

## METHODS FOR PREDICTING SPINAL ARTHRODESIS

Leandro Fabian Almeida Escobar<sup>1</sup>, Deborah Ribeiro Carvalho<sup>2</sup>, Egon Walter Wildauer<sup>1</sup>

<sup>1</sup> Universidade Federal do Paraná – UFPR

<sup>2</sup> Pontifícia Universidade Católica do Paraná – PUCPR

leandro.escobar@ufpr.br, ribeiro.carvalho@pucpr.br, egon.wildauer@ufpr.br

**Abstract:** This article presents the construction of models for predicting the need for Spinal Arthrodesis (fusion) based on Multiple Linear Regression and Decision Tree, and compares the outcomes and accuracy. The experiments involved 8,016 orthopedic patients, of which 34 underwent Spinal Arthrodesis (fusion) The Multiple Linear Regression Equation was based on Microsoft Excel and the Decision Tree on algorithm J48. The respective accuracies were compared and the understanding of the results was assessed by orthopedic specialists. Linear Regression and Decision Tree showed sensitivity of 77.8% and 60.0%, specificity of 99.4% and 99.7% and areas under the ROC curve of 73.5% and 74%, respectively. The Decision Tree, in turn, was said to be more understandable by the specialists.

**Keywords:** Data mining; predictign models; Decision trees; Linear regression; pinal Arthrodesis

### 1. INTRODUCTION

The number of chronic patients is growing due to population aging and the higher levels of stress in the workplace (DALAGASSA, 2009). Thus, health specialists must be properly informed about disease prevention strategies or tactics for cost reduction, aiming at the improvement of the care and the health of these patients, as well as the optimization of the resources used in the treatments provided (OLIVEIRA, 2005; CANCIAN, 2006).

Degenerative diseases of the spine are becoming more frequent with population aging (FELIPE e CHUEIRE, 2006), which challenges officials of the department of health who have to ensure a better quality of life for the population with increasingly scarce resources (SANTANA et al, 2006; DALLAGASSA, 2009).

As a result of the disability caused by spinal diseases, several types of treatment are imposed. Initially, the treatment involves drugs, waistcoats, blocks and manual therapies. However, as such indications fail, surgical intervention is needed. Spinal Arthrodesis is a technique that seeks to stabilize the affected region of the spine through bone fusion with or without graft materials. This technique has been used for a long time with positive results and has become standard procedure in many pathological conditions, which represented a great advance, though young patients had to use it for a very long period (FELIPE e CHUEIRE, 2006; BENATO et AL, 2009).

Lied et al (2010) investigated a set of 258 patients with degenerative disc disease and found that, before surgical treatment for spinal fusion, 48% of the individuals were unable to work for an average period of 5 months, demonstrating the severity of the condition and justifying treatment by Spinal Arthrodesis.

However, although widely recommended, Spinal Arthrodesis may have failures, especially if improperly indicated (FELIPE e CHUEIRA, 2006) and the greater the number of surgical procedures performed, the greater the risk of failure (AOTA e HIRABAYASHI, 1995). Such risks are associated with postoperative infectious complications, internal injuries, prolonged surgical time and consequent increase in direct costs of treatment (BENATO, 2009; FALAVIGNA et al 2009), morbidity of surgery and anesthetic risks (LANDIM et al, 2008).

The scarcity of available resources, on the other hand, requires efficiency in identifying trends for development of serious illnesses and the development of strategies aimed to carry out preventive efforts, by obtaining information that help identify the pertinent population and anticipate possible diseases (DALAGASSA, 2009).

One possible strategy is to provide the specialists with tools that replicate models to support the prediction of events, exploring databases and the available information systems, and that combine statistical and probabilistic capacity of machine learning and the current knowledge of health specialists. Thus, methods are designed to explore the abundance of available data in information systems in health and facilitate the analysis of recognition of trends, individual or collective problems and provide timely support to decision making in undesirable situations (TRINDADE, 2005; TARGINO, 2009), such as the need for Spinal Arthrodesis.

Despite the ease and low cost of storing large volumes of data, the benefits depend on the ability to extract useful information and obtain trends and correlations in such databases, through the use of analysis and recognition of patterns that support the elaboration of policies and decision making by specialists (FAYADD, 1996).

Exploratory data analysis, as well as methods based on machine learning may facilitate the identification of relevant information in large databases, providing predictive models for planning preventive actions or decision making regarding the target population (RAMÓN, 2007; DALAGASSA, 2009; MEYFROIDT, 2009).

The use of predictive models for support to decision making based on data related to patients' health records is being explored in various medical specialties. Ramon et al (2007) use machine learning algorithms to anticipate problems in 1,548 patients of Intensive Care Units, aiming to find patterns that support the intensive care specialist in the diagnosis and treatment to be adopted. Meyfroidt et al (2009) use machine learning algorithms to predict the overall condition, risk exposure and length of stay of patients in Intensive Care Units. Dallagassa (2009), in turn, uses Decision Trees to predict the risk of Type 2 Diabetes Mellitus in beneficiaries of a complementary health insurance plan, aimed to developing planning and prevention policies.

