Principal Investigator

First Name: Peter
Last Name: Higgins
Degree: Ph.D., M.D., M.Sc.
Primary Affiliation: University of Michigan
E-mail: higginsibdteam@gmail.com
Phone number: 734-647-2564
Address: 6510 MSRB 1
1150 w. Medical Center Dr.
City: Ann Arbor
State or Province: Mi
Zip or Postal Code: 48109
Country: USA

General Information

Key Personnel (in addition to PI):

First Name: Kay
Last name: Sauder
Degree: B.S.
Primary Affiliation: University of Michigan
SCOPUS ID:

First Name: Akbar
Last name: Waljee
Degree: M.D., M.Sc.
Primary Affiliation: University of Michigan
SCOPUS ID:

First Name: Yiwei
Last name: Zhang
Degree: M.S.
Primary Affiliation: University of Michigan
SCOPUS ID:

First Name: Ji
Last name: Zhu
Degree: Ph.D.
Primary Affiliation: University of Michigan
SCOPUS ID:

First Name: Boang
Last name: Liu
Degree: BS
Primary Affiliation: University of Michigan
SCOPUS ID:

First Name: Zachary
Last name: Bearinger
Degree:
Primary Affiliation: University of Michigan
SCOPUS ID:

Are external grants or funds being used to support this research?: External grants or funds are being used to support this research.
Project Funding Source: NIH R01 GM097117

Conflict of Interest

http://yoda.yale.edu/system/files/skmbt_22314121710230.pdf
http://yoda.yale.edu/system/files/yoda_project_coi_form_boang_liu.pdf
http://yoda.yale.edu/system/files/yoda_project_coi_form_bearinger.pdf

Certification

Certification: All information is complete; I (PI) am responsible for the research; data will not be used to support litigious/commercial aims.
Data Use Agreement Training: As the Principal Investigator of this study, I certify that I have completed the YODA Project Data Use Agreement Training
Associated Trial(s):

1. NCT00094458 - Multicenter, Randomized, Double-Blind, Active Controlled Trial Comparing REMICADE® (infliximab) and REMICADE plus Azathioprine to Azathioprine in the Treatment of Patients with Crohn's Disease Naive to both Immunomodulators and Biologic

What type of data are you looking for?: Individual Participant-Level Data, which includes Full CSR and all supporting documentation

Research Proposal

Project Title

Can Machine Learning Algorithms using General Labs Predict Biologic Remission for Patients on Thiopurines in the SONIC trial?

Narrative Summary:

We have applied machine learning to common lab values to predict biologic remission in patients with inflammatory bowel disease in our local cohort. Now, we want to externally validate this by applying our algorithm to general labs from week 6 and week 10 in SONIC to predict which patients will achieve Biologic Remission at week 26 and at week 50.

Scientific Abstract:

Background:
Machine learning applied to common lab results, including the CBC with differential and the comprehensive chemistry panel (plus age in days at time of blood draw), can predict biologic remission in patients with inflammatory bowel disease (ThioMon 2.0).
Objective:
Our primary objective is to determine if the ThioMon algorithm can predict which patients will reach biologic remission as determined by CRP and endoscopy scores by week 26.
Study Design:
For each subject, we will determine immunosuppression scores from week 6 and week 10 lab values. We will see if these scores can accurately predict which patients reached either biologic remission or clinical remission at week 26 and week 50.
Participants:
Subjects randomized in the SONIC trial who received thiopurines.

Main Outcome Measure(s):
1) The ability of ThioMon to predict biologic remission (based on CRP and mucosal healing) at week 26.
2) The ability of ThioMon to predict biologic remission (based on CRP) at week 50.
3) The ability of ThioMon to predict clinical remission (based on CDAI) at weeks 26 and 50.
Statistical Analysis:
We will use a Student's T test or Mann-Whitney U test, to compare the following:
1.) Immunosuppression scores between biologic remission (BR) and non-BR groups at week 26 and week 50
2.) Immunosuppression scores between clinical remission (CR) and non-CR groups at week 26 and week 50
We will also report the AuROC, sensitivity, specificity, NPV, and PPV for each of the comparisons above, using an Immunosuppression score of 100 as the cut point.

