Internal Fixation Evaluation: A Machine Learning Approach

Ray Hashemi[1], Cameron Coates[2], Azita Bahrami[3], Mahmood Bahar[4], Alexander Tyler[5], Nicholas Tyler[6], and William Gibson[1]

[1]Department of Computer Science
[2]Engineering Program
[6]Department of Biology
Armstrong Atlantic University
Savannah, GA 31419, USA

[3]IT Consultation
Savannah, GA

[4]Department of Basic Sciences
Garmsar Branch, Islamic Azad University
Garmsar, Iran

[5]Conway Regional Health System
2302 College Avenue, Conway, AR 72034

Abstract- *A major dilemma currently faced by orthopedic surgeons is whether to: retain or remove an internal fixation from patient's body after the bone rebuilds itself. The difficulty stems from the fact that for both options there are total of ten major side-effects. The goal of this research effort is to generate a set of rules of thumb by which a decision can be rapidly reached. The goal was met by (1) creating organic synthetic data using likelihood measures, (2) Calculating systematically the confidence interval of the risk factors for every side-effect, (3) developing a special type of neural network to pre-process the organic patients' records in reference to the ten side-effects, (4) applying Wilcoxon's statistical model to conclude the rules of thumb, and (5) verifying the rules of thumb using a domain expert which resulted in 87% accuracy.*

Key Words: *Internal fixation, Organic synthetic record, Likelihood measures, Neural network, Risk factors, retaining and removal side-effects.*

1. Introduction

Internal fixation means stabilizing and joining the ends of fractured bones by mechanical devices such as metal plates, pins, rods, wires or screws. An example is shown in Figure 1.

The major question facing orthopedic surgeons is: Is it more beneficial to retain or remove an internal fixation from patient's body after fracture healing has occurred? One may ask why this is a major question. Because both options may result in significant side-effects.

Figure 1: An example of internal fixation borrowed from [1].

In the case of retaining fixation, the major adverse effects are: *Metallosis* (adverse reaction of the soft tissue in the body caused by the presence of excess metal ions due to a nearby metallic implant), *carcinogenesis* (implant presence may turn normal cells into cancer cells), *Re-fracture leading to complicated revision*, and *Localized Osteopenia* (decreases in bone mineral density in the vicinity of the implant due to contact with the implant itself and/or stress shielding effects of the implant).

In the case of removing the fixation, the major adverse effects are: *Re-fracture risk* (occurrence of a new fracture in the same area at some point in time), *Iotrogenic fracture risk* (additional fracture complications that may occur during removal of the fixation caused by the activities of the operating physician), *Anesthetic complications* (complications resulting from improper application of anesthesia due to human error, equipment failure, or pre-existing patient related factors such as cardiovascular or respiratory disease), *Nerve damage* (adverse effect on nerves that are at or close to the fracture site being affected by regional anesthesia or during the surgery), *Infection* (contamination of the blood due to bacteria), and *Hematoma* (excessive blood leak into tissues where it does not belong caused by the damaged wall of a blood vessel, artery, vein, or capillary) [2, 3, 4, 5].

An experienced domain expert always uses his/her experience to solve a domain-based problem intuitively. In fact, gained experience is manifested in form of rules of thumb that enable the domain expert in his endeavor. The goal of this research effort is to generate a set of rules of thumb by which a decision for a patient with internal fixation can be rapidly reached.

The rest of the paper is organized as follow: The Previous Works is the subject of section 2. Methodology is presented in section 3. Results and Discussion are covered in section 4. Conclusion and Future Research are the subjects of section 5.

2. Previous Works

The closest research activities to the system that is presented and discussed in this paper is the work of Hanson et al. [6]. Henson and his research team reported the results of a survey in which 655 orthopedic surgeons from 65 countries participated. Among other things, the survey tried to get the answer to the question of when removal of an internal fixation is preferred by the surgeons. Fifty-eight percent of the participants are against the removal of internal fixation partially because they do not believe in the severity of the side-effects associated with the retained implants. Forty-eight percent of the surgeons believe there are more risk in removal than retained implants. However, in the case that patient is a child the removal of internal fixation is highly considered. The justification for this exceptional case is that the children have a growing skeleton.

The reader needs to be reminded, the results are extracted from a survey and do not have any clinical trial foundation. In contrast, we try to investigate and build a decision support system that evaluates both removing and retaining of the internal fixation based on the synthetic patients' data using a machine learning approach. To the best of our knowledge there is no report of such a system in literature.

