Original Article
http://mjiri.iums.ac.ir

Medical Journal of the Islamic Republic of Iran (MJIRI)
Iran University of Medical Sciences

Prediction of unwanted pregnancies using logistic regression, probit regression and discriminant analysis

Farzad Ebrahimzadeh[1], Ebrahim Hajizadeh[2], Nasim Vahabi[3], Mohammad Almasian[4]
Katayoon Bakhteyar*[5]

Received: 16 January 2015 Accepted: 11 May 2015 Published: 19 September 2015

Abstract

Background: Unwanted pregnancy not intended by at least one of the parents has undesirable consequences for the family and the society. In the present study, three classification models were used and compared to predict unwanted pregnancies in an urban population.

Methods: In this cross-sectional study, 887 pregnant mothers referring to health centers in Khorramabad, Iran, in 2012 were selected by the stratified and cluster sampling; relevant variables were measured and for prediction of unwanted pregnancy, logistic regression, discriminant analysis, and probit regression models and SPSS software version 21 were used. To compare these models, indicators such as sensitivity, specificity, the area under the ROC curve, and the percentage of correct predictions were used.

Results: The prevalence of unwanted pregnancies was 25.3%. The logistic and probit regression models indicated that parity and pregnancy spacing, contraceptive methods, household income and number of living male children were related to unwanted pregnancy. The performance of the models based on the area under the ROC curve was 0.735, 0.733, and 0.680 for logistic regression, probit regression, and linear discriminant analysis, respectively.

Conclusion: Given the relatively high prevalence of unwanted pregnancies in Khorramabad, it seems necessary to revise family planning programs. Despite the similar accuracy of the models, if the researcher is interested in the interpretability of the results, the use of the logistic regression model is recommended.

Keywords: Unwanted Pregnancy, Logistic regression, Discriminant Analysis, Probit Regression, Khorramabad.

Cite this article as: Ebrahimzadeh F, Hajizadeh E, Vahabi N, Almasian M, Bakhteyar K. Prediction of unwanted pregnancies using logistic regression, probit regression and discriminant analysis. *Med J Islam Repub Iran* 2015 (19 September). Vol. 29:264.

Introduction

Unplanned or unwanted pregnancy is a type of pregnancy which is unintended by at least one of the parents (1,2). Every year, 75 million cases of unwanted pregnancies occur in the world and 50 million cases of unwanted pregnancies lead to abortion, of which 20 million take place under unsafe conditions (3-5). The prevalence of unwanted pregnancies in the USA, Japan, and Tanzania, has been reported as 48%, 46.2%, and 23.7%, respectively (6-8). Studies in Iran indicate that despite the easy availability of contraception, 400 to 500 thousand cases of unwanted pregnancies occur each year, 19% of which end in abor-

[1]. Biostatistics Instructor, Department of Statistics and Epidemiology, Faculty of Health and Nutrition, Lorestan University of Medical Sciences, Khorramabad, & PhD Candidate in Biostatistics, Department of Biostatistics, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran. farzadebrahimzadeh2012@gmail.com, ebrahimzadeh@modares.ac.ir
[2]. Associate Professor, Department of Biostatistics, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran. hajizadeh@modares.ac.ir
[3]. PhD Candidate in Biostatistics, Department of Biostatistics, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran. nasim.vahabi@modares.ac.ir
[4]. English Language Instructor, Department of the English Language, Faculty of Medicine, Lorestan University of Medical Sciences, Khorramabad, Iran. almasian2@gmail.com
[5]. (**Corresponding author**) Midwifery Instructor, Department of Public Health, Faculty of Health and Nutrition, Lorestan University of Medical Sciences, Khorramabad, Iran. k_bakhteyar@yahoo.com

tion (9-11). Based on reports published by the Iran Ministry of Health, 1 in every 4 pregnancies has been unwanted according to the parents (5,12). Based on two systematic reviews conducted in Iran, the prevalence of unwanted pregnancies has been estimated as 29.7% and 30.6% (1,2).

Unwanted pregnancies has possibly undesirable consequences for the child, the parents, and the society, the most common of which is attempting illegal and unsafe abortions which is one of the main causes of mortality and disability among mothers. Additionally, due to unpleasant feelings and emotions of the mother during pregnancy, the mother experiences more stress and nervous strain, and it may even have effects such as depression, suicide, and a reduction in the quality of life of the mother (3,10,13). The most important effects of unwanted pregnancies on the child can be pre-term birth and underweight babies, anorexia, hyperactivity, abuse and negligence by the parents, and finally infant mortality (5,12,3-16).

Social sanctions, sex-based social inequalities, opposition of spouses, unavailability of modern contraceptive methods, inadequate planning, lack of awareness of the role of counseling, lack of adequate skills among healthcare employees are among factors that can lead to an incompatibility between the recommended contraceptive method and the conditions of the individual and can lead to the disuse or failure of the contraceptive methods (4).

