Norma Pugh
October 24, 2008

Integrity Check for the Inflammatory Bowel Disease Genetic Consortium (IBDGC) Phenotype Analysis File

As a partial check of the integrity of the IBDGC phenotype analysis dataset archived in the NIDDK data repository, a set of tabulations was performed to verify that published results can be reproduced using the archived dataset. Analyses were performed to duplicate results for the data published by Nguyen et al [1] in the *American Journal of Gastroenterology* in May 2006. The results of this integrity check are described below. The full text of the *American Journal of Gastroenterology* article can be found in Attachment 1, and the SAS code for our tabulations is included in Attachment 2. Attachment 3 includes confirmatory Stata code provided by the DCC, and Attachment 4 provides a complete description of the genotypic data. The genotypic data have not been replicated, as they are not included in the NIDDK Data Repository at RTI. They are housed in the NIDDK Genetics Repository at Rutgers.

The intent of this DSIC is to provide confidence that the data distributed by the NIDDK repository is a true copy of the study data. Our intent is ***not*** to assess the integrity of the statistical analyses reported by study investigators. As with all statistical analyses of complex datasets, complete replication of a set of statistical results should not be expected on a first (or second) exercise in secondary analysis. This occurs for a number of reasons including differences in the handling of missing data, restrictions on cases included in samples for a particular analysis, software coding used to define complex variables, etc. Experience suggests that most discrepancies can ordinarily be resolved by consultation with the study data coordinating center (DCC), however this process is labor-intensive for both DCC and Repository staff. It is thus not our policy to resolve every discrepancy that is observed in an integrity check. Specifically, we do not attempt to resolve minor or inconsequential discrepancies with published results or discrepancies that involve complex analyses, *unless NIDDK Repository staff suspect that the observed discrepancy suggests that the dataset may have been corrupted in storage, transmission, or processing by repository staff*. We do, however, document in footnotes to the integrity check those instances in which our secondary analyses produced results that were not fully consistent with those reported in the target publication.

Background. The NIDDK Inflammatory Bowel Disease Genetics Consortium (IBDGC) consists of investigators from seven sites in the U.S. and Canada who have recruited a large sample of inflammatory bowel disease patients, their relatives, and control subjects. All of the individuals in this sample have been evaluated according to a standardized protocol for clinical traits related to IBD, and have donated blood samples as a source of DNA. The IBDGC investigators are conducting genetic linkage and association studies to identify genes influencing predisposition to IBD [2].

The Nguyen paper examines racial differences in Inflammatory Bowel Disease (IBD) in family history, disease location, and extraintestinal manifestations (EIMs) that may reflect underlying genetic variations and have important implications for diagnosis and management of the disease [1].

The IBDGC analysis file provided to the repository was the most current version at the time of delivery. It is not the version that was used to produce the published results. In order to perform this replication, the repository was provided a supplemental list of the patient IDs analyzed in the publication.

The DCC indicated that one patient requested to be removed from the repository. In addition, two others either had no samples collected (or there was a problem in sample processing), and so are not part of the repository. This accounts for the total of 1,123 patients analyzed in this replication versus the total of 1,126 patients analyzed in the publication. Additionally, per the publication, the replication analysis was restricted to all unrelated index subjects (‘control’ variable = 0) with a confirmed diagnosis of IBD (‘diag’

variable = 1, 2 or 3) and self-identified race of African American, Hispanic, or Non-Hispanic White ('hispanic' and 'race' variables).

Finally, Attachment 2 provides the SAS code that was used by the data repository for the replication analysis. The DCC has reviewed and confirmed that the appropriate variables were used. Attachment 3 documents the DCC-provided Stata code used to clarify the definition of various variables and aid in the resolution of discrepancies. The Stata analysis was run using a copy of the data housed at the Data Repository.

Demographic and Baseline Characteristics. Table 1 in the publication [1] reports on demographic and baseline characteristics. Table A lists the variables we used in our replication.

Table A: Variables Used to Replicate Table 1

Table Variable	Variables Used in Replication
Race distribution	hispanic, race
Age at study entry	yob
Age at diagnosis	diag_yr, yob
Female	sex=2
Diagnosis: CD	diag=1
Diagnosis: UC	diag=3
Diagnosis: Indeterminate colitis	diag=2
Family history of IBD	chld_cd>0 or chld_uc>0 or chld_ibd>0 or fthr_ibd in(1,2,3) or mthr_ibd in(1,2,3) or sib_cd>0 or sib_uc>0 or sib_ibd>0 or fam_hist=1
Family history of IBD among CD patients	Family history (as defined above) and diag=1
Family history of IBD among UC patients	Family history (as defined above) and diag=3
% Siblings affected	sib_cd, sib_uc, sib_ibd, sib_unf
Packs/day at diagnosis	no_cigar/20, where smoking=1
Smoking at diagnosis: Current	smoking=1
Smoking at diagnosis: Former	smoking=2
Smoking at diagnosis: Never	smoking=3
Appendectomy (2 yr prior to IBD)	diag_yr, app_yr
Note: 'age at study entry' is not collected on the phenotype forms, and therefore is not included in the analysis dataset. Per advice from the DCC, an approximate age was calculated by subtracting 'year of birth' from 2005. Total patient enrollment took place from 2003-2007.	

In Table B, we compare the results calculated from the archived dataset to the results published in Table 1, Patient Demographics by Race/Ethnicity of the NIDDK-IBDGC Repository. As Table B shows, the results are similar.

Table B: Comparison of Values Computed in Integrity Check to Reference Article Table 1 Values

Table	Group: All			Group: White			Group: AA			Group: Hispanic		
Variable	**Nguyen et al (2006)**	**Integrity Check**	**Diff**	**Nguyen et al (2006)**	**Integrity Check**	**Diff**	**Nguyen et al (2006)**	**Integrity Check**	**Diff**	**Nguyen et al (2006)**	**Integrity Check**	**Diff**
Race	1,126	1,123	3	830	829	1	127	126	1	169	168	1
Age at study entry in yr, mean (SD)	36.1 (14.4)	37.1 (14.4)	1 (0)	36.7 (14.9)	37.7 (14.9)	1 (0)	36.1 (14.6)	37.1 (14.5)	1.0 (0.1)	33.3 (10.6)	34.2 (10.6)	0.9 (0)
Age at diagnosis in yr, mean (SD)	26.4 (12.8)	26.3 (12.8)	0.1 (0)	26.6 (13.3)	26.6 (13.3)	0	25.4 (13.2)	25.4 (13.2)	0	25.9 (9.8)	25.9 (9.8)	0
Female, no. (%)	584 (51.9)	584 (52.0)	0 (0.1)	427 (51.5)	428 (51.6)	1 (0)	80 (63.0)	80 (63.5)	0 (0.5)	77 (45.6)	76 (45.2)	1 (0.4)
Diagnosis, no. (%):												
CD	697 (61.9)	695 (61.9)	2 (0)	510 (61.5)	508 (61.3)	2 (0.2)	81 (63.8)	80 (63.5)	1 (0.3)	106 (62.7)	107 (63.7)	1 (1.0)
UC	396 (35.2)	397 (35.4)	1 (0.2)	299 (36.0)	302 (36.4)	3 (0.4)	35 (27.6)	35 (27.8)	0 (0.2)	62 (36.7)	60 (35.7)	2 (1.0)
IC	33 (2.9)	31 (2.8)	2 (0.1)	21 (2.5)	19 (2.3)	2 (0.2)	11 (8.7)	11 (8.7)	0	1 (0.6)	1 (0.6)	0
Fam Hx IBD	288 (25.6)	288 (25.6)	0	237 (28.6)	237 (28.6)	0	23 (18.1)	23 (18.3)	0 (0.2)	28 (16.6)	28 (16.7)	0 (0.1)
CD patients	191 (27.4)	191 (27.5)	0 (0.1)	156 (30.6)	156 (30.7)	0 (0.1)	16 (19.8)	16 (20.0)	0 (0.2)	19 (17.9)	19 (17.8)	0 (0.1)
UC patients	88 (22.2)	88 (22.2)	0	74 (24.8)	74 (24.5)	0 (0.3)	5 (14.3)	5 (14.3)	0	9 (14.5)	9 (15.0)	0 (0.5)
% Siblings	3.8	3.8	0	4.6	4.6	0	2.5	2.5	0	1.3	1.3	0
Packs/day at diagnosis	0.12	0.12	0	0.15	0.15	0	0.07	0.07	0	0.04	0.04	0

Table B: Comparison of Values Computed in Integrity Check to Reference Article Table 1 Values (cont.)

Table	Group: All			Group: White			Group: AA			Group: Hispanic		
Variable	**Nguyen et al (2006)**	**Integrity Check**	**Diff**	**Nguyen et al (2006)**	**Integrity Check**	**Diff**	**Nguyen et al (2006)**	**Integrity Check**	**Diff**	**Nguyen et al (2006)**	**Integrity Check**	**Diff**
Smoking at diagnosis, no. (%):												
Current	192 (17.2)	193 (17.2)	1 (0)	161 (19.5)	162 (19.5)	1 (0)	20 (16.1)	20 (15.9)	0 (0.2)	11 (6.6)	11 (6.6)	0
Former	140 (12.6)	138 (12.3)	2 (0.3)	117 (14.2)	116 (14.0)	1 (0.2)	15 (12.1)	15 (11.9)	0 (0.2)	8 (4.8)	7 (4.2)	1 (0.6)
Never	783 (70.2)	781 (69.6)	2 (0.6)	547 (66.3)	546 (65.9)	1 (0.4)	89 (71.8)	88 (69.8)	1 (2.0)	147 (88.6)	147 (87.5)	0 (1.1)
Appendectomy (2 yr prior to IBD)	47 (4.4)	47 (4.2)	0 (0.2)	32 (4.0)	32 (3.9)	0 (0.1)	3 (2.9)	3 (2.4)	0 (0.5)	12 (7.1)	12 (7.1)	0

Notes:
(1) AA = African American, CD = Crohn's disease, UC = ulcerative colitis, IC = indeterminate colitis, IBD = inflammatory bowel disease

CD and UC Phenotype. Table 2 in the publication [1] reports on the unadjusted distribution of CD location and behavioral pattern, as well as UC disease extent. Table C lists the variables we used in our replication.

Table C: Variables Used to Replicate Table 2

Table Variable	Variables Used in Replication
Race distribution	hispanic, race
Crohn's disease: Disease Involvement: Upper gastrointestinal	diag=1 and (jejunal=1 or gi=1)
Crohn's disease: Disease Involvement: Esophagogastroduodenal	diag=1 and gi=1
Crohn's disease: Disease Involvement: Jejunum	diag=1 and jejunal=1
Crohn's disease: Disease Involvement: Ileum	diag=1 and ileal=1
Crohn's disease: Disease Involvement: Colorectal	diag=1 and colorect=1
Crohn's disease: Disease Involvement: Perianal disease	diag=1 and perianal=1
Best-fit disease site: Ileum	*Note:*
Best-fit disease site: Ileo-colon	*'Best-fit' items have not been replicated. They are*
Best-fit disease site: Colon	*generated by other variables and differ slightly from the*
Best-fit disease site: Upper gastrointestinal	*Montreal classification.*
Disease pattern: Inflammatory	diag=1 and behavior=1
Disease pattern: Stricturing	diag=1 and behavior=2
Disease pattern: Penetrating	diag=1 and behavior=3
UC Disease extent: Proctitis	diag=3 and (proctit=1 & left=2 & extensiv=2)
UC Disease extent: Left-sided colitis	else diag=3 and (left=1 & extensiv=2)
UC Disease extent: Extensive colitis	else diag=3 and extensiv=1

In Table D, we compare the results calculated from the archived dataset to the results published in Table 2, CD and UC Phenotype by Race/Ethnicity of the NIDDK-IBDGC Repository. As Table D shows, the results are similar.

Table D: Comparison of Values Computed in Integrity Check to Reference Article Table 2 Values

Table	**Group: All**			**Group: White**			**Group: AA**			**Group: Hispanic**		
Variable	**Nguyen et al (2006)**	**Integrity Check**	**Diff**	**Nguyen et al (2006)**	**Integrity Check**	**Diff**	**Nguyen et al (2006)**	**Integrity Check**	**Diff**	**Nguyen et al (2006)**	**Integrity Check**	**Diff**
Crohn's disease												
Disease involvement												
Upper gastrointestinal	96 (15.8)	96 (15.8)	0	72 (16.1)	72 (16.1)	0	17 (25.8)	17 (26.2)	0 (0.4)	7 (7.4)	7 (7.3)	0 (0.1)
Esophagogastroduodenal	54 (8.8)	55 (9.0)	1 (0.2)	38 (8.5)	39 (8.7)	1 (0.2)	14 (20.0)	14 (20.3)	0 (0.3)	2 (2.1)	2 (2.1)	0
Jejunum	57 (9.1)	56 (8.9)	1 (0.2)	44 (9.4)	43 (9.1)	1 (0.3)	8 (12.9)	8 (13.1)	0 (0.2)	5 (5.2)	5 (5.1)	0 (0.1)
Ileum	530 (79.0)	530 (79.2)	0 (0.2)	396 (80.0)	395 (80.1)	1 (0.1)	50 (67.6)	50 (68.5)	0 (0.9)	84 (82.4)	85 (82.5)	1 (0.1)
Colorectal	459 (68.3)	458 (68.2)	1 (0.1)	326 (65.6)	325 (65.4)	1 (0.2)	63 (77.8)	62 (77.5)	1 (0.3)	70 (74.5)	71 (74.7)	1 (0.2)
Perianal disease	231 (33.5)	227 (33.0)	4 (0.5)	146 (28.7)	142 (28.1)	4 (0.6)	32 (40.0)	32 (40.5)	0 (0.5)	53 (52.5)	53 (52.0)	0 (0.5)
Disease pattern												
Inflammatory	332 (48.2)	332 (48.1)	0 (0.1)	242 (48.0)	242 (47.9)	0 (0.1)	42 (53.2)	41 (52.6)	1 (0.6)	48 (45.3)	49 (45.8)	1 (0.5)
Stricturing	155 (22.5)	154 (22.3)	1 (0.2)	108 (21.4)	108 (21.4)	0	22 (27.9)	21 (26.9)	1 (1)	25 (23.6)	25 (23.4)	0 (0.2)
Penetrating	202 (29.3)	204 (29.6)	2 (0.3)	154 (30.6)	155 (30.7)	1 (0.1)	15 (19.0)	16 (20.5)	1 (1.5)	33 (31.1)	33 (30.8)	0 (0.3)

Table D: Comparison of Values Computed in Integrity Check to Reference Article Table 2 Values (cont)

Table	Group: All			Group: White			Group: AA			Group: Hispanic		
Variable	**Nguyen et al (2006)**	**Integrity Check**	**Diff**	**Nguyen et al (2006)**	**Integrity Check**	**Diff**	**Nguyen et al (2006)**	**Integrity Check**	**Diff**	**Nguyen et al (2006)**	**Integrity Check**	**Diff**
Ulcerative Colitis												
Disease extent												
Proctitis	23 (6.2)	24(6.4)	1 (0.2)	17 (5.9)	18 (6.2)	1 (0.3)	4 (13.8)	4 (13.8)	0	2 (3.6)	2 (3.8)	0 (0.2)
Left-sided colitis	109 (29.4)	109 (29.2)	0 (0.2)	90 (31.4)	90 (30.9)	0 (0.5)	10 (34.5)	10 (34.5)	0	9 (16.4)	9 (17.0)	0 (0.6)
Extensive colitis	239 (64.4)	240 (64.3)	1 (0.1)	180 (62.7)	183 (62.9)	3 (0.2)	15 (51.7)	15 (51.7)	0	44 (80.0)	42 (79.2)	2 (0.8)
Notes: (1) AA = African American												

Surgical History. Table 3 in the publication [1] reports on the surgical history of IBD patients. Table E lists the variables we used in our replication of these variables.

Table E: Variables Used to Replicate Table 3

Table Variable	Variables Used in Replication
Race distribution	hispanic, race
CD: # surgeries: abdominal	op_ad
CD: # surgeries: perianal	op_pd
CD: # surgeries: bowel resection	surg_br=1
CD: # surgeries: bowel diversion	surg_div=1
CD: # surgeries: penetrating disease	surg_af=1
CD: # surgeries: perianal disease	surg_pf=1
Time to first surgery	surgery, diag_yr, review, surg_yr
% surgery free: at 2 years	surgery, diag_yr, review, surg_yr
% surgery free: at 5 years	surgery, diag_yr, review, surg_yr
% surgery free: at 10 years	surgery, diag_yr, review, surg_yr
UC: Colectomy	surgery=1 and diag=3
UC: Dysplasia	surg_dys=1
UC: Chronic refractory disease	surg_chr=1
UC: Fulminant colitis	surg_acu=1 and diag=3

In Table F, we compare the results calculated from the archived dataset to the results published in Table 3, Surgical History of IBD Patients in the NIDDK-IBDGC Repository by Race/Ethnicity. As Table F shows, the results are similar. The DCC provided confirmatory Stata code (Attachment 3) in order to clarify the definition of various variables and aid in the resolution of discrepancies. Specifically, 'time to first surgery' and the '% surgery free at *xx* years' variables have been confirmed by the Stata code.

