Analysis of the Quality of Life After an Endoscopic Thoracic Sympathectomy: A Business Intelligence Approach

Diana Gonçalves, Maribel Yasmina Santos
Algoritmi Research Centre
University of Minho
Guimarães, Portugal
pg13247@uminho.pt, maribel@dsi.uminho.pt

Jorge Cruz
Medicine Faculty
University of Lisbon
Lisboa, Portugal
costacruzjorge@gmail.com

Abstract— **Primary hyperhidrosis, a disorder characterized by an excessive sweating, has been treated by endoscopic thoracic sympathectomy. As a consequence of the surgery, patients improved their overall quality of life. Their day-by-day activities are not affected, or are less affected, by this disorder, and their emotional state verifies a significant improvement, from a situation of shame and self-punishing to what we could say a normal life. This paper presents the analysis of the quality of life of 227 patients that were treated by an endoscopic thoracic sympathectomy. The study was based on the use of business intelligence technologies, which allowed the storage, the analysis and the reporting of all the relevant findings. In technological terms, this paper illustrates the database and data analysis developments needed in a specific healthcare application domain. For data storage, a data mart was designed addressing the relevant attributes. For data analysis, on-line analytical processing and data mining technologies were used to show the evolution of the patients' health condition and the incidence of complications or side effects as consequence of the surgery.**

Keywords-primary hyperhidrosis, endoscopic thoracic sympathectomy, business intelligence, on-line analytical processing, data mining.

I. Introduction

Primary hyperhidrosis, a disorder of the sympathetic autonomous nervous system, affects around 1% of the global population [1]. It is characterized by an excessive sweating of the face, palms, armpits and foots, either in all of these locals or in some of them. This excessive sweating provokes several problems to the individual, who see his/her life completely influenced by this disorder.

Endoscopic upper-thoracic sympathectomy has been considered the treatment of choice for primary palmar hyperhidrosis [2]. This surgery is a minimal invasive procedure of thoracic sympathetic blockage and consists of the bilateral ablation of the second and third thoracic sympathetic ganglia, affecting the sympathetic nervous outflow to the arms and elsewhere [3-4].

Being a definitive treatment for this disorder, the results of the surgery have revealed a high degree of patient satisfaction [2]. To show the increase of the quality of life and the incidence of complications and side effects after the surgery, this paper shows the analysis of data collected from 227 patients. Although several studies have been conducted in order to evaluate the improvement in the patients' health condition [1-6], the analysis presented in this paper exploits a new perspective in data analysis, using business intelligence concepts and applying business intelligence techniques to store, analyze and report the several findings.

A business intelligence system was specifically developed for this study. It integrates data collection, data analysis and data reporting mechanisms. The data collection was carried out using a simplified version of the SF-36 Health Survey Questionnaire [7], the SF-6D Questionnaire. In this study, this questionnaire was extended, integrating several questions related with the beginning of the primary hyperhidrosis, the complications associated with the surgery and the side effects verified by the patients.

The collected data were analyzed with on-line analytical processing and data mining techniques. The results achieved are twofold: i) they show an increase of the overall health condition and quality of life of the patients; and ii) they demonstrate the applicability of the used data analysis concepts and techniques.

This paper is organized as follows. Section 2 gives an overall overview of the context in which this work was carried out. Section 3 presents the business intelligence system developed to analyze the quality of life after an upper-thoracic sympathectomy. Section 4 presents the results achieved using the implemented business intelligence system. Section 5 concludes with some remarks about the work undertaken and also some guidelines for future work.

II. Patients and Related Data

Data from 227 patients were collected and analyzed in this study. An adapted version of the SF-6D questionnaire was made available through the Internet to 504 patients that were submitted to an upper-thoracic sympathectomy in two Portuguese health institutions: the *Hospital Santa Maria* in Lisbon and the *Instituto Cardiovascular de Lisboa*, also in Lisbon. A letter containing a description with the objectives of this work, a set of instructions, a login and a password (for an on-line answering) was sent to each patient. From this population of 504 patients, an answering rate of 48% was obtained (30 questionnaires did not reach their destination due to wrong addresses or changes of addresses).

