Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Madeline Bertha Date

The Characterization of Antimicrobial and Autoimmune Antibodies in African-Americans with Crohn's Disease

By

Madeline McGuire Bertha
Master of Science

Clinical Research

_____
Subramanian Kugathasan, MD
Advisor

_____
Beau B. Bruce, MD, MS, PhD
Committee Member

_____
Amita Manatunga, PhD
Committee Member

Accepted:

_____
Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

_____
Date

The Characterization of Antimicrobial and Autoimmune Antibodies in African-Americans with Crohn's Disease

By

Madeline McGuire Bertha
B.A. Georgetown University, 2008

Advisor:
Subramanian Kugathasan, MD

An abstract of
A thesis submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science
in Clinical Research
2015

Abstract

The Characterization of Antimicrobial and Autoimmune Antibodies in African-Americans with Crohn's Disease

By
Madeline Bertha

Background: Crohn's disease (CD) is a chronic, relapsing and remitting, inflammatory disease that affects the small bowel and/or colon. Recent studies have identified the role of antimicrobial and autoimmune antibodies in characterizing disease phenotype, location, complications and severity among Caucasians with CD. Despite these advances, very little is known about the nature of CD among African Americans (AA). This study will explore the relationship between serological antibodies and disease phenotype in AAs with CD and assess for their interaction with % African ancestry.

Methods: AAs with CD were enrolled as participants of 2 large IBD databases (GENESIS and BIG). Data on clinical features were obtained by patient interview or retrospective chart review. Serological levels of IgA ASCA, IgG ASCA, anti-OmpC, anti-CBir1, and pANCA were measured using enzyme linked immunosorbent assays. Genotyping was performed using Illumina immunochip technology. Multiple imputation by chained equations was performed to account for data missing at random. Logistic regression was used to calculate adjusted odds ratios (OR) for associations between serological markers and complicated disease, ileal disease, and disease requiring surgery.

Results: 358 patients were included in the analysis. The majority of our patients had inflammatory, non-complicated type disease (58.4%), perianal disease (55.7%), and disease involving the colon (86.8%). On multivariable analysis, IgG ASCA (comparing the 75%ile to the 25%ile) was associated with complicated disease (OR: 3.21; 95% confidence interval (CI) 1.98,5.21), ileal disease (OR: 2.42; 95% CI: 1.30, 9.02), and surgery (OR: 3.13; 95% CI: 1.72 , 5.69). IgA ASCA was associated with ileal disease (OR: 3.42; 95% CI: 1.30, 9.02). pANCA was associated with ileal disease (OR: 0.81; 95% CI: 0.68 , 0.97). Anti-OmpC was associated with surgery (OR:2.33; 95% CI 1.28, 4.25).

Conclusions: Anti-OmpC is an independent risk factor for surgery in AA. There is no interaction between % African ancestry and serological levels, however we had inadequate power to investigate this hypothesis. The diagnostic value of serological antibodies may vary significantly among different ethnicities and, therefore, should be interpreted differently in AA compared with Caucasians.

The Characterization of Antimicrobial and Autoimmune Antibodies in African-Americans with Crohn's Disease

By

Madeline McGuire Bertha
B.A. Georgetown University, 2008

Advisor:
Subramanian Kugathasan, MD

A thesis submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science
in Clinical Research
2015

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

APPENDIX

INTRODUCTION

Crohn's disease (CD) is a chronic, relapsing and remitting, inflammatory disease that affects the small bowel and/or colon. It often presents as a diverse disorder and can have varying disease course and clinical manifestations (1). Significant advances in our understanding of disease pathogenesis have been made over the past fifteen years as we have begun to expose the relationship that exists between genetic, microbial and immune response in individuals with CD (2). Recent studies have identified the role of serological antibodies (targeting microbial antigens and autoantibodies) in characterizing disease phenotype, location, complications and severity among Caucasians with CD (3). They have identified an association between the presence of ASCA IgA, ASCA IgG, anti-OmpC, anti-CBir1, and pANCA with the manifestation of complicated disease, disease located in the ileum and disease requiring surgery in Caucasian populations of Western European descent (3-5). These studies bring us closer to the development of tools for risk stratification and treatment selection (2).

Despite these advances, very little is known about the nature of CD within the African American (AA) population; to date, there has only been one study examining serology in AAs. This analysis was limited to ASCA alone and was limited by its reliance on self-reported ethnicity (6). There is a need for further characterization of these biomarkers in AAs, as the diagnostic values of such serological indicators are known to vary significantly among different ethnic groups and geographic regions (3). We explored relationship between serological anti-microbial and autoimmune antibodies and disease phenotype in AAs with CD, while controlling for genetic ancestry.

BACKGROUND

Inflammatory bowel disease (IBD) is a chronic, inflammatory disorder of the gastrointestinal tract that often presents with abdominal pain, malabsorption and rectal bleeding. It is comprised of two major disorders: ulcerative colitis (UC) and Crohn's disease (CD). UC is limited to the colon, whereas CD is known to affect any part of the gastrointestinal tract. CD is further characterized by the presence of transmural inflammation on histology, with endoscopy generally revealing a pattern of discontinuous lesions, cobblestoning, stricturing and apthous or linear ulcerations (7, 8). In North America, the incidence of CD is up to 20 cases per 100,000 person-years with the onset typically occurring in children and young adults (15-25 years) (1, 9). For years CD has been accepted as a heterogeneous disorder with varying presentation and clinical manifestations. However, recent advances in the characterization of the genetic, serological and microbial markers related to CD suggest that this disorder constitutes a collection of disease subtypes with both common and distinguishing characteristics (10). While the etiopathogenesis remains unclear, it is increasingly understood that CD is, in part, due to the complex interactions between the immune system, enteric bacteria and host genotype (11).

Integral to this advancement is the understanding and characterization of serological immune responses to microbial antigens and autoantibodies; specifically, antibodies to the *E. coli* outer-membrane porin C (OmpC), the bacterial flagellin (CBir 1), *Saccharomyces cerevisiae* (ASCA), and antineutrophilic cytoplasmic antibodies (pANCA) (3). It is believed that the development of serological antibodies in CD reflects a general, and likely genetically determined, loss of tolerance toward bacterial and fungal

flora (12). Numerous studies have evaluated the ability of these markers to differentiate patients with IBD from healthy controls, as well as differentiate UC from CD, with conflicting results. However, other studies that examined their ability to determine disease phenotype and predict a complicated disease course among patients with CD have revealed more promising results: ASCA positivity has been shown to be associated with disease located in the small bowel, stricturing and/or penetrating disease, a higher risk for surgery, and early disease onset; anti-CBir1 positivity has been associated with disease located in the small bowel, stricturing and/or penetrating disease, and early disease onset; anti-OmpC positivity has been associated with disease located in the small bowel, structuring and/or penetrating disease, and a higher risk for surgery; and pANCA positivity has been associated with a more benign disease course and disease located in the colon (3, 13). The ability of serological antibodies, in combination with other genetic, environmental and microbial data, to help predict risk of disease progression and outcome could prove key to the management of patients with CD and may lend itself to the development of more personalized care. However, these associations have been derived predominantly from studies of Caucasian populations and likely vary between different ethnicities. The associations of these antibodies with disease phenotypes need to be assessed in populations of non-Western European descent in order to determine their utility apart from the populations in which they have been developed (2, 3).

Despite the high prevalence of CD in the United States, the racial and ethnic aspects of IBD remain poorly studied (14). There have been very few trials designed to look at the phenotypic presentation of CD in AAs. Current data is conflicting and limited but suggests that AAs *may* have increased ileocolonic disease, increased fistulizing

perianal disease, increased extraintestinal manifestations, and may be more likely to require surgery, particularly when fistulizing perianal disease is present, compared with Caucasians (14-17). In regards to serology, there has only been one study designed to look at immune response to microbial antigens in AAs. It found that ASCA has a similar sensitivity but a lower specificity for CD, as well as being associated with ileal involvement, complicated behavior and surgery in AAs with CD. In addition to only focusing on ASCA, this study was further limited by its reliance on self-reported ethnicity which can lead to misclassification (6). Due to differences in environmental and genetic influences among various ethnic populations, there is a need to further delineate the diagnostic value of these biomarkers among the AA community (3).

Our study was designed to quantify serological levels of IgA ASCA, IgG ASCA, anti-Ompc, anti-CBir1, and pANCA among AA's with CD and to determine whether these levels are a risk factor for complicated disease, disease located in the ileum and disease requiring surgery among AA with CD, while controlling for established risk factors. Additionally, we controlled for genetic ancestry and assessed for interaction between % African admixture and serological levels for each of the above mentioned outcomes.

