İstanbul Üniversitesi İşletme Fakültesi Dergisi
Istanbul University Journal of the School of Business Administration
Cilt/Vol:40, Sayı/No:1, 2011, 60-71
ISSN: 1303-1732 – www.ifdergisi.org © 2011

Dedicated to the 40th Anniversary of the Journal

Statistical modeling of breast cancer using differential equations

Chris P. Tsokos[1]
Department of Mathematics and Statistics,
University of South Florida, Tampa, FL 33620, USA

Young Xu[2]
Department of Mathematics, Readford
University, Radford, VA 24142, USA

Abstract

The object of the present study is to develop differential equations that will characterize the behavior of the tumor as a function of time. Having such differential equations, the solution of which once plotted will identify the rate of change of tumor size as a function of age. The structures of the differential equations characterize the growth of breast cancer tumor. Once we have developed the differential equations and their solutions, we proceed to validate the quality of the differential system and discuss its usefulness.

Keywords: *Differential Equation (DE), Rate of Change (ROC), Short and Long Term Prediction of Rate of Change, Statistical Modeling*

Diferansiyel denklemler kullanarak meme kanserinin istatistik modellemesi

Özet

Sunulan bu çalışmanın amacı, tümor davranışını zamanın bir fonksiyonu olarak niteleyen diferansiyel denklemler geliştirmektir. Bu tanıma uyan farklı diferansiyel denklemler elde edilmiş ve grafiksel olarak gösterilen denklemin sonucu, yaşın fonksiyonu olarak tümör büyüklüğünün değişim oranını belirlemede kullanılmıştır. Diferansiyel denklemlerin yapıları meme kanseri tümörünün gelişimini tanımlamaktadır. Diferansiyel denklemler ve çözümleri geliştirildikten sonra diferansiyel sistemin kalitesi doğrulanmış ve yöntemin faydası tartışılmıştır.

Anahtar Sözcükler: *Diferansiyel Denklem, Değişim Oranı, Değişim Oranının Kısa ve Uzun Vadede Tahmini, İstatistik Modelleme*

1. Introduction

Breast cancer (malignant breast neoplasm) is cancer originating from breast tissue, most commonly from the inner lining of milk ducts or the lobules that supply the ducts with milk, from Sariego [1]. There are different types of breast cancer. The object of the present study is to develop a differential equation that characterizes the behavior of the tumor as a function of age. With respect to the present study, we will address several questions:

- What is the mathematical characterization of the growth of the breast cancer tumor as a function of age?
- Is the analytical behavior of breast tumor size (TS) uniform over all age?

[1] *ctsokos@usf.edu (C.P.Tsokos)*

[2] *Xyoung007@hotmail.com (Y.Xu)*

- If the mathematical behavior of the tumor size as a function of age is not uniform, can we identify the age intervals where the sizes of the tumor have the same analytical growth behavior?
- Can we identify and justify their mathematical behavior of the size of tumor as a function of age over these age intervals?
- Can we develop a differential equation in characterizing the change of breast tumor size as a function of age over these age intervals?

In the present analysis, we used real data that we obtained from Surveillance Epidemiology and End Results (SEER) Program supported by NIH [2]. SEER collects information on incidence, survival, and prevalence from specific geographic areas representing 26 percent of the U.S. population and compiles reports on all of these items plus cancer mortality for the entire U.S.

The proposed differential equations are useful in predicting the rate of change of the cancer tumor size for a specified age of interested. The quality of the differential equation is statistically evaluated using residual analysis. Finally, the usefulness of our findings is discussed.

2. Historical Review

Many researchers had been working on various aspects of breast cancer. Here, we have summarized some of the most recent and relevant researches to the present study. Madigan et al. [3] studied the risk factors for breast cancer patients in the United States. Winchester [4] presented the relationship between breast cancer and age. Feig and Hendrick [5] analyzed the risk related to women aged 40 to 49 who undergo mammography procedure. Venturi [6] mentioned the key role for iodine in breast diseases. Fyles et al. [7] studied the behavior of Tamoxifen with or without breast radiation for women 50 years of age or older with early breast cancer detection. Jayasinghe [8] studied the behavior of age as one factor for the breast cancer patients. Chlebowski et al. [9] introduced the relationship between interim efficacy for female nutrition and cancer. Boffetta et al. [10] conducted research on the relation between alcohol drinking and cancer. Ibrahim et al. [11] presented a decision tree analysis for competing risks in breast cancer. Buchholz [12] discussed the benefit of radiation therapy for early-stage breast cancer patients. Xu, Kepner and Tsokos [13] have introduced a statistical model of breast cancer tumor size that is used to identify the attributable variables and significant interactions and ranking their influences.