3. Methodology

The goal of this research is met in four steps of (1) Creating synthetic patients' records, (2) Introducing a new neural network for pre-processing the patient's records, (3) Applying Wilcoxon statistical model to the pre-processed records, and (4) obtaining the rules of thumb. Each step is covered in detail in the following four subsections.

3.1 Creating Synthetic Patients' Records

Let $A = \{a_1, . . . , a_n\}$ and $B = \{b_1, . . ., b_m\}$ be the set of side-effects for retaining and removing internal fixations, respectively. Comparing the sets A and B is not possible because side-effects in A are different from side-effects in B. To make A and B comparable, we introduce a new set of side-effects that belongs to both retaining and removing of the internal fixations. The new set of side-effects is $S = \{s_1, . . . ,, s_{(n+m)}\}$ where, $S = A \cup B$ and $|S| = |A| + |B|$. For the investigation in hand n+m = 10.

Table 1: Patient attributes and categories

Cat.	**Age**	**Weight**
1	<16	Underweight
2	16-35	Normal
3	35-60	Over Weight
4	>60	Obese
	Physical Activity Level	**Health Problems**
1	Sedentary	None
2	Moderate	Low
3	Very Active	Medium
4	Elite Athlete	Serious

A patient is modeled by four attributes of *Age*, *Weight*, *Physical activity level*, and *Health problems*. Each attribute has four possible categorical values of 1, 2, 3, and 4, shown in Table 1.

Let us look at two synthetic patients' records that are in the same age group and having the same weight, but Physical Activity Level for the first record is *sedentary* while for the second one is *elite athlete.* Let us assume the value for the forth attribute (Health problems) of both records is the same and it says the patients have *serious* health problems. Common sense suggests that it is less likely for an elite athlete to have serious health problems. Therefore, the first record looks more *organic* than the second one.

Table 2: Likelihood values for all possible patterns of set (Age, Weight, Physical Activity Level).

Age	**Weight**	**Physical Activity Level**			
		1	**2**	**3**	**4**
1	**1**	0.92	**0.02**	0.02	0.0
	2	0.05	0.1	0.84	0.01
	3	0.9	0.1	0.0	0.0
	4	0.99	0.01	0.0	0.0
2	**1**	0.8	0.15	0.05	0.0
	2	0.1	0.3	0.5	0.1
	3	0.78	0.15	0.05	**0.02**
	4	0.98	0.02	0.0	0.0
3	**1**	0.9	0.1	0.0	0.0
	2	0.01	0.64	0.35	0.0
	3	0.85	0.15	0.0	0.0
	4	0.95	0.05	0.0	**0.0**
4	**1**	0.98	**0.02**	0.0	0.0
	2	0.1	0.8	0.1	0.0
	3	0.9	0.1	0.0	0.0
	4	0.95	0.05	0.0	0.0

Another point that needs to be made is that the values for attributes Age and Weight will also correlate with the patient being an elite athlete. For example if the age value for both patients is 4 (i.e. >60) the likelihood of the person being an elite athlete is zero.

To create organic synthetic patient records, the domain expert assigns a *likelihood* value, in the range of [0-1), to every possible combinations of attribute values (*patterns*). This is done systematically by assigning likelihood values to all the possible combinations of attributes of {Age, Weight, and Physical Activity Level} and {Age, Weight, Health Problems}, Tables 2 and 3.

For each Age value, one of the likelihood values is designated as the *threshold* by domain expert and it is shown in bold. Therefore, there are four threshold values for each one of the Tables 2 and 3.

A patient record that is composed of four values is checked against both Tables 2 and 3. If one of the patterns for {Age, Weight, Physical Activity Level} or {Age, Weight, Health Problems}, has the likelihood value less than or equal to the corresponding threshold value, the record is dismissed because it is not organic.

The tables for likelihood values are used by the algorithm ORGANIC for creating organic patient records.

Table 3: Likelihood values for all possible patterns of set (Age, Weight, Health Problems).

