Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 32 (2007) 987–994
Expert Systems with Applications
www.elsevier.com/locate/eswa

Comparison of logistic regression model and classification tree: An application to postpartum depression data

Handan Ankarali Camdeviren [a], Ayse Canan Yazici [b,*], Zeki Akkus [c], Resul Bugdayci [d], Mehmet Ali Sungur [a]

[a] Biostatistics Department, Faculty of Medicine, Mersin University, Mersin, Turkey
[b] Biostatistics Department, Faculty of Medicine, Baskent University, Baglica Campus, 06530 Ankara, Turkey
[c] Biostatistics Department, Faculty of Medicine, Dicle University, Diyarbakir, Turkey
[d] Public Health Department, Faculty of Medicine, Mersin University, Mersin, Turkey

Abstract

In this study, it is aimed that comparing logistic regression model with classification tree method in determining social-demographic risk factors which have effected depression status of 1447 women in separate postpartum periods. In determination of risk factors, data obtained from prevalence study of postpartum depression were used. Cut-off value of postpartum depression scores that calculated was taken as 13. Social and demographic risk factors were brought up by helping of the classification tree and logistic regression model. According to optimal classification tree total of six risk factors were determined, but in logistic regression model 3 of their effect were found significantly. In addition, during the relations among risk factors in tree structure were being evaluated, in logistic regression model corrected main effects belong to risk factors were calculated. In spite of, classification success of maximal tree was found better than both optimal tree and logistic regression model, it is seen that using this tree structure in practice is very difficult. But we say that the logistic regression model and optimal tree had the lower sensitivity, possibly due to the fact that numbers of the individuals in both two groups were not equal and clinical risk factors were not considered in this study. Classification tree method gives more information with detail on diagnosis by evaluating a lot of risk factors together than logistic regression model. But making correct selection through constructed tree structures is very important to increase the success of results and to reach information which can provide appropriate explanations.

Keywords: Classification and regression trees; Logistic regression model; Cross-Validation; Postpartum depression; Diagnostic models

1. Introduction

Classification methods are commonly used in medicine particularly with the purpose of diagnosing (Harper et al., 2003). Usability of these methods increases parallel with developments in statistical packet programs. These methods usually evaluate more than one variable together and are examined in multivariate analyses group. If dependent variable consists of two (binary) or more (multinomial) categories, taking more than one risk factor or predictor variables together into the model with the purpose of estimating the values of dependent variable or correct classifying that will be increased the success in classification. Classification models are being used commonly with this purpose in discriminant analysis, logistic regression analysis, cluster analysis and neural network (Breiman, Friedman, Olshen, & Stone, 1984; Cappelli, Mola, & Siciliano, 1998; Hosmer & Lemeshow, 1989).

Logistic regression and Classification Trees (CT) are the models being used for estimating class membership of categorical dependent variable without getting any assumption

* Corresponding author. Tel.: +90 312 2341010/1637 (O).
E-mail addresses: acyazici@baskent.edu.tr, aysecanan@yahoo.com (A.C. Yazici).

0957-4174/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2006.02.022

on independent variables (Breiman et al., 1984; Buntine, 1992; Cappelli, Mola, & Siciliano, 2002; Hosmer & Lemeshow, 1989; Kerby, 2003; Olaru & Wehenkel, 2003; Siciliano & Mola, 2000; Terin, Schmid, Griffith, D'Agostino, & Sekler, 2003). These methods are very popular in machine learning applications, computer science (data structures), botany (classification), and psychology (decision theory) and are also used as prognostic models in medicine. Nowadays, logistic regression models are used commonly with the purpose of determining risk factors in medical researches and diagnose. In last a few years CTs are attractive because they provide a symbolic representation that lends itself to easy interpretation by humans (Abu-Hanna & de Keizer, 2003; Breiman et al., 1984; Fu, 2004; Kline et al., 2003; Robnik-Sikonja, Cukjati, & Kononenko, 2003).

The aim of this study is to examine logistic regression and CT methods comparatively in term of results obtained. In direction of this purpose, summarized theoretical explanations belong to both two methods were made and results obtained by examining effects of some social-demographic features on postpartum depression with these methods were compared controversially.