Lied et al (2010) explore data from 258 patients with degenerative disc disease, using statistical methods, and develop a Multiple Linear Regression function to predict the outcomes of Anterior Cervical. Jakola et al (2010) use Multiple Linear Regression and Logistics Regression in the assessment of the factors that threaten the chances of improvement of pain and functional status, predicting the postoperative conditions for patients with symptoms of lumbar stenosis and aged above 70 years, with the aim of supporting specialists and patients in the decision to be made: surgical treatment or maintenance physical therapy.

Evidently, the predictive ability of the model adopted must be carefully assessed for its real value in decision making.

Regarding the assessment of the quality of models for support to decision making in healthcare, Trindade (2005) organized criteria related to the ability to detect true cases, exclude false positives, facilitate understanding, detect problems quickly and be widely accepted.

The strategy adopted by Mani (1999), although it is used to assess the accuracy of predictive models in relation to dementia, can also be used in the prediction of Spinal Arthrodesis, in which the specificity has a greater weight because it indicates the ability to predict true negative cases, determining the performance of the model, because the decision that a surgical procedure (in this case, Spinal Arthrodesis) is not indicated for a given patient must be totally reliable.

Dalagassa (2009) expressed the need to assess the usefulness of patterns extracted according to the specialist interest, proposing that such patterns are assessed for their obvious importance because they contemplate the current knowledge; for their interest because they can be used in routine medical practice, or for their innovative character, complementing the specialist knowledge and shedding new light on the subject.

Therefore, any model proposed for predicting the indication of Spinal Arthrodesis should be assessed for its accuracy, to ensure its predictive capacity, and also for its acceptance and ease of use by specialists, demonstrating its usefulness and possibility of application in health care routine.

This study aims to elaborate two predictive models on the chance of indication of Spinal Arthrodesis, with the purpose of supporting the specialists in the elaboration of policies for prevention or planning medical treatments. Based on data related to procedures carried out in orthopedic patients, exploratory analyzes; one Multiple Linear Regression function and one Decision Tree are presented, whose usefulness and accuracy were assessed by specialists.

## 2. MATERIALS AND METHODS

### Selection and preparation of data

This exploratory study was conducted on a database of 55,814 patients that were beneficiaries of a complementary health insurance plan and who underwent the surgical procedures between 1999 and 2004.

The selected patients underwent at least one procedure out of a set of 13 procedures referred by an specialist in the medical field as related to spinal diseases (Chart 1) and, thus, to Spinal fusion surgery (arthrodesis).


Chart 1 Procedures for the spinal column (indicated by a specialist)

Procedure Code	Procedure
c25060082	FUNCTIONAL RECOVERY AFTER SURGERY OR AFTER IMMOBILIZATION ASSOCIATED WITH ORTHOPEDIC CONDITION
c32020015	X RAY OF CERVICAL SPINE :A.P.-LAT.-T.O.OR FLEXION
c32020023	X RAY OF CERVICAL DORSAL COLUMN:A.P.-LAT.-T.O.-OBLIQUE
c32020040	X RAY OF DORSAL COLUMN:A.P.-LATERAL
c32020066	X RAY OF SACRAL DORSUM COLUMN

c36010022	MRI OF CERVICAL SPINE
c36010049	MRI OF SACRAL DORSUM COLUMN
c49030086	DORSAL DISC HERNIATION : MICROSCOPY TREATMENT
c49030116	CERVICAL DISC HERNIATION: POSTERIOR SURGICAL TREATMENT
c52010392	REMOVAL OF MATERIAL AFTER SYNTHESIS
c52010465	SPINAL OSTEOMYELITIS – SURGICAL TREATMENT
c52250091	REMOVAL OF BONE GRAFT
c92010016	ORTHOSIS/PROSTHESIS

For the application of the classification modes, 4 different procedures were grouped into a single indication of Spinal Arthrodesis (Chart 2), as recommended by the involved specialist and because they constitute different techniques with the same general outcome, which is vertebral fusion. This indicator is used as dependent variable in Linear Regression and as target attribute for the algorithm of the Decision Tree.

Chart 2 Procedures grouped with attribute that indicate spinal arthrodesis.

Procedure Code	Procedure		<b>SPINAL FUSION (ARTHRODESIS)</b>
52010015	POSTERIOR SPINAL FUSION		
52010023	ANTERIOR OR POSTEROLATERAL SPINAL FUSION		
52010163	SPINAL CORD DECOMPRESSION ASSOCIATED WITH ARTHRODESIS		
52010368	OSTEOTOMY OF THE SPINE - ANTERIOR OR POSTERIOR		

Thus, we obtained data on age, gender, number of times each procedure was performed and an indicator for Spinal Arthrodesis, where the value 1 means YES and 0 (zero) means NO, as shown in chartChart 3.

Chart 3 Example of data preparation

Gender	Age Range	Age	c32020023	c52010465	c52250091	c36010022	c25060082	c32020015	c32020066	c36010049	c49030086	c52010392	c49030116	c32020040	c92010016	c52010163	Spinal Arthrodesis
M	41 to 50	43	4	0	0	3	1	4	2	0	0	0	0	0	0	1	1
F	51 to 60	59	4	0	0	1	0	0	1	1	0	0	0	0	0	0	0

## Quantitative predictive model

The predictive model was elaborated with the use of Microsoft Excel, in which the dispersion of data was analyzed to check for data linearity, indicating that changes in the values of independent variables are (or are not) directly interrelated and that the number of unexplained waste is smaller, allowing for a more accurate predictive model (HAIR, 1998; CORRAR et al, 2009).