Additionally, the DCC has noted that published results include approximately 50 observations in which the 'number of surgery' variables (abdominal and perianal) should not have been answered. These observations represent cases where the overall surgery variable is not recorded as 'yes'. Therefore, it is unclear how the 'number of surgery' responses should be interpreted and these observations have been excluded from the archived dataset.

Finally, variables labeled as "# surgeries" in the publication are not always "# surgeries". Yes/No surgery variables are actually summarized for bowel resection, bowel diversion, penetrating disease and perianal disease.

Table F: Comparison of Values Computed in Integrity Check to Reference Article Table 3 Values

Table	Group: All			Group: White			Group: AA			Group: Hispanic		
Variable	**Nguyen et al (2006)**	**Integrity Check**	**Diff**	**Nguyen et al (2006)**	**Integrity Check**	**Diff**	**Nguyen et al (2006)**	**Integrity Check**	**Diff**	**Nguyen et al (2006)**	**Integrity Check**	**Diff**
CD												
No. surgeries, mean (SD)												
For abdominal CD	1.4 (1.3)	1.5 (1.2)	0.1 (0.1)	1.6 (1.3)	1.6 (1.3)	0	0.8 (1.1)	1.4 (1.2)	0.6 (0.1)	1.1 (1.0)	1.3 (1.0)	0.2 (0)
For perianal CD	0.5 (1.2)	0.6 (1.2)	0.1 (0)	0.5 (1.1)	0.5 (1.1)	0	0.3 (0.7)	0.6 (0.9)	0.3 (0.2)	0.7 (1.7)	0.8 (1.9)	0.1 (0.2)
Bowel resection	335 (71.0)	334 (71.8)	1 (0.8)	263 (78.7)	262 (76.2)	1 (2.5)	27 (35.1)	27 (58.7)	0 (23.6)	45 (73.8)	45 (60.0)	0 (13.8)
Bowel diversion	46 (10.0)	46 (10.1)	0 (0.1)	24 (7.4)	24 (7.1)	0 (0.3)	9 (12.0)	9 (20.0)	0 (8.0)	13 (22.4)	13 (17.6)	0 (4.8)
Surgery for penetrating disease	98 (21.5)	100 (22.1)	2 (0.6)	74 (22.9)	76 (22.7)	2 (0.2)	10 (13.3)	10 (22.2)	0 (8.9)	14 (24.1)	14 (19.2)	0 (4.9)
Surgery for perianal disease	107 (23.0)	104 (22.5)	3 (0.5)	77 (23.4)	76 (22.3)	1 (1.1)	15 (20.0)	15 (33.3)	0 (13.3)	15 (24.6)	13 (17.1)	2 (7.5)
Median time to 1st surgery (yr)	9.5	See Attachment 3		9.8	See Attachment 3		11.4	See Attachment 3		6.6	See Attachment 3	
% Surgery-free												
At 2 yr	86.1%			84.6%			94.8%			86.7%		
At 5 yr	67.7%			67.7%			76.3%			59.8%		
At 10 yr	48.0%			49.6%			58.8%			26.1%		

Table F: Comparison of Values Computed in Integrity Check to Reference Article Table 3 Values (cont.)

Table	Group: All			Group: White			Group: AA			Group: Hispanic		
Variable	**Nguyen et al (2006)**	**Integrity Check**	**Diff**	**Nguyen et al (2006)**	**Integrity Check**	**Diff**	**Nguyen et al (2006)**	**Integrity Check**	**Diff**	**Nguyen et al (2006)**	**Integrity Check**	**Diff**
Ulcerative colitis												
Colectomy	71 (18.0)	70 (17.6)	1 (0.4)	47 (15.8)	47 (15.6)	0 (0.2)	4 (11.4)	4 (11.4)	0	20 (32.3)	19 (31.7)	1 (0.6)
Dysplasia	4 (3.0)	4 (5.2)	0 (2.2)	3 (3.5)	3 (5.9)	0 (2.4)	1 (3.7)	1 (16.7)	0 (13.0)	0 (0)	0 (0)	0
Chronic refractory disease	54 (40.6)	53 (69.7)	1 (29.1)	35 (40.7)	35 (68.6)	0 (27.9)	1 (3.9)	1 (20.0)	2 (16.1)	18 (85.7)	17 (85.0)	1 (0.7)
Fulminant colitis	8 (5.9)	8 (10.5)	0 (4.6)	6 (6.9)	6 (12.0)	0 (5.1)	0 (0)	0 (0)	0	2 (9.5)	2 (10.0)	0 (0.5)
Notes: (1) AA = African American, CD = Crohn's disease												

Extraintestinal Manifestations. Table 4 in the publication [1] reports on the distribution of EIMs in all patients. Table G lists the variables we used in our replication of these variables.

Table G: Variables Used to Replicate Table 4

Table Variable	Variables Used in Replication
Race distribution	hispanic, race
Sacroiliitis	j_si=1
Ankylosing spondylitis	j_as=1
Uveitis	eye_uv=1
Erythma nodosum	skin_en=1
Pyoderma gangrenosum	skin_py=1
Primary sclerosing cholangitis	liv_psc=1

In Table H, we compare the results calculated from the archived dataset to the results published in Table 4, Extraintestinal Manifestations in the NIDDK-IBDGC Repository by Race/Ethnicity. As Table H shows, the results are similar.

Table H: Comparison of Values Computed in Integrity Check to Reference Article Table 4 Values

Table	Group: All			Group: White			Group: AA			Group: Hispanic		
Variable	**Nguyen et al (2006)**	**Integrity Check**	**Diff**	**Nguyen et al (2006)**	**Integrity Check**	**Diff**	**Nguyen et al (2006)**	**Integrity Check**	**Diff**	**Nguyen et al (2006)**	**Integrity Check**	**Diff**
Sacroiliitis	23 (2.1)	23 (2.1)	0	14 (1.7)	16 (2.0)	2 (0.3)	7 (5.8)	6 (5.3)	1 (0.5)	2 (1.2)	1 (0.6)	1 (0.6)
Ankylosing spondylitis	14 (1.3)	12 (1.1)	2 (0.2)	12 (1.5)	11 (1.4)	1 (0.1)	2 (1.7)	1 (0.9)	1 (0.8)	0 (0)	0 (0)	0
Uveitis	25 (2.2)	24 (2.2)	1 (0)	13 (1.6)	13 (1.6)	0	10 (8.1)	9 (8.0)	1 (0.1)	2 (1.2)	2 (1.2)	0
Erythema nodosum	48 (4.3)	46 (4.3)	2 (0)	28 (3.4)	29 (3.6)	1 (0.2)	2 (1.7)	2 (1.8)	0 (0.1)	18 (10.7)	15 (9.3)	3 (1.4)
Pyoderma gangrenosum	23 (2.1)	22 (2.0)	1 (0.1)	16 (1.9)	16 (2.0)	0 (0.1)	3 (2.5)	3 (2.7)	0 (0.2)	4 (2.4)	3 (1.9)	1 (0.5)
Primary sclerosing cholangitis	20 (1.8)	20 (1.9)	0 (0.1)	11 (1.3)	11 (1.4)	0 (0.1)	4 (3.2)	4 (3.5)	0 (0.3)	5 (3.0)	5 (3.1)	0 (0.1)
Note: AA = African American												

Notes

1. The few discrepancies documented in this report are likely due to database corrections made after the data freeze used for the publication. The analysis dataset housed at the repository is the most current, updated dataset and contains more than twice the number of patients as was analyzed for the publication. The DCC has confirmed all variables used in our replication analysis and, further, has used a copy of the data sent to the repository to conduct their own replication of several variables (see Attachment 3).

2. In addition to the analysis dataset examined in this replication analysis (PHENOTYP), there exist genotypic data from the IBDGC cohort. Attachment 4 provides a complete description of the genotypic data, which is housed in the NIDDK Genetics Repository at Rutgers. These data have not been replicated, as they are not included in the NIDDK Data Repository at RTI.

3. The SAS datasets provided to the NIDDK Data Repository are provided as transport files. In order to use SAS Viewer, limit CPU resources and increase performance when using these datasets, they must be converted back to an un-archived state. One method to do this is via PROC COPY, as follows:

```
/* Location of IBDGC SAS Data Files */
libname indata xport 'R:\IBDGC\Phenotype data\phenotypes.xpt';
libname infmts xport 'R:\IBDGC\Phenotype data\formats.xpf';

/* Location for Un-archived IBDGC SAS Data Files */
libname sasdb 'R:\IBDGC\Phenotype data\Unarchived';

/* Create SAS-Readable File & Formats */
proc copy in=indata out=sasdb; run;
proc copy in=infmts out=sasdb; run;

/* Output Formats */
data fmts; set sasdb.formats;
proc format cntlin=fmts; run;
```

References

1. Geoffrey C. Nguyen, M.D. et al, **Inflammatory Bowel Disease Characteristics Among African Americans, Hispanics, and Non-Hispanic Whites: Characterization of a Large North American Cohort**, American Journal of Gastroenterology 2006; 101:1012-1023.

2. NIDDK Website: IBDGC page. Inflammatory Bowel Disease Genetic Consortium (IBDGC) : NIDDK.

ATTACHMENT 1

Full Text of Article

Geoffrey C. Nguyen, M.D. et al, **Inflammatory Bowel Disease Characteristics Among African Americans, Hispanics, and Non-Hispanic Whites: Characterization of a Large North American Cohort**, American Journal of Gastroenterology 2006; 101:1012-1023.

American Journal of Gastroenterology
© 2006 by Am. Coll. of Gastroenterology
Published by Blackwell Publishing

ISSN 0002-9270
doi: 10.1111/j.1572-0241.2006.00504.x

Inflammatory Bowel Disease Characteristics Among African Americans, Hispanics, and Non-Hispanic Whites: Characterization of a Large North American Cohort

Geoffrey C. Nguyen, M.D.,[1] Esther A. Torres, M.D.,[2] Miguel Regueiro, M.D.,[3] Gillian Bromfield, M.Sc.,[4] Alain Bitton, M.D., F.R.C.P.(C.),[5] Joanne Stempak,[6] Themistocles Dassopoulos, M.D.,[1] Philip Schumm,[4] Federico J. Gregory, M.D.,[2] Anne M. Griffiths, M.D.,[7] Stephen B. Hanauer, M.D.,[4] Jennifer Hanson,[3] Mary L. Harris, M.D.,[1] Sunanda V. Kane, M.D., M.S.P.H.,[4] Heather Kiraly Orkwis,[3] Raymond Lahaie, M.D.,[8] Maria Oliva-Hemker, M.D.,[9] Pierre Pare, M.D., F.R.C.P.(C.), F.A.C.G.,[10] Gary E. Wild, M.D., Ph.D., F.R.C.P.(C.),[11] John D. Rioux, Ph.D.,[12,13] Huiying Yang, M.D., Ph.D.,[14] Richard H. Duerr, M.D.,[3,15] Judy H. Cho, M.D.,[4] A. Hillary Steinhart, M.D.,[6] Steven R. Brant, M.D.,[1,16] and Mark S. Silverberg, M.D., Ph.D.[6]

[1]*Harvey M. and Lyn P. Meyerhoff Inflammatory Bowel Disease Center, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland,* [2]*Department of Medicine, University of Puerto Rico School of Medicine, Medical Sciences Campus, San Juan, Puerto Rico,* [3]*Inflammatory Bowel Disease Center and Division of Gastroenterology, Hepatology and Nutrition, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania,* [4]*The University of Chicago, Chicago, Illinois,* [5]*Royal Victoria Hospital, McGill University, Montreal, Quebec, Canada,* [6]*Department of Medicine, Mount Sinai Hospital Inflammatory Bowel Disease Centre, University of Toronto, Toronto, Canada,* [7]*Department of Pediatrics, Hospital for Sick Children, University of Toronto, Toronto, Canada,* [8]*Hôpital St. Luc, Centre Hospitalier, Université de Montréal, Quebec, Canada,* [9]*Pediatric Gastroenterology, Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, Maryland,* [10]*Hôpital du Saint-Sacrement, Quebec, Canada,* [11]*Montreal General Hospital, McGill University, Quebec, Canada,* [12]*The Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, Massachusetts,* [13]*The Montréal Heart Institute and the Université de Montréal, Quebec, Canada,* [14]*Division of Medical Genetics and Inflammatory Bowel Disease Center, Cedar-Sinai Medical Center, Los Angeles, California,* [15]*Department of Human Genetics, University of Pittsburgh Graduate School of Public Health, Pittsburgh, Pennsylvania,* [16]*Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland*

OBJECTIVES: Inflammatory bowel disease (IBD), comprising primarily of Crohn's disease (CD) and ulcerative colitis (UC), is increasingly prevalent in racial and ethnic minorities. This study was undertaken to characterize racial differences in disease phenotype in a predominantly adult population.

METHODS: Phenotype data on 830 non-Hispanic white, 127 non-Hispanic African American, and 169 Hispanic IBD patients, recruited from six academic centers, were abstracted from medical records and compiled in the NIDDK-IBD Genetics Consortium repository. We characterized racial differences in family history, disease location and behavior, surgical history, and extraintestinal manifestations (EIMs) using standardized definitions.

RESULTS: African American CD patients were more likely than whites to develop esophagogastroduodenal CD (OR = 2.8; 95% CI: 1.4–5.5), colorectal disease (OR = 1.9; 95% CI: 1.1–3.4), perianal disease (OR = 1.7; 95% CI: 1.03–2.8), but less likely to have ileal involvement (OR = 0.55; 95% CI: 0.32–0.96). They were also at higher risk for uveitis (OR = 5.5; 95% CI: 2.3–13.0) and sacroiliitis (OR = 4.0; 95% CI: 1.55–10.1). Hispanics had higher prevalence of perianal CD (OR = 2.9; 95% CI: 1.8–4.6) and erythema nodosum (3.3; 95% CI: 1.7–6.4). Among UC patients, Hispanics had more proximal disease extent. Both African American and Hispanic CD patients, but not UC patients, had lower prevalences of family history of IBD than their white counterparts.

CONCLUSIONS: There are racial differences in IBD family history, disease location, and EIMs that may reflect underlying genetic variations and have important implications for diagnosis and management of disease. These findings underscore the need for further studies in minority populations.

(Am J Gastroenterol 2006;101:1012–1023)

INTRODUCTION

Inflammatory bowel disease (IBD) comprises two major subtypes, Crohn's disease (CD) and ulcerative colitis (UC), that are based on clinical, endoscopic, histopathological, and radiological characteristics. The term indeterminate colitis is reserved for patients with IBD of the colon in whom clinical features do not allow differentiation between CD and UC. For research purposes, the Vienna classification was devised to categorize CD into phenotypic subgroups based on age of onset, disease location, and disease behavior (1). UC, on the other hand, is classified by extent of colonic involvement: proctitis, left-sided colitis, and pancolitis. Both IBD subtypes may have extraintestinal manifestations (EIMs) that include ocular disease, arthritis, dermatological lesions, and primary sclerosing cholangitis (PSC) (2–4). Characterization of differences in disease presentation has important implications for diagnosis and management of IBD (5).

There is a dearth of research on racial and ethnic variations in IBD phenotype, and the available literature is conflicting. Several retrospective studies have cited racial differences in clinical phenotype. Simsek and Schuman (6) reported that EIMs such as PSC and arthritis were more common in African Americans than white UC patients. A case series of 43 African American IBD patients from Howard University documented a high prevalence of perianal and fistulizing disease (7). Similarly, Goldman *et al.* (8) described a higher rate of perianal disease and fistulas among African Americans than in whites (67% *vs* 48% for perianal disease and 35% *vs* 20% for fistulizing disease).

Straus *et al.* found no differences in age of onset, disease severity, surgical history, or presence of stricturing disease in their multi-center survey of CD patients (9). The study did not use a standard classification system such as the Vienna criteria nor did it assess for racial differences in disease location, family history, or EIMs. Similarly, Kugathasan *et al.* found no differences in disease location and behavior among African Americans, Hispanics, and white CD patients, although this study was restricted to children (10). Deveaux *et al.* found trends toward lower prevalence of family history and small bowel involvement comparing 38 African American to 140 white patients who had undergone CD surgery, though the sample size was small (11).

