http://jhmi.sums.ac.ir
J Health Man & Info 2019, 6(2), 56–65

JHMI
Journal of Health Management and Informatics

Original Article

Investigating and Modeling the Reasons of Percutaneous Coronary Intervention Patients to Participate Rarely in Cardiac Rehabilitation: A Data Mining Approach

Seyedeh Tara Zamir[1], Mohammad Mehdi Sepehri[1,*], Hassan Aghajani[2], Morteza Khakzar Bafruei[3], Toktam Khatibi[1]

[1]Faculty of Industrial and Systems Engineering, Tarbiat Modares University, Tehran, Iran
[2]Department of Cardiology, Tehran University of Medical Sciences, Tehran, Iran
[3]Department of Industrial Engineering, Technology Development Institute (ACECR), Tehran, Iran

Abstract

Introduction: The high prevalence of cardiovascular diseases has caused many health problems in countries. Cardiac Rehabilitation Programs (CRPs) is a complementary therapy for Percutaneous Coronary Intervention (PCI) patients. However, PCI patients hardly attend CRPs. This study aims to decipher the reasons why PCI patients rarely participate in CRPs after PCI.

Methods: A cross-sectional study was used. The parameters affecting the attendance of the patients at CRPs were identified by using the previous studies and opinions of experts. A questionnaire was designed based on the identified parameters and distributed among PCI patients who were referred to Tehran Heart Center Hospital. According to data mining approach, 184 samples were collected, cleaned by replacing the corresponding mean value, and classified with three algorithms (Decision Trees, *k*-Nearest Neighbor (*k*NN), and Naïve Bayes) by using SPSS 19 and Weka 3.6.

Results: The obtained results by decision trees were superior with the average accuracy of 82%, while *k*NN and Naïve Bayes obtained 81.2% and 78%, respectively. Results of J48 classification algorithm and regression showed that lack of physician's advice was the most significant reason for non-participation of PCI patients in CRPs ($P< .001$). Other factors were family and friends' encouragement, paying expenses by insurance, awareness of the benefits of the CRPs, and comorbidity, respectively.

Conclusion: Results of the best model can help medical centers to increase the knowledge about the factors that affect participation of PCI patients in CRPs. Therefore, related centers can provide better services for patients.

Keywords: Cardiovascular Disease, Percutaneous Coronary Intervention, Cardiac Rehabilitation Programs, Data Mining, Classification, Regression

Article History:
Received: 14 October 2018
Accepted: 25 December 2018

Please cite this paper as:
Zamir ST, Sepehri MM, Aghajani H, Khakzar Bafruei M, Khatibi T. Investigating and Modeling the Reasons of Percutaneous Coronary Intervention Patients to Participate Rarely in Cardiac Rehabilitation: A Data Mining Approach. J Health Man & Info. 2019; 6(2): 56-65.

***Correspondence to:**
Mohammad Mehdi Sepehri, Faculty of Industrial and Systems Engineering, Room 906, North Engineering Building, Tarbiat Modares University, Jalal Al-e Ahmad Highway, P.O. Box: 14115-133, Tehran, Iran
Tel/Fax: +98-21-82883379
Email: mehdi.sepehri@modares.ac.ir

Introduction

Cardiovascular diseases are the issues much discussed in previous studies. The high prevalence and mortality of these diseases have caused many socioeconomic and health problems in both industrialized and developing countries (1). According to the Institute for Health Metrics and Evaluation (IHME), the Ischemic Heart Disease is the major reason for death in Iran that causes the most premature deaths in 2016 (2). Studies showed that increased level of physical activity and exercise could prevent primary and secondary Coronary Heart Diseases (CHD). These activities are considered as Cardiac Rehabilitation (CR) (3, 4). CR is a non-drug and complementary treatment that should be used along with drug therapy for patients having Coronary Artery Disease (CAD). Such programs are safe, useful and have benefits like increased physical performance; reduced re-attack mortality; delay in the incidence of Acute Myocardial Infarction (AMI); increased quality of life, social status and job opportunity; improved mood; and reduced anxiety and depression (5-8). Despite these advantages and recommendations, the referral rate of patients to the cardiac rehab centers is very low (9, 10). In particular, Percutaneous Coronary Intervention (PCI), formerly known as angioplasty with stent, is a non-surgical procedure that uses a catheter (a thin flexible tube) to place a small structure -called a stent- to open blood vessels in the heart (11). Although CR Programs (CRPs) are effective in reducing the risk factors according to the international guidelines and psychosocial problems, patients probably do not participate in these programs routinely due to the nature of PCI and short stay in hospitals (10, 12).The same problem is observed in Iran so that most cardiac patients ignore rehabilitation after PCI, Coronary Artery Bypass Graft (CABG), etc. According to our investigations and after consultations with the relevant specialists, we found that patients who had undergone CABG were more and those who had undergone PCI were less likely to attend CRPs. In this paper, we studied the reasons for the presence or absence of PCI patients at CRPs.

