



PROJECT PLAN

ENDOMINE

DIABETIC TREATMENT KNOWLEDGE EXTRACTION FROM BLOOD TEST DATA OF JGH ENDOCRINOLOGY DEPARTMENT

Projet #

Version 0.1

Auteur : Fodil Belghait - ETS

Date of issue: Mai 7th 2013

Review Date:



Table of Content

1. Introduction	3
2. Project Background	3
3. Project Scope	3
4. Objectives, Approach and Output	3
4.1 Metformine and lactic acidosis incidence	3
4.1.1 Objective	3
4.1.2 Approach	4
4.2 Diabetic patients' clustering	4
4.2.1 Objective	4
4.2.2 Approach	5
4.3 Data mining predictive model	5
4.3.1 Objective	5
4.3.2 Approach	5
4.4 Association rules between metformin and observed side effects	6
4.4.1 Objective	6
4.4.2 Approach	6
5. Project Outputs	7
6. Project Outcomes	7
7. Stakeholder Analysis	7
8. Risk Analysis	8
9. Technical Development	9
10. Intellectual Property Rights	9
11. Project Management	9
11.1 Project Working Committee	9
11.2 Project Team	10
12. Budget	10
13. Project Work Breakdown structure and timelines	10
14. Project progress follow up	3
15. Appendix A : Project Status Report	4
16. Appendix B: Issue register	6

1. Introduction

The purpose of this document is to present the scope and the objectives of the Endomine research project, the required steps to successfully execute and control the project, the communication plan that ensures the information distribution among the project stakeholder and to document the project scheduled baselines. This is a living document during the project period. Included information is expected to be reviewed as more information about the project becomes available.

2. Project Background

The Metformin is a biguanide that has been used to treat the type 2 diabetes (T2D) for more than 40 years. Phenformin hydrochloride is an earlier biguanide that was used to treat T2D and it has been withdrawn from the market because it was associated with a reported case of lactic acidosis especially for patients with renal failure. The Metformin has been also associated with the increased risk of lactic acidosis because of its association with the phenoformin. The options other than metformin are expensive and with other type of risk such as substantial risk of low glucose.

The true incidence of metformin associated with lactic acidosis is not known. The Endomine project addresses the relationship that exists between the treatment of the T2D with metformin and the lactic acidosis risk.

3. Project Scope

The scope of this project is to apply the data mining techniques on the blood test results in order to identify any existing patterns between the usage of metformin/Glucofage on type 2 diabetes and the increased risk of lactic acidosis especially for patients with severe acute condition, such as renal failure and identify groups of patients with renal failure problems and have not developed any side effect when treated with metformin/Glucofage.

4. Objectives, Approach and Output

4.1 Metformine and lactic acidosis incidence

4.1.1 Objective

The main objective is to perform a statistic analysis of the blood tests' data in order to identify evidence either for the existence or not of direct incidence of metformin and the increased risk of lactic acidosis for the type 2 diabetics in general and for those with renal failure problems in particular.

This objective intends to provide statistic evidence on the direct incidence of the usage of the metformin/Glucophage on the T2D patients that have renal problems and the lactic acidosis.

4.1.2 Approach

- **Step 1:** Gather the data of all diabetic patients without consideration to the diabetes type nor to the problem of the renal failure
- **Step 2:** Establish the criteria to use in the data cleanup (time the test has been taken, the machines used to take the tests, etc. then clean the data from incomplete records, inconsistent data, etc.). This step is very important, because it determines the quality of the data to be used in the analysis and it will have a direct impact on the quality of the obtained result
- **Step 3:** Identify the criteria and the lab tests measures to use in order to identify the patients' categories. Diabetic versus non diabetic, how to identify the patients treated with metformin versus other medications, what element to consider in order to confirm the renal failure problem in the blood tests, and what test result and measure to consider in order to identify the patients with lactic acidosis.
- **Step 4:** Proceed to a statistics analysis of the data in order to identify the percentages of each category of patients:
 - Diabetic patients
 - Diabetic patients treated with metformin
 - Diabetic patients treated with other medication than metformin
 - Diabetic patients with lactic acidosis
 - Diabetic patients with renal failure and lactic acidosis
 - Diabetic type 2 with metformin , renal failure and lactic acidosis
 - Etc...
- **Step 5 :** **Calculate the correlation coefficient between the following two variables:**
 - Diabetic treated with Metformin
 - Diabetic treated with metformin and having the lactic acidosis and renal problem
- **Step 6 :** Analyze the obtained results

4.2 Diabetic patients' clustering

4.2.1 Objective

Identify groups of Type 2 Diabetic patients treated with metformin/Glucophage and share the same type of response to the treatment.

