

Dataset Integrity Check for The  
Predicting Response to Standardized  
Pediatric Colitis Therapy (PROTECT)  
Study

Prepared by NIDDK-CR  
August 31, 2022

## Contents

1 Standard Disclaimer .....	2
2 Study Background .....	2
3 Archived Datasets .....	2
4 Statistical Methods .....	2
5 Results .....	3
6 Conclusions .....	3
7 References .....	3
Table A: Variables used to replicate Table 1 – Characteristics of PROTECT treated population by initial therapy .....	4
Table B1: Comparison of values computed in integrity check to reference article Table 1 (Total and 5-ASA) .....	6
Table B2: Comparison of values computed in integrity check to reference article Table 1 (Oral CS and IV CS) .....	9
Attachment A: SAS Code .....	12

## **1 Standard Disclaimer**

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 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.

## **2 Study Background**

This multicenter open-label study was designed to evaluate the safety and efficacy of standardized initial therapy using either mesalamine or corticosteroids then mesalamine for the treatment of children and adolescents newly diagnosed with ulcerative colitis. The study investigated the hypothesis that response to the initial 4 weeks of therapy as well as specific clinical, genetic, and immune parameters determined during the initial course of therapy predicted severe disease as reflected by need for escalation of medical therapy or surgery.

Participants were assigned to one of two initial therapeutic plans (mesalamine only or prednisone/liquid equivalent prednisolone followed by mesalamine) depending upon initial disease severity determined by the validated multi-dimensional Pediatric Ulcerative Colitis Activity Index (PUCAI). Biospecimens were obtained at diagnosis, and subsequently following the initiation of therapy at weeks 4, 12, and 52 (blood and stool at weeks 4 and 12; blood, stool, and colonic tissue at week 52). Follow-up clinic visits were conducted for a minimum of 1 year to a maximum of 5 years depending on when the participant enrolled. Adherence to mesalamine dosing was monitored using a state of the art electronic Medication Event Monitoring System (MEMS®).

## **3 Archived Datasets**

All data files, as provided by the Data Coordinating Center (DCC), are located in the PROTECT folder in the data package. For this replication, variables were taken from the “primary\_manuscript\_nidkk\_full.sas7bdat” dataset.

## **4 Statistical Methods**

Analyses were performed to replicate results for the data in the publication by Hyams et al. [1]. To verify the integrity of the data, only descriptive statistics were computed.

## 5 Results

For Table 1 in the publication [1], Characteristics of PROTECT treated population by initial therapy, Table A lists the variables that were used in the replication, and Tables B1 and B2 compare the results calculated from the archived data files to the results in Table 1. The results of the replication are within expected variation of the published results.

## 6 Conclusions

The NIDDK Central Repository is confident that the PROTECT data files to be distributed are a true copy of the study data.

## 7 References

[1] Hyams JS, Davis S, Mack DR, Boyle B, Griffiths AM, LeLeiko NS, Sauer CG, Keljo DJ, Markowitz J, Baker SS, Rosh J, Baldassano RN, Patel A, Pfefferkorn M, Otley A, Heyman M, Noe J, Oliva-Hemker M, Rufo P, Strople J, Ziring D, Guthery SL, Sudel B, Benkov K, Wali P, Moulton D, Evans J, Kappelman MD, Marquis A, Sylvester FA, Collins MH, Venkateswaran S, Dubinsky M, Tangpricha V, Spada KL, Britt A, Saul B, Gotman N, Wang J, Serrano J, Kugathasan S, Walters T, Denson LA. Factors associated with early outcomes following standardised therapy in children with ulcerative colitis (PROTECT): a multicentre inception cohort study. *The Lancet: Gastroenterology & Hepatology*, 2(12), 855-868, December 2017. doi: [https://doi.org/10.1016/S2468-1253\(17\)30252-2](https://doi.org/10.1016/S2468-1253(17)30252-2)