Different studies have introduced various factors influencing the attendance at CRPs. Banerjee et al. pointed out that the rates of participation of eligible patients in CRP are unfortunately in the range of 8.7-50%. Based on their research, women, aged and ethnic minorities, rarely referred for rehabilitation (13). A limited number of studies have focused on the identification of barriers to participate in CRPs in the South Asian population. Results showed that physician's advice was a strong factor in patients' decision making. Other factors were ease of transportation, awareness of the existence of such programs, wrong perception of harmfulness of rehab after AMI, and family and social support (13).

From another glance, identified barriers can be classified into Personal and Contextual ones. In previous studies, Personal barriers were classified as *Low knowledge about rehabilitation services* such as Usefulness of rehabilitation programs, Lack of awareness, Little encouragement, Vague explanations about rehab, Little interest; *Beliefs about heart diseases* such as Perceived lack of control over the disease and Low sense of control over future health, Perceived unpredictable, inevitable, uncontrollable condition for risk reduction of heart disease; *Negative views about services* such as Physician's disrespect towards the patient, Insufficient time for consultations, Giving inadequate comments about recovery, Lack of local available services, Unresponsive services for older people or ethnic minorities; *Self and identity* such as Imagine himself/herself different from other people who should participate in rehab, Conflict with patient's priorities, Perceived rehab programs are unnecessary, and Comorbidity; *Financial condition* like Low income; and *Demands on women* like Priorities of women's lives from their view and those around them. Contextual barriers were classified as *Long distances to Cardiac Rehab programs* such as Poor transport facilities, and Long distance from rural places; and *Lack of support from family* means Overprotect or Lack of any support (12-15).

To investigate the reasons why patients rarely participate in CRPs, some researchers have used statistical approaches (14, 16-18); but based on our knowledge, data mining approach was rarely used for extracting hidden patterns from this kind of data. However, according to a definition of data mining, it can help to analyze the available datasets to find hidden relations among the data and declare the findings in the new forms that would be understandable and useful for owners of data (19). Data mining is more flexible than statistics in terms of methods used to mine the data. Although data mining is based on mathematics, most of the data mining approaches use heuristics methods to solve the real world problems. Also, statistics uses a few records of a sample, while data mining uses data encompassing the whole sample. Moreover, statistics is mathematical-oriented and uses numeric data, but data mining deals with categorical and different kinds of data such as medical data, sounds, text, etc. Besides, in statistics, a hypothesis is made and the collected data test the hypothesis; however, data mining can work without a hypothesis and discover hidden patterns from data (20). In medicine, data mining approaches were used for managing Parkinson's disease, prediction of ischemic stroke, prediction of breast cancer survivability, prediction of kidney disease, etc (21-24).

Methods

Design

In this paper, a cross-sectional survey design was used. A questionnaire was designed based on the items extracted from previous studies, as illustrated in Table 1, and the opinions of experts. Next, the designed questionnaire was distributed among PCI patients referred to Tehran Heart Center Hospital for follow-up. Then, data mining was used to discover the effects of hidden patterns on the behavior of patients to participate in CRPs. Also, logistic regression method was used to interpret the results as the subjects assigned to the variable categories were random subsets of the original sample.

Data Collection

Knowledge Discovery and Data Mining

Fayyad et al. (1996) proposed a definition of Knowledge Discovery in Databases (KDD) as "the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in Data". According to the classical KDD methodologies, data mining is the knowledge extraction step in KDD process including the data

Table 1: Factors extracted using previous researches

Questions' Segmentation	Extracted Factors	Factors extracted from these References
Demographic Information	Gender	(12, 15, 26)
	Degree	Experts
	Age	(15, 26)
	Job	Experts
	Marital Status	Experts
	Location	(15, 28)
Patient Comments	Benefit of Program	(12, 15, 29)
	Interest	(15)
	Priority	(15, 27)
	Advice	(15, 27)
	Mental Status	(15, 27, 29)
	Service	(15)
	Environmental Condition	(15, 30)
	Transportation	(15, 28)
Financial Conditions and Comorbidities	Financial Status	(15, 27)
	Comorbidity	(18, 26)

Experts included Cardiologist, Fellowship of cardiac rehabilitation, Cardiovascular fellowship, Psychiatrist, Gynecologist, General practitioner, Internist, Physiotherapist, Insurance specialist.

selection, preprocessing, and proper interpretation of the results. Therefore, KDD process includes 5 steps including Data selection, Preprocessing, Data mining, Interpretation/Evaluation, and Using of the discovered knowledge which will be used in the next sections (25).

Gathering Data

To collect the samples and assess the willingness of patients to take part in CRPs after PCI, we chose Tehran Heart Center Hospital due to the diversity of strata going there for treatment and follow-up from August 20, 2014 to November 30, 2014. Diversity means presence of different social groups from all parts of Iran. To gather the data, we designed an appropriate questionnaire.