Given the widely varying reports in the literature, we sought to characterize the IBD phenotypes of African Americans and Hispanics compared to whites in a predominantly adult population. The study of these minorities is particularly relevant given that African Americans and Hispanics represent the two largest minority groups in the United States, and that the latter is the most rapidly growing ethnic minority. Epidemiology studies in the 1960s and 1970s had shown that the incidence of IBD in African Americans was significantly less than in whites. However, over the last two decades, the incidence of IBD hospitalizations in African Americans has approached that in whites (12, 13). In a recent study of insured patients in Puerto Rico, the prevalence of CD and UC was 41.4 per 100,000 and 62.2 per 100,000, respectively, which are similar to reported IBD prevalence in whites in a large health maintenance organization (13, 14). Our understanding of the natural history of IBD has arisen from studies in predominantly white populations and may not be readily generalizable to minorities. We hypothesize that there are likely racial differences in family history of IBD, disease location and behavior, and EIMs. Therefore, the aims of this multicenter review are two-fold: (1) to clinically characterize a large North American repository of racially diverse IBD patients and (2) to determine if phenotype varies by race and ethnicity.

METHODS

Subjects

We conducted a multi-center retrospective record analysis of IBD patients registered in the Inflammatory Bowel Disease Genetics Consortium (IBDGC) to determine whether there were racial differences in IBD phenotype. This consortium was commissioned by the NIDDK (the National Institute of Diabetes and Digestive and Kidney Diseases) to serve as a resource for pooling genetic specimens and clinical phenotype data on IBD subjects recruited through six genetic research centers in the United States and Canada. These included the Broad Institute of the Massachusetts Institute of Technology and Harvard University, Cedars-Sinai Hospital in Los Angeles, Johns Hopkins University, the University of Chicago, the University of Pittsburgh, and the University of Toronto. Johns Hopkins recruited exclusively African American patients and coordinated additional recruitment of a small number (less than 10) of African American patients from satellite recruitment centers at Howard University and the University of North Carolina. Cedars-Sinai in Los Angeles in conjunction with the University of Puerto Rico recruited Hispanic IBD patients from Puerto Rico. The other four centers recruited primarily white IBD patients from their respective IBD centers, with the exception of the Broad Institute which recruited IBD patients from the Quebec Inflammatory Bowel Disease Genetic Consortium (QIGC) which included the following centers in the province of Quebec: Royal Victoria Hospital (RVH), Montreal General Hospital (MGH), Jewish General Hospital (JGH), Hopital Maisonneuve Rosemont (HMR), Hôpital Hotel Dieu de Montréal (HDM), Hôpital St. Luc (HSL), Centre Hospitalier Universitaire du Quebec–Hôpital St. Sacrement (CHUQ–HS), Centre Hospitalier Universitaire de Sherbrooke (CHUS), Hôpital Sainte Justine (HSJ), Montreal Children's Hospital (MCH), Complexe Hospitalier de la Sagamie–Chicoutimi. Each center and satellite center recruited patients that had confirmed diagnoses of IBD. A requirement was that, for each patient, the centers had to recruit either an age- and geographically matched healthy (*i.e.*, non-IBD) control (spouse, significant other, friend, or population control) or to have both parents recruited as genetic controls. The latter circumstance in which

both parents of an affected subject or index case are recruited as genetic controls regardless of whether or not they have IBD is referred to as a case-trio. Patients previously recruited for IBD genetic studies were excluded. Recruitment was not focused on any specific characteristics of IBD (*e.g.*, surgery history, disease complications, or IBD family history).

We extracted clinical data on all unrelated index subjects with a confirmed diagnosis of IBD from the NIDDK-IBDGC's central repository database. CD, UC, and IC were diagnosed based on a standardized set of clinical, endoscopic and/or radiological and histopathological criteria. Subjects were prospectively identified and recruited at an academic center or one of its satellite clinics from January 10, 2003 to April 15, 2005. In addition to demographic data, clinical phenotype data on age at diagnosis, disease location and extent, clinical behavioral pattern (for CD), tobacco use, family history, EIMs, and surgical history were collected. Family history of IBD was defined as IBD (CD or UC), by patient report, in a first- or second-degree relative, which included grandparents, avuncular relatives, or half-siblings. Smoking was defined as smoking an average of at least 1 cigarette daily for a period of at least 3 months prior to diagnosis. Data on disease characteristics were extracted from medical, surgical, endoscopic, radiological, and histopathological (both endoscopic biopsies and surgical specimens) records by experienced IBD clinicians on the Consortium Phenotyping Committee using standardized definitions, forms, and procedures detailed in an operations manual (available upon request).

Designation of Race and Ethnicity

In the NIDDK database, race was designated as White, Black/African American, American Indian/Alaskan Native, Asian, or Other by self-report. Ethnicity was also self-reported as Hispanic or non-Hispanic. Our study included the following racial/ethnic groups: non-Hispanic whites, non-Hispanic African Americans, and Hispanics. All Hispanic subjects in this analysis reside on the island of Puerto Rico. One African American Hispanic living in Puerto Rico was classified as Hispanic for analysis.

Disease Site Designation

Charts were reviewed for evidence of IBD, from medical, radiology, and pathological reports. For a given patient, all areas of macroscopic involvement throughout the course of disease were recorded. These categories included: esophagogastroduodenal, jejunal, ileal, colonic, and perianal diseases. Designation of a disease site as "yes" or "no" required evaluation by upper or lower endoscopy reports, barium X-rays, operative reports, or pathology resection specimen reports. In cases where there was missing information regarding evaluation of a portion of the GI tract, disease status was classified as "unknown." Perianal disease was defined as a history of perianal or perineal abscess(es) and/or fistula, anal canal ulcers, anal stenosis or chronic edematous, and violaceous skin tags, but did not include anal hemorrhoids or fissures.

For comparison with other population studies, the same individuals were placed in a "best-fit" category for disease site. Patients were classified into one of four best-fit groups: (1) ileal: all patients with any gross involvement of the ileum without colonic or proximal small bowel involvement; (2) ileocolonic: involvement of both the ileum and colon without proximal small bowel disease; (3) colon-only: involvement limited to the colon or rectum, and (4) upper gastrointestinal: involvement of the esophagus, stomach, duodenum, or jejunum with or without disease in the ileum or colon. In contrast to the Vienna classification which categorizes disease site prior to resection, we modified the criteria with disease location based on the maximum extent of disease at any point during the course of the disease. Subjects had to have available information on ileal, colonic, and upper GI disease in order to be included in the best-fit analysis.

Because of our concern that the proportion of patients who had an evaluation for upper GI disease may have varied by center and could potentially bias our racial comparisons, we performed a sensitivity analysis excluding centers where more than 15% of patients did not undergo evaluation for upper GI disease.

Macroscopic disease extent in UC was classified as: (1) proctitis, defined as inflammation extending up to but not beyond 15 cm from the anorectal junction; (2) left-sided colitis, defined as inflammation extending up to the splenic flexure; and (3) extensive colitis, defined as inflammation extending proximal to the splenic flexure. Disease site designation required evaluation of the entire colon as confirmed by colonoscopy, barium enema, or colectomy gross pathology reports.

Extraintestinal Manifestations

Erythema nodosum (EN) was defined as raised, tender, red or violet subcutaneous nodules that are 1–5 cm in diameter and most commonly located on the extensor surfaces of the extremities, particularly over the anterior tibial area. Subjects were classified as having pyoderma gangrenosum (PG) if they had documentation of ulcerative disease of the skin, occurring as single or multiple lesions, most commonly on the legs, especially the pretibial area. Uveitis was defined as intraocular inflammation with documentation of typical findings, preferably a slit lamp examination. A designation of sacroiliitis required radiological documentation of typical findings of sacroiliac inflammation, narrowing or sclerosis, and/or inflammation or fusion of vertebral bodies. Classification of ankylosing spondylitis required documentation of back pain and progressive stiffness of the spine in addition to radiological documentation of findings of sacroiliac inflammation, narrowing or sclerosis, and/or inflammation or fusion of vertebral bodies. PSC was documented with typical dye cholangiographic or MRCP findings in someone with no other known causes of secondary cholangitis. Abnormal liver enzymes or liver biopsy alone was not sufficient evidence of PSC. Because the operating definitions for peripheral

arthritis were revised during the study period, this EIM was not included in the analysis.

Clinical Disease Behavior

Disease behavior or type was considered cumulatively to the time of the most recent visit. Patients were assigned to only one category, following the Vienna classification guidelines. CD behavior was classified as: (1) *inflammatory* or non-penetrating non-stricturing disease was defined as uncomplicated inflammatory disease without evidence of *stricturing* or *penetrating* complications; (2) *stricturing* disease was defined as occurrence of constant luminal narrowing demonstrated by radiological, endoscopic, or surgical examination combined with prestenotic dilation and/or obstructive signs or symptoms but without evidence of abdominal penetrating disease; and (3) *penetrating*, which was defined as the occurrence of bowel perforation, intra-abdominal fistulas, inflammatory masses, and/or abscesses at any time in the course of the disease, and not secondary postoperative intra-abdominal complication. Perianal and rectovaginal fistulae were not included in the *penetrating* category, in contrast to the Vienna classification. Subjects with both stricturing and penetrating disease were classified as penetrating disease. Because many patients with inflammatory CD ultimately develop strictures or penetrating disease over time, we performed a sensitivity analysis of disease behavior in a subgroup of patients with disease for at least 5 yr (15).

Surgical History

Surgery for complication or treatment of CD was confirmed by medical records and classified in non-exclusive categories: (1) bowel resection or stricturoplasty for stricturing disease or resection of disease complicated by a fistula or abscess; (2) diversions, defined as a diverting ostomy that is performed prior to definitive surgery, such as resection, or to allow healing of perineal disease; (3) surgery for an abdominal fistula or abscess, which included surgical fistulotomy, surgical resection of a complicated fistula, surgical drainage of an abscess, or percutaneous drainage of an intra-abdominal abscess that is followed by surgical resection of involved intestine; and (4) surgery for perianal fistula including placement of a Seton, fistulotomy, or intestinal diversion to permit healing of perineal disease. Of note, a simple incision and drainage of a perianal abscess performed using only local anesthetic was not considered a surgical intervention.

Surgery for complication of UC was confirmed by medical records and, when available, the indication for surgery was recorded as: (1) dysplasia or cancer, which did not include incidental findings of neoplasia during surgery for another indication; (2) chronic continuous disease, when neither dysplasia nor fulminant disease was indicated; and (3) acute fulminant disease defined as an acute or subacute onset of severe colitis within a period of 2–12 wks in an individual with previously undiagnosed or quiescent colitis.

Statistical Analysis

Statistical analysis was performed with Stata 8.2 software (StataCorp, TX). Analysis was performed on unrelated IBD patients. For pedigrees with multiple affected family members, the index case was selected for analysis. Comparisons of group means were made using analysis of variance with Bonferroni correction. Categorical patient characteristics among the different race groups were compared using χ^2 or Fisher's exact test of proportions, and Bonferroni adjustments were not made. The proportion of siblings affected by IBD was calculated by dividing the total number of affected siblings within an entire racial/ethnic cohort by the total number of siblings in that cohort.

Multiple logistic regression models were constructed to determine the independent effect of race on the relative odds of ileal disease, perianal disease, penetrating disease, uveitis, sacroiliitis, and EN. The relative odds of the EIMs were adjusted for age at study entry, gender, and disease subtype (CD *vs* UC or IC). The logistic regression model for perianal disease was controlled for age at study entry and age at diagnosis. The odds ratio of abdominal penetrating disease was adjusted for attained age, age at diagnosis, tobacco use, and family history. The relative odds of the remaining dichotomous outcomes (esophagogastroduodenal, ileal, and colorectal involvement) were adjusted for age at study entry, age at diagnosis, gender, tobacco use, and family history of IBD. Multiple linear regression was implemented to determine the independent effect of race on the number of CD-related surgeries after controlling for the above covariates.

To obtain time-to-event data for survival analysis, we retrospectively reconstructed a cohort of CD patients for whom we had dates of first surgery and obtained the time from diagnosis to first surgery. Survival curves for time free of surgery were constructed for whites, African Americans, and Hispanics using the Kaplan–Meier method. Global differences between the curves were tested for statistical significance using the log-rank test. The cumulative proportion of patients free of CD-related surgeries were then calculated at 2, 5, and 10 yr. We used a Cox proportional hazard model to determine the relative hazard of first surgery for both African Americans and Hispanics compared to whites, while adjusting for age at diagnosis, gender, smoking, family history, best-fit disease location, and disease behavior category. Median times to first surgery were estimated by modeling the data to a Weibull distribution.

Ethical Considerations

Prior to data collection, the Institutional Review Boards at Broad-MIT, Cedars-Sinai Hospital (Los Angeles), the University of Puerto Rico Medical Sciences Campus, the Johns Hopkins University, the University of Chicago, the University of Pittsburgh, and Mount Sinai Hospital at the University of Toronto approved the patient and data collection protocols for their respective sites.

Table 1. Patient Demographics by Race/Ethnicity of the NIDDK-IBDGC Repository

Characteristics	All No. (%)	Race/Ethnicity White No. (%)	AA No. (%)	Hispanic No. (%)
Race distribution	1,126	830	127	169
Age at study entry in yr, mean (SD)	36.1 (14.4)	36.7 (14.9)	36.1 (14.6)	33.3 (10.6)[a]
Age at diagnosis in yr, mean (SD)	26.4 (12.8)	26.6 (13.3)	25.4 (13.2)	25.9 (9.8)
Female, no. (%)	584 (51.9)	427 (51.5)	80 (63.0)[b]	77 (45.6)[c]
Diagnosis, no. (%)				
CD	697 (61.9)	510 (61.5)	81 (63.8)	106 (62.7)
UC	396 (35.2)	299 (36.0)	35 (27.6)	62 (36.7)
Indeterminate colitis	33 (2.9)	21 (2.5)	11 (8.7)[b]	1 (0.6)
Family history of IBD[d]	288 (25.6)	237 (28.6)	23 (18.1)[b]	28 (16.6)[c]
Among CD patients	191 (27.4)	156 (30.6)	16 (19.8)[b]	19 (17.9)[c]
Among UC patients	88 (22.2)	74 (24.8)	5 (14.3)	9 (14.5)
% Siblings affected	3.8	4.6	2.5	1.3[c]
Packs/day at diagnosis[f]	0.12	0.15	0.07	0.04[c]
Smoking at diagnosis, no. (%)				
Current	192 (17.2)	161 (19.5)	20 (16.1)	11 (6.6)[c,g]
Former	140 (12.6)	117 (14.2)	15 (12.1)	8 (4.8)[c,g]
Never	783 (70.2)	547 (66.3)	89 (71.8)	147 (88.6)[c,g]
Appendectomy (2 yr prior to IBD)	47 (4.4)	32 (4.0)	3 (2.9)	12 (7.1)

AA = African American; IBD = inflammatory bowel disease; CD = Crohn's disease; UC = ulcerative colitis.
[a]Bonferroni-adjusted $p < 0.05$ comparing Hispanics to whites.
[b]$p < 0.05$, comparing African Americans to whites.
[c]$p < 0.01$, comparing Hispanics to African Americans.
[d]Family history of IBD refers to a subject who has a first- or second-degree relative with either CD or UC. The data is further stratified into family history of IBD among only CD patients and only UC patients.
[e]$p < 0.01$, comparing Hispanics to whites.
[f]Average number of packs of cigarettes per day smoked at time of diagnosis among smokers and nonsmokers. The average smoker smoked 0.7 packs per day.
[g]$p < 0.001$, comparing Hispanics to whites.

RESULTS

After screening for eligibility, a total of 1,126 eligible subjects with IBD (830 whites, 127 African Americans, and 169 Hispanics) were included in this study. The age at study registration ranged from 3 to 85 yr, with an interquartile range from age 26 (25th percentile) to 45 yr (75th percentile). Descriptive analysis of racial differences in patient demographics showed that Hispanics were younger at study entry than white subjects (33.3 *vs* 36.7 yr, $p < 0.05$). Hispanics were also less likely to have been smoking at the time of IBD diagnosis, and as a group smoked fewer packs per day (ppd) than whites (0.04 *vs* 0.15 ppd, $p < 0.01$; Table 1). There were more female African American subjects compared to whites (63.0% *vs* 51.5%). The distribution of disease subtype (CD *vs* UC) was consistent among all racial and ethnic groups with a 3:2 ratio, respectively. The average age at diagnosis in the total study population was 26.4 yr and was similar among all races (Table 1). There were no differences in appendectomy rates at least 2 yr prior to IBD diagnosis between whites and African Americans or Hispanics.