**Table A:** Variables used to replicate Table 1 – Characteristics of PROTECT treated population by initial therapy

<b>Table Variable</b>	<b>dataset.variable</b>
Initial therapy	primary_manuscript_niddk_full.initial_trt_c4
Age (mean ± SD)	primary_manuscript_niddk_full.ageyr primary_manuscript_niddk_full.age_ge12
Female (%)	primary_manuscript_niddk_full.female
Non-white (%)	primary_manuscript_niddk_full.race_nonwhite
Hispanic/Latino (%)	primary_manuscript_niddk_full.hispanic
Weight z-score (mean ± SD)	primary_manuscript_niddk_full.bl_waz
Height z-score (mean ± SD)	primary_manuscript_niddk_full.bl_haz
BMI z-score (mean ± SD)	primary_manuscript_niddk_full.bl_bmiz
Hospitalized at baseline (%)	primary_manuscript_niddk_full.bl_hosp
Disease extent	primary_manuscript_niddk_full.paris_class
PUCAI (mean ± SD)	primary_manuscript_niddk_full.puc primary_manuscript_niddk_full.puc_c4
Abdominal Pain	primary_manuscript_niddk_full.puc_abdominal_pain_C2
Diarrhea	primary_manuscript_niddk_full.puc_diarrhea_c2
Rectal bleeding	primary_manuscript_niddk_full.puc_rectal_bleed_C2
Total Mayo score (mean ± SD)	primary_manuscript_niddk_full.total_mayo_eef primary_manuscript_niddk_full.total_mayo_c3
Total Mayo score % ≥ 11	primary_manuscript_niddk_full.total_mayo_ge11
Mayo endoscopy sub-score (mean ± SD)	primary_manuscript_niddk_full.bl_endo_mayo
Relative rectal sparing	primary_manuscript_niddk_full.rectal_inflam
Macroscopic patchiness	primary_manuscript_niddk_full.macro_patch_c2
Cecal patch	primary_manuscript_niddk_full.peri_inflam_c2
Non-specific macroscopic gastritis	primary_manuscript_niddk_full.macro_gastritis
Microscopic gastritis	primary_manuscript_niddk_full.micro_gastritis
Rectal biopsy eosinophilic inflammation (count > 32 hpf)	primary_manuscript_niddk_full.eosgrade_c2
Rectal biopsy surface villiform changes	primary_manuscript_niddk_full.cbh_villiform
Rectal biopsy basal plasmacytosis	primary_manuscript_niddk_full.cbh_bplasmac
Hemoglobin g/dL	primary_manuscript_niddk_full.hb_cv primary_manuscript_niddk_full.hb_c2 primary_manuscript_niddk_full.hb_lt12
Platelet count (x10 <sup>9</sup> /L)	primary_manuscript_niddk_full.plt_si primary_manuscript_niddk_full.plt_c2
White blood count (x10 <sup>9</sup> /L)	primary_manuscript_niddk_full.wbc_cv primary_manuscript_niddk_full.wbc_c2
ESR (mm/hr)	primary_manuscript_niddk_full.esr_cv primary_manuscript_niddk_full.esr_c3
CRP/hsCRP	primary_manuscript_niddk_full.crp_uln primary_manuscript_niddk_full.crp_2uln
Albumin (g/dL)	primary_manuscript_niddk_full.alb_cv primary_manuscript_niddk_full.alb_c2
Fecal calprotectin (mcg/g)	primary_manuscript_niddk_full.calprotectin

	primary_manuscript_niddk_full.calprotectin_c3
Fecal OPG (pg/mL)	primary_manuscript_niddk_full.opg primary_manuscript_niddk_full.opg_c3
25-OH Vitamin D (ng/mL)	primary_manuscript_niddk_full.vitd primary_manuscript_niddk_full.vitd_c3
ANCA	primary_manuscript_niddk_full.anca_pos primary_manuscript_niddk_full.anca_gt100
OmpC	primary_manuscript_niddk_full.ompc primary_manuscript_niddk_full.ompc_ge12
CBir1	primary_manuscript_niddk_full.cbir1_pos