A correlation analysis was also performed in order to assess whether there was a relationship between independent variables. If they do not remain constant and change, indicating autocorrelations, it is difficult to obtain an accurate predictive model. (HAIR, 1998; CALEGARI-JAQUES, 2003; CORRAR et al 2009).

For obtaining a statistical predictive model, Multiple Linear Regression was used, identifying the cause-and-effect relationship between quantitative variables and the construction of a mathematical model that contemplates this relationship (CALEGARI-JAQUES, 2003), thus, determining a mathematical function that can be used to predict the behavior of a dependent variable based on the values of independent variables (HAIR, 1998; CORRAR et al. 2009). The prediction sought in this study is the indication of Arthrodesis.

The linear regression learning was conducted with 5,344 records of the database (2/3 of the total). Age and number of procedures performed were defined as independent variables, and the indicator Spinal Arthrodesis (0/1), as the dependent variable, and the following equation was obtained:

$$y = m + \sum_{i=1}^n c_i x_i$$

In order to verify the accuracy of the quantitative predictive model obtained, the Linear Regression expression was imposed to 2,672 records (1/3 of the total), set aside for model tests, generating a value for dependent variable that could be rounded to 1, indicating the Spinal Arthrodesis, or to 0, indicating a trend of not performing Spinal Arthrodesis. For the validation of the Linear Regression model, the value obtained from the expression was compared to the original values observed in the database.

## Decision tree

For estimation of the model using decision tree, the WEKA's J48 Decision Tree Algorithm, an implementation of the C4.5 algorithm developed by Quinlan (1993), was used (Waikato Environment for Knowledge Analysis; <http://www.cs.waikato.ac.nz/ml/weka>). This algorithm discovers patterns in the form of a classifier represented by nodes that hierarchically relate concepts, ending in attributes that represent the classes in the databases (VIANNA et al, 2010), thus representing the occurrence (1) or non-occurrence (0) of Spinal Arthrodesis. The accuracy of the predictive model was obtained by WEKA using cross-validation based on 10 fragments of the database, where one part is set aside for the tests and the others are used for iterative learning algorithm.

The option for the WEKA application for obtaining the decision tree was based on criteria suggested by Vianna et al (2010): a) it is a friendly and easy to use tool; b) it is a free software e c) can effectively extract database classification models, with the ability to generate accuracy indicators by model tests.

Production rules of the IF-THEN type were extracted from the Decision Tree, to communicate how the machine learning method comes to a particular classification, making it clearer and validating the recognized pattern (LOPES, 2007).

### Comparison of model predictive capacity

The areas under the ROC curve of each model were compared to assess the ability of the models to classify the instances obtained as positive and negative (RAMON, 2007), and, thus, to distinguish patients with indication of Spinal Arthrodesis from those without such indication. For the purpose of enhancing the assessment of the predictive capacity of the models, specificity, ability to identify real positive cases and sensitiveness, ability to identify real negatives (MANI, 1999) were also compared.

### Understanding of the model

The usefulness of the models was assessed by the understanding indicated by the specialists interviewed. A survey instrument containing the results of the correlations for the case of the Multiple Linear Regression model and the production rules obtained with the Decision Tree were used in the interviews. The specialists indicated in a specific form (Appendix A) the ease of reading and use of the models and the degree of interest on the outcomes generated: the correlations or production rules.

The interest on rules and correlations was rated as:

- a) Of little interest because they reflect the common knowledge;
- b) Interesting because they could be used in the routine of the specialist;
- c) Surprising because they complemented the knowledge of the specialist.

Two spine surgeons and one physiotherapist specialized in postoperative recovery of spine surgery, all of them with experience of more than 15 years in the area, assessed the results and revealed their impressions through the survey instrument.

## 4. RESULTS

Based on the aggregation of the database, by age range and patient gender, it can be seen that spinal procedures are prevalent in women, with 69.9%, accounting for 5,606 occurrences (figure 1) compared to 30.1% for men, with 2,410 occurrences. The prevalent age groups are 21-60 years, with 75.1% of the cases (chart 4).

