REVSTAT – Statistical Journal
Volume 13, Number 1, March 2015, 63–78

LONGITUDINAL ANALYSIS OF TUMOR MARKER CEA OF BREAST CANCER PATIENTS FROM BRAGA'S HOSPITAL

Authors: Ana Borges
– CIICESI, ESTGF — Polytechnic Institute of Porto, DMA-ECUM, Minho's University, Portugal
aib@estgf.ipp.pt

Inês Sousa
– Department of Mathematics and Applications, Minho's University, Centre of Mathematics (CMAT), Minho's University, Portugal
isousa@math.uminho.pt

Luís Castro
– Senology Unit of Braga's Hospital, Portugal
luis.castro@hospitaldebraga.pt

Received: October 2014 Revised: November 2014 Accepted: December 2014

Abstract:

- Allied to an epidemiological study of population of the Senology Unit of Braga's Hospital that have been diagnosed with malignant breast cancer, we describe the progression in time of repeated measurements of tumor marker Carcinoembryonic antigen (CEA). Our main purpose is to describe the progression of this tumor marker as a function of possible risk factors and, hence, to understand how these risk factors influences that progression. The response variable, values of CEA, was analyzed making use of longitudinal models, testing for different correlation structures. The same covariates used in a previous survival analysis were considered in the longitudinal model. The reference time used was time from diagnose until death from breast cancer. For diagnostic of the models fitted we have used empirical and theoretical variograms. To evaluate the fixed term of the longitudinal model we have tested for a changing point on the effect of time on the tumor marker progression. A longitudinal model was also fitted only to the subset of patients that died from breast cancer, using the reference time as time from date of death until blood test.

Key-Words:

- *breast cancer; carcinoembryonic antigen; longitudinal models.*

AMS Subject Classification:

- 62P10, 62J10.

1. INTRODUCTION

Oncological diseases are the second highest cause of death in Portugal, and they have a big social impact in patients and their families [12]. In Europe breast cancer is the tumor with highest incidence in women [1]. In Portugal there are not many published studies on breast cancer. However, Pinheiro *et al.* (2003) ([12]) refer that, since 1995, mortality due to breast cancer has been decreasing in Portugal. They argue that this improvement is a consequence of earlier diagnostic and better quality of treatment.

According to results presented by the European Cancer Observatory [5], the estimated incidence for Breast Cancer in Portuguese women in 2012 is 85.6% and the estimated mortality rate due to this type of cancer is 18.4%, both values are quite lower than the European average (94.2% and 23.1% respectively). At the moment, the existing recommendations and guidelines from the National Health Service are mainly based on European studies. However, it is not clear that the behavior of the disease is similar among European countries. Therefore, it is of great importance the continuous investment on statistical and epidemiological studies in oncological diseases for understanding the progression of the disease in Portugal.

This study aims to answer at least some of the questions on a specific Portuguese population, particularly the population of the Senology Unit of Braga's Hospital, located in the north of Portugal, that were diagnosed with malignant breast cancer.

The tumor marker Carcinoembryonic antigen (CEA) is usually used for therapy monitoring in advanced disease ([6]), although recent reports, *e.g.* Fiorella *et al.* (2001) ([6]), discourage its routine use because of low sensitivity. The authors conclude that its use should be considered as an inefficient method of follow-up evaluation for breast cancer patients, and it provides no additional value when used in combination with another tumor marker Carcinoma Antigen 15-3 (CA 15-3). Nevertheless, as Sturgeon *et al.* ([16]) point out, on occasion, it can be informative when levels of CA 15-3 remain below the cutoff point.

Since it is a usual medical procedure to be alert for possible tumor recurrence in the case of detecting a rise in levels of this marker above a certain reference value, our main purpose is to describe the progression of this tumor marker, on patients who were followed and treated in this Unit, as a function of possible risk factors. We intend to estimate on average the time to the increase of this tumor marker, and to characterize the degree of heterogeneity between patients.

2. METHODOLOGY

2.1. Motivation and data set

Data were collected directly from the medical records of each patient, listed in the computer system of Braga's Hospital — Glintt HS. We therefore have access to baseline and clinical history of each patient (a roll of information such as diagnosis; pre-surgery, post-surgery, group meetings; follow-up and medical exams). The authorization to collect and use of senology data was approved by the Ethical Committee of Hospital de Braga.

