

# Using the Analysis of Logistic Regression Model in Auditing and Detection of Frauds

Engin Boztepe<sup>1\*</sup>, Hayrettin Usul<sup>2</sup>

<sup>1</sup>Ardahan University, Ardahan, Turkey

<sup>2</sup>İzmir Kâtip Çelebi University, Turkey

\*Corresponding author: [enginboztepe@ardahan.edu.tr](mailto:enginboztepe@ardahan.edu.tr)

## Abstract

Fraud is defined as intentional actions in which one or more people, including from the management, employees, or the third parties, venture to obtain an unjust or illegal benefit. According to the researches, the average cost of fraud was determined as 5% of total incomes. The fraud, which has the results like a financial iceberg besides the direct losses, causes damages like loss of reputation, and adverse effects of customer relations. Auditing and detection of fraud, which has such vast effects, is of great importance.

In this study, we have developed a model that is designed for detecting mistreatments with logistic regression and the abuses in the performance-based salary system in the health sector. For this, some imaginary surgery data were added into the actual data of laparoscopic cholecystectomy operations performed in a public hospital in 2015, and to distinguish this fictitious data, the success of the generated logistic regression model was tested. Consequently, it shows that the model had 83.30% of the success rate for detecting the false data added to real data.

**Keywords:** fraud, fraud audit, logistic regression analysis, health sector frauds

**JEL Classification:** M41, M42

## Introduction

The word “Fraud” was derived from the Latin word “Fraus” in English, which was used for damage, misconduct, cheating concepts (Silwerstone & Sheetz, 2007, pp. 3-4). The word “Fraud” means cheating, swindle, corruption, and pretenses (Tureng, 2014).

Fraud is open or implicit cheating that brings illegal financial advantage for its doer, and its victims are public and private sectors or individuals (Levi et al., 2007, p. 9). According to Rezaee, the word “fraud” is a general concept used to mean

cheating and damaging another person and may be defined as purposeful action to cheat or swindle a person (Rezaee, 2010, p. 4). Fraud is the hiding of a material truth knowingly and wilfully to cause damage to others (Pedneault, 2009, p. 2).

Pehlivanlı defined fraud as capturing or using the assets of an individual or an institution unjustly, in other words, using the assets of an institution where an employee works improperly by trying to obtain unfair profit (Pehlivanlı, 2011, p. 3).

According to Bozkurt, fraud is defined as the activities done for cheating to obtain an unfair benefit in a way that will cause harm to individuals or groups. He also claimed that, as seen in the definitions as mentioned earlier, it is hard to make one single comprehensive definition of fraud, and it is a broad and extensive concept and has various meanings (Bozkurt, 2011, p. 60).

According to Simmons, when the following conditions occur, it is possible to accept that there is fraud (Simmons, 2015);

- When an individual or an institution makes a statement that is contrary to the truth about a vital reality or event;
- When the victim relies on an incorrect declaration or acts in line with this declaration;
- When the victim experiences losses in terms of money and/or property as a result of the trust in the incorrect declaration.

To make a general definition of fraud, it is possible to say that fraud is the hidden and purposeful behavior or act that has the purpose of harming an individual or a group or obtaining an unfair or illegal benefit. The aim of the person or people who perform fraud is to gain benefits from the assets that belong to others, although they do not have the right to these assets. Those who make mistakes damage someone; however, they do not have their benefits (Dumanoğlu, 2005, p. 349).

In general, the following items are included in the fraud concept:

- Fraud proceeds secretly;
- The swindler obtains benefit from the fraud action;
- The action contains a deliberate will;
- The victim of the fraud is somehow cheated;
- The victim, person, or the institutions face losses in any case (Bozkurt, 2011, pp. 60-61).

The most crucial property of fraud is the fact that a person or an institution is somehow cheated. On the very basis of cheating, there is trust in the swindler. The

trust factor is the essential element in the success of fraud. As long as trust does not exist, it is an extremely low probability that the fraud achieves success.

Since the person who is defrauded does not trust in the swindler, no one can deceive any other person. Without trust, fraud is not likely to happen (Çelik, 2007, p. 34). The probability of fraud increases in an excessive trust atmosphere. For this reason, swindlers spend too much time to obtain the trust of their target people.

### **The Cost of Fraud**

Determining the cost of fraud is an essential part of understanding the depth of the problem (ACFE, 2012, p. 8).

Although determining the cost of the fraud is very important, it is not possible to make a cost assessment due to various reasons (Pehlivanlı, 2011, p. 31).