Family History

The prevalence of family history of IBD (CD or UC) among all subjects was 25.6% (95% CI: 23.0–28.0%). Among the entire study population, white IBD subjects had a higher prevalence of family history of IBD compared to African Americans (28.6% *vs* 18.1%, $p = 0.01$) and Hispanics (28.6% *vs* 16.6%, $p = 0.001$). However, when stratified by disease type (CD *vs* UC), it was only among CD patients that white subjects were more likely to have a family history of IBD compared to African Americans (30.6% *vs* 19.8%, $p < 0.05$) and to Hispanics (30.6% *vs* 17.9%, $p < 0.01$). There were no statistically significant racial differences in prevalence of family history among UC patients. The proportion of siblings in the Hispanic cohort affected with IBD was lower than that in whites (1.3% *vs* 4.6%, $p = 0.004$), while there was no statistically significant difference between African Americans and whites (2.5% *vs* 4.6%, $p = 0.13$).

CD Phenotype

The unadjusted distribution of CD location and behavioral pattern for all subjects, whites, African Americans, and Hispanics are shown in Table 2. Compared to white subjects, African Americans had a lower prevalence of ileal involvement (67.6% *vs* 80.0%, $p = 0.02$) with an odds ratio for ileal involvement of 0.55 (95% CI: 0.32–0.96) after adjustment for age, age at diagnosis, gender, smoking at diagnosis, and family history. In addition to being African American compared to white, smoking at time of study entry was associated with an odds ratio of 1.86 (95% CI: 1.1–3.1) for ileal involvement, and being diagnosed at age >40 yr was associated with an odds ratio of 0.40 (95% CI: 0.18–0.91) compared to individuals who were diagnosed at age <20 yr. Gender and family history were not statistically significant covariates, but were kept in the model because they are known to be associated with ileal disease. Though the difference in upper gastrointestinal

Table 2. CD and UC Phenotype by Race/Ethnicity of the NIDDK-IBDGC Repository

Characteristics	All No. (%)	Race/Ethnicity White No. (%)	AA No. (%)	Hispanic No. (%)
Crohn's disease				
Disease involvement				
Upper gastrointestinal (N = 609)	96 (15.8)	72 (16.1)	17 (25.8)	7 (7.4)[a]
Esophagogastroduodenal (N = 611)	54 (8.8)	38 (8.5)	14 (20.0)[b]	2 (2.1)[a]
Jejunum (N = 629)	57 (9.1)	44 (9.4)	8 (12.9)	5 (5.2)
Ileum (N = 671)	530 (79.0)	396 (80.0)	50 (67.6)[c]	84 (82.4)[d]
Colorectal (N = 672)	459 (68.3)	326 (65.6)	63 (77.8)[c]	70 (74.5)
Perianal disease (N = 689)	231 (33.5)	146 (28.7)	32 (40.0)[c]	53 (52.5)[e]
Best-fit disease site (N = 579)				
Ileum	156 (26.9)	126 (29.2)	10 (16.1)[b]	20 (23.3)[a,f]
Ileo-colon	217 (37.5)	158 (36.7)	14 (22.6)[b]	45 (52.3)[a,f]
Colon	110 (19.0)	75 (17.4)	21 (33.9)[b]	14 (16.3)[a,f]
Upper gastrointestinal	96 (16.6)	72 (16.7)	17 (27.4)[b]	7 (8.1)[a,f]
Disease pattern (N = 689)				
Inflammatory	332 (48.2)	242 (48.0)	42 (53.2)	48 (45.3)
Stricturing	155 (22.5)	108 (21.4)	22 (27.9)	25 (23.6)
Penetrating	202 (29.3)	154 (30.6)	15 (19.0)[c]	33 (31.1)
Ulcerative colitis				
Disease extent (N = 371)				
Proctitis	23 (6.2)	17 (5.9)	4 (13.8)	2 (3.6)[a,g]
Left-sided colitis	109 (29.4)	90 (31.4)	10 (34.5)	9 (16.4)[a,d]
Extensive colitis	239 (64.4)	180 (62.7)	15 (51.7)	44 (80.0)[a,g]

Phenotype data for some individuals were unknown or insufficiently documented in the medical record. The total number of subjects with complete data with respect to a given phenotype is indicated in italicized brackets.
AA = African American
[a] $p < 0.05$, comparing Hispanics to whites.
[b] $p < 0.01$, comparing African Americans to whites.
[c] $p < 0.05$, comparing African Americans to whites.
[d] $p < 0.05$, comparing Hispanics to African Americans.
[e] $p < 0.01$, comparing Hispanics to whites.
[f] $p < 0.001$, comparing Hispanics to African Americans.

involvement between African Americans and whites was only marginally significant (25.8% *vs* 16.1%, $p = 0.05$), the prevalence of esophagogastroduodenal disease, which did not include jejunal disease, was notably greater in African Americans (20% *vs* 8.5%, $p = 0.003$) with an adjusted odds ratio of 2.8 (95% CI: 1.4–5.5). Colorectal disease was also more common in African Americans than whites (77.8% *vs* 65.6%, $p = 0.03$) with an adjusted odds ratio of 1.9 (95% CI: 1.1–3.4). As a consequence, disease location based on a best-fit classification using exclusive categories was significantly different between whites and African Americans ($\chi^2 = 17.4$; $p = 0.001$). Hispanics, in contrast, had a lower prevalence of esophagogastroduodenal involvement than whites (2.1% *vs* 8.5%, $p = 0.03$) with an adjusted odds ratio of 0.23 (95% CI: 0.05–1.0). The overall pattern of best-fit disease location between white and Hispanics was also statistically different ($\chi^2 = 8.8$; $p = 0.03$).

Sensitivity analysis showed that only two recruitment centers had greater than 15% of patients who did not undergo evaluation for upper GI disease (range: 2.3–30.9%). After exclusion of these centers from analysis, we obtained similar results with the prevalence of esophagogastroduodenal disease being higher in African Americans compared to whites (20% *vs* 7.1%, $p = 0.001$). The differences between African Americans and whites in best-fit site location, reflecting less ileal disease and more upper GI disease, also remained statistically significant ($p = 0.001$).

Perianal disease was more common in African Americans than whites (40% *vs* 28.7%, $p < 0.05$) with an adjusted odds ratio of 1.7 (1.03–2.8). Hispanics also had a higher prevalence of perianal disease compared to white subjects (52.5% *vs* 28.7%, $p < 0.001$) with odds ratio of 2.9 (95% CI: 1.8–4.6) after controlling for confounders.

In terms of disease behavior, African Americans were less likely to have abdominal penetrating disease than whites (19% *vs* 30.6%, $p = 0.03$) with an odds ratio of 0.55 (95% CI: 0.3–1.0) after adjustment for age, age at diagnosis, gender, smoking at diagnosis, family history of IBD. When we performed a subgroup analysis of patients with disease duration of at least 5 yr, we derived similar results comparing the prevalence of abdominal penetrating disease in African Americans to whites (20% *vs* 35.8%, $p = 0.02$). In addition, after 5 yr since diagnosis, African Americans were more likely to have stricturing disease (38.2% *vs* 24.4%, $p = 0.03$). Disease behavior was not significantly different between whites and Hispanics.

Table 3. Surgical History of IBD Patients in the NIDDK-IBDGC Repository by Race/Ethnicity

Characteristics	All No. (%)	Race/Ethnicity White No. (%)	AA No. (%)	Hispanic No. (%)
CD				
No. surgeries, mean (SD)				
For abdominal CD (N = 439)	1.4 (1.3)	1.6 (1.3)	0.8 (1.1)[a]	1.1 (1.0)[b]
For perianal CD (N = 425)	0.5 (1.2)	0.5 (1.1)	0.3 (0.7)	0.7 (1.7)
Bowel resection (N = 472)	335 (71.0)	263 (78.7)	27 (35.1)[a]	45 (73.8)[c]
Bowel diversion (N = 458)	46 (10.0)	24 (7.4)	9 (12.0)	13 (22.4)[d]
Surgery for penetrating disease (N = 456)	98 (21.5)	74 (22.9)	10 (13.3)	14 (24.1)
Surgery for perianal disease (N = 465)	107 (23.0)	77 (23.4)	15 (20.0)	15 (24.6)
Median time to 1st surgery (yr)	9.5	9.8	11.4	6.6
% Surgery-free				
At 2 yr	86.1%	84.6%	94.8%	86.7%
At 5 yr	67.7%	67.7%	76.3%	59.8%
At 10 yr	48.0%	49.6%	58.8%	26.1%
Ulcerative colitis				
Colectomy (N = 395)[e]	71 (18.0)	47 (15.8)	4 (11.4)	20 (32.3)[b]
Dysplasia (N = 134)	4 (3.0)	3 (3.5)	1 (3.7)	0 (0)
Chronic refractory disease (N = 133)	54 (40.6)	35 (40.7)	1 (3.9)[f]	18 (85.7)[g]
Fulminant colitis (N = 137)	8 (5.9)	6 (6.9)	0 (0)	2 (9.5)

AA = African American; CD = Crohn's disease.
[a] $p < 0.001$, comparing African Americans to whites.
[b] $p < 0.01$, comparing Hispanics to whites.
[c] $p < 0.001$, comparing Hispanics to African Americans.
[d] $p < 0.001$, comparing Hispanics to whites.
[e] Indication for surgery was only available for about a third of UC patients who underwent surgery.
[f] $p < 0.001$, comparing African Americans to whites.
[g] $p < 0.001$, comparing Hispanics to whites.

UC Phenotype

While the UC disease extent was not significantly different between African Americans and whites, Hispanics had more extensive and proximal distribution of disease compared to whites (Fisher's exact test, $p = 0.04$; see Table 2). The overall prevalence of IBD-related colorectal dysplasia or cancer was 3.5% among all UC patients and was similar among all races: whites (3.1%), African Americans (5.7%), and Hispanics (4.1%).

Surgical History

Compared to whites, African Americans were less likely to have undergone bowel resection for CD (35.1% *vs* 78.7%, $p < 0.001$). This association was persistent with an odds ratio of 0.14 (95% CI: 0.06–0.31) for bowel resection in African Americans compared to whites even after adjustment for confounders including disease site and behavior. In the multivariate analysis, a higher odds of undergoing bowel resection was observed among females compared to males (OR = 2.9; 95% CI: 1.5–5.6) and those diagnosed between ages of 20 and 40 yr, compared to those younger than 20 yr (OR = 2.1; 95% CI: 1.1–4.0). The average number of surgeries for abdominal CD since diagnosis was also significantly lower in African Americans than whites (0.8 *vs* 1.6, $p < 0.001$) with an adjusted difference of −0.6 (95% CI: −0.9 to −0.2) after controlling for age, age at diagnosis, gender, smoking at diagnosis, family history, disease location, and behavior (Table 3). Interestingly, Hispanics were more likely to have undergone bowel diversion for CD (22.4% *vs* 7.4%, $p = 0.001$). They, however, had fewer total surgeries for abdominal CD than whites with an average difference of −0.4 (95% CI: −0.74 to −0.02) following adjustment for the same confounders. The number of surgeries for perianal CD was similar among all racial groups (Table 3).

The time interval from IBD diagnosis to first abdominal surgery of any type for all CD patients were calculated and the cumulative proportion of individuals who remained free of IBD-related surgery for all three racial groups are depicted with Kaplan–Meier curves as shown in Figure 1. The cumulative proportion of individuals with CD who were free from surgery at 2, 5, and 10 yr after diagnosis are shown in Table 3. At 5 yr, the proportion of surgery-free CD patients were: 68% (95% CI: 62–72%) for whites, 76% (95% CI: 62–86%) for African Americans, and 60% (95% CI: 46–71%) for Hispanics. The survival functions representing the three racial groups were of marginal statistical significance using the log-rank test ($p = 0.05$). The median survival times free from surgery by race were: 9.5 yr for all racial/ethnic groups, 9.8 yr for whites, 11.4 yr for African Americans, and 6.6 yr for Hispanics. The Cox proportional hazards model showed no statistically significant increased risk of CD-related surgery for African Americans (hazard ratio 0.91; 95% CI: 0.55–1.50) or Hispanics (hazard ratio 1.38; 95% CI: 0.90–2.11) compared to white subjects. In this multivariate analysis, the following covariates were also statistically significant: being female compared to males (OR = 1.34; 95% CI: 1.01–1.78); colonic involvement only compared with ileal disease only (OR = 0.59; 95% CI: 0.36–0.93); stricturing behavior

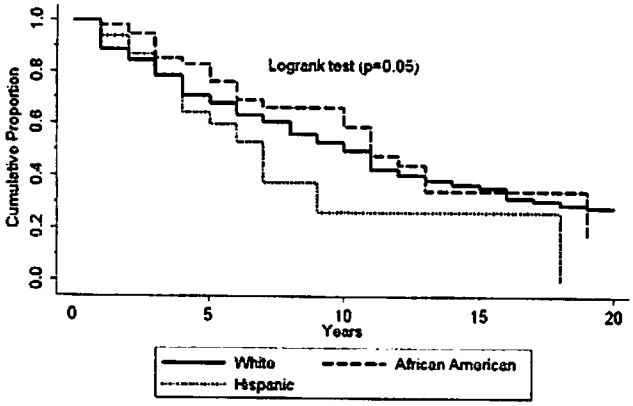

Figure 1. Separate Kaplan–Meier curves of CD patients stratified by race are depicted indicating cumulative survival free of first CD-related surgical intervention at specified time interval following diagnosis of CD. White CD patients are represented by a solid line. African Americans by a long dashed line, and Hispanics by a dotted line. The three survival curves are marginally statistically different as determined by the log-rank test ($p = 0.05$).

compared to inflammatory (OR = 2.75; 95% CI: 1.85–4.1); and penetrating behavior compared to inflammatory (OR = 3.86; 95% CI: 2.65–6.0).

There was considerable racial variation in the proportion of individuals who received colectomy for chronic refractory UC. While African Americans were less likely to undergo colectomy for chronic disease (3.9% *vs* 40.7%, $p < 0.001$), Hispanics had higher prevalence of surgery for chronic UC (85.7% *vs* 40.7%, $p < 0.001$). Furthermore, Hispanics had a considerably higher rate of colectomy for any indication than white subjects (32.3% *vs* 15.8%, $p < 0.01$).

Extraintestinal Manifestations

The most profound racial differences in phenotype were observed in the distribution of EIMs. The distribution of EIMs in all patients and subjects categorized by race are shown in Table 4. Compared to whites, African Americans had more than a four-fold greater prevalence of uveitis (8.1% *vs* 1.6%, $p < 0.001$) with an age- and sex-adjusted odds ratio of 5.5 (95% CI: 2.3–13.0). African Americans were also more likely than whites to have a diagnosis of sacroiliitis with an adjusted odds ratio of 4.0 (95% CI: 1.6–10.1). Hispanic patients had a higher prevalence of EN than white subjects (10.7% *vs* 3.4%, $p < 0.001$) with an adjusted odds ratio of 3.3 (95% CI: 1.7–6.4). In the multivariate analysis, gender and disease type were also significant prognostic factors for the development of EN with odds ratios of 2.9 (95% CI: 1.5–5.7) for females compared to males and 2.4 (95% CI: 1.1–4.9) for CD compared to UC, respectively. There were no statistically significant racial differences in the prevalence of PG, ankylosing spondylitis, and PSC.

DISCUSSION

We characterized the clinical phenotype of a large North American cohort of predominantly adult IBD patients. Our study revealed racial variations in disease location, family and surgical history, and EIMs. African Americans compared to whites with CD had less ileal disease but more esophagogastroduodenal and perianal disease than whites. Hispanics compared to whites had a dramatically higher prevalence of perianal disease. Hispanic UC patients had more extensive proximal disease than their white counterparts. Race and ethnicity appeared to have the greatest effect on the development of uveitis, sacroiliitis, and EN. These findings contribute to the sparse literature on the epidemiology and disease presentation of IBD in racial and ethnic minorities that have been underrepresented in previous population-based studies.