From the information gathered in the medical reports we were able to collect more than 50 variables that can be grouped into two categories: (i) explanatory variables at individual level, which are a group of demographic characteristics that include a set of prognostic factors reported by Rodrigues (2011) ([14]), for example: age, menopause, age at first full term pregnancy; (ii) explanatory variables at tumor level, that include characteristics of the tumor, some of them important prognostic factors which were already reported in the literature and resumed by Fitzgibbons *et al.* (2000) ([7]) and Cianfrocca and Goldstein (2004) ([3]), such as TNM stage, histological type of tumor, hormonal receptors or vascular or lymphatic invasion, among others.

We collected data from 577 female patients diagnosed with a malignant tumor in the period of 2008 until 2012 (or before, but alive at 2008 and all patients at follow up on group meetings at 2008). Patients at follow up on group meetings were diagnosed as late as 1998. Patients' age at the time of diagnosis varies between 20 and 89 years. However patients with no information regarding tumor markers CEA were excluded for the present analysis, as well patients with no follow up information. We handled all missing values as missing completely at random ([10]).

For the longitudinal analysis of the tumor marker CEA, we considered data of 532 patients. Since 19 patients had bilateral breast cancer, and bilateral breast cancer is treated as independent case in this study, it translates into a total number of 551 cases analyzed. The total number of deaths from breast cancer is 54. There were 4166 measurements of tumor marker CEA, with a number of observations per patient varying between 1 and 23 measurements, as shown in Figure 1. The median number of measurements per person is 7.

It is an unbalanced study for the tumor marker, since patients measurements were not made at the same moment.

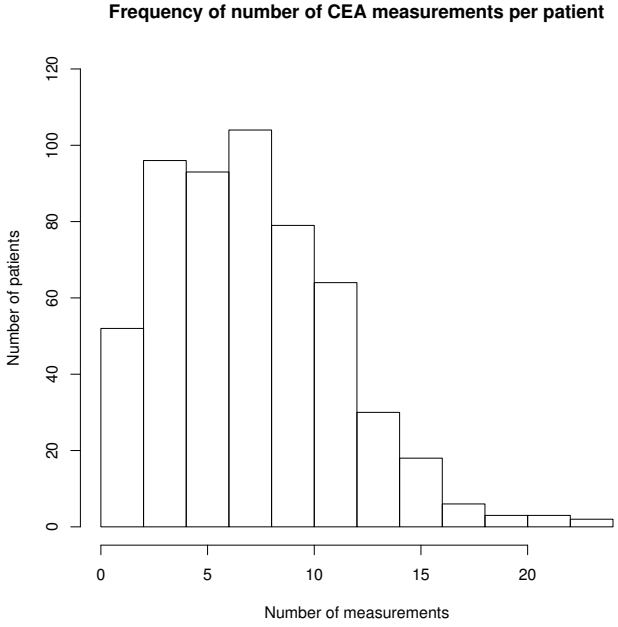

Figure 1: Histogram for the number of measurements per patients for tumor marker CEA.

2.2. Statistical methodology

The response variable, value of CEA, was analyzed making use of longitudinal models as defined in Diggle et al (2002) ([4]), where different correlation structures were tested.

The same covariates used in the survival model, previous adjusted in an earlier study ([2]), were tested in the longitudinal model fitted. The reference time used was time, in years, since diagnose of breast cancer. We have used the reference value of 5,0 ng/mL ([14]) for the response variable. According to the usual medical procedures, physicians stay alert to a possible recurrence of breast cancer for patients that present values of CEA above this reference value.

In general, we denote each patient in this analysis by the subscript $i = 1, ..., n$. Repeated tumor marker measurements for each patient i, at corresponding time t_{ij}, are denoted by Y_{ij}, where $j = 1, ..., m_i$. Note that for this particular study, measurement times are not common to all subjects (unbalanced study). Let $N = \sum_i^n m_i$, be the total number of measurements in the data set.

