

# Determining factors influencing survival of breast cancer by fuzzy logistic regression model

Roya Nikbakht, Abbas Bahrapour

Department of Biostatistics and Epidemiology, Modeling in Health Research Center, Faculty of Health, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran

**Background:** Fuzzy logistic regression model can be used for determining influential factors of disease. This study explores the important factors of actual predictive survival factors of breast cancer's patients. **Materials and Methods:** We used breast cancer data which collected by cancer registry of Kerman University of Medical Sciences during the period of 2000–2007. The variables such as morphology, grade, age, and treatments (surgery, radiotherapy, and chemotherapy) were applied in the fuzzy logistic regression model. Performance of model was determined in terms of mean degree of membership (MDM). **Results:** The study results showed that almost 41% of patients were in neoplasm and malignant group and more than two-third of them were still alive after 5-year follow-up. Based on the fuzzy logistic model, the most important factors influencing survival were chemotherapy, morphology, and radiotherapy, respectively. Furthermore, the MDM criteria show that the fuzzy logistic regression have a good fit on the data (MDM = 0.86). **Conclusion:** Fuzzy logistic regression model showed that chemotherapy is more important than radiotherapy in survival of patients with breast cancer. In addition, another ability of this model is calculating possibilistic odds of survival in cancer patients. The results of this study can be applied in clinical research. Furthermore, there are few studies which applied the fuzzy logistic models. Furthermore, we recommend using this model in various research areas.

**Key words:** Breast cancer, fuzzy logistic regression, mean degree of membership, survival

**How to cite this article:** Nikbakht R, Bahrapour A. Determining factors influencing survival of breast cancer by fuzzy logistic regression model. J Res Med Sci 2017;22:135.

## INTRODUCTION

### Epidemiology of breast cancer

Cancer is a leading cause of death in the world.<sup>[1]</sup> The burden of cancer is rising gradually throughout the world. Among all types of cancer, lung cancer is the most prevalent cancer (worldwide) in men. When compared to men, in women, breast cancer is the first leading cause of cancer death in developed and developing countries.<sup>[2,3]</sup> Regarding the global cancer statistics report in 2011, breast cancer included around 23% of total cancer. Besides, the incidence rate of breast cancer had the maximum rate (89.9%) in western European countries based on standardized age (2011).<sup>[4]</sup>

Nearly 7% of woman diagnosed with breast cancer are at an age younger than 40 years.<sup>[5]</sup> Expected number

of deaths for breast cancer, US, 2015, is estimated near 40,290 women.<sup>[6]</sup> Southern Africa had the highest rate of breast cancer incidence mortality (19.3%). In addition, western Asia countries had the highest incidence rate and mortality rate of breast cancer 31.8% and 18.9%, respectively.<sup>[4]</sup> When compared to other countries, breast cancer is the second common cancer in Iranian females (21.4%).<sup>[7,8]</sup>

### Etiology and predictive factors of breast cancer

The main predictive risk factors of breast cancer are "age, geographic area of residence, age at first birth, certain indicators of ovarian activity, history of benign breast disease, and familiar history of breast cancer."<sup>[9]</sup> Breast cancer treatment usually involves a combination of surgery, radiation therapy, chemotherapy, and hormone therapy. In addition, prognosis and selection of therapy can determine by age, menopausal status,

#### Access this article online

Quick Response Code:



Website:  
www.jmsjournal.net

DOI:  
10.4103/jrms.JRMS\_405\_17

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

**For reprints contact:** reprints@medknow.com

**Address for correspondence:** Prof. Abbas Bahrapour, Department of Biostatistics and Epidemiology, Modeling in Health Research Center, Faculty of Health, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran. E-mail: abahrapour@yahoo.com

**Received:** 13-05-2017; **Revised:** 02-09-2017; **Accepted:** 10-10-2017

stage of disease, histologic, and nuclear grade of the primary tumor.<sup>[10,11]</sup> Some studies indicated that survival of breast cancer's patients depends on tumor size and biological factors.<sup>[12]</sup> In some patients who received the radiotherapy before chemotherapy, the actual 5-year breast failure rate was 4% in 99 of them. In other patients (54 people) who received chemotherapy sequentially then radiotherapy without concurrent chemotherapy, this rate was 8%. Furthermore, the failure rate was 6% in 116 patients with receiving concurrent chemotherapy and radiotherapy.<sup>[13]</sup>