2. Material and methods

2.1. Sampling procedure

This cross-sectional study was conducted in 2001, in the province of Mersin in southern Turkey on the coast of the Mediterranean. In this region, there were 58,094 women aged between 15 and 44. A multi-step, stratified (for age groups) cluster sampling method was used. *In the first step*, seven of the 20 primary health centers in Mersin Provincial Center were randomly selected. Single women and pregnant women were excluded. *In the second step*, women were separated into groups according to postpartum periods. As there is no consistently identified grouping method for postpartum periods, the time periods arbitrarily selected were: (i) 0–2 months, (ii) 3–6 months, (iii) 7–12 months, (iv) 13 months and more. *In the third step*, women were selected systematically from each group, depending on weight and age groups.

Estimating PPD prevalence as 15%, a sample size of 1477 would represent a population of 58,094 people with a reliability of 95%. We planned to reach 1550 women for four groups. The 68 women who could not be found at home after two visits and the 35 women who didn't want to participate were excluded, leaving 1447 (93.4%) women (Buğdaycı, Şaşmaz, Tezcan, Kurt, & Öner, 2004; Engindeniz, Küey, & Kültür, 1997).

2.2. Statistical analysis

2.2.1. Classification trees

The CT has a tree structure in which an internal node denotes a variable, the branches of a node denote value (or value ranges) of the corresponding risk factor and a leaf denotes a (dominant) class. The CT construction is achieved by recursively partitioning sets beginning with the whole dataset. Each partitioning of a set is based on a corresponding value partitioning of some risk factors. In each of the recursive iterations, the aim is to find the risk factor, along with its value-partitioning, that can result in subsets which are maximally homogeneous (pure) in their class value. The first node where division starts is called family node, the nodes which continue division are called child node and the nodes where division finishes or homogeneity occurs are called terminal node (Abu-Hanna & de Keizer, 2003; Fu, 2004; Lewis, 2004).

2.2.2. Logistic regression models

In logistic regression models, dependent variable is always in categorical form and has two or more levels. Independent variables may be in numerical or categorical form. The binary multiple logistic regression model is defined as below:

$$g(x) = \ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \ln\left[\frac{P(y=1|x)}{P(y=0|x)}\right] = \beta_0 + \sum_{i=1}^{p} \beta_i x_i$$

The log-likelihood function is used for estimating regression coefficients (β_i) in model. Coefficients are obtained by iterative methods. Exponential value of regression coefficients (e^{β}) gives odds ratio and this value reflects the effect of risk factor in the disease and the interpreted values are odds ratios. The Wald test is used commonly, in hypothesis test of model coefficients. In addition, after model obtained a classification table is obtained as in other classification methods.

In CT and logistic regression model, seven risk factors are used. Total 1447 women were being included into the study were called Learning Sample.

In calculations, EPI-INFO 6.0 (Dean, Dean, Burton, & Dicker, 1990) and Statistica® 6.0 (STATISTICA AFA) statistical packet programs were used.

3. Results

Information about the characteristics and descriptive statistics belong to social-demographic risk factors are being included into the study were given as frequencies, percent, Mean ± SD in Table 1.

3.1. Results of CT

When Table 2 is examined, there are 30 different tree structures for this data set. Complexity of tree structures decreases from Tree 1 to Tree 30. The number of terminal nodes is used as complexity measurements. In selection of optimal tree structure, it is considered that cost-complexity measures are balanced and minimum. In condition that it is balanced, predictive accuracy of tree increases. Through the tree structures given in Table 2, the tree numbered 27 balancing the cost of misclassification (Cross-Validation

Table 1
Descriptive statistics of risk factors according to groups (number (%) or Mean ± SD)