Determining the total cost of the fraud attracts the attention of any sections of the society. While administrators want to know how significant the risk of abuse is for their companies, the professionals who fight against misconduct need the data to base their budgets and to cover their performance measurements; the media and the general public, on the other hand, are curious about how much the white-collar criminals cost us. Unfortunately, due to its nature, most of the costs of misconduct remain secret. Since secrecy constitutes the fundamental element of many misconduct types, some of them never appear to the public; and most of those which are detected are either not measured or reported. In addition to this, many misconduct types cause indirect costs like losses in productivity and reputation and depending on these, damages of customers and the cost to investigate the problems that cause the fraud and some other expenses that incur due to the efforts to resolve the issues that lead to fraud. The result is like a giant financial iceberg; although some parts of the direct losses are seen, there is a great deal of damage that we cannot see (ACFE, 2014, p. 8).

There are several reasons for not being able to determine the real costs of fraud. Even if the fraud is revealed, companies act on as if nothing matters and do not dismiss their employees to avoid losses in terms of reputation. Companies do not want to proceed with legal steps, and sometimes think that dismissing the relevant employee will prevent the recurrence of such fraud (Rezaee, 2010, p. 8).

Since there is secrecy in fraud's nature, it is not possible to determine the economic losses that are affected by the fraud. When the victims who are exposed to fraud detect these actions, it is possible to see the economic size of the losses; however, it

is not possible to determine the fraud actions that are successful and are not detected.

In the ACFE 2010 Report, 1843 fraud specialists were asked to determine the estimated cost of fraud. According to the answers given, 5% of the incomes of companies appear before us as the cost of fraud processes for the companies (ACFE, 2010, p. 8). Similarly, in the ACFE 2012 Report, 5% of the company incomes were computed as the cost of fraud actions for companies (ACFE, 2012, p. 8). When the ACFE 2012 Report is examined, it is seen that 51.4% of the 1388 files that were examined had a size of 200.000 US Dollars and below; and in addition to this, 25,3% had an economical capacity at and above 1.000.000 US Dollars.

In the ACFE 2014 Report, it was seen that the 5% rate was protected. To determine the exciting effect of this situation, when the relevant percentage is compared with the gross world product for 2013, which is 73,87 trillion dollars, it was seen that there was a global loss of 3,7 trillion dollars due to the fraud actions (ACFE, 2014, p. 8). Fraud organizations have effects in many fields like financial, operational, and psychological. However, although monetary losses are essential, fraud also has some destructive effects like the losses of reputation and the disruption in customer relations (ACL, 2010, p. 1).

### **The Additional Payment System Based on Performance Applied in The Health Sector and Fraud Proceedings to Abuse the System**

In this part, information has been provided on an additional payment system based on performance, which is applied in the health sector in Turkey, and on possible misconduct proceedings in the health sector.

#### **Performance-based Additional Payment System in Health Sector**

Different from other sectors, the health sector defines when and in which amount it will need qualities that are specific to it and is among the most basic public services. Being labor-intensive, not being substituted and delayed, unequal supply and demand balance, information asymmetry, high specialization level, not being able to stock, and being random make health system distinctive from other goods and services (Çakır & Sakaoğlu, 2014, pp. 2-3).

Health service that has the health service providers at different levels, their knowledge and the infrastructure of the health sector being at different levels, the need of the applicant in health sector, and the relation between the healthcare provider and the person who needs the health service to be established on information asymmetry and similar complex relations avoid that the health service is defined as a simple relation that is based on mutual concerns. Because of these properties of the health services, pricing and the repayment methods are among the essential topics, and several payment systems have been developed in this field (Aydın & Demir, 2007, p. 11).

It is possible to examine the payment methods used in health institutions in two main groups. One of these is the payments made to institutions that provided health services as institutions (hospitals, health centers, public health centers, medical imaging centers, dispensaries, etc.); and the other one is the payments made to people who provide health service as individuals (doctor, nurse, etc.) (Akçakanat, 2013, p. 85).

The performance-based payment system depends on the payment made to health institutions or individuals that provide healthcare services as much as the services they produce.

The performance-based additional payment system in the health sector aims to reward employees at a rate of their successes to increase activity and quality (Şişaneci, 2009, p. 21).

According to the system, the points received by doctors in return for the proceedings are collected at the end of each month, and the individual performances of doctors are measured. By considering the days on which all the doctors worked at hospitals, the performance score average for that specific period is determined for that institution by taking the arithmetic average of the performance scores. The individual performance scores of the other healthcare staff like hospital administrators and laboratory branch doctors whose own performance scores are not computed directly are computed by multiplying specific coefficients by the average of the performance of the institution. These coefficients are determined by considering the titles, duties, working conditions and durations, and the factors like working in risky departments. The net performance scores of all the staff are found by multiplying individual performance scores by self-employed coefficients, active working day coefficient, staff title coefficient, and similar parameters. The additional scores received when different duties are performed like tenders, and purchase commissions are added to the performance scores. The scores of the staff are multiplied by the monetary coefficient determined each

month to determine the additional payment amounts that will be paid according to performances (Aydın & Demir, 2007, pp. 59-68).