METHODS

Study Design and Hypothesis:

This cross sectional study tested the hypothesis that serological levels of IgA ASCA, IgG ASCA, anti-OmpC, anti-CBir1, and pANCA are a risk factor for complicated disease, disease located in the ileum and disease requiring surgery among AA with CD. The Institutional Review Board at Emory University, Children's Hospital of Atlanta, Atlanta VA Medical Center Children's Hospital of Philadelphia, Cincinnati Children's Hospital Medical Center, University Hospitals Case Western Medical Center, University of Maryland School of Medicine, Vanderbilt-Monroe Carell Jr. Children's Hospital, UT Southwestern, UNC Chapel Hill, University of Chicago Children's Hospital, LSU Health Science Center, Cooks Medical Center, and Willis-Knighton Physician Network approved the study, and informed consent was obtained from all participants.

Study Population:

The study population consisted of over 500 individuals, self-identified as AAs and enrolled as participants of 2 large IBD databases between August 2011 and March 2014. The GENESIS database (Gene Discoveries in Subjects with CD of African Descent) enrolled patients from 12 participating sites. This database contained clinical, serologic and genetic data on 414 African Americans with CD. Blood samples for DNA purification and sera were successfully collected from all study subjects. All clinical information was obtained at the time of blood collection by the consenting physician. The BIG database (Genetic Analysis of Children and Adults With and Without Inflammatory Bowel Disease) enrolled self-identified AAs from Children's Hospital of Atlanta, Emory University Hospital, the VA, and Grady Memorial Hospital. This database contains

clinical, serologic and genetic data on 113 AAs with CD. Blood samples for DNA purification and sera were successfully collected from all study subjects at the time of enrollment. All clinical information was obtained by retrospective chart review.

To be an eligible case, all CD participants must have a confirmed diagnosis of CD, based on standard diagnostic criteria (18), readily available serological results, well-documented disease behavior and may not be genetically related to any other case. Based on this criteria, we excluded 125 participants from GENESIS and 21 participants from BIG due to lack of available serology results. Additionally, we excluded 3 participants from GENESIS and 20 participants from BIG because they were either genetically related to another participant or had a missing chart and a chart review was unable to be performed. In summary, a total of 358 AA with CD were included in our study (**Figure 1**).

Clinical Characteristics of CD Patients:

Patient demographics, date of diagnosis, disease duration, disease location, disease behavior, surgical history, extraintestinal manifestations (EIM), smoking history, autoimmune history, family history, and history of biologic medication use were obtained either at the time of blood draw or via retrospective chart review. CD phenotyping was performed in accordance with the Montreal Classification for adults and Paris Classification for children (7, 8). For disease location, patients were classified into one of four mutually exclusive groups: L1 (terminal ileal disease +/- limited cecal disease), L2 (colonic disease), L3 (ileocolonic disease), L4 (isolated upper disease). The definition of location of CD was based on “macroscopic appearance of mucosal ulceration, anywhere along the GI tract (with the exception of the mouth) or bowl wall thickening on

radiography. The presence of mucosal erythema and/or granularity was not sufficient to be considered evidence of involvement" (8). In addition, the presence of upper gastrointestinal (UGI) disease was also documented, as many patients present with UGI manifestations in addition to ileal and/or colonic disease. Patients were classified into four groups: 0 (no disease), L4a (upper disease proximal to the Ligament of Treitz), L4b (upper disease distal to the Ligament of Treitz and proximal to the distal 1/3 ileum), and L4ab (8). For disease behavior, patients were categorized into four groups: B1 (non-stricturing non-penetrating disease), B2 (structuring disease), B3 (penetrating disease), and B23 (both structuring and penetrating disease, either at the same moment in time or separately over a period of time) (8). Complicated disease was defined as B1, B2, or B2B3. History of surgery was defined as any confirmed, documented surgical procedure necessary for the treatment or management of CD. Surgeries related to perianal disease were not included. Extraintestinal manifestations were defined as any documented history of large joint arthritis, small joint arthritis, iritis/uveitis, pyoderma gangrenosum, erythema nodosum, sacro-ileitis, or primary sclerosing cholangitis. Current smoking was positive if, at enrollment, the subject was a smoker and/or lived with someone who smokes cigarettes. History of smoking was positive if the subject had a history of at least 6 months of consistent exposure to smoke. Autoimmune history was positive if the patient had a documented history of rheumatoid arthritis, idiopathic diabetes mellitus, asthma/atopy, multiple sclerosis, lupus, psoriasis, ankylosing spondylitis, or autoimmune thyroid disease. Family history was positive if the patient's birth mother or father had a history of CD or UC. History of biologic use was positive if there was any current or past biologic use as documented in the medical records or reported by the patient. Disease

duration was calculated as the difference between the year of enrollment and the year of diagnosis. Age at enrollment was calculated as the difference between the year of enrollment and the year of birth. Age at diagnosis was calculated as the difference between the year of diagnosis and the year of birth.

Serological Analysis:

Blood samples were collected at the time of enrollment. Sera were measured for expression of ASCA IgG, ASCA IgA, anti-OMPC, anti-CBir1 and pANCA antibodies in a blinded fashion by an enzyme-linked immunosorbent assay (ELISA). The tests were run at Prometheus Laboratories or Cedars-Sinai using previously described protocols and standards (19). Antibody levels were measured relative to the Cedars-Sinai Laboratory or Prometheus Laboratory standard and were expressed in ELISA units (EU/mL).

All serological values were treated as continuous variables except for on bivariable analysis, where they were analyzed as categorical variables. All cut off values were determined according to the manufacturer's guidelines. IgA ASCA was classified as negative (0.0 – 20.0 EU/ml) and positive (≥ 20.0 EU/ml). IgG ASCA was classified as negative (0.0 – 40.0 EU/ml) and positive (≥ 40.0 EU/ml). Anti-OmpC was classified as negative (0.0 – 23.0 EU/ml) and positive (≥ 23.0 EU/ml). Anti-CBir1 was classified as negative (0.0 – 25.0 EU/ml) and positive (≥ 25.0 EU/ml). pANCA was classified as negative (0.0 – 30.0 EU/ml) and positive (≥ 30.0 EU/ml).

Genotyping:

DNA samples were derived from whole blood. All DNA samples were genotyped for NOD2/CARD15 single nucleotide polymorphsims (SNPs) rs2066844, rs2066846, and rs5743293 using the Illumina Immunochip and the genotyped cells were

made by using GenomeStudio version 2011.1 and Genotyping Module version 1.9.4. Samples were genotyped at Cedars-Sinai Medical Center Genetics Institute.

Patient genotypes were divided into two groups NOD2: positive and negative. NOD2 positive patients included those who carried two NOD2 mutant alleles or a single NOD2 mutant allele. NOD2 negative patients included those who carried only wild-type alleles.

African Ancestry Estimation:

Because AAs are well-modeled as linear combinations of West African and European ancestries, to estimate the locus-specific local ancestry, we chose the WINPOP model in LAMP package for its fast computation with low error rate, WINPOP takes allele frequencies from ancestral populations (YRI and CEU from HapMap), as input, and outputs the local ancestry estimate for each sample at each SNP as proportion of YRI at values of 0, 0.5, or 1 (20-22). The global YRI ancestry for each sample was estimated by using ADMIXTURE which only requires sample genotypes and number of ancestral populations as input and outputs estimated proportion from each ancestral population with a numeric value ranging from 0 to 1. (23).

Statistical Analysis:

Statistical analyses were performed using SAS 9.3 (SAS Institute Inc., Cary, NC) and R 3.1.1 (R Foundation for Statistical Computing, Vienna, Austria). Statistical tests were 2-sided, and a p-value less than 0.05 was considered statistically significant. The Shapiro-Wilk Test was used to test for normality on all continuous variables. Descriptive statistics are presented as median (interquartile range) and percent (95% confidence interval) for continuous and categorical outcomes, respectively.

Bivariate Analysis:

Two separate bivariable analyses were performed to determine associations between outcome, exposure and predictor variables. Our first bivariable analysis assessed the association of serological markers (IgA ASCA, IgG ASCA, anti-OmpC, anti-CBir1, and pANCA) as categorical variables with our four outcomes of interest (complicated disease, any ileal disease, isolated ileal disease, and disease requiring surgery) and our other predictor variables. Non-parametric Mann-Whitney Tests were used for all continuous variables and chi-squared or Fisher's Exact Tests were used for binomial variables. Our second bivariable analysis assessed the association between the categorical outcome variables and each predictor variable using simple logistic regression.
All independent terms were tested for linearity with the log odds of the outcome (alpha=0.05) using 5-knot restricted cubic splines according to the method of Harrell (24). One-knot linear splines were created based on review of the plot of the restricted cubic spline and subject matter knowledge for those variables that did not meet this assumption.

Multiple Imputation:

Multiple imputation by chained equations (MICE) was employed for multivariable analysis in order to account for data presumed to be missing at random (MAR). This method of handling missing data results in valid statistical inferences that correctly reflect the uncertainty due to missing values (25). Age at diagnosis, disease duration, gender, family history of IBD, personal history of autoimmune disorder, presence of extraintestinal manifestations, current smoking status, disease location, disease behavior, presence of upper disease modifiers, presence of perianal disease,

surgical history, time until surgery, history of biologic use, presence of NOD2 mutation, % African admixture, IgA ASCA, IgG ASCA, PANCA, CBir1, PANCA, and all variables interacted with % African admixture were included in the imputation regression models along with the relevant interaction terms. 100 complete datasets were created using stochastic regression. Analyses were performed on each 100 complete datasets and the resulting estimates were then combined (25). Imputed data sets were only used for multivariable modeling. All univariable and bivariate analyses were performed on the original dataset.