For the analysis, we began with an exploratory analysis and point estima-

tion by modeling a saturated ordinary least square (OLS) ([4]) model with the variables that had shown significant effect on patients' survival, given by:

$$Y_{ij} = \mu_{ij} + \varepsilon_{ij} \ , \tag{2.1}$$

where $E[Y_{ij}] = \mu_{ij}$ and ε_{ij} are N independent and identically distributed (i.i.d.) realizations of $N(0, \xi^2)$.

Since the OLS model assumes independence between any two measurements, from the same or different subject, it is important to consider different models in the context of longitudinal analysis, that take into account the correlation that usually exists in the measurements of the same subject.

A longitudinal model was also fitted only to the subset of patients that died, using the reference time, in years, from blood tests until date of death.

To model the correlation structure for each model we analyzed the empirical variogram of OLS residuals from the saturated model for the mean response ([4]). These patterns suggested the existence of variability between subjects (as random effects), and a possible variability within subjects (serial correlation). Hence, maintaining the same mean structure we compared two nested models with different covariance structures with three components, such as: (i) random effects, exponential serial correlation and measurement error; (ii) random effects, Gaussian serial correlation and measurement error.

In many medical studies it is important to consider not only random effects but also a possible variability within subjects as it may have important medical implications. In fact, Liang and Zeger (1986) ([9]) alert that treating the correlation as a nuisance may be less appropriate when the time course of the outcome for each subject is of primary interest or when the correlation itself has scientific relevance.

Both longitudinal models are given by:

$$Y_{ij} = \mu_{ij} + U_i + W_i(t_{ij}) + Z_{ij} \ , \tag{2.2}$$

where U_i are n i.i.d. realizations of $N(0, \nu^2)$, representing the random effects at individual level, $W_i(t_{ij})$ is a continuous time Gaussian Process with $E[W_i(t_{ij})] = 0$ and $\mathrm{Var}(W_i(t_{ij})) = \sigma^2$ and, Z_{ij} are N i.i.d. realizations of $N(0, \tau^2)$, representing the measurement error (variability non specified).

To model the fixed term of the longitudinal model, μ_{ij}, we have tested for a changing point δ on the effect of time on the tumor markers. In practice, the changing point is the moment where there is an alteration on the slope of the linear marker's progression, on average. Considering δ the changing point, we

have $E[Y_{ij}] = \mu_{ij}$ with:

$$\mu_{ij} = \begin{cases} X_{ij}\beta + \alpha_1 t_{ij}\,, & \text{if } t_{ij} < \delta\ , \\ X_{ij}\beta + \alpha_2(t_{ij} - \delta)\,, & \text{if } t_{ij} \geq \delta\ , \end{cases} \tag{2.3}$$

where X_{ij} represents the vector of covariates, β the vector of unknown regression coefficients, α_1 and α_2 the coefficients representing the slope before and after the changing point, respectively.

For parameter estimation we use the maximum likelihood method, whose associated likelihood function is given by:

$$L(\theta; Y) = \prod_{i=1}^{n}\prod_{j=1}^{m_i} \frac{1}{2\pi|V_{ij}|} \exp\left\{-\left(\frac{1}{2}\right)(y_{ij} - \mu_{ij})V_{ij}^{-1}(y_{ij} - \mu_{ij})^T\right\}\ , \tag{2.4}$$

where V_{ij} is the variance/covariance positions on the variance/covariance matrix of all data.

We then conducted a backward elimination to delete variables not significant, until the mean structure was well defined with only significant covariates.

What distinguishes these two longitudinal models is how two different realizations of W_i are correlated in time. That is, if we consider the correlation among $W_i(t_{ij})$, let say between $W(t)$ and $W(t-u)$, determined by the autocorrelation function $\rho(u)$, we will have for the REE model $\rho(u) = \exp(-\frac{1}{\phi}.|u|)$, and for the REG model $\rho(u) = \exp(-\frac{1}{\phi}.u^2)$, where ρ is the range parameter that specifies the rate at which the correlation stables.

The validation of the correlation structure was made by graphical comparison between the empirical variogram and the theoretical fitted ones, and comparing their maximized log likelihood values.