For determining actual predictive survival factors in breast cancer, the fuzzy logistic regression model can be used. In addition, one of the abilities of this model is predicting the status of the new patients by possibilistic odds (fuzzy logistic regression was based on the possibilistic odds approach). A few studies applied this model for survival of cancer patients. In this study, we conducted the fuzzy logistic regression model (new statistical regression model) for determining important factors influencing survival of breast cancer patients and predicting the status of the new patient. The main purpose of this study is to introduce a new statistical model and applying the results in the clinical research.

## MATERIALS AND METHODS

### Patients

This study used data of breast cancer registered by cancer registry of Kerman (the largest province of Iran) during the period of 2000–2007. The aim of the study was to determine important factors in survival of breast cancer patients. There were 1311 patients with breast cancer (both gender). Males and some patients with incomplete information excluded from the study. Finally, 924 patients remained but we studied just 71 patients of them (we used Lingo software for solving inequalities of fuzzy logistic regression model, but the Lingo software can solve inequalities in a sample of 71). Fuzzy logistic regression was performed with a binary dependent variable (survival status which had two states: alive or dead) and some independent variables such as age, morphology, grade, and treatments (radiotherapy, chemotherapy, and surgery).

The data of death results from breast cancer were available. Therefore, we determined 5-year survival from incidence (diagnosed with a disease during a given period of time) to death.

Types of morphology were neoplasm malignant, carcinoma, and infiltrating duct carcinoma. Radiotherapy, chemotherapy, and surgery were treatments approaches. Some cancer patients received treatments while some others did not. In other words, in some cases, radiotherapy,

chemotherapy, and surgery used to cure or control the disease. Tumor grade had three levels (I = well-differentiated, II = moderate-differentiated, III = poor-differentiated). Furthermore, the status of patients who died from 2000 to 2007 was zero.

### Statistical analysis

The fuzzy logistic regression is an important tools for evaluating the relationship between independent variable (crisp or fuzzy) and fuzzy binary outcome (The status of some patients was one, and some were zero with a probability of  $\mu$  and  $1-\mu$ , respectively).

At first step, we fitted a logistic regression model to the data. After that, we used predicted probabilities as  $\mu_i$  for modeling fuzzy logistic regression. Note that, possibilistic was used instead of probability in fuzzy logistic regression.

In the mentioned model  $\mu_i = Poss(Y_i = 1)$  is the possibility of

having the related property and  $\frac{\mu_i}{1-\mu_i}$  shows possibilistic

odds. Then, we applied the Fuzzy logistic regression on data. To estimate fuzzy coefficients, R and Lingo software were used but Lingo software could not support data with large sample sizes.

Suppose that  $X = (x_1, x_2, \dots, x_n)$  denoted a vector of explanatory variables, then the formulation of Fuzzy logistic regression with fuzzy coefficients is defined as:

$$\tilde{W}_i = \tilde{b}_0 + \tilde{b}_1 x_{i1} + \dots + \tilde{b}_n x_{in} \quad (i = 1, \dots, n) \quad \text{Eq (1)}$$

That,  $\tilde{W}_i = \ln\left(\frac{\mu_i}{1-\mu_i}\right)$  and  $\exp(\tilde{W}_i[x])$  is possibilistic odds of survival of breast cancer for each patient. Also  $(\tilde{b}_1, \tilde{b}_2, \dots, \tilde{b}_n)$   $\tilde{b}_0$  are the slopes of explanatory variables and intercept respectively [more details exist in Appendix 1].