### **Fraud Proceedings Observed in Health Sector**

Although most of the staff providing health services act honestly and within ethical rules, various mischiefs are performed by a minor group that damages the health system. The perpetrators of the offenses observed in the health system may be classified under three main groups: health service providers, the insured ones and workplace owners, and health insurance providers (Şişaneci, 2009, pp. 20-21).

The fraud proceedings observed in the health sector may be listed as follows (Jing et al., 2008, pp. 275-276):

- Showing that a health service has been provided although it is not provided in fact (like fake surgeries);
- Separation, in other words, billing each step of one single proceeding as separate steps;
- Performing the procedures that are not necessary for medical terms to receive more money from the insurance company;
- Billing the treatment proceedings with invoices that are more than the actual service fees;
- Repetitive billing;
- Obtaining equipment from certain suppliers with higher prices;
- Using the social security and private health insurances of other people by those who do not have the right to this;
- Unnecessary hospital stays;
- The operations that are performed to increase the circulating capital, although these are not necessary.

When we consider the frauds observed in the health sector in general terms, it is possible to classify some of them as management frauds, and some of them as employee frauds. Since essential, it is possible to consider them in both groups.

### **Determining the Fraud Proceedings in Health Sector with Logistic Regression Method**

As the latest item in the present study, the issue of whether or not the “showing unperformed service as performed” that might be made in performance-based additional payment system applied in health system can be determined with the selected variables was examined with Logistic Regression Analysis; and the issue of using the information on risk factors, diagnosis and interventional processes of the diseases to determine the misconduct as a clear selection criterion was tested. When the variables were selected, it was cared for to choose the variables like the blood test results, which cannot be manipulated by swindlers and other variables that are sent directly to the database from the test laboratory.

The universe of the study consists of laparoscopic cholecystectomy operations performed in 2015 in a health service facility with the 150-bed capacity of the Turkey Public Hospitals Institution. The data were obtained with the permission of the Turkey Public Hospitals Institution.

In the application step of the study, the Logistic Regression Method was used to determine whether the “showing unperformed service as performed” fraud could be used as a fraud for the misconduct of the performance-based payment system applied in the health sector. The most important reason for using this analysis method is that the dataset that consisted of independent variables did not have a normal distribution, and the dependent variable was binary.

### **The Purpose of the Study**

The purpose of the study is to determine whether the fraud “showing unperformed service as performed” in the health sector could be predicted with the Logistic Regression Method. In this respect, the risk factors of the diseases and the real operations/unreal operations will be determined with the Logistic Regression Model that will be formed by using the diagnosis and operation data as variables ( $Y = 1$  probability). In case the model produces successful predictions, auditors will be released from the burden of unnecessary data inspection during audits and perform audits on smaller samples in shorter times. In this way, the quality of audits will increase, and the costs of audits will decrease.

### **The Limitations of the Study**

In the present study, the data of the laparoscopic cholecystectomy operations performed in 2015 in a healthcare facility with the 150-bed capacity of the Turkey Public Hospitals Institution will be used after they are checked. The data were obtained with the permission of the Turkey Public Hospitals Institution. All of the data that fit the above-mentioned criteria were included in the model.

The required tests were necessary to make the diagnosis of laparoscopic cholecystectomy and the data of 12 patients who had the risk factors but did not receive gall bladder operation were added to the model to test the model. The data of the unoperated patients were accepted as unreal operations.

One hundred nineteen patients who met the variables defined for the model and who applied with the diagnosis of gall bladder diseases and 12 unoperated patients who were in the above-mentioned healthcare facility in 2015 were included in the study. In this way, the primary study group consisted of 131 patients, 12 of whom were unoperated.

### **The Methodology of the Study**

In many socio-economic studies conducted to determine the cause and effect relation, some of the variables examined were in the form of two-level data like positive/negative, successful/unsuccessful, yes/no, and satisfied/dissatisfied. In case the dependent variable consists of two-level or multiple-level categoric data, the Logistic Regression Analysis has a vital place in examining the cause-effect relation between the dependent variables and the independent variable(s) (Agresti, 1996, p. 103).

Since the data used in studies conducted in the field of Social Sciences are generally categoric, the Logistic Regression Analysis is a method used frequently in this field. There are many study topics whose dependent variable has two categories. For example, determining which party the voters will vote for, success/failure at work, recovery/not recovery of a patient, death/living of a patient, guilty/not guilty about a person, or a tumor is benign/malign, etc.

The dependent variable is encoded as 0 and 1 in the two-category Logistic Regression Model. The primary aim of Logistic Regression Analysis is to form a regression equation that might be used in predicting to which group the units belong (Büyüköztürk, Şekercioğlu, & Çoklu, 2012, p. 57).



The Logistic Regression Analysis, which is firstly used in biomedical practice, has been used increasingly in the field of social sciences in the past two decades. It has become a popular tool in job applications in recent years. For example, credit scoring application models have been developing by using some variables like the invoice size of a person, the habit of paying the bills on time, annual income, profession, mortgage and debt liabilities (Agresti, 2002, p. 165).

