

# i2b2 to Optimize Patients Enrollment

Mauro BUCALO<sup>a,1</sup>, Matteo GABETTA<sup>a</sup>, Lorenzo CHIUDINELLI<sup>b</sup>, Cristiana LARIZZA<sup>c</sup>, Antonio BELLASI<sup>b</sup>, Alberto ZAMBELLI<sup>b</sup> and Nicola BARBARINI<sup>b</sup>

<sup>a</sup>BIOMERIS (BIOMedical Research Informatics Solutions), Pavia, Italy

<sup>b</sup>Hospital "ASST Papa Giovanni XXIII", Bergamo, Italy

<sup>c</sup>Laboratory for Biomedical Informatics, University of Pavia, Italy

**Abstract.** i2b2 data-warehouse could be a useful tool to support the enrollment phase of clinical studies. The aim of this work is to evaluate its performance on two clinical trials. We developed also an i2b2 extension to help in suggesting eligible patients for a study. The work showed good results in terms of ability to implement inclusion/exclusion criteria, but also in terms of identified patients actually enrolled and high number of patients suggested as potentially enrollable.

**Keywords.** patients enrollment, patients recruitment, clinical data-warehouse

## 1. Introduction

Patients enrollment is a critic phase for a clinical study: complexities of inclusion/exclusion criteria, patient data not easily accessible, volume and heterogeneity of data. Due to these factors, the enrollment phase often leads to delays in the clinical trial, also linked to costs increase.

i2b2 (Informatics for Integrating Biology and the Bedside) is an open-source software aimed at creating data warehouses to support clinical research [1]. Given that the i2b2 web interface ("query tool") allows to easily select cohorts of patients with certain inclusion/exclusion criteria, i2b2 can be a very useful tool for predicting potential patient accrual but also to support the enrollment phase [2].

The objective of present work is the evaluation of i2b2 in supporting the patient enrollment phase of a study. To this aim we developed also a preliminary version of a specific i2b2 extension to suggest eligible patients for a study.

## 2. Methods

The proposed approach starts from an i2b2 project including all patient data of a center with also Natural Language Processing procedures in place for the data extraction from clinical reports. The approach consists in translating the inclusion / exclusion criteria of a study into i2b2 queries. We defined as "query block" a portion of a criterion that can be directly implementable with a query through the i2b2 Query Tool. So an inclusion/exclusion criterion is a complex logical combination of many "query blocks". We developed an i2b2 extension, consisting of a cell and a plugin, which allows to

---

<sup>1</sup> Corresponding Author, Mauro BUCALO; E-mail: mauro.bucalo@biomeris.it

create a study and associate a list of inclusion and exclusion criteria with it. It is also possible to associate “query blocks” defined by the i2b2 Query Tool to each criterion.

The extension allows to associate each patient-criterion pair with an eligibility score that represents the measure of how much the patient meets the criterion. The scores relating to all the criteria contribute to determine the patient's eligibility score for the entire study. The eligibility score can be calculated by attributing different weights to the different met criteria. Given that each patient is assigned an eligibility score, it is necessary to consider a threshold of this score to classify a patient as "eligible".

Finally, the plugin can show the list of patients who best meet the combination of the inclusion and exclusion criteria of a single study.

### 3. Results and Discussion

The described approach was tested using the i2b2 system installed at the ASST Papa Giovanni XXIII (HPG23) which includes both an i2b2 project with data from the entire hospital and an i2b2 oncology project for oncology patients. The approach was evaluated on two breast cancer studies involving the center: (i) COMPLEMENT-1 enrolling 14 patients during 2017; (ii) BioItaLEE enrolling 13 during 2018.

A first result is that i2b2 allowed to implement 18 out of 20 criteria for COMPLEMENT-1 study and 21 out of 22 criteria for BioItaLEE study.

In this work, a simplified eligibility score was implemented by attributing a weight of -1 to the exclusion criteria and +1 to the exclusion criteria. According to the clinician involved in the project, a weight of 0.7 was attributed to a criterion relating to the advanced stage of the disease, due to many missing data. Given a score threshold of 85% to define a patient as “eligible”, the proposed approach suggested 11 of the 14 patients enrolled in COMPLEMENT-1 and 12 of the 13 enrolled in BioItaLEE.

Finally, considering the same score threshold, the tool suggested as eligible for each study respectively more than 190 and 130 patients.

### 4. Conclusions

This work describes the use of i2b2 with an extension developed specifically to help support the patient enrollment phase. The preliminary results shown are certainly promising. The system is able to implement almost all the criteria of the tested studies. Furthermore, although obtained on only two studies, we demonstrated the ability of the proposed approach based on i2b2 system to suggest patients who could actually be enrolled. In conclusion, the high number of patients suggested as potentially enrollable shows that the proposed approach could help speed up the patient enrollment phase.

*Acknowledgements:* This preliminary work was funded by Novartis Farma s.p.a.

### References

- [1] Kohane IS, Churchill SE, Murphy SN. A translational engine at the national scale: informatics for integrating biology and the bedside. *J Am Med Inform Assoc.* 2012;19(2):181-5
- [2] London JW, Balestrucci L, Chatterjee D, Zhan T. Design-phase prediction of potential cancer clinical trial accrual success using a research data mart. *J Am Med Inform Assoc.* 2013 Dec;20(e2):e260-6