2.1. The Breast Cancer Data

From the SEER data we have randomly selected information of 1,000 breast cancer patients. The age of the patients ranges from 33 to 85 years old. However, from the age of 33 to 40 year old patients the data is not complete. Thus, our analysis is focused from the age of 41 to 85 year of age. The scatter diagram given by Figure 1 is obtained by averaging breast tumor sizes as a function of the age of the breast cancer patients.

For better analytical characterization, we decided that our analysis should be based on partitioning the scatter plot into three age intervals. Age Group I will be patitioned into the age interval from 41 to 58 and Age Group II are III from 59 to 73 and 74 to 85 year old, respectively. The data tree diagram in Figure 2 identifies the sample sizes and the age intervals.

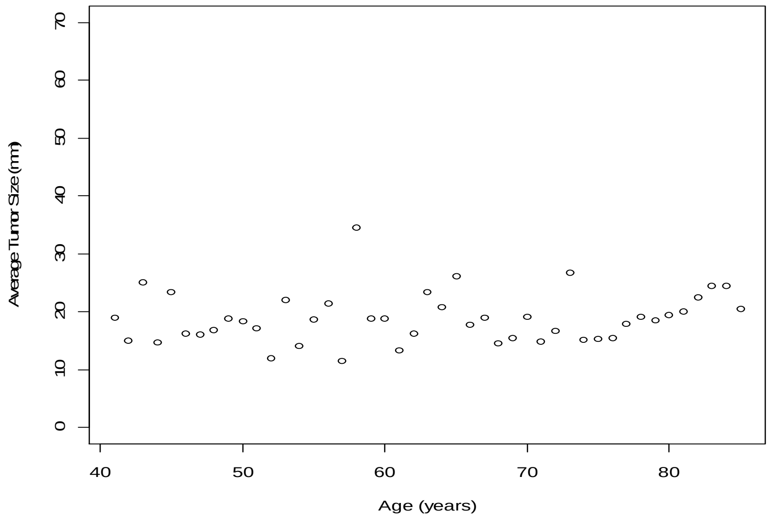

Figure 1 Breast Cancer Patients' Tumor Size from Age 41 to Age 85

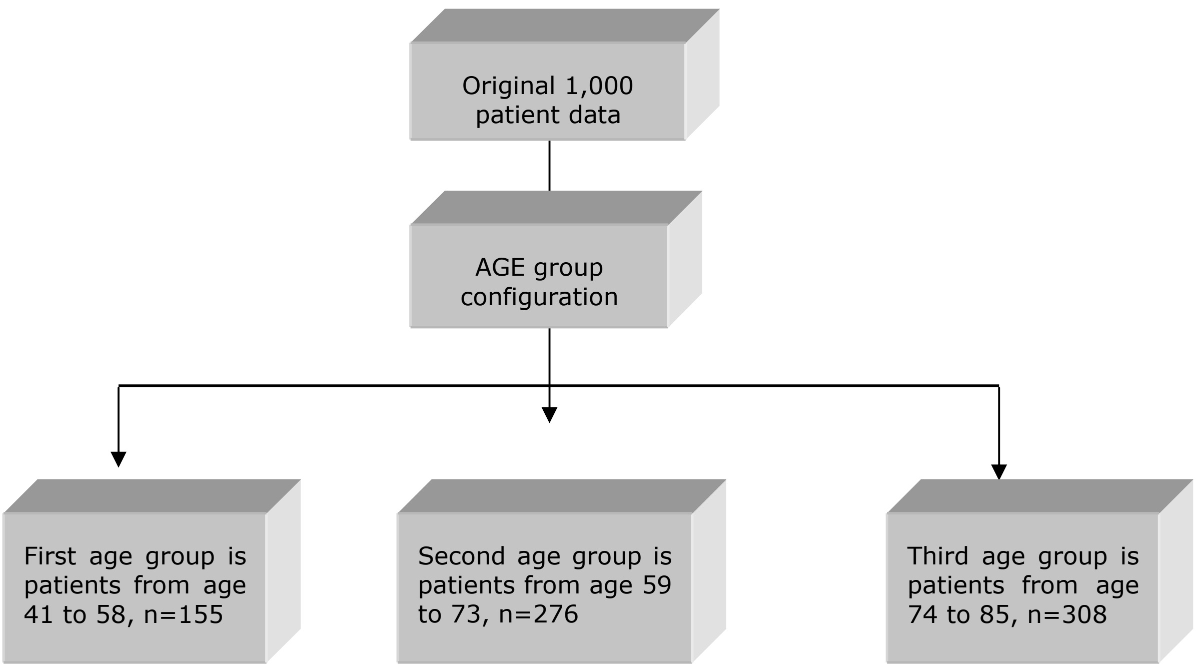

Figure 2 Breast Cancer Data Tree Diagram

3. Developments of the Differential Equations

Age Group I

This age group consists of 155 breast cancer patients from 41 to 58 years of age. The scatter diagram of this data is given in Figure 3.

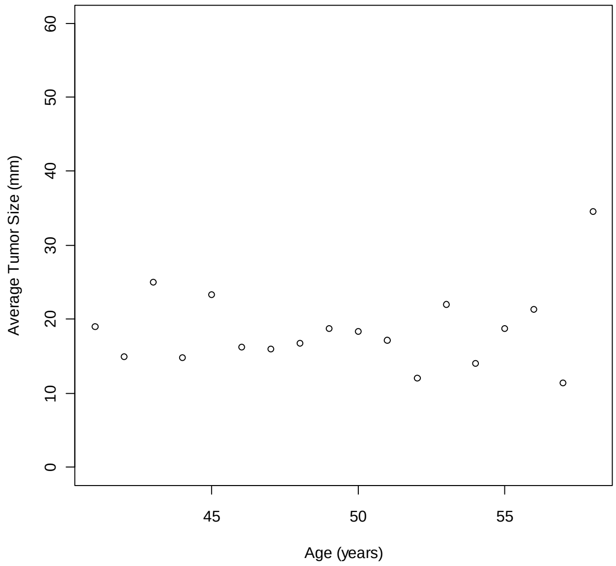

Figure 3 Breast Cancer Patients' Tumor Size from Age 41 to Age 58

The mathematical function that characterizes the breast cancer tumor size behavior in the given age group is given by (1).