Risk factors	Category	Non-depression ($n = 906$)	Depression ($n = 541$)
Occupation of women	Housewife	754 (%83.2)	468 (%86.5)
	Others	152 (%16.8)	73 (%13.5)
Education of women	No literate	26 (%2.9)	23 (%4.3)
	Literate	13 (%1.4)	22 (%4.1)
	Primary school	379 (%41.8)	245 (%45.3)
	Junior high school	110 (%12.1)	79 (%14.6)
	High school	266 (%29.4)	126 (%23.3)
	University	112 (%12.4)	46 (%8.5)
Education of husband	No literate	9 (%1.0)	5 (%0.9)
	Literate	3 (%0.3)	2 (%0.4)
	Primary school	262 (%28.9)	210 (%38.9)
	Junior high school	142 (%15.7)	100 (%18.5)
	High school	333 (%36.8)	155 (%28.7)
	University	157 (%17.3)	68 (%12.6)
Occupation of husband	Unemployed	39 (%4.3)	56 (%10.4)
	Employed	867 (%95.7)	484 (%89.6)
Postpartum months	0–8 week	164 (%18.1)	67 (%12.4)
	3–6 month	208 (%23.0)	120 (%22.2)
	7–12 month	240 (%26.5)	135 (%25.0)
	⩾13 month	294 (%32.5)	219 (%40.5)
Age of women		27.6 ± 5.45	27.4 ± 5.45
Age of marriage		21.1 ± 3.77	20.3 ± 3.55

Table 2
Cost-complexity measures of all possible trees

All possible trees	Terminal nodes number	CV cost	CV std. error	Resubstitution cost	Node complexity
Tree 1	328	0.445059	0.013065	0.123704	0.000000
Tree 2	321	0.434692	0.013032	0.125086	0.000197
Tree 3	297	0.434692	0.013032	0.130615	0.000230
Tree 4	292	0.429164	0.013012	0.131997	0.000276
Tree 5	244	0.428473	0.013009	0.148583	0.000346
Tree 6	237	0.416724	0.012961	0.151348	0.000395
Tree 7	232	0.416724	0.012961	0.153421	0.000415
Tree 8	216	0.416033	0.012958	0.160332	0.000432
Tree 9	202	0.416724	0.012961	0.166551	0.000444
Tree 10	187	0.415342	0.012954	0.173462	0.000461
Tree 11	174	0.413269	0.012945	0.179682	0.000478
Tree 12	167	0.412578	0.012942	0.183138	0.000494
Tree 13	151	0.411196	0.012935	0.191431	0.000518
Tree 14	141	0.407049	0.012915	0.196959	0.000553
Tree 15	135	0.403594	0.012898	0.200415	0.000576
Tree 16	127	0.405667	0.012908	0.205252	0.000605
Tree 17	75	0.383552	0.012783	0.241189	0.000691
Tree 18	69	0.380788	0.012765	0.246717	0.000921
Tree 19	59	0.381479	0.012770	0.256393	0.000968
Tree 20	47	0.376641	0.012738	0.268832	0.001037
Tree 21	44	0.373186	0.012714	0.272287	0.001152
Tree 22	37	0.374568	0.012724	0.280581	0.001185
Tree 23	28	0.366966	0.012670	0.293020	0.001382
Tree 24	24	0.371113	0.012700	0.299931	0.001728
Tree 25	20	0.366966	0.012670	0.307533	0.001900
Tree 26	15	0.360055	0.012619	0.317899	0.002073
***Tree 27**	**9**	**0.369730**	**0.012690**	**0.333794**	**0.002649**
Tree 28	4	0.377332	0.012743	0.355218	0.004285
Tree 29	3	0.378715	0.012752	0.360746	0.005529
Tree 30	1	0.373877	0.012719	0.373877	0.006565

cost = CV cost and Resubstitution cost), the complexity parameter (a penalty for additional terminal nodes) and the number of terminal nodes (T), and marked "*" was used in classification. In this tree, it is seen that the CV cost, the Resubstitution cost and the complexity parameter values are minimum and in addition to this the Resubstitution cost value is the nearest one to CV cost ± 1 SE boundaries. In the tree structures including a lot of terminal nodes, the CV cost value got higher values than the Resubstitution cost. Only in optimal tree structure, the CV cost and the Resubstitution cost have made the most appropriate balance constructed (Table 2). In the tree structure including one terminal node, these two misclassification ratios got values that are equal to each other.