**Table B1:** Comparison of values computed in integrity check to reference article Table 1 (Total and 5-ASA)

	Total Publication (n=428)	Total DSIC (n=431)	Diff. (n=3)	5-ASA Publication (n=136)	5-ASA DSIC (n=136)	Diff. (n=0)
Age (mean ± SD)	12.7 ± 3.3	12.6 ± 3.3	0.1 ± 0.0	12.8 ± 3.3	12.8 ± 3.3	0.0 ± 0.0
% ≥ 12 years n (%)	289 (68)	290 (67)	1 (1)	92 (68)	92 (68)	0 (0)
Female (%)	212 (50)	214 (49.6)	2 (0.4)	69 (51)	69 (51)	0 (0)
Non-white (%)	69/420 (16)	70/423 (16.5)	1/3 (0.5)	21/134 (16)	21/134 (16)	0/0 (0)
Hispanic/Latino (%)	38/424 (9)	38/427 (8.9)	0/3 (0.1)	18/134 (13)	18/134 (13)	0/0 (0)
Weight z-score (mean ± SD)	-0.1 ± 1.2	-0.1 ± 1.2	0.0 ± 0.0	0.0 ± 1.1	0.0 ± 1.1	0.0 ± 0.0
Height z-score (mean ± SD)	0.1 ± 1.0	0.1 ± 1.0	0.0 ± 0.0	0.1 ± 1.0	0.1 ± 1.0	0.0 ± 0.0
BMI z-score (mean ± SD)	-0.2 ± 1.3	-0.2 ± 1.3	0.0 ± 0.0	0.0 ± 1.2	0.0 ± 1.2	0.0 ± 0.0
Hospitalized at baseline (%)	166 (39)	167 (39)	1 (0)	4 (3)	4 (3)	0 (0)
Disease extent						
Proctosigmoiditis	29 (7)	29 (7)	0 (0)	22 (16)	22 (16)	0 (0)
Left-sided colitis	44 (10)	45 (10)	1 (0)	28 (21)	28 (21)	0 (0)
Extensive/Pancolitis/Unassessable	355 (83)	357 (83)	2 (0)	86 (63)	86 (63)	0 (0)
PUCAI (mean ± SD)	50.0 ± 19.9	50.0 ± 20.0	0.0 ± 0.1	31.1 ± 13.3	31.1 ± 13.3	0.0 ± 0.0
% 10-30 (Mild)	102 (24)	104 (24)	2 (0)	84 (62)	84 (62)	0 (0)
% 35-60 (Moderate)	185 (43)	186 (43)	1 (0)	50 (37)	50 (37)	0 (0)
% ≥ 65 (Severe)	141 (33)	141 (33)	0 (0)	2 (1)	2 (1)	0 (0)
Abdominal Pain	354 (83)	355 (82)	1 (1)	97 (71)	97 (71)	0 (0)
Diarrhea	393 (92)	395 (92)	2 (0)	104 (76)	104 (76)	0 (0)
Rectal bleeding	398 (93)	401 (93)	3 (0)	117 (86)	117 (86)	0 (0)
Total Mayo score (mean ± SD)	7.8 ± 2.5	7.8 ± 2.5	0.0 ± 0.0	5.5 ± 1.8	5.5 ± 1.8	0.0 ± 0.0
% 0-6	125 (29)	127 (29)	2 (0)	99 (73)	99 (73)	0 (0)
% 7-9	183 (43)	183 (42)	0 (1)	35 (26)	35 (26)	0 (0)
% ≥ 10	120 (28)	121 (28)	1 (0)	2 (1)	2 (1)	0 (0)
Total Mayo score % ≥ 11	71 (17)	71 (16)	0 (1)	0 (0)	0 (0)	0 (0)

	Total Publication (n=428)	Total DSIC (n=431)	Diff. (n=3)	5-ASA Publication (n=136)	5-ASA DSIC (n=136)	Diff. (n=0)
Mayo endoscopy sub-score (mean ± SD)	2.2 ± 0.7	2.2 ± 0.7	0.0 ± 0.0	1.8 ± 0.6	1.8 ± 0.6	0.0 ± 0.0
% Mayo 1	59 (14)	60 (14)	1 (0)	43 (32)	43 (32)	0 (0)
% Mayo 2	224 (52)	226 (52)	2 (0)	79 (58)	79 (58)	0 (0)
% Mayo 3	145 (34)	145 (34)	0 (0)	14 (10)	14 (10)	0 (0)
Relative rectal sparing	38/427 (9)	38/430 (9)	0/3 (0)	17/136 (13)	17/136 (13)	0/0 (0)
Macroscopic patchiness	37/427 (9)	37/430 (9)	0/3 (0)	23/136 (17)	23/136 (17)	0/0 (0)
Cecal patch	29/397 (7)	29/400 (7)	0/3 (0)	17/135 (13)	17/135 (13)	0/0 (0)
Non-specific macroscopic gastritis	114/420 (27)	114/423 (27)	0/3 (0)	33/131 (25)	33/131 (25)	0/0 (0)
Microscopic gastritis	242/418 (58)	244/421 (58)	2/3 (0)	59/129 (46)	59/129 (46)	0/0 (0)
Rectal biopsy eosinophilic inflammation (count > 32 hpf)	210/367 (57)	212/369 (57)	2/2 (0)	64/116 (55)	64/116 (55)	0/0 (0)
Rectal biopsy surface villiform changes	135/364 (37)	136/366 (37)	1/2 (0)	31/115 (27)	31/115 (27)	0/0 (0)
Rectal biopsy basal plasmacytosis	176/336 (52)	178/338 (53)	2/2 (1)	51/111 (46)	51/111 (46)	0/0 (0)
Hemoglobin g/dL	N = 402	N = 405	3	N = 122	N = 122	0
Mean ± SD	11.4 ± 2.2	11.4 ± 2.2	0.0 ± 0.0	12.5 ± 1.8	12.5 ± 1.8	0.0 ± 0.0
% < 10 g/dL	98 (24)	98 (24)	0 (0)	8 (7)	8 (7)	0 (0)
% < 12 g/dL	231 (58)	233 (58)	2 (0)	46 (38)	46 (38)	0 (0)
Platelet count (x10 <sup>9</sup> /L)	N = 399	N = 402	3	N = 121	N = 121	0
Median (P25, P75)	372 (303, 462)	371.5 (303, 461)	0.5 (0, 1)	319 (251, 406)	319 (251, 406)	0 (0, 0)
> 500	77 (19)	77 (19)	0 (0)	10 (8)	10 (8)	0 (0)
White blood count (x10 <sup>9</sup> /L)	N = 397	N = 400	3	N = 119	N = 119	0
Median (P25, P75)	9.2 (7.2, 12.1)	9.2 (7.2, 12.0)	0 (0, 0.1)	7.8 (6.5, 9.3)	7.8 (6.5, 9.3)	0 (0, 0)
% > 12 x 10 <sup>9</sup> /L	100 (25)	100 (25)	0 (0)	10 (8)	10 (8)	0 (0)
ESR (mm/hr)	N = 385	N = 388	3	N = 118	N = 118	0
Median (P25, P75)	25 (12, 42)	24.5 (12, 42)	0.5 (0, 0)	15 (8, 24)	14.5 (8, 24)	0.5 (0, 0)
% ≤ 20 mm/hr	165 (43)	167 (43)	2 (0)	79 (67)	79 (67)	0 (0)
% > 40 mm/hr	102 (26)	102 (26)	0 (0)	8 (7)	8 (7)	0 (0)
CRP/hsCRP	N = 315	N = 317	2	N = 89	N = 89	0
% > ULN	144 (46)	144 (45)	0 (1)	18 (20)	18 (20)	0 (0)
% > 2x ULN	97 (31)	97 (31)	0 (0)	10 (11)	10 (11)	0 (0)