Moreover ultimately, mean degree of membership (MDM) used for checking goodness of fit in fuzzy logistic regression which is a value between 0 and 1. As MDM is closer to one, the model has a good fit to the data.

$$MDM = \frac{1}{n} \sum_{i=1}^n W(w) = \frac{1}{n} \sum_{i=1}^n \exp\left(\tilde{W}_i\left(\frac{\mu_i}{1-\mu_i}\right)\right) \quad \text{Eq (2)}^{[14,15]}$$

## RESULTS

Based on the results of Table 1, Mean (SD) of age was 47.9 (12.5) in patients with breast cancer. Nearly 41% of patients were in neoplasm and malignant group. Furthermore, 2.8% and 56.3% of patients were in carcinoma and infiltrating duct carcinoma, respectively. More than two-third of patients were still alive after 5-year survival.

Most of the patients had a moderate grade (90.1%). Patients based on their conditions received surgery (54.9%), radiotherapy (46.9%), and chemotherapy (19.7%).

We obtained the fuzzy logistic regression model as follows:

$$\tilde{W}_i = (0.5598, 1.387)_T + (0.000001, 0.0924)_T \text{age} + 0.3828 \text{surgery} + (0.1238, 2.3870)_T \text{radiotherapy} + 0.5676 \text{chemotherapy} + 0.1960 \text{grade} + 0.5131 \text{morphology}$$

Eq (3)

The coefficient of some variables such as age and radiotherapy were fuzzy and were not interpretable. Therefore, we used defuzzification approach for fuzzy coefficients (more information about defuzzification exists in Appendix1).

After defuzzification, coefficients of variables were compared for detecting the importance of variables. According to obtained model, the most important factors influencing survival were chemotherapy, morphology, and radiotherapy, in order (Eq [3]) [Table 2].

**Table 1: The basic and clinical characteristics of study participants**

Variable	Frequency (%)
Morphology	
Neoplasm, malignant	29 (40.8)
Carcinoma	2 (2.8)
Infiltrating duct carcinoma	40 (56.3)
Surgery	
No	32 (45.1)
Yes	39 (54.9)
Radiotherapy	
No	38 (53.3)
Yes	33 (46.5)
Chemotherapy	
No	57 (80.3)
Yes	14 (19.7)
Status	
Dead	11 (15.5)
Alive	60 (84.5)
Grade	
I	4 (5.6)
II	64 (90.1)
III	3 (4.2)
Age	
Mean (SD)	47.9 (12.5)

SD=Standard deviation

**Table 2: Coefficients defuzzification**

Variable	Fuzzy coefficient	Defuzzification
Age	0.000001, 0.0924	0.27
Radiotherapy	0.1238, 2.3870	0.51
Constant	0.5598, 1.3870	1.01

In addition, the value of possibilistic odds of all cases determined. The possibilistic odds of five patients are reported in Table 3.

For example, a possibilistic odds of survival of breast cancer for the first patient is 0.84. We can also interpret possibilistic odds for other patients in similar way. Furthermore, possibilistic odds for a new case can be obtained by Eq.<sup>[3]</sup>

For instance, suppose we have a new case with following variables:

Age = 45 years old, Grade II, morphology = Infiltrating duct carcinoma, gave all treatments (surgery, radiotherapy, and chemotherapy). The possibilistic odds of this patient can be calculated as follows:

$$\tilde{W}_i = (0.5598, 1.387)_T + (0.000001, 0.0924)_T \times 45 + 0.3828 \times 1 + (0.1238, 2.3870)_T \times 1 + 0.5676 \times 1 + 0.1960 \times 2 + 0.5131 \times 3 = (3.57, 7.93)_T$$

Therefore, possibilistic odds of survival of breast cancer for this case were 0.73 after difuzzification.

Finally, model performance was assessed by MDM. The value of MDM was 0.86 which it showed fuzzy logistic regression had a good fit on the data.

## DISCUSSION

In this study, we determined the importance of factors may affect survival of breast cancer, namely, fuzzy logistic regression model. The study results demonstrated the important factors were chemotherapy, morphology, and radiotherapy. Furthermore, the fuzzy logistic regression had a good fit on the data (MDM = 0.86). Such findings were consistent with similar study that pointed out in pre- and post-menopausal patients with high risk of breast cancer. The short-term surgical with combination of chemotherapy was effective.<sup>[16]</sup> In another study, the factor which influenced the results of 5-year survival was number of axillary lymph nodes involved (not by menopausal status). Furthermore, the combination chemotherapy at full dose is vital for achieving clinical benefit.<sup>[17]</sup> In similar study, the combination of radiotherapy with chemotherapy declined rates of loco regional after modified radical mastectomy.<sup>[18]</sup>

