

Co-Morbidity in Bariatric Patients

A new approach in quantifying the severity

T.P.A. BEISHUIZEN (0791613)
Biomedical Engineering - Computational Biology
Data Engineering - Information Systems
Eindhoven, University of Technology
Email: `t.p.a.beishuizen@student.tue.nl`

November 3, 2017

Contents

1	Introduction	2
1.1	Data Sets	2
1.2	Co-morbidity Marker Relation	4
2	First Proposal	4
2.1	Research Question	4
2.2	Hypothesis	6
2.3	Methods	6
3	Proposal Comparisons	6
4	Final Proposal	6

1 Introduction

Worldwide the number of bariatric surgeries is increasing. Although initially thought otherwise, this type of surgery has added benefits on top of losing weight, the primary reason. Among those benefits the remission of metabolic co-morbidities can be named. Due to binary labelling of those co-morbidities, valuable information is lost, while this labelling is not clearly defined either. To obtain more and better results, this binary labelling can be replaced by a continuous severity score. Ruben Deneer conducted a research on trying to achieve a successful replacement.[1]

The available data used for the research stemmed from the Catharina Hospital in Eindhoven (Subsection 1.1). This extensive data set was a combination of two data sets and consisted of 41 markers measured pre- and post-surgery for 2367 patients that underwent gastric sleeve or gastric bypass surgery. These 41 markers are divided in several sub-panels that each describe different processes in the body. Conditions for these co-morbidities that were tested for the severity score were type II diabetes mellitus (T2DM), hypertension and dyslipidemia. Extensive literature research was done to connect them with 41 markers to find possible relations (Subsection 1.2). [1]

To better specify the research on the data, a main goal was created: *to use data mining techniques to develop score that can objectively quantify the severity of co-morbidities present in bariatric patients based on biomarkers, both before and after surgery*. This means that it is not the goal to predict the outcome of the two types of surgery, but it is to quantify the improvement of the co-morbidities before and after the surgery. [1]

Since the data is not available other than within the hospital, a figurative second study was conducted to complement Ruben Deneer's research. This study consists of three parts. The first part was creating a research question, hypothesis and an own analysis approach without the knowledge of Deneer's way. Since no data is available, this can only be made globally. It should be detailed enough, though, to compare it with Deneer's own approach. Secondly the two are compared and remarks are given on the original approach. At last both are combined to create a final proposal for a possible follow-up research.

1.1 Data Sets

Two data sets were used in the study. The first one is called "The Dutch Audit for Treatment of Obesity" (DATO). This data set is a national database that houses all registrations and health statuses of pre- and post treatment bariatric surgery patients in the Netherlands. Before surgery, the co-morbidities T2DM, hypertension and dyslipidemia were given a binary label of "Yes/No". After surgery they were given one of the following labels:

1. **Cured** No co-morbidity any more
2. **Improved** Less affected by co-morbidity
3. **Same** No change in co-morbidity status
4. **Worse** More affected by co-morbidity
5. **Denovo** Diagnosed co-morbidity while not present before surgery
6. **Not present** No co-morbidity present

The second data set came from a laboratory database, stored in health records. This extensive data set consisted of 3 clinical and 38 blood markers measured pre- and 6, 12 and 24 months post-surgery. The tests pre-surgery had some additional markers on top of the 41 ones. These markers can be divided in the following categories: (Table 1.1) Complete blood count, liver function, kidney function, inflammation, lipid spectrum, coagulation, glucose metabolism, thyroid function and at last minerals and vitamins. The data sets of the patients that underwent bariatric surgery can be extracted from these. [1]

Table 1: The markers present in the bariatric laboratory data set [1]

	Before Surgery/Pre-Op/Screening	After Surgery/Post-Op/Follow-up
Complete blood count	hemoglobin	hemoglobin
	hematocrit	hematocrit
	erythrocytes	erythrocytes
	mean corpuscular hemoglobin	mean corpuscular hemoglobin
	mean corpuscular volume	mean corpuscular volume
	thrombocytes	thrombocytes
Liver function	leukocytes	leukocytes
	bilirubin	bilirubin
	aspartate aminotransferase	aspartate aminotransferase
	alanine aminotransferase	alanine aminotransferase
	lactate dehydrogenase	lactate dehydrogenase
	alkaline phosphatase	alkaline phosphatase
Kidney function	gamma-glutamyltransferase	gamma-glutamyltransferase
	urea	urea
	creatinine	creatinine
	potassium	potassium
	sodium	sodium
	calcium	calcium
Inflammation	phosphate	phosphate
	albumin	albumin
Lipid spectrum	C-reactive protein	C-reactive protein
	total cholesterol	total cholesterol
	high-density lipoprotein-cholesterol	high-density lipoprotein-cholesterol
	total/high-density cholesterol ratio	total/high-density cholesterol ratio
	low-density lipoprotein-cholesterol	low-density lipoprotein-cholesterol
Coagulation	triglycerides	triglycerides
	prothrombin time	prothrombin time
Glucose metabolism	hemoglobin A1c (IFCC)	hemoglobin A1c (IFCC)
	glucose	glucose
	insulin	-
	C-peptide	-
Thyroid function	parathyroid hormone	parathyroid hormone
	thyroid-stimulating hormone	-
	free T4	-
	cortisol	-
Minerals and vitamins	iron	iron
	ferritin	ferritin
	folic acid	folic acid
	zinc	-
	magnesium	-
	vitamin A	-
	vitamin B1	vitamin B1
	vitamin B6	vitamin B6
	25-OH vitamin D	25-OH vitamin D
	vitamin B12	vitamin B12

A smaller part of the research is to combine these two data sets. Some challenges arise when doing so. Such a challenge is obviously to find the right connection between them, using the survey and lab data of the same patient. What can the markers say about the severity of co-morbidities?