$$T(x) = 2.016 \times 10^{6} - 2.518 \times 10^{5} + 1.307 \times 10^{4}x^{2} - 3.605 \times 10^{2}x^{3} + 5.581x^{4} - 4.593 \times 10^{-2}x^{5} + 1.571 \times 10^{-4}x^{6}, \ 41 \leq x \leq 58 \quad (1)$$

We check the quality of the fitting by residual analysis of the breast cancer tumor size in Table 1.

Table 1 Residual Analysis of Breast Cancer Tumor Size in Group I

Age	Actual value	Fitted value	Residual
41	19	18.5811	.4189
42	15	17.824	-2.824
43	25	19.0416	5.9584
44	14.75	19.9738	-5.2238
45	23.33	19.856	3.4773
46	16.25	18.8271	-2.577
47	16	17.4501	-1.45
48	16.8	16.3467	.4533
49	18.75	15.9438	2.8062
50	18.33	16.3342	1.9991
51	17.125	17.2499	-.1249

Mean of Residual	7.31957e-17
Standard Deviation of Residual (SD)	3.975954
Standard Error of Residual (SE)	0.9371414

Thus, based on the residual analysis we can conclude that the analytical behavior of the tumor size of breast cancer patients given by (1) is a good fit. Figure 4 shows the actual polynomial over the scatter data.

Now we proceed to identify the differential equation for the first age group. Let X represent the patients' age in term of years and the according tumor size is a function, T(x), in term of mille meter (mm) then the instantaneous rate of change (IROC) of tumor size is the derivative of the tumor size function with respect to time (T′(x)).