The variogram ([4]) of a stochastic process $Y(t)$ is given by:

$$V(u) = \frac{1}{2}\,\mathrm{Var}\big\{Y(t) - Y(t-u)\big\}\ , \qquad u \geq 0\ . \tag{2.5}$$

For a stationary process, the autocorrelation function, $\rho(u)$, and the variance of $Y(t)$, σ^2, are related by:

$$\gamma(u) = \sigma^2\{1 - \rho(u)\}\ . \tag{2.6}$$

The estimation of the empirical variogram is based on the calculation of the observed half-squared-differences between pair of residuals, $\nu_{ij} = \frac{1}{2}(r_{ij} - r_{ik})^2$, and the corresponding time-differences, $u_{ijk} = t_{ij} - t_{ik}$, where $r_{ij} = Y_{ij} - \mu_{ij}$, and $j < k = 1, ..., m_i$.

The autocorrelation function at any lag u is estimated from the sample variogram by:

$$\hat{\rho}(u) = 1 - \frac{\hat{\gamma}(u)}{\hat{\sigma}^2} \ , \tag{2.7}$$

where $\hat{\gamma}(u)$ is the average of all the ν_{ij} corresponding to that particular value of u, and $\hat{\sigma}^2$ is the estimated process variance.

The entire analysis was performed using R software ([15]), in particular making use of both *nlme* ([11]) and *JoineR* ([13]) packages.

3. RESULTS

Since the normality assumption of the response variable failed, we used a log-transformation of the tumor marker CEA values. It is, in fact, a usual transformation in biological markers. The spaghetti plot (Figure 2) presents the progression of the CEA values for each patient, against the reference, and the non parametric smooth spline line, indicating the average trend of progression. The smooth spline suggests that, on average, the marker progression stays below the reference value with a non accentuated slope in its increase. However, it is possible to see that there are individuals with values above the reference value of log (5.0) ng/mL. Nevertheless a linear modeling approach appears to be reasonable. Also, it does not point out to the existence of a changing point in its progression in time.

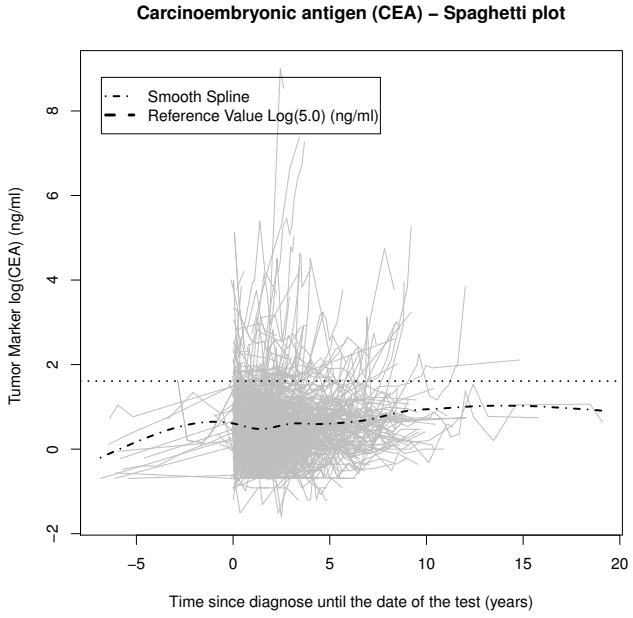

Figure 2: Spaghetti plot for tumor marker CEA values.

In fact, after fitting several saturated parametric models considering various changing points' values, its existence was not significant in the mean time trend of the tumor progression.

Table 1 presents the estimated parameters of the fitted longitudinal model that best represent the tumor marker progression in time, and compares the estimates to those obtained by fitting the simple OLS model and the respective log Likelihood values.

Table 1: Estimated parameters values for General Linear Model and Longitudinal Model.