Given the importance of the issue and given the fact that the occurrence of unwanted pregnancies is affected by various and unique individual and social factors in each society, it is necessary to investigate the determining factors in each society separately. Using methods known as classification models makes it possible to predict the risk of unwanted pregnancies in the individuals based on a set of characteristics and variables related to the parents. The most common classification methods that can be used for this purpose include logistic regression, discriminant analysis, and probit regression. Regardless of their differences in terms of estimation methods and calculation algorithms, these methods and models are different from each other in terms of accuracy of classification, interpretability of the result, and the availability of statistical applications (17-19).

A review on unwanted pregnancies indicates that, in the majority of the cases, only the prevalence of the issue has been reported and most studies applied the independent t-test and chi square test to determine the relationship of each independent variables with the occurrence of unwanted pregnancy (3,5,10,12,13,20-26). In a few studies, it has been attempted to investigate the simultaneous effect of a set of variables on unwanted pregnancies using classification models, but due to the limitations in the selection of the statistical population, the results of these studies are mostly not generalizable (3,4,9,11,27). The statistical models used in previous studies include simple logistic regression (3,4,6,8,11,27-29), multinomial logistic regression (9,30), probit regression (27), and log-linear models (31), the last of which is not of course among the classification models.

Given the importance of the issue of unwanted pregnancies and the fact that, up to now, no comprehensive comparative study has been conducted in Western Iran to predict unwanted pregnancies, this study tried to determine the best classification model for the prediction of unwanted pregnancies to detect the most important determinants of unwanted pregnancy in the urban population of Khorramabad, using logistic regression, discriminant analysis, and probit regression.

Methods

Data

The study population in this cross-sectional research included all pregnant mothers referred to health centers in Khorramabad, Iran in 2012 to receive medical care during pregnancy. A sample consisting of 887 pregnant women were selected using stratified and cluster sampling

methods. First, the urban health centers in Khorramabad were divided into three strata of the north, center, and south, and two health centers (cluster heads) were randomly selected from the north, two from the south and three from the central area of the city. Next, mothers who referred to these health centers to receive medical care were entered into the sample in the order of referral.

The instrument was a researcher-designed data gathering form which was completed by the pregnant mothers; for illiterate or low literate mothers it was completed through an interview. The form included the name of the health center, the age of the couple, parity, the time interval between the present pregnancy and the previous one (pregnancy spacing), the number of living male and female children, the contraceptive method(s) used, educational attainment of the couple, income level of the household, the occupation of the couple, the residential substructure area, and the ownership of a personal vehicle, and some questions about whether the present pregnancy is wanted or unwanted from the viewpoint of at least one of the parents.

It is necessary to mention that all stages of recording and analyzing the data were conducted anonymously to preserve the confidentiality of the data, and in cases where it was necessary to interview mothers, female interviewers were employed.

Statistical Analysis

Logistic Regression: It is a regression model used to analyze binary response (success and failure) variables which is a member of the family of linear generalized models and uses the logit function as the link function. Logistic regression is capable of making predictions, estimating the coefficients, and the effect of each independent variable, and is also used for classification and recognition purposes (32).

Probit Regression: It is another regression model used to analyze two-state response variables. This model belongs to the family of generalized linear models and uses the probit function as the link function.

Discriminant Analysis: It is a method of classification and attribution of new observations into pre-defined groups. The attribution of observations to groups is done using the discriminant function. Fischer's discriminant function is one of the most widely used functions is discriminant analysis (33).

To predict unwanted pregnancy, and to compare the logistic regression, discriminant analysis, and probit regression models, SPSS version 21 was used. To select the main independent variables among various variables, the results of univariate analysis (chi-square test) were used and finally such variables as the mothers' age, the mother's educational attainment, the number of living male children, parity, pregnancy spacing, the contraceptive method used and the household income were selected. It is necessary to mention that, to avoid multicollinearity between the independent variables, the variables of the number of living female children, the age of the husband, and the educational attainment of the husband were excluded. Additionally, to avoid multicollinearity, the variables parity and pregnancy spacing were merged together into a new variable called "parity and pregnancy spacing".

In logistic regression, to select the studied variables for inclusion in the final model, the forward selection method was used. The significance level for including a variable in the model was set at 0.15 and the significance level for excluding a variable was set at 0.1. In discriminant analysis, Fischer's discriminant function was used to allocate new observations to pre-defined groups. To do the probit regression model, a generalized linear model with a binary probit link function was used. It is necessary to mention that in all three models, probability values less than 0.05 were considered as statistically significant. To compare the aforementioned models indicators such as sensitivity, specificity, the area under the ROC curve, and the percentage of correct predictions were used.