This objective aims to classify the T2D patients by their type of response to the treatment; this will help the physicians to have an idea on what treatment to give to their patients based on their profiles.

4.2.2 Approach

Step 1: Data sampling and preparation: In this step we isolate all the data of the T2D patients with disregards to the medication they take and to other problems they may have. The quality of the data will be revised if the sample includes other data than the one that have been used in the first experience.

Step 2: Data clustering model design: In this step we will establish the criteria and the data attributes to use to build the data clusters such as: the medication we want to consider in this analysis, the blood tests attribute to use to measure the patients response to the medication, the patients information we want to consider in the patient profile, etc. We will also identify the data clustering technique to use in the analysis.

Step 3: Data mining: Design, develop and execute the defined data clustering model

Step 4: Results validation: The results obtained in the step of data mining will be validated against any available published results and with the project stakeholders. The data clustering model will be refined and re-executed based on the validation results.

Step 5: results interpretation: The final validated results will be presented to the project stakeholder for interpretation and comparison with any published result for similar studies.

4.3 Data mining predictive model

4.3.1 Objective

Develop a predictive data mining tool that allows predicting the response of the type 2 diabetic patient to the treatment by the metformin

4.3.2 Approach

- **Step 1: Data Sampling and preparation:** The objective of this step is to identify the data that will be used to build the model (training data) and the one that will be used to validate the model (testing data). The sampled data will be cleaned and transformed for the purpose of the data clusters' building.
- **Step 2: Design and building the prediction model:** At this step we will choose the appropriate data mining predictive technique to use select three different algorithms and apply them to the set of data selected in the data mining sampling and preparation step. We will compare the quality of the results and the performance of each of them in order to select one to retain for our experience. We will also select a tool to implement the built prediction model.

- **Step 3: Results analysis and validation:** The result produced by the built model will be presented to the physician for validation. The feedback of this step will be used to refine the design of the data mining model. The output of this step will be the final data mining predictive model that can be used by physicians.
- **Step 4: Results interpretation:** The results generated by the last version of the data mining model will be presented to the physician for validation and interpretation.

4.4 Association rules between metformin and observed side effects

4.4.1 Objective

Identify the potential harmful interaction between the treatment of T2D by the metformin and the usage of other medication. This objective aims to identify the possible combinations between: T2D, metformin, other medication taken conjointly with metformin, and observed side effects.

4.4.2 Approach

- **Step 1: Data preparation and sampling:** The purpose of this step is to identify, prepare and transform the set of data to be used to apply the data mining model in order to extract the potential correlation that may exist between the data.
- **Step 2: Association rule extraction model design and implementation:** In this step, we will analyze the data in order to choose the appropriate data mining technique that will be most suitable for the context of blood test results. This step intends to identify the list of the attributes we want to consider in the research of the existing association rules, the validation rules to use in order to filter the generated correlations.
- **Step 3 : Result evaluation :** In this step, considering the rules validation criteria identified in the previous step, we will evaluate the quality and the relevance of the association rules generated, this is a recursive step, that means that we will tweak and refine the attribute, generate the association rules with the data mining model and validate results with the physicians, the process will continue until we arrive to a set of associated rules evaluated as good quality and relevant to the physicians.
- **Step 4: Result interpretation and conclusion:** This step consists in interpreting and giving a meaning to the correlation generated by the model. The objective is to identify the dangerous combination between some medication, metformin on the diabetic patients.

5. Project Outputs

Key project output will include:

- Statistical analysis reports that will show the responsibility degree of metformin in the lactic acidosis for the T2D that are affected with renal failure.
- List of categories of type 2 diabetic patients classified by their response to metformin treatment. This list includes a set of information that may explain the T2D patient response to metformin treatment.
- A tool that helps predict the patients response to the metformin treatment.
- The list of dangerous correlation between the usages of analyzed medication, metformin and side effects.

6. Project Outcomes

Key project outcome will include

- A set of tool kit to the physicians that can help them identify the appropriate treatment for the T2D patients.
- An improved understanding of the incidence of the metformin in the lactic acidosis.