Multivariable Modeling Approach:

A multivariate logistic regression model was used to evaluate the primary associations among serologic levels with the presence of complicated disease, ileal disease disease and disease requiring surgery. All relevant demographic, phenotypic, genetic, and serological variables except history of surgery, history of biologic use, and outcome of interest were included in the initial model. Fast backward variable selection using Wald chi-square of individual factors at the 0.1 significance level was performed until only statistically significant variables remained (26). Odds ratio's and 95% confidence intervals were calculated using the 75^{th} percentile to the 25^{th} percentile for all continuous variables as a way to assess which variables have the largest impact over the same range. All initial and final models were tested for interaction with % African admixture.

RESULTS

Patient Characteristics:

Data from 358 unrelated AA participants with CD were analyzed. The demographic and phenotypic characteristics of these participants are characterized in **Table 1**. Serological levels for IgA ASCA, IgG ASCA, anti-OmpC, anti-CBir1, and pANCA were available on all participants. Greater than 15% of data was missing for % African admixture, current smoking status, presence of perianal disease and biologic use at diagnosis **(Table A2)**. The median age at diagnosis was 15 years (IQR: 12 – 22.5), with a median disease duration of 4 years (IQR: 1- 10). Ileocolonic disease was observed in 53.6% of patients. Perianal disease was observed in 55.7%. Disease behavior was complicated (either stricturing, penetrating or both) in 41.6%. Biologic agents were used at some point in their disease course in 69.9% of patients. Nod2 positivity was observed in 5.9% of patients.

The median level of IgA ASCA was 7.1 EU/ml (IQR: 2.3 – 26.7). The median level was IgG was 21.1 EU/ml (IQR: 5.9 – 53.5). The median level anti-OmpC was 15.7 EU/ml (IQR: 9.9 – 27.7). The median level of anti-CBir1 was 24.2 EU/ml (IQR: 14.4 - 46.1). The median level of pANCA was 16.4 EU/ml (IQR: 10.3 – 23.4). The frequency of IgA and IgG ASCA positivity was 34.9% and 32.7%, respectively. The frequency of anti-OmpC positivity was 31.3%. The frequency of anti-CBir1 positivity was 49.4%. The frequency of pANCA positivity was 17.0%.

Patient Demographic and Phenotypic Characteristics According to Serological Status:

Among the 358 AA patients with CD 155 were ASCA positive, 112 were anti-OmpC positive, 177 were positive for anti-CBir1, and 61 were positive for PANCA. The demographic and phenotypic characteristics of participants according to serological status are displayed in **Table 2**. On bivariate analysis, the following characteristics were associated with ASCA positivity ($P<0.05$): personal history of autoimmune disease ($P=0.02$), EIM ($P=0.02$), complicated disease ($P<0.0001$), ileal disease ($P<0.0001$) and disease requiring surgery ($P=0.002$). Anti-OmpC positivity was significantly associated with age at enrollment ($P=0.005$), disease duration ($P=0.0004$) and disease requiring surgery ($P=0.002$). Anti-CBir1 positivity was associated with age at enrollment ($P=0.001$), age at diagnosis ($P=0.008$), disease duration ($P=0.06$), % African admixture ($P=0.004$), current smoke exposure ($P=0.002$) and isolated ileal disease ($P<0.0001$). pANCA positivity was significantly associated with EIM ($P=0.01$), complicated disease ($P=0.003$) and ileal disease ($P=0.01$).

Patient Characteristics and Serological levels Associated with Complicated Disease:

One-knot linear splines were created for the variables disease duration, ASCA IgG and CBir1 at 10 years, 15 EU/ml, 20 EU/ml, and 20 EU/ml, respectively (**Figure 2**). **Table 3** displays the univariate logistic regression of the outcome complicated disease with all serological markers and other predictors of interest. Complicated disease was significantly associated with age at diagnosis, disease duration, ileal involvement, colonic involvement, IgG ASCA, anti-OmpC, anti-CBir1, and pANCA. **Table 4** displays the crude odds ratios comparing the 75^{th} percentile to the 25^{th} percentile (IQR) for all continuous variables. IgG ASCA (OR: 3.83; CI: 2.5,6.0) had the largest odds ratio of the

serological markers, followed by anti-OmpC (OR:2.77; CI: 1.8,4.2), ASCA IGA (OR: 1.86; CI:1.5,2.4), anti-Cbir1 (OR: 1.54; CI: 1.1,2.0) and pANCA (OR: 0.79; CI: 0.7,0.9).

On multivariable analysis (**Table 5**), disease duration (OR:6.44; CI:3.18,13.04) IgA ASCA (OR:3.21; CI: 1.98,5.21) and ileal disease (OR: 3.23; CI: 1.79,5.88) were significantly associated with complicated disease. Interaction with % African admixture was not significant.

Patient Characteristics and Serological levels Associated with Ileal Disease:

One-knot linear splines were created for the variables age at diagnosis and IgA ASCA at 15 years and 40 EU/ml, respectively **(Figure 3)**. On univariable analysis (**Table 3**), the following variables were found to be associated with the presence of any ileal disease: age at diagnosis, nod2 positivity, upper disease, IgA ASCA and IgG ASCA. **Table 4** displays the crude odds ratios comparing the 75%ile to the 25%ile (IQR) for all continuous variables. IgA ASCA (OR: 6.01; CI: 2.5,13.7) had the largest odds ratio, followed by IgG ASCA (OR: 3.21; CI: 2.0,5.1)**,** and pANCA (OR:0.76; CI:0.6,0.9). All variables in Table 4 were included in the original multivariable model. The final multivariable model (**Table 6**) showed that the presence of ileal disease was significantly associated with perianal disease (OR: 2.83; CI: 1.50,5.33), upper intestinal disease (OR: 2.33; CI: 1.26,4.29), IgA ASCA (OR: 3.42; CI: 1.30, 9.02); IgG ASCA (OR: 2.42; CI: 1.33,4.39), and pANCA (OR: 0.81; CI: 0.68,0.97). Interaction with % African admixture was assessed and found to be insignificant.

Patient Characteristics and Serological levels Associated with Surgery:

One-knot linear splines were created for age at diagnosis, disease duration, IgA ASCA, IgG ASCA, and anti-OmpC at 25 years, 10 years, 40 EU/ml, 20 EU/ml, and 20

EU/ml, respectively **(Figure 4)**. On univariable analysis age at diagnosis, disease duration, current smoking history, ileal involvement, upper intestinal disease, IgA ASCA, IgG ASCA and anti-OmpC were found to be associated with surgery (**Table 3).** Table 4 displays the crude odds ratios comparing the 75%ile to the 25%ile (IQR) for all continuous variables. Anti-OmpC (OR: 4.17; CI: 2.5,6.9) had the largest odds ratio, followed by IgG ASCA (OR: 3.3; CI: 2.0,5.4) and IgA ASCA (OR: 3.19; CI: 1.7,6.0). All variables in Table 4 were included in the original multivariable model. The final multivariable model (**Table 7**) showed that disease requiring surgery was significantly associated with Nod2 positivity (OR: 3.27; CI: 1.07,9.96), disease duration (OR: 6.80; CI: 3.08,15.05), IgG ASCA (OR: 3.13; CI: 1.72,5.69), and anti-OmpC (OR: 2.33; CI: 1.28,4.25). An additional model was created (**Table A1)** to assess whether the significant findings in the previous model remained significant after controlling for the potential confounders age at diagnosis, family history, current smoking status, and the presence of ileal disease. This model showed that disease requiring surgery remained significantly associated with disease duration (OR: 6.95; CI: 3.06,15.81), IgG ASCA (OR: 2.54; CI: 1.33,4.84) and anti-OmpC (OR: 2.25; CI: 1.20,4.20). Nod2 positivity was no longer significant (OR: 2.61; CI: 0.81,8.39). Interaction with % African admixture was assessed for all models and was not significant.

DISCUSSION

Our study is the largest AA cohort to report on the prognostic potential of serum biomarkers in CD and the first study to evaluate anti-OmpC, anti-CBir1, and pANCA in AA's with CD. The results of our study showed: (i) the majority of our patients had inflammatory, non-complicated type disease, perianal disease, and disease involving the colon (ii) anti-OmpC is an independent risk factor for surgery in AA (iii) There were threshold effects for the level of these serological markers in the AA with CD population.

Our study showed that the majority of our patients had inflammatory, non-complicated type disease (58.4%), perianal disease (55.7%) and disease involving the colon (86.8%). Among participants with colonic disease, 53.6% had disease involving both the colon and small bowel. Our observations regarding disease location and disease behavior concur with the findings of both Mahid et al. and Hou et al., who published systematic reviews on the epidemiology and phenotypic presentation of IBD in AA (15, 27). These findings are also consistent with those found in Caucasian populations, which is contrary to the earlier literature, which reported that AA have a more severe course with a different disease distribution (14, 16).