Results

In this study, the average mothers' age (±standard deviation) was 30.1 (±4.21) with an age range of 16 to 52. About 89.5% (793 individuals) of the mothers were younger than 35 years old and 88.1% (777 individuals) of their husbands were younger than 40 years old. The average age difference (±SD) between the husband and wife was 5.1 (±3.4) with a range of -8 to 15 years. Almost 31.1% (267 individuals) had a living male child, and 29.7% (255 individuals) a living female child. From among pregnant mothers, 47.1% (417) were pregnant for the first time, 35.7% (317 individuals) had been pregnant twice, and 17.2% (153 individuals) three times or more. Pregnancy spacing was more than four years in 34.8% (301 individuals) of the mothers. Almost 37.4% (332 individuals) used the natural traditional contraceptive methods and 15.3% (136 individuals) used no contraception. About 19.8% (176 individuals) of the pregnant mothers had a university degree and 23.7% (210 individuals) of their husbands also had a university degree. About 91% (807 individuals) were housewives and 51.3% (454 individuals) of their husbands were self-employed. About 23% (206 cases) of the families had a monthly income of less than 7.5 million Rials (about 214 US dollars). Almost 48% (418 cases) of the families lived in rented houses and 34% (302 cases) lived in houses with substructures less than 100 square meters. And finally 42.2% (373 cases) of the studied population had a personal vehicle. From among all of the pregnancies, 25.3% (224 cases) were unwanted. To investigate the relationship between the variables and the occurrence of unwanted pregnancies, the results of the univariate analysis is presented in Table 1.

Given the results of the univariate analysis, the occupation of the mother (p=0.664), the occupation of the husband (p=0.264), the age difference between the husband and wife (p=0.439), the substructure of the residence (p=0.338), and ownership of a personal vehicle (p=0.149) had no significant relationships with the occurrence of unwanted pregnancies. Variables such as the mothers and fathers' age, the educational attainment of the mother and her husband, the number of living male or female children, parity, pregnancy spacing, the contraception used before pregnancy, and the monthly household income had significant relationships.

The Findings of Logistic Regression: The results of the logistic regression model are presented in Table 3. The variables of parity and pregnancy spacing, the type of contraceptive used before pregnancy, the household income, and the number of living male children had a significant relationship with the occurrence of unwanted pregnancies. In addition, given the values of the odds ratios, the variable parity and pregnancy spacing had the highest impact on the occurrence of unwanted pregnancies, such that the odds ratio of unwanted pregnancy among mothers with a parity of three or more and a pregnancy spacing of less than 4 years was 6.88 times greater than among mothers who were pregnant for the first time (p<0.001). This model showed that the odds ratio of unwanted pregnancy among mothers with two or more than two living male children was 2.68 times greater than among mothers who had not any living male children (p=0.010). Also, among mothers who had used oral contraceptives, the odds ratio of unwanted pregnancy was about 2.37 times greater than the mothers who had not used any contraception (p=0.005) and also, among the couples who had used the condom, the odds ratio of unwanted pregnancy was almost 2.1 times more than the couples who had not used any contraceptive methods (p=0.035). Finally, the odds ratio of unwanted pregnancy among couples who had a monthly income of about 7.50 to 9.99 million Rials (about 214-285 USD) was 41% less than the couples with monthly income of less than 7.50 million Rials (less than about 214 USD) (p=0.013). It is necessary to mention that, the effect of such variables as the mother's age and educational attainment were not

Table1. The frequency distribution of all pregnancies and unwanted pregnancies among women referred to the health centers in Khorramabad in 2012