Age	**Weight**	**Health Problems**			
		1	**2**	**3**	**4**
1	**1**	0.25	0.25	0.25	0.25
	2	0.99	**0.01**	0.0	0.0
	3	0.85	0.15	0.0	0.0
	4	0.55	0.45	0.0	0.0
2	**1**	0.1	0.3	0.3	0.3
	2	0.88	0.02	0.0	0.10
	3	0.5	0.05	0.05	0.4
	4	**0.02**	0.38	0.1	0.5
3	**1**	0.7	0.1	0.1	0.1
	2	0.01	0.03	0.15	0.3
	3	0.85	0.1	0.15	0.1
	4	**0.01**	0.4	0.19	0.4
4	**1**	0.3	0.2	0.2	0.3
	2	0.4	0.1	0.3	0.2
	3	0.2	0.2	0.3	0.3
	4	**0.0**	0.4	0.3	0.3

Algorithm ORGANIC

Given: Table 2, Table 3, set V = {1, 2, 3, 4), four threshold values for Table 2 ($T_{2,1}$, $T_{2,2}$, $T_{2,3}$, $T_{2,4}$), four threshold values for Table 3 ($T_{3,1}$, $T_{3,2}$, $T_{3,3}$, $T_{3,4}$), and four attributes of Age, Weight, Physical Activity Level, and Health Problems.

Objective: Creation of an organic patient record.

Step 1: Randomly generate two values (i, j) from set V and assign them to attributes Age and Weight.

Step 2: Randomly generate a value (k) from set V.
If Table2(i, j, k) $\le T_{2,i}$
Then go to Step2;
Else Assign k to attribute Physical Activity Level;

Step 3: Randomly generate a value (ℓ) from set V.
If Table3(i, j, ℓ) $\le T_{3,i}$
Then go to Step3;
Else Assign ℓ to attribute Health Problems;

Step 4: End;

Algorithm ORGANIC recognizes only 113 pattern out of the 256 possible patterns as organic.

To each value in an organic patient's record, 10 side-effects are related. Let v be a possible value for one of the attributes of patient's record and s_i be one of the side-effects. There is a risk factor (probability) involved with s_i in reference to v that is denoted as $P(s_i|v)$. The $P(s_i|v)$ and its confidence interval are calculated using the following procedure.

Procedure: A population, G, of 1000 patients with internal fixation was randomly created. Each patient had four random values (borrowed from set {1, 2, 3, 4}) for the four attributes of Age, Weight, Physical Activity Level, and Health Problems. Population G included only organic patients' records

Table 4: Risk Factors

Side Effect *	Independent Variables			
	Age			
	1	2	3	4
M	0.7..0.9	0.5..0.7	0.3..0.5	0.2..0.3
C	0..0.3	0.3..0.5	0.5..0.7	0.7..0.9
RC	0.4..0.6	0.4..0.6	0.5..0.7	0.7..0.9
LO	0.2..0.4	0.1..0.2	0.3..0.5	0.2..0.4
RR	0.1..0.3	0.3..0.5	0.1..0.2	0.3..0.5
LF	0.1..0.3	0.05..0.1	0.5..0.7	0.3..0.5
AC	0.3..0.5	0.1..0.2	0.1..0.2	0.3..0.5
ND	0.3..0.5	0.1..0.2	0.1..0.2	0.3..0.5
I	0.4..0.6	0.3..0.5	0.3..0.5	0.4..0.6
H	0.1..0.2	0.2..0.3	0.2..0.3	0.4..0.6
	Weight			
M	0	0	0	0
C	0	0	0	0
RC	0.1..0.2	0.0..0.05	0.3..0.4	0.1..0.2
LO	0.1..0.2	0.0..0.05	0.1..0.2	0.1..0.2
RR	0.3..0.5	0.1..0.2	0.2..0.4	0.3..0.5
LF	0.2..0.4	0.2..0.4	0.3..0.5	0.4..0.7
AC	0.0..0.2	0.0..0.1	0.1..0.3	0.2..0.5
ND	0.2..0.4	0.2..0.4	0.3..0.5	0.3..0.5
I	0.2..0.4	0.2..0.4	0.2..0.4	0.3..0.5
H	0.0..0.2	0.0..0.2	0.1..0.3	0.1..0.3
	Physical Activity Level			
M	0	0	0	0
C	0	0	0	0
RC	0.0..0.05	0.05..1	0.95..0.99	0.3..0.5
LO	0.1..0.2	0.05..0.07	0.1..0.2	0.05..0.1
RR	0.0..0.2	0.2..0.4	0.3..0.5	0.5..0.8
LF	0.2..0.4	0.0..0.2	0.0..0.1	0.0..0.1
AC	0.2..0.4	0.1..0.3	0.0..0.2	0.0..0.2
ND	0.2..0.4	0.1..0.2	0.1..0.2	0.1..0.2
I	0.2..0.4	0.3..0.5	0.2..0.3	0.0..0.2
H	0.1..0.3	0.1.0.4	0.0..0.2	0.0..0.2
	Health Problems			
M	0	0.2..0.3	0.05..0.1	0.1..0.2
C	0	0	0	0
RC	0	0.3..0.4	0.1..0.2	0.2..0.3
LO	0.0..0.1	0.05..0.1	0.1..0.2	0.1..0.2
RR	0	0.3..0.4	0.1..0.2	0.2..0.3
LF	0	0.2..0.3	0.05..0.1	0.1..0.2
AC	0.0..0.2	0.05..0.1	0.1..0.2	0.2..0.3
ND	0	0	0	0
I	0.01..0.04	0.05..0.1	0.05..0.1	0.1..0.2
H	0.1..0.2	0.05..0.1	0.2..0.3	0.3..0.4