	Total Publication (n=428)	Total DSIC (n=431)	Diff. (n=3)	5-ASA Publication (n=136)	5-ASA DSIC (n=136)	Diff. (n=0)
Albumin (g/dL)	N = 422	N = 425	3	N = 133	N = 133	0
Mean ± SD	3.7 ± 0.7	3.7 ± 0.7	0.0 ± 0.0	4.0 ± 0.7	4.0 ± 0.7	0.0 ± 0.0
% < 3.5 g/dL	138 (33)	139 (33)	1 (0)	25 (19)	25 (19)	0 (0)
Fecal calprotectin (mcg/g)	N = 239	N = 240	1	N = 74	N = 74	0
Median (P25, P75)	2352 (1202, 3928)	2350 (1202, 3917)	2 (0, 11)	1692 (851, 3651)	1672 (851, 3631)	20 (0, 0)
% ≥ 250 mcg/g	226 (95)	227 (95)	1 (0)	64 (86)	64 (86)	0 (0)
Fecal OPG (pg/mL)	N = 178	N = 179	1	N = 53	N = 53	0
Median (P25, P75)	424 (31, 3259)	417 (31, 3259)	7 (0, 0)	119 (31, 1677)	119 (31, 1677)	0 (0, 0)
% > 1000 pg/mL	70 (39)	70 (39)	0 (0)	15 (28)	15 (28)	0 (0)
25-OH Vitamin D (ng/mL)	N = 393	N = 395	2	N = 125	N = 125	0
Median (P25, P75)	28.5 (23.9, 34.8)	28.5 (23.9, 34.8)	0 (0, 0)	28.1 (23.8, 34.0)	28.1 (23.8, 34.0)	0 (0, 0)
% < 20 ng/mL	42 (11)	42 (11)	0 (0)	14 (11)	14 (11)	0 (0)
ANCA	N = 397	N = 399	2	N = 125	N = 125	0
% Positive titer	259 (65)	260 (65)	1 (0)	73 (58)	73 (58)	0 (0)
% Titer ≥ 100 EU/ml	75 (19)	76 (19)	1 (0)	17 (14)	17 (14)	0 (0)
OmpC	N = 397	N = 399	2	N = 125	N = 125	0
Median (P25, P75)	7.4 (5.3, 11.2)	7.4 (5.3, 11.2)	0 (0, 0)	7.2 (5.1, 11.2)	7.2 (5.1, 11.2)	0 (0, 0)
% Positive titer	21 (5)	21 (5)	0 (0)	8 (6)	8 (6)	0 (0)
% ≥ 12	81 (20)	82 (21)	1 (1)	25 (20)	25 (20)	0 (0)
Cbir1	N = 397	N = 399	2	N = 125	N = 125	0
% Positive titer	76 (19)	77 (19)	1 (0)	27 (22)	27 (22)	0 (0)

