# Important parts Thesis Ruben Deneer

**Introduction**

It has become clear that bariatric surgery has a positive effect that extends beyond weight loss and includes improvement in diabetes, hypertension, dyslipidemia and reduction of overall mortality

there is no uniform definition of how patients are identified as having co-morbidities

to use data mining techniques to develop a score that can objectively quantify the severity of co-morbidities present in bariatric patients based on biomarkers, both before and after surgery

However, in this study the goal is not to predict the pre-operative probability of remission but to quantify the severity of co-morbidities and monitor improvement in a patient before and after surgery.

**Background**

There are different types of bariatric surgery but this study is limited to primary gastric sleeve (SG) and gastric bypass (RYGB) surgery

Initially SG was labeled as a restrictive procedure because the weight loss was attributed to a reduction in stomach size. However, it later became apparent that modifications of gastrointestinal hormones play a significant role.

In similar fashion to SG, RYGB also has a beneficial effect on co-morbidities.

To assess the metabolic health of patients before and after bariatric surgery, the Catharina Hospital in Eindhoven makes use of an extensive bariatric laboratory panel. This panel includes blood tests with markers related to complete blood count, kidney function, liver function, inflammation, lipid spectrum, coagulation, glucose metabolism, thyroid function, mineral and vitamin status

Both diabetes and dyslipidemia can be diagnosed from the results of the markers contained in the glucose metabolism and lipid spectrum sub-panels, respectively.

From literature, evidence was found to conclude that hypertension can be observed from changes in several biomarkers (i.e. leukocytes, CRP, urea, GGT, etc.) contained in the bariatric lab panel.

**Material and Methods**

All data of

patients that underwent primary gastric sleeve or bypass surgery (no revision cases) at

the Catharina Hospital in Eindhoven, in the period of 14-02-007 to 24-02-2016 were

extracted from multiple databases.

The DATO database is a national database that contains

registrations regarding the pre- and post-treatment health status of patients that underwent

bariatric surgery in the Netherlands.

On 01-10-2011 the Catharina Hospital in Eindhoven

introduced a bariatric laboratory panel containing a set of blood tests with (bio)markers

(see section 2.2 for more details). Pre-operatively the panel consists of 47 markers and a

subset of 42 markers is measured during follow-up (6, 12, 24, 36 months after surgery).

Allowed 90 days as a maximum time window between a lab set and

DATO entry, pairs that had more than 90 days between the lab and

DATO date were dropped.3 – Kind of strange this was done

Removed patients that did not have a screening lab-DATO-pair. This pair had to

be present because patients had to be compared to their pre-operative (baseline)

status, both in terms of lab measurements and co-morbidity status. – Why? If it was just to find out a score

Although there are follow-ups at 36, 48 and 60 months, only the 6, 12 and 24 month

follow-ups are included because not enough data was available of the 36- and 48-month

follow-up (see figure 3.2). – Same question

At follow-up a patient can have one of six labels: "Cured",

"Improved", "Same", "Worse", "Denovo" or "Not present" as described in section 3.1.3.

If the screening label would be missing and the next label would be "Same" it would

be impossible to determine whether the co-morbidity is still present or never was. – Shouldn’t it be Not present then?

A straightforward approach to solving a multilabel

classification problem is the binary-relevance (BR) method [45]. In this method

a classifier is trained independently for each label and given, a new unseen patient,

each classifier predicts whether the respective co-morbidity is present.

Project singularities

Data omission

During preprocessing, numerous biomarkers and patients are omitted. Some of these omissions were correctly justified, however some were done a bit too rigorously.

* Removing data sets for which lab set and DATO set were too far apart. While the lab set is extensive and much information, the DATO set is small and has many binary values which could be estimated easily.
* Removing double pairs for predefined measuring moments. These are good data sets that should definitely be used.
* Removing pairs not on the predefined measuring moments. These still can be used for the tests and for regression.
* The markers that are only available in pre-surgery measurements. These might be very useful in regression.
* Even though LDL-cholesterol might be wrong sometimes as a marker, it still shows a very specific way of combining all cholesterol markers.
* DATO registrations missing can be estimated quite easily.

Scoring system

* Ordinal is not really an answer to the original goal it seems. It is not that different from the original idea that checks out every co-morbidity separately. So why only limit yourself to 0, 1 or multiple?
* I understand the idea of combining all three co-morbidities to one score. However it seems more logical to create three scales for every co-morbidity and add those together. Even though they have overlap, they still have their own indicators most likely