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

2016-1176

General Information

Key Personnel (in addition to PI): **First Name:** Akbar
Last name: Waljee
Degree: MD
Primary Affiliation: University of Michigan
SCOPUS ID:

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

First Name: Kay
Last name: Sauder
Degree: BS
Primary Affiliation: University of Michigan
SCOPUS ID:

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

Are external grants or funds being used to support this research?: No external grants or funds are being used to support this research.
How did you learn about the YODA Project?: Conference
higgins_yoda_coi.pdf

akbar_yoda_coi.pdf
boang_yoda_coi.pdf
kay_yoda_coi.pdf
yumu_coi.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): NCT01369329 - A Phase 3, Randomized, Double-blind, Placebo-controlled, Parallel-group, Multicenter Study to Evaluate the Safety and Efficacy of Ustekinumab Induction Therapy in Subjects With Moderately to Severely Active Crohn's Disease Who Have Failed
NCT01369342 - A Phase 3, Randomized, Double-blind, Placebo-controlled, Parallel-group, Multicenter Study to Evaluate the Safety and Efficacy of Ustekinumab Induction Therapy in Subjects With Moderately to Severely Active Crohn's Disease (UNITI-2)
NCT01369355 - A Phase 3, Randomized, Double-blind, Placebo-controlled, Parallel-group, Multicenter Study to Evaluate the Safety and Efficacy of Ustekinumab Maintenance Therapy in Subjects With Moderately to Severely Active Crohn's Disease
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 ustekinumab?

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 screening, week 0, and week 3 in UNITI 1 and UNITI 2 to predict which patients will achieve Biologic and/or clinical Remission at week 6, week 8, and IMUNITI week 44.

Scientific Abstract:
Background:
Machine learning applied to common lab results, including the CBC 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, fecal calprotectin, and CDAI scores at week 6, week 8 and IMUNITI week 44.
Study Design:
For each subject, we will determine an immunosuppression score for each subject from lab values taken at screening, week 0, and week 3 (in UNITI1, UNITI 2). We will see if these scores can accurately predict which patients reached either biologic remission or clinical remission at week 6, week 8, and IMUNITI week 44.
Participants:
Subjects randomized to ustekinumab in UNITI1 or UNITI2 and in IMUNITI.
Main Outcome Measure(s):
1) The ability of ThioMon to predict biologic remission (based on fecal calprotectin) at week 6 and IMUNITI week 44.
2) The ability of ThioMon to predict biologic remission (based on CRP) at weeks 6, 8, and IMUNITI week 44.

3) The ability of ThioMon to predict clinical remission (based on CDAI) at weeks 6, 8, and IMUNITI week 44.
Statistical Analysis:
We will use a T test or Mann-Whitney U test, to compare the following:
1.) Immunosuppression scores between biologic remission (BR) and non-BR groups at week 6, 8, and IMUNITI week 44.
2.) Immunosuppression scores between clinical remission (CR) and non-CR groups at week 6, 8, and IMUNITI week 44.

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 learning 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 Ustekinumab Treatment Success in IBD Patients
Numerous patients with inflammatory bowel disease (IBD) require treatment with ustekinumab. Physicians often monitor the efficacy and safety of this low cost medication by following blood counts and blood chemistry. Additionally, we will monitor ustekinumab serum concentration levels at weeks 3 and 6.
Dr. Higgins' research group previously 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 Ustekinumab
We would like to use our ThioMon 2.0 algorithmic model to predict which patients will achieve Biologic Remission at week 6, week 8, and IMUNITI week 44. This ability to predict the efficacy of ustekinumab treatment in patients with IBD will transform the way in which this therapeutic class is monitored (7). Being able to identify patients who respond to ustekinumab and optimize their therapy could significantly reduce costs to patients, insurers, and society for IBD care. For patients who benefit from 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 screening, week 0, and week 3 to predict which patients will reach biologic remission by weeks 6, 8, and IMUNITI week 44.
Our secondary objectives are to determine if the ThioMon 2.0 algorithm can use lab values from screening, week 0, and week 3 to predict which patients will reach clinical remission by weeks 6, 8, and IMUNITI week 44.
We hypothesize that our ThioMon 2.0 predictive algorithm will accurately predict which patients achieve Biologic Remission at weeks 6, 8, and 44 based on lab values from screening, week 0, and week 3.

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 populations

Research Methods

Data Source and Inclusion/Exclusion Criteria to be used to define the patient sample for your study:
We would like subjects randomized to ustekinumab in UNITI1 or UNITI2 and in IMUNITI (IMUNITI dataset is not

available in YODA list to select, however, we received confirmation that this dataset is available) . We are requesting this subject population because we want to build our machine learning model with data from subjects that were on active study drug.

We will obtain the following data:
1.) All labs from all study visits for each subject.
2.) Age (in days) of each subject at each study visit
5.) CRP results from all study visits for each subject
6.) Fecal caplrotectin results from all study visits for each subject.
7.) Clinical Remission status (including CDAI) for each subject at all study visits.
8.) Gender, race, and medication doses for each subject.
9.) Ustekinumab serum concentration levels at all study visits.

Main Outcome Measure and how it will be categorized/defined for your study:
Biologic Remission at week 6, week 8, and IMUNITI week 44 will be our main outcome. It is defined as a CRP <0.5 mg/dL and/or a fecal calprotectin <175.
(We may exclude subjects that already achieved this outcome at baseline).
Our secondary outcomes will be :
Clinical Remission (defined by CDAI) at weeks 6, 8, and IMUNITI week 44.

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 screening, week 0 and week 3.
2.) A comprehensive metabolic panel (COMP CHEM) from screening, week 0 and week 3.
3.) Age (in days) of each subject at screening, week 0, and week 3
4.) We may include other lab results that are available to enhance this model.

We will use the Random Forest Machine learning approach to calculate this score.

We will also calculate an immunosuppression score at week 6 and 8 using: age (in days), CBCPD, and COMP CHEM, to determine if these later time points are more accurate at predicting week 44 in the IMUNITI trial.

Other Variables of Interest that will be used in your analysis and how they will be categorized/defined for your study:
We will look at ustekinumab serum concentration levels from each subject at weeks 3, week 6, week 8, and IMUNITI week 44 to see if we can correlate these levels to each subject's "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 6, week 8, and IMUNITI week 44.
2.) Immunosuppression scores between clinical remission (CR) and non-CR groups at week 6, week 8, and IMUNITI week 44.
3.) Immunosuppression scores between subjects with an elevated ustekinumab serum concentration level and subjects without an elevated ustekinumab serum concentration level.

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 ustekinumab dose, immunosuppression score, infliximab dose, age, gender, and other demographics to predict ustekinumab serum concentration.

Project Timeline:
Project start date: February 1, 2017 (or when data received)
Analysis completion date: August 1, 2017
First manuscript draft: October 1, 2017
Date of expected manuscript submission: December 1, 2017
Date of results reported back to the YODA Project: February 1, 2018

Dissemination Plan:
Study manuscript, target audience: gastroenterologists.
Likely journals: NEJM, Gut, Gastroenterology, American Journal of Gastroenterology.
Oral presentations at DDW 2018.

Upon request, we can allow others to use this algorithm through our portal.

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

Supplementary Material: ustekinumab_data_research_proposal.docx
response_to_reviewers.docx