**Table B2:** Comparison of values computed in integrity check to reference article Table 1 (Oral CS and IV CS)

	Oral CS Publication (n=144)	Oral CS DSIC (n=144)	Diff. (n=0)	IV CS Publication (n=148)	IV CS DSIC (n=148)	Diff. (n=0)
Age (mean ± SD)	12.5 ± 3.3	12.5 ± 3.3	0 ± 0	12.7 ± 3.2	12.7 ± 3.2	0 ± 0
% ≥ 12 years n (%)	96 (67)	96 (67)	0 (0)	101 (68)	101 (68)	0 (0)
Female (%)	61 (42)	61 (42)	0 (0)	82 (55)	82 (55)	0 (0)
Non-white (%)	18/140 (13)	18/140 (13)	0/0 (0)	30/146 (21)	30/146 (21)	0/0 (0)
Hispanic/Latino	9/143 (6)	9/143 (6)	0/0 (0)	11/147 (7)	11/147 (7)	0/0 (0)
Weight z-score (mean ± SD)	0.0 ± 1.2	0.0 ± 1.2	0.0 ± 0.0	-0.3 ± 1.2	-0.3 ± 1.2	0.0 ± 0.0
Height z-score (mean ± SD)	0.2 ± 0.9	0.2 ± 0.9	0.0 ± 0.0	0.0 ± 1.0	0.0 ± 1.0	0.0 ± 0.0
BMI z-score (mean ± SD)	-0.2 ± 1.3	-0.2 ± 1.3	0.0 ± 0.0	-0.4 ± 1.4	-0.4 ± 1.4	0.0 ± 0.0
Hospitalized at baseline (%)	14 (10)	14 (10)	0 (0)	148 (100)	148 (100)	0 (0)
Disease extent						
Proctosigmoiditis	4 (3)	4 (3)	0 (0)	3 (2)	3 (2)	0 (0)
Left-sided colitis	14 (10)	13 (10)	1 (0)	2 (1)	2 (1)	0 (0)
Extensive/Pancolitis/Unassessable	126 (88)	126 (88)	0 (0)	143 (97)	143 (97)	0 (0)
PUCAI (mean ± SD)	50.4 ± 13.8	50.4 ± 13.8	0.0 ± 0.0	66.9 ± 13.7	66.9 ± 13.7	0.0 ± 0.0
% 10-30 (Mild)	15 (10)	15 (10)	0 (0)	3 (2)	3 (2)	0 (0)
% 35-60 (Moderate)	96 (67)	96 (67)	0 (0)	39 (26)	39 (26)	0 (0)
% ≥ 65 (Severe)	33 (23)	33 (23)	0 (0)	106 (72)	106 (72)	0 (0)
Abdominal pain	122 (85)	122 (85)	0 (0)	135 (91)	135 (91)	0 (0)
Diarrhea	142 (99)	142 (99)	0 (0)	147 (99)	147 (99)	0 (0)
Rectal bleeding	136 (94)	136 (94)	0 (0)	145 (98)	145 (98)	0 (0)
Total Mayo score (mean ± SD)	7.9 ± 1.7	7.9 ± 1.7	0.0 ± 0.0	9.8 ± 1.7	9.8 ± 1.7	0.0 ± 0.0
% 0-6	23 (16)	23 (16)	0 (0)	3 (2)	3 (2)	0 (0)
% 7-9	93 (65)	93 (65)	0 (0)	55 (37)	55 (37)	0 (0)
% ≥ 10	28 (19)	28 (19)	0 (0)	90 (61)	90 (61)	0 (0)
Total Mayo score % ≥ 11	6 (4)	6 (4)	0 (0)	65 (44)	65 (44)	0 (0)