* **M**: Metallosis, **C**:carcinogenesi, **RC**: Re-fracture leading to complicated revision, **LO**: Localized Osteopenia, **RR**: Re-fracture risk, **LF**: Lotrogenic fracture risk, **AC**: Anesthetic complications, **ND**: Nerve Damage, **I**:nfection, and **H**: Hematoma.

.

In addition, one side-effect, s_i, was added to G. Value of s_i for a given record was randomly assigned and it was either 0 or 1. Zero means the patient did not suffer from s_i and "1" means otherwise.

Let K be the number of patients in G suffering from s_i and let L be the number of patients within K with value v for a designated attribute. In addition, Let M be the number of patients in G with value v for the same attribute. The following probabilities can be calculated for the population: $P(s_i) = K/|G|$, $P(v|s_i) = L/K$ and $P(v) = M/|G|$. Out of these probability one can calculate $P(s_i|v)$ using formula (1):

$$P(s_i|v) = \frac{P(v|s_i).P(s_i)}{P(v)} \qquad (1)$$

The process was repeated 10 times and each time a new randomly generated population was created and $P(s_i|v)$ was calculated for the new G.

Let μ be the mean for $P(s_i|v)$ and let the ten values of $P(s_i|v)$ be denoted as x_j, j = 1 to 10 . The objective here is to establish a confidence interval for μ. The estimated mean is $\bar{X} = \frac{1}{10}\sum_1^{10} x_j$ and the estimated variance of the probabilities is $S^2 = \frac{1}{9}\sum_1^{10}(x_j - \bar{X})^2$. The (1-α)% confidence interval for μ is given by formula (2):

$$\bar{X} - t_{u,\alpha/2}\,\frac{S}{\sqrt{10}} \le \mu \le \bar{X} + t_{u,\alpha/2}\,\frac{S}{\sqrt{10}} \qquad (2)$$

where $t_{u,\alpha/2}$ is the t-value with u = 9 degrees of freedom from the student t probability distribution, leaving an area of α/2 to the right.

End of procedure.

We used α = 0.05 for the calculation of the confidence intervals. For s_i = *Metallosis*, the obtained interval value for v = 1 of the attribute Age is [0.7, 0.9]. This means, P(*Metallosis* | 1) for 95% of the patient populations falls in the range of [0.7-0.9].

The above procedure was repeated to obtain the interval of the risk factor for all possible values of the four attributes and for all 10 side-effects. The results are shown in Table 4. Some of the intervals in the table have zero as a value and it means there is no risk factor for the corresponding side-effect. The confidence intervals of zeros are not produced by the procedure and they are suggested by the domain expert.

3.2 A Neural Network for Pre-Processing the Patients' Records

Each patient's record has four risk factors for every side-effect which their sum represents the *strength* of the side-effect for the record. In addition, each side-effect influences decisions for both retaining and removing of the internal fixation. This influence is represented by a *weight* assigned to each side-effect by the expert. The side-effects' weights are shown in Table 5.

Table 5: The weights for the side-effects' weights.