A second challenge would be to define what to do with non-matching data. In the pre-treatment for example more markers were used than in the post-treatment. These missing ones might be more useful for scoring the co-morbidity severity than the known markers. There also might be measurements that were missing or corrupted, whereas other markers might still be useful enough for a result. How to tackle this missing data challenge should be defined properly.

1.2 Co-morbidity Marker Relation

For some of the markers it is clear they affect the severity score of a co-morbidity. Co-morbidity is diagnosed from this markers and therefore a relation is imminent. For others more subtle relations are present, that in first sight would not be recognised. Both are important for assigning a co-morbidity severity score.

All three co-morbidities can be derived from specific markers. T2DM is known for higher glucose and insulin levels in the blood, while hardly finding any C-peptides. Ideally these would be used to find T2DM, however these values vary significantly over the day. Therefore help from other markers should be helpful. Hypertension is best measured with blood pressure measurements (BP). These however are not reliable for the nervous patients in the hospital and other markers must be found for help. Thirdly dyslipidemia obviously can be found using markers in the lipid spectrum (Table 1.1), however there may be other markers contributing to dyslipidemia detection, as well.[1]

For T2DM, hypertension and to a lesser extend dyslipidemia, more knowledge on relations between them and markers would benefit the creation of a severity score. Literature research is done to understand more of those subtle relations and know that those exist when trying to create a morbidity score. The outcome of this literature search shows that quite some relations can be found between co-morbidities and the markers (Table 2). For most it is definitely worth to investigate any relation.[1]

2 First Proposal

To start of a research proposal, the first thing that needs to be done is to create a research question that fits the intended goal. This main question can then divided in several sub questions for explaining the steps how to answer it. Next hypotheses should be made for the main and sub questions. Thirdly the methods of how to answer the research questions are discussed.

2.1 Research Question

As stated earlier (Section 1) the goal was to develop a severity score for the co-morbidities using data mining. A research question for this goal would then be:

Can a severity score for co-morbidities present in bariatric patients based on biomarkers be developed?

This main research question includes all aspects of the goal. It can however be specified in several different sub-questions. A first sub-question would be about the relation between the co-morbidities and the biomarkers, which is a major part on making this score.

Which relations are present between the co-morbidities and the biomarkers?

A second sub-question should be about retrieving the score for the co-morbidities from the biomarkers. This score retrieval should use the outcome of the first sub-question to efficiently find

Table 2: Relations found in literature between the co-morbidities and markers.[1]. The number shows how many citations were found that state a relation exist, nothing means no relation. A 'D' means that the disease is diagnosed from those markers.

	Diabetes	Hypertension	Dyslipidemia
hemoglobin	3	1	
hematocrit	2	2	
erythrocytes	1	1	
MCH		1	
MCV		1	
thrombocytes			
leukocytes	4	3	1
bilirubin	3	1	1
ASAT	1	1	1
ALAT	4	1	1
LD	1		
Alkaline phosphatae	2		
gamma GT	4	2	1
urea	1	3	
creatinine		2	1
potassium		2	
sodium		1	
calcium	2	1	1
phosphate			
albumin			
CRP	3	4	1
cholesterol		2	D
HDL-cholesterol		2	D
chol/HDL ratio			D
LDL-cholesterol		2	D
triglycerides		2	D
prothrombin time			
hemoglobin A1c (IFCC)	D		
glucose	D	2	
parathormone	1	2	
iron			
ferritin	3	1	1

the correct score.

How can a severity score of co-morbidities be obtained from biomarkers?

This second questions mainly focuses on techniques to achieve a sufficient answer to the main research question. When both sub-questions are sufficiently answered, then the main one should definitely be, as well.

2.2 Hypothesis

The data set available for this research is quite extensive. The number of biomarkers is high, many relations between them and the co-morbidities are known and a high number of patients has been tested for both. These pre-conditions are very beneficial to finding positive answers for the research questions.

Knowing many relations between the co-morbidities and biomarkers beforehand implies that a lot of them should be present in the data set. to the sub-question which relations are present, the answer would be two-fold. First strong relations between the co-morbidities and on the other hand biomarkers that help diagnosing them. This would for example be the case for dyslipidemia and biomarkers of the lipid spectrum. Weaker relations are expected for all found in literature (Table 2).

As for the question how a score can be defined from biomarkers, several different options are possible. When knowing there is a relation between biomarkers and the co-morbidity, with machine learning combined with regression a function can be made predicting a score outcome for the biomarkers. This score is then compared with the categorical labelling of co-morbidities to find out if they match. This way, through training the model and afterwards testing if that model could efficiently give a good score to the co-morbidities.

With the hypotheses for both sub-questions the main research question can be answered, too. The hypothesis for this one would be that it is possible to find a severity score with the biomarkers. A model can made using the biomarkers as an input and retrieving a severity score as output.

2.3 Methods

Since the materials are all explained in Deneer's thesis, that and most of the pre-processing part is skipped in methods.

3 Proposal Comparisons

4 Final Proposal

References

- [1] R. Deneer, "Scoring co-morbidity severity in bariatric patients based on biomarkers: a data mining approach," 2017.