The SF-6D questionnaire is based on the short form health survey questionnaire (SF-36), a promising instrument for measuring health perception in a general population [7].

The SF-36 instrument integrates eight health concepts: physical functioning, bodily pain, role limitations due to physical health problems, role limitations due to personal or emotional problems, emotional well-being, social functioning, energy/fatigue, and general health perceptions. The SF-36 questionnaire was already used in several studies for evaluating the quality of life in individuals submitted to an endoscopic thoracic sympathectomy for primary hyperhidrosis treatment [5-6].

A validated Portuguese version of the SF-6D was used [8]. This version was upgraded to include specific information about the disorder in study and the complications and side effects verified as consequence of the surgery. Each patient answered to two questionnaires. One associated to his/her health condition prior the surgery and the other related with his/her health condition in a post-surgery scenario. Personal information included in the questionnaires, like name, address, birth date and gender of the patients, was of optional answering. Information related to the surgeon's name, surgery date and patient's age at surgery was also of optional answering. Although optional, the majority of the patients provided this information allowing the characterization of the population in study. Only personal questions were answered in natural language. All other answers were selected from a fixed and pre-formatted set of options.

III. BUSINESS INTELLIGENCE SYSTEM

A Business Intelligence (BI) system integrates operational data with analytical tools in order to support the decision-making process. This system has as main objective to make available high quality information [9]. The designation BI is relatively new and emerged to replace terms such as Decision Support System used from more than thirty years [10]. Although the label *Business* could give the impression of some strong link to the commercial or business area of an organization, BI systems are of generic use and very useful when the collection, storage and analysis of data are needed to provide information to support the decision-making process (either in individual or corporate level).

In a BI system, several technologies need to be integrated in order to support two fundamental components of these systems: data storage and data analysis. For data storage, Data Warehouse Systems (DWS) allow the development of decision support data models [11]. DWS are usually implemented to store large amounts of data. Although it is not the case of this study, it is proposed a decision support data model for this application domain. In this study, the available data are stored in a data mart, which is based on the proposed model, and analyzed with On-Line Analytical Processing (OLAP) and Data Mining (DM) technologies [12]. OLAP technology supports the analysis of data over the different perspectives considered in the decision support data model. DM technology integrates statistic and artificial intelligence concepts to identify useful models, patterns or trends on data. For the analysis of the quality of life after an endoscopic thoracic sympathectomy, a BI system was designed and implemented. In terms of design, Figure 1 presents the system architecture.
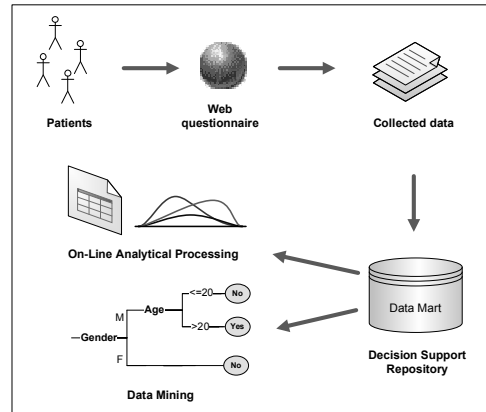

Figure 1. BI system architecture

All the contacted patients had available an on-line platform (through the Web) to answer the questionnaires (the prior and the post-surgery). The collected data were verified in order to see if any errors were present. As answers were limited to a fixed set of valid options, no errors were found in data. A few cases of missing data fields were detected, mainly on personal information. In these cases, all fields were marked with the "Unknown" or "Not know" labels. For some missing data fields on continuous attributes, the average value was calculated and assigned to the missing values [12]. After the data verification process, the collected data were loaded into the data mart in order to support the analysis process. At this phase, only one source dataset was used, the associated with the 227 patients, although the inclusion of data from other sources is envisaged. The data mart model was defined with the decision process in mind: i) know the characteristics of the patients, ii) verify the occurred side effects and complications, and iii) see the evolution of the patients' quality of life. The data model integrates six main vectors of analysis, each one of them represented by one of the fact tables present in the model. The constellation schema is shown in Figure 2, where the fact tables are: Hyperhidrosis, Hyperhidrosis Drawbacks, Compensatory Hyperhidrosis, Emotional State, Side Effects, and Quality of Life. In the BI system, the data mart that constitutes the proof-of-concept to the proposed model was implemented in MS ACCESS®. This choice allowed a simple exchange of data among the project team members during the prototype implementation.