Studies have been conducted frequently in recent years about the logistic model. The logistic model is used commonly especially in the banking and medicine field. The logistic regression analysis has an important place in categoric data analysis because it has several advantages in use when compared with other multi-variable analyses, and its use has become common in recent periods. Normality assumption is sought in linear regression analysis and discriminant analysis; however, there is no such condition in logistic regression analysis. It is not a condition for Logistic Regression Analysis to have homogenous variance-covariance matrices. Since these assumptions cannot be met in practice in some situations, the use of the logistic model has become more convenient. In addition to its easy use, Logistic Regression Analysis enables researchers to easily interpret numerical data, which is another reason for its being preferred more (Kocabaş, 2014, p. 8).

### **The Variables Used in the Study**

The patients who were operated were taken as the dependent variable and were encoded as 0, and the patients who were not operated were encoded with 1. It is expected the model that would be formed would detect the 12 patients who were not operated in any way. The IBM SPSS 22 Software was used in the estimation of the Logistic Regression Model.

**Age:** with advanced age, the risk of gall bladder disease increases in all ethnic groups and the risk increases as high as 4-fold over the age of 40 (Stinton & Shaffer, 2012, p. 176); (Khan, Harrison, Basset, & Bates, 2009, p. 41).

**Gender:** in all previous studies, it was shown that the gallbladder stones are observed more frequently in females than in males (Şahin, Erbilien, & Hasanoğlu, 1997, p. 76; Lee, Han, & Min, 2009, p. 456).

**Murphy Sign:** the classical physical examination finding in the initial stage of the acute cholecystitis is the “Murphy” sign. It is defined as the cut-off of inspiration due to the pain felt during deep palpation in the upper-right quadrant (Akdur, 2009,

p. 6). In acute cholecystitis, the Murphy Finding has generally a diagnostic property (Özden A., 2015, p. 283).

**USG:** ultrasonography is an important imaging method in assessing the gall bladder and relevant diseases and is the gold standard (Yıldırım et al., 2008, p. 85). The specificity and sensitivity of the USG is more than 90% in defining acute cholecystitis (Akçal, 2002, p. 143).

**Fasting Blood Sugar:** it is accepted normal if the Fasting Blood Sugar is between 74-106 in adults (Erciyes University, 2016, p. 27). Studies conducted so far have shown that there is a relation between high Fasting Blood Sugar levels and gall bladder diseases (Yıldırım et al., 2008, p. 85).

**Total Cholesterol:** it is accepted normal if the total cholesterol values are between 70-200 in adults; and high total cholesterol levels have been considered as a risk factor for gall bladder diseases (Demir et al., 2010, pp. 51-52).

**SGOT (AST):** it is a liver enzyme, and the values between 0-40 is accepted as usual in males and 0-32 in females (Erciyes University, 2016, p. 18). High levels of it have been considered as a risk factor for gall bladder diseases.

**Triglyceride:** if the Triglyceride levels are between 40-160 in males, and if the Triglyceride levels are between 40-130 in females, they are considered normal; and high triglyceride levels are considered among risk factors for gall bladder diseases (Kayaçetin, Karaaslan, & Köylü, 2003, p. 73).

**Alkaline Phosphatase (Alp):** it is considered to be normal between 40-130 in males and between 35-130 in females (Erciyes University, 2016, p. 17). It reaches very high levels in bile duct obstructions (Duman & Erden, 2004, p. 257).

**Body Mass Index:** studies conducted so far have revealed that obesity is a significant risk factor in gall bladder diseases (Çavuş & Karaca, 2013, p. 154); (Khan et al., 2009, p. 43).

**Surgery Time:** it is expected that the surgeries performed with laparoscopic methods last 30-75 minutes.

**Hospitalization Period:** after the surgery, the patient is usually discharged the other day; however, it is expected that the hospitalization period is between 1-4 days (Yıldız, Terzi, & Bardakçı, 2010, p. 74).

### Logistic Regression Analysis Application

**Table 1.** Data Distribution Table

		n	%
<b>Group</b>	Operated	119	90,8%
	Not Operated	12	9,2%
	Total	131	100,0%

The data of the main study group consisted of the data of 131 patients 12 of whom were not operated.

90,8% of the patients included in the study were operated; 9,2% were not operated.

According to the analyses results, there is a statistically significant difference between the Fasting Blood Sugar values of the Operated Patients and Not Operated Patients ( $p < 0,05$ ). According to the average values, the Fasting Blood Sugar values of the Not Operated Patients are lower than those of the Operated Patients.

According to the analyses results, there is a statistically significant difference between the Total Cholesterol values of the Operated Patients and Not Operated Patients ( $p < 0,05$ ). According to the average values, the Total Cholesterol values of the Not Operated Patients are lower than those of the Operated Patients.