Retaining		Removing	
Side Effect	**Influence Weight**	**Side Effect**	**Influence Weight**
Metallosis (M)	1.2	Re-Fracture within 18 months (R)	1.2
carcinogenesis (C)	1	Lotrogenic fracture (L)	1
Re-Fracture and Revision (r)	1.6	Anesthetic Complications (A)	1
Osteopenia (O)	1	Nerve Damage (N)	0.3
		Infection (I)	0.5
		Hematoma (H)	0.4

A new neural network (Figure 2) is developed that is able to (1) calculate the strength of each side-effect for every patient's record (2) treat all the side-effects pertaining to retaining decision as one entity and all the side-effects related to the removing decision as another entity, and (3) deliver a quantitative influence for each entity on their corresponding decisions. The output of the neural net is directly used by the Wilcoxon model.

The neural network is made up of three layers. The first layer, input layer, accepts a patient's record. Considering the fact that each patient's record has four values, the number of nodes in the first layer is four.

The i-th node (for i = 1 to 4) of the input layer is made up of a look-up table with four columns and ten rows. Each column represents risk factors for one of the four possible values of the i-th attribute. Each node uses its input as column index and then the ten values of the selected column serve as the output of the node. The weight matrix, W, for all the connections between the first and the second layer are initialized with value of one.

The second layer, hidden layer, has ten nodes representing the ten side-effects. Each node has four inputs from the first layer and produces one output using the following formula:

$$\sigma_j = \sum_{i=1}^{4} Input_{i,j} * w_{i,j} \quad \textit{(For j = 1 to 10)} \qquad (1)$$

where, $Input_{i,j}$ is the input from $node_i$ of the input layer to $node_j$ of the hidden layer and $w_{i,j}$ is the weight for the connections between $node_i$ of the input layer and $node_j$ of the hidden layer.

The third layer, output layer, has only two nodes of A and B. Node A receives only the output of the first four nodes of the hidden layer and the node B receives the output of the last six nodes of the hidden layer. The output for the nodes A and B are calculated using formulas 2 and 3.

$$x = \sum_{j=1}^{4} \sigma_j * u_j \quad (2)$$
$$y = \sum_{j=5}^{10} \sigma_j * u_j \quad (3)$$

where, σ_j is the input from $node_i$ of the hidden layer to the corresponding node in the output layer and u_j is the weight for the connections between $node_j$ of the hidden layer and the corresponding node in the output layer and its value is borrowed from Table 5.

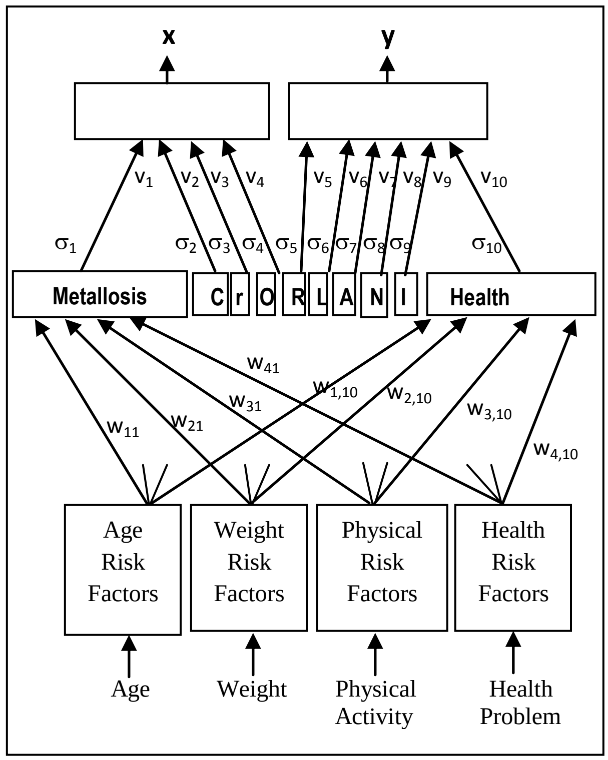

Figure 2: The neural Network

The output of the first and the second nodes in the third layer are referred to as x and y and represent the final outcome of a pre-processed patient's record.

.

3.3 The Wilcoxon Statistical Model

Let S1 and S2 be two samples of a population (S1 and S2 are not necessarily distinct) and E1 and E2 be two experimental events (either of the two events can be null but not both of them). Let S1 be exposed to E1 and S2 to E2. In addition, let the differences observed between S1 and S2 after exposure be D. If D is not significant, the null hypothesis (that E1 and E2 did not have lasting effects on S1 and S2) is rejected and the alternative hypothesis that E1 or E2 has a lasting effect on its corresponding sample is accepted.