The Hyperhidrosis fact table allows the storage of all the information associated with the hyperhidrosis incidence verified in the patient before and after the surgery. The Hyperhidrosis Drawbacks is used to characterize the beginning of the disorder in the patient and when he/she started to feel uncomfortable with it. The Compensatory Hyperhidrosis table stores the incidence of compensatory hyperhidrosis in the patients both one month and six months after the surgery. The Emotional State table characterizes the emotional feelings of the patients, namely how they are affected by this disease. The Side Effects table stores the temporary and/or permanent side effects verified in the patients as consequence of the surgery. The Quality of Life table receives the answers to the

questions that are related with the SF-6D questionnaire and the quality of life measure calculated from the given answers. The quality of life measures were obtained using the SF-6D utility made available by Professor John Brazier [7].
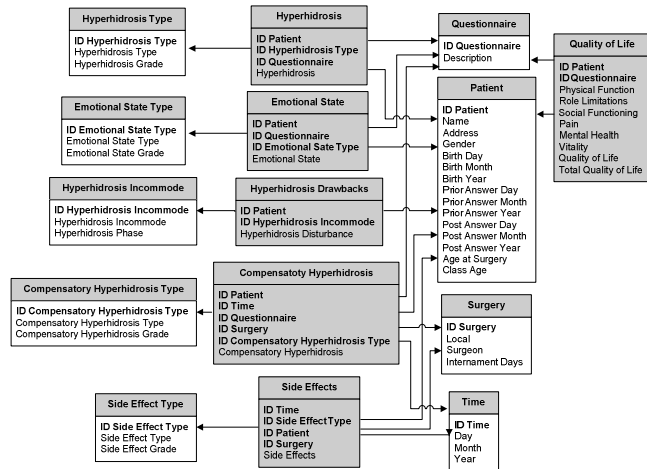

Figure 2. Data mart model

Each one of these six fact tables is linked to a set of dimension tables allowing the analysis of the available data in different perspectives. In the defined data model (Figure 2) it is possible to verify the links among the fact tables and the dimension tables and so, the analyses allowed in each case. If we look at the Hyperhidrosis fact table, we can see that this table is linked to the Hyperhidrosis Type, Patient and Questionnaire dimension tables, meaning that the Hyperhidrosis answers can be analyzed by the locals where the hyperhidrosis is verified (face, hands, armpits or feet), the grade of incidence associated to each one of these locals (inexistent, irrelevant, moderate or excessive), the associated patients (with information such as age, gender, age at surgery, among other attributes) and the questionnaire, which allows the analysis of the answers before or after the surgery, or the evolution from one stage to another.

IV. DATA ANALYSIS WITH OLAP AND DATA MINING

After the presentation of the proposed BI system and the data model that stores the data collected in this study, this section presents the results obtained from the analysis of the available data using OLAP and DM technologies. OLAP technology is used to analyze the several fact tables along different perspectives. DM technology is used to find trends or patterns in data. For OLAP, the pivot tables and pivot charts facilities of MS Excel® were used.