Variable	Category	Total pregnancies*	Unwanted pregnancies*	p**
Mother's age (years)	< 20	7.1 (63)	20.6 (13)	<0.001
	20-34	82.4 (730)	23.0 (168)	
	≥ 35	10.5 (94)	45.7 (43)	
Husband's age (years)	< 25	9.4 (84)	16.7 (14)	0.050
	25-39	78.7 (698)	25.0 (174)	
	≥ 40	11.9 (106)	32.1 (34)	
Age difference between husband and wife (years)	< 0	(71)	18.3 (13)	0.439
	0-4	(369)	24.4 (90)	
	5-9	(335)	26.6 (89)	
	≥10	(107)	28.0 (30)	
The number of living male children	0	68.9 (592)	18.8 (111)	<0.001
	1	24.0 (206)	34.0 (70)	
	≥ 2	7.1 (61)	67.2 (41)	
The number of living female children	0	70.3 (604)	21.2 (128)	0.003
	1	23.2 (199)	34.7 (69)	
	≥ 2	6.5 (56)	44.6 (25)	
Parity	1	47.1 (417)	15.8 (66)	<0.001
	2	35.7 (317)	22.4 (71)	
	≥3	17.2 (153)	56.9 (87)	
Pregnancy spacing (years)	First pregnancy	48.3 (417)	15.8 (66)	<0.001
	<2	7.2 (62)	51.6 (32)	
	2 - 4	9.6 (83)	42.2 (35)	
	≥ 4	34.9 (301)	29.6 (89)	
Mother's educational attainment	Illiterate	2.8 (25)	60.0 (15)	<0.001
	high school diploma or lower	77.4 (686)	24.8 (170)	
	University degree	19.8 (176)	22.2 (39)	
Husband's educational attainment	Illiterate	2.7 (24)	58.3 (14)	0.005
	high school diploma or lower	73.6 (653)	24.3 (159)	
	University degree	23.7 (210)	24.3 (51)	
Mother's occupation	Housewife	91.0 (807)	25.4 (205)	0.664
	Employee	6.9 (61)	21.3 (13)	
	Self-employed	2.1 (19)	31.6 (6)	
Husband's occupation	Unemployed	2.8 (25)	36.0 (9)	0.264
	Worker	16.0 (142)	29.6 (42)	
	Employee	29.8 (264)	22.7 (60)	
	Self-employed	51.3 (454)	24.9 (113)	
Monthly household income (Rials)	<7.5 million	23.2 (206)	32.5 (67)	0.004
	7.5 million-9.9 million	51.3 (455)	20.9 (95)	
	≥10 million	25.5 (226)	27.4 (62)	
Contraceptive method	None	15.3 (136)	16.2 (22)	<0.001
	Natural/Traditional	37.4 (332)	21.4 (71)	
	Condoms	13.2 (117)	29.1 (34)	
	Oral Contraceptive Pills	26.0 (231)	34.6 (80)	
	Other	8.0 (71)	23.9 (17)	
Ownership status of the residence	Personal	27.0 (237)	26.2 (62)	0.860
	Parent's	25.5 (224)	24.1 (54)	
	Rented	47.6 (418)	25.6 (107)	
The substructure of the residence (m^2)	< 100	34.4 (302)	28.1 (85)	0.338
	100-199	53.8 (473)	24.3 (115)	
	≥200	11.8 (104)	22.1 (23)	
Ownership of a personal vehicle	Yes	42.2 (373)	23.1 (86)	0.149
	No	57.8 (511)	26.4 (135)	

*: The numbers in each cell indicates % (frequency).
**: The statistical test used is chi-square.

statistically related to the occurrence of unwanted pregnancy.

The Findin gs of Probit Regression: The results of fitting the probit regression model are presented in Table 3. Similar to the logistic regression model, the variables of parity and pregnancy spacing, the type of contraception used before pregnancy, the income level of the household, the number of living male children were found to have

Table 2. The effect of the studies variables on the occurrence of unwanted pregnancies among pregnant women referring to health centers in Khorramabad in 2012 using the logistic regression model

Variable	Category	Odds ratio	95% Confidence interval	p
Mother's age (years)	< 20	Reference		
	20-34	0.695	0.346 – 1.395	0.306
	≥35	0.873	0.350 – 2.178	0.771
Parity and pregnancy spacing	First pregnancy	Reference		
	Second pregnancy with a spacing of < 4 years	3.735	2.061 – 6.769	< 0.001
	Second pregnancy with a spacing of ≥4 years	0.894	0.506 – 1.582	0.701
	Third pregnancy or higher with a spacing of < 4 years	6.877	2.879 – 16.427	< 0.001
	Third pregnancy or higher with a spacing of ≥4 years	3.750	1.931 – 7.283	< 0.001
The number of living male children	0	Reference		
	1	1.133	0.695 – 1.848	0.616
	≥2	2.684	1.265 – 5.693	0.010
Contraceptive method	None	Reference		
	Natural/Traditional	1.594	0.886 – 2.869	0.120
	Condoms	2.099	1.052 – 4.187	0.035
	Oral Contraceptive Pills	2.365	1.300 – 4.303	0.005
	Other	0.979	0.433 – 2.209	0.958
Mother's educational attainment	Illiterate	Reference		
	High school diploma and lower	0.797	0.301 – 2.110	0.647
	Bachelor's degree and higher	0.714	0.248 – 2.057	0.532
Household income	< 7.5 million Rials	Reference		
	7.5 million to 9.9 million Rials	0.590	0.389 – 0.896	0.013
	≥10 million Rials	0.781	0.475 – 1.282	0.328

Table 3. The effect of the studies variables on the occurrence of unwanted pregnancies among pregnant women referring to health centers in Khorramabad in 2012 using the probit regression model