As an example, two groups of patients (S1 and S2) who have the same sickness have been selected. One group (S1) is treated by a new drug (E1) but the second group (S2) is not treated at all (E2= $\emptyset$). If the differences observed between the two groups (D) is not significant then the drug is not effective on the sickness; otherwise it is.

For the problem that in hand, S1 and S2 are the same (the same sample of population). E1 and E2 are all the risk factors for side-effects of *retaining* and *removing* internal fixation, respectively. The null hypothesis (H_0) is that the there is no preference in either retaining or removing the internal fixation and the alternative hypothesis (H_1) is that there is a preference.

The significance of D may be determined by Wilcoxon [7, 8] statistics, or paired t test [9]. The former one is chosen for two reasons: (1) it is used for population with non-Gaussian distribution and (2) it adapts to arbitrary sample size.

The Wilcoxon model is applied to either reject or accept the null hypothesis when the null hypothesis is rejected, then H_1 suggests that one of the decisions (retaining or removing) is preferred. The following algorithm, DECISION, is used to determine the preferred one.

The algorithm DECISION works with a set of randomly generated population, S, for a given pattern. In this population, the randomly generated risk factor for each side-effect is in the range prescribed by Table 4 for the pattern. The difference between side-effects for retaining and removing the internal fixation of the population S is calculated using formula (4).

$$D = \sum_{i=1}^{|S|} E_{1,i} - \sum_{i=1}^{|S|} E_{2,i} \qquad (4)$$

Based on the significance of D, the decision outcome for a valid pattern is: *No-Preference*, *Retaining*, or *Removing*.

Algorithm DECISION

Given: E1 and E2 (All the risk factors for side-effects of *retaining* and *removing* internal fixation). A null Hypothesis (H_0) that says no preference between E1 and E2. A population S that includes randomly generated patients' records for a given valid pattern. Wilcox critical value table and the neural network of Figure 2.

Objective: Evaluate the internal fixation.

Step 1: Pre-process S using the neural network; X=$\sum_1^{|S|} x_i$ and Y=$\sum_1^{|S|} y_i$;

Step 2: Using the Wilcoxon nomenclature, X and Y are considered total absolute value of positive and negative ranks, respectively;

Step 4: R = Min(X, Y); and the sign of R is borrowed from D.

Step 4: Use Wilcoxon critical value table to get the sum of total rank (ρ) using number of patients (df) and confidence level α=0.05.

Step 5: If (ρ >R)
Then /*H0 is true */
Decision← No-Preference;
Else hypothesis is false;
If(R is positive)
Then Decision← Retaining;
Else Decision← Removing;

Step 6: End;

Step 5 needs further explanation. In the case of null hypothesis rejection, if R=X, it means the rank value for X (i.e. retaining) was smaller than rank value for Y. Therefore, total risk probability for retention is less than the total risk probability for removal. Thus, decision is for "retaining". If R =Y using the same reasoning, decision is for "removing".

3.4. Rules of Thumb

Each pattern with a decision can be presented as an *if-then* rule. For example, if for the pattern PAT = "1223" the decision is *retaining*, then the following rule with four conditions can be generated:

If (Age = 1) ^ (Weight = 2) ^
(Physical Activity Level = 2) ^
(Health Problems = 3)
then Decision = Retaining.

Therefore, we use the terms attribute and condition interchangeably.

Table 6: Patterns with mixture of decisions

#	Pattern	Decision	#	Pattern	Decision
1	1 2 1 3	Remove	6	1 3 3 3	Remove
2	2 1 4 1	Retain	7	2 4 4 3	Remove
3	3 3 3 3	Retain	8	2 3 4 1	Retain
4	4 3 3 2	Retain	9	1 2 2 4	Retain
5	2 1 2 3	Remove	10	3 2 1 3	Retain

Rules of thumb are the generalization of retaining and removing rules that are compact and easy to remember. We generalize the rules using a modified Dropping Condition Approach [10]. In this approach, a minimum subset of conditions (attributes) in a given rule is kept such that the values for the subset of attributes can be found only in the removing rules, for example, and not in any of the retaining rules.