Questionnaire Design

The questionnaire options were designed as five-point Likert-type scale (1=disagree; 2=partly disagree; 3=no comment; 4=partly agree; 5=agree). At last, the questionnaire was designed in three sections: (1) Demographic Information, (2) Patient's Comments (involving closed questions using five-point Likert scale), and (3) Financial Conditions beside Comorbidities. Table 1 also shows the segmentation of the designed questions. In total, 44 questions were designed.

Ethical Considerations

This study was under the authorization of Tehran Heart Center Hospital. The patients were informed that participation was voluntary and their personal information would be kept anonymous.

Data Analysis

IBM SPSS Statistics for Microsoft Windows, version 19, was used for analysis of the data. To carry out data mining, we used RStudio-0.97.551 and Weka 3.6 software. We ran binary logistic regression in SPSS and j48 algorithm for decision tree in Weka. Also, we used *k*NN and Naïve Bayes classifiers in Weka.

Questionnaire Validation

To measure the validity of the questionnaire, it was first given to a limited number of PCI patients (20 people) in Tehran Heart Center PCI Clinic. They were asked if they could understand the expressions without any ambiguity. Finally, we recognized it was necessary to change some terms and sentences. After reformation, 47 questions were designed and the new questionnaire was given to PCI patients in the clinic for three months. To collect more samples from patients who had gone to rehab after PCI, we attended Tehran Heart Center Rehabilitation Unit. It is necessary to mention that because a few PCI patients go to rehab, we could collect only 26 samples from them. At last, 184 samples were collected (158 samples from patients not attending CRPs and 26 samples from those attending CRPs).

To measure the questionnaire reliability, we used Cronbach's alpha. It was calculated for each parameter using SPSS software. The alpha for the questions related to "Advice" and "Transportation" was over 0.8 that shows the proper reliability. The alpha for “Benefit of Program” and “Environmental Condition” parameters was 0.7 after deleting one question from each parameter. This α shows an acceptable reliability. The alpha for “Priority” and “Mental Status” was 0.6 that shows an almost good reliability. The amount of alpha for these parameters can be changed if more samples are collected from different centers.

Data Mining Process

For data preprocessing, we used data cleaning technique, so that the missing values were replaced by the corresponding mean (mode for categorical data) values (31). For model construction by data mining, we used classification technique to predict the patients’ status about going or not going to rehabilitation. Classification includes two phases: construction of a classification model using training dataset and evaluation of the model using a testing dataset (31, 32).

Since the real data is imbalanced, it is necessary to balance the data to achieve reliable results in the classification. In this research, we had one class label in our dataset that shows two categories (classes): (1) PCI patients who had not gone to rehab (C1=No), and (2) those who had gone (C2=Yes). The samples of C1 class were much more than those of C2 class (158 versus 26). Thus, the oversampling method was used to change the class imbalance of our dataset. This method uses some processes to replicate the minority class examples (31, 33). We balanced our data with R software.

To choose the most relevant features in the dataset and discard any other feature as irrelevant and redundant information, feature selection algorithms should be used. One model of it is filter model that was used in this research. As the filter model uses independent evaluation criteria without involvement of any classification algorithm, it will not be affected by these algorithms and has computational efficiency (34).

Classification Algorithms

Machine learning methods like Decision Trees, Bayesian Networks, and *k*-Nearest Neighbor (*k*NN) are very suitable to build a simple and interpretable classification model (35). Decision tree is like a flowchart in which each internal node denotes a test on an attribute, each branch shows an outcome of the test, and each terminal node represents a class label. Also, the topmost node is the root node. The main criterion truncate is forming leaves with the highest possible purity, which in turn forms branches that provide the greatest distinction from the upper branch. This technique represents the attribute priority with respect to the target variable, and is proper for knowledge discovery. Representation of the obtained knowledge in trees is easy to understand by humans. Decision trees can easily be converted to classification rules (31, 36). According to these explanations, the J48 was the decision tree algorithm in the Weka software used in this study. This algorithm is an implementation of the C4.5 algorithm included in the Weka software (32). Bayesian classifiers are the statistical classifiers. They can predict the probability of the class membership. Bayesian classifiers have minimum error rate when compared with other classifiers. However, it is not always true practically (31).

Learning by analogy is the basic concept in the nearest-neighbor classifiers. In other words, a given test tuple compares with training tuples that are similar to it. The training tuples are described by *n* attributes. “Closeness” is defined in terms of a distance metric, such as Euclidean distance.

Model Validation

It is possible to make several classifiers during the model construction and select the best one based on their accuracy. The Confusion Matrix is a useful tool for analyzing how well our classifier can work and predict samples of different classes (Table 2) (31, 32).

The accuracy, Precision and Recall measures of a classifier are defined as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

The Holdout is one of the methods for assessing reliable classifier accuracy estimates. In this method, the dataset is divided into two independent sets randomly: a training set and a test set. Usually, two-thirds of the data are allocated to a training set, and one-third to the test set. The training set is used to build a model. Then, the model accuracy is evaluated by the test set. By Random Subsampling, the Holdout method is repeated k times. The overall accuracy is estimated with the average accuracy of each iteration (31).