Regarding the morphological factors, we found that through fuzzy logistic regression model-these factors had an important role at survival status in cancer patients. Similarly, morphological assessments studies showed that specific morphological characteristics strongly associated with “basal-like breast carcinoma” and can provide helpful

**Table 3: The observed and the estimated fuzzy outputs of breast cancer**

ID	The possibility of survive $\mu_i$	Membership function of fuzzy estimated output $\tilde{W}_i$	Predicted possibilistic odds $\tilde{W}(w_i) = \exp(\tilde{W}_i \cdot (\frac{\mu_i}{1 - \mu_i}))$
1	0.92	2.61, 6.82	0.84
2	0.95	2.61, 8.76	0.75
3	0.93	1.47, 7.85	0.81
4	0.81	2.49, 4.90	0.87
5	0.91	2.49, 6.56	0.76

information of prognosis of breast cancer.<sup>[19]</sup> Morphological assessment of the difference has been shown in numerous studies to provide useful information in breast cancer<sup>[20]</sup> that shows the power of fuzzy logistic regression for exploring the influential factors.

This method is applied in other areas such as diabetic's patients, lupus, and tuberculosis. In one of them, the fuzzy logistic regression model introduced as a new possibilistic model for determining the diabetic status.<sup>[14]</sup> Another study measured the association between tuberculosis and smoking through this model.<sup>[21]</sup> Furthermore, Pourahmad applied fuzzy logistic regression in the suspected cases to systematic lupus erythematosus disease.<sup>[22]</sup>

In this research, we introduced fuzzy logistic regression, a new statistical method, for determining breast cancer's survival factors. By this model which extracts possibilistic odds of survival and predicts status it is possible to calculate the odds for survival of the new patient which can be introduced. Furthermore, the results of this study can be applied in clinical research.

## CONCLUSION

According to our findings, we recommended fuzzy logistic regression model: first, this model is useful for determining the survival of the breast cancer's patients regarding to real data. Another advantage of Fuzzy logistic regression model is that possibilistic odds of survival status can be calculated for a new case.

## Financial support and sponsorship

Nil.

## Conflicts of interest

The authors have no conflicts of interest.