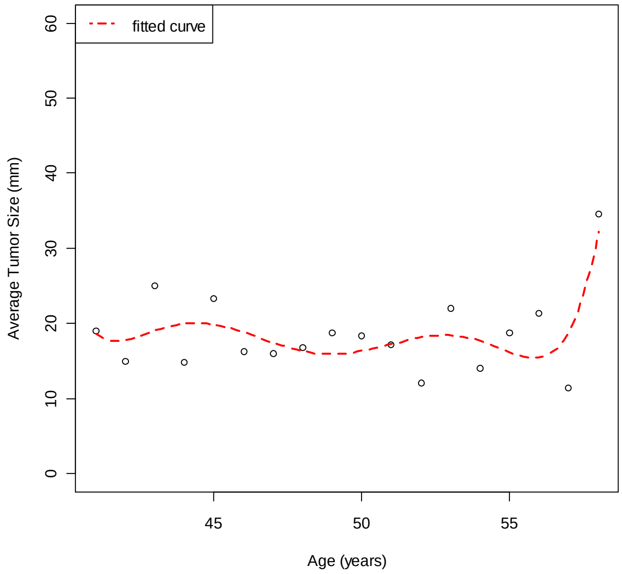

Figure 4 Breast Cancer Patients' Tumor Size from Age 41 to Age 58 with Curve Fitting

The differential equation is given by (2).

$$T(x) + T'(x) = 1.764 \times 10^6 - 2.2456 \times 10^5 x + 1.198 \times 10^4 x^2 - 3.38 \times 10^2 x^3 + 5.351x^4 - 4.499 \times 10^{-2} x^5 + 1.571 \times 10^{-4} x^6, \quad 41 \le x \le 58 \tag{2}$$

Thus, the solution of (2) is given by (3). Therefore, if one is interest in obtaining the change of rate of the breast cancer tumor size for a desired age in age group I, he can evaluate the solution of the differential equation at the desired age.

$$\frac{d(T(x))}{d(x)} = -2.518 \times 10^5 + 2.613 \times 10^4 x - 1.082 \times 10^3 x^2 + 22.3221 \times x^3 - 0.2297x^4 + 0.942 \times 10^{-3} x^5, \quad 41 \le x \le 58 \tag{3}$$

Figure 5 is a representation of the solution to (3). We proceed to evaluate the results given by the solution to the differential equation. We evaluate the accuracy of the results from the differential equation as follows. For example, at age of 41 to 42, the solution to the differential equation estimate the change of rate is -0.04074, where the observed actual rate of change is given by -0.210526 which is obtained from $ROC = \frac{\text{current year} - \text{previous year}}{\text{previous year}}$. The difference of the two constitutes the first rate of change residual (ROC residual). Table 2 gives the 10 estimates of the solution of the differential equation.

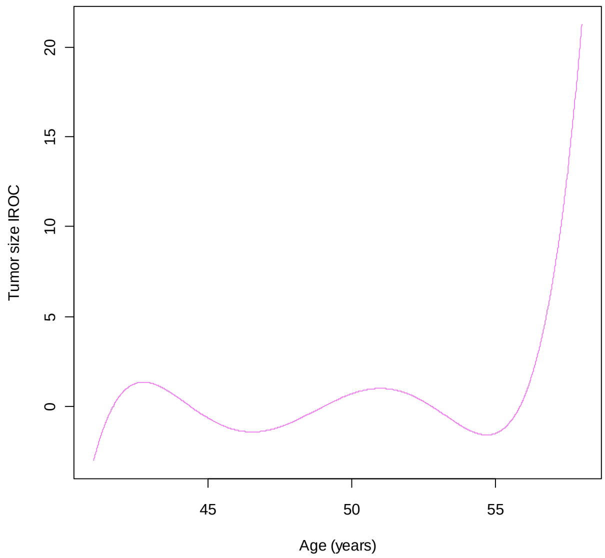

Figure 5 Breast Cancer Patients' Tumor Size IROC from Age 41 to Age 58

Table 2 Residual Analysis of ROC of Breast Cancer Tumor Size in Group I

Age	Empirical ROC	DE IROC	ROC Residual
42	-0.210526	-0.04074	-0.16978212
43	0.666667	0.0683115	0.598355
44	-0.410000	0.0489545	-0.458955
45	0.581921	-0.0058972	0.587818
46	-0.303571	-0.0518198	-0.251752
47	-0.015385	-0.0731362	0.057752
48	0.050000	-0.0632339	0.113234
49	0.116071	-0.0246478	0.140719
50	-0.022222	0.024487	-0.046709
51	-0.065909	0.0560584	-0.121967

Mean of Residual	0.1012375
Standard Deviation of Residual	0.4946078
Standard Error of Residual	0.1165802

Based on the above results, we can conclude that the differential equation gives fairly accurate rate of the change of the breast tumor size as a function of age.