Figure 1 Distribution of spinal column procedures by gender

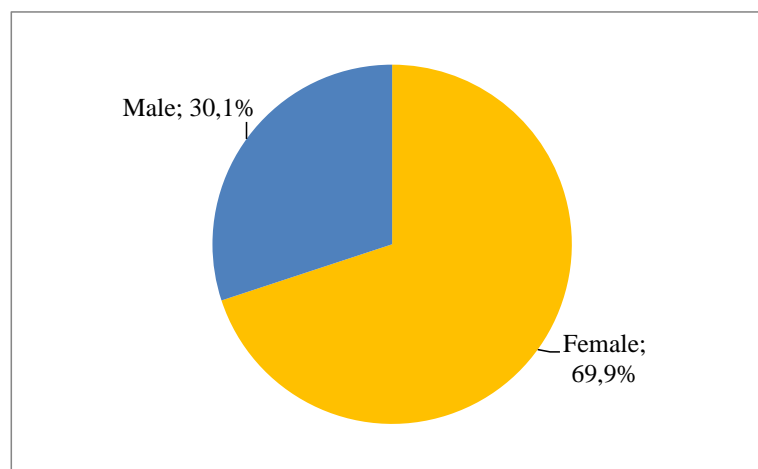




Chart 4 Distribution of procedures by age range and gender

Age Range	Female	%	Male	%	Total	%
0 to 10	78	1.40%	68	2.80%	146	1.80%
11 to 20	308	5.50%	258	10.70%	566	7.10%
21 to 30	326	5.80%	119	4.90%	445	5.60%
31 to 40	1,194	21.30%	510	21.20%	1,704	21.30%
41 to 50	1,884	33.60%	652	27.10%	2,536	31.60%
51 to 60	1,287	23.00%	492	20.40%	1,779	22.20%
61 to 70	366	6.50%	220	9.10%	586	7.30%
71 to 80	129	2.30%	76	3.20%	205	2.60%
81 to 90	34	0.60%	14	0.60%	48	0.60%
91 or above	-	0.00%	1	0.00%	1	0.00%
<b>Total</b>	<b>5.606</b>		<b>2.410</b>		<b>8.016</b>	

Dispersion analysis (Figure 2) shows that the variables have a linear relationship and, thus, changes in the values of independent variables may lead to changes in the dependent variable, i.e., the indicator of Spinal Arthrodesis, indicating the possibility of a predictive model that is sufficiently discriminant.

Figure 2 Dispersion of data

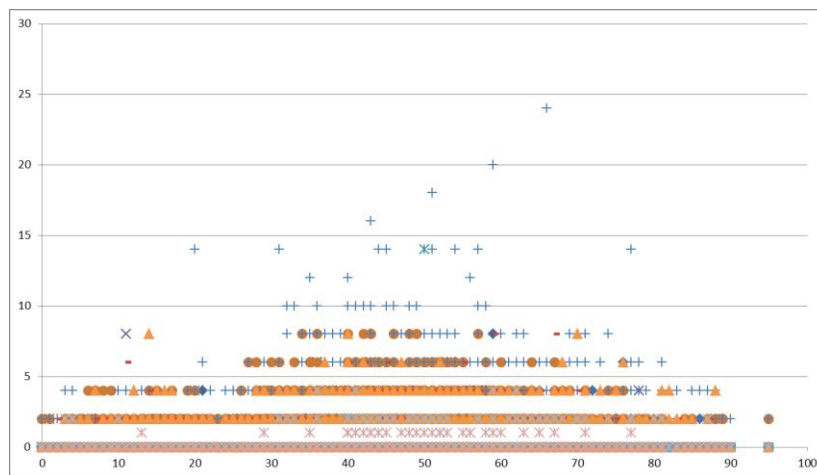


Chart 4 demonstrates the correlation between the attributes, stressing the association between the variables that impact the occurrence of Spinal Arthrodesis. Based on correlation coefficients above 0.15, Spinal Arthrodesis is more statistically correlated with the surgical treatment of Osteomyelitis (c52010465) and Herniated Cervical Disc (c49030116), MRI of the Cervical spine (c36010022) or MRI of the sacral dorsum column (c36010049) and with Bone Graft Removal (c52250091).

Chart 5 Correlation between orthopedic procedures and Spinal Arthrodesis

	Age	c32020023	c52010465	c52250091	c36010022	c25060082	c32020015	c32020066	c36010049	c49030086	c52010392	c49030116	c32020040	c92010016
Age														
c32020023	0,07													
c52010465	0,01	0,05												
c52250091	0,00	0,03	0,14											
c36010022	0,02	0,10	0,05	0,10										
c25060082	0,02	0,03	0,00	0,04	0,07									
c32020015	-0,01	-0,08	-0,01	0,04	0,10	0,00								
c32020066	0,10	-0,16	-0,01	0,14	-0,06	0,03	-0,17							
c36010049	0,07	-0,03	0,00	0,11	0,07	0,16	-0,05	0,15						
c49030086	0,03	-0,01	0,00	0,04	-0,01	0,45	-0,02	0,06	0,25					
c52010392	0,00	-0,01	0,00	0,17	0,00	0,00	-0,02	0,14	0,04	0,00				
c49030116	0,00	0,05	0,00	0,00	0,05	0,00	0,05	0,00	0,00	0,00	0,00			
c32020040	-0,03	-0,02	0,00	-0,01	-0,03	-0,01	0,01	-0,03	-0,02	-0,02	0,01	0,00		
c92010016	0,05	-0,04	0,00	0,06	0,06	0,01	-0,03	-0,03	0,00	0,00	0,00	0,00	-0,03	
Spinal Arthrodesis	0,02	0,08	0,17	0,55	0,20	0,10	0,07	0,13	0,17	0,05	0,14	0,17	0,01	0,04

The linear regression learning conducted with 5,344 records of the database (2/3 of the total). The procedures carried out are the independent variables and the indicator of Spinal Arthrodesis (1/0) is the dependent variable resulted in the equation:

$$\text{Spinal Arthrodesis} = -0.0114 + 0.0001 * \text{Age} + 0.0078 * \text{c32020023} + 0.6119 * \text{c52010465} + 0.3354 * \text{c52250091} + 0.0442 * \text{c36010022} + 0.1239 * \text{c25060082} + 0.0055 * \text{c32020015} + 0.0039 * \text{c32020066} + 0.0109 * \text{c36010049} + -0.015 * \text{c49030086} + 0.2502 * \text{c52010392} + 0.9279 * \text{c49030116} + 0.0026 * \text{c32020040} + 0.0035 * \text{c92010016} + 0.0493$$

The Decision Tree obtained from the 8,016 records with the WEKA J48 algorithm is represented in Chart 6.