Overall characterization of the NIDDK repository suggests that the total IBD study population was similar to other referral-based patients with respect to clinical disease behavior and disease location according to the Vienna classification (16). Our findings of less abdominal penetrating disease in African Americans was in contrast to a report by Kugathasan *et al.* who found no racial differences in disease behavior

Table 4. Extraintestinal Manifestations in the NIDDK-IBDGC Repository by Race/Ethnicity

Characteristics	All No. (%)	Race		
		White No. (%)	AA No. (%)	Hispanic No. (%)
Sacroiliitis (N = 1,106)	23 (2.1)	14 (1.7)	7 (5.8)[a]	2 (1.2)[b]
Ankylosing spondylitis (N = 1,106)	14 (1.3)	12 (1.5)	2 (1.7)	0 (0)
Uveitis (N = 1,114)	25 (2.2)	13 (1.6)	10 (8.1)[c]	2 (1.2)[d]
Erythema nodosum (N = 1,109)	48 (4.3)	28 (3.4)	2 (1.7)	18 (10.7)[d,e]
Pyoderma gangrenosum (N = 1,112)	23 (2.1)	16 (1.9)	3 (2.5)	4 (2.4)
Primary sclerosing cholangitis (N = 1,117)	20 (1.8)	11 (1.3)	4 (3.2)	5 (3.0)

Data on the presence of extraintestinal manifestations were not available for all individuals. The total number of subjects with complete data is in italicized brackets.
AA = African American.
[a] $p < 0.01$, comparing African Americans to whites.
[b] $p < 0.05$, comparing Hispanics to African Americans.
[c] $p < 0.001$, comparing African Americans to whites.
[d] $p < 0.01$, comparing Hispanics to African Americans.
[e] $p < 0.001$, comparing Hispanics to whites.

(10). Their study comprised, however, exclusively of children, while ours included predominantly adults. Younger age of CD onset which independently predisposes to abdominal penetrating disease, may have attenuated racially driven differences. Similarly, the same authors did not find racial variations in upper gastrointestinal and ileal disease that were characterized in our study. Disease location may also be influenced by earlier age of onset. Deveaux *et al.* found nonsignificant trends toward less small bowel disease comparing African American to white CD patients who had undergone surgery (11). Our results support previous descriptive studies reporting that African Americans have a greater prevalence of perianal disease (8, 17). Also, consistent with the findings by Straus *et al.* (9) and some (6) but not all earlier retrospective series (8), we found that African Americans had no increased risk of first CD-related surgery, after adjustment for duration of disease. The data were, however, suggestive that Hispanics may have higher occurrence of surgical intervention over the duration of disease.

The lower prevalence of family history among African Americans and Hispanics compared to whites is also consistent with previous studies (10, 17). Though these findings may be due to underlying genetic differences among races, ascertainment bias may also be a contributing factor. The overall prevalence of family history in our IBD repository is higher than those reported by population-based studies (18). The African American group had a lower percentage of parent-affected offspring trios recruited compared to whites (9% *vs* 38%), while the percentage in Hispanics was not significantly different than whites. Although there may be a concern that trios recruited would potentially have a greater frequency of IBD family history as a motivation for participation, this was not observed. Among white patients, the prevalence of family history was not different between trios and nontrios (29% *vs* 28%). In addition, no patients were excluded from study entry because they were unable to provide parental controls.

We believe it is most likely that racial differences observed in this study may reflect underlying genetic and biological variations. There is emerging evidence that genetic susceptibility is a strong determinant of clinical phenotype (19). The most consistent finding is the association between the *NOD2* mutation and certain CD phenotypes such as small bowel involvement, earlier age of onset, and fibrostenotic disease (20). Several studies have also shown that there are racial and ethnic variations in the allele frequencies of various mutations within the *NOD2/CARD15* gene. For example, the allele frequency of Gly908Arg is higher among Jewish individuals with CD (10.2%) than non-Jewish CD patients (4.3%). Conversely, non-Jewish CD patients (10.7%) have a significantly higher allele frequency of the Arg702Trp mutation than do Jewish CD patients (2.6%) (21). In certain racial groups, there is no association between the *NOD2/CARD15* gene and disease susceptibility. For example, among 350 Japanese patients with CD, 272 with UC, and 292 controls, none were carriers for any of the three major white risk alleles (22). In African American patients with CD, the frequency of the mutant *NOD2/CARD15* allele is significantly lower than that observed in white CD patients and unaffected white individuals (10, 18). In addition, the specific mutations in the *NOD2/CARD15* gene are unique to African American CD patients (18).

The increased prevalence of uveitis and sacroiliitis in African Americans compared to white IBD patients is also a convincing finding that may share common biological pathways with other inflammatory disorders. Uveitis is frequently associated with other chronic, granulomatous disorders such as sarcoidosis (23), for which African Americans have increased predisposition (24). The increased prevalence of uveitis in both CD and sarcoidosis may elude to a common pathway in immune dysregulation such as polymorphisms in the tumor necrosis factor, which plays a role in granuloma formation (25, 26). Our findings that African Americans were more likely to have uveitis and sacroiliitis is consistent with another study that showed a higher prevalence of EIMs among African American children with UC compared to other races (27).

Our results also potentially implicate the contributory role of environmental factors in the observed racial differences in UC phenotype. Hispanic UC patients had more extensive UC than whites and had higher rates of colectomy. These findings were associated with higher rates of nonsmoking at diagnosis among Hispanics compared to whites, and is consistent with tobacco being a predictor of disease phenotype and risk factor for surgical intervention (28, 29).

In addition to tobacco use, there were racial and ethnic differences in gender and attained age in our study cohort. Specifically, there is a slight female predominance of IBD among African Americans (63%) compared to the distribution of females among whites (51%) and Hispanics (45%). Because some of the differences attributed to race may actually reflect gender differences, we adjusted for this covariate in multivariate analysis. Gender was shown to be independently associated with the presence of EN, bowel resection, and time to surgery. The Hispanic cohort was also slightly younger than whites at study entry, and this was accounted for in multivariate analysis.

The main limitation of our study is that African Americans were recruited primarily at Johns Hopkins and Hispanics at Cedars-Sinai/University of Puerto Rico. Because white IBD patients were not recruited from those centers for this current study, we are unable to statistically adjust for site-specific factors that may affect phenotype. The influence of site of recruitment is, however, minimized by use of standardized definitions of CD and UC phenotype and a detailed manual developed by the Phenotyping Steering Committee that is based on the widely accepted Vienna classification. There may be concern that the ascertainment and diagnosis of certain phenotypes such as esophagogastroduodenal involvement may vary by center. However, all the centers involved in this study are tertiary referral centers experienced in the management and diagnosis of IBD. Our sensitivity analysis, which excluded centers with a high percentage of

individuals who did not undergo evaluation for upper gastrointestinal disease, yielded similar results. Moreover, white patients from the NIDDK repository are phenotypically similar to a cohort of 338 historical white CD patients from Johns Hopkins, where African Americans were predominantly recruited, with respect to the prevalences of family history of IBD (32%) (30), esophagogastroduodenal disease (6.6%), ileal involvement (76.9%), perianal disease (31.5%), and uncomplicated inflammatory behavior pattern (44.2%) (31). Our findings are further strengthened by a retrospective review of 58 African American and 187 white pediatric IBD cases at Johns Hopkins that preceded patients recruited to the NIDDK-IBDGC from our institution (17). This study showed that African Americans were less likely than whites to have a family history of IBD (17.5% *vs* 36.4%, $p = 0.006$) and that the odds for developing perianal disease was 1.8 times higher for African Americans than whites, mirroring our own results.

There may be, however, center-specific differences in health utilization and preferences toward conservative *versus* aggressive disease management that may have been reflected in racial differences in surgical history. Hispanic UC patients had a higher rate of colectomy compared to whites, corresponding to greater disease extent. Because we do not have measures of disease activity, it is unclear whether the difference in surgical intervention is due to racial differences in disease severity or due to region-specific physician preferences for earlier surgery. Puerto Rico has universal health coverage and that may partially explain the higher percentage of clinically indicated surgery among Hispanics for refractory UC compared to whites who experience a much more heterogenous health-care access. Thus, ethnic differences in colectomy with respect to Hispanics may not be generalizable to Hispanics living in the United States. These differences may also represent referral bias. The University of Puerto Rico receives referrals from the most complex and severe IBD cases from the community and is where practically all total proctocolectomies with ileal pouch anal anastamoses are performed. In contrast to Hispanics, African Americans CD patients had a lower rate of bowel resection and surgery for abdominal penetrating disease than whites. The lower prevalence of colectomy for chronically active UC among African Americans was even more dramatic compared to white patients. Though the difference in CD-related bowel resection may be partially explained by phenotypic differences such as less ileal disease and penetrating disease among African Americans, racial differences persisted after adjustment for disease site and behavior. Thus, we must explore other explanations such as racial disparities in health utilization, which may play an important role. Jha *et al.* have recently shown that there are growing disparities between African Americans and whites in the utilization of various surgical procedures (32). Thus, African Americans may have similar UC disease severity and indications for surgery as whites but may not be referred appropriately for colectomy. These racial differences in surgical intervention require further confirmation, and the potential role of patient preferences, physician bias, and health-care access need to be further explored.

Another limitation of this study is that the Vienna classification system does not adequately address the dynamic nature of CD behavior pattern. It is difficult to clinically distinguish between fibrostenosing disease and inflammatory stenosis at one given point in time. Additionally, clinical CD behavior tends to evolve from inflammatory disease to stricturing or penetrating disease over the duration of disease (33). However, our subgroup analysis of CD patients with disease duration of at least 5 yr yielded similar racial differences in penetrating disease. Interestingly, African Americans developed more stricturing disease than whites as CD behavior evolved with disease duration (38.1% *vs* 24.4%, $p = 0.03$). It is unclear whether this is due to biological differences or racial and ethnic disparities in access to therapy. Another criticism of the Vienna classification is that it categorizes fistulae that seemingly form secondary to distal fibrostenosing disease as "fistulizing" disease. This classification may, thus, underestimate the prevalence of stricturing CD, and this must be taken into consideration when comparing findings to other studies that classify secondary fistula as fibrostenotic disease.

One of the particular strengths of this study is the large diverse study population arising from multiple centers throughout North America, allowing sufficient power to detect racial differences while adjusting for various confounders. Our study findings will be greatly strengthened by future studies comparing African Americans to concurrent white IBD patients from the same center to further minimize potential confounding by regional factors, particularly for surgical outcomes. Further studies are warranted to characterize mutant allele frequencies of IBD susceptibility genes such as *NOD2/CARD15* in adult minority populations to determine correlation to phenotype. Overall, we have demonstrated the utility of a large multi-center phenotyping database.

In conclusion, we have demonstrated that there are racial differences in IBD phenotype that were previously poorly characterized in the adult population. Our findings underscore the potential importance of race and ethnic background as a source of heterogeneity of IBD phenotype in genetic studies. Racial variations in phenotype also have potential implications for diagnosis and management of IBD and its complications. Characterization of phenotypes in minority groups facilitates timely diagnosis of certain complications such as uveitis or perianal disease through understanding of which racial or ethnic groups are a particular risk. Furthermore, racial differences in disease phenotype may have implications for variations in response to medical therapy. One retrospective study has evaluated infliximab response specifically in Hispanics and found no differences compared to previous reports in predominant white populations (34). However, the vast majority of clinical trials in IBD have been underrepresented by minority groups. For example, both the National Cooperative Crohn's Disease Study (35) and the more

recent Accent I trial (36) have enrolled 5% or less African Americans, without adequate subgroup analysis to determine the influence of race on therapeutic response. With certain manifestations such as perianal disease being more common among African Americans and Hispanics, it is important to assess whether biological agents such as infliximab yield the same effectiveness as in whites (37, 38). Future observational studies of racial differences in response to medical therapy are necessary to determine whether concerted efforts are warranted to adequately represent racial and ethnic minorities in future trials. With the projection that, by 2050, half of the U.S. population will be represented by ethnic minorities, we must direct research efforts toward elucidating clinical disease patterns in these understudied populations. A better understanding of phenotype distributions in various racial and ethnic groups will potentiate more effective management of disease in an increasingly diverse IBD community.

ACKNOWLEDGMENTS

The authors of this work are indebted to the patients with IBD and to their families, who gave of themselves to make this work possible. This study was funded by the National Institutes of Health Grants U01DK62413 (Cedars-Sinai at Los Angeles), U01 DK062420 (University of Pittsburgh), U01DK62422 (University of Chicago), U01DK062423 (University of Toronto), U01DK62429 (Data Coordinating Center at the University of Chicago), U01DK62431 (Johns Hopkins), and U01DK62432 (Broad Institute). This research was also supported by the Burroughs Wellcome Clinical Scientist Awards in Translational Research (University of Chicago), the Eli and Edythe L. Broad Foundation (University of Chicago), and NIH-General Clinical Research Center Grants 5M01RR000055 (University of Chicago) and RR0052 (Johns Hopkins).

We gratefully acknowledge Kim Isaacs, M.D. (University of North Carolina), Duane Smoot, M.D. (Howard University), Theodore M. Bayless, M.D. (Johns Hopkins University), Carmen Cuffari, M.D. (Johns Hopkins University), Lisa Datta, M.S. (Johns Hopkins University), and Patricia Ushry (Johns Hopkins University) for their assistance in recruitment of African American patients. We thank the Quebec Inflammatory Bowel Disease Genetic Consortium and the following individuals for their participation in this study: Edmond Jean Bernard, M.D., F.R.C.P. (Hôpital Hôtel Dieu), Albert Cohen, M.D., F.R.C.P.(C.) (Jewish General Hospital), Guy Aumais, M.D., C.S.P.Q., F.R.C.P. (Hôpital Maisonneuve Rosemont), Gilles Jobin, M.D., F.R.C.P., M.Sc., Diane Langelier, M.D. (Centre Hospitalier Universitaire de Sherbrooke), Colette Deslandres, M.D. (Hôpital Sainte-Justine), Daniel Gaudet, M.D., Ph.D. (Centre Hospitalier de la Sagamie-Chicoutimie), and Sylviane Forget, M.D., F.R.C.P.(C.) (Montreal Children's Hospital). We would like to express our gratitude to the following physicians from the University of Toronto: Zane Cohen, M.D., Robin McLeod, M.D., and Gordon R. Greenberg, M.D. We are also grateful to Yin Yao, Ph.D. (Johns Hopkins University), for her statistical and methodological review of the manuscript.

Since the submission of our article, Basu et al. have published results from a smaller study supporting our findings that inflammatory EIMs including uveitis and arthritis are significantly increased in African American compared to white CD patients (Am J Gastroenterol 2005;100:2254–61).

STUDY HIGHLIGHTS

What Is Current Knowledge

- The incidence of inflammatory bowel disease in certain minorities approaches that of whites.
- The frequency of certain Crohn's disease susceptibility genes such as *NOD2/CARD15* vary by race.

What Is New Here

- The prevalence of esophagogastroduodenal and ileal involvement in Crohn's disease varies among races.
- Perianal disease is more common in African American and Hispanic Crohn's disease patients than in their white counterparts.
- There are racial differences in the prevalence of certain extraintestinal manifestations particularly uveitis.
- Family history of inflammatory bowel disease is less common in African Americans and Hispanics compared to whites.

Reprint requests and correspondence: Steven R. Brant, M.D., 1503 E. Jefferson Street, Rm. B-136, Baltimore, MD 21205.
Received September 9, 2005; accepted November 11, 2005.

REFERENCES

1. Gasche C, Scholmerich J, Brynskov J, et al. A simple classification of Crohn's disease: Report of the Working Party for the World Congresses of Gastroenterology, Vienna 1998. Inflamm Bowel Dis 2000;6:8–15.
2. Su CG, Judge TA, Lichtenstein GR. Extraintestinal manifestations of inflammatory bowel disease. Gastroenterol Clin North Am 2002;31:307–27.
3. Mintz R, Feller ER, Bahr RL, et al. Ocular manifestations of inflammatory bowel disease. Inflamm Bowel Dis 2004;10:135–9.
4. Raj V, Lichtenstein DR. Hepatobiliary manifestations of inflammatory bowel disease. Gastroenterol Clin North Am 1999;28:491–513.
5. Gasche C, Grundtner P. Genotypes and phenotypes in Crohn's disease: Do they help in clinical management? Gut 2005;54:162–7.
6. Simsek H, Schuman BM. Inflammatory bowel disease in 64 black patients: Analysis of course, complications, and surgery. J Clin Gastroenterol 1989;11:294–8.
7. Paul H, Jr, Barnes RW, Reese VE, et al. Crohn's disease in black patients. J Natl Med Assoc 1990;82:709–12.
8. Goldman CD, Kodner IJ, Fry RD, et al. Clinical and operative experience with non-Caucasian patients with Crohn's disease. Dis Colon Rectum 1986;29:317–21.