	OLS Model		**REE Model**	
	Estimate	**p-value**	**Estimate**	**p-value**
Intercept	0.0689	0.7170	0.7355	0.0405
Time	−0.1304	<0.0001	−0.1049	0.0038
Tumor stage (III or IV)	0.2132	<0.0001	0.2655	0.0038
Primary tumor size (Tx or T1 or T2 or T3 or Tis)	−0.2063	0.2660	−1.0383	0.0023
Age at diagnosis	0.0095	<0.0001	0.0117	<0.0001
Venous vascular invasion (Yes) * Time	0.1355	<0.0001	0.0967	0.0175
Tumor degree (G3) * Time	0.1281	<0.0001	0.1179	<0.0001
Estrogen receptor expression (positive) * Time	0.1548	<0.0001	0.1455	<0.0001
$\hat{\nu}^2$			0.2849	
$\hat{\sigma}^2$			0.3295	
$\hat{\phi}$			2.1912	
$\hat{\tau}^2$			0.0239	
$\hat{\xi}^2$	0.6770			
Log Likelihood	−3792.429		−1853.366	

The fixed part of the longitudinal model, which describes the mean progression of the marker, is composed by the following significant covariates on the intercept component of the model: tumor stage (0/I/II versus III/IV), primary tumor size (Tx/T1/T2/T3/Tis versus T4), and age at diagnosis. The intercept component of the model, in this particular case, means that a patient with a tumor stage of 0, I or II, a T4 primary tumor size at an earlier age of diagnosis will start the progression of the tumor marker with a value of 0.7355, on a logarithmic scale.

A patient with a tumor on stage III or IV implicates an increasing of the log value of the tumor marker by an increment of 0.2655, comparing to those

with a tumor on stage 0, I or II. Also, a tumor that presented a primary tumor size different from the classification T4 has a decrease in the starting point of the marker value by an increment of -1.0383. The age at diagnosis affects the log value of the marker at a rate of 0.0117 per year of age at diagnosis.

The covariates that affect the slope (-0.1049) of the linear progression of the tumor are: images of vascular invasion (Yes versus No), Bloom-Richardson degree of differentiation (Gx/G1/G2 versus G3) and estrogen receptors expression (Positive versus Negative).

According to the estimated values, cases that present a venous vascular invasion of the tumor, a tumor degree G3 and a positive estrogen receptor expression increase the progression slope at a rate of, respectively, 0.0967, 0.1179 and 0.1455.

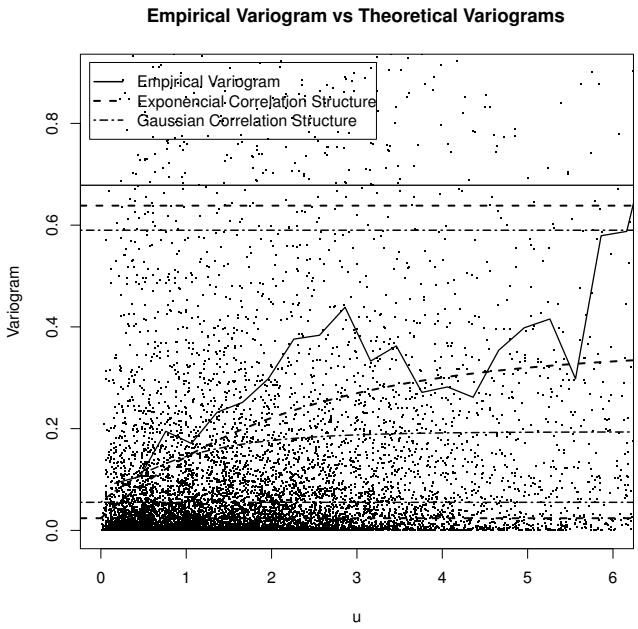

Figure 3: Superposition of empirial variogram and theoretical variogram.

The correlation structure that best represents the variability of the data is, in fact, the one that incorporates random effects at individual level with $\hat{\nu}^2 \approx 0.2849$, an exponential correlation structure to describe the variability within patients with $\rho(u) = \exp(\frac{-1}{2.1912}.|u|)$ and $\hat{\sigma}^2 \approx 0.3295$, and a measurement error with variance $\hat{\tau}^2 \approx 0.0239$. That fact can be easily accessed by the superposition of the theoretical fitted variogram of both exponential and Gaussian correlation structures with the empirical variogram (Figure 3).

When fitting the saturated general linear model for the subset of patients who died from breast cancer, we detected a changing point at 2 years before death. The smooth spline of the spaghetti plot (Figure 4) is consistent with that result and informs a transposition of the reference value nearly after that changing point.