Numerous studies have linked the presence and magnitude of serological antibodies with complicated disease, ileal involvement, and earlier disease onset in Caucasian populations. Dassopoulos et al. conducted the first study to assess ASCA levels in AA with CD. They reported that, like Caucasians, ASCA was independently associated with ileal involvement, complicated behavior and surgery in AA with CD (6).

Our study confirms and expands upon the findings of Dassopoulos et al. by revealing an association of (i) IgA ASCA with disease located in the ileum (ii) IgG

ASCA with complicated disease, disease located in the ileum and disease requiring surgery (iii) pANCA with disease located in the colon and (iv) anti-OmpC with disease requiring surgery. The association between pANCA and disease located in the colon is consistent with previously reported findings among white populations, confirming the "UC-like" effect of pANCA in CD (3). Literature on anti-OmpC has found that its presence and/or magnitude, as represented by quartiles, is associated with the need for surgery in Caucasian populations (5, 28, 29). However, our study is the first to report that anti-OmpC is *independently* associated with disease requiring surgery in the AA population. Moreover, we performed a secondary analysis to assess whether anti-OmpC remained significant after controlling for age at diagnosis, disease duration, family history, current smoke exposure, NOD2 positivity, disease location, and IgG ASCA. Anti-OmpC remained statistically significant with an odds ratio of 2.25 (CI: 1.2,4.2) **(Table A1**).

Additionally, our study is the first study to treat serological markers as continuous, numerical values, as well as to address their issue of non-linearity. In doing so, we found that the majority of the serological markers of interest appear to have threshold effects. In other words, the odds of having a particular outcome of interest plateaus once levels reach a particular value. This novel finding is especially important because, to date, the vast majority of the literature analyzes serological markers as binary or nominal variables based on cutoffs derived from a potentially different population. This may result in the misclassification of patient status with respect to disease severity and complications. Furthermore, this discovery has the potential to alter the way we

interpret the results of these serological tests, which may have direct implications on clinical practice.

Our analysis was unique in that we were that we were the first study to control for race using % African admixture. We hypothesized that there would be an interaction between % African admixture and the level of autoimmune and antimicrobial serological levels. However, we found none. These findings were likely secondary to the fact that our population had minimal admixture variation (IQR: 0.8,0.9) resulting in inadequate power to investigate this hypothesis. Further studies may benefit from assessing interaction using % European admixture (the inverse of % African admixture) among Caucasians, AA, and those individuals who identify with both ethnicities. This analysis would allow for the variability and power required in order to truly assess the interaction between serological markers and race.

The limitations of our study include the substantial amount of missing data on our participant population. Table A2 shows the distribution of our missing data. While we were only missing greater than 15% of data on four variables, eleven additional variables were missing values for anywhere between 1.4% and 13.97% of the participants. As a result, our sample size decreased by over 50% when we performed multivariable analysis with our original model. We addressed this limitation by multiple imputation by chained equations (MICE), a principled method of dealing with missing data (25). Additionally, our study was limited by the fact that we did not have any demographic or phenotypic information on those patients who chose not to participate. For this reason, our study may have some degree of selection bias, which we are unable to control for.

These limitations notwithstanding, the results of our study suggest that the diagnostic value of anti-microbial and autoimmune antibodies may vary among different ethnic and geographic populations due to differences in environmental and genetic influences (3). Consequently, serological antibodies should not only be interpreted differently in AA compared with Caucasians but these differences likely need to be considered and incorporated into clinical practice. Additionally, our data strongly suggests that serological markers should be analyzed as continuous variables and not binary or ordinal variables, as previous literature suggests.

To expand upon the information gained from this study, we can continue to explore the relationship that exists between antimicrobial and autoimmune antibodies and disease complications in participants with CD by performing a prospective cohort study of AA and Caucasians with newly diagnosed CD. In this way we can begin to assess causality and to truly understand the differences in serological levels that exist between races. Our findings may also be useful to develop of a risk score based on combinations of clinical, genetic, serological and microbiome data for individualized risk stratification and treatment selection.

REFERENCES

1. Loftus EV, Jr. Clinical epidemiology of inflammatory bowel disease: Incidence, prevalence, and environmental influences. *Gastroenterology* 2004;126(6):1504-17.
2. Gerich ME, McGovern DP. Towards personalized care in IBD. *Nature reviews Gastroenterology & hepatology* 2014;11(5):287-99.
3. Prideaux L, De Cruz P, Ng SC, et al. Serological antibodies in inflammatory bowel disease: a systematic review. *Inflammatory bowel diseases* 2012;18(7):1340-55.
4. Papp M, Altorjay I, Dotan N, et al. New serological markers for inflammatory bowel disease are associated with earlier age at onset, complicated disease behavior, risk for surgery, and NOD2/CARD15 genotype in a Hungarian IBD cohort. *The American journal of gastroenterology* 2008;103(3):665-81.
5. Mow WS, Vasiliauskas EA, Lin YC, et al. Association of antibody responses to microbial antigens and complications of small bowel Crohn's disease. *Gastroenterology* 2004;126(2):414-24.
6. Dassopoulos T, Nguyen GC, Talor MV, et al. NOD2 mutations and anti-Saccharomyces cerevisiae antibodies are risk factors for Crohn's disease in African Americans. *The American journal of gastroenterology* 2010;105(2):378-86.
7. Satsangi J, Silverberg MS, Vermeire S, et al. The Montreal classification of inflammatory bowel disease: controversies, consensus, and implications. *Gut* 2006;55(6):749-53.
8. Levine A, Griffiths A, Markowitz J, et al. Pediatric modification of the Montreal classification for inflammatory bowel disease: the Paris classification. *Inflammatory bowel diseases* 2011;17(6):1314-21.
9. Molodecky NA, Soon IS, Rabi DM, et al. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology* 2012;142(1):46-54.e42; quiz e30.
10. Silverberg MS, Satsangi J, Ahmad T, et al. Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: report of a Working Party of the 2005 Montreal World Congress of Gastroenterology. *Canadian journal of gastroenterology = Journal canadien de gastroenterologie* 2005;19 Suppl A:5a-36a.
11. Reiff C, Kelly D. Inflammatory bowel disease, gut bacteria and probiotic therapy. *International journal of medical microbiology : IJMM* 2010;300(1):25-33.
12. Sartor RB. Mechanisms of disease: pathogenesis of Crohn's disease and ulcerative colitis. *Nature clinical practice Gastroenterology & hepatology* 2006;3(7):390-407.
13. Kuna AT. Serological markers of inflammatory bowel disease. *Biochemia medica* 2013;23(1):28-42.
14. Agrez MV, Valente RM, Pierce W, et al. Surgical history of Crohn's disease in a well-defined population. *Mayo Clinic proceedings* 1982;57(12):747-52.
15. Hou JK, El-Serag H, Thirumurthi S. Distribution and manifestations of inflammatory bowel disease in Asians, Hispanics, and African Americans: a

systematic review. *The American journal of gastroenterology* 2009;104(8):2100-9.
16. Sonnenberg A, McCarty DJ, Jacobsen SJ. Geographic variation of inflammatory bowel disease within the United States. *Gastroenterology* 1991;100(1):143-9.
17. El-Tawil AM. Inflammatory bowel disease and African Americans. *Acta gastro-enterologica Belgica* 2009;72(2):271-2; author reply 3.
18. Bernstein CN, Fried M, Krabshuis JH, et al. Inflammatory bowel disease: a global perspective. In: Organisation WG, ed. *World Gastroenterology Organisation Global Guidelines*, 2009.
19. Landers CJ, Cohavy O, Misra R, et al. Selected loss of tolerance evidenced by Crohn's disease-associated immune responses to auto- and microbial antigens. *Gastroenterology* 2002;123(3):689-99.
20. Price AL, Tandon A, Patterson N, et al. Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. *PLoS genetics* 2009;5(6):e1000519.
21. Pasaniuc B, Sankararaman S, Kimmel G, et al. Inference of locus-specific ancestry in closely related populations. *Bioinformatics (Oxford, England)* 2009;25(12):i213-21.
22. Seldin MF, Pasaniuc B, Price AL. New approaches to disease mapping in admixed populations. *Nature reviews Genetics* 2011;12(8):523-8.
23. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome research* 2009;19(9):1655-64.
24. Harrell F. *Regression Modeling Strategies, with Applications to Linear Models, Survival Analysis, and Logistic Regression*. New York: Springer; 2001.
25. Azur MJ, Stuart EA, Frangakis C, et al. Multiple imputation by chained equations: what is it and how does it work? *International journal of methods in psychiatric research* 2011;20(1):40-9.
26. Lawless JF. Efficient Screening of Nonnormal Regression Models. *Biometrics* 1978;34(2):318-27.
27. Mahid SS, Mulhall AM, Gholson RD, et al. Inflammatory bowel disease and African Americans: a systematic review. *Inflammatory bowel diseases* 2008;14(7):960-7.
28. Dubinsky MC, Kugathasan S, Mei L, et al. Increased immune reactivity predicts aggressive complicating Crohn's disease in children. *Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association* 2008;6(10):1105-11.
29. Ferrante M, Henckaerts L, Joossens M, et al. New serological markers in inflammatory bowel disease are associated with complicated disease behaviour. *Gut* 2007;56(10):1394-403.