When the results obtained from other tree structures are generated, in Tree 18 and the other tree structures which are more complex (the number of terminal nodes $\geqslant$69) classification success of individuals with postpartum depression rose up over 50% but no significant variation in classification success of individuals without depression was observed (Table 2).

The optimal tree structure numbered 27 obtained at the end of the pruning was drawn clearly in Fig. 1.

In Fig. 1, the nodes that demonstrated with dark colored squares are child node, the nodes that demonstrated with gray colored squares are terminal node. In this condition, in the optimal tree is being constructed, there are totally 17 nodes that consist of eight child nodes and nine terminal nodes. ID placing at the left top of the corner of nodes shows the number of that node (Node #), and N placing at the right top of the corner shows the total number of the individuals placing in that node (Fig. 1). The first node placing in the tree is rote node and the first discriminator split this node into two child nodes as husbands' education levels of women are being included into the study. Total 734 women whose husbands' education levels are maximum 8 years were allocated to the left child node; total 713 women whose husbands' education levels are more than 8 years were allocated to the right child node. Four hundred and sixteen of people placing in the left child node are individuals without depression and 318 of them are individuals with depression, and this node was divided again by helping of a second discriminator because of that the purity rule was not provided in this node or this node was not homogeneous enough. The right child node numbered 3 is a terminal node any more. Four hundred and ninety of the women placing in this node are individuals without depression and 223 of them are individuals with depression. According to the purity rule in this node was named group of individuals without depression. The discriminator is being used for making pure the node numbered 2 that is not pure yet is women's husbands' occupations. Six hundred and fifty five of 734 women whose husbands work were allocated to the left child node, 79 women whose husbands do not work