**Table 2.** Measurement Values in Distinction of the Groups

GROUP		n	Ave.	sd	t	p
Age	Operated	119	52,53	14,38	-0,306	0,760
	Not Operated	12	53,83	10,07		
Fasting Blood Sugar	Operated	119	105,73	29,15	2,492	0,014*
	Not Operated	12	84,58	9,86		
Total Cholesterol	Operated	119	239,32	45,33	3,378	0,006*
	Not Operated	12	168,08	71,62		
SGOT (AST)	Operated	119	33,11	36,60	1,096	0,275
	Not Operated	12	21,35	18,42		
Triglyceride	Operated	119	191,52	49,64	1,472	0,143
	Not Operated	12	168,72	64,97		
Alkalane Phosphatase (Alp)	Operated	119	216,63	95,36	1,608	0,110
	Not Operated	12	169,72	105,98		
Body Mass Index	Operated	119	27,07	4,18	1,673	0,097
	Not Operated	12	24,92	4,89		
Surgery Time (Minutes)	Operated	119	50,47	14,50	-0,497	0,620
	Not Operated	12	52,67	15,55		
Hospitalization Period (Days)	Operated	119	3,18	1,81	1,048	0,296
	Not Operated	12	2,58	2,39		

\*p&lt;0,05

**Table 3.** Gender, Murphy Finding and USG Distribution in Group Separation

Gender, Murphy Finding and USG Distribution Table in Group Separation		GROUP			
		Operated		Not Operated	
		n	%	n	%
Gender	Male	24	20,2%	7	58,3%
	Female	95	79,8%	5	41,7%
Murphy Finding	None	23	19,3%	11	91,7%
	Yes	96	80,7%	1	8,3%
USG	None	14	11,8%	9	75,0%
	Yes	105	88,2%	3	25,0%

79,8% of the Operated Patients were female, and 41,7% of the Not Operated Patients were Female. There was Murphy Finding in 80,7% of the Operated Patients and in 8,3% of the Not Operated Patients. There was USG findings in 88,2% of the Operated Patients; and in 25,0% of the Not Operated Patients. In this context, the Murphy Finding and negative USG are more in Not Operated Patients when compared with Operated Patients.

**Table 4.** Variables and Definitions

Dependent Variable	Group (0: Operated Patient- 1: Not Operated Patient)
Independent variables	Gender (0: Male- 1: Female)
	Age
	Fasting Blood Sugar
	Total Cholesterol
	SGOT (AST)
	Triglyceride
	Alkaline Phosphatase (Alp)
	Body Mass Index
	Surgery Time (Minutes)
	Hospitalization Period (Days)
	Murphy Finding (0: No- 1: Yes)
	USG (0: No- 1: Yes)

**Table 5.** Fitness (Significance) of Fit Test of the Model

		$X^2$	sd	p
Step 1	Step	64,073	11	0,000
	Block	64,073	11	0,000
	Model	64,073	11	0,000

Since the p value was found to be significant as a result of the analysis ( $p < 0,05$ ), the Chi-Square value is also significant. In this context, all Logistic Regression coefficients aside from the constant term are different from zero; in other words, the model coefficients are significant. According to this result, the Independent Variables affect the Dependent Variable.

**Table 6.** Hosmer - Lemeshow Goodness of Fit Test

Step	$X^2$	sd	p
1	1,395	8	0,994

Hosmer - Lemeshow Goodness of Fit Test was not found to be statistically significant ( $p > 0,05$ ). In this context, it is shown that the model-data fitness is at an adequate level. Since the p value that was computed according to this test was higher than 0,05, the Goodness of Fit condition is met for the model.

**Table 10.** The Level of the Effect of the Independent Variables on Dependent Variable

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	16,159	0,387	0,845

According to Cox & Snell, the Independent Variables explain 39% of the total change in the Dependent Variable and explain 85% according to Nagelkerke.

**Table 7.** The Variables in the Model

	<b>B</b>	<b>sh</b>	<b>Wald</b>	<b>sd</b>	<b>p</b>	<b>Coefficient</b>
Gender	3,060	2,072	2,181	1	0,140	21,320
Age	0,202	0,185	1,187	1	0,276	1,224
<b>Fasting Blood Sugar</b>	<b>-0,124</b>	<b>0,063</b>	<b>3,890</b>	<b>1</b>	<b>0,049*</b>	<b>0,884</b>
Total Cholesterol	-0,038	0,020	3,792	1	0,051	0,962
SGOT (AST)	0,011	0,029	0,154	1	0,695	1,011
<b>Murphy Finding</b>	<b>4,610</b>	<b>2,356</b>	<b>3,829</b>	<b>1</b>	<b>0,050*</b>	<b>100,446</b>
Triglyceride	0,000	0,021	0,000	1	0,998	1,000
AlkalanePhosphatase (Alp)	0,012	0,012	1,070	1	0,301	1,012
<b>USG</b>	<b>5,230</b>	<b>2,627</b>	<b>3,964</b>	<b>1</b>	<b>0,046*</b>	<b>186,768</b>
Body Mass Index	-0,233	0,281	0,692	1	0,406	0,792
Surgery Time (Minutes)	-0,019	0,068	0,077	1	0,782	0,981
Hospitalization Period (Days)	0,469	0,438	1,150	1	0,284	1,599