9. Straus WL, Eisen GM, Sandler RS, et al. Crohn's disease: Does race matter? The Mid-Atlantic Crohn's Disease Study Group. Am J Gastroenterol 2000;95:479–83.
10. Kugathasan S, Loizides A, Babusukumar U, et al. Comparative phenotypic and CARD15 mutational analysis among African American, Hispanic, and White children with Crohn's disease. Inflamm Bowel Dis 2005;11:631–8.
11. Deveaux PG, Kimberling J, Galandiuk S. Crohn's disease: Presentation and severity compared between black patients and white patients. Dis Colon Rectum 2005;48: 1404–9.
12. Calkins BM, Lilienfeld AM, Garland CF, et al. Trends in incidence rates of ulcerative colitis and Crohn's disease. Dig Dis Sci 1984;29:913–20.
13. Kurata JH, Kantor-Fish S, Frankl H, et al. Crohn's disease among ethnic groups in a large health maintenance organization. Gastroenterology 1992;102:1940–8.
14. Torres EA, De JR, Perez CM, et al. Prevalence of inflammatory bowel disease in an insured population in Puerto Rico during 1996. P R Health Sci J 2003;22:253–8.
15. Cosnes J, Cattan S, Blain A, et al. Long-term evolution of disease behavior of Crohn's disease. Inflamm Bowel Dis 2002;8:244–50.
16. Zankel E, Rogler G, Andus T, et al. Crohn's disease patient characteristics in a tertiary referral center: Comparison with patients from a population-based cohort. Eur J Gastroenterol Hepatol 2005;17:395–401.
17. Eidelwein A, Fiorino K, Thompson R, et al. Inflammatory bowel disease (IBD) in African American children, in: North American Society of Pediatric Gastroenterology, Hepatology, and Nutrition Annual Meeting. October 20–22, 2005, Salt Lake City, Utah. J Pediatr Gastroenterol Nutr 2005;41:539–40 (abstract).
18. Bonen DK, Cho JH. The genetics of inflammatory bowel disease. Gastroenterology 2003;124:521–36.
19. Cummings JR, Jewell DP. Clinical implications of inflammatory bowel disease genetics on phenotype. Inflamm Bowel Dis 2005;11:56–61.
20. Economou M, Trikalinos TA, Loizou KT, et al. Differential effects of NOD2 variants on Crohn's disease risk and phenotype in diverse populations: A metaanalysis. Am J Gastroenterol 2004;99:2393–404.
21. Roth MP, Petersen GM, McElree C, et al. Familial empiric risk estimates of inflammatory bowel disease in Ashkenazi Jews. Gastroenterology 1989;96:1016–20.
22. Inoue N, Tamura K, Kinouchi Y, et al. Lack of common NOD2 variants in Japanese patients with Crohn's disease. Gastroenterology 2002;123:86–91.
23. Merrill PT, Kim J, Cox TA, et al. Uveitis in the southeastern United States. Curr Eye Res 1997;16:865–74.
24. Rybicki BA, Major M, Popovich J Jr, et al. Racial differences in sarcoidosis incidence: A 5-year study in a health maintenance organization. Am J Epidemiol 1997;145: 234–41.
25. Shikama Y, Kobayashi K, Kasahara K, et al. Granuloma formation by artificial microparticles in vitro. Macrophages and monokines play a critical role in granuloma formation. Am J Pathol 1989;134:1189–99.
26. Kasahara K, Kobayashi K, Shikama Y, et al. The role of monokines in granuloma formation in mice: The ability of interleukin 1 and tumor necrosis factor-alpha to induce lung granulomas. Clin Immunol Immunopathol 1989;51:419–25.
27. White JM, O'Connor S, Winter HS, et al. Different patterns of inflammatory bowel disease between African-American children and other racial/ethnic groups in a large, multicenter registry of 1640 children, in: Digestive Disease Week and the 105th Annual Meeting of the American Gastroenterological Association. May 15–20, 2004, New Orleans, Louisiana, USA. Gastroenterology 2004;126:A457 (abstracts).
28. Mokbel M, Carbonnel F, Beaugerie L, et al. Effect of smoking on the long-term course of ulcerative colitis. Gastroenterol Clin Biol 1998;22:858–62.
29. Fraga XF, Vergara M, Medina C, et al. Effects of smoking on the presentation and clinical course of inflammatory bowel disease. Eur J Gastroenterol Hepatol 1997;9:683–7.
30. Picco MF, Bayless TM. Tobacco consumption and disease duration are associated with fistulizing and stricturing behaviors in the first 8 years of Crohn's disease. Am J Gastroenterol 2003;98:363–8.
31. Original data courtesy of Michael F. Picco, M.D., Ph.D., and Theodore M. Bayless, M.D.
32. Jha AK, Fisher ES, Li Z, et al. Racial trends in the use of major procedures among the elderly. N Engl J Med 2005;353:683–91.
33. Louis E, Collard A, Oger AF, et al. Behaviour of Crohn's disease according to the Vienna classification: Changing pattern over the course of the disease. Gut 2001;49:777–82.
34. Sanchez JM, Maldonado JC, Torres EA, et al. Infliximab in Hispanics: Characterization of response to infliximab in an ethnic minority with Crohn's disease. P R Health Sci J 2005;24:11–7.
35. Summers RW, Switz DM, Sessions JT, Jr, et al. National Cooperative Crohn's Disease Study: Results of drug treatment. Gastroenterology 1979;77:847–69.
36. Hanauer SB, Feagan BG, Lichtenstein GR, et al. Maintenance infliximab for Crohn's disease: The ACCENT I randomised trial. Lancet 2002;359:1541–9.
37. Present DH, Rutgeerts P, Targan S, et al. Infliximab for the treatment of fistulas in patients with Crohn's disease. N Engl J Med 1999;340:1398–405.
38. Ardizzone S, Maconi G, Colombo E, et al. Perianal fistulae following infliximab treatment: Clinical and endosonographic outcome. Inflamm Bowel Dis 2004;10:91–6.