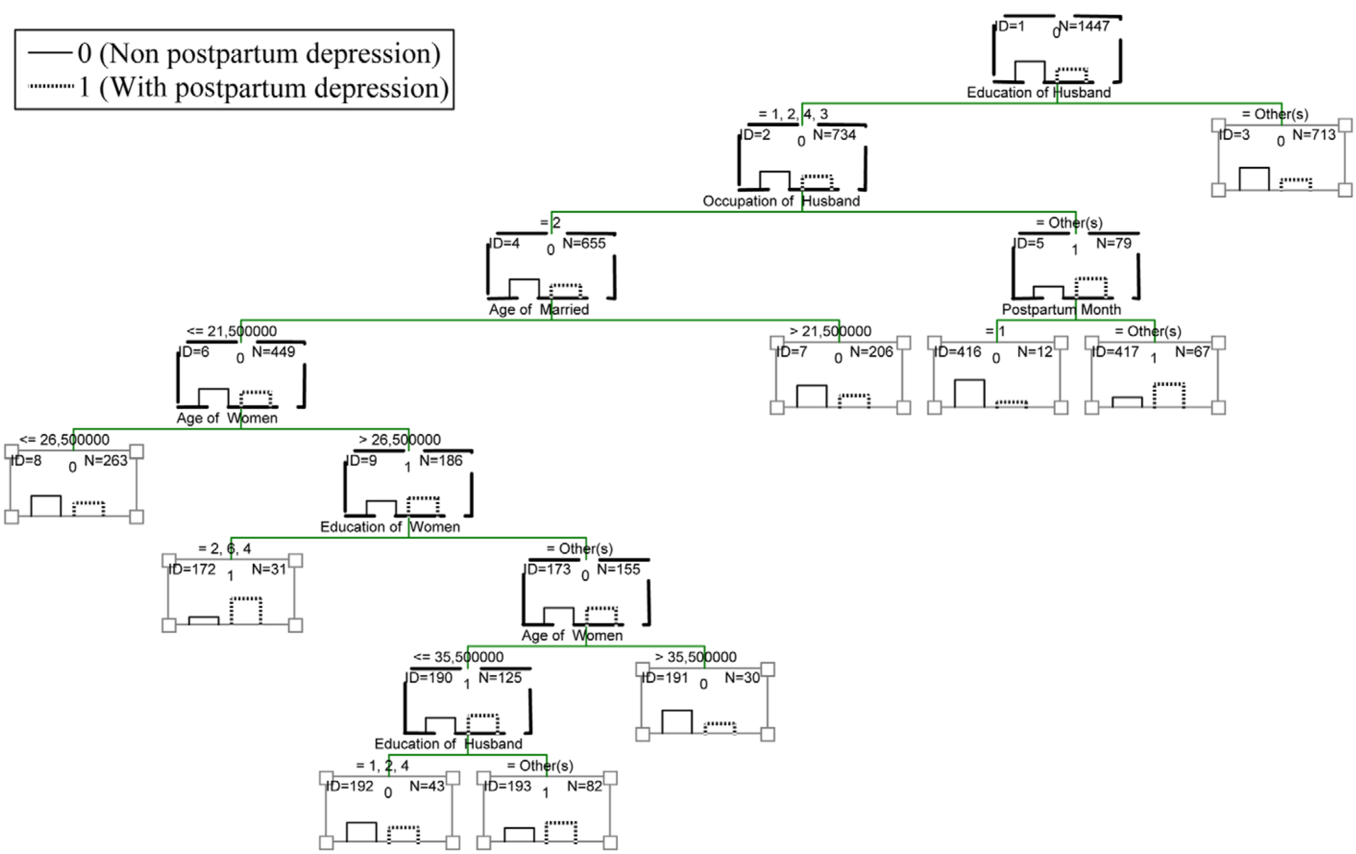

Fig. 1. Optimum tree for postpartum depression data.

were allocated to the right child node, and the tree advanced one more step because of that homogeneity enough was not provided (Fig. 1). The left child node containing total 655 women was split again according to the women's ages of marriage and the right node numbered 7 containing total 206 women came into being as terminal node. The age of marriage's cut-off value discriminating best in this step is 21.5. One hundred and thirty seven of 206 women are individuals without depression, 69 of them are individuals with depression, and according to the purity rule this node was named group of individuals without depression, too. The right child node numbered 5 containing total 79 individuals was split into two terminal nodes numbered 416 and 417, and in this discrimination postpartum months were used as discriminator. According to this discriminator the individuals who are in 0–8 weeks period after delivery were allocated to the left terminal node numbered 416, the others were allocated to the right terminal node numbered 417. There are total 12 individuals in the left terminal node, 10 of these are women without depression, two of these are women with postpartum depression, and this node is allocated as group of individuals without depression. There are total 67 individuals in the right terminal node, and 20 of these are women without depression, 47 of these are women with postpartum depression. In this condition, this node is allocated as group of individuals with depression. The child node # 6 was split again according to the ages of women, and the left terminal node # 8 and the right child node # 9 occurred. In this division, the most appropriate cut-off value belong to ages of women was determined as 26.5. The terminal node # 8 containing total 263 individuals were allocated as a group of individuals without depression because of that 162 of these 263 individuals are individuals without depression, 101 of them are individuals with depression. The right child node # 9 was split into two nodes # 172 and 173 according to the women's education levels. The left node # 172 is terminal node and occurred from ones whose education levels were literate, secondary school and high school, the right child node # 173 occurred from ones whose education levels were not literate and were graduated from high school. In the left node that is terminal, there are total 31 women and seven of them are individuals without depression, 24 of them are individuals with depression. In this condition, the terminal node # 172 was named group of individuals with depression. The child node # 173 was split again into two according to the women's ages, and the left child node # 190 and the right terminal node # 191 occurred. In this division, the most appropriate cut-off value belonging to ages of women was determined as 35.5. The terminal node # 191 containing total 30 individuals was named group of individuals without depression because 21 of these 30 individuals are individuals without depression, nine of them are individuals with depression. As last, total 192 women in the child node # 190 whose ages are equal to 35.5 and less than it, were split again into two according to their husbands' education levels, and ones whose husbands are not literate, are literate and are graduated from high school, were allocated to the left terminal node # 192, and the others were allocated to the right terminal node # 193. The left terminal node was named group of individuals without depression because the numbers of individuals without depression existing in this node are more than the numbers of individuals with depression, and the right terminal node was named group of individuals with depression because the individuals with depression in this node are more.