Table 2: Confusion Matrix

Predicted Class		**Yes**	**No**	**Total**
Actual Class	Yes	TP	FN	P
	No	FP	TN	N
	Total	P'	N'	P+N

TP (True Positive) and TN (True Negative) show the true classified samples of positive and negative classes, while FP (False Positive) and FN (False Negative) show the false classified samples of positive and negative classes, respectively (31,32)

Results

Demographic Results

The demographic results of the collected data are shown in Table 3.

As seen in Table 3, Free Job, Housewives, and Retired were the majorities of the participants' job. 90% of the responders were married and more than 48% of them were above 60 years old. The ratio of men to women was 70/30 and more than 35% of them had under diploma degree. 57.8% of the participants lived in Tehran, while 41.7% were residing near Tehran and other provinces.

Data Mining Results

Feature selection with the filter model led to ranking and selecting a series of features among the existing factors. 11 features were selected from 16 due to their higher rankings. The higher ranking features were Location, Benefit of Program2, Priority5, Advice3, Environmental Condition1, Service1, Transportation2, Financial Status5, Financial Status8, Comorbidity, and Go to Rehabilitation (as class label).

After feature selection, decision tree, *k*NN, and Naïve Bayes classifiers were run for model

Table 3: Details of demographic characteristics

Variable	**Category**	**Number (Total=184)**	**Percent %**
Gender	Male	130	70.7
	Female	54	29.3
Degree	Illiterate	18	9.8
	under diploma	65	35.3
	Diploma	63	34.2
	Above Diploma	13	7.1
	Bachelor	19	10.3
	Master	6	3.3
	PhD and above	0	0
Age	30-39	3	1.6
	40-49	20	10.9
	50-59	71	38.6
	Above 60	90	48.9
Job	Physician	0	0
	Engineer	8	4.3
	Teacher (school or university)	7	3.8
	Employee	8	4.3
	Free	49	26.6
	Retired	47	25.5
	Student	0	0
	Housewife	48	26.1
	Farmer	10	5.4
	Other	7	3.8
Marital Status	Single	9	4.9
	Married	166	90.2
	Divorced	9	4.9
Location	Tehran	107	57.8
	Near Tehran	19	10.3
	Other (Provinces, Cities, and Villages)	58	31.4

construction and evaluation of data. Individually, the model was trained using the training set; then, the trained classifier was tested on the test set. This process was performed 30 times on different samples generated randomly in each iteration. The averages of accuracy, precision, and recall measures were calculated for 30 models built using test set based on equations (1), (2), and (3). The average results are shown in Table 4.

For *k*NN, the higher value of *k* usually decreases the effect of noise on the classification; however, the separation between classes will be less distinct. Therefore, different values of k=1,3,5,7,9 were used. According to the mean of accuracy, the best neighbor was for k=1 (accuracy=81.2%).

By comparing the accuracy of applied algorithms on the collected samples of Tehran Heart Center Hospital, the mean of accuracy for decision tree had the best value (82%). It means that classifying the samples with decision tree was done correctly up to 82%. Therefore, it seems that decision tree classifier is the best algorithm to evaluate the data properly. The best built tree is shown in Figure 1.

The related confusion matrix of the proper tree showed that 68 samples of PCI patients that did not go to rehab were classified correctly in class (a), but one sample was classified inappropriately in class (b). Also, for C2, 68 samples were classified correctly in class C2=b=Yes, but one sample was placed in class (a) incorrectly.

The accuracy of this proper (the best) tree obtained by Weka visualization was 98.55%. According to the generated tree, the physician's advice (Advice 3) factor is located at the root. Therefore, factors contributing to the patients' participation in CRPs with priority and in order are as follows: physician's advice, family and friends' encouragement (Environmental-Condition1), paying expenses by insurance (Financial-Status5), awareness of the benefits of the CRPs (Benefit of Program2), and existence of other diseases in addition to the heart disease (Comorbidity).

Table 4: The average results for holdout method (repeated 30 times)

	Decision Tree	*k*NN					Naïve Bayes
		k=1	k=3	k=5	k=7	k=9	
Accuracy (%)	**82**	**81.2**	78.2	74	71.2	69.9	**78**
Recall (Sensitivity) (%)	82.2	81.6	78.2	74	71.2	69.9	78
Precision (%)	82.3	82	84.1	84.3	84.3	83.9	80