	Oral CS Publication (n=144)	Oral CS DSIC (n=144)	Diff. (n=0)	IV CS Publication (n=148)	IV CS DSIC (n=148)	Diff. (n=0)
Mayo endoscopy sub-score (mean ± SD)	2.2 ± 0.6	2.2 ± 0.6	0.0 ± 0.0	2.6 ± 0.5	2.6 ± 0.5	0.0 ± 0.0
% Mayo 1	13 (9)	13 (9)	0 (0)	3 (2)	3 (2)	0 (0)
% Mayo 2	88 (61)	88 (61)	0 (0)	57 (39)	57 (39)	0 (0)
% Mayo 3	43 (30)	43 (30)	0 (0)	88 (59)	88 (59)	0 (0)
Relative rectal sparing	15/144 (10)	15/144 (10)	0/0 (0)	6/147 (4)	6/147 (4)	0/0 (0)
Macroscopic patchiness	8/144 (6)	8/144 (6)	0/0 (0)	6/147 (4)	6/147 (4)	0/0 (0)
Cecal patch	10/136 (7)	10/136 (7)	0/0 (0)	2/126 (2)	2/126 (2)	0/0 (0)
Non-specific macroscopic gastritis	28/143 (20)	28/143 (20)	0/0 (0)	53/146 (36)	53/146 (36)	0/0 (0)
Microscopic gastritis	86/142 (61)	86/142 (61)	0/0 (0)	97/147 (66)	97/147 (66)	0/0 (0)
Rectal biopsy eosinophilic inflammation (count > 32 hpf)	82/125 (66)	82/125 (66)	0/0 (0)	64/126 (51)	64/126 (51)	0/0 (0)
Rectal biopsy surface villiform changes	47/125 (38)	47/125 (38)	0/0 (0)	57/124 (46)	57/124 (46)	0/0 (0)
Rectal biopsy basal plasmacytosis	65/118 (55)	65/118 (55)	0/0 (0)	60/107 (56)	60/107 (56)	0/0 (0)
Hemoglobin g/dL	N = 133	N = 133	0	N = 147	N = 147	0
Mean ± SD	11.5 ± 2.1	11.5 ± 2.1	0.0 ± 0.0	10.5 ± 2.2	10.5 ± 2.2	0.0 ± 0.0
% < 10 g/dL	28 (21)	28 (21)	0 (0)	62 (42)	62 (42)	0 (0)
% < 12 g/dL	77 (58)	77 (58)	0 (0)	108 (73)	108 (73)	0 (0)
Platelet count (x10 <sup>9</sup> /L)	N = 133	N = 133	0	N = 145	N = 145	0
Median (P25, P75)	376 (318, 464)	376 (318, 464)	0 (0, 0)	411 (339, 509)	411 (339, 509)	0 (0, 0)
> 500	28 (21)	28 (21)	0 (0)	39 (27)	39 (27)	0 (0)
White blood count (x10 <sup>9</sup> /L)	N = 133	N = 133	0	N = 145	N = 145	0
Median (P25, P75)	9.2 (7.0, 11.0)	9.2 (7.0, 11.1)	0 (0, 0.1)	11.3 (8.7, 14.9)	11.3 (8.7, 14.9)	0 (0, 0)
% > 12 x 10 <sup>9</sup> /L	23 (17)	23 (17)	0 (0)	67 (46)	67 (46)	0 (0)
ESR (mm/hr)	N = 125	N = 125	0	N = 142	N = 142	0
Median (P25, P75)	25 (12, 39)	25 (12, 39)	0 (0, 0)	38 (21, 57)	37.5 (21, 57)	0.5 (0, 0)
% ≤ 20 mm/hr	53 (42)	53 (42)	0 (0)	33 (23)	33 (23)	0 (0)
% > 40 mm/hr	29 (23)	29 (23)	0 (0)	65 (46)	65 (46)	0 (0)
CRP/hsCRP	N = 102	N = 102	0	N = 124	N = 124	0
% > ULN	44 (43)	44 (43)	0 (0)	82 (66)	82 (66)	0 (0)
% > 2x ULN	27 (26)	27 (26)	0 (0)	60 (48)	60 (48)	0 (0)

	Oral CS Publication (n=144)	Oral CS DSIC (n=144)	Diff. (n=0)	IV CS Publication (n=148)	IV CS DSIC (n=148)	Diff. (n=0)
Albumin (g/dL)	N = 142	N = 142	0	N = 147	N = 147	0
Mean ± SD	3.7 ± 0.6	3.7 ± 0.6	0.0 ± 0.0	3.4 ± 0.7	3.4 ± 0.7	0.0 ± 0.0
% < 3.5 g/dL	37 (26)	37 (26)	0 (0)	76 (52)	76 (52)	0 (0)
Fecal calprotectin (mcg/g)	N = 75	N = 75	0	N = 90	N = 90	0
Median (P25, P75)	2663 (1202, 3664)	2662 (1202, 3663)	1 (0, 1)	3202 (1495, 4384)	3202 (1495, 4384)	0 (0, 0)
% ≥ 250 mcg/g	73 (97)	73 (97)	0 (0)	89 (99)	89 (99)	0 (0)
Fecal OPG (pg/mL)	N = 59	N = 59	0	N = 66	N = 66	0
Median (P25, P75)	318 (31, 1638)	318 (31, 1638)	0 (0, 0)	1491 (169, 6830)	1491 (169, 6830)	0 (0, 0)
% > 1000 pg/mL	18 (31)	18 (31)	0 (0)	37 (56)	37 (56)	0 (0)
25-OH Vitamin D (ng/mL)	N = 132	N = 132	0	N = 136	N = 136	0
Median (P25, P75)	29.9 (25.2, 36.4)	29.9 (25.2, 36.4)	0 (0, 0)	27.4 (22.4, 33.0)	27.4 (22.4, 33.0)	0 (0, 0)
% < 20 ng/mL	10 (8)	10 (8)	0 (0)	18 (13)	18 (13)	0 (0)
ANCA	N = 135	N = 135	0	N = 137	N = 137	0
% Positive titer	90 (67)	90 (67)	0 (0)	96 (70)	96 (70)	0 (0)
% Titer ≥ 100 EU/ml	28 (21)	28 (21)	0 (0)	30 (22)	30 (22)	0 (0)
OmpC	N = 135	N = 135	0	N = 137	N = 137	0
Median (P25, P75)	7.3 (5.7, 10.2)	7.3 (5.7, 10.2)	0 (0, 0)	7.5 (5.3, 11.6)	7.5 (5.3, 11.6)	0 (0, 0)
% Positive titer	7 (5)	7 (5)	0 (0)	6 (4)	6 (4)	0 (0)
% ≥ 12	25 (19)	25 (19)	0 (0)	31 (23)	31 (23)	0 (0)
CBir1	N = 135	N = 135	0	N = 137	N = 137	0
% Positive titer	27 (20)	27 (20)	0 (0)	22 (16)	22 (16)	0 (0)