As a summary, six of nine terminal nodes obtained from the optimal tree are determined as a group of women without depression (Node #: 8, 192, 7, 416, 3 and 191), and three of them are determined as a group of women with depression (Node #: 172, 193 and 417). In this condition, it can be told that in the conditions summarized below the postpartum depression occurs:

(a) Depression risk is less in the women whose husbands' education levels are more than 8 years (Node # 3).
(b) Depression risk is less in the women whose husbands' education levels are maximum 8 years and whose husbands work, if the ages of marriage are 21.5 and more (Node # 7).
(c) Depression risk is less in the women whose husbands' education levels are maximum 8 years, whose husbands do not work and who spend maximum 8 weeks after delivery (Node # 416).
(d) In addition, depression risk increases in the women whose husbands' education levels are maximum 8 years, whose husbands do not work and who spend more than 8 weeks after delivery (Node # 417).
(e) Depression risk decreases in the women whose husbands' education levels are maximum 8 years, and whose husbands work if the ages of marriage are maximum 21.5 and the ages when the measurements were taken are maximum 26.5 (Node # 8).
(f) Depression risk increases as the women's education levels increase, in the women whose husbands' education levels are maximum 8 years, and whose husbands work, if the ages of marriage are maximum 21.5 and the ages when the measurements were taken are older than 26.5 (Node # 172).
(g) Depression risk decreases as the women's education levels decrease, in the women whose husbands' education levels are maximum 8 years, and whose husbands work, if the ages of marriage are maximum 21.5 and the ages when the measurements were taken are older than 35.5 (Node # 191).
(h) Postpartum depression risk increases as women's husbands' education levels increase (Node # 193) and decreases as women's husbands' education levels decrease (Node # 192) in the women whose husbands' education levels are maximum 8 years, and whose husbands work if the ages of marriage are maximum 21.5, women's education levels are low and the ages when the measurements were taken are maximum 35.5.

The successes of correct classification the women of optimal tree and maximal tree the being used in the study were given in Tables 3 and 4 respectively. When Tables 3 and 4 are examined as comparative, no significant variation in the success of correct classification of the individuals without depression was observed (93.68% for optimal tree and 92.16% for maximum tree), but when maximal tree (Tree 1) is used, a significant increasing in the success of correct classification of the individuals with depression was observed (22% for optimal tree and 80.04% for maximal tree). According to this result, it can be told that the maximal tree's success of diagnosing of postpartum depression is rather high. However, the maximal tree is very complex (the numbers of terminal node are 328), and its appropriateness with other data sets decreases because of its excessive appropriateness with Learning sample. In addition we say that optimal tree had the lower sensitivity, possibly due to the fact that numbers of the individuals with or without depression were not equal and clinical risk factors were not considered in this study.

Table 3
Cost matrix or classification table for optimal tree

Predicted	Observed		Total
	Non-depression	With depression	
Non-depression	905	422	1327
With depression	61	119	180
Total	966	541	1447

Specificity 94% (=905/966), Sensitivity 22% (=119/541), Resubstitution cost 34% (=(422 + 61)/1447), Total accuracy rate 71% (=1024/1447).

Table 4
Cost matrix or classification table for maximum tree

Predicted	Observed		Total
	Non-depression	With depression	
Non-depression	835	108	943
With depression	71	433	504
Total	906	541	1447

Specificity 92% (=835/906), Sensitivity 80% (=433/541), Resubstitution cost 12% (=(108 + 71)/1447), Total accuracy rate 88% (=1268/1447).

