Principal Investigator

First Name: Sai Phanindra
Last Name: Venkatapurapu
Degree: PhD
Primary Affiliation: AstraZeneca
E-mail: saiphanindra.venkatapurapu@astrazeneca.com
State or Province: MD
Country: United States

General Information

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?: Internet Search

Conflict of Interest

https://yoda.yale.edu/wp-content/uploads/2025/04/YODA_COI_SPV.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

1. NCT03464136 - CNTO1275CRD3007 - A Phase 3b, Multicenter, Randomized, Blinded, Active-Controlled Study to Compare the Efficacy and Safety of Ustekinumab to That of Adalimumab in the Treatment of Biologic Naïve Subjects With Moderately-to-Severely Active Crohn's Disease
2. NCT01369355 - CNTO1275CRD3003 - 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

Validation of a hybrid mechanistic-ML model of Crohn's Disease to predict clinical and endoscopic outcomes in individual patients

Narrative Summary:

Crohn's disease (CD) is a chronic, relapsing inflammatory disease of the gastrointestinal (GI) tract with a prevalence of about 320 cases in 100,000 people in Europe and North America. Even though there are multiple approved therapies, long-term remission rates are low in Crohn's Disease patients with about 50% of patients requiring surgery within 10 years. Quantitative systems pharmacology models that provide mechanistic understanding of a disease are increasingly being used at various stages of drug development, for gaining insights into both safety and efficacy of a treatment. For efficacy predictions, QSP models that predict clinical scores, in addition to biomarkers, have applicability.

Scientific Abstract:

Background:
In Crohn's Disease, SES-CD (Simple Endoscopic Score for Crohn's Disease) is used to measure the endoscopic activity in patients by evaluating endoscopy images. However, due to the invasive nature of endoscopic evaluations, objective assessments of gut damage is infrequent. Non-invasive tools to predict changes in gut mucosal condition with treatment are needed.

Objective:
The study aims to create an algorithm to predict patient-specific clinical and endoscopic outcomes from a mathematical/QSP model of Crohn's disease using de-identified patient data.

Study Design:
This is a retrospective study where the deidentified patient data from the requested studies will be used to develop and validate a mechanistic-ML model of Crohn's disease. The primary outcome measures include SES-CD and CDAI scores over time. Secondary outcome measures include endoscopic and clinical response measures along with changes in biomarkers over time.

Participants:
Patients from the requested studies who have the primary predictor variables, SES-CD and/or CDAI score data along with the component scores will be included in the study.

Statistical Analysis:
Descriptive statistics will be used to summarize baseline characteristics as well as clinical/endoscopic outcomes. Dichotomous variables will be presented as percentages along with 95% CI. Continuous variables will be reported as means with SD or medians with interquartile ranges. The machine learning algorithm's accuracy to predict the clinical scores such as SES-CD and CDAI using the QSP model outputs will be assessed.

Brief Project Background and Statement of Project Significance:

Crohn's disease (CD) is a chronic, relapsing inflammatory disease of the gastrointestinal (GI) tract with a prevalence of about 320 cases in 100,000 people in Europe and North America. Even though there are multiple approved therapies, long-term remission rates are low in Crohn's Disease patients with about 50% of patients requiring surgery within 10 years. Quantitative systems pharmacology models that provide mechanistic understanding of a disease are increasingly being used at various stages of drug development, for gaining insights into both safety and efficacy of a treatment. For efficacy predictions, QSP models that predict clinical scores, in addition to biomarkers, have applicability in late-stage drug development [1, 2].

Typically clinical scores constitute subjective assessments either by patients or physicians which make it challenging to predict using a mechanistic approach [3]. In Crohn's Disease, for example, SES-CD (Simple Endoscopic Score for Crohn's Disease) is used to measure the endoscopic activity in patients by evaluating endoscopy images. In addition to being an endpoint in clinical trials, endoscopic healing is one of the primary treatment goals in IBD as recommended by STRIDE consensus. However, given the invasive nature of endoscopic evaluations, physicians are often required to make treatment decisions based on limited objective information about the state of the

patient's gastrointestinal tissue while aiming to achieve mucosal healing. Non-invasive tools to predict changes in gut mucosal condition with treatment are needed.

We developed a computational platform for simulating mucosal health and the level of inflammation in patients with CD under different treatment scenarios [1]. The platform estimates changes in mucosal health and inflammatory activity over a continuous timeline of weeks to years. In order to estimate outcomes such as SES-CD and CDAI (Crohn's Disease Activity Index) at a patient level from the immunological model outputs, we would need de-identified individual patient data. Patient data from the following studies will be analyzed to create machine learning models to connect the immunological outputs from the Crohn's disease model to clinical scores [3]. The data will also be used to validate the model response to different therapies.

Impact for patients: Besides their application in drug development, computational tools that can predict endoscopic outcomes can be a promising form of decision support that can predict outcomes and patient progress in Crohn's disease.

Specific Aims of the Project:

This is a retrospective study where the deidentified patient data from the requested studies will be used to develop and validate a mechanistic-ML model of Crohn's disease. The primary objectives of this study are to
A) Create a machine learning algorithm to connect the immunological outputs from a model of Crohn's Disease to the clinical scores, and
B) Validate the model predictions

We will be using demographic and lifestyle data, disease history, baseline disease information, treatment history, lab results and symptoms to inform the model and predict the time course clinical scores and biomarkers. To validate the model, we would need time course SES-CD and CDAI score data split by their components for each gut segment, subject to availability. Please see the sections below for more information on data requirements.