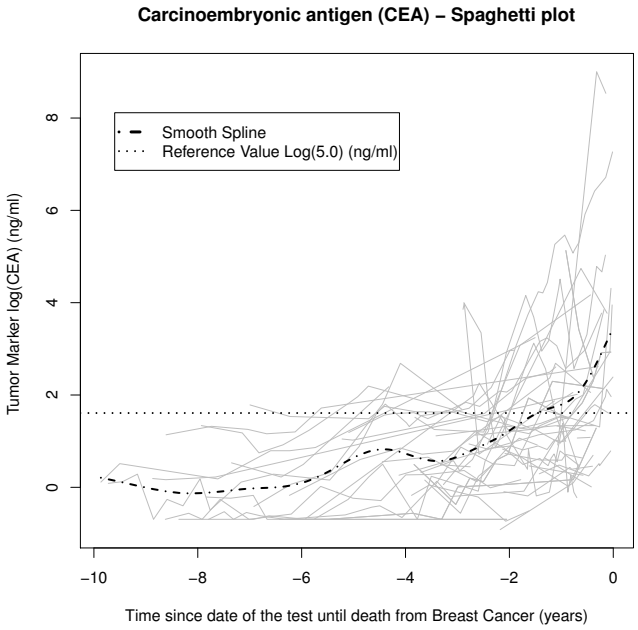

Figure 4: Spaghetti plot for tumor marker CEA values of patients who died from breast cancer.

Note that, as we are analyzing the marker values from date of blood tests until death, we are dealing with duration at a negative scale.

Table 2 summarizes and compares the estimated parameters for the longitudinal model which best fitted the data with those of the general linear model (OLS Model). As expected, the presence of venous vascular invasion has an increasing effect on the average CEA linear progression in time, as it is related to a worst prognostic case in the previous survival analysis ([2]).

Contradictory results are the decreasing effect of a bilateral type of tumor and the presence of lymphatic invasion and the increasing effect of a positive estrogen and HER-2neu expression. The mentioned covariates have a statistical significant effect on the intercept component of the model (1.4622). Bilateral cancer cases have a decrease of 0.5981 on the intercept component, and a case with lymphatic invasion a decrease of 0.7322 compared to those with no lymphatic invasion. A case that presents images of vascular invasion increases of the start value of the tumor marker by an increment of 0.7322, comparing to those that do

not present any image. A positive estrogen receptor expression has an increasing effect on the intercept component by 1.2177, compared to those with a negative expression. A positive expression of HER-2neu has an increment of 0.4882.

Table 2: Estimated parameters values for General Linear Model and Longitudinal Model, for the patients who died from breast cancer.

	OLS Model		**REG Model**	
	Estimate	**p-value**	**Estimate**	**p-value**
Intercept	2.0376	<0.0001	1.8507	<0.0001
Time before changing point (2 years before death)	0.2540	<0.0001	0.2128	<0.0001
Time after changing point (2 years before death)	0.9453	<0.0001	0.8815	<0.0001
Bilateral (Yes)	−0.9290	<0.0001	−0.5981	0.0471
Lymphatic invasion (Yes)	−0.8821	<0.0001	0.7769	<0.0001
Venous vascular invasion (Yes)	1.0350	<0.0001	0.7769	0.0266
Estrogen receptor expression (positive)	1.5675	<0.0001	1.2177	<0.0001
$\hat{\nu}^2$			0.2404	
$\hat{\sigma}^2$			0.8239	
$\hat{\phi}$			0.3762	
$\hat{\tau}^2$			0.0415	
$\hat{\xi}^2$	1.2499			
Log Likelihood	−1089.503		−621.695	

For this subset, the correlation structure that best represent the variability of the data is the structure that incorporates random effects at individual level with $\hat{\nu}^2 \approx 0.2404$, a Gaussian correlation structure to describe the variability within patients with $\rho(u) = \exp(\frac{-1}{0.3762.}u^2)$ and $\hat{\sigma}^2 \approx 0.8239$, and a measurement error with variance $\hat{\tau}^2 \approx 0.0415$. The superposition of the theoretical variogram of both exponential and Gaussian correlation structures with the empirical variogram (Figure 5) validates the choice of an exponential correlation structure.

Both REE and REG models were compared with a longitudinal model only with an intercept random effect component U_i, and the serial correlation component $W_i(t_{ij})$ shown to be significant in the models. This result reinforces the need to take into account correlation within subject measurements.