Brief Project Background and Statement of Project Significance:

Machine Learning Methods
Machine learning is a group of methods that optimize splits in datasets to predict important outcomes(1). Applications of machine leaning have improved the analysis of gene microarrays(2), proteomics results from mass spectrophotometry(3), predictions in financial markets(4), and algorithms to optimize signal and reduce noise in images(5). Machine learning is often applied by businesses to identify customers for a product based on their purchasing history, as in Amazon.com recommendations for books, or Google optimized searches. More recently, machine learning has been applied to clinical problems in which large complex datasets are available.
Machine Learning Predicts Thiopurine Treatment Success in IBD Patients
Numerous patients with inflammatory bowel disease (IBD) require treatment with thiopurines. Physicians often monitor the efficacy and safety of this low cost medication by following blood counts and blood chemistry. Alternatively, a commercially available test measures the amount of 6TGN metabolite present in a patient's blood. Dr. Higgins' research group used the machine learning approach to build an algorithm that predicted a patient's response to thiopurine treatment (6). This algorithm was optimized to predict clinical response.
In this study of ThioMon 1.0, the machine learning algorithm based on blood metabolites was 86% accurate in predicting clinical response to thiopurine treatment, while commercially available blood metabolite measurements were only 59% accurate.
Since clinical symptoms can be subjective and not always indicative of inflammation present, the Machine learning algorithm has been improved by using objective biological evidence of inflammation to determine patient response to thiopurine treatment.
The ThioMon 2.0 algorithm (for biologic response) was developed using 3,269 patient cases, and is significantly more accurate than the 6-TGN metabolites for predicting biologic response to thiopurines. This algorithm is significantly more accurate ($P<0.0001$) than metabolites in predicting biologic remission. These algorithms have been patented by the Regents of the University of Michigan.
Machine Learning to Predict Clinical Remission for IBD Patients Treated with Thiopurines
We would like to use our ThioMon 2.0 algorithmic model to predict which patients will achieve Biologic Remission at week 26 and week 50. This ability to predict the efficacy of thiopurine treatment in patients with IBD will transform that way in which this therapeutic class is monitored. Thiopurines cost roughly $700/year, much less than to biologic therapies like infliximab, which costs over $25,000 per year (7). Being able to identify patients who respond to thiopurines and optimize their therapy could significantly reduce costs to patients, insurers, and society for IBD care. For patients who benefit from thiopurine optimization, achieving biologic remission is associated with fewer steroid prescriptions, hospitalizations, and surgeries.

Specific Aims of the Project:

Our primary objective is to determine if the ThioMon 2.0 algorithm can use lab values from week 10 to predict which patients will reach biologic remission by week 26.

Our secondary objectives are to determine if the ThioMon 2.0 algorithm can use lab values from week 6 and also week 10 to predict which patients will reach biologic remission by week 50 or reach clinical remission at weeks 26 and 50. Also, to determine if we can correlate subjects' Immunosuppression Scores to their IFX (infliximab) levels and HACA (human anti-chimeric antibody) levels for subjects in cohort 3.

We hypothesize that our ThioMon 2.0 predictive algorithm will accurately predict which patients achieve Biologic Remission at week 26 based on lab values from week 10.

What is the purpose of the analysis being proposed? Please select all that apply.

New research question to examine treatment effectiveness on secondary endpoints and/or within subgroup populationsConfirm or validate previously conducted research on treatment effectiveness

Research Methods

Data Source and Inclusion/Exclusion Criteria to be used to define the patient sample for your study:

We would like subjects randomized into cohort 1 (azathioprine only) and cohort 3 (combination therapy with both infliximab and azathioprine). We will look at these aims in both cohorts. We expect that could be a larger effect in cohort 1, as any effect in cohort 3 could be confounded by the efficacy of infliximab. We will exclude participants who did not have active ulceration or an elevated CRP at baseline.