We can utilize the mathematical expressions (2) and (3) with the correction factor of the mean of residual to estimate the rate of the tumor growth for future age.

Age Group II

This group consists of 276 patients from 59 to 73 years of age. The scatter diagram of the data with the best fit is given in Figure 6.

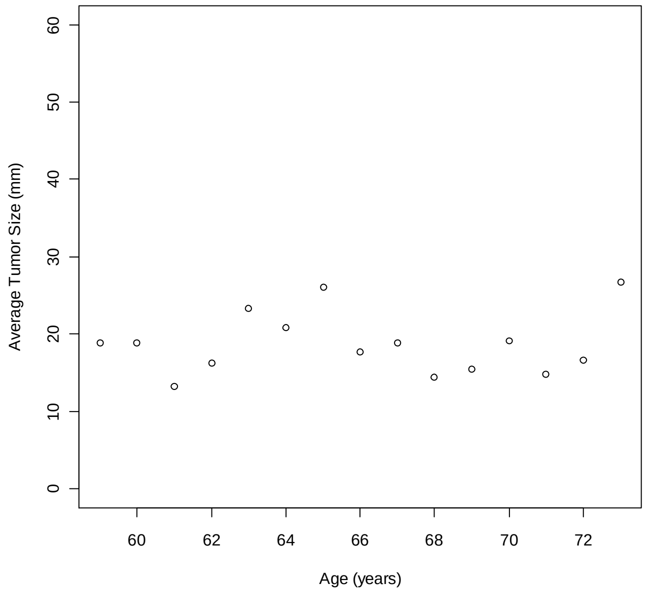

Figure 6 Breast Cancer Patients' Tumor Size from Age 59 to Age 73

The mathematical function that characterizes the breast cancer tumor size behavior in the given age group is given by (4).

$$T(x) = 1.749 \times 10^5 - 1.0788 \times 10^4 x + 2.491 \times 10^2 x^2 - 2.551x^3 + 9.776 \times 10^{-3} x^4, \;\; 59 \le x \le 73 \qquad (4)$$

We check the quality of the fitting by residual analysis of the breast cancer tumor size in Table 3.

Table 3 Residual Analysis of Breast Cancer Tumor Size in Group II

Age	Actual value	Fitted value	Residual
58	18.875	18.81482	0.06018
59	18.8235	16.72068	2.10285
60	13.2381	16.88538	-3.64728
61	16.2105	18.2	-1.989647
62	23.3333	19.79	3.5423886
63	20.8	21.018	-0.21818

Mean of Residual	1.484996e-17
Standard Deviation of Residual	2.706574
Standard Error of Residual	0.6988345

Thus, based on the residual analysis we can conclude that the analytical behavior of the tumor size of breast cancer patients given by (4) is a good fit. Figure 7 below shows the actual polynomial over the scatter data.

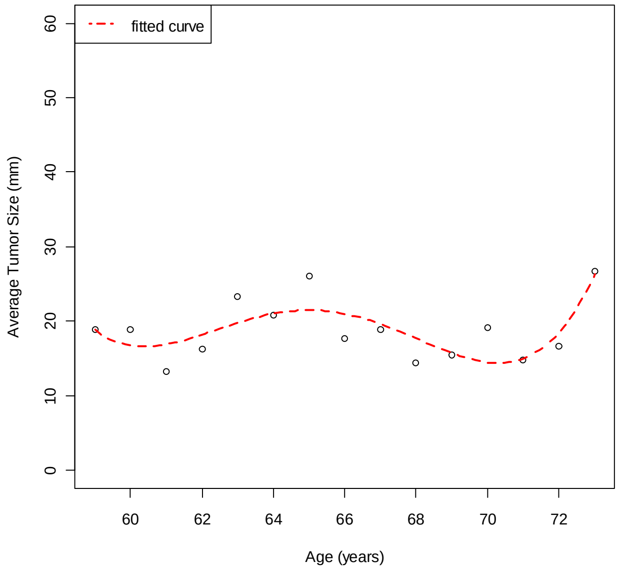

Figure 5 Breast Cancer Patients' Tumor Size from Age 59 to Age 73 with Curve Fitting

Now we proceed to identify the differential equation for the second age group. The differential equation is given by (5).