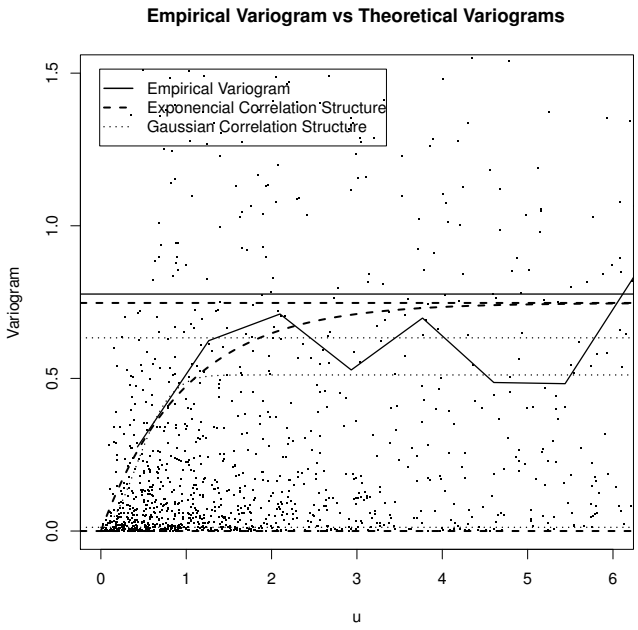

Figure 5: Superposition of empirial variogram and theoretical variogram, for patients who died from breast cancer.

4. DISCUSSION

An abrupt rise in values of CEA tumor marker, over a reference value, is an alert sign to a possible recurrence of breast cancer.

When analyzing all patients that were diagnosed with breast cancer, in our study, the only variables that have a statistically significant effect on the linear progression of the tumor marker are: tumor stage (III/IV versus 0/I/II), primary tumor size (Tx/T1/T2/T3/Tis versus T4), age at diagnosis, venous vascular invasion (Yes versus No), tumor degree (Gx/G1/G2 versus G3) and estrogen receptor expression (positive versus negative). As expected, a III or IV tumor stage, a T4 type of tumor, a G3 type of tumor, the presence of venous vascular invasion and age at diagnosis have an increasing effect on the average tumor marker progression in time, as they are related to a worst prognostic case ([2]). One unexpected result was the fact that a positive expression of the estrogen receptor has an increasing effect on that progression, contradicting the results from a previous survival analysis ([2]), where the same patients' cases of positive estrogen receptor shown a lower probability of dying from breast cancer than those who presented a negative expression.

It was detected a changing point on the linear progression of the tumor marker for the subset of patients that died from breast cancer two years before the death. This means that, at that point, there is an abrupt rise on the rate of its progression.

The risk factors for the progression of the marker, for that subset of patients are: bilateral (Yes versus No), lymphatic invasion (Yes versus No), venous vascular invasion (Yes versus No), estrogen receptor expression (positive versus negative) and HER-2neu expression (positive versus negative). As expected, the presence of venous vascular invasion has an increasing effect on the average CEA linear progression in time, as it is related to a worst prognostic case in the previous survival analysis ([2]). A bilateral type of tumor and the presence of lymphatic invasion have a decreasing effect. A positive estrogen and HER-2neu expression has an increasing effect. These two last results contradict the results from the previous survival analyses ([2]) since bilateral cases and lymphatic invasion are related to lower survival probability and, a positive estrogen and Her-2neu expression are both related to a higher probability of survival.

For both models fitted, the fact that the estimated variance of the measurement error is quite lower that the estimated variance of the OLS model, means that the fitted REE longitudinal model explains the variability of the data mainly by means of variability between patients and within patients assigning a very low value for measurement error (or *white noise* as usually mentioned in literature).

The fact that, when comparing the REE and the REG models to a longitudinal model with only an intercept random effect, the component the serial correlation was significant, stresses the importance incorporating a variability component that translates within subject measurements correlation, in this type of biological data.

The presented longitudinal analysis of this tumor marker, in combination with the previous survival analysis is going to be proceeded, in future work, with a joint modeling of the longitudinal and survival process of the present data.