**\*Variables Significant**

According to the results of the analysis, Fasting Blood Sugar Variable affect the status of being a fake patient. When the coefficient values are considered (0,884), it

is seen that Fasting Blood Sugar increases the being a fake patient status at a rate of 0,884-fold. Since the B Coefficient of the Fasting Blood Sugar is negative (-0,124), the probability of the person for being a fake patient increases as the Fasting Blood Sugar decreases. In this respect, Fasting Blood Sugar being low is an effective factor in terms of the person's being a fake patient.

Murphy Finding affects being a fake patient. When the coefficient value is concerned (100,446), it is seen that Murphy Finding being not present increases the probability of being a fake patient at a rate of 100,446-fold. For this reason, Murphy Finding has a very high contribution in detecting fake patients. Since the B coefficient of Murphy Finding is positive (4,610), it increases the probability of being a fake patient. In this respect, the lack of Murphy Finding is an efficient factor in terms of determining a fake patient.

USG affects the being a fake patient status. When the coefficient value is considered (186,768), it is seen that the lack of USG increases the probability of being a fake patient at a rate of 186,768-fold. For this reason, USG has a very high contribution in detecting fake patients. Since the B Coefficient of the USG is positive (5,230), it increases the probability of being a fake patient. In this respect, the lack of USG is an effective factor in terms of being a fake patient.

According to the result of the analysis, although Total Cholesterol Variable does not affect being a fake patient status at a significant level, the **p** value is very close to significance limit ( $p=0,051$ ). For this reason, this variable must not be ignored in terms of detecting fake patients. When the coefficient value of the Total Cholesterol is considered, it is seen that it increases the being a fake patient status at a rate of 0,962-fold. Since the B coefficient of Total Cholesterol is negative (-0,038), the probability of being a fake patient increases as the Total Cholesterol of the individual decreases. In this respect, low Total Cholesterol level is an effective factor in terms of being a fake patient.

The values in the B Coefficients column and the equation of the model is as follows.

$$L = \ln\left(\frac{P}{1-P}\right) = Z = -0,124 \times \text{Fasting Blood Sugar} + 4,610 \times \text{Murphy Finding} + 5,230 \times \text{USG}$$

While the risk of male patients to be fake patients is very high, the risk of female patients to be fake patients is low. In this respect, male patients must be considered in terms of being fake patients.

While the risk for patients who do not have Murphy Finding to be fake patients is very high, the risk for patients who have Murphy Finding to be fake patients is low. In this respect, the patients who do not have Murphy Finding must be considered in terms of being fake patients.

According to analysis result, the success of the model in detecting Operated Patients was 99,2%; and the success in detecting Not Operated Patients was 83,3%. The general success level of the Model is 97,7%. In this respect, the success of the model in detecting the patients who are not operated is very high.

**Table 8.** Risk Coefficients for Gender, Murphy Finding and USG

		Not Operated		All Patients		Risk Coefficients
		n	%	n	%	
<b>Gender</b>	Male	7	58,3%	31	23,7%	<b>2,46</b>
	Female	5	41,7%	100	76,3%	<b>0,55</b>
<b>Murphy Finding</b>	No	11	91,7%	34	26,0%	<b>3,53</b>
	Yes	1	8,3%	97	74,0%	<b>0,11</b>
<b>USG</b>	No	9	75,0%	23	17,6%	<b>4,26</b>
	Yes	3	25,0%	108	82,4%	<b>0,30</b>

While the risk for patients who do not receive USG to be fake patients is very high, the risk of patients who receive USG is low. In this respect, the patients who do not receive USG must be considered in terms of being fake patients.

**Table 9.** Classification Rate Tables

		GROUP		Oran
		Operated	Not Operated	
<b>GROUP</b>	Operated	118	1	99,2
	Not Operated	2	10	83,3
General Rate				97,7

### Conclusion

This study dealt with a field study on the misconducts and frauds in health sector, which is one of the most important problems of efficient source use in the social security system. Misconducts are becoming a major problem with each passing day, and affect the service quality in a direct way, and worst of all, causes unfair profit for swindlers. In this context, it is important to understand the misconducts and frauds, detect them and prevent them because they affect social security system deeply.