3.2. Results of logistic regression model

Seven risk factors used in CT analysis were examined by being taking the model and it is seen that occupation of Husband, Postpartum Mounts and woman's age of marriage from these factors effect postpartum depression statistically significant (Table 5). When categories of these 3 risk factor were evaluated, it is seen that postpartum depression is 2.1 multiple more in the women whose husbands do not work, women who spent 3–6 months after delivery have 1.52 multiple significantly higher depression risk than women who spent 0–8 weeks, and women who spent 13 months after delivery have 1.75 multiple much more depression risk than women who spent 0–8 weeks. In addition, depression risk decreased 0.953 multiple as woman's

Table 5
Odds ratios and its 95% *Confidence interval* of risk factors in logistic regression model

Risk factors	Category	OR (95% Confidence interval)	*P*
Occupation of women	Housewife **(reference)**	**1.0**	–
	Others	1.18 (0.78–1.78)	0.429
Education of women	No literate **(reference)**	1.0	–
	Literate	2.2 (0.87–5.59)	0.098
	Primary school	0.90 (0.48–1.693)	0.747
	Junior high school	1.16 (0.58–2.33)	0.666
	High school	0.87 (0.44–1.73)	0.702
	University	0.78 (0.35–1.77)	0.560
Education of husband	No literate **(reference)**	**1.0**	–
	Literate	2.39 (0.27–20.9)	0.430
	Primary school	2.41 (0.71–8.20)	0.158
	Junior high school	2.36 (0.68–8.20)	0.178
	High school	1.57 (0.45–5.42)	0.473
	University	1.62 (0.46–5.74)	0.458
Occupation of husband	Unemployed	*2.10 (1.35–3.27)*	*0.001***
	Employed **(reference)**	**1.0**	–
Postpartum months	0–8 Week **(reference)**	**1.0**	–
	3–6 month	*1.52 (1.05–2.20)*	*0.027**
	7–12 month	1.37 (0.95–1.97)	0.088
	⩾13 month	*1.75 (1.24–2.47)*	*0.001**
Age of women		1.01 (0.99–1.04)	0.317
Age of marriage		*0.953 (0.92–0.98)*	*0.009**

* $p < 0.05$.
** $p < 0.01$.

Table 6
Cost matrix or classification table for logistic regression model

Predicted	Observed		Total
	Non-depression	With depression	
Non-depression	859	454	1313
With depression	47	87	134
Total	906	541	1447

Specificity 95% (=87/541), Sensitivity 16% (=859/906), Resubstitution cost 34.6% (=(454 + 47)/1447), Total accuracy rate 65.4% (=(859 + 87)/1447).

age of marriage increased and this decreasing was found statistically significant (Table 5).

In addition, logistic regression model's classification success of individuals with or without depression was given in Table 6. When classification success of this model were evaluated, it was rather successful with 95% success of discriminate the individuals without depression. But its determination success of the individuals with depression (16%) was found low.

3.3. Comparison of CT and logistic regression model

It is found that three of seven risks factors examined in Logistic Regression Model effect depression significantly, in spite of this six of them are effective in formation of optimal tree. Three risk factors being found significantly in Logistic Regression Model were also found significantly in CT method. But adjusted main effects of these factors in logistic regression model were obtained, in spite of this in CT method of these three factors with each other and other risk factors were brought up. In this condition, the results obtained from CT method are more detail and explain biologic structure better. For instance, while in logistic regression model depression was found significantly high in women whose husbands do not work, in CT method while it is seen that depression risk increases significantly in women whose husbands' education levels are less than 8 years and if their husbands do not work, a relationship between depression and their husbands' working condition couldn't been determined in women whose husbands' education levels are more than 8 years. In addition, maximal tree classified individuals with depression more successful than both optimal tree and logistic regression model. But this tree structure is rather complex and its harmony to new data gets is low. In spite of this, in optimal tree, classification success of individuals with depression and total correct classification success of tree were found a little higher than logistic regression model, but other classification success were found similar.

4. Conclusion

In this study, social-demographic risk factors of postpartum depression occurred in women after delivery by using CT and logistic regression model. As it is determined in other some researchers, it is seen that the postpartum depression risk increases particularly depending on the time after two months period after in this study, too (Buğdaycı et al., 2004; Heh & Fu, 2003). In addition, it was observed that depression risk is parallel with the increasing of the women's ages, the increasing in the women's education levels and the increasing in women's husbands' education levels increases depression risk in the women who married early, if the ages at the moment of giving birth are middle age or more.