Chart 6 Decision Tree obtained with J48

```

BONE GRAFT REMOVAL <= 0: N
BONE GRAFT REMOVAL > 0
| LUMBAR MRI <= 0
| | MRI OF CERVICAL SPINE <= 0: N
| | MRI OF CERVICAL SPINE > 0
| | | X-RAY OF SACRAL DORSUM COLUMN <= 0: S
| | | X-RAY OF SACRAL DORSUM COLUMN > 0: N
| MRI OF SACRAL DORSUM COLUMN > 0
| | X RAY OF CERVICAL SPINE A.P. – LAT. –TO. –OBLIQUE <= 4: S
| | X RAY OF CERVICAL SPINE A.P. – LAT. –TO. –OBLIQUE > 4: N
    
```

**Comparison of tests of the models obtained**

The model obtained with linear regression reached a sensitiveness of 77.8% and specificity of 99.4%, calculated from the metrics obtained in the tests (Chart 8).



Chart 7 Result of the tests of the multiple linear regression model

Result	Arthrodesis	
	Present (+)	Absent (-)
Positive	7	15
Negative	2	2.648

The Decision Tree obtained reached a sensitiveness of 60.0% and specificity of 99.7% (Chart 6).

Chart 8 Results of Decision Tree tests

Result	Arthrodesis	
	Present (+)	Absent (-)
Positive	12	22
Negative	8	7.975

The area under the ROC Curve (Receiver Operating Characteristic), demonstrating the total results obtained by the statistical model and by Decision Tree correspond, respectively, to 73.5% and 74.0% (chart 7).

Chart 9 Comparison of tests of the predictive models obtained

	Linear Regression	Decision Tree
Sensitiveness	77.8%	60.0%
Specificity	99.4%	99.7%
AUC (Area under the ROC Curve)	73.5%	74.0%

### Assessment of models by specialists

The responses obtained showed that the model obtained with the Decision Tree is more understandable than the linear regression model. However, none of the specialists indicated that the Decision Tree is easy to use.

Regarding the correlations obtained with the statistical model:

The correlation of Bone Graft Removal with Arthrodesis was surprising for two specialists;

The correlation of Osteomyelitis Treatment with Arthrodesis was surprising for two specialists;

The correlation of Posterior Surgery for Herniated Disc n with Arthrodesis was interesting for the three specialists.

Regarding the production rules obtained with Data Mining:

Patients who underwent one or more BONE GRAFT REMOVALS, MRI OF THE SACRAL DORSUM COLUMN AND MRI OF THE CERVICAL SPINE will undergo SPINAL ARTHRODESIS: it was found surprising by two specialists and interesting for one specialist.

When asked about which model they would adopt, two specialists said they preferred the Decision Tree.

Regarding support to decision making, one specialist does not use analysis methods, one specialist analyzes data from spreadsheets and one specialist performs data mining through a specific system developed for this purpose.

## **5. DISCUSSION**

Health databases are a complex universe because of their frequency and the absence of data. Thus, the proper selection of the predictive model is a critical task, because each method can generate different results depending on the quality and availability of data, according to Lopes (2007).

Statistical analysis of data contributes to the preliminary understanding of the database, in particular the prevalence of instances (patients) who did not undergo Spinal Arthrodesis over those who did not undergo such procedure.

Considering the frequency of patients in terms of age, it is evident that spinal diseases affect those who are economically active and, thus, the close monitoring of these individuals is recommended, especially if we consider the mean time of absence from professional activities in patients with spinal problems, identified by Lied et al(2010). In this regard, the models implemented can be used as catalysts for the decision making process regarding the adoption of Spinal Arthrodesis.

Concerning the structure of the models obtained, the mathematical equation for Multiple Linear Regression, although long, is easily implemented in spreadsheets. The Decision Tree, however, can be more easily explained because of the production rules obtained, and communication is facilitated thanks to its visual appeal.

In turn, the Decision Tree generated maintained the attributes for which correlation analysis identified the highest levels of correlation with the dependent variable (indication of Spinal Arthrodesis). Thus, although the correlation of data provides a complete view of the indexes generated, the Decision Tree generates a simpler model, using a smaller number of attributes to be analyzed for the prediction, which may accelerate the decision making process, by requiring less information for discrimination between the classes Arthrodesis=YES and Arthrodesis=NO.

The correlation matrix generated demonstrates higher correlation indexes of Spinal Arthrodesis with surgical treatment of Osteomyelitis (c52010465: 0.17) and Herniated Cervical Disc (c49030116: 0.17), MRI of Cervical Spine (c36010022: 0.20) or RI of the sacral dorsum column (c36010049: 0.17) and with Bone Graft Removal (c52250091: 0.55), indicating the possibility of predicting the occurrence of Spinal Arthrodesis for the purposes of elaboration of prevention or assistance policies, through the estimate of the volume of occurrences of Spinal Arthrodesis in relation to the profile of patients and the identified occurrences.

The Decision Tree incorporated the same attributes, except the Surgical Treatment of Herniated Cervical Disc, showing that there is a relationship between this model and the correlation analysis obtained, because of the information gain in the database.