In terms of OLAP analyses, the first one verifies when occurred the beginning of the disorder and when the patients started to feel uncomfortable with it. Figure 3 shows a pivot chart with the obtained results, differentiating the patients through gender. For female individuals, the beginning of the disorder is verified mostly in the childhood. For male, it is mostly verified in the adolescence. With respect to the drawbacks associated with the disease, it is in the adolescence that both female and male start to feel uncomfortable.
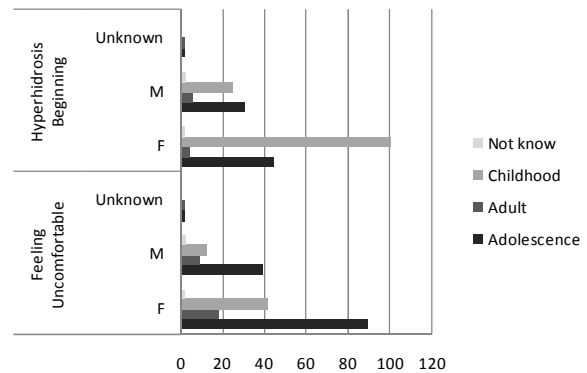

Figure 3. Beginning of the disorder and its drawbacks

The characterization of the hyperhidrosis can be achieved looking at Figure 4, where a pivot table is presented. The data are now presented in a tabular form to give a clear overview of the several values. The number of answers in the prior-surgery and post-surgery is equal to 213, since 199 individuals answered to both questionnaires, 14 answered only to the prior-surgery, and others 14 to the post-surgery. With respect to the answers before the surgery, almost all types of hyperhidrosis are considered severe (excessive), particularly in what concerns to palmar and plantar hyperhidrosis. In terms of a moderate incidence, it is verified in the armpits. Facial hyperhidrosis is, in most of the cases, inexistent or not relevant. After surgery, the number of cases of palmar hyperhidrosis decreases considerably. Plantar hyperhidrosis also presents a lower number of cases, although this surgery does not apply to this type of hyperhidrosis.

Count Hyperhidrosis		Types of Hyperhidrosis				
Questionnaire	Hyperhidrosis Grade	Facial Hyperhidrosis	Armpits Hyperhidrosis	Palmar Hyperhidrosis	Plantar Hyperhidrosis	Grand Total
Prior Surgery	Excessive	5	99	179	150	433
	Inexistent	93	13	7	5	118
	Irrelevant	48	29	9	12	98
	Moderate	26	52	14	36	128
	Not know	41	20	4	10	75
Prior Surgery Total		213	213	213	213	852
Post-Surgery	Definitive	9	24	21	83	137
	Inexistent	136	79	151	24	390
	Not know	29	19	9	19	76
	Temporary	39	91	32	87	249
Post-Surgery Total		213	213	213	213	852
Grand Total		426	426	426	426	1704

Figure 4. Hyperhidrosis characterization

The emotional state of the patients is seriously affected by this disease, as individuals see their everyday lives affected by this disorder. Figure 5 and Figure 6 show the pivot charts with the evolution of the emotional state from a prior-surgery to a post-surgery scenario. The fear of compliments and the need to hide the disease are present in most of the patients in an excessive way, followed by the incomprehension feeling concerning the disease, the insecurity, the socialization difficulties and the shame of talk about the disease in an expressive way. In a post-surgery scenario, all these feelings are mostly inexistent or irrelevant to the patients.
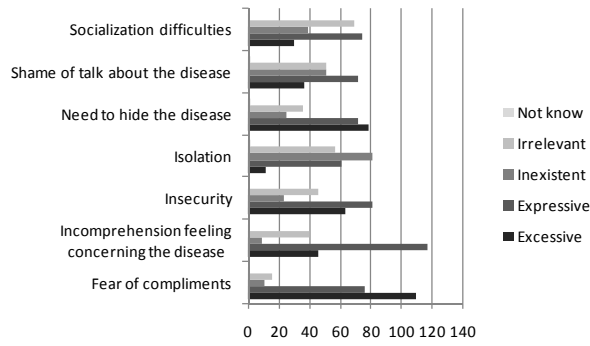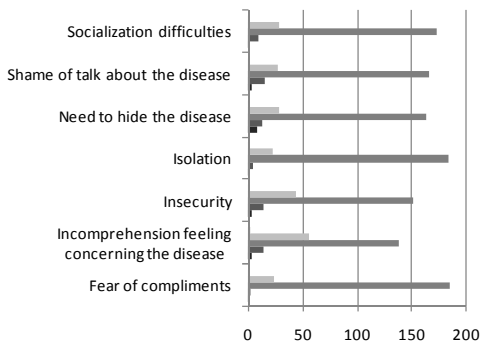
Figure 5. Emotional state (before surgery)

