Statistical Analysis Plan

1. Descriptive statistics (Percentages, Means) to describe the baseline symptom characteristics of children newly diagnosed with CD will be run.

2. We will look at the distribution of patient scores on each PCDAI question. Patients receive a score of 0, 5, or 10 on most questions (0, 2.5, or 5 on others). We will compute and compare the percentage of patients who receive low, moderate, or high scores on each PCDAI question. Chi-square tests may be used to compare distribution of scores for patients with active versus inactive disease at follow-up.

3. Change on the PCDAI in relation to other measures of disease state including Global Assessments of Change (Physician, Parent, Patient) when available, and a pediatric disease-specific measure of health related quality of life completed by patients—the IMPACT-III, will be examined. The IMPACT-III asks participants about some of their symptoms (e.g., How much has your stomach been hurting you in the past two weeks? Response range: Not very much, to Not at all), and the degree to which they impair daily functioning. We will compare a subjective patient-reported measure to a clinician reported measure of disease state, on some aspects of disease activity. We will look at correspondence between patient-physician and/or parent-physician scores via Pearson and/or Spearman correlations (when appropriate) to further understand response and disease state.

4. Scores for subjective items (Abdominal pain, Stooling pattern, Patient well-being), Lab values (Hematocrit, Erythrocyte Sedimentation Rate, and Albumin), and other indicators of disease state (Weight, Height, Abdominal Tenderness, Perirectal disease, and Extra-intestinal manifestations) will be examined together to look at distribution of scores, and redundancy of items.

5. Drivers of change, in other words, which questions on the PCDAI change more in response to therapy, will be examined.

6. Once drivers of change are identified, we will further examine this change through linear regression. This will show the percentage of variance in scores accounted for by each of the PCDAI questions. This may further identify redundancy or removal of items.

7. Logistic regression will be carried out to look at different PCDAI models in order to determine the best combination of variables that correctly classify response and remission.

8. Psychometric properties of the newly proposed instrument will be examined and compared to previous PCDAI versions.

9. Lastly, Receiver Operating Characteristic (ROC) curves will be run to propose new clinical endpoints that establish cutoff values to define response and remission. This will be essential for use in clinical trials in order to define eligibility for studies, and ultimately response to therapy.