Table 1: Demographic and Phenotypic Status among AA with CD (N=358)

Variable	Median (IQR) or n(%)
African Admixture*	0.83 (0.8 - 0.9)
Age (IQR) -- yr	20 (15 - 36)
Age at Diagnosis (IQR) --	15 (12 - 22.5)
Age Category at Diagnosis	
< 10 years	39 (11.1%)
10 ≥ x < 18 years	183 (52.1%)
≥ 18 years	129 (36.8%)
Disease Duration (IQR) -- yr	4 (1 - 10)
Male sex -- no. (%)	181 (50.6%)
Family History of IBD -- no. (%)	27 (8.2%)
History of Autoimmune Disorder -- no. (%)	96 (27.7%)
Extraintestinal Manifestations -- no. (%)	95 (27.7%)
Smoking -- no. (%)	
Current*	84 (28.8%)
Ever	135 (42.4%)
Disease Location -- no. (%)	
L1: Ileal	42 (13.2%)
L2: Colonic	106 (33.2%)
L3: Ileocolonic	171 (53.6%)
L4: Isolated Upper	0 (0.0%)
Upper Disease Modifier[1]-- no. (%)	
No Disease	198 (63.6%)
L4A	73 (23.5%)
L4B	20 (6.4%)
L4AB	20 (6.4%)
Disease Behavior -- no. (%)	
B1: inflammatory	209 (58.4%)
B2: stricturing	6 (18.2%)
B3: penetrating	37 (10.3%)
B2B3: stricturing and penetrating	47 (13.1%)
Complicated Disease -- no. (%)	149 (41.6%)
Perianal Disease* -- no. (%)	156 (55.7%)
Surgical History -- no. (%)	128 (40.1%)
Biologic Use -- no. (%)	
At diagnosis*	55 (19.0%)
Ever	229 (69.9%)
Nod2 Positivity	20 (5.9%)
Serological Response (level)	
ASCA IgA	7.1 (2.3 - 26.7)
ASCA IgG	21.1 (5.9 - 53.5)
anti-OmpC	15.7 (9.9 - 27.7)
anti-CBir1	24.2 (14.4 - 46.1)
pANCA	16.4 (10.3 - 23.4)
Serological response[2] -- no. (%)	
ASCA IgA	125 (34.9%)
ASCA IgG	117 (32.7%)
anti-OmpC	112 (31.3%)
anti-CBir1	177 (49.4%)
pANCA	61 (17.0%)

*Missing greater than 15% of the data

[1] L4A: disease proximal to the ligament of treitz; L4B: disease distall to the ligament of treitz but proximal to the distal 1/3 of the ileum

[2] Results are classified as negative or positive according to the manufacturer's determined cut offs. A negative response was defined as: ASCA IgA < 20 EU/ml; ASCA IgG < 40 EU/ml; anti-OmpC < 23 EU/ml; anti-CBir1 < 25 EU/ml; pANCA < 17.46 EU/ml

AA: African-American; CD: Crohn's disease; ASCA: anti-Saccharomyces cervisiae antibody; anti-OmpC: E.coli outer-membrane porin C; anti-Cbir1: flagellin-like bacterial antigen; pANCA: perinuclear anti-nuclear cytoplasmic antibody

Table 2: Demographic and Phenotypic Characteristics according to Serological Status in AA with CD

Predictors	ASCA[2] Positive (N=155)	ASCA[2] Negative (N=203)	P-Value	anti - OmpC Positive (N=112)	anti - OmpC Negative (N=246)	P-Value	anti - Cbir1 Positive (N=177)	anti - Cbir1 Negative (N=181)	P-Value	pANCA Positive (N=61)	pANCA Negative (N=297)	P-value
Age (IQR)[1] -- yr	21 (16 - 35)	18 (15 - 37)	0.11	25 (16-40)	18 (15-33)	0.005	18 (15 - 28)	22 (16-44)	0.001	18 (16 - 39)	20 (15 - 36)	0.61
Age at Diagnosis[1]-- yr	16 (13 - 22)	15 (11 - 24)	0.25	15 (12 -23.5)	15 (11.5-21.5)	0.44	15 (12 - 19)	16 (12 -25)	0.008	16 (12 - 24)	15 (12 -22)	0.31
Disease Duration (IQR)[1] -- yr	5 (1.5 - 11)	4 (1 - 9)	0.24ǂ	7 (2 - 13)	3.5 (1-8)	0.0004ǂ	4 (1 - 9)	5.5 (1 - 11)	0.06ǂ	5 (2 - 8)	4 (1 - 10)	0.78ǂ
Male sex -- no. (%)	81 (52.3%)	100 (49.2%)	0.57	57 (50.9%)	124 (50.4%)	0.93	93 (53.5%)	88 (48.6%)	0.46	25 (41.0%)	156 (53.5%)	0.10
NOD2 Positive -- no. (%)	12 (8.7%)	8 (4.7%)	0.16	4 (4.2%)	16 (7.5%)	0.28ǂ	8 (5.4%)	12 (7.5%)	0.45	2 (3.9%)	18 (7.0%)	0.41ǂ
African Admixture[1*]	0.80 (0.74 - 0.88)	0.84 (0.76 - 0.89)	0.41	0.84 (0.78 - 0.90)	0.82 (0.62 - 0.88)	0.06	0.84 (0.79 - 0.90)	0.82 (0.71 - 0.93)	0.004	0.83 (0.76 - 0.89)	9.83 (0.76 - 0.89)	0.80
Family History of IBD -- no. (%)	11 (7.9%)	16 (8.5%)	0.83	7 (6.9%)	20 (8.8%)	0.56	12 (7.3%)	15 (9.1%)	0.55	4 (7.0%)	23 (8.4%)	0.72ǂ
Autoimmune Disesase -- no. (%)	33 (21.5%)	63 (32.5%)	0.02	31 (28.4%)	65 (27.3%)	0.83	45 (26.3%)	51 (29.0%)	0.58	14 (24.6%)	82 (28.3%)	0.57
EIM -- no. (%)	32 (21.5%)	63 (32.5%)	0.02	33 (31.4%)	62 (26.1%)	0.31	43 (25.2%)	52 (30.2%)	0.30	24 (40.7%)	71 (29.1%)	0.01
Current Smoke Exposure -- no. (%)	35 (28.2%)*	49 (29.2%)	0.86	27 (29.0%)	57 (28.6%)	0.95	30 (20.6%)	54 (37.0%)	0.002	14 (27.5%)	70 (29.1%)	0.82
Upper Disease[3] -- no. (%)	50 (36.2%	63 (26.4%)	0.97	32 (33.7%)	81 (37.5%)	0.52	56 (37.3%)	57 (35.4%)	0.72	18 (33.3%)	95 (37.0%)	0.61
Perianal Disease -- no. (%)	78 (61.4%)	78 (51.0%)	0.08	59 (64.8%)	97 (51.3%)	0.03	85 (57.4%)	71 (53.8%)	0.54	18 (42.9%)	138 (58.0%)	0.07
Outcomes												
[4]Complicated Disease -- no.(%)	85 (54.8%)	64 (31.5%)	<0.0001	57 (50.9%)	92 (37.4%)	0.02	82 (46.3%)	67 (37.0%)	0.07	15 (24.6%)	134 (45.1%)	0.003
[5]Ileal Disease -- no.(%)	110 (81.5%)	103 (56.0%)	<0.0001	68 (68.0%)	145 (66.2%)	0.75	109 (66.1%)	104 (67.5%)	0.78	28 (50.1%)	185 (70.1%)	0.01
Isolated Ileal Disease -- no.(%)	35 (11.1%)	46 (14.7%)	0.35	24 (12.0%)	57 (13.7%)	0.68	27 (9.1%)	54 (17.5%)	0.03	12 (10.9%)	69 (13.6%)	0.59
Surgery -- no.(%)	72 (49.3%)	56 (32.4%)	0.002	56 (56.0%)	72 (32.9%)	<0.0001	62 (38.5%)	66 (41.7%)	0.55	16 (28.6%)	112 (42.6%)	0.05

ǂ: Fisher's exact test; [1]median (IQR); [2]ASCA positivity is defined as being either IgA or IgG ASCA positive; [3]upper Disease is defined as L4A, L4B or L4AB positivity; [4]Complicated disease is defined as B2 or B3 or B2B3 disease behavior; [5]ileal disease is defined as the (L1) isolate ileal or (L3) ilealcolonic disease