ACKNOWLEDGMENTS

This work is funded by FEDER funds through the Operational Program for Competitiveness Factors – COMPETE and National Funds through FCT – Foundation for Science and Technology under the project EXPL/MAT-STA/0313/2013. The author Ana Borges has a PhD scholarship from FCT SFRH/BD/74166/2010.

REFERENCES

[1] Boyle, P. and Ferlay, J. (2004). Cancer incidence and mortality in Europe, *Annals of Oncology*, **16**, 3, 481–488.

[2] Borges, A.; Sousa, I. and Castro, L. (2013). *Breast cancer survival at Braga's hospital – Portugal*. In "Theoretical and Applied Issues in Statistics and Demography" (C.H. Skiadas, Ed.), Book of Abstracts - ASMDA2013, 34–35.

[3] Cianfrocca, M. and Goldstein, L.J. (2004). Prognostic and predictive factors in early-stage breast cancer, *The Oncologist*, **9**, 6, 606–616.

[4] Diggle, P.J.; Heagerty, P.; Liang K.-Y. and Zeger, S.L. (2002). *Analysis of Longitudinal Data*, University Oxford Press.

[5] *European Cancer Observatory*, (http://eco.iarc.fr/).

[6] Fiorella, G.; Ferroni, P.; Carlini, S.; Mariotti, S.; Spila, A.; Aloe, S.; D'Alessandro, R.; Carone, M.D.; Cicchetti, A.; Ricciotti, A.; Venturo, I.; Perri, P.; Filippo, F.; Cognetti, F.; Botti, c. and Roselli, M. (2001). A re-evaluation of carcinoembryonic antigen (CEA) as a serum marker for breast cancer: a prospective longitudinal study, *Clinical Cancer Research*, **7**, 8, 2357–2362.

[7] Fitzgibbons, P.L.; Page, D.L.; Weaver, D.; Thor, A.D.; Allred, D.C. and Clark, G.M. (2000). Prognostic factors in breast cancer, College of American Pathologists Consensus Statement 1999, *Archives of Pathology & Laboratory Medicine*, **124**, 7, 966–978.

[8] Harris, L.; Fritsche, H.; Mennel, R.; Norton, L.; Ravdin, P.; Taube, S.; Somerfield, MR.; Hayes, D.F. and Bast, R.C. Jr (2007). American Society of Clinical Oncology: American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer, *Journal of Clinical Oncology*, **25**, 5287–5312.

[9] Liang, K.Y. and Zeger, S.L. (2007). Longitudinal data analysis using generalized linear models, *Biometrika*, **73**, 13–22.

[10] Little, R.J.A. and Rubin, D.B. (2002). *Statistical Analysis with Missing Data*, 2nd Edition, Wiley-Interscience.

[11] Pinheiro, J.; Bates, D.; DebRoy, S.; Sarkar, D. and R Core Team (2014). nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-118, http://CRAN.R-project.org/package=nlme.

[12] Pinheiro, P.S.; Tyczynski, J.E.; Bray, F.; Amado, J.; Matos, E. and Parkin, D.M. (2003). Cancer incidence and mortality in Portugal, *European Journal of Cancer*, **39**, 17, 2507–2520.

[13] Philipson, P.; Sousa, I.; Diggle, P., Williamson, P.; Kolamunnage-Dona, R.; Henderson, R. and R Core Team (2014). JoineR: Joint modelling of repeated measurements and time-to-event data. R package version: 1.0-3, http://CRAN.R-project.org/package=joineR.

[14] Rodrigues, V. (2011). *Chapter 34*. In "Manual de Ginecologia", Permanyer, Portugal, 175–191.

[15] R Development Core Team (2008). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, Austria. URL http://www.R-project.org.

[16] Sturgeon, C.M.; Duffy, M.J.; Stenman, U.H.; Lilja, H.; Brünner, N.; Chan, D.W.; Babaian, R.; Bast, R.C. Jr.; Dowell, B. and Esteva, F.J. (2008). National Academy of Clinical Biochemistry *et al.*: National Academy of Clinical Biochemistry laboratory medicine practice guidelines for use of tumor markers in testicular, prostate, colorectal, breast and ovarian cancers, *Clinical Chemistry Journal*, **54**, e11–e79.