The accuracy of the models was also similar, indicating that for the concerned set of data the level of prediction was similar, tending slightly to Multiple Linear Regression, because of its higher sensitiveness (77.8%) compared to 60.0% of the Decision Tree. Although both models are more suitable for the prediction of negative

cases– specificity, corroborating the assumptions for predictive capacity made by Mani et al (1999).

Nevertheless, the Decision Tree is more efficient. The J48 demanded the same effort in data preparation, but required a smaller number of operations for obtaining the predictive model. While Linear Regression requires care in the analysis of data accuracy for its validation, no preliminary analyzes are required for obtaining a model using Decision Tree, which has proven to be a method that, although less effective for true positive prediction, is more efficient.

The specialists interviewed reported greater ease for reading the Decision Tree compared to Linear Regression, and, thus, a greater potential to adopt the Decision Tree, according to Targino (2009). However, according to the specialists, such ease is only partial, which suggests that the representation of the results obtained with Data Mining needs a more sophisticated visual treatment to become more friendly and, thus, more understandable by the users of a system of support to decision making based on this technique.

The production rules obtained with the Decision Tree were found to be more surprising for the specialists, to the detriment of the statistical correlations which were found to be interesting, but not surprising, indicating that the model based on statistics showed in fact patterns related to the common knowledge and that Data Mining discovered more interesting patterns.

Finally, despite the different predictive capacities, the models obtained can be applied for the planning of policies of prevention and quality of life in complementary health insurance plans, since there is the possibility of predicting a surgical treatment via spinal fusion based on data related to the authorization of the procedures. Administratively, the complementary health insurance companies may plan and manage their budgets effectively, through the predictive models generated.

However, the preference for the Decision Tree, despite its lower sensitiveness, shows that Data Mining plays a key role in the support to decision making in health care because of its ability to find surprising patterns, which are concealed in the large volumes of data stored by information systems.

## **6. CONCLUSION**

The development of predictive models to indicate the possibility of carrying out Spinal Arthrodesis procedures, based on the attributes identified as relevant by the specialist is feasible. From the two models presented, tools can be developed to support decision making and planning in health care, contributing to the reduction of uncertainty as to the use of this kind of procedure in the future.

The predictive capacity of the models can be explored to support policies and strategies aimed to the rational use of resources and the improvement of the quality of life of patients, allowing monitoring the risks of treatments by Spinal Arthrodesis, based on information on the routines of these procedures and the number of times they were conducted, though without access to medical diagnoses.

However, the capacity to predict true positives of the two models requires more in-depth analysis of the two classes (Arthrodesis = Yes or Arthrodesis = No).

Finally, the Decision Tree generates a friendlier model and facilitates the extraction of production rules for use in systems of support to decision making, demonstrating cause and effect chains that have been identified as more interesting by

the specialists. However, its abstraction capacity hides, by pruning, those attributes whose influence is irrelevant for the prediction of the target attribute (or dependent variable). So, the user will be in doubt when making the decision because he/she won't notice the original set of attributes when analyzing the tree, tending to adopt the statistical model that allows visualization of all the attributes selected in the expression of the Linear Regression.

## REFERENCES

AOTA Y, Kumano K, HIRABAYASHI S. **Post fusion instability at the adjacent segments after rigid pedicle screw fixation for degenerative lumbar spinal disorders.** Journal of Spinal Disord 1995; V8 (6), 464-73.

BENATO, Marcel Luiz; ZANINELL, Ed Marcelo; SOLER I GRAELLS, Xavier; SONAGLI, Marcos Andre. **Artrodese Cervical em três e quatro níveis com dispositivo intersomático não associado a placa cervical.** COLUNA/COLIMNA, 2009. V8, 143-147.

CALEGARI-JAQUES, Sidia M. **Bioestatística: princípios e aplicações.** 2003. Artmed. Porto Alegre.

CORRAR, Luiz J.; PAULO, Edilson; DIAS FILHO, José Maria. **Análise Multivariada.** 2009. Atlas. São Paulo.

DALLAGASSA, Marcelo Rosano. **Concepção de uma metodologia para identificação de beneficiários com indicativos de Diabetes Melitus tipo 2.** Pontifícia Universidade Católica do Paraná. Curitiba, 2009. Dissertação de Mestrado. 105 f.

FALAVIGNA Asdrubal; RIGHESSE Orlando; TELES Alisson R.; KLEBER Fabrício D. **Evolução clínica e funcional dos pacientes com infecção após artrodese de coluna lombar.** OLUNA/COLUMNA. 2009. V8, 171-177.

FELIPE, Fernando M. R.; CHUEIRE, Alceu G. **Doença do Nível Adjacente após Artrodese da Coluna Lombar: Revisão da Literatura.** Arquivos da Ciência em Saúde. 2006. V13, 95-100.

HAIR, Joseph F.; TATHAN, Ronald L.; ADERSON, Rolph E.; BLACK, William. **Análise Multivariada de Dados.** 1998. Prentice Hall. São Paulo.