[*] Missing greater than 15% of the data

AA: African-American; CD: Crohn's disease; ASCA: anti-Saccharomyces cervisiae antibody; anti-OmpC: E.coli outer-membrane porin C; anti-Cbir1: flagellin-like bacterial antigen; pANCA: perinuclear anti-nuclear cytoplasmic antibody

Table 3: Univariate Logistic Regression for Complicated Disease, Ileal Disease and Disease Requiring Surgery

		Complicated				Any Ileal Disease				Surgery		
Covariate		**β_1**	**SE**	**P- Value**		**β_1**	**SE**	**P- Value**		**β_1**	**SE**	**P- Value**
Age at Diagnosis[1]		0.02	0.01	0.04								
					≤ 15 yrs	0.16	0.05	0.001	**≤ 25 yrs**	0.11	0.25	<0.0001
					> 15 yrs	-0.18	0.06	0.002	**> 25 yrs**	-0.13	0.04	0.0004
Disease Duration[1]						0.006	0.014	0.70				
	≤ 10 yrs	0.18	0.04	<0.0001					**≤ 10 yrs**	0.26	0.04	<0.0001
	> 10 yrs	-0.18	0.05	0.0002					**> 10 yrs**	-0.21	0.06	0.0006
Gender		0.15	0.21	0.94		-0.25	0.24	0.29		0.17	0.23	0.46
NOD2 Pos		0.97	0.48	0.05		2.14	1.04	0.04		0.86	0.47	0.07
Family History of IBD		0.44	0.4	0.27		0.82	0.51	0.12		0.30	0.42	0.48
History of Autoimmune Disease		0.09	0.24	0.7		-0.29	0.27	0.28		0.05	0.26	0.86
Extraintestinal Manifestations		0.12	0.24	0.63		-0.08	0.27	0.76		0.39	0.26	0.13
Current Smoker		0.08	0.26	0.75		-0.39	0.28	0.16		0.57	0.28	0.04
Ileal Involvement		1.30	0.27	<0.0001		--	--	--		0.98	0.28	0.0006
Colonic Involvement		0.87	0.34	0.01		--	--	--		0.6	0.34	0.08
Upper Disease Modifier		-0.30	0.24	0.22		0.62	0.28	0.02		-0.61	0.26	0.02
Perianal Disease		0.30	0.24	0.22		-0.41	0.27	0.13		0.05	0.26	0.84
African Admixture		-0.33	0.86	0.7		-1.05	1.02	0.31		0.33	0.92	0.72
Immune Response												
ASCA IgA[1]		0.02	0.004	<0.0001								
					≤ 40 EU/ml	0.05	0.01	<0.0001	**≤ 40 EU/ml**	0.03	0.009	0.0003
					> 40 EU/ml	-0.06	0.02	0.002	**> 40 EU/ml**	-0.04	0.01	0.01
ASCA IgG[1]						0.02	0.005	<0.0001				
	≤ 15 EU/ml	0.12	0.03	<0.0001					**≤ 20 EU/ml**	0.10	0.02	<0.0001
	> 15 Euml	-0.11	0.03	0.002					**> 20 EU/ml**	-0.10	0.02	<0.0001
anti-OmpC						-0.001	0.005	0.83				
	≤ 20 EU/ml	0.12	0.02	<0.0001					**≤ 20 EU/ml**	0.14	0.03	<0.0001
	> 20 EU/ml	-0.11	0.03	<0.0001					**> 20 EU/ml**	-0.15	0.03	<0.0001
anti-CBir1[1]						0.007	0.003	0.06		-0.0004	0.003	0.86
	≤ 20 EU/ml	0.07	0.03	0.009								
	> 20 EU/ml	-0.07	0.03	0.01								
P-ANCA		-0.02	0.006	0.006		-0.02	0.007	0.001		-0.01	0.006	0.08

SE: standard error; AA: African-American; CD: Crohn's disease; ASCA: anti-Saccharomyces cervisiae antibody; anti- OmpC: E.coli outer-membrane porin C; anti- Cbir1: flagellin-like bacterial antigen; pANCA: perinuclear anti-nuclear cytoplasmic antibody; OR: odds ratio; CI: confidence interval

[1] one-knot linear splines were created based on review of restricted cubic spline and subject matter knowledge

Table 4: Crude Odds Ratio for Complicated Disease, Ileal Disease and Disease Requiring Surgery

	Complicated				Any Ileal Disease				Surgery			
Covariate		**OR[2]**	**95% CI**	**P- Value**		**OR[2]**	**95% CI**	**P- Value**		**OR[2]**	**95% CI**	**P- Value**
Age at Diagnosis	Comparing 22.5 years to 12 years	1.22	1.01 , 1.47	0.04	Comparing 22.5 years to 12 years[1]	1.48	1.13 , 1.95	0.006	Comparing 22.5 years to 12 years[1]	2.87	1.77 , 4.65	0.0001
Disease Duration	Comparing 10 years to 1 year[1]	5.10	2.27 , 9.57	<0.0001	Comparing 10 years to 1 year	1.05	0.82 , 1.35	0.7	Comparing 10 years to 1 year[1]	10.11	21.39	<0.0001
Gender	Female to male	1.02	0.67 , 1.55	0.94	Female to male	0.78	0.49 , 1.24	0.29	Female to male	1.18	0.76 , 1.85	0.46
NOD2 Pos	Positive to negative	2.64	1.02 , 6.81	0.05	Positive to negative	8.53	1.11 , 65.35	0.04	Positive to negative	2.37	0.94 , 6.01	0.07
Family History of IBD	Positive to negative	1.55	0.71 , 3.42	0.27	Positive to negative	2.28	0.83 , 6.24	0.11	Positive to negative	1.35	0.58 , 3.13	0.48
History of Autoimmune Disease	Positive to negative	1.1	0.68 , 1.77	0.7	Positive to negative	0.75	0.45 , 1.27	0.28	Positive to negative	1.05	0.63 , 1.74	0.86
Extraintestinal Manifestations	Positive to negative	1.12	0.70 , 1.81	0.63	Positive to negative	0.92	0.54 , 1.56	0.76	Positive to negative	1.47	0.89 , 2.44	0.13
Current Smoker	Positive to negative	1.09	0.65 , 1.81	0.75	Positive to negative	0.68	0.39 , 1.17	0.16	Positive to negative	1.78	1.03 , 3.05	0.04
Ileal Involvement	Positive to negative	3.7	2.17 , 6.25	<0.0001		--	--	--	Positive to negative	2.7	1.54 , 4.76	0.0006
Colonic Involvement	Positive to negative	0.42	0.21 , 0.81	0.01		--	--	--	Positive to negative	0.55	0.28 , 1.08	0.08
Upper Disease Modifier	Positive to negative	0.74	0.46 , 1.19	0.22	Positive to negative	1.86	1.08 , 3.2	0.02	Positive to negative	0.54	0.33 , 0.91	0.02
Perianal Disease	Positive to negative	0.74	0.46 , 1.19	0.22	Positive to negative	1.5	0.89 , 2.55	0.13	Positive to negative	0.98	0.58 , 1.57	0.84
% African Admixture	Comparing 9.0% to 8.0%	0.96	0.77 , 1.20	0.7	Comparing 9.0% to 8.0%	0.87	0.67 , 1.13	0.31	Comparing 9.0% to 8.0%	1.05	0.82 , 1.33	0.72
Immune Response												
ASCA IgA	Comparing 26.7 EU/ml to 2.3 EU/ml	1.86	1.47 , 2.36	<0.0001	Comparing 26.7 EU/ml to 2.3 EU/ml[1]	6.01	2.64 , 13.65	<0.0001	Comparing 26.7 EU/ml to 2.3 EU/ml[1]	3.19	1.70 , 6.00	< 0.0001
ASCA IgG	Comparing 53.5 EU/ml to 5.9 EU/ml[1]	3.83	2.46 , 5.95	<0.0001	Comparing 53.5 EU/ml to 5.9 EU/ml	3.21	2.02 , 5.08	<0.0001	Comparing 53.5 EU/ml to 5.9 EU/ml[1]	3.3	2.02 , 5.39	< 0.0001
anti-OmpC	Comparing 27.7 EU/ml to 9.9 EU/ml[1]	2.77	1.81 , 4.24	<0.0001	Comparing 27.7 EU/ml to 9.9 EU/ml	0.98	0.83 , 1.16	0.83	Comparing 27.7 EU/ml to 9.9 EU/ml[1]	4.17	2.54 , 6.85	<0.0001
anti-CBir1	Comparing 46.1 EU/ml to 14.4 EU/ml[1]	1.54	1.16 , 2.04	0.01	Comparing 46.1 EU/ml to 14.4 EU/ml	1.23	0.99 , 1.52	0.06	Comparing 46.1 EU/ml to 14.4 EU/ml	0.99	0.84 , 1.16	0.86
pANCA	Comparing 23.4 EU/ml to 10.3 EU/ml	0.79	0.67 , 0.93	0.006	Comparing 23.4 EU/ml to 10.3 EU/ml	0.76	0.64 , 0.89	0.001	Comparing 23.4 EU/ml to 10.3 EU/ml	0.87	0.75 , 1.02	0.08

