa is 1.65 times worst than stage I; stage IIb 2.75 times; stage IIIa in the case of multiplicative model bT. Additive model a+T produces a loss of 51.5 months for stage IIa; 84 for stage IIb and 84.5 for IIIa comparing with stage I.

The algorithm from Theorem 3.4 has no practical problems for few tenth of patients on each group. Even here for values moderately higher we need 1000 iterations for b and further to each b we have to find all values of a



a) one month interval, a=62 months

b) three months interval, a=21x3=63 months

c) six months interval; a=10.5x6=63 months

d) twelve months interval, a=5.5x6=66 months

FIGURE 3. Model $a + T$, graphic of $\chi^2(a)$ and survival curves with actuarial approximation for one, three, six and twelve months.

as in Theorem 3.3 and so the total number is up to 1 000 000. If we have on average 3-4 seconds for an iterations on a Pentium IV at 3000 Mhz we obtain the solution in 100 hours that is 4 days. We need other methods to identify the solution. The first is suggested by the previous examples where we worked with intervals of one month, 3, 6 and 12 months. For intervals of 12 months the solution of the pairs (a, b) derived from Theorem 3.4 is $\{\{0.76, 6.47\}, \{0.78, 6.44\}, \{0.79, 6.42\}, \{0.80, 6.40\}, \{0.80, 6.4\}, \{0.80, 6.40\}\}$ and for the intervals of length 6 the solutions $\{\{0.45, 0.75\}, \{0.45, 0.73\}, \{0.45, 0.70\}, \{0.45, 0.68\}, \{0.46, 0.65\}\}$. Unfortunately $\chi^2(a, b)$ for all these values give us large values. The explanation resides perhaps in the fact that we can accept the two models from above and $a + bT$ is not the best option.

REFERENCES

[1] Breslow, N. E. and Day N.E. *Statistical Methods in cancer research: volume II – The design and analysis of cohort studies*, IARC scientific publications, Lyon, 1987

[2] Collett, D., *Modelling survival data in medical research*, Chapman & Hall / CRC, London, 2003

[3] Crowley, J. and Ankerst, D. P., *Handbook of statistics in clinical oncology, 2-nd edition*, Chapman and Hall, London, 2006

TABLE 1. Results for operable breast cancer.

Categories	Model bT	Model $a + T$
Clinical TNM stage		
I	1	0
IIa	1.65	51.5
IIb	2.75	84
IIIa	2.65	84.5
Pathological node status		
pN0	1	0
pN1-3	2.58	84
pN4-10	2.9	93.5
pN11+	4.7	115.5
Pathological tumor status		
pT1	1	0
pT2	1.6	49.5
pT3	2.1	76.5

[4] DeVita, V. T., Hellman, S. and Rosenberg, S. A., *Cancer principles & Practice of Oncology, 7-th edition*, Lippincot Williams & Wilkins, New York, 2004

[5] Esteve, J., Benhamou, S. and Raymond, L., *Methodes statistiques en epidemiologie descriptive*, INSERM, Paris, 1993

[6] Gill, R. D. and Keiding, N., *Statistical Models Based on Counting Processes*, Springer Verlag, London, 1993

[7] Hill, C., Com-Nouguee, C., Kramar, A. et al. *Analyse statistiquedes donees de Survie*, Flamarion, Paris, 1990

[8] Vitoc, C., Ghilezan, N., Todor, N. et al. *Ten Years Results for Operable Breast Cancer: Cancer Institute Cluj Experience – 1995-1996 Period. Abstract at http:www.srrom.ro*, Radioterapie si Oncologie Medicala, **13** (2007), No. 2, 219-228

[9] Wolfram, S., *Mathematica, A System for Doing Mathematics by Computers*, Addison-Wesley, New York, 1991

CANCER INSTITUTE "ION CHIRICUŢĂ" CLUJ-NAPOCA
DEPARTAMENT OF BIOSTATISTICS AND INFORMATICS
REPUBLICII 34-36, 400015, CLUJ-NAPOCA, ROMANIA
E-mail address: todor@iocn.ro

APPLIED INFORMATION COMPANY
REPUBLICII 107, 400489, CLUJ-NAPOCA, ROMANIA
E-mail address: gsaplacan@yahoo.com

UNIVERSITY OF MEDICINE AND PHARMACY "IULIU HAŢIEGANU" CLUJ-NAPOCA
EMIL ISAC 13, 400023, CLUJ-NAPOCA, ROMANIA
E-mail address: dan_rad31@yahoo.com