The fact that fraud proceedings might be detected by selecting true variables to ensure that social security system provides more quality services far from misconduct constituted the starting point of the study. For the purpose of examining the validity of this hypothesis, the Logistic Regression Analysis, which is one of the Multi-Variable Prediction Methods, was used in the study.

The scope of the study consisted of the data of laparoscopic cholecystectomy (gall bladder) operations performed in a 150-bed capacity healthcare facility of Turkey Public Hospitals Institutions in 2015. The data were obtained with the permission of Turkey Public Hospitals Institutions. All the data that met the criteria given above and that belonged to the year 2015 were included in the study. In order to test the model, the data of 12 patients who were not operated for gall bladder but who had the necessary tests for the diagnosis of laparoscopic cholecystectomy and risk factors were added to these data; and the data of these patients were accepted as not performed operations. In this way, the main universe of the study consisted of a total of 131 patients 12 of whom were not operated. The Dependent Variable of the analyses were the real operations, which were encoded as "0", and the fake operations were encoded as "1". In order to establish a Logistic Regression Model, the health literature was made use of, and the Independent variables were determined as gender, age, Fasting Blood Sugar, Total Cholesterol, SGOT (AST), triglyceride, Alkaline Phosphates (ALP), Body Mass Index, operation time, patient's hospitalization period, Murphy Finding and USG. As a result of the analysis, the Fasting Blood Sugar, USG and Murphy Finding variables were found to be significant and the Logistic Regression Model was thus established.

According to the analysis result, the Model classified 118 of the 119 Operated Patients as being operated; and 1 as Not Operated; and thus, the accurate rate of the classification of the Operated Patients was 99,2%. The established Model classified 10 of the 12 Not Operated Patients as Not Operated, and 2 as Operated; and classified Not Operated Patients accurately at a rate of 83,3%. In this context, the general success level of the Model was determined as 97,7%. In this context, the success of the Model in detecting fake patients is extremely high.

The fraud detection methods that are developed by using Logistic Regression are generally intended to detect the financial balance frauds or credit card frauds. The topic that is dealt with in the present study is intended to detect fraud proceedings in health sector.

With the established Logistic Regression Model, it was aimed to reduce the work of auditors who have to examine thousands of data to detect fraud proceedings in



the form of showing some operations as being performed although they are not performed in reality, which is a fraud that might be observed in health sector.

In the present study, only the data on gall bladder surgeries were used. However, there are different risk factors and diagnosis data for each disease or surgery. For this reason, by using the data on the disease or surgery whose auditing is performed, a specific model may be created for the disease or surgery. The Model seems beneficial because it may be used for every disease or surgery, and it decreases auditing costs, reduces the mistaken decision-making risks, and makes the decision-making process to become systematic by using statistical methods.

## References

- ACFE. (2010). "2010 Global Fraud Study". Association Of Certified Fraud Examiners.
- ACFE. (2012). "2012 Global Fraud Study". Association Of Certified Fraud Examiners.
- ACFE. (2014). "2014 Global Fraud Study". Association Of Certified Fraud Examiners .
- ACL. (2010). "Fraud Detection Using Data Analytics in The Banking Industry". ACL.
- Agresti, A. (1996). "An Introduction to Categorical Data Analysis". New Jersey: John Wiley & Sons, Inc.,.
- Agresti, A. (2002). "Categorical Data Analysis". New Jersey: John Wiley & Sons, Inc.
- Akçakanat, T. (2013). "Üniversite Hastanelerinde Çalışan Öğretim Üyelerinin Performansa Dayalı Ek Ödeme Sistemine Yönelik Tutumları Üzerine Bir Araştırma". Süleyman Demirel University, Phd Thesis
- Akçal, T. (2002). "Akut ve Kronik Kolesistit". *İstanbul University Hepatobiliary System and Pancreas Diseases Symposium Series*, 28, pp. 141-147.
- Akdur, A. (2009). "Komplike Kolesistit Olgularında Erken Laparoskopik Kolesistektomi İle Perkütan Kolesistostomi Sonrası İnterval Kolesistektomi Sonuçlarının Karşılaştırılması". Başkent University, Medical Faculty, Medical Speciality Thesis, Ankara
- Aydın, S., & Demir, M. (2007). "Sağlıkta Performans Yönetimi Performansa Dayalı Ek Ödeme Sistemi". Ankara: Republic of Turkey, Ministry of Health, Sağlıkta Dönüşüm Serisi.
- Bozkurt, N. (2011). "İşletmelerin Kara Deliği Hile: Çalışan Hileler". İstanbul: Alfa Basım Yayım.
- Büyüköztürk, Ş., Şekercioğlu, G., & Çoklu, Ö. (2012). "Sosyal Bilimler için Çok Değişkenli İstatistik: SPSS ve LISREL Uygulamaları". İstanbul: Pegem Akademi.
- Çakır, Ö., & Sakaoğlu, H. (2014, Ocak). "Sağlık Çalışanlarının Performansa Dayalı Ek Ödeme Sisteminde Ücret Adaleti Algısı: Tepecik Eğitim ve Araştırma Hastanesi Örneği". *Journal of Labour Relations*, 5(1), 1-21.
- Çavuş, B., & Karaca, Ç. (2013). "Safra Taşı Hastalığı". *Journal of Internal Medicine* (20), 151-160.
- Çelik, A. (2007). "Inventory Fraud And A Case Study ". Marmara University Social Science Institute, Msc Thesis.