Variable	Category	Estimated Coefficient	Std. Error	p
Mother's age (years)	< 20	Reference		
	20-34	-0.20	0.204	0.335
	≥35	-0.06	0.271	0.837
Parity and pregnancy spacing	First pregnancy	Reference		
	Second pregnancy with a spacing of < 4 years	0.77	0.180	< 0.01
	Second pregnancy with a spacing of ≥4 years	-0.08	0.163	0.632
	Third pregnancy or higher with a spacing of < 4 years	1.14	0.267	< 0.01
	Third pregnancy or higher with a spacing of ≥4 years	0.78	0.202	< 0.01
The number of living male children	0	Reference		
	1	0.07	0.147	0.634
	≥2	0.59	0.232	0.011
Contraceptive method	None	Reference		
	Natural/Traditional	0.26	0.166	0.124
	Condoms	0.43	0.199	0.033
	Oral Contraceptive Pills	0.50	0.177	0.004
	Other	0.03	0.231	0.900
Mother's educational attainment	Illiterate	Reference		
	High school diploma and lower	-0.15	0.302	0.622
	Bachelor's degree and higher	-0.22	0.325	0.508
Household income	< 7.5 million Rials	Reference		
	7.5 million to 9.9 million Rials	-0.30	0.124	0.016
	≥10 million Rials	-0.12	0.147	0.413

a significant relationship with the occurrence of unwanted pregnancies.

The Findings of Discriminant Analysis: Given the estimation of the standardized coefficients of the discriminant function, the general form of the model was obtained as is presented below, in which x_1 is the age of the mother, x_2 is the gravidity and pregnancy spacing, x_3 is the number of living male children, x_4 is the contraception used, x_5 is the educational attainment of the mother, and x_6 is the income level.

$F = 0.068\ X1 + 0.453\ X2 + 0.559\ X3 + 0.128\ X4 - 0.052\ X5 - 0.138\ X6$

Table 4. The values of sensitivity, specificity, the area under the ROC curve, and the percentage of correct predictions of the logistic regression, discriminant analysis, and probit regression models in the prediction of wanted and unwanted pregnancies among pregnancy women referring to health centers in Khorramabad in 2012

Method	AUC*	Sensitivity (%)	Specificity (%)	Correct prediction (%)
Logistic regression	0.735	73.1	64.5	70.7
Probit regression	0.733	72.5	64.2	70.4
Linear discriminant analysis	0.680	65.2	58.5	60.2

*: Area under curve

To compare the three aforementioned methods, the values of indices of sensitivity, specificity, the area under the ROC curve, and the percentage of correct predictions of each model are presented in Table 4. The area under the ROC curve was 0.735 for the logistic regression model and the percentage of correct predictions of this model was 70.7%. Given the reported values, the logistic regression model in general was more capable to predict and classify wanted and unwanted pregnancies.

Discussion

In the present study, it was attempted to predict the occurrence of unwanted pregnancies in Khorramabad, Iran based on variables such as the mothers' age and educational attainment, number of living male children, parity and pregnancy spacing, the type of contraceptive used before pregnancy, and the income level of the household, using logistic regression, probit regression, and linear discriminant analysis models. Based on results of the logistic and probit regression models, variables such as parity and pregnancy spacing, contraceptive methods, household income and number of living male children were related to unwanted pregnancy and the mother's age was not related to this main outcome. And finally, based on the indicator of the area under the ROC curve, the best methods were logistic regression and probit regression and finally linear discriminant analysis in order, and of course the results of logistic regression and probit regression were very similar.

The most important advantage of the logistic regression method, as a model-based method, is that it is capable to test and compare the impact of each independent variable on the occurrence of unwanted pregnancies. As a result, it offers desirable interpretability via a model which is presented in the format of a closed form (17-19,34,35). As an example, based on this study, it can be stated that low income level of the household can increase the odds ratio of occurrence of unwanted pregnancies. Despite the mentioned advantages, the logistic regression model is highly affected by high correlations between independent variables and the presence of nonlinear relationships (19).

In practice, the logistic regression model is more commonly used in observational studies, while the probit model is more frequently used in experimental studies. However, these two models present very similar results and the most important difference between them occurs when the relationship between an independent variable and a dependent variable is examined at the extreme points of the independent variable, in which case the logistic regression model offers better predictions (36,37). In a simulation study by Cakmakyapan et al (2013), the highest distinction between logistic regression and the probit model was observed when the relationship between the dependent variable and the other independent variables is strong, which in this case, for sample sizes greater than 500, the Pearson residuals of the logistic regression model is less than the probit model and as a result better fitting occurs, while in cases where the sample size is less than 100, better fitting has been observed with the probit model. It is necessary to mention that not much difference has been observed between the two models in other sample sizes (38).

In this study, the weakest predictions

were made by discriminant analysis, because it is a parametric method and is dependent on the assumption of normality of several variables, and as a result while this assumption is not true, it cannot be considered a suitable method for classification purposes. The results of simulation studies indicate that when most of the variables are qualitative and have a small number of levels, discriminant analysis performs more poorly than logistic regression (35). Another problem of discriminant analysis is the uninterpretability of the coefficients of the discriminant function (17-19,35).