AA: African-American; CD: Crohn's disease; ASCA: anti-Saccharomyces cervisiae antibody; anti-OmpC: E.coli outer-membrane porin C; anti-Cbir1: flagellin-like bacterial antigen; pANCA: perinuclear anti-nuclear cytoplasmic antibody; OR: odds ratio; CI: confidence interval

[1] one-knot linear splines were created based on review of restricted cubic spline and subject matter knowledge; [2] OR for cotinuous variables were calculated by comparing the 75%ile to the 25%ile

Table 5: Adjusted Odds Ratio for Complicated Crohn's Disease

		OR[2]	95% CI
Disease Duration[1]	Comparing 10 years to 1 year	6.44	3.18 , 13.04
IgG ASCA[1]	Comparing 53.5 EU/ml to 5.9 EU/ml	3.21	1.98 , 5.21
Ileal Disease	Positive to negative	3.23	1.79 , 5.88

ASCA: anti-Saccharomyces cervisiae antibody; OR: Odds Ratio; CI: Confidence Interval

[1] one-knot linear splines were created based on review of restricted cubic spline and subject matter knowledge

[2] OR were calculated by comparing the 75%ile to the 25%ile

Table 6: Adjusted Odds Ratio for the Presence of Ileal Disease

		OR[2]	95% CI
Perianal Disease	Positive to negative	2.83	1.50 , 5.33
Upper Disease	Positive to negative	2.33	1.26 , 4.29
IgA ASCA[1]	Comparing 26.7 EU/ml to 2.3 EU/ml	3.42	1.30 , 9.02
IgG ASCA	Comparing 53.5 EU/ml to 5.9 EU/ml	2.42	1.33 , 4.39
pANCA	Comparing 23.4 EU/ml to 10.3 EU/ml	0.81	0.68 , 0.97

ASCA: anti-Saccharomyces cervisiae antibody; pANCA: perinuclear anti-nuclear; OR: Odds Ratio; CI: Confidence Interval

[1] one-knot linear splines were created based on review of restricted cubic spline and subject matter knowledge

[2] OR were calculated by comparing the 75%ile to the 25%ile

Table 7: Adjusted Odds Ratio for Disease Requiring Surgery

		OR[2]	95% CI
Nod 2 Positive	Positive to negative	3.27	1.07 , 9.96
Disease Duration[1]	Comparing 10 years to 1 year	6.80	3.08 , 15.05
IgG ASCA[1]	Comparing 53.5 EU/ml to 5.9 EU/ml	3.13	1.72 , 5.69
anti-OmpC[1]	Comparing 27.7 EU/ml to 9.9 EU/ml	2.33	1.28 , 4.25

ASCA: anti-Saccharomyces cervisiae antibody; anti-OmpC: E.coli outer-membrane porin C; OR: Odds Ratio; CI: Confidence Interval

[1] one-knot linear splines were created based on review of restricted cubic spline and subject matter knowledge

[2] OR were calculated by comparing the 75%ile to the 25%ile

Figure 1: Sample selection for analysis of serological markers in AA with CD

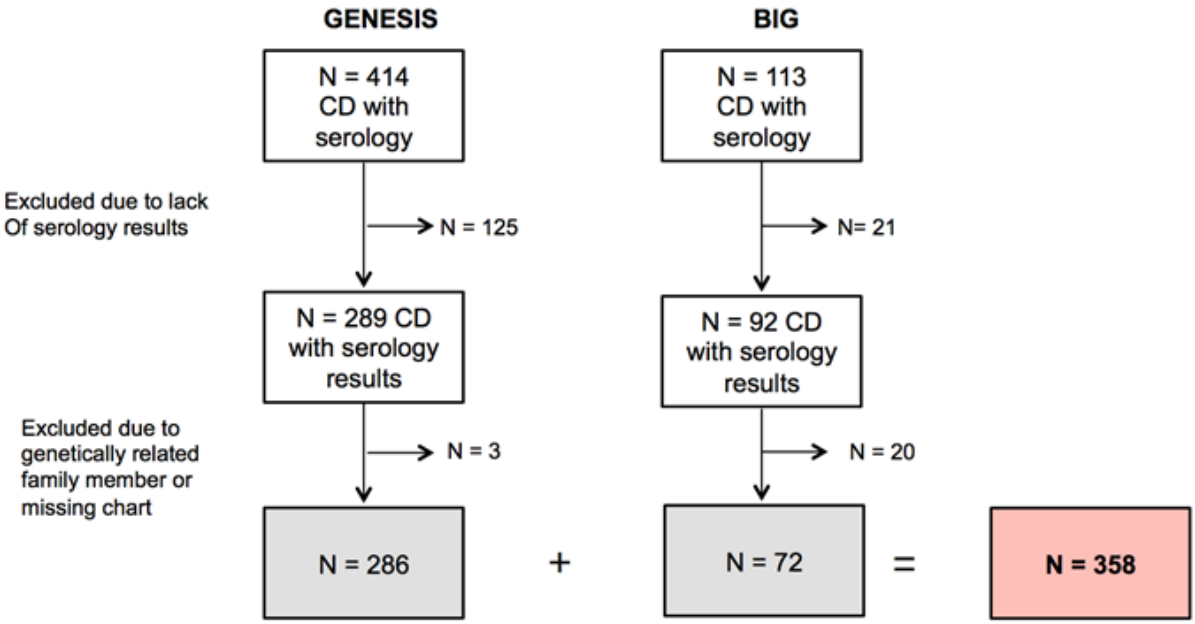


Figure 2: One-knot linear spline selection for the outcome "complicated disease"

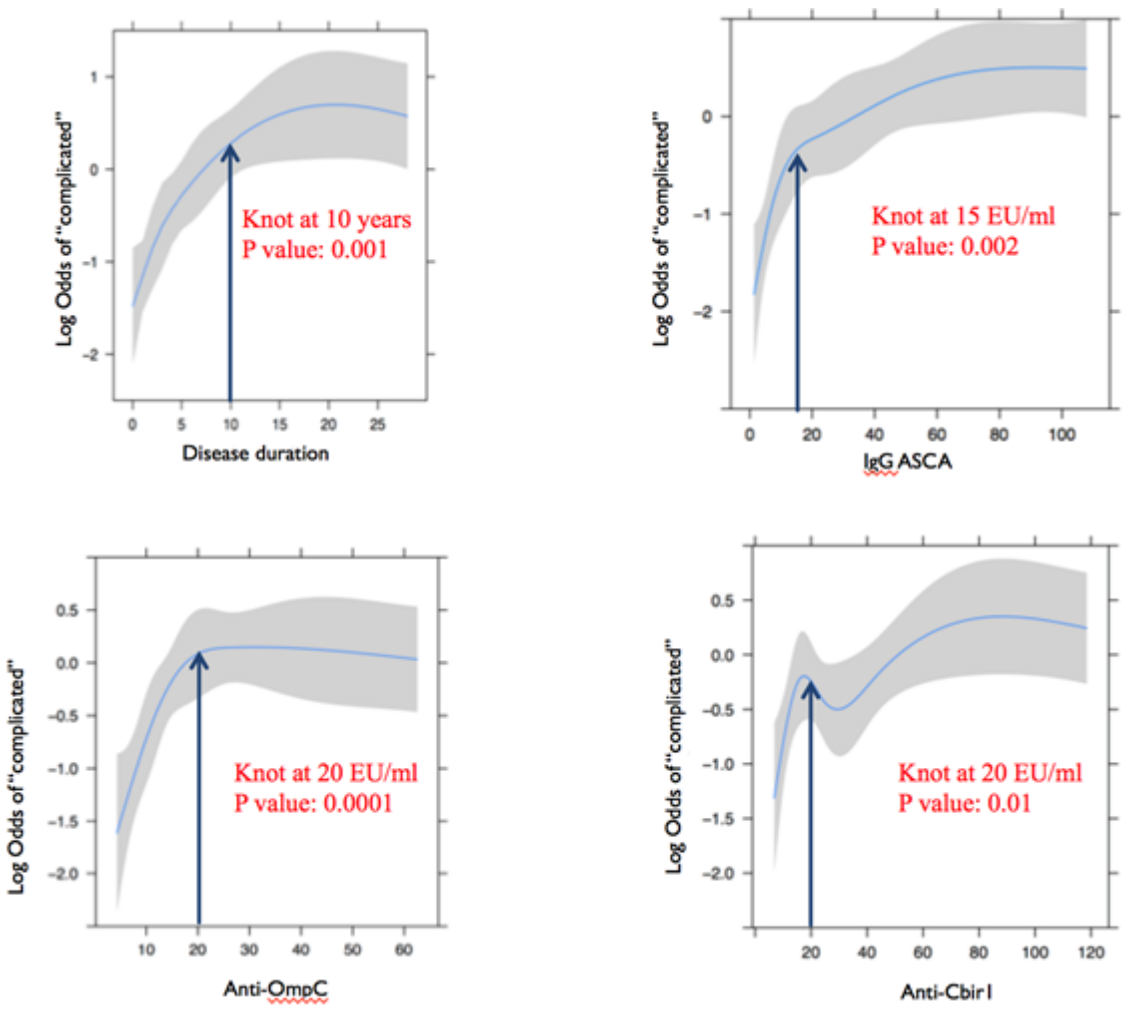


Figure 3: One-knot linear spline selection for the outcome "ileal disease"