7. Stakeholder Analysis

Stakeholder	Function	Role	Eff.ort per month	Contribution	Contact
Dr. Mark Trifiro	Chief of the Division of Endocrinology (JGH)	Project Sponsor	8 hours	Medium	mark.trifiro@mcgill.ca
Dr. Elizabeth Mac Namara	SMBD-Jewish General Hospital, Program Director	Data owner and Project Partner	4 hours	Medium	elizabeth.macnamara@mcgill.ca
Shaun Selcer	Physician at McGill University Health Centre	Medical knowledge support	16 hours	High	shaun.selcer@mail.mcgill.ca
Dr. Shaun Eintracht	Medical Biochemist, Dept. of Diagnostic Medicine	Data access and understanding support	8 hours	High	seintracht@jgh.mcgill.ca
Dr. Alain April	Software Engineering Professor at ETS	1. Manage ETS student 2. Project Manager	8 hours	Medium	alain.april@etsmtl.ca



DR. Christian Desrosier	Software Engineering Professor at ETS	Data mining expert	8 hours	Medium	christian.desrosiers@etsmtl.ca
Fodil Belghait	PHD Student	Main project resource	32 hours	High	fbelghait@hotmail.com
Information systems services staff		Remote access to the data	8 hours	High	

8. Risk Analysis

Risk	Probability (1-5)	Severity (1-5)	Score (P x S)	Action to Prevent/Manage Risk
Availability of the medical resources to help identify the requirements	3	5	15	Prepare a plan of meetings in advance with the identified resources
The main resource working on this project is a part time student and is working full time. The limited access to the data may have a major impact on the progress of the project	5	5	25	Implement a secure remote access to the data
Limited understanding of the medical terms of concepts by the main resource	3	2	6	Identify support resources accessible with email contact in order to provide medical terms understanding support when required
Any constraint related to the access to the data	3	5	15	Identify the required data to access and the responsible person of this data as first step of the project then define and implement a cleared process to access this data without constraint
Key stakeholders do not buy in to/support the project	3	3	9	<ul style="list-style-type: none"> In early stage of the project gather key project stakeholders in project steering group and seek commitment to project Ensure regular project status communication to all stakeholders, Solicit opinions and feedback on project direction frequently. Develop a communication plan for the project.
Stakeholders' expectations higher than what can be delivered.	2	2	4	<ul style="list-style-type: none"> Clearly define the project scope and outputs Communicate and prioritize the user requirements to the project committee

9. Technical Development

In this project we intend to use the waterfall SDLC model enriched with rapid prototyping aspect. During the course of the project life cycle, we will review the development process with the project working group whenever required.

The following are the main steps of this model:

- Gather Requirements
- Plan and design solutions
- Develop and Implement solution
- Test, review and validate solution

The approaches described above are inspired from the waterfall SDLC model.

Each phase will interact with the other phases, and visual mock-ups and usability check points will be used to ensure transparent and effective delivery. Maintaining open communication channels during the development will ensure that a usable solution is realized.

Technologies used in the project will include most likely the following: Oracle store procs, Java, XML, JSP, and J2EE technologies, and open source data mining tools.

10. Intellectual Property Rights

- **Term 1:** ETS resources will access to depersonalized data. All the data remain the exclusive property of the JGH
- **Term 2:** The produced knowledge from the project effort remain a shared property between the ETS resources and the JGH resources
- **Term 3:** Any source code generated in the project is a shared property between the ETS resources and the JGH resources Project Resources

11. Project Management

11.1 Project Working Committee

The Project Steering Committee provides a formal approval to the main project documents, results and any change requests. It ensures that the project outcomes are broadly in accordance with the project scope.

The Project Steering Committee is composed with:

- Dr. Mark Trifiro / JGH
- Dr. Elizabeth Mac Namara / JGH
- Dr Alain April / ETS
- Dr Christian Desrosier / ETS

11.2 Project Team

The Project Team's role is to co-ordinate the project participant effort during the whole project life, escalate any project showstopper issues to the project steering committee and deliver the work packages.

The Project Team is composed of:

Stakeholder	Function	Role	Eff.ort per month	Contribution	Contact
Shaun Selcer	Physician at McGill University Health Centre	Medical knowledge support	16 hours	High	shaun.selcer@mail.mcgill.ca
Dr. Shaun Eintracht	Medical Biochemist, Dept. of Diagnostic Medicine	Data access and data understanding support	8 hours	High	seintracht@jgh.mcgill.ca
Fodil Belghait	PhD Student	<ul style="list-style-type: none"> • R&D • Developer 	32 hours	High	fbelghait@hotmail.com
Information systems services staff		Remote access to the data	8 hours	High	

12. Budget

This project is done in the context of a PhD research thesis at ETS, no budget is associated to this project

13. Project Work Breakdown structure and timelines

In this section we will present the tasks, the associated deliverables and resources responsibilities.