## Attachment A: SAS Code

```
libname dsic "X:\NIDDK\niddk-dr_studies6\PROTECT\private_created_data\PROTECT_V1\Data\Primary  
Manuscript Dataset";
```

```
/******/  
/* DSIC for PROTECT Study */  
/******/
```

```
*create temp dataset;  
data prim; set dsic.primary_manuscript_niddk_full;  
run;
```

```
/******/  
/* Age */  
/******/
```

```
*Total;  
proc means data=dsic.primary_manuscript_niddk_full n mean std;  
var ageyr;  
run;
```

```
proc freq data=dsic.primary_manuscript_niddk_full;  
tables age_ge12;  
run;
```

```
*5-ASA, Oral, IV CS;  
proc sort data=prim;  
by initial_trt_c4;  
run;
```

```
proc means data=prim n mean std;  
var ageyr;  
by initial_trt_c4;  
run;
```

```
proc freq data=prim;  
tables age_ge12*initial_trt_c4/norow nopercnt;  
run;
```

```
/******/  
/* Female */  
/******/
```

```
proc freq data=prim;  
tables female*initial_trt_c4/norow nopercnt;  
run;
```

```
/******/
```

```

/* Non-White */
/*****/

proc freq data=prim;
tables race_nonwhite*initial_trt_c4/norow;
run;

/*****/
/* Hispanic/Latino */
/*****/

proc freq data=prim;
tables hispanic*initial_trt_c4/norow ;
run;

/*****/
/* Weight z-score */
/*****/
*total;
proc means data=prim n mean std;
var bl_waz;
run;

*5-ASA, Oral, IV CS;
proc means data=prim n mean std;
var bl_waz;
by initial_trt_c4;
run;

/*****/
/* Height z-score */
/*****/
*total;
proc means data=prim n mean std;
var bl_haz;
run;

*5-ASA, oral, IV CS;
proc means data=prim n mean std;
var bl_haz;
by initial_trt_c4;
run;

/*****/
/* BMI z-score */
/*****/
*total;
proc means data=prim n mean std;
var bl_bmiz;

```

```

run;
*5-ASA, oral, IV cs;
proc means data=prim n mean std;
var bl_bmiz;
by initial_trt_c4;
run;

/*****/
/* Hospitalized at baseline */
/*****/
proc freq data=prim;
tables bl_hosp*initial_trt_c4/norow;
run;

/*****/
/* Disease Extent */
/*****/
proc freq data=prim;
tables paris_class*initial_trt_c4/norow ;
run;

/*****/
/* PUCAI */
/*****/
*total;
proc means data=prim n mean std;
var pucai;
run;

*5-ASA, oral, IV CS;
proc means data=prim n mean std;
var pucai;
by initial_trt_c4;
run;

*categorical PUCAI;
proc freq data=prim;
tables puc_c4*initial_trt_c4/norow;
run;

/*****/
/* Abdominal Pain */
/*****/
proc freq data=prim;
tables puc_abdominal_pain_C2*initial_trt_c4/norow;
run;

/*****/
/* Diarrhea */

```

```

/*****/
proc freq data=prim;
tables puc_diarrhea_c2*initial_trt_c4/norow;
run;

/*****/
/* Rectal Bleeding */
/*****/
proc freq data=prim;
tables puc_rectal_bleed_c2*initial_trt_c4/norow;
run;

/*****/
/* Toal Mayo Score */
/*****/
*total;
proc means data=prim n mean std;
var total_mayo_eef;
run;

*5-ASA, oral, IV CS;
proc means data=prim n mean std;
var total_mayo_eef;
by initial_trt_c4;
run;

*categorical Mayo Score;
proc freq data=prim;
tables total_mayo_c3*initial_trt_c4/norow;
run;

/*****/
/* Total Mayo score GE 11 */
/*****/
proc freq data=prim;
tables total_mayo_ge11*initial_trt_c4/norow;
run;

/*****/
/* Mayo endoscopy sub-score */
/*****/
*total;
proc means data=prim n mean std;
var bl_endo_mayo;
run;

proc means data=prim n mean std;
var bl_endo_mayo;
by initial_trt_c4;