## REFERENCES

- Parkin DM, Pisani P, Ferlay J. Global cancer statistics. *CA Cancer J Clin* 1999;49:33-64, 1.
- Travis WD, Lubin J, Ries L, Devesa S. United States lung carcinoma incidence trends: Declining for most histologic types among males, increasing among females. *Cancer* 1996;77:2464-70.
- Fentiman IS, Fourquet A, Hortobagyi GN. Male breast cancer. *Lancet* 2006;367:595-604.
- Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D, *et al.* Global cancer statistics. *CA Cancer J Clin* 2011;61:69-90.
- Anders CK, Johnson R, Litton J, Phillips M, Bleyer A. Breast cancer before age 40 years. *Seminars in oncology* 2009; 36: 237-249.
- DeSantis, Carol E, Fedewa, Stacey A, Goding Sauer, Ann Kramer, *et al.* Breast cancer statistics, 2015: Convergence of incidence rates between black and white women. *CA: a cancer journal for clinicians* 2016; 66: 31-42.
- Sadjadi A, Nouraei M, Mohagheghi MA, Mousavi-Jarrahi A, Malekezadeh R, Parkin DM, *et al.* Cancer occurrence in Iran in 2002, an international perspective. *Asian Pac J Cancer Prev* 2005;6:359-63.
- Babu GR, Samari G, Cohen SP, Mahapatra T, Wahbe RM, Mermash S, *et al.* Breast cancer screening among females in Iran and recommendations for improved practice: A review. *Asian Pac J Cancer Prev* 2011;12:1647-55.
- MacMahon B, Cole P, Brown J. Etiology of human breast cancer: A review. *J Natl Cancer Inst* 1973;50:21-42.
- Simpson JF, Gray R, Dressler LG, Cobau CD, Falkson CI, Gilchrist KW, *et al.* Prognostic value of histologic grade and proliferative activity in axillary node-positive breast cancer: Results from the eastern cooperative oncology group companion study, EST 4189. *J Clin Oncol* 2000;18:2059-69.
- Frkovic-Grazio S, Bracko M. Long term prognostic value of Nottingham histological grade and its components in early (pT1N0M0) breast carcinoma. *Journal of clinical pathology* 2002;55: 88-92.
- Bundred NJ. Prognostic and predictive factors in breast cancer. *Cancer Treat Rev* 2001;27:137-42.
- Recht A, Come SE, Gelman RS, Goldstein M, Tishler S, Gore SM, *et al.* Integration of conservative surgery, radiotherapy, and chemotherapy for the treatment of early-stage, node-positive breast cancer: Sequencing, timing, and outcome. *J Clin Oncol* 1991;9:1662-7.
- Pourahmad S, Ayatollahi S, Taheri S. Fuzzy logistic regression: A new possibilistic model and its application in clinical vague status. *Iran J Fuzzy Syst* 2011;8:1.
- Pourahmad S. Fuzzy Logistic Regression Models with their Application in Medical. Shiraz, University of Medical Sciences: LAP LAMBERT Academic Publishing; 2013.
- Cooper RG, Holland JF, Glidewell O. Adjuvant chemotherapy of breast cancer. *Cancer* 1979;44:793-8.
- Bonadonna G, Valagussa P. Dose-response effect of adjuvant chemotherapy in breast cancer. *N Engl J Med* 1981;304:10-5.
- Cuzick J, Stewart H, Peto R, Baum M, Fisher B, Host H, *et al.* Overview of randomized trials of postoperative adjuvant radiotherapy in breast cancer. *Cancer Clin Trials* 1988;111:108-29.
- Fulford LG, Easton DF, Reis-Filho JS, Sofronis A, Gillett CE, Lakhani SR, *et al.* Specific morphological features predictive for the basal phenotype in grade 3 invasive ductal carcinoma of breast.



- Histopathology 2006;49:22-34.
20. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: Experience from a large study with long-term follow-up. *Histopathology* 1991;19:403-10.
21. Zainuddin NH, Safiuh LM. A Relationship between Tuberculosis and Smoking: Fuzzy Logistic Regression Approach. *Empowering Science, Technology and Innovation Towards a Better Tomorrow*; 2011.
22. Pourahmad S, Ayatollahi SM, Taheri SM, Agahi ZH. Fuzzy logistic regression based on the least squares approach with application in clinical studies. *Comput Math Appl* 2011;62:3353-65.

## APPENDIX 1

**Fuzzy number definitions** - A fuzzy number is a quantity whose value is imprecise, as is the case with “ordinary” numbers. Let  $X$  be a nonempty set of elements  $x$ , each fuzzy subset of  $x$  is defined by membership function as  $\tilde{A}: X \rightarrow [0,1]$ . The grade of membership of  $X$  in fuzzy subset for each  $x \in X$  identifies by  $\tilde{A}(x)$ .

We define one of the basic types of fuzzy numbers used in practice. A triangular fuzzy number is defined by three numbers and there are two types of triangular fuzzy number: symmetric and nonsymmetric.

**Triangular Fuzzy Number** - Let be a triangular  $\tilde{A}$  fuzzy number on  $X$  that can be written as  $(s_L, m, s_R)_T$  where  $m$  identifies the mode of triangular fuzzy number and  $s_L, s_R$  are left and right bound of triangular fuzzy number respectively.  $\tilde{A}(x)$  is the membership function as  $\tilde{A}: X \rightarrow [0,1]$  that identifies the grade of membership of  $X$  in fuzzy subset for each  $x \in X$ .<sup>[15]</sup>

$$\tilde{A}(x) = \begin{cases} 0 & x < s_L \\ L((x - s_L) / (m - s_L)) & s_L \leq x \leq m \\ R((s_R - x) / (s_R - m)) & m \leq x \leq s_R \\ 0 & x > s_R \end{cases} \tag{Eq 4}$$

If the left and right bound are equal, then  $\tilde{A}(x)$  is called a symmetrical fuzzy number. On the other hand the support of  $A$  is the set of elements  $\{x \in X \mid s_L < x < s_R\}$ . When,  $s_L = s_R = m$  it is a nonfuzzy number.<sup>[16]</sup>

In this paper, fuzzy form of  $W$  and  $b$  introduced as  $\tilde{W}_i = (f_{is}^c(x), f_i^c(x))_T$  and  $\tilde{b}_j = (s_j, m_j)$ .