JAKOLA, Asgeir S; SØRLIE, Andreas; GULATI, Sasha; NYGAARD, Øystein P; LYDERSEN, Stian; SOLBERG, Tore. **Clinical outcomes and safety assessment in elderly patients undergoing decompressive laminectomy for lumbar spinal stenosis: a prospective study.** BMC Surgery. 2010. n10, p34-41

LIED, Bjarne; ROENNING, Paal Andre; SUNDSETH, Jarle; HELSETH, Eirik. **Anterior cervical discectomy with fusion in pacientes with cervical disc degeneration: a prospective outcome study in 258 patients.** BMC Surgery. 2010. n10, p10-19.

LOPES, Lucilene. **Aprendizagem de máquina baseada na combinação de classificadores em bases de dados da Área da Saúde.** Pontifícia Universidade Católica do Paraná. Curitiba, 2007. Dissertação de Mestrado. 117 f..

MANI, Subramani, SHANKLE, William R., DICK, Malcolm B.; PAZZANI, Michael J. **Two-Stage Machine Learning Model for Guideline Development.** Artificial Intelligence in Medicine Journal. 1999. V16 p51-71

MEYFROIDT Geert; GÜIZA, Fabian; RAMON, Jan; BRUYNOOGHE, Maurice. **Machine learning techniques to examine large patient databases**. Best Practice & Research Clinical Anaesthesiology, 2009. N 23 p 127–143.

QUINLAN, John Ross. **C4.5: Programs for Machine Learning**. Morgan Kaufmann Publishers. 1993

RAMON Jan, FIERENS Daan; GÜIZA Fabián; MEYFROIDT Geert; BLOCKEEL Hendrik; BRUYNOOGHE Maurice; VAN DEN BERGHE, Greet. **Mining data from intensive care patients**. Advanced Engineering Informatics, 2007. n21 p243–256.

SANTANA, Ádamo L; FRANCÊS, Carlos Renato L; COSTA, João C. W. A.; TARSO, Paulo de. **Observatório de Saúde da Amazônia – Sistema de Suporte à Decisão para Gestores de Saúde Pública da Amazônia**. Environmental and Health World Congress, 2006.

TARGINO, Maria das Graças. **Informação em saúde: potencialidades e limitações**. Informação & Informação. V14, n1, 2009

TRINDADE, Carla Machado da. **Identificação do Comportamento das Hepatites Virais a partir da Exploração de Bases de Dados de Saúde Pública**. Dissertação (Mestrado em Tecnologia em Saúde). Pontifícia Universidade Católica do Paraná. 2005

VIANNA, Rossana Cristina Xavier; MORO, Claudia Maria Cabral de Barra; MOYSÈS, Samuel Jorge; CARVALHO, Deborah Ribeiro; NIEVOLA, Julio Cesar. **Mineração de Dados e Características da Mortalidade Infantil**. Caderno Saúde Pública, 2010. Rio de Janeiro, p 532 – 542.

## APPENDIX A – INSTRUMENT FOR ASSESSMENT OF THE MODELS OBTAINED

### Objective

Investigate the acceptance of different predictive models for the need of performing Spinal Arthrodesis based on the orthopedic procedures that the patients have undergone.

Record of the number of orthopedic procedures performed in 8,016 patients of which 34 underwent Spinal Arthrodesis.

### Assessed models

Model A: Linear Regression (Statistics)

Model B: Data Mining (Artificial Intelligence)

### 1. Indicate your opinion on each of the correlations resulting from the statistical model

**SPINAL OSTEOMYELITIS – SURGICAL TREATMENT is strongly correlated with SPINAL ARTHRODESIS**

- ( ) It is of little interest because it reflects the common knowledge
- ( ) It is interesting because I can use it in my routine practice
- ( ) It is very interesting because it surprises me or complements my knowledge

**BONE GRAFT REMOVAL is strongly correlated with SPINAL ARTHRODESIS**

- ( ) It is of little interest because it reflects the common knowledge

- ( ) It is interesting because I can use it in my routine practice  
 ( ) It is very interesting because it surprises me or complements my knowledge

**MRI OF THE CERVICAL SPINE is strongly correlated with SPINAL ARTHRODESIS**

- ( ) It is of little interest because it reflects the common knowledge  
 ( ) It is interesting because I can use it in my routine practice  
 ( ) It is very interesting because it surprises me or complements my knowledge

**MRI OF THE SACRAL DORSUM COLUMN SPINE is strongly correlated with SPINAL ARTHRODESIS**

- ( ) It is of little interest because it reflects the common knowledge  
 ( ) It is interesting because I can use it in my routine practice  
 ( ) It is very interesting because it surprises me or complements my knowledge

**CERVICAL DISC HERNIATION: POSTERIOR SURGICAL TREATMENT is strongly correlated with SPINAL ARTHRODESIS**

- ( ) It is of little interest because it reflects the common knowledge  
 ( ) It is interesting because I can use it in my routine practice  
 ( ) It is very interesting because it surprises me or complements my knowledge