```



```

run;

proc freq data=prim;
tables bl_endo_mayo*initial_trt_c4/norow;
run;

/*****/
/* Relative rectal sparing */
/*****/
proc freq data=prim;
tables rectal_inflam_c2*initial_trt_c4/norow;
run;

/*****/
/* macroscopic patchiness */
/*****/
proc freq data=prim;
tables macro_patch_c2*initial_trt_c4/norow;
run;

/*****/
/* Cecal Patch */
/*****/
proc freq data=prim;
tables peri_inflam_c2*initial_trt_c4/norow;
run;

/*****/
/* non-specific macroscopic gastritis */
/*****/
proc freq data=prim;
tables macro_gastritis*initial_trt_c4/norow;
run;

/*****/
/* microscopic gastritis */
/*****/
proc freq data=prim;
tables micro_gastritis*initial_trt_c4/norow;
run;

/*****/
/* Rectal biopsy eosinophilic inflammation */
/*****/
proc freq data=prim;
tables eosgrade_c2*initial_trt_c4/norow;
run;

/*****/

```

```

/* Rectal biopsy surface villiform changes */
/*****/
proc freq data=prim;
tables cbh_villiform*initial_trt_c4/norow;
run;

/*****/
/* Rectal biopsy basal plasmacytosis */
/*****/
proc freq data=prim;
tables cbh_bplasmac*initial_trt_c4/norow;
run;

/*****/
/* Hemoglobin g/dL */
/*****/
*total;
proc means data=prim n mean std;
var hb_cv;
run;

*5-ASA, oral, IV CS;
proc means data=prim n mean std;
var hb_cv;
by initial_trt_c4;
run;

proc freq data=prim;
tables (hb_c2 hb_lt12)*initial_trt_c4/norow;
run;

/*****/
/* Platelet Count */
/*****/
proc means data=prim n median q1 q3;
var plt_si;
run;

proc means data=prim n median q1 q3;
var plt_si;
by initial_trt_c4;
run;

proc freq data=prim;
tables plt_c2*initial_trt_c4/norow;
run;

/*****/
/* White Blood count */

```

```

/*****/
proc means data=prim n median q1 q3;
var wbc_cv;
run;

proc means data=prim n median q1 q3;
var wbc_cv;
by initial_trt_c4;
run;

proc freq data=prim;
tables wbc_c2*initial_trt_c4/norow;
run;

/*****/
/* ESR (mm/hr) */
/*****/
proc means data=prim n median q1 q3;
var esr_cv;
run;

proc means data=prim n median q1 q3;
var esr_cv;
by initial_trt_c4;
run;

proc freq data=prim;
tables esr_c3*initial_trt_c4/norow;
run;

/*****/
/* CRP/hsCrp */
/*****/
proc freq data=prim;
tables (crp_uln crp_2uln)*initial_trt_c4/norow;
run;

/*****/
/* Albumin (g/dL) */
/*****/
proc means data=prim n mean std;
var alb_cv;
run;

proc means data=prim n mean std;
var alb_cv;
by initial_trt_c4;
run;

```

```

proc freq data=prim;
tables alb_c2*initial_trt_c4/norow;
run;

/*****/
/* Fecal calprotectin */
/*****/
proc means data=prim n median q1 q3;
var calprotectin;
run;

proc means data=prim n median q1 q3;
var calprotectin;
by initial_trt_c4;
run;

proc freq data=prim;
tables calprotectin_c3*initial_trt_c4/norow;
run;

/*****/
/* Fecal OPG */
/*****/
proc means data=prim n median q1 q3;
var opg;
run;

proc means data=prim n median q1 q3;
var opg;
by initial_trt_c4;
run;

proc freq data=prim;
tables opg_c3*initial_trt_c4/norow;
run;

/*****/
/* 25-OH Vitamin D */
/*****/
proc means data=prim n median q1 q3;
var vitd;
run;

proc means data=prim n median q1 q3;
var vitd;
by initial_trt_c4;
run;

proc freq data=prim;

```

```

tables vitd_c3*initial_trt_c4/norow;
run;

/*****/
/* ANCA */
/*****/
proc freq data=prim;
tables (anca_pos anca_gt100)*initial_trt_c4/norow;
run;

/*****/
/* OmpC */
/*****/
proc means data=prim n median q1 q3;
var ompc;
run;

proc means data=prim n median q1 q3;
var ompc;
by initial_trt_c4;
run;

proc freq data=prim;
tables (ompc_pos ompc_ge12)*initial_trt_c4/norow;
run;

/*****/
/* Cbir1 */
/*****/
proc freq data=prim;
tables cbir1_pos*initial_trt_c4/norow;
run;

```