ATTACHMENT 2

SAS Code for Tabulations from the IBDGC Phenotype Analysis Dataset in the NIDDK Repository

```
/*************************************************************************************/
/*
/* Program: R:\05_Users\Norma\IBDGC\PhenotypeData\NguyenPaper\Update\table1_updated.sas
/* Author:  Norma Pugh
/* Date:    28 July 2008
/* Revised: 03 October 2008 per DCC comments.
/* Revised: 24 October 2008 per DCC comments re: calculation of '% affected siblings'.
/* Purpose: Replicate table 1 results.
/*
/*************************************************************************************/
/* DATA SOURCE */
libname sasdb 'R:\05_Users\Norma\IBDGC\PhenotypeData';

/*******************/
/* INCLUDE FORMATS */
/*******************/
data fmts; set sasdb.formats;
proc format cntlin=fmts; run;
proc format; value yn 1='1=Y' 2='2=N'; run;

/************************/
/* GET STUDY POPULATION */
/************************/
data table1; merge sasdb.phenotyp(in=x1) sasdb.patients(in=x2); by suid; if x1 & x2;
 /* Define age at study entry */
 /* NOTE: Enrollments were from 2003-2005. Use enrollment year of 2005, per P.Schumm e-
mail. */
 ageentry=2005-yob; label ageentry='Age at study entry (yr)';

 /* Define age at diagnosis */
 agediag=diag_yr-yob; label agediag='Age at diagnosis (yr)';

 /* Define racial group */
 if race=1 & hispanic=2 then racegrp='2_White';
 if race=2 then racegrp='3_AA';
 if hispanic=1 and race^=2 then racegrp='4_Hispanic';
 label racegrp='Racial group';

 /* Define family history variables */
 label famhx='1st/2nd degree relative w/ IBD(CD or UC)'
       numsibaff='# Sibling(s) affected'
       numsib='# Sibling(s)';

 if chld_cd>0 or chld_uc>0 or chld_ibd>0 or fthr_ibd in(1,2,3) or mthr_ibd in(1,2,3) or
sib_cd>0 or sib_uc>0 or sib_ibd>0 or fam_hist=1 then famhx=1;
 else famhx=2;

 numsibaff=sib_cd+sib_uc+sib_ibd;
 numsib=sib_cd+sib_uc+sib_ibd+sib_unf;

 /* Define packs/day */
 if smoking=1 then numpacks=no_cigar/20; else numpacks=0;

 /* Define appendectomy 2 yrs prior to IBD */
 app2yr=2;

 if diag_yr-app_yr>=2 then app2yr=1;
```

```
 /* Output patients w/ IBD & appropriate racial group */
 if control=0 & diag in(1,2,3) & racegrp>'';
run;

/************************/
/* CHECK FOR DUPLICATES */
/************************/
data check_dups; set table1; keep suid; run;
proc sort data=check_dups nodup; by suid; run;

/***********************************************/
/* CREATE 'ALL' GROUPING FOR RACE & DENOMINATORS */
/***********************************************/
data table1; set table1; output; racegrp='1_All'; output; run;

proc freq data=table1 noprint; tables racegrp / out=denom(drop=percent
rename=(count=denom)); run;
proc print data=denom; title'Denominators: Racial group'; run;

/******************/
/* GET STATISTICS */
/******************/
proc sort data=table1; by racegrp; run;
proc means data=table1 n mean stddev; by racegrp; var ageentry agediag; title'Means/SDs:
Ages'; run;

proc freq data=table1 noprint; tables racegrp*sex / out=outsex(drop=percent
rename=(count=numer)); format sex sex.; run;
data outsex; merge outsex denom; by racegrp; pct=(numer/denom)*100; run;
proc print data=outsex(where=(sex=2)); title'Frequency counts: Female'; run;

proc freq data=table1 noprint; tables racegrp*diag / out=outdiag(drop=percent
rename=(count=numer)); format diag diag.; run;
data outdiag; merge outdiag denom; by racegrp; pct=(numer/denom)*100; run;
proc print data=outdiag; title'Frequency counts: Diagnosis'; run;

proc freq data=table1 noprint; tables racegrp*famhx / out=outfamhx(drop=percent
rename=(count=numer)); format famhx yn.; run;
data outfamhx; merge outfamhx denom; by racegrp; pct=(numer/denom)*100; run;
proc print data=outfamhx; title'Frequency counts: Family History'; run;

proc freq data=table1(where=(diag=1)); tables racegrp*famhx; format famhx yn.;
 title'Frequency counts: Family History for Patient Diagnosis of CD';
run;

proc freq data=table1(where=(diag=3)); tables racegrp*famhx; format famhx yn.;
 title'Frequency counts: Family History for Patient Diagnosis of UC';
run;

proc means data=table1 n sum; by racegrp; var numsibaff numsib;
 title'Sums: # Sibs Affected, # Sibs Total';
 title'Calculate sample proportion as # affected/# total';
run;

proc means data=table1 n mean; by racegrp; var numpacks; title'Means: Packs of cigarettes
per day AMONG SMOKERS & NON-SMOKERS'; run;
```

```
proc means data=table1(where=(smoking=1 & racegrp='1_All')) n mean; by racegrp; var
numpacks; title'Means: Packs of cigarettes per day AMONG SMOKERS'; run;

proc freq data=table1; tables racegrp*(smoking app2yr); format smoking smoking. app2yr
yn.;
 title'Frequency counts: Smoking at diagnosis, Appendectomy (2 yr prior to IBD)';
run;
```

```
/*******************************************************************************************/
/*
/* Program: R:\05_Users\Norma\IBDGC\PhenotypeData\NguyenPaper\Update\table2.sas
/* Author:  Norma Pugh
/* Date:    29 July 2008
/* Revised: 03 October 2008 per DCC comments.
/* Purpose: Replicate table 2 results.
/*
/*******************************************************************************************/
/* DATA SOURCE */
libname sasdb 'R:\05_Users\Norma\IBDGC\PhenotypeData';

/******************/
/* INCLUDE FORMATS */
/******************/
data fmts; set sasdb.formats;
proc format cntlin=fmts; run;
proc format; value yn 1='1=Y' 2='2=N'; value uc 1='1=Proctitis' 2='2=Left'
3='3=Extensive'; run;

/************************/
/* GET STUDY POPULATION */
/************************/
data table2; merge sasdb.phenotyp(in=x1) sasdb.patients(in=x2); by suid; if x1 & x2;

 /* Define racial group */
 if race=1 & hispanic=2 then racegrp='2_White';
 if race=2 then racegrp='3_AA';
 if hispanic=1 and race^=2 then racegrp='4_Hispanic';
 label racegrp='Racial group';

 /* Define Crohn's dx: Dx involvement: Upper gastro  */
 if jejunal=1 or gi=1 then crohndx_gastro=1; else crohndx_gastro=2;
 label crohndx_gastro='CD: Dx involvement: Upper GI';

 /* Define disease pattern */
 inflam=2; strict=2; penetrate=2;

 if behavior=1 then inflam=1;
 if behavior=2 then strict=1;
 if behavior=3 then penetrate=1;

 label inflam='Dx pattern: inflammatory'
       strict='Dx pattern: stricturing'
       penetrate='Dx pattern: penetrating';

 /* Define UC Disease Extent: Proctitis, Left-sided Colitis or Extensive Colitis */
 if (proctit=1 & left=2 & extensiv=2) then uc_dx_ext=1;
 else if (left=1 & extensiv=2) then uc_dx_ext=2;
 else if (extensiv=1) then uc_dx_ext=3;

 label uc_dx_ext='UC Dx Xtent:1=Proctitis,2=Left,3=Extensv';

 /* Output patients w/ IBD & appropriate racial group */
 if control=0 & diag in(1,2,3) & racegrp>'';
run;
```

```
/************************/
/* CHECK FOR DUPLICATES */
/************************/
data check_dups; set table2; keep suid; run;
proc sort data=check_dups nodup; by suid; run;

/*****************************************************************************/
/* CREATE 'ALL' GROUPING FOR RACE & CD DISEASE INVOLVEMENT UPPER GI DENOMINATORS */
/*****************************************************************************/
data table2_cd_upper; set table2;
 if diag=1;
 if (jejunal in(1,2) & gi in(1,2)) or (jejunal=1 & gi=3) or (gi=1 & jejunal=3); output;
racegrp='1_All'; output;
run;

proc freq data=table2_cd_upper noprint; tables racegrp / out=denom_cd_upper(drop=percent
rename=(count=denom)); run;
proc print data=denom_cd_upper; title'CD Dx Involvement Upper GI Denominators: Racial
group'; run;

/*****************************************************************************/
/* CREATE 'ALL' GROUPING FOR RACE & CD DISEASE INVOLVEMENT ESOPH. DENOMINATORS */
/*****************************************************************************/
data table2_cd_esoph; set table2;
 if diag=1 & gi in(1,2); output; racegrp='1_All'; output;
run;

proc freq data=table2_cd_esoph noprint; tables racegrp / out=denom_cd_esoph(drop=percent
rename=(count=denom)); run;
proc print data=denom_cd_esoph; title'CD Dx Involvement Esoph. Denominators: Racial
group'; run;

/*****************************************************************************/
/* CREATE 'ALL' GROUPING FOR RACE & CD DISEASE INVOLVEMENT JEJUNUM DENOMINATORS */
/*****************************************************************************/
data table2_cd_jej; set table2;
 if diag=1 & jejunal in(1,2); output; racegrp='1_All'; output;
run;

proc freq data=table2_cd_jej noprint; tables racegrp / out=denom_cd_jej(drop=percent
rename=(count=denom)); run;
proc print data=denom_cd_jej; title'CD Dx Involvement Jejunal Denominators: Racial
group'; run;


/*****************************************************************************/
/* CREATE 'ALL' GROUPING FOR RACE & CD DISEASE INVOLVEMENT ILEUM DENOMINATORS */
/*****************************************************************************/
data table2_cd_il; set table2;
 if diag=1 & ileal in(1,2); output; racegrp='1_All'; output;
run;

proc freq data=table2_cd_il noprint; tables racegrp / out=denom_cd_il(drop=percent
rename=(count=denom)); run;
proc print data=denom_cd_il; title'CD Dx Involvement Ileum Denominators: Racial group';
run;
```

```
/*****************************************************************************/
/* CREATE 'ALL' GROUPING FOR RACE & CD DISEASE INVOLVEMENT COLORECTAL DENOMINATORS */
/*****************************************************************************/
data table2_cd_colo; set table2;
 if diag=1 & colorect in(1,2); output; racegrp='1_All'; output;
run;

proc freq data=table2_cd_colo noprint; tables racegrp / out=denom_cd_colo(drop=percent
rename=(count=denom)); run;
proc print data=denom_cd_colo; title'CD Dx Involvement Colorectal Denominators: Racial
group'; run;

/*****************************************************************************/
/* CREATE 'ALL' GROUPING FOR RACE & CD DISEASE INVOLVEMENT PERIANAL DX DENOMINATORS */
/*****************************************************************************/
data table2_cd_peri; set table2;
 if diag=1 & perianal in(1,2); output; racegrp='1_All'; output;
run;

proc freq data=table2_cd_peri noprint; tables racegrp / out=denom_cd_peri(drop=percent
rename=(count=denom)); run;
proc print data=denom_cd_peri; title'CD Dx Involvement Perianal Dx Denominators: Racial
group'; run;

/*****************************************************************/
/* CREATE 'ALL' GROUPING FOR RACE & CD DISEASE PATTERN DENOMINATORS */
/*****************************************************************/
data table2_cd_dxpatt; set table2;
 if diag=1 & behavior in(1,2,3); output; racegrp='1_All'; output;
run;

proc freq data=table2_cd_dxpatt noprint; tables racegrp /
out=denom_cd_dxpatt(drop=percent rename=(count=denom)); run;
proc print data=denom_cd_dxpatt; title'CD Dx Pattern Denominators: Racial group'; run;

/*************************************************/
/* CREATE 'ALL' GROUPING FOR RACE & UC DENOMINATORS */
/*************************************************/
data table2_uc; set table2;
 if diag=3 & proctit in(1,2) & left in(1,2) & extensiv in(1,2); output; racegrp='1_All';
output;
run;

proc freq data=table2_uc noprint; tables racegrp / out=denom_uc(drop=percent
rename=(count=denom)); run;
proc print data=denom_uc; title'UC Denominators: Racial group'; run;

/******************/
/* GET STATISTICS */
/******************/
%macro table2(indata,var,denom,title);
 proc freq data=&indata noprint; tables racegrp*&var / out=outdx(drop=percent
rename=(count=numer)); format &var yn.; run;
 data outdx; merge outdx &denom; by racegrp; pct=(numer/denom)*100; run;
 proc print data=outdx(where=(&var=1)); title"Frequency counts: &title"; run;
%mend table2;
```

```
%table2(table2_cd_upper,crohndx_gastro,denom_cd_upper,%str(CD Disease Involvement - Upper
gastrointestinal));
%table2(table2_cd_esoph,gi,denom_cd_esoph,%str(CD Disease Involvement -
Esophagogastroduodenal));
%table2(table2_cd_jej,jejunal,denom_cd_jej,%str(CD Disease Involvement - Jejunum));
%table2(table2_cd_il,ileal,denom_cd_il,%str(CD Disease Involvement - Ileum));
%table2(table2_cd_colo,colorect,denom_cd_colo,%str(CD Disease Involvement - Colorectal));
%table2(table2_cd_peri,perianal,denom_cd_peri,%str(CD Disease Involvement - Perianal
disease));

%table2(table2_cd_dxpatt,inflam,denom_cd_dxpatt,%str(CD Disease Pattern - Inflammatory));
%table2(table2_cd_dxpatt,strict,denom_cd_dxpatt,%str(CD Disease Pattern - Stricturing));
%table2(table2_cd_dxpatt,penetrate,denom_cd_dxpatt,%str(CD Disease Pattern -
Penetrating));

proc freq data=table2_uc noprint; tables racegrp*uc_dx_ext / out=outdx(drop=percent
rename=(count=numer)); format uc_dx_ext uc.; run;
data outdx; merge outdx denom_uc; by racegrp; pct=(numer/denom)*100; run;
proc print data=outdx; title"Frequency counts: UC Disease Extent"; run;
```

```
/*************************************************************************************/
/*
/* Program: R:\05_Users\Norma\IBDGC\PhenotypeData\NguyenPaper\Update\table3.sas
/* Author:  Norma Pugh
/* Date:    29 July 2008
/* Revised: 03 October 2008 per DCC comments.
/* Purpose: Replicate table 3 results.
/*
/*************************************************************************************/
/* DATA SOURCE */
libname sasdb 'R:\05_Users\Norma\IBDGC\PhenotypeData';

/******************/
/* INCLUDE FORMATS */
/******************/
data fmts; set sasdb.formats;
proc format cntlin=fmts; run;
proc format; value yn 1='1=Y' 2='2=N'; run;

/************************/
/* GET STUDY POPULATION */
/************************/
data table3; merge sasdb.phenotyp(in=x1) sasdb.patients(in=x2); by suid; if x1 & x2;
 /* Define racial group */
 if race=1 & hispanic=2 then racegrp='2_White';
 if race=2 then racegrp='3_AA';
 if hispanic=1 and race^=2 then racegrp='4_Hispanic';
 label racegrp='Racial group';

 /* Define surgery variables */
 uc_colectomy=2;

 if surgery=1 then surgtime=surg_yr-diag_yr;
  else if surgery=2 then surgtime=review-diag_yr;

 if surgery=1 & diag=3 then uc_colectomy=1;

 label surgtime='Time to 1st surgery (yrs)'
       uc_colectomy='UC: Colectomy surgery';

 /* Output patients w/ IBD & appropriate racial group */
 if control=0 & diag in(1,2,3) & racegrp>'';
run;

/************************/
/* CHECK FOR DUPLICATES */
/************************/
data check_dups; set table3; keep suid; run;
proc sort data=check_dups nodup; by suid; run;

/*********************************************************************/
/* CREATE 'ALL' GROUPING FOR RACE & ABDOMINAL CD SURGERY DENOMINATORS */
/*********************************************************************/
data table3_abcd; set table3;
 if surgery=1 & diag=1 & op_ad>.; output; racegrp='1_All'; output;
run;
```

```
proc freq data=table3_abcd noprint; tables racegrp / out=denom_abcd(drop=percent
rename=(count=denom)); run;
proc print data=denom_abcd; title'Abdominal CD Surgery: Racial group'; run;

/*********************************************************************/
/* CREATE 'ALL' GROUPING FOR RACE & PERIANAL CD SURGERY DENOMINATORS */
/*********************************************************************/
data table3_percd; set table3;
 if surgery=1 & diag=1 & op_pd>.; output; racegrp='1_All'; output;
run;

proc freq data=table3_percd noprint; tables racegrp / out=denom_percd(drop=percent
rename=(count=denom)); run;
proc print data=denom_percd; title'Perianal CD Surgery: Racial group'; run;

/*********************************************************************/
/* CREATE 'ALL' GROUPING FOR RACE & BOWEL RESECTION SURGERY DENOMINATORS */
/*********************************************************************/
data table3_br; set table3;
 if surgery=1 & surg_br in(.,1,2); output; racegrp='1_All'; output;
run;

proc freq data=table3_br noprint; tables racegrp / out=denom_br(drop=percent
rename=(count=denom)); run;
proc print data=denom_br; title'Bowel Resection Surgery: Racial group'; run;

/*********************************************************************/
/* CREATE 'ALL' GROUPING FOR RACE & BOWEL DIVERSION SURGERY DENOMINATORS */
/*********************************************************************/
data table3_div; set table3;
 if surgery=1 & surg_div in(.,1,2); output; racegrp='1_All'; output;
run;

proc freq data=table3_div noprint; tables racegrp / out=denom_div(drop=percent
rename=(count=denom)); run;
proc print data=denom_div; title'Bowel Diversion Surgery: Racial group'; run;

/*********************************************************************/
/* CREATE 'ALL' GROUPING FOR RACE & PENETRATING DX SURGERY DENOMINATORS */
/*********************************************************************/
data table3_af; set table3;
 if surgery=1 & surg_af in(.,1,2); output; racegrp='1_All'; output;
run;

proc freq data=table3_af noprint; tables racegrp / out=denom_af(drop=percent
rename=(count=denom)); run;
proc print data=denom_af; title'Penetrating Dx Surgery: Racial group'; run;

/*********************************************************************/
/* CREATE 'ALL' GROUPING FOR RACE & PERIANAL DX SURGERY DENOMINATORS */
/*********************************************************************/
data table3_pf; set table3;
 if surgery=1 & surg_pf in(.,1,2); output; racegrp='1_All'; output;
run;

proc freq data=table3_pf noprint; tables racegrp / out=denom_pf(drop=percent
rename=(count=denom)); run;
```

```
proc print data=denom_pf; title'Perianal Dx Surgery: Racial group'; run;

/*********************************************************/
/* CREATE 'ALL' GROUPING FOR RACE & SURGERY DENOMINATORS */
/*********************************************************/
data table3_surg; set table3;
 if surgery=1; output; racegrp='1_All'; output;
run;

proc freq data=table3_surg noprint; tables racegrp / out=denom_surg(drop=percent
rename=(count=denom)); run;
proc print data=denom_surg; title'Surgery Denominators: Racial group'; run;

/*********************************************************/
/* CREATE 'ALL' GROUPING FOR RACE & OVERALL DENOMINATORS */
/*********************************************************/
data table3_all; set table3; output; racegrp='1_All'; output; run;

proc freq data=table3_all noprint; tables racegrp / out=denom_all(drop=percent
rename=(count=denom)); run;
proc print data=denom_all; title'Overall Denominators: Racial group'; run;

/*********************************************************/
/* CREATE 'ALL' GROUPING FOR RACE & COLECTOMY DENOMINATORS */
/*********************************************************/
data table3_col; set table3;
 if diag=3; output; racegrp='1_All'; output;
run;

proc freq data=table3_col noprint; tables racegrp / out=denom_col(drop=percent
rename=(count=denom)); run;
proc print data=denom_col; title'Colectomy Denominators: Racial group'; run;

/*********************************************************/
/* CREATE 'ALL' GROUPING FOR RACE & DYSPLASIA DENOMINATORS */
/*********************************************************/
data table3_dys; set table3;
 if surgery=1 & surg_dys in(1,2); output; racegrp='1_All'; output;
run;

proc freq data=table3_dys noprint; tables racegrp / out=denom_dys(drop=percent
rename=(count=denom)); run;
proc print data=denom_dys; title'Dysplasia Denominators: Racial group'; run;

/*********************************************************************/
/* CREATE 'ALL' GROUPING FOR RACE & CHRONIC REFRACTORY DX DENOMINATORS */
/*********************************************************************/
data table3_chr; set table3;
 if surgery=1 & surg_chr in(1,2); output; racegrp='1_All'; output;
run;

proc freq data=table3_chr noprint; tables racegrp / out=denom_chr(drop=percent
rename=(count=denom)); run;
proc print data=denom_chr; title'Chronic Refractory Dx Denominators: Racial group'; run;
```

```
/*****************************************************************/
/* CREATE 'ALL' GROUPING FOR RACE & FULMINANT COLITIS DENOMINATORS */
/*****************************************************************/
data table3_acu; set table3;
 if surgery=1 & surg_acu in(1,2); output; racegrp='1_All'; output;
run;

proc freq data=table3_acu noprint; tables racegrp / out=denom_acu(drop=percent
rename=(count=denom)); run;
proc print data=denom_acu; title'Fulminant colitis Denominators: Racial group'; run;


/******************/
/* GET STATISTICS */
/******************/
proc sort data=table3_abcd; by racegrp; run;
proc sort data=table3_percd; by racegrp; run;
proc sort data=table3_surg; by racegrp; run;
proc sort data=table3_all; by racegrp; run;

proc means data=table3_abcd n mean stddev; by racegrp; var op_ad;
 title'Means/SDs: # Surgeries: abdominal';
run;

proc means data=table3_percd n mean stddev; by racegrp; var op_pd;
 title'Means/SDs: # Surgeries: perianal';
run;

%macro table3(indata,var,denom,title);
 proc freq data=&indata noprint; tables racegrp*&var / out=outsurg(drop=percent
rename=(count=numer)); format &var yn.; run;
 data outsurg; merge outsurg &denom; by racegrp; pct=(numer/denom)*100; run;
 proc print data=outsurg(where=(&var=1)); title"Frequency counts: &title"; run;
%mend table3;

%table3(table3_br,surg_br,denom_br,%str(Bowel resection surgery));
%table3(table3_div,surg_div,denom_div,%str(Bowel diversion surgery));
%table3(table3_af,surg_af,denom_af,%str(Penetrating disease surgery));
%table3(table3_pf,surg_pf,denom_pf,%str(Perianal disease surgery));

%table3(table3_col(where=(diag=3)),uc_colectomy,denom_col,%str(UC Surgery: Colectomy));
%table3(table3_dys(where=(diag=3)),surg_dys,denom_dys,%str(UC Surgery: Dysplasia));
%table3(table3_chr(where=(diag=3)),surg_chr,denom_chr,%str(UC Surgery: Chronic refractory
disease));
%table3(table3_acu(where=(diag=3)),surg_acu,denom_acu,%str(UC Surgery: Fulminant
colitis));
```

```
/*************************************************************************************/
/*
/* Program: R:\05_Users\Norma\IBDGC\PhenotypeData\NguyenPaper\Update\table4.sas
/* Author:  Norma Pugh
/* Date:    29 July 2008
/* Revised: 03 October 2008 per DCC comments.
/* Purpose: Replicate table 4 results.
/*
/*************************************************************************************/
/* DATA SOURCE */
libname sasdb 'R:\05_Users\Norma\IBDGC\PhenotypeData';

/******************/
/* INCLUDE FORMATS */
/******************/
data fmts; set sasdb.formats;
proc format cntlin=fmts; run;

/************************/
/* GET STUDY POPULATION */
/************************/
data table4; merge sasdb.phenotyp(in=x1) sasdb.patients(in=x2); by suid; if x1 & x2;
 /* Define racial group */
 if race=1 & hispanic=2 then racegrp='2_White';
 if race=2 then racegrp='3_AA';
 if hispanic=1 and race^=2 then racegrp='4_Hispanic';
 label racegrp='Racial group';

 /* Output patients w/ IBD & appropriate racial group */
 if control=0 & diag in(1,2,3) & racegrp>'';

 /* Output patients w/ complete data, per table note */
 if j_si in(1,2) & j_as in(1,2) & eye_uv in(1,2) & skin_en in(1,2) & skin_py in(1,2) &
liv_psc in(1,2);
run;

/************************/
/* CHECK FOR DUPLICATES */
/************************/
data check_dups; set table4; keep suid; run;
proc sort data=check_dups nodup; by suid; run;

/*************************************************/
/* CREATE 'ALL' GROUPING FOR RACE & DENOMINATORS */
/*************************************************/
data table4; set table4; output; racegrp='1_All'; output; run;

proc freq data=table4 noprint; tables racegrp / out=denom(drop=percent
rename=(count=denom)); run;
proc print data=denom; title'Denominators: Racial group'; run;

/******************/
/* GET STATISTICS */
/******************/
%macro table4(var,out,title);
 proc freq data=table4 noprint; tables racegrp*&var / out=&out(drop=percent
rename=(count=numer)); format &var &var.; run;
```

```
 data &out; merge &out denom; by racegrp; pct=(numer/denom)*100; run;
 proc print data=&out(where=(&var=1)); title"Frequency counts: &title"; run;
%mend table4;

%table4(j_si,out_j_si,%str(Sacroiliitis (j_si)));
%table4(j_as,out_j_as,%str(Ankylosing spondylitis (j_as)));
%table4(eye_uv,out_eye_uv,%str(Uveitis (eye_uv)));
%table4(skin_en,out_skin_en,%str(Erythema nodosum (skin_en)));
%table4(skin_py,out_skin_py,%str(Pyoderma gangrenosum (skin_py)));
%table4(liv_psc,out_liv_psc,%str(Primary sclerosing cholangitis (liv_psc)));
```

