



Comprehensive Research Experience for Medical Students
Alexandra Yeo Summer Studentship Research Program 2019

Supervisor/Project Information Form

Due February 20 2019 by email to crems.programs@utoronto.ca

Supervisor Name: Dr. Lisa Hicks

Project Title: Deriving structured clinical data from free text notes in patients with JAK2 V617F+ polycythemia rubra vera, essential thrombocytosis and primary myelofibrosis

Hospital/Research Institution: St. Michael's Hospital/ Li Ka Shing Knowledge Institute

Email: HicksL@smh.ca or Petruccij@smh.ca

Field of Research (2 keywords): Quality improvement, hematology

Department: Hematology/Oncology

School of Graduate Studies Appointment (IMS, LMP, IHPME etc)? Yes/No: If YES, please name:

Project Title: Deriving structured clinical data from free text notes in patients with JAK2 V617F+ polycythemia rubra vera, essential thrombocytosis and primary myelofibrosis

Brief Project Description (< 300 words):

Myeloproliferative neoplasms are hematopoietic stem cell-derived clonal disorders characterized by the proliferation of one or more myeloid cell lineages leading to disease entities such as polycythemia rubra vera, essential thrombocytosis and primary myelofibrosis. These MPN disorders are often characterized by the JAK2 V617F tyrosine kinase mutation and share clinical features such as increased risk of thrombotic events and progression to acute myeloid leukemia. As our population ages, the number of patients diagnosed with JAK2 V617F mutated MPNs has grown and there is an increased need to have a centralized database to aid in physician-led research to improve the quality of care for patients with MPNs in Ontario.

We are conducting a pilot study to evaluate the accuracy, precision, and efficiency of natural language processing (NLP) technology to translate unstructured electronic free text data from physician clinic notes at St. Michael's Hospital into structured data that can be encoded and

inputted into a database. The ultimate aim is to use NLP technology to include other hospital institutions in Ontario and have a centralized database.

Evaluation of NLP technology will occur in two outpatient malignant hematology clinics at St. Michael's Hospital.

Indexed output from the NLP technology will be evaluated against the 'gold standard' of manual chart review and data extraction. Measurements of test accuracy will be performed by testing the sensitivity and specificity of each tool. Reliability of the NLP tool will be compared to manual chart review using Cohen's kappa statistical methods. Sub-analysis will be conducted to evaluate whether the NLP tool performs better for some data types than others. Work logs from those performing data extractions through NLP technology and manual chart review will be kept in order to compare the efficiency of both tools.