In a study by Mohammadpour Asl et al which was conducted on 1576 women living in Tabriz, Iran in 2004, the prevalence of unwanted pregnancies was 26.7%. The logistic regression method showed that factors affecting unwanted pregnancies included higher maternal age and higher numbers of living children. Moreover, the mother's educational attainment, and the economic status of the household were determined as not significant (3). But the results of the logistic and probit regression models in the present study showed that the income level was significant, while the mothers' age was not significant. This difference could be due to the different age and income composition of the mothers in the two studies.

In a study by Pourheidari et al conducted on women living in Shahrood, Iran in 2004-2005, the prevalence of unwanted pregnancies was 31%. Using logistic regression, they found out that the higher number of living children and higher maternal age were risk factors in unwanted pregnancies. In the present study, logistic and probit regression models showed that the number of living male children (as an optional independent variable) was associated with the occurrence of unwanted pregnancies, while maternal age had no significant relationship. The reason for this difference could be due to the relatively different age composition of mothers studied in the aforementioned study and the present one (4).

In a study by Amani et al conducted on 328 women living in Ardabil, Iran, the prevalence of unwanted pregnancies was 60.7%, and using logistic regression, unwanted pregnancy had a significant relationship with parity and the contraceptive method, but no significant relationship was reported with the maternal age (11), and this is in line with the results obtained from logistic and probit regression models in the present study.

In a study by Vakili et al conducted in 2010 on women residing in Yazd, Iran, the prevalence of unwanted pregnancies was about 24.5%, and based on logistic regression, the higher number of living male children affected the occurrence of unwanted pregnancies, but the variables of maternal age and mother's educational level were not found to be significant (28). Of course, based on the results of logistic and probit regression models, similar results were obtained in the present study.

In a study by Goto et al conducted on Japanese women in 2002, 46.2% of women had previously experienced unwanted pregnancies, and logistic regression showed that higher parity and the higher number of living children had significant relationships with the occurrence of unwanted pregnancies (6). This result is in line with the findings of logistic and probit regression models in the present study. However, the prevalence of unwanted pregnancies was obviously higher among Japanese women.

In a study by Calvert et al conducted on Tanzanian mothers in 2013 using logistic regression, the relationship between advanced maternal age and mother's low educational attainment with the occurrence of unwanted pregnancies was found to be significant, which of course does not correspond with the results obtained from logistic and probit regression models in the present study (8). The reason for this issue can be attributed to the completely different composition of independent variables present in the two studies.

In a study by Sadat Hashemi et al conducted on women living in Tehran, Iran in 2003, using statistical methods such as lo-

gistic regression, probit, linear discriminant analysis, and artificial neural networks, the effect of factors such as maternal age and the number of living male children on the occurrence of unwanted pregnancies were determined. In the end, artificial neural networks, logistic regression, and probit were presented as the best methods in order and discriminant analysis was reported as the worst method (27). In the present study, despite the absence of artificial neural networks as a data mining technique, a completely similar order was obtained.

In another study by Sadat Hashemi et al conducted on women residing in Tehran, Iran in 2005, the prevalence of unwanted pregnancies was determined as 31.1%, and using artificial neural networks and multinomial logistic regression and taking into account such variables as the age of the women, the number of living children, educational attainment of the woman, and the type of contraception used before pregnancy, the predictive power of the two methods were compared and the artificial neural network was recommended as the better method. The four-level nature of the dependent variable and the relatively different composition of independent variables in this study can be a probable reason for the different results (30).

In a study by Khalajabadi Farahani et al on 4141 women living in Tehran in 1996, the prevalence of unwanted pregnancies was mentioned as 31.1%, and using multinomial logistic regression method, the factors of advanced maternal age and the number of living male children were determined as effective on the occurrence of unwanted pregnancies (9). The results of the aforementioned study are in line with the results of the logistic and probit regression models in the present study about the effect of the number of living male children, but disagree with the present study in terms of the significance of maternal age. This contradiction can be due to the different four-level classification of the variable of unwanted pregnancies in that study.

In a research by Faghihzadeh et al conducted in 2003 on women living in Tehran, the prevalence of unwanted pregnancies was 38.2%, and using log-linear models, statistically significant relationships were reported between higher maternal age and pregnancy spacing and unwanted pregnancies (31). This finding nearly corresponds to the results of logistic and probit regression models in the present study about the effect of low pregnancy spacing, but it does not correspond to the results of the present study about maternal age. This lack of correspondence could be attributed to the difference in selected samples in the two studies in terms of underlying and demographic factors and also the different composition of independent variables used in modeling.

In a study by Finer and Zolna conducted in 2011 on women residing in the USA, the prevalence of unwanted pregnancies was reported as 48%, in which no statistical tests or models were applied to the data, due to the very large sample size. In this study, the prevalence of unwanted pregnancies was obviously higher among women younger than 25 years old and those with higher parities (7). The relationship between low age and the increase in the occurrence of unwanted pregnancies contradicts the results of the present study. The reason for this issue could be attributed to the cultural difference between the American mothers and the mothers studied in the present study.