Bold values show the best results obtained in this research

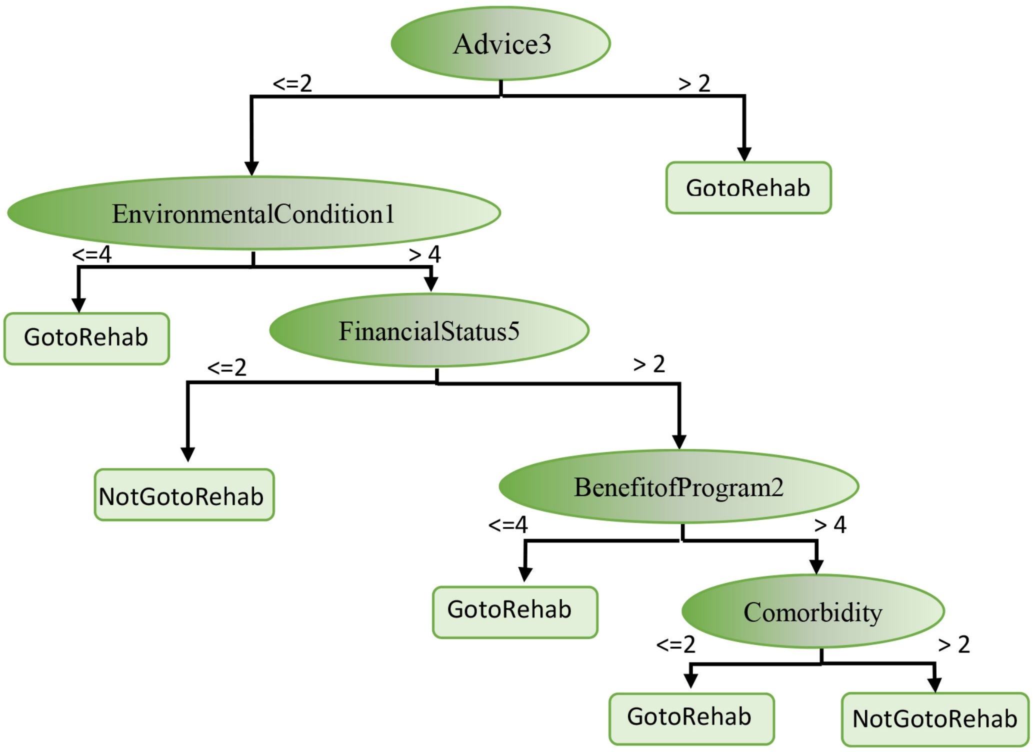

Figure 1: The proper (best) built tree

Regression Results

A logistic regression was used to model the chance of an outcome based on individual characteristics. Because chance is a ratio, what will be actually modeled is the logarithm of the chance given by: (37)

$$Log(\frac{\pi}{1-\pi}) = \beta_0 + \beta_1\chi_1 + \beta_2\chi_2 + \ldots + \beta_m\chi_m \quad (4)$$

where π indicates the probability of an event (attendance in rehab), and β_i are the regression coefficients associated with the reference group and X_i are the explanatory variables. Due to the non-linear model of our case, the parameter values were estimated by using the logarithmic likelihood ratio method. This method was used for the linearization of the likelihood function and easy mathematical operation to achieve the desired relation (38). The estimated model for predicting of "to go" or "not to go to rehab" is shown in equation 5.

For statistical analysis, we used bootstrap method as a resampling technique. It samples the training tuples uniformly with replacement. Binary logistic regression was used for all factors illustrated in the decision tree. A P<0.001 was used as the level of significance. Financial Status5 and Comorbidity were removed from the regression results. Enter method of binary logistic regression was used to access the best model of prediction. As a result, it is shown that three factors (advice3, Benefit of Program 2 and Environmental Condition 1) are significant (Table 5). Results showed that the model prediction for the percentage of samples that had gone to rehab was 82.6, while for those who had not gone was 89.2. Overall, the regression model could predict the results very close to the accuracy of the best built decision tree (82%). Therefore, by using both models (made by decision tree and logistic regression) and based on these three factors, it can be predicted that a patient would go to the rehabilitation or not. The coefficients of the regression model show that by increasing the amount of the Advice3 variable and decreasing the other two variables, the patient's probability of referral to rehabilitation can be increased. The results of the decision tree verify the outputs of the regression model and keep the two variables –Financial Status 5 and Comorbidity- in the tree.

Discussion

In this study, a questionnaire was designed based on parameters affecting on the patients' attendance at CRPs and distributed among PCI patients who were referred to Tehran Heart Center Hospital for follow-up. After performing data preprocess and feature selection, we used J48 decision tree algorithm, *k*NN, and Naïve Bayes for classification. By comparing the accuracy of the applied algorithms on the samples, the mean of the accuracy for decision tree had the best value (82%). The best tree with the accuracy of 98.55% was selected for interpretation. This tree was analyzed based on the intrinsic properties of the decision tree algorithms which put the factors from the root to the leaves, respectively. Logistic regression model was used to model the chance of an outcome based on individual characteristics of the factors. It could predict the results very close to the accuracy of the best built decision tree (82%).

By comparing the obtained results with previous studies, the situation (behavior/willingness of patients) can be explained better. In previous studies, awareness of the benefits of rehabilitation programs has an effect on participation (15, 16, 29). In this research, 109 from 184 patients agreed (59.2%) with the question about "I am aware of the benefits of rehab programs.", but they did not participate in such programs.