In diagnosing studies, using more than one variable together will increase the diagnose success. In these kinds of researches, CT method gives successful results in term of evaluating these variables together and bringing up relations between variables (Abu-Hanna & de Keizer, 2003; Kline et al., 2003; Olaru & Wehenkel, 2003).

References

Abu-Hanna, A., & de Keizer, N. (2003). Integrating classification trees with local logistic regression in Intensive Care prognosis. *Artificial Intelligence in Medicine, 29*(1–2), 5–23.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. Wadsworth, Belmont, CA.

Buğdaycı, R., Şaşmaz, T., Tezcan, H., Kurt, A. Ö., & Öner, S. (2004). A cross sectional prevalence study of depression at various times after delivery in Mersin province in Turkey. *Journal of Women's Health, 13*(1), 65–68.

Buntine, W. L. (1992). Learning classification trees. *Statistics and Computing, 2*, 63–73.

Cappelli, C., Mola, F., & Siciliano, R. (2002). A statistical approach to growing a reliable honest tree. *Computational Statistics & Data Analysis, 38*(3), 285–299.

Cappelli, C., Mola, F., & Siciliano, R. (1998). An alternative pruning method based on the impurity-complexity measure. In R. Payne & P. Green (Eds.), *Proceedings in computational statistics*. Heidelberg: Physica-Verlag.

Dean, A. D., Dean, J. A., Burton, J. H., & Dicker, R. C. (1990). A word processing, database and statistics program for epidemiology on microcomputers. Centers for Disease Control, Atlanta, GA, USA.

Engindeniz, A. N., Küey, L., & Kültür, S. (1997). Edinburg dogum sonrası depresyon ölçegi Türkce formu geçerlilik ve güvenirlik calişmasi. In *Bahar Sempozyumlari 1:23–26 Nisan* (pp. 51–52). Antalya.

Fu, C. Y. (2004). Combining loglinear model with classification and regression tree (CART): An application to birth data. *Computational Statistics & Data Analysis, 45*(4), 865–874.

Harper, P. R., Sayyad, M. G., de Senna, V., Shahani, A. K., Yajnik, C. S., & Shelgikar, K. M. (2003). A systems modeling approach for the prevention and treatment of diabetic retinopathy. *European Journal of Operational Research, 150*, 81–91.

Heh, S. S., & Fu, Y. Y. (2003). Effectiveness of informational support in reducing the severity of postnatal depression in Taiwan. *Journal of Advanced Nursing, 42*(1), 30.

Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression*. New York, USA: John Wiley.

Kerby, D. S. (2003). CART analysis with unit-weighted regression to predict suicidal ideation from Big Five traits. *Personality and Individual Differences, 5*, 249–261.

Kline, J. A., Hernandez-Nino, J., Newgard, C. D., Cowles, D. N., Jackson, R. E., & Courtney, D. M. (2003). Use of pulse oximetry to predict in-hospital complications in normotensive patients with pulmonary embolism. *The American Journal of Medicine, 115*(3), 203–208.

Lewis, R. (2004). *An introduction to classification and regression tree cart analysis*. California: Academic Emergency Medicine (pp. 1–14).

Olaru, C., & Wehenkel, L. (2003). A complete fuzzy decision tree technique. *Fuzzy Sets and Systems, 138*(2), 221–254.

Robnik-Sikonja, M., Cukjati, D., & Kononenko, I. (2003). Comprehensible evaluation of prognostic factors and prediction of wound healing. *Artificial Intelligence in Medicine, 29*(1–2), 25–38.

Siciliano, R., & Mola, F. (2000). Multivariate data analysis and modeling through classification and regression trees. *Computational Statistics & Data Analysis, 32*(3–4), 285–301.

STATISTICA AFA (release 6.0). StatSoft, Inc./USA, All rigths reserved. 1984–2002: www.statsoft.com, 2300 E. 14th St., Tulsa, OK.

Terin, N., Schmid, C. H., Griffith, J. L., D'Agostino, R. B., & Sekler, H. P. (2003). External validity of predictive models: A comparison of logistic regression, classification trees, and neural Networks. *Journal of Clinical Epidemiology, 56*(8), 721–729.