Patients with Crohn's disease may benefit from this study in two ways: 1) the validated and published model will help researchers in drug development, 2) computational tools, developed using the validated model as the basis, that can predict endoscopic outcomes can be a promising form of decision support to predict outcomes and progress in Crohn's disease patients.

Study Design:

Meta-analysis (analysis of multiple trials together)

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

Participant-level data meta-analysis

Meta-analysis using only data from the YODA Project

Develop or refine statistical methods

Research on clinical prediction or risk prediction

Research Methods

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

Inclusion criteria: Patients from the requested studies who have the primary predictor variables, SES-CD and/or CDAI score data along with the component scores will be included in the study.

Exclusion criteria: None

Primary and Secondary Outcome Measure(s) and how they will be categorized/defined for your study:

Our study will focus on predicting SES-CD endoscopic score along with its components over time in individual patients post-treatment using the mechanistic-ML model of Crohn's Disease. All available time points for each individual will be used to prepare the training and validation datasets. Regarding SES-CD data, we request the components - size of ulcers, ulcerated surface, affected surface area and narrowing for each gut segment along with the total SES-CD for each patient at available time points. Similar component data for CDAI would be helpful, subject to availability.

Primary outcome measures:
1. Endoscopic outcomes:
A. SES-CD score over time
B. Individual time course SES-CD score data split by its components for each gut segment
2. Endoscopic response (percentage, 95% CI)
3. Endoscopic remission (percentage, 95% CI)

Secondary outcome measures:
1.Clinical outcomes:
A. CDAI over time
B. Individual time course CDAI score data split by its components
2. Clinical response (percentage, 95% CI)
3. Clinical remission (percentage, 95% CI)
4. Change in Fecal calprotectin over time (continuous variable, mg/L)
5. Change in CRP over time (continuous variable, ug/g)

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

Primary predictor/independent variables for the analysis include:
1. Demographic and lifestyle data: Age, Gender, Body weight, Smoking status
2. Disease history: Duration of Crohn's disease, Disease location classification, History of surgery
3. Baseline clinical data: SES-CD, CDAI, FCP (ug/g), CRP (mg/L)
4. Treatment history: Previous exposure to biologics, steroids and other SoC
5. Lab tests: Immune cell counts (if available), cytokine levels (if available)
6. Symptoms: Diarrhea, Abdominal pain reported during visits (if available)

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

Gender, age and ethnicity would also be of interest to run responder analysis and understand the characteristics of patients that are associated with response to specific biologics.

Statistical Analysis Plan:

Descriptive statistics will be used to summarize baseline characteristics (e.g. disease activity and patient demographics) as well as clinical and endoscopic outcomes. Dichotomous variables will be presented as proportions or percentages along with 95% CI. Continuous variables will be reported as means with standard deviations or medians with interquartile ranges. Linear imputation will be performed on continuous variables with missing data. Data will be analyzed using Python.

The processed data will then be used to create datasets for the study by combining all the features listed in the proposal. Each row in the final dataset will represent an individual patient and the columns would be the independent variables and outcomes. The final dataset will be randomly split into a training, validation and test sets in the ratio of 70/20/10. These datasets will be used to

develop the machine learning algorithm (decision trees/regression models similar to the approach used in Shim et al.) for linking immunology outcomes to clinical scores and to validate the overall performance of the hybrid mechanistic-ML model, The ML algorithm will be developed on the training set and the hyper parameters are tuned on the validation set. Finally, the overall performance of the mechanistic-ML model will be evaluated on the Test dataset.

Software Used:

Python

Project Timeline:

Target Analysis Start Date: 09/02/2025
Estimated Analysis Completion Date: 04/30/2026

Dissemination Plan:

We aim to present results of the work at Digestive Disease Week (DDW) or Crohn's & Colitis Congress or American Conference on Pharmacometrics (ACoP) and publish in Nature Systems Biology or Clinical Pharmacology and Therapeutics.

Bibliography:

1. Venkatapurapu, S. P., Iwakiri, R., Udagawa, E., Patidar, N., Qi, Z., Takayama, R., Kumar, K., Sato, Y., Behar, M., Offner, P., Dwivedi, G., Miyasaka, H., Suzuki, R. K., Hamada, A. L., D’Alessandro, P. M., & Fernandez, J. (2022). A computational platform integrating a mechanistic model of Crohn’s disease for predicting temporal progression of mucosal damage and healing. Advances in Therapy, 39(7), 3225--3247. https://doi.org/10.1007/s12325-022-02144-y
2. Rogers, K. V., Martin, S. W., Bhattacharya, I., Singh, R. S. P., & Nayak, S. (2021). A dynamic quantitative systems pharmacology model of inflammatory bowel disease: Part 1 – Model framework. Clinical and Translational Science, 14(1), 239--248. https://doi.org/10.1111/cts.12849
3. Shim, J. V., Rehberg, M., Wagenhuber, B., van der Graaf, P. H., & Chung, D. W. (2025). Combining mechanistic modeling with machine learning as a strategy to predict inflammatory bowel disease clinical scores. Frontiers in Pharmacology, 16, 1479666. https://doi.org/10.3389/fphar.2025.1479666