$$T(x) + T'(x) = 1.6414 \times 10^5 - 1.029 \times 10^4 x + 2.4141 \times 10^2 x^2 - 2.511x^3 + 9.776 \times 10^{-3} x^4, \quad 59 \le x \le 73 \qquad (5)$$

The instantaneous rate of change of breast cancer patients' tumor size as a function of time is given analytically by

$$\frac{d(T(x))}{d(x)} = -0788 \times 10^4 + 4.9812 \times 10^2 x - 7.652x^2 + 0.391x^3, \; 59 \le x \le 73 \qquad (6)$$

A graphical display of (6) is given by Figure 6.

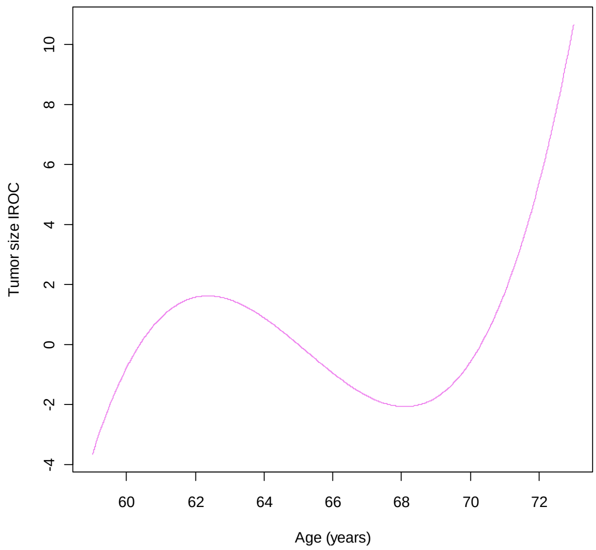

Figure 6 Breast Cancer Patients' Tumor Size IROC from Age 59 to Age 73

The residual analysis we performed on the proposed differential equation of tumor size is given in Table 4. We will only keep 5 residual data.

As seen from the table below the residuals are small and so is the standard error. These results attest to the good quality of the proposed model for tumor size.

Table 4 Residual Analysis of ROC of Breast Cancer Tumor Size in Group II

Age	Empirical ROC	DE IROC	ROC Residual
59	-0.4529	-0.40292	-0.04998
60	-0.00273	-0.2513	0.24857
61	-0.29673	0.005681	-0.30241
62	0.224536	0.164981	0.059555
63	0.439394	0.161232	0.278162

Mean of Residual	0.01485848
Standard Deviation of Residual	0.1969682
Standard Error of Residual	0.05462916

Age Group III

This group consists of 308 patients from 73 to 85 years of age. The scatter diagram of the data with the best fit is given in Figure 7.

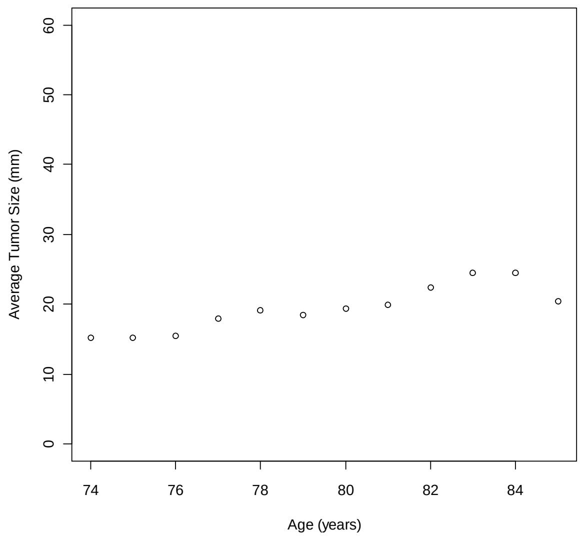

Figure 7 Breast Cancer Patients' Tumor Size from Age 74 to Age 85

The mathematical function that characterizes the breast cancer tumor size behavior in the given age group is given by (7).

$$T(x) = -2.93789 \times 10^5 + 1.4954 \times 10^4 x - 2.853 \times 10^2 x^2 + 2.4166x^3 - 7.672 \times 10^{-3} x^4, \quad 74 \le x \le 85 \tag{7}$$

We check the quality of the fitting by residual analysis of the breast cancer tumor size in Table 5.