**1.1. Equation obtained from the data:**

$$\begin{aligned}
 \text{Spinal Arthrodesis} &= 0.0114 + 0.0001 \times \text{Age} \\
 &+ 0,0078 \times \text{X RAY OF CERVICAL SPINE:A.P.-LAT.-T.O.-OBLIQUE} \\
 &+ 0,6119 \times \text{SPINAL OSTEOMYELITIS -SURGICAL TREATMENT} \\
 &+ 0,3354 \times \text{BONE GRAFT REMOVAL} \\
 &+ 0,0442 \times \text{MRI OF THE CERVICAL SPINE} \\
 &+ 0,1239 \times \text{POST OPERATIVE FUNCTIONAL RECOVERY} \\
 &+ 0,0055 \times \text{X RAY OF CERVICAL SPINE:A.P.-LAT.-T.O.OR FLEXION} \\
 &+ 0,0039 \times \text{X-RAY OF SACRAL DORSUM COLUMN} \\
 &+ 0,0109 \times \text{MRI OF THE CERVICAL SPINE} \\
 &+ -0,015 \times \text{DORSAL DISC HERNIATION: MICROSCOPY TREATMENT} \\
 &+ 0,2502 \times \text{REMOVAL OF MATERIAL AFTER SYNTHESIS} \\
 &+ 0,9279 \times \text{CERVICAL DISC HERNIATION: POSTERIOR SURGICAL TREATMENT} \\
 &+ 0,0026 \times \text{X RAY OF THE DORSAL COLUMN :A.P.-LATERAL} \\
 &+ 0,0035 \times \text{ORTHOSIS/PROSTHESIS} \\
 &+ 0,0493
 \end{aligned}$$

**The model (formula and correlations) generated by Linear Regression is easy to read and understand.**

- ( ) I totally disagree  
 ( ) I partially disagree  
 ( ) I cannot express an opinion  
 ( ) I partially agree  
 ( ) I totally agree



**2. Indicate your opinion for each of the rules obtained by Data Mining**

**Rule 1:** Patients who did not undergo BONE GRAFT REMOVALS WILL NOT undergo SPINAL ARTHRODESIS

- It is of little interest because it reflects the common knowledge
- It is interesting because I can use it in my routine practice
- It is very interesting because it surprises me or complements my knowledge

**Rule 2:** Patients who underwent one or more BONE GRAFT REMOVAL, MRI OF THE SACRAL DORSUM COLUMN AND MRI OF THE CERVICAL SPINE will undergo SPINAL ARTHRODESIS

- It is of little interest because it reflects the common knowledge
- It is interesting because I can use it in my routine practice
- It is very interesting because it surprises me or complements my knowledge

**Rule 3:** Patients who underwent one or more BONE GRAFT REMOVAL and more than one MRI OF THE CERVICAL SPINE and did not undergo MRI of the SACRAL DORSUM COLUMN nor X RAY OF SACRAL DORSUM COLUMN, will undergo SPINAL ARTHRODESIS

- It is of little interest because it reflects the common knowledge
- It is interesting because I can use it in my routine practice
- It is very interesting because it surprises me or complements my knowledge

**Rule 4:** Patients who underwent one or more BONE GRAFT REMOVAL and MRI of the CERVICAL SPINE, did not undergo MRI OF THE SACRAL DORSUM COLUMN and did not undergo X RAY OF THE SACRAL DORSUM COLUMN will **NOT** undergo SPINAL ARTHRODESIS

- It is of little interest because it reflects the common knowledge
- It is interesting because I can use it in my routine practice
- It is very interesting because it surprises me or complements my knowledge

**Rule 5:** Patients who underwent one or more BONE GRAFT REMOVAL and MRI OF THE SACRAL DORSUM COLUMN and up to 4 X RAY OF THE CERVICAL SPINE A.P - LAT. -TO. -OBLIQUE will undergo SPINAL ARTHRODESIS

- It is of little interest because it reflects the common knowledge
- It is interesting because I can use it in my routine practice
- It is very interesting because it surprises me or complements my knowledge

**Rule 6:** Patients who underwent one or more BONE GRAFT REMOVAL, one or more MRI OF THE SACRAL DORSUM COLUMN and more than 4 X RAY OF THE CERVICAL SPINE A.P - LAT. -TO. -OBLIQUE will undergo SPINAL ARTHRODESIS

- It is of little interest because it reflects the common knowledge
- It is interesting because I can use it in my routine practice
- It is very interesting because it surprises me or complements my knowledge

The model of rules generated by Data Mining is easy to read and understand.

- I totally disagree
- I partially disagree
- I cannot express an opinion
- I partially agree
- I totally agree

**For my needs and considering my work routine, I would use the model**

- Linear Regression
- Decision Tree
- I would not use the models presented

**Indicate the option that best reflects your activity in the health area**

- Expert or similar in a Complementary Health Insurance Company
  - Medical specialist
  - Nurse
  - Intensive care physician
  - Intensive care nurse
  - Manager in a Complementary Health Insurance Company
  - Other. Specify
- 

**How long have you been active in the health area?** \_\_\_\_\_

**Do you use any system to record or consult information about patients?**

- Electronic Medical Record System implemented in the company / hospital.
- Spreadsheets (Excel or similar).
- Electronic documents (Excel or similar).
- Printed chart.
- I don't use any.

**On decision making in your work routine**

- I use statistical analyzes automated by a specific system (ready to use).
- I use data mining results automated by a specific system (ready to use).
- I perform my own statistical analyzes from spreadsheets or similar
- I carry out Data Mining experiments from specific systems (Weka, Apriori, C4.5 or similar).
- I don't use data analyzes for decision making.

**If interested, please inform an e-mail to which the results of this research can be sent as soon as they are systematized:**

Name: \_\_\_\_\_

E-mail : \_\_\_\_\_