Some of advantages in this study includes: using a large sample size (i.e. 887 persons) in the study and a highly-accurate prediction approach in order to model the unwanted pregnancy data. On the other hand, one of the main limitations of the present study is paying exclusive attention to the population of women residing in urban areas. Thus, it is recommended that the population of women living in rural areas be also included. Also, given the importance of statistical modeling in identification of the most important determining factors in unwanted pregnancies and given the fact that the determinants of unwanted pregnancies are different among multipa-

rous and nulliparous mothers, it is recommended that a comparative investigation of the different risk factors be conducted among the two aforementioned groups of women in a separate study.

Conclusion

Given the relatively high prevalence of unwanted pregnancies among women in Khorramabad, Iran, it is inevitable that further studies should be conducted in this field and also, the revision of the family planning programs and preconception training programs women will be necessary. The main training content may be focused on the selection of the ideal number of children and their gender based on the viewpoints of the parents, adequate and appropriate spacing between births, and the selection of appropriate contraceptive methods for all mothers (including low-income mothers). From the statistical viewpoint, it is recommended that, in a simulation study, different techniques of classification be compared based on criteria such as the type and number of independent variables, their distribution, and the sample size.

Acknowledgments

This article is part of a research project titled “The Comparison of Logistic Regression, Probit, and Linear Discriminant Analysis Models in the Prediction of Unwanted Pregnancies in Khorramabad, Iran” approved by the Lorestan University of Medical Sciences in 2012 and supported by Lorestan University of Medical Sciences.

References

1. Moosazadeh M, Nekoei-moghadam M, Emrani Z, Amiresmaili M. Prevalence of unwanted pregnancy in Iran: a systematic review and meta-analysis. Int J Health Plan. 2013.

2. Najafi F, Iran-far SH, Rezayi M, Iran-far KH. Systematic Review and Meta Analysis of Prevalence of Unwanted Pregnancy in Iran, 1995-2006. J Kermanshah Univ Med Sci 2011; 16(57):280-287.

3. Mohammadpoorasl A , Rostami F, Ivanbagha R, Torabi SH. Prevalence of Unwanted Pregnancy and Multivariate Analysis of Its Correlates in Tabriz, 2004. Med Sci J Islamic Azad Univ 2005; 15(4):201-206.

4. Pourtaheri M, Souzani A, Shamailan N. Prevalence of Unwanted Pregnancies and Their Correlate in Pregnant Woman in Shahroud, Iran. Payesh 2007; 6(1):63-70.

5. Khousheh-Mehri G, EbrahimTaheri G, Hatami Z, Safaari M. Prevalence of Unwanted Pregnancy and Its Related Factors in Women Referring to Health Centers in South of Tehran. J Nurs Midwifery Fac, Shahid Beheshti Univ Med Sci 2008; 16(59):26-32.

6. Goto A, Yasumura S, Reich MR, Fukao A .Factors associated with unintended pregnancy in Yamagata, Japan. Soc Sci Med 2002; 54(7):1065-79.

7. Lawrence BF, Mia RZ. Unintended Pregnancy in the United States: Incidence and Disparities, 2006. Contraception 2011; 84(5):478-485.

8. Calvert C, Baisley K, Doyle A, Maganja K, Changalucha J, Watson-Jones D, et al. Risk Factors for Unplanned Pregnancy Among Young Women in Tanzania. J FamPlan Reprod Health Care 2013; 39:1-2.

9. Khalaj-Abadi-Faraahani F, Sadat-Hashemi SM. Factors Influencing Unwanted Pregnancies in Tehran. Hakim 2002; 5(3):201-206.

10. Kazemi SA, Kousha A, Taddayon P, Mousavi-nasab N. The Causes of Unwanted Pregnancy in 500 Pregnant Women Which Referring to Zanjan Hospitals. J Zanjan Univ Medi Sci 2000; 37:39-45.

11. Amani F, Bashiri J, Nahan-Moghadam N, Tabraei Y. Application of Logistic Regression Model in Surveying effective Causes of Unwanted Pregnancy .Qom Univ Med Sci J 2010; 4(1):32-36.

12. Shahbazi A, Ghorbani R, Akbari-far M. A Survey on the Prevalence of Unwanted Pregnancy and Some Related Factors in Pregnant Women Who Referred to the Medical Laboratories in Semnan. Koomesh 2006; 7(3-4 (20)):133-137.

13. Mansouri A, Hosseini Sh, Dadgar S. Unexpected Pregnancy and Relative Factors in Pregnant Women Referring to Mashhad Maternity Wards in 2004. J Birjand Univ Med Sci 2009; 16(38):65-71.