Figure 6. Emotional state (after surgery)

The analysis of the compensatory hyperhidrosis at the body part that is affected also shows interesting results. Looking at the data available for the incidence of compensatory hyperhidrosis 6 months after the surgery, and partitioning the data into two subsets, until December of 2006 and after this date, Figure 7 and Figure 8 show the obtained pivot charts. The unsupportable, excessive and moderate compensatory hyperhidrosis decreased in all cases (abdominal, dorsal and legs), after the change of the surgical procedure, contributing to the success of this treatment and also to the improvement of the quality of life of the individuals.
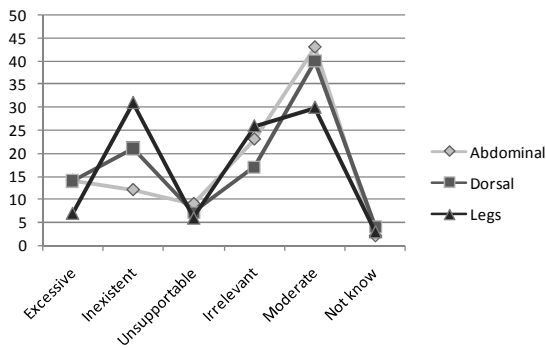

Figure 7. Compensatory hyperhidrosis: 6 months after the surgery and surgery date until December 2006
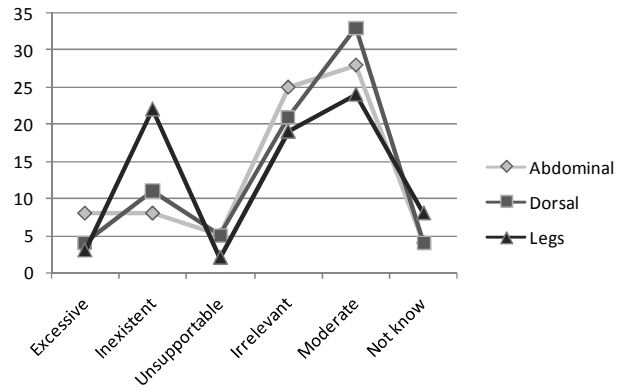

Figure 8. Compensatory hyperhidrosis: 6 months after the surgery and surgery date from January 2007

In several of the presented analyses, the improvement of the quality of life of the individuals has been demonstrated by the changes in their emotional state and in the decrease of incidence of unsupportable, excessive and moderate compensatory hyperhidrosis. Figure 9 presents another view of this reality, showing the measure that result of the SF-6D questionnaire. The several answers to this component of the questionnaires are converted into a numerical value between 0 and 1, representing 1 the best possible health condition (a perfect health) and 0 the worst one [13].
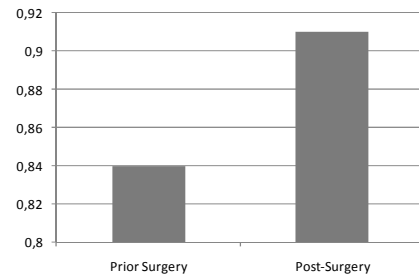

Figure 9. Measure of the quality of life (SF-6D questionnaire)

After the analysis of data with OLAP technology, the results achieved using DM are now presented. This technology can be used to understand or to describe the data, and to identify predictive models that could be used to support decisions or to know what is most likely to happen. In the context of this work, we are concerned with the identification of models that have predictive capacities, as the understanding and description of data were already achieved with the OLAP technology.