To provide an example of the generalization approach, let the patterns of Table 6 have a mixture of retaining and removing decisions. The four values in each pattern represent the attributes Age, Weight, Physical Activity Level, and Health Problems, respectively. Values for any of the attributes cannot exclusively identify patterns for one of the decisions. However, values for Age and Health Problems, collectively, can identify all the patterns of the removing decision and none of the retaining decision. One may conclude that if combination values for Age and Health Problems can represent the patterns of removing decision by inclusion, it can also represent the patterns of retaining rules by exclusion. As a result, the following general rule can represent both removing and retaining rules:

If ((Age = 1) $\vee$ (Age = 2)) $\wedge$
(Health Problems = 3)
Then Decision = Removing;
Else Decision = Retaining;

The above rule is compact and easy to remember.

4. Results and Discussion

There is a total of 256 patterns (four attributes and four possible values for each attribute) from which 113 of them are valid patterns (using algorithm ORGANIC).

Let PAT = "1223" be a valid pattern. Each value in this pattern has a range of risk factors for the ten side-effects. Therefore, one can generate M number of patients' records for which the pattern is the same but the risk factors for each value of the pattern may be different. Analysis of the M records using Algorithm Decision produces one of the following three decisions for PAT: *no-preference*, *retaining*, and *removing*.

Results revealed that no patterns has the decision no-preference, 92 patterns have the decision retaining, and 21 patterns have the decision removing. The generalization of the rules generated the following rule of thumb:

If *(Weight* $\neq 4$*)* $\wedge$
((Physical Activity Level = 3) $\vee$
(Physical Activity Level = 4))
Then *Decision = Removing;*
Else *Decision = Retaining;*

To validate the above rule of thumb, (1) decisions generated for all 113 valid patterns shared with domain expert and 87% of the decisions were confirmed by the expert.

5. Conclusion and Future Research

An extra effort has been dedicated to the creation of the organic patient records. Use of likelihood measures and confidence interval to determine the range of risk factors for each side-effect are crucial in support of having randomly generated organic patients' records. The outcome of the verification of the rule of thumb by the domain expert is a good indicator for: (1) the quality of the organic patient's records and (2) viability of the presented methodology.

As future research, the collection of data from the real patients and assembling a team of domain experts for verification of the rules of thumb generated by the presented methodology are in progress.

6. References

1. http://en.wikipedia.org/wiki/Internal_fixation, retrieved in November 2011.

2. SB Keel, KA Jaffe, G Petur Nielsen, AE Rosenberg, "Orthopaedic implant related sarcoma: A study of twelve cases," Mod Pathol, 14(10), 969-977, 2001.

3. LB Signorello, W Ye, JP Fryzek, L. Lipworth, JF Fraumeni Jr., WJ Blot, JK McLaughlin,O. Nyrén, "Nationwide study of cancer risk among hip replacement patients in Sweden," J Natl Cancer Inst, 93(18), 1405-1410, 2001.

4. ML Busam, RJ Esther, WT Obremskey, "Hardware Removal: Indications and Expectations," J Am Acad Orthop Surg, 14(2),113-120, 2006.

5. W. Jamil, M, Allami, MZ Choudhury, C. Mann, T. Bagga, A. Roberts, "Do orthopaedic surgeons need a policy on the removal of metalwork? A descriptive national survey of practicing surgeons in the United Kingdom", Injury, Int. J. Care Injured, 2008, 39(3), 362-367, 2008.

6. B. Hanson, ML CVD Werken, and D. Stengel, "Surgeons' beliefs and perceptions about removal of orthopaedic implants", BMC Musculoskeletal Disorders 2008, 9:73.

7. M. Hollander and D. A. Wolfe, Nonparametric Statistical Methods, 2nd Ed., A Wiley Interscience Publication. 1999.

8. G.W. Corder and D.I. Foreman, Nonparametric Statistics for Non-Statisticians: A Step-by-Step Approach, New Jersey: Wiley, 2009.

9. H. Motulsky, Intuitive Biostatistics, Oxford University Press Inc, 1995.

10. R. Hashemi, A. Tyler, A. Bahrami, "Use of Rough Sets as a Data Mining Tool for Experimental Bio-Data", "Computational Intelligence in Biomedicine and Bioinformatics: Current Trends and Applicationjs", Springer-Verlag Publisher, June 2008, pp. 69-91.