14. Gipson JD, Koenig MA, Hindin MJ. The Effects of Unintended Pregnancy on Infant, Child, and Parental Health: A Review of the Literature .Stud Family Plann 2008; 39(1):18-38.

15. Enayati M, Abd-Alrahimi F. A Comparison of Mental Health and Marital Satisfaction between 'Wanted' and 'Unwanted' Pregnancy Women in Ahwaz. New Find Psychol 2008; 2(6):66-80.

16. Bayat H, Asefzadeh S. Prevalence of Unintended Pregnancies and Its Relationship with Low Birth Weight. J Birjand Univ Med Sci 2004; 11(1):28-31.

17. Dreiseitl S, Ohno-Machado L. Logistic Regression and Artificial Neural Network Classification Models: A methodology Review. J Biomed Inform 2002;35:352–359.

18. Harper PR. A Review and Comparison of Classification Algorithms for Medical Decision Making. Health Policy 2005;71:315–331.
19. Kiang MY. A comparative assessment of classification methods. Decis Support Syst 2003; 35:441-54.
20. Yassaee F. Prevalence of Unwanted Pregnancy for Women Which Referring to Mahdieh Hospital. J Med Res, Shahid Beheshti Univ Med Sci 2001; 26(2):133-136.
21. Abazari F, Arab M, Abbasszadeh A. Relationship of Unwanted Pregnancy and Fertility Behavior in Pregnant Women who Visited Maternity Wards of Kerman Hospitals. J Reprod Infert 2003;4(13):39-46.
22. Robabi H, Asan-sarani H, Azarkish F, Dastfan Z, Dashipor A. The Survey of Factors Associated with Unwanted Pregnancy among Women Referring to Health Care Centers of Iranshahr in 2007. Iranian J Obs Gyn Infert 2011;714(4):32-40.
23. Rakhshani F, Ansari Moghaddam AR, Tehrani H. Prevalence of Unwanted Pregnancy and Associated Factors in Zahedan, 1999. J Res Med Sci 2003; 8(3):40-43.
24. Khalili M, Shoohani B, Soltani O, Pour-Najaf A. A Study of Factors Effecting Unwanted Pregnancy among the Women Referring to Illam Health Centers, 2003. J Illam Univ Med Sci 2004; 12(42-43):18-24.
25. Yazdani F. Comparison of the Some Factors in Women with Unwanted and Planned Pregnancy. Family Health 2012;1(2):19-26.
26. Najafian M, Karami KB, Cheraghi M, Jafari RM. Prevalence of Some Factors Relating with Unwanted Pregnancy, in Ahwaz, Iran, 2010. Int Scholarly Res Netw 2011;4:30-34.
27. Sadat-Hashemi M, Kazemnejad A, Kavehei B. Use of Various Artifical Neural Networks for Prediction of Unwanted Pregnancy and their Comparison Using Traditional Statistical Methods. J Shahid Sadoughi Univ Med Sci Health Serv 2003; 11(1):10-15.
28. Vakili M, Shahbazi H, Dehghani MH. The Prevalence of Unintended Pregnancies and its Related Demographic Factors in Hospitals of Yazd City, 2008. Toloo-e-Behdasht 2010;9(4):23-36.
29. Noroozi A, Khoram-Roodi R, Sharifi Sh, Tahmasebi R. Prevalence of Unwanted Pregnancy and Its Related Factors in the Women Covered by Health Centers in Bushehr Province in 2003. Iranian South Med J 2005; 8(1):83-89.
30. Sadat-Hashemi SM, Kazemnejad A, Lucas C, Badie K. Predicting the type of pregnancy using artificial neural networks and multinomial logistic regression: a comparison study. Neural Comput Appl 2005; 14(3):198-202.
31. Faghihzadeh S, Babaee Rochee G, Lmyian M, Mansourian F, Rezasoltani P. Factors associated with unwanted pregnancy. J Sex Marital Ther 2003; 29(2):157-64.
32. Agresti A. Categorical Data Analysis, third Edition. Florida: John Wiley & Sons 2013, pp:119-121.
33. Bishop CM. Patern Recognition and Machine Learning. New York: Springer, 2006, pp:186-189.
34. Tu JV. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. J Clin Epidemiol 1996; 49(11):1225-31.
35. Pohar M, Blas M, Turk S. Comparison of Logistic Regression and Linear Discriminant Analysis: A Simulation Study. Metodološki zvezki 2004; 1(1):143-61.
36. Horowitz JL, Savin NE. Binary Response Models: Logits, Probits and Semiparametrics. J Econ Perspect Fall 2001; 15(4):43-56.
37. Ayatollahi SM, Poorahmad S, Vakili M, Heydari T. Probit Models & its Application in Medical Data. Andishe-ye Amari 2005; 10:36-46.
38. Cakmakyapan S, Goktas A. A Comparison of Binary Logit and Probit Models with a Simulation Study. J SocEcon Stat 2013;2(1): 1-17.