For the identification of predictive models, an algorithm for the induction of decision trees [14] is used. It requires the identification of the input parameters and the output parameter. The input attributes are used in the learning process to predict the output parameter. In order to test the confidence of the identified models, the sample data is divided in two: the training dataset, used to identify the model, and the test dataset, used to verify the accuracy of the

model. In all DM tasks, presented afterwards, the partition used was 60% for training and 40% for testing. This partition was the one that maximized the performance of the obtained models. The tool used in the DM process was the SAS Enterprise Miner® v9.1.

The first DM task analyzed the emotional state of the individuals in order to predict if they have been submitted to the surgery, or not, attending to the grade associated with each emotional state. The resulting decision tree is presented in Figure 10. Through the analysis of Figure 10 it is possible to verify that if the Emotional State Grade is equal to Inexistent, the patient is in a post-surgery scenario in 83.7% of the cases. If the Emotional State Grade is equal to Expressive or Excessive, the patient is in a prior-surgery scenario in 91.6% of the cases. However, when the Emotional State Grade has the values Irrelevant or Not know, it is necessary to analyze the Emotional State Type in order to be possible to get a decision. The values Incomprehension feeling concerning the disease, Fear of compliments and Insecurity lead to a nearly tie scenario, where the patient could be already submitted to the surgery, or not. If the Emotional State Type is related with Need to hide the disease, Socialization difficulties, Shame of talk about the disease or Isolation, the patient is in 71.2% of the cases in a prior-surgery situation.
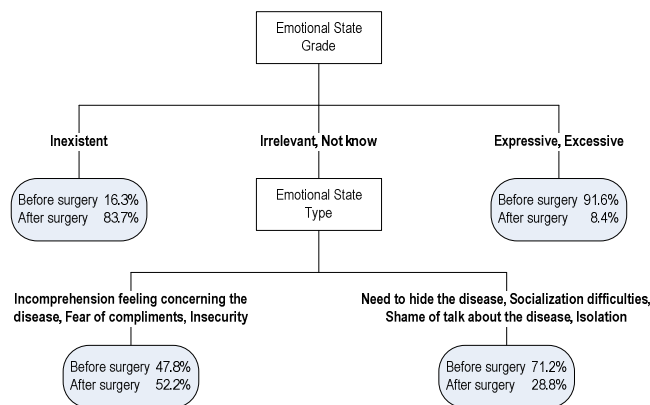

Figure 10. Decision tree for the emotional state

The analysis of the decision tree associated with the emotional state help us to understand the emotions associated with the individuals that are affected by this disorder and the negative impact that it has on their lives.

In order to know if this decision tree has any prediction capabilities, or not, its accuracy was tested. Figure 11 shows the obtained results. The Misclassification Rate column indicates the wrong predictions in the train dataset, while the column Valid:Misclassification Rate presents the wrong predictions in the test dataset (the one that is used to validate the prediction capability of the model). The obtained model presents an accuracy of 82.6%.

Misclassification Rate	Valid:Misclassification Rate
0.1900503074	0.1735121542

Figure 11. Misclassification rate for the emotional state decision tree

The second analysis predicts the incidence of side effects as a consequence of the surgery. The incidence of side effect is low, but the three analyzed complications present different behaviors, as the percentage of incidence is different among them. Figure 12 presents the obtained decision tree. When the Side Effect Type is associated with Horner's Syndrome or with Hemothorax, one of these will be verified in 1.2% of the patients. In the case of Pneumothorax, the percentage of incidence is related with the number of days the patients stayed at the hospital. A permanency of 1 or 2 days in the Hospital could verify an occurrence of Pneumothorax in 11.5% of the surgeries, while a long stay (3, 4, 5, 6 or 7 days) might be associated to a Pneumothorax episode in 16% of the cases.
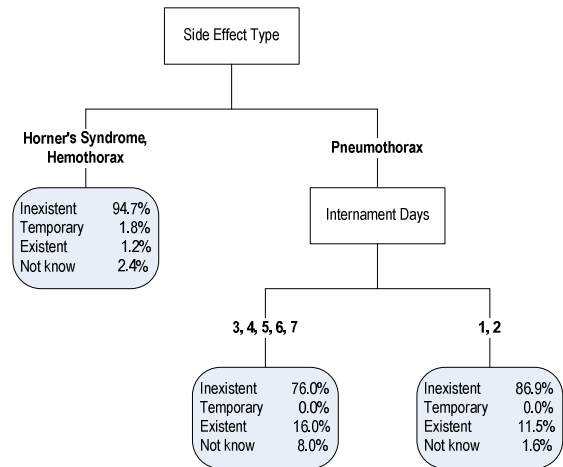