**Possibilistic odds** - If  $\tilde{A}$  is a fuzzy number and  $\tilde{W}_i$  is its membership function and  $f(x) = \exp(x)$ , then  $f(\tilde{A}) = \exp(\tilde{A})$  is a fuzzy number with following membership function:

$$\exp(\tilde{A}) = \begin{cases} \tilde{A}(\ln(x)) & x > 0 \\ 0 & \text{otherwise} \end{cases} \tag{Eq 5}$$

The following formula is used for calculating  $\exp(W(x))$ (possibilistic odds).

$$\exp(\tilde{W}_i(x)) = \tilde{W}_i(\ln(x)) = \begin{cases} 1 - \frac{f_i^c(x) - \ln(x)}{f_{is}^L(x)}, f_i^c(x) - f_{is}^L(x) \leq \ln(x) \leq f_i^c(x) \\ 1 - \frac{\ln(x) - f_i^c(x)}{f_{is}^R(x)}, f_i^c(x) \leq \ln(x) \leq f_i^c(x) + f_{is}^R(x) \end{cases} \tag{Eq 6}$$

The membership function of fuzzy estimated output is as follow:

$$\tilde{W}_i(w_i) = \begin{cases} 1 - \frac{f_i^c(x) - w_i}{f_{is}^L(x)}, f_i^c(x) - f_{is}^L(x) \leq w_i \leq f_i^c(x) \\ 1 - \frac{w_i - f_i^c(x)}{f_{is}^R(x)}, f_i^c(x) \leq w_i \leq f_i^c(x) + f_{is}^R(x) \end{cases} \tag{Eq 7}$$

We supposed two assumptions for estimating

1. For each observation,  $h$  is membership degree. so:

$$W_i(w_i) \geq h \text{ and } w_i = \ln\left(\frac{\mu_i}{1 - \mu_i}\right), h \in (0,1) \tag{Eq 8}$$

2. Minimization is as bellow equations:

$$Z = n(s_0^L + s_0^R) + \sum_{j=1}^n ((s_j^L + s_j^R) \sum_{i=1}^m x_{ij}) \tag{Eq (9)}$$

$$1 - \frac{f_i^c(x) - w_i}{f_{is}^L(x)} \geq h \Rightarrow (1-h)s_0^L + (1-h) \sum_{j=1}^n s_j^L x_{ij} - m_0^c - \sum_{j=1}^n m_j^c x_{ij} \geq -w_i \tag{Eq (10)}$$

$$1 - \frac{w_i - f_i(x)}{f_{is}(x)} \geq h \Rightarrow (1-h)s_0^R + (1-h) \sum_{j=1}^n s_j^R x_{ij} - m_0^c - \sum_{j=1}^n m_j^c x_{ij} \geq w_i \tag{Eq (11)}$$

The following formula is used for calculating exp (W(x)) (possibilistic odds).

$$\exp(\widetilde{W}_i(x)) = \widetilde{W}_i(\ln(x)) = \begin{cases} 1 - \frac{f_i^c(x) - \ln(x)}{f_{is}^L(x)}, & f_i^c(x) - f_{is}^L(x) \leq \ln(x) \leq f_i^c(x) \\ 1 - \frac{\ln(x) - f_i^c(x)}{f_{is}^L(x)}, & f_i^c(x) \leq \ln(x) \leq f_i^c(x) + f_{is}^R(x) \end{cases} \tag{Eq (12)}$$

**Defuzzification** - Changing fuzzy number to real number is called defuzzification. Defuzzification formula for  $\widetilde{W}_i$  is defined:

$$def(\widetilde{W}) = \frac{\int_x \widetilde{W}(x) dx}{\int_x \widetilde{W}(x) dx} \tag{Eq (13)}$$