ATTACHMENT 3

Confirmatory Stata Code Provided by the DCC for Selected Tabulations from the IBDGC Phenotype Analysis Dataset in the NIDDK Repository

```
  ___  ____  ____  ____  ____ tm
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/   10.1   Copyright 1984-2008
  Statistics/Data Analysis            StataCorp
                                      4905 Lakeway Drive
     Special Edition                  College Station, Texas 77845 USA
                                      800-STATA-PC        http://www.stata.com
                                      979-696-4600        stata@stata.com
                                      979-696-4601 (fax)

60-user Stata for Macintosh (network) license expires 30 Jun 2009:
       Serial number:  81910043852
         Licensed to:  Phil Schumm
                       Department of Health Studies

Notes:
      1.  (-m# option or -set memory-) 10.00 MB allocated to data
      2.  (-v# option or -set maxvar-) 5000 maximum variables
      3.  Command line editing disabled
      4.  Stata running in batch mode


running /usr/local/bin/profile.do ...
--------------------------------------------------------------------------------
----------
       log:  /tmp/statalog_092908_162121.log
  log type:  text
 opened on:  29 Sep 2008, 16:21:21

. do reports/nguyen-replication

. // $Id: nguyen-replication.do 453 2008-09-29 21:18:59Z pschumm $
.
. // replicate results from Geoff Nguyen's paper, using dataset submitted to RTI
.
. version 10

. clear

.
. // grab data submitted to RTI, excluding those subjects that were not part of
. // Geoff's analysis
. insheet using raw/nguyen-sample.txt
(1 var, 1123 obs)

. isid suid, so
(data now sorted by suid)

. tempfile nguyen_sample

. save `"`nguyen_sample'"'
file /var/folders/h6/h6Fm5HMPHWmZr-ei8VIDEk+++TI/-Tmp-//St02364.000002 saved

.
. insheet using tmp/ibdgc-rti-1.0/phenotypes.txt, clear
(74 vars, 4761 obs)
```

```
. isid suid, so
(data now sorted by suid)

. merge suid using `"`nguyen_sample'"'

. keep if _merge==3
(3638 observations deleted)

. drop _merge

.
. // generate race/ethnicity variable for use below
. gen ethgroup = 1 if race==1 & hispanic==2
(294 missing values generated)

. replace ethgroup = 2 if race==2
(126 real changes made)

. replace ethgroup = 3 if hispanic==1 & race!=2
(168 real changes made)

. lab def ethgroup 1 "White" 2 "Black" 3 "non-Black Hispanic"

. lab val ethgroup ethgroup

. tab ethgroup

          ethgroup |      Freq.     Percent        Cum.
-------------------+-----------------------------------
             White |        829       73.82       73.82
             Black |        126       11.22       85.04
non-Black Hispanic |        168       14.96      100.00
-------------------+-----------------------------------
             Total |      1,123      100.00

.
.
. // proportion of affected siblings (Table B)
. // note: here we ignore sibs with unknown affection status, as Geoff did
. gen no_affected_sibs = sib_cd+sib_uc+sib_ibd

. gen no_sibs = sib_cd+sib_uc+sib_ibd+sib_unf

. ratio no_affected_sibs / no_sibs if ethgroup=="Black":ethgroup

Ratio estimation                    Number of obs    =     126

     _ratio_1: no_affected_sibs/no_sibs

------------------------------------------------------------
             |             Linearized
             |      Ratio   Std. Err.     [95% Conf. Interval]
-------------+----------------------------------------------
    _ratio_1 |   .0252101   .0119009      .0016567    .0487634
------------------------------------------------------------

.
```

```
.
. // UC disease extent (Table D)
. // switch to maximal-extent coding
. gen uc_dis_extent = 1 if proctit==1 & left==2 & extensiv==2
(1098 missing values generated)

. replace uc_dis_extent = 2 if left==1 & extensiv==2
(114 real changes made)

. replace uc_dis_extent = 3 if extensiv==1
(263 real changes made)

. lab def uc_dis_extent 1 "proctitis" 2 "left-sided colitis" 3 "extensive colitis"

. lab val uc_dis_extent uc_dis_extent

. tab uc_dis_extent ethgroup if diag==3, col

+-------------------+
| Key               |
|-------------------|
|     frequency     |
| column percentage |
+-------------------+

                   |             ethgroup
     uc_dis_extent |     White      Black  non-Black |     Total
-------------------+---------------------------------+----------
         proctitis |        18          4          2 |        24
                   |      6.19      13.79       3.77 |      6.43
-------------------+---------------------------------+----------
left-sided colitis |        90         10          9 |       109
                   |     30.93      34.48      16.98 |     29.22
-------------------+---------------------------------+----------
 extensive colitis |       183         15         42 |       240
                   |     62.89      51.72      79.25 |     64.34
-------------------+---------------------------------+----------
             Total |       291         29         53 |       373
                   |    100.00     100.00     100.00 |    100.00


.
.
. // CD surgery (Table F)
. replace op_ad = 0 if surgery==2
(655 real changes made)

. replace op_pd = 0 if surgery==2
(655 real changes made)

. sum op_ad if diag==1 & ethgroup=="Black":ethgroup

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       op_ad |        79     .721519     1.13156          0          5

. sum op_pd if diag==1 & ethgroup=="Black":ethgroup
```

```
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       op_pd |        79    .3164557    .7079096          0          4

.
.
. // compute time till first surgery (Table F)
. gen time_to_surgery = (surg_yr - diag_yr) if surgery==1
(674 missing values generated)

. replace time_to_surgery = (review - diag_yr) if surgery==2
(654 real changes made)

. stset time_to_surgery, fail(surgery==1) if(diag==1)

     failure event:  surgery == 1
obs. time interval:  (0, time_to_surgery]
 exit on or before:  failure
               if:  diag==1

--------------------------------------------------------------------------------
     1123  total obs.
       20  event time missing (time_to_surgery>=.)              PROBABLE ERROR
      427  ignored per request (if(), etc.)
      133  obs. end on or before enter()
--------------------------------------------------------------------------------
      543  obs. remaining, representing
      263  failures in single record/single failure data
     3571  total analysis time at risk, at risk from t =         0
                             earliest observed entry t =         0
                                  last observed exit t =        35

.
. // median as estimated by Weibull model
. streg, dist(weibull)

         failure _d:  surgery == 1
   analysis time _t:  time_to_surgery

Fitting constant-only model:

Iteration 0:   log likelihood = -572.49241
Iteration 1:   log likelihood = -569.20507
Iteration 2:   log likelihood = -569.20227
Iteration 3:   log likelihood = -569.20227

Fitting full model:
Iteration 0:   log likelihood = -569.20227

Weibull regression -- log relative-hazard form

No. of subjects =          543                     Number of obs   =       543
No. of failures =          263
Time at risk    =         3571
                                                   LR chi2(0)      =      0.00
Log likelihood  =   -569.20227                     Prob > chi2     =         .
```

```
------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
-------------+----------------------------------------------------------------
       /ln_p |   .1249559   .0471767     2.65   0.008     .0324913    .2174205
-------------+----------------------------------------------------------------
           p |   1.133098   .0534558                      1.033025    1.242867
         1/p |   .8825358   .0416351                      .8045916    .9680309
------------------------------------------------------------------------------

. predict median, median
(option median time assumed; predicted median time)
(580 missing values generated)

. sum median

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      median |       543    9.449808           0   9.449808   9.449808

.
. // survivor function at 2, 5, and 10 years
. sts list, at(2 5 10)

         failure _d:  surgery == 1
   analysis time _t:  time_to_surgery

              Beg.                       Survivor      Std.
    Time     Total     Fail             Function     Error     [95% Conf. Int.]
-------------------------------------------------------------------------------
       2       454       74              0.8604    0.0151     0.8278    0.8873
       5       269       78              0.6746    0.0222     0.6290    0.7159
      10       132       59              0.4801    0.0269     0.4266    0.5316
-------------------------------------------------------------------------------
Note:  survivor function is calculated over full data and evaluated at
       indicated times; it is not calculated from aggregates shown at left.

.
. // summarize time till first surgery separately by ethgroup
. xi: streg i.ethgroup, dist(weibull)
i.ethgroup        _Iethgroup_1-3      (naturally coded; _Iethgroup_1 omitted)

         failure _d:  surgery == 1
   analysis time _t:  time_to_surgery

Fitting constant-only model:

Iteration 0:   log likelihood = -572.49241
Iteration 1:   log likelihood = -569.20507
Iteration 2:   log likelihood = -569.20227
Iteration 3:   log likelihood = -569.20227

Fitting full model:

Iteration 0:   log likelihood = -569.20227
Iteration 1:   log likelihood = -566.49388
```

```
Iteration 2:   log likelihood = -566.39486
Iteration 3:   log likelihood = -566.39475
Iteration 4:   log likelihood = -566.39475

Weibull regression -- log relative-hazard form

No. of subjects =          543                     Number of obs   =       543
No. of failures =          263
Time at risk    =         3571
                                                   LR chi2(2)      =      5.62
Log likelihood  =   -566.39475                     Prob > chi2     =    0.0604

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
_Iethgroup_2 |   .8198009   .1739502    -0.94   0.349     .5408731    1.242571
_Iethgroup_3 |    1.46634   .2658532     2.11   0.035     1.027799    2.091996
-------------+----------------------------------------------------------------
       /ln_p |   .1382062   .0474798     2.91   0.004     .0451475    .2312648
-------------+----------------------------------------------------------------
           p |   1.148212   .0545169                      1.046182    1.260193
         1/p |   .8709191    .041351                      .7935293    .9558565
------------------------------------------------------------------------------

. predict median_by_ethgroup, median
(option median time assumed; predicted median time)
(580 missing values generated)

. tabstat median_by_ethgroup, by(ethgroup)

Summary for variables: median_by_ethgroup
     by categories of: ethgroup

        ethgroup |      mean
-----------------+----------
           White |  9.659752
           Black |  11.48468
non-Black Hispan |  6.921323
-----------------+----------
           Total |  9.447864
----------------------------

. sts list, at(2 5 10) by(ethgroup)

         failure _d:  surgery == 1
   analysis time _t:  time_to_surgery

              Beg.                      Survivor      Std.
    Time     Total     Fail             Function     Error     [95% Conf. Int.]
-------------------------------------------------------------------------------
White
       2       331       61              0.8452    0.0183     0.8054    0.8774
       5       201       54              0.6734    0.0255     0.6206    0.7206
      10       108       42              0.4931    0.0305     0.4320    0.5512
Black
       2        55        3              0.9476    0.0295     0.8461    0.9828
       5        37        9              0.7641    0.0602     0.6201    0.8594
```

```
      10        18         6              0.5889    0.0796     0.4175    0.7254
non-Black Hispanic
       2        68        10              0.8700    0.0385     0.7713    0.9280
       5        31        15              0.6053    0.0636     0.4692    0.7168
      10         6        11              0.2826    0.0754     0.1479    0.4337
-------------------------------------------------------------------------------
Note:  survivor function is calculated over full data and evaluated at
       indicated times; it is not calculated from aggregates shown at left.

.
.
. // fulminant colitis (Table F)
. tab ethgroup surg_acu if diag==3, row

+----------------+
| Key            |
|----------------|
|   frequency    |
| row percentage |
+----------------+

                   |            surg_acu
          ethgroup |         1          2          3 |     Total
-------------------+---------------------------------+----------
             White |         6         40          1 |        47
                   |     12.77      85.11       2.13 |    100.00
-------------------+---------------------------------+----------
             Black |         0          4          0 |         4
                   |      0.00     100.00       0.00 |    100.00
-------------------+---------------------------------+----------
non-Black Hispanic |         2         17          0 |        19
                   |     10.53      89.47       0.00 |    100.00
-------------------+---------------------------------+----------
             Total |         8         61          1 |        70
                   |     11.43      87.14       1.43 |    100.00


.
end of do-file
```

ATTACHMENT 4

Description of the Genotypic Data Housed in the NIDDK Genetics Repository at Rutgers

NIDDK IBDGC Crohn's Disease Genome-Wide Association Study

:Study name: NIDDK IBDGC Crohn's Disease Genome-Wide Association Study
:Study report name: NIDDK IBD Genetics Consortium Crohn's Disease Genome-Wide Association Study
:Version: $Revision: 432 $
:Date: $Date: 2008-03-28 17:46:54 -0500 (Fri, 28 Mar 2008) $

Description

This dataset contains data from a genome-wide association study performed with 968 IBD-affected cases and 995 unrelated controls using the Illumina HumanHap300 Genotyping BeadChip. Cases were selected to have Crohn's disease with ileal involvement, and controls were matched to cases based on sex and year of birth. Subjects were drawn from two cohorts: (1) persons with non-Jewish, European ancestry (561 cases and 563 controls), and (2) persons with Jewish ancestry (407 cases and 432 controls). Genotyping was performed at the Feinstein Institute for Medical Research.

Seven-hundred and fifty-four of the samples (468 cases and 286 controls) were taken from the NIDDK IBD Genetics Consortium cell line repository. These samples are identified in the file dbGaP_consent.txt. The subject IDs for these individuals may be used to request corresponding samples for follow-up research through the repository. In addition, complete phenotype data for these individuals are available, together with the Consortium's phenotyping manual and the forms used to collect the data. The remaining 1,209 samples were obtained from pre-existing collections ascertained through Cedars-Sinai Medical Center, Johns Hopkins University, University of Chicago, University of Montreal, University of Pittsburgh, University of Toronto, and the New York Health project (controls only). For these samples, only sex, cohort (Jewish vs. non-Jewish), and age at diagnosis (cases only) are available.

Two-hundred and three individuals from among the pre-existing samples did not provide consent to release their genotype data (designated as consent group 2 in the file dbGaP_consent.txt). Thus, individual genotype data are only provided for 1,760 samples. To compensate for this, we have provided summary results for each SNP. These are based on a stratified analysis testing case/control association. Fifty-one samples had a call rate less than 93% and were therefore excluded from this analysis, leaving an overall sample size of 1,963 - 51 = 1,912.

X Chromosome Heterozygosity

Nine samples have X chromosome heterozygosity that is neither consistent nor inconsistent with their phenotypic sex. One of these (3019572) was found to have Turner Syndrome. The remaining 8 (3001651, 3002191, 3003302, 3003339, 3005474, 3006537, 3014051, 3017976) have heterozygosity ranging from 35-76%.

Type

Case-control study, stratified by ancestry (European non-Jewish vs. Jewish).

Disease Name(s)

- Crohn's disease
- Ulcerative colitis
- Inflammatory bowel disease

Inclusion/Exclusion Criteria

Cases were selected to have Crohn's disease with ileal involvement. After genotyping was completed and updated phenotype information was obtained, two (non-repository) cases were identified as having Indeterminate Colitis instead of Crohn's disease, and one Crohn's case was found not to have ileal involvement (these three samples are nonetheless included in the dataset).

Diagnosis of IBD required (i) one or more of the following symptoms: diarrhea, rectal bleeding, abdominal pain, fever or complicated perianal disease; (ii) occurrence of symptoms on two or more occasions separated by at least 8 weeks or ongoing symptoms of at least 6 weeks' duration and (iii) objective evidence of inflammation from radiologic, endoscopic and histologic evaluation. Ileal Crohn's disease involvement was defined as mucosal ulceration, cobblestoning, stricturing or bowel wall thickening from endoscopy reports, barium X-rays, operative reports and/or pathology resection specimen reports. Individuals with either 'ileal only' or 'ileocolonic' were included.

Within each of the non-Jewish and Jewish cohorts, controls were matched to cases based on sex and year of birth. In addition, controls were required to meet the following inclusion criteria: (1) no history of IBD among 1st and 2nd degree relatives, (2) never been diagnosed with IBD, and (3) never experienced chronic diarrhea, unexplained rectal bleeding, or unexplained weight loss.

Relevant Publications

[1] R. H. Duerr, K. D. Taylor, S. R. Brant, J. D. Rioux, M. S. Silverberg, M. J. Daly, A. H. Steinhart, C. Abraham, M. Regueiro, A. Griffiths, T. Dassopoulos, A. Bitton, H. Yang, S. Targan, L. W. Datta, E. O. Kistner, L. P. Schumm, A. T. Lee, P. K. Gregersen, M. M. Barmada, J. I. Rotter, D. L. Nicolae, and J. H. Cho. A genome-wide association study identifies il23r as an inflammatory bowel disease gene. Science, 314(5804):1461╤1463, 2006. PMID: 17068223.

[2] J. D. Rioux, R. J. Xavier, K. D. Taylor, M. S. Silverberg, P. Goyette, A. Huett, T. Green, P. Kuballa, M. M. Barmada, L. W. Datta, Y. Y. Shugart, A. M. Griffiths, S. R. Targan, A. F. Ippoliti, E.-J. Bernard, L. Mei, D. L. Nicolae, M. Regueiro, L. P. Schumm, A. H. Steinhart, J. I. Rotter, R. H. Duerr, J. H. Cho, M. J. Daly, and S. R. Brant. Genome-wide association study identifies new susceptibility loci for crohn disease and implicates autophagy in disease pathogenesis. Nat Genet, 39(5):596╤604, 2007. PMID: 17435756.

Attribution
==========

:Principal Investigator: Judy H. Cho, MD
:Affiliation: Department of Medicine, Yale University, New Haven
:Funding Source:

    NIDDK DK62431 (Steven R. Brant), DK62422 (Judy H. Cho), DK62420 (Richard H. Duerr),
    DK62432 (John D. Rioux), DK62423 (Mark S. Silverberg), DK62413 (Kent D. Taylor),
    and DK62429 (Judy H. Cho)