Figure 12. Decision tree for the side effects

The analysis of the accuracy of the presented decision tree is shown in Figure 13. The obtained model presents a high confidence, producing the right decision in 91% of the cases.

Misclassification Rate	Valid:Misclassification Rate
0.1174934726	0.0901960784

Figure 13. Misclassification rate for the side effects decision tree

The last model presented in this paper predicts the gender of the patient attending to the grade of incidence of compensatory hyperhidrosis. As already mentioned, this is one of the major drawbacks associated to the surgery. In the obtained decision tree (Figure 14), an interesting pattern emerged. When the Compensatory Hyperhidrosis Grade is Excessive or Unsupportable, the Age of the individual help us to identify his/her gender. If the Age is between 15 and 29 years old, we are talking about a Male individual in 62.1% of the situations. But, if the Age is less than 15 and greater than 29 years old, they are Female in 84.7% of the cases. In this decision tree we have an explicit relationship between the age and the incidence of compensatory hyperhidrosis (in what gender concerns).

However, it is recommended a deeper analysis in order to investigate this relationship. The result could help to increase the patients' satisfaction with respect to the surgery.
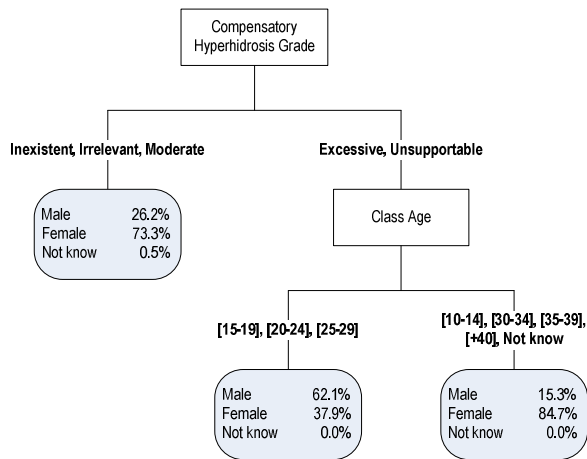
Figure 14. Decision tree for the compensatory hyperhidrosis

Once again, the accuracy of the model was verified. The results are shown in Figure 15. In this case, the model presents an accuracy of 72.9%. Although we verified a decrease of confidence with respect to this model, the findings explicit in the model open new possibilities of analysis. In this case, the collection of more data and the replication of the study could help to improve the accuracy of these results or could show that there is no relevant relationship between these two variables.

Misclassification Rate	Valid:Misclassification Rate
0.258934592	0.2709807887

Figure 15. Misclassification rate for the compensatory hyperhidrosis decision tree

V. CONCLUSIONS

This paper presented a business intelligence system for the analysis of data related with a set of patients submitted to a specific surgery, an endoscopic thoracic sympathectomy. The objective was to analyze the improvement in the quality of life of these individuals, looking at their emotional state and secondary side effects.

The implemented system integrated a data mart for the storage of data, and on-line analytical processing (OLAP) and data mining (DM) technologies for data analysis.

Data from 227 patients were collected using web questionnaires. Each patient answered to two questionnaires. One associated with his/her health condition after the surgery, and the other related to his/her condition in a post-surgery scenario.

The several undertaken analyses, in terms of OLAP, allowed the verification of the beginning of the disorder, the incidence of hyperhidrosis before and after the surgery, the changes in the emotional state of the individuals, the occurrence of compensatory hyperhidrosis as consequence of the surgery and the improvement of the overall quality of life associated to this dataset.