Table 5 Residual Analysis of Breast Cancer Tumor Size in Group III

Age	Actual value	Fitted value	Residual
74	15.16667	14.53395	0.63271232
75	15.26667	16.05426	-0.78759789
76	15.48387	16.69901	-1.2151395
77	17.9	17.06703	0.83296694
78	19.11765	17.57305	1.5445921
79	18.45833	18.44768	0.01065

Mean of Residual	-6.780781e-18
Standard Deviation of Residual (SD)	0.929734
Standard Error of Residual (SE)	0.2683911

Thus, based on the residual analysis we can conclude that the analytical behavior of the tumor size of breast cancer patients given by (7) is a good fit. Figure 8 shows the actual polynomial over the scatter data.

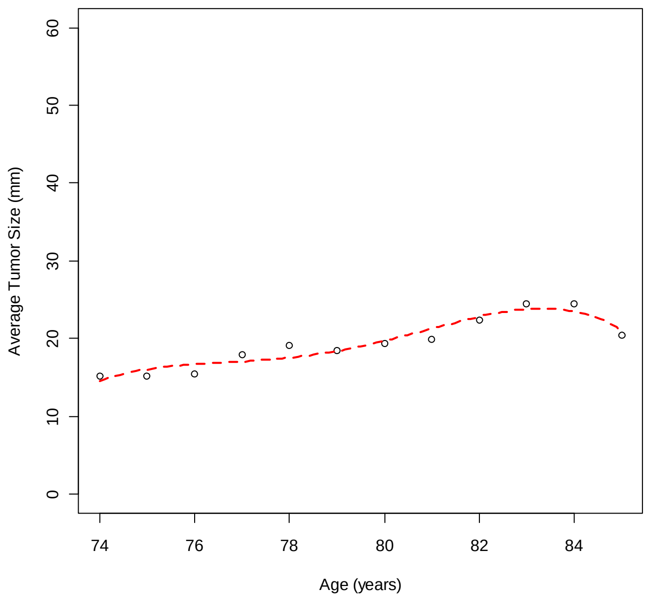

Figure 8 Breast Cancer Patients' Tumor Size from Age 74 to Age 85 with Curve Fitting

Now we proceed to identify the differential equation for the third age group. The differential equation is given by (8).

$$T(x) + T'(x) = -2.7883 \times 10^{7} - 1.438 \times 10^{4}x - 2.78 \times 10^{2}x^{2} - 2.3859x^{3} - 7.672 \times 10^{-3}x^{4}, \quad 74 \le x \le 85 \tag{8}$$

Thus, the solution of (8) is given by (9). Therefore, if one is interested in obtaining the change of rate of the breast cancer tumor size for a desired age in age group III, he can evaluate the solution of the differential equation at the desired age.

$$\frac{d(T(x))}{d(x)} = -1.4954 \times 10^{4} - 5.705 \times 10^{2}x + 7.24988x^{2} - 0.03068 \times x^{3}, \;\; 74 \le x \le 85 \tag{9}$$

A graphical display of (9) is given by Figure 9.

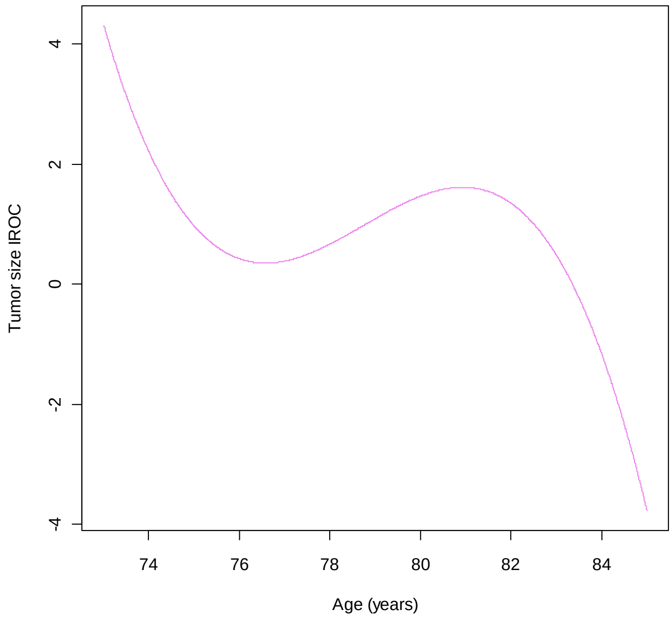

Figure 9 Breast Cancer Patients' Tumor Size IROC from Age 74 to Age 85

The residual analysis we performed on the proposed differential equation of tumor size is given in Table 6. We will only keep 5 residual data.