Studies have showed that physician's advice has the most effect on referral to CRPs (15, 27). In this study, 110 out of 158 patients (69.6%) of those who did not go

$$Log\left(\frac{\pi}{1-\pi}\right) = 98.058 + 0.841 * (Advice3) - 2.761 * (BenefitofProgram2) - 17.182 * (EnvironmentalCondition1) \quad (5)$$

Table 5: Bootstrap for Variables in the Equation

		B	Bootstrap[a]				
			Bias	Std. Error	Sig (2-tailed)	95% Confidence Interval	
						Lower	Upper
Step 1	BenefitofProgram2	-2.761	-5.842	8.218	0.016	-19.824	-1.134
	Advice3	0.841	0.150	1.118	0.000	0.478	1.683
	EnvironmentalCondition1	-17.182	0.027	1.168	0.000	-18.344	-15.392
	Constant	98.058	28.771	41.232	0.004	88.065	185.787

[a]Bootstrap results are based on 2000 bootstrap samples

to CRPs declared that the doctor did not advise them to go to rehab after PCI. Previous studies suggested that family encouragement and accompaniment has an effect on attendance (15, 27, 30). Here, we saw 109 out of 158 patients (69%) who declared that their family were there for them; however, they did not go to CRPs.

Financial status of the family was previously considered as the main factor that encouraged the patients to go to rahab (14, 15, 27). In this research, among 15 patients with salaries over $430, 7 people did not go to CRPs, but 8 patients went. Also, among 166 people who earned less than $430, 148 patients did not attend CRPs. Also, 150 out of 184 patients (81.5%) said that if insurance companies paid for rehab, they would go to CRPs. Therefore, financial condition has an important role in attendance.

In previous studies, comorbidities contributed to CRPs underutilization (14, 26). In this research, 80 patients mentioned that they would go to CRPs in spite of comorbidities, while 78 patients believed that they would not go to CRPs if they had any comorbidity. Many of the patients thought that some comorbidities like skin diseases or depression were barriers to attending CRPs. Therefore, awareness and training should be taken into account. Also, it is necessary to ask some questions like "Do you think you should not go to CRPs after PCI because of your comorbidities?" during the hospitalization process. If the answer is "yes", sufficient information about benefits of CRPs can be provided to them.

On the other hand, only 10% of cardiac surgery hospitals are doing proper and complete rehabilitation in Iran (39, 40). A solution to the problems of economic situation of families and distance can be suggested as running CRPs at home because based on studies, home-based CRPs can improve the quality of life as well (39, 40).

For the physician's advice factor as the main reason of patients going or not going to rehab, doctors should write a recommendation for eligible patients and advise them to go to rehab.

These results can be given to Tehran Heart Center Hospital and used for other related hospitals as well.

Using the findings of this study in the health information management systems of hospitals would be effective in improvement and management of the disease in related patients. In addition, in order to achieve the goals of rehabilitation, after national training, information should be given to the health management systems for the modification of their programs. Moreover, the results of this research can improve the quality of healthcare services and modify some of the standards of clinical services.

Besides, At last, It should be pointed out that if PCI surgery costs are reimbursed after completing of CRPs, the patients would attend those programs and participation rate will increase significantly. Moreover, the costs of CRPs increase the level of families' expenses. Therefore, insurance companies are recommended to cover the costs of these programs effectively.

Also, women and men may want different things from CRPs based on their needs and social positions. Previous studies revealed that CRPs must be adaptable to individual preferences, including modular formats for every patient to help him/her do his/her affairs (12). Therefore, it should be considered a way of encouraging PCI patients to attend CRPs.

It should be pointed out that some patients did not contribute to our survey. Among the difficulties of this research, it can be noted that we had to collect information directly using the questionnaire because of the lack of a comprehensive database about the behavior and views of patients and our data was real which has its own problems to handle for obtaining reliable results.

Conclusion

In conclusion, because the results showed that physician's advice was the most effective factor on patient's attendance in CRPs, there is a need to make the related physicians aware of that. Overall, results of the best built model of this study can help medical centers to increase their awareness about the factors that affect participation of PCI patients in CRPs, improve the quality of services, promote health, and prevent additional costs for patients.

Acknowledgements

We gratefully thank Dr. Mostafa Nejatian (MD), director of Cardiac Rehabilitation Unit of Tehran Heart Center for his help during data collection and giving useful information to us. Also, the authors would like to thank Dr. Alireza Atashi (PhD) from the E-Health department at Tehran University of Medical Sciences for his valuable comments to improve the quality of the manuscript.

Conflict of Interest: None declared.