In terms of DM models, three models were identified: related with the emotional sate, with the side effects, and with the compensatory hyperhidrosis grade. Two of these models present high accuracies, helping to support the decision making process. The other one presents a lower accuracy but integrates findings that need further investigation due to their relevance in the association of the gender with the incidence of compensatory hyperhidrosis.

Related to future work, the collection of more data is envisaged. Namely, data associated with patients of other hospitals and health institutions, and also new patients of the analyzed health institutions. In terms of data exploration, a deeper study of the incidence of compensatory hyperhidrosis is needed in order to decrease the incidence of this surgery side effect.

REFERENCES

[1] T. M. Dewey, M. A. Herbert, S. L. Hill, S. L. Prince, and M. J. Mack, "One-Year Follow-Up After Thoracoscopic Sympathectomy for Hyperhidrosis: Outcomes and Consequences," The Annals of Thoracic Surgery, vol. 81, 2006, pp. 1227-1233.

[2] D. H. Yoon, Y. Ha, Y. G. Park, and J. W. Chang, "Thoracoscopic limited T-3 sympathicotomy for primary hyperhidrosis: prevention for compensatory hyperhidrosis," Journal of Neurosurgery: Spine, vol. 99, 2003, pp. 39-43.

[3] J. Cruz, J. Sousa, A. G. Oliveira, and L. Silva-Carvalho, "Effects of endoscopic thoracic sympathectomy for primary hyperhidrosis on cardiac autonomic nervous activity," The Journal of Thoracic and Cardiovascular Surgery, vol. 137, number 3, 2009, pp. 664-669.

[4] J. M. Cruz, M. Fonseca, F. J. Pinto, A. G. Oliveira, and L. S. Carvalho, "Cardiopulmonar effects following endoscopic thoracic sympathectomy for primary hyperhidrosis," European Journal of Cardio-thoracic Surgery, vol. 36, 2009, pp. 491-496.

[5] S. Elia, G. Guggino, D. Mineo, G. Vanni, A. Gatti, and T. C. Mineo, "Awake one stage bilateral thoracoscopic sympathectomy for palmar hyperhidrosis: a safe outpatient procedure," European Journal of Cardio-thoracic Surgery, vol. 28, number 2, 2005, pp. 312-317.

[6] K. Kumagai, H. Kawase, and M. Kawanishi, "Health-Related Quality of Life After Thoracoscopic Sympathectomy for Palmar Hyperhidrosis," The Annals of Thoracic Surgery, vol. 80, 2005, pp. 461-466.

[7] J. E. Brazier, R. Harper, N. M. B. Jones, A. O'Cathian, K. J. Thomas, T. Usherwood, and L. Westlake, "Validating the SF-36 health survey questionnaire: new outcome measure for primary care," BMJ, vol. 305, July 1992, pp. 160-164.

[8] P. L. Ferreira, L. Noronha e Ferreira, "A medição de preferências em saúde na população portuguesa," (in Portuguese) Revista Portuguesa de Saúde Pública, vol. 24, number 2, 2006, pp. 5-14.

[9] W. F. Cody, J. T. Kreulen, V. Krishna, W. S. Spangleret, "The integration of business intelligence and knowledge management," IBM Systems Journal, vol. 41, number 4, 2002, pp. 697-713.

[10] S. Alter, "A Work System View of DSS in its Fourth Decade," Proceedings of AMCIS 2002 - Americas Conference on Information Systems, Dallas, 2002.

[11] W. H. Inmon, "The Data Warehouse and Data Mining," Communications of the ACM, vol. 39, number 11, 1996, pp. 49-50.

[12] J. Han, and M. Kamber, "Data Mining: Concepts and Techniques," Morgan Kaufmann Publishers, 2001.

[13] S. A. Kharroubi, J. E. Brazier, J. Roberts, and A. O'Haganb, "Modelling SF-6D health state reference data using a nonparametric Bayesian method," Journal of Health Economics, vol. 26, 2007, pp. 597–612.

[14] J. R. Quinlan, "Induction of decision trees," Machine Learning, vol. 1, number 1, 1986, pp. 81-106.