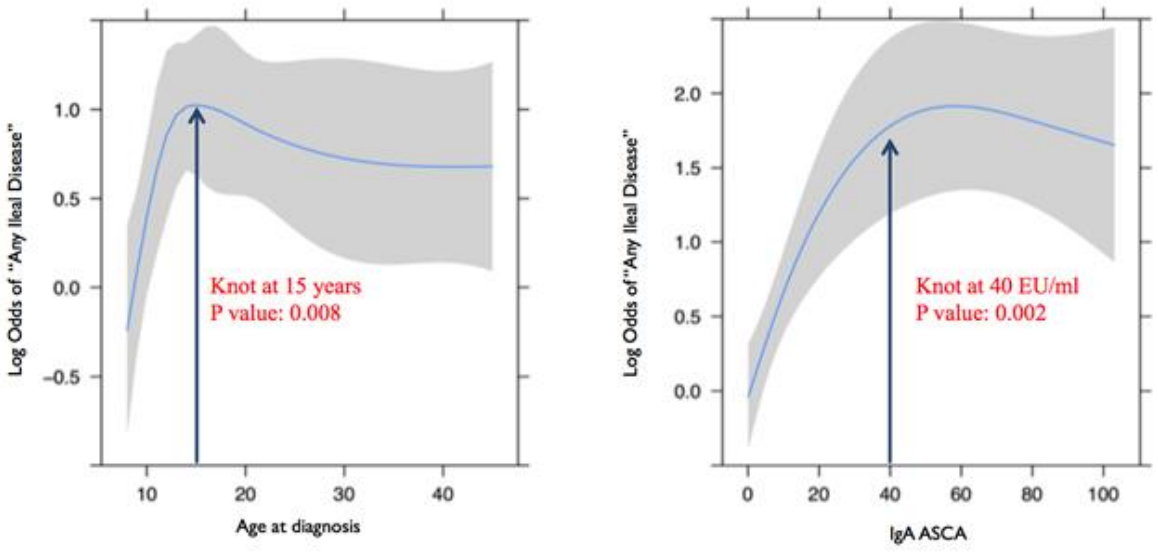
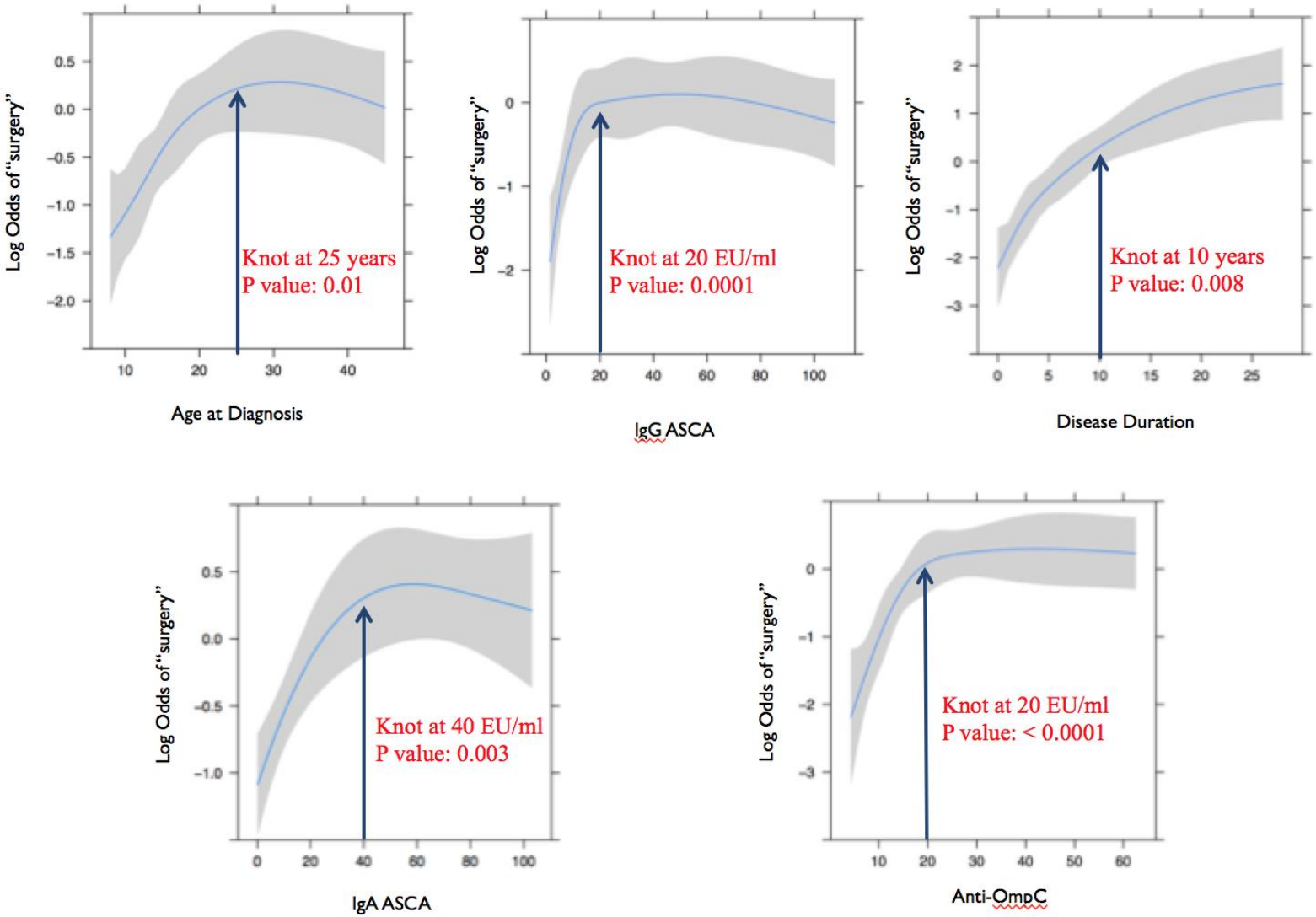


Figure 4: One-knot linear spline selection for the outcome "surgery"



APPENDIX:

Table 1: Adjusted Odds Ratio for Disease Requiring Surgery after Controlling for Potential Confounders

		OR[2]	95% CI
Age at Diagnosis	Comparing 22.5 years to 12 years	1.33	0.73 , 2.43
Ileal Disease	Positive to negative	2.44	1.15 , 5.26
Family History	Positive to negative	0.84	0.29 , 2.41
Current Smoker	Positive to negative	1.18	0.60 , 2.30
Nod2 Positive	Positive to negative	2.61	0.81 , 8.39
Disease Duration[1]	Comparing 10 years to 1 year	6.95	3.06 , 15.81
IgG ASCA[1]	Comparing 53.5 EU/ml to 5.9 EU/ml	2.54	1.33 , 4.84
anti-OmpC[1]	Comparing 27.7 EU/ml to 9.9 EU/ml	2.25	1.20 , 4.20

ASCA: anti-Saccharomyces cervisiae antibody; anti-OmpC: E.coli outer-membrane porin C; OR: Odds Ratio; CI: Confidence Interval

[1] one-knot linear splines were created based on review of restricted cubic spline and subject matter knowledge

[2] OR were calculated by comparing the 75%ile to the 25%ile

Table 2: Distribution of Missing Data

Variable	% Missing
% African Admixture	25.42%
Age at Diagnosis (IQR) -- yr	1.68%
Disease Duration (IQR) -- yr	1.40%
Male sex -- no. (%)	0.00%
Family History of IBD -- no. (%)	7.82%
History of Autoimmune Disorder -- no. (%)	3.07%
Extraintestinal Manifestations -- no. (%)	4.19%
Smoking -- no. (%)	
Current	11.17%
Ever	18.44%
Disease Location -- no. (%)	10.89%
Upper Disease Modifier-- no. (%)	13.13%
Complicated Disease -- no. (%)	0.00%
Perianal Disease -- no. (%)	21.79%
Surgical History -- no. (%)	10.89%
Biologic Use -- no. (%)	
At diagnosis	19.27%
Ever	8.38%
Nod2 Positivity	13.97%
Serological Response (level)	
ASCA IgA	0.00%
ASCA IgG	0.00%
anti-OmpC	0.00%
anti-CBir1	0.00%
pANCA	0.00%

ASCA: anti-Saccharomyces cervisiae antibody; anti-OmpC: E.coli outer-membrane porin C; anti-Cbir1: flagellin-like bacterial antigen; pANCA: perinuclear anti-nuclear cytoplasmic antibody