We will obtain the following data:
1.) A complete blood count with differential and platelet count (CBCPD) and a comprehensive metabolic panel (COMP CHEM) from week 6 and week 10 for each subject.
2.) IFX (infliximab) levels and HACA (human anti-chimeric antibody) levels from each subject in cohort 3.
3.) Age (in days) of each subject at week 6.
4.) Age (in days) of each subject at week 10.
5.) CRP results from baseline, week 26, and week 50 for each subject
6.) Endoscopy scores (or reports) from baseline and week 26 for each subject (Biologic remission [BR] = normal CRP and no ulceration on endoscopy)
7.) Clinical Remission status (including CDAI) for each subject at week 26 and week 50
8.) Gender, race, and medication doses for each subject

Main Outcome Measure and how it will be categorized/defined for your study:

Biologic Remission at week 26 will be our main outcome. It is defined as a CRP <0.5 mg/dL and/or the absence of ulceration on endoscopy.

Our secondary outcomes will be :
Biological Remission at week 50, Clinical Remission (defined by CDAI) at week 26, and Clinical Remission at week 50.

Main Predictor/Independent Variable and how it will be categorized/defined for your study:

We will calculate an "Immunosuppression Score" for each subject using the following data:

1.) A complete blood count with differential and platelet count (CBCPD) from week 6 and week 10.
2.) A comprehensive metabolic panel (COMP CHEM) from week 6 and week 10.
3.) Age (in days) of each subject at week 6.
4.) Age (in days) of each subject at week 10.

Other Variables of Interest that will be used in your analysis and how they will be categorized/defined for your study:

We will also look at IFX (infliximab) levels and HACA (human anti-chimeric antibody) levels from each subject in cohort 3 to see if we can correlate these levels to each subjects' "Immunosuppression Score."

Statistical Analysis Plan:

We will use a student’s T test or Mann-Whitney U test, as appropriate to compare the following:
1.) Immunosuppression scores between biologic remission (BR) and non-BR groups at week 26 and week 50
2.) Immunosuppression scores between clinical remission (CR) and non-CR groups at week 26 and week 50
3.) Immunosuppression scores between subjects with an elevated IFX level and subjects without
4.) Immunosuppression scores between subjects with an elevated HACA level and subjects without
We will also report the AuROC, sensitivity, specificity, NPV, and PPV for each of the comparisons above, using an Immunosuppression score of 100 as the cut point.
We will also explore multivariate models using thiopurine dose, immunosuppression score, infliximab dose, age, gender, and other demographics to predict infliximab levels and HACA levels.

Project Timeline:

Project start date: January 1, 2015 (or when data received)
Analysis completion date: March 1, 2015
First manuscript draft: April 1, 2015
Date of expected manuscript submission: May 1, 2015
Date of results reported back to the YODA Project: May 1, 2015

Dissemination Plan:

Study manuscript, target audience: gastroenterologists.
Likely journals: NEJM, Gut, Gastroenterology, American Journal of Gastroenterology.
Oral presentations at UEGW 2015 and ECCO 2016, or we might make the late-breaking deadline for DDW 2015.

Bibliography:

1. Breiman L. Classification and regression trees. Belmont, Calif.: Wadsworth International Group; 1984

2. Zhu J, Hastie T. Classification of gene microarrays by penalized logistic regression. Biostatistics. 2004;5:427-443

3. Ulintz PJ, Zhu J, Qin ZS, et al. Improved classification of mass spectrometry database search results using newer machine learning approaches. Mol Cell Proteomics. 2006;5:497-509

4. Wang L, Zhu J. Financial Market Forecasting Using a Two-Step Kernel Learning Method for Support Vector Regression. Annals of Operations Research. 2008;174:103-120

5. Wang L, Zhu J. Image Denoising via Solution Paths. Annals of Operations Research. 2008;174:3-17

6. Waljee AK, Joyce JC, Wang S, et al. Algorithms outperform metabolite tests in predicting response of patients with inflammatory bowel disease to thiopurines. Clin Gastroenterol Hepatol. 2010;8:143-150

7. Bonafede MM, Gandra SR, Watson C, et al. Cost per treated patient for etanercept, adalimumab, and infliximab across adult indications: a claims analysis. Adv Ther. 2012;29:234–248