Table 6 Residual Analysis of ROC of Breast Cancer Tumor Size in Group III

Age	Empirical ROC	DE IROC	Residual
74	0.0065934	-0.030347	0.036941
75	0.014227	0.0967895	-0.082562
76	0.156042	0.090693	0.06534849
77	0.06802498	0.04762	0.0204083
78	-0.0344872	0.0187968	-0.053284

Mean of Residual	0.001199553
Standard Deviation of Residual (SD)	0.04160096
Standard Error of Residual (SE)	0.01200916

As seen from Table 6 the residuals are small and so is the standard error. These results attest to the good quality of the proposed model for tumor size.

4. Usefulness of the Proposed Differential Equation

We can conclude from our extensive statistical analysis that all of the four parts of the differential equations have good quality.

This model is useful for a number of reasons:

1. It can be used to identify the rate of change of the growth of the breast cancer tumor size.
2. One can also use the proposed differential equation systems to generate various scenarios of the tumor size as a function of different values of the age.
3. People can use these differential equation systems to predict the rate of change of the growth of tumor for different ages.

5. Conclusions & Discussion

In the present study, we extract a random sample 1,000 breast cancer patients from SEER data base and develop differential equations to obtain information about the rate of growth of breast cancer tumor. We found the breast cancer tumor size is not uniform over all age. The sample data was partitioned into three intervals groups as a function of age for better analytical tractability, that is, the age group from 41 to 58, age group from 59 to 73 and age group from 74 to 85. For each age group, we develop a differential equation that can be used to obtain the rate of growth of the malignant tumor size. We justified the mathematical behavior of the function we proposed by residual analysis.

References

[1] J. Sariego, Breast cancer in the young patient. *The American Surgeon*, 76, 12, 1397–1400 (2010).

[2] U.S. National Institutes of Health, http://seer.cancer.gov (2010).

[3] M.P. Madigan, R.G. Ziegler, J. Benichou, C. Byrne, and R.N. Hoover, Proportion of breast cancer cases in the United States explained by well-established risk factors. Journal of the National Cancer Institute. 87, 22, 1681–1685 (1995).

[4] D.P. Winchester, The National Cancer Data Base report on breast carcinoma characteristics and outcome in relation to age. *Cancer*, 78, 1838 (1996).

[5] SA. Feig and RE. Hendrick, Radiation risk from screening mammography of women aged 40–49 years. *Journal of National Cancer Institute Monographs*, 22, 22, 119–124 (1997).

[6] S. Venturi, Is there a role for iodine in breast diseases? *The Breast*, 10, 5, 379–382 (2001).

[7] A.W. Fyles, D.R. McCready, and L.A. Manchul, Tamoxifen with or without breast irradiation in women 50 years of age or older with early breast cancer. *New England Journal of Medicine*, 351, 963-970 (2004).

[8] U.W. Jayasinghe, Is age at diagnosis an independent prognostic factor for survival following breast cancer? *Anz Journal of Surgery*, 75, 762 (2005).

[9] R.T. Chlebowski, GL. Blackburn, and CA. Thomson, Dietary fat reduction and breast cancer outcome: interim efficacy results from the Women's Intervention Nutrition Study. *Journal of the National Cancer Institute*, 98, 24, 1767–1776 (2006).

[10] P. Boffetta, M. Hashibe, C. La Vecchia, W. Zatonski, and J. Rehm, The burden of cancer attributable to alcohol drinking. *International Journal of Cancer*, 119, 4, 884–887 (2006).

[11] N.A. Ibrahim, A. Kudus, and I. Daud, Decision tree for competing risks survival probability in breast cancer study. *International Journal of Biomedical Sciences*, 3, 1 (2008).

[12] T. Buchholz, Radiation therapy for early-stage breast cancer after breast-conserving surgery. *The New England Journal of Medicine*, 360, 1, 63-70 (2009).

[13] Y. Xu, J. Keper, and C.P. Tsokos, Identify attributable variables and interactions in breast cancer. *Journal of Applied Sciences*, 11, 6, 1033-1038 (2011).