- Demir, M., Doru, İ., Ataseven, H., & Albayrak, Y. (2010). "Prevalence and Risk Factors of Gallbladder Polyps in Erzurum Region". *Cumhuriyet Medical Journal*, 33(1), 48-52.
- Duman, C., & Erden, B. (2004). "Birinci Basamak Sağlık Hizmetlerine Yönelik Biyokimyasal Laboratuvar Verilerinin Kısa Yorumu". *Journal Of Continuous Medical Education*, 13(7), 256-262.
- Dumanoğlu, S. (2005). "Hata ve Hile Ayrımı: Hile Denetimi". *Marmara University, Journal of Faculty of Business Administration*, 10(1), 349.
- Erciyes Üniversitesi. (2016). "*Test Guide for Biochemistry Laboratories*". Kayseri: Erciyes University Health Practice and Research Center..
- Jing, L., Kuei-Ying, H., Jionghua, J., & Jianjun, S. (2008). "A Survey On Statistical Methods For Health Care Fraud Detection". *Health Care Management Science*, 11(3), 275-287.
- Kayaçetin, E., Karaaslan, H., & Köylü, R. (2003). "Bölgemizde Safra Kesesi Taşı Hastalığı ve Risk Faktörlerinin Araştırılması, Investigation Of Bile Duct Stone Disease And Risk Factors In Our Region". *EndoscopyJournal*, 15(3), 70-76.
- Khan, H., Harrison , M., Basset, E., & Bates, T. (2009). A 10-year follow-up of a longitudinal study of gallstone prevalence at necropsy in South East England. *Digestive Diseases and Sciences*, 59(2736), 40-62.
- Kocabaş, E. (2014). "Logistic Regression And An Application Of Banking Data". Yıldız Technic University, Social Science Institute, Msc Thesis.
- Lee, H., Han, H., & Min, S. (2009). "The association between body mass index and the severity of cholecystitis." *The American Journal of Surgery*, 4(197), 455-458.
- Levi, M., & diğ. (2007). "*The Nature, Extent and Economic Impact of Fraud in the UK, Report for the association of Chief Police Officers*". London: Economic Crime Portfolio.
- Özden, A. (2015, Aralık). "Murphy Sign". *Current Gastroenterology Journal*, 19(4), 283-285.
- Pedneault, S. (2009). "*Fraud 101*". John Wiley & Sons, Inc.
- Pehlivanlı, D. (2011). "*Hile Denetimi Metodoloji ve Raporlama*". İstanbul: Beta Yayınları.
- Rezaee, Z. (2010). "*Financial Statement Fraud: Prevention and Detection*". New York: John Wiley and Sons.
- Silverstone, H., & Sheetz, M. (2007). "*Forensic Accounting and Fraud Investigation for Non- Experts*". New Jersey: John Wiley & Sons Inc.
- Simmons, M. (2015). 10.02.2015 "Recognizing The Elements of Fraud": <http://facilitatedcontrols.com>
- Stinton, L., & Shaffer, E. (2012, Nisan). "Epidemiology of Gallbladder Disease: Cholelithiasis and Cancer". *Gut and Liver*, 6(2), 172-187.
- Şahin, M., Erbilin, M., & Hasanoğlu, A. (1997). "Safra Taşları Ve Risk Faktörleri - Bile Stones and Risk Factors". *Journal Of Turgut Ozal Medical Center*, 4(1), 72-75.
- Şişaneci, İ. (2009). "Sağlık Sisteminde Veri Madenciliği İle Suistimal Tespiti - Abuse detection with data minig in healthcare system". Kocaeli, Gebze: Gebze Higher Technology Institute Msc Thesis
- Tureng. (2014). 13.12.2014 <http://tureng.com/search/fraud>

- Yıldırım, B., & diğ. (2008). "Tokat İli Erişkinlerinde Kolelitiazis Sıklığı ve Olası Risk Faktörleri-Gallstone Disease Prevalence In The Province Of Tokat And Possible Risk Factors", *Academic Gastroenterology Journal*, 7(2), 83-86.
- Yıldız, F., Terzi, A., & Bardakçı, O. (2010). "Laparoskopik Kolesistektomi Deneyimimiz: 312 Olgunun Değerlendirilmesi-Laparoscopic Cholecystectomy, Experience with 312 Consecutive Patients". *Harran University Medical faculty Journal*, 7(3), 73-76.