References

1. Claes J, Buys R, Budts W, Smart N, Cornelissen VA. Longer-term effects of home-based exercise interventions on exercise capacity and physical activity in coronary artery disease patients: A systematic review and meta-analysis. *European*

journal of preventive cardiology*. 2017;24(3):244-56. doi: 10.1177/2047487316675823.
2. IHME [Internet]. Iran: What causes the most deaths? What causes the most premature death? c2016. Available from: http://www.healthdata.org/iran
3. Kachur S, Chongthammakun V, Lavie CJ, De Schutter A, Arena R, Milani RV, et al. Impact of cardiac rehabilitation and exercise training programs in coronary heart disease. *Prog Cardiovasc Dis*. 2017;60(1):103-14. doi: 10.1016/j.pcad.2017.07.002.
4. Anderson L, Oldridge N, Thompson DR, Zwisler AD, Rees K, Martin N, et al. Exercise-Based Cardiac Rehabilitation for Coronary Heart Disease: Cochrane Systematic Review and Meta-Analysis. *J Am Coll Cardiol*. 2016;67(1):1-12. doi: 10.1016/j.jacc.2015.10.044.
5. Dalal HM, Doherty P, Taylor RS. Cardiac rehabilitation. *BMJ*. 2015;351:h5000. doi: 10.1136/bmj.h5000.
6. Menezes AR, Lavie CJ, Milani RV, Forman DE, King M, Williams MA. Cardiac rehabilitation in the United States. *Prog Cardiovasc Dis*. 2014;56(5):522-9. doi: 10.1016/j.pcad.2013.09.018.
7. Blumenthal JA, Sherwood A, Smith PJ, Watkins L, Mabe S, Kraus WE, et al. Enhancing Cardiac Rehabilitation With Stress Management Training: A Randomized, Clinical Efficacy Trial. *Circulation*. 2016;133(14):1341-50. doi: 10.1161/CIRCULATIONAHA.115.018926.
8. Herber OR, Smith K, White M, Jones MC. 'Just not for me' - contributing factors to nonattendance/noncompletion at phase III cardiac rehabilitation in acute coronary syndrome patients: a qualitative enquiry. *J Clin Nurs*. 2017;26(21-22):3529-42. doi: 10.1111/jocn.13722.
9. Beatty AL, Bradley SM, Maynard C, McCabe JM. Referral to Cardiac Rehabilitation After Percutaneous Coronary Intervention, Coronary Artery Bypass Surgery, and Valve Surgery: Data From the Clinical Outcomes Assessment Program. *Circ Cardiovasc Qual Outcomes*. 2017;10(6). doi: 10.1161/CIRCOUTCOMES.116.003364.
10. Thomas RJ, Balady G, Banka G, Beckie TM, Chiu J, Gokak S, et al. 2018 ACC/AHA Clinical Performance and Quality Measures for Cardiac Rehabilitation: A Report of the American College of Cardiology/American Heart Association Task Force on Performance Measures. *Circ Cardiovasc Qual Outcomes*. 2018;11(4):e000037. doi: 10.1161/HCQ.0000000000000037.
11. Heart [Internet]. Percutaneous coronary intervention (PCI or angioplasty with stent). [Accessed 01 November 2017]. Available from: https://www.heartandstroke.ca/heart/treatments/surgery-and-other-procedures/percutaneous-coronary-intervention
12. Angus JE, King-Shier KM, Spaling MA, Duncan AS, Jaglal SB, Stone JA, et al. A secondary meta-synthesis of qualitative studies of gender and access to cardiac rehabilitation. *J Adv Nurs*. 2015;71(8):1758-73. doi: 10.1111/jan.12620.
13. Banerjee AT, Grace SL, Thomas SG, Faulkner G. Cultural factors facilitating cardiac rehabilitation participation among Canadian South Asians: a qualitative study. *Heart Lung*. 2010;39(6):494-503. doi: 10.1016/j.hrtlng.2009.10.021.
14. Aragam KG, Dai D, Neely ML, Bhatt DL, Roe MT, Rumsfeld JS, et al. Gaps in referral to cardiac rehabilitation of patients undergoing percutaneous coronary intervention in the United States. *J Am Coll Cardiol*. 2015;65(19):2079-88. doi: 10.1016/j.jacc.2015.02.063.
15. Clark AM, King-Shier KM, Thompson DR, Spaling MA, Duncan AS, Stone JA, et al. A qualitative systematic review of influences on attendance at cardiac rehabilitation programs after referral. *Am Heart J*. 2012;164(6):835-45 e2. doi: 10.1016/j.ahj.2012.08.020.
16. Dahhan A, Maddox WR, Krothapalli S, Farmer M, Shah A, Ford B, et al. Education of Physicians and Implementation of a Formal Referral System Can Improve Cardiac Rehabilitation Referral and Participation Rates after Percutaneous Coronary Intervention. *Heart Lung Circ*. 2015;24(8):806-16. doi: 10.1016/j.hlc.2015.02.006.
17. Gaalema DE, Cutler AY, Higgins ST, Ades PA. Smoking and cardiac rehabilitation participation: Associations with referral, attendance and adherence. *Prev Med*. 2015;80:67-74. doi: 10.1016/j.ypmed.2015.04.009.
18. Gaalema DE, Savage PD, Rengo JL, Cutler AY, Elliott RJ, Priest JS, et al. Patient Characteristics Predictive of Cardiac Rehabilitation Adherence. *J Cardiopulm Rehabil Prev*. 2017;37(2):103-10. doi: 10.1097/HCR.0000000000000225.
19. Hand DJ, Mannila H, Smyth P. Principles of data mining (adaptive computation and machine learning): MIT Press Cambridge, MA; 2001.
20. Yoo I, Alafaireet P, Marinov M, Pena-Hernandez K, Gopidi R, Chang JF, et al. Data mining in healthcare and biomedicine: a survey of the literature. *J Med Syst*. 2012;36(4):2431-48. doi: 10.1007/s10916-011-9710-5.
21. Arslan AK, Colak C, Sarihan ME. Different

medical data mining approaches based prediction of ischemic stroke. *Comput Methods Programs Biomed*. 2016;130:87-92. doi: 10.1016/j.cmpb.2016.03.022.
22. Miljkovic D, Aleksovski D, Podpečan V, Lavrač N, Malle B, Holzinger A. Machine learning and data mining methods for managing Parkinson's disease. Machine Learning for Health Informatics: Springer; 2016. p. 209-20.
23. Chaurasia V, Pal S. Data mining techniques: to predict and resolve breast cancer survivability. 2017. ; 3: 10-22.
24. Vijayarani S, Dhayanand S. Data mining classification algorithms for kidney disease prediction. *International Journal on Cybernetics & Informatics (IJCI)*. 2015;4(4):13-25.doi: 10.5121/ijci.2015.4402.
25. Ltifi H, Ben Mohamed E, ben Ayed M. Interactive visual knowledge discovery from data-based temporal decision support system. *Information Visualization*. 2016;15(1):31-50.doi: 10.1177/1473871614567794.
26. Clark AM, Sharp C, Macintyre PD. The role of age in moderating access to cardiac rehabilitation in Scotland. *Ageing & Society*. 2002;22(4):501-15. doi: 10.1017/s0144686x02008760.
27. Dunlay SM, Witt BJ, Allison TG, Hayes SN, Weston SA, Koepsell E, et al. Barriers to participation in cardiac rehabilitation. *Am Heart J*. 2009;158(5):852-9. doi: 10.1016/j.ahj.2009.08.010.
28. Leung YW, Brual J, Macpherson A, Grace SL. Geographic issues in cardiac rehabilitation utilization: a narrative review. *Health & place*. 2010;16(6):1196-205. doi: 10.1016/j.healthplace.2010.08.004.
29. Tod AM, Lacey EA, McNeill F. 'I'm still waiting...': barriers to accessing cardiac rehabilitation services. *J Adv Nurs*. 2002;40(4):421-31.doi: 10.1046/j.1365-2648.2002.02390.x.
30. Grace SL, Abbey SE, Shnek ZM, Irvine J, Franche RL, Stewart DE. Cardiac rehabilitation I: review of psychosocial factors. *Gen Hosp Psychiatry*. 2002;24(3):121-6.
31. Han J, Pei J, Kamber M. Data mining: concepts and techniques. 3rd ed. New York: Elsevier, 2011.
32. Marcano-Cedeño A, Chausa P, García A, Cáceres C, Tormos JM, Gómez EJ. Data mining applied to the cognitive rehabilitation of patients with acquired brain injury. *Expert Systems with Applications*. 2013;40(4):1054-60.doi: 10.1016/j.eswa.2012.08.034.
33. Torgo L. Data mining with R: learning with case studies: Chapman and Hall/CRC; 2011.doi: 10.1201/9780429292859.
34. Shilaskar S, Ghatol A. Feature selection for medical diagnosis: Evaluation for cardiovascular diseases. *Expert Systems with Applications*. 2013;40(10):4146-53.doi: 10.1016/j.eswa.2013.01.032.
35. Fan C-Y, Chang P-C, Lin J-J, Hsieh J. A hybrid model combining case-based reasoning and fuzzy decision tree for medical data classification. *Applied Soft Computing*. 2011;11(1):632-44.doi: 10.1016/j.asoc.2009.12.023.
36. Sayadi M, Zibaeenezhad M, Taghi Ayatollahi SM. Simple Prediction of Type 2 Diabetes Mellitus via Decision Tree Modeling. *International Cardiovascular Research Journal*. 2017;11(2). : 71-76.
37. Sperandei S. Understanding logistic regression analysis. *Biochem Med (Zagreb)*. 2014;24(1):12-8. doi: 10.11613/BM.2014.003.
38. Bard Y. Nonlinear parameter estimation. New York: Academic Press, 1974. 351p.
39. Ghalamghash R, Keyhani M, Gousheh B, BAZR AA, Barzegari M, Hoseini A. The importance of phase I cardiac rehabilitation and evaluation of its performance in Tehran city hospitals. *J Med Counc Islam Repub Iran*. 2006;24(2):123-32.
40. Saeidi M, Mostafavi S, Heidari H, Masoudi S. Effects of a comprehensive cardiac rehabilitation program on quality of life in patients with coronary artery disease. *ARYA